



POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Milan Court of Justice Immigration Section Processes Analysis

TESI DI LAUREA MAGISTRALE IN
MANAGEMENT ENGINEERING - INGEGNERIA GESTIONALE

Author: **Gabriele Liuzzo**

Student ID: 995956
Advisor: Professor Barbara Pernici
Academic Year: 2022-2023

Abstract

Immigration in Italy is a phenomenon of great social importance, the year 2023 registered one of the highest number of migrant arrivals on Italian territory ever, most of whom applied for international protection. The juridical process for obtaining it is articulated and requires the collaboration of several entities, among which the Immigration Section of the Courts stands out, it has the duty to ensure that applicants for international protection receive the right assessment. The objective of this thesis is to develop a methodology that allows for the systematic analysis of the performance of the justice delivered by the Milan Immigration Section, one of the 17 operating in Italy. Estimating the quality of justice service is complex, both because of the numerous elements that affect it and because of the difficulty in expressing it numerically. To achieve this purpose, in this thesis the data provided by the Court of Milan are examined through the use of process mining techniques: a set of methods belonging to the broader field of data science that allows for the systematic analysis of data generated by the computerized documentation of processes. These techniques make it possible to render schematically the progress of processes, validate their compliance with set standards, highlight critical issues and thus enable their improvement. Overall, it offers the possibility of monitoring the juridical load and its progress over time. Moreover, a machine learning model is also developed and discussed with the goal of predicting the total duration of the cases at their beginnings. In summary, this thesis aims to develop a standardized methodology to monitor the juridical load and performance of the Immigration Section in order to improve its efficiency.

Keywords: jurisdiction, Court, immigration, process mining, data analysis

Abstract in lingua italiana

L'immigrazione in Italia è un fenomeno di grande rilevanza sociale, nell'anno 2023 si è registrato uno dei più elevati numeri di arrivi di migranti sul suolo Italiano di sempre, la maggior parte dei quali ha presentato domanda di protezione internazionale. Il procedimento giuridico per il suo ottenimento è articolato e richiede la collaborazione di diverse entità, tra le quali spicca la Sezione Immigrazione dei Tribunali, l'ente che ha il compito di assicurare la giusta valutazione ai richiedenti di protezione internazionale. L'obiettivo di questa tesi è elaborare una metodologia che consenta l'analisi sistematica della performance dalla Sezione Immigrazione di Milano, una delle 17 operanti in Italia. Misurare la qualità del servizio di giustizia è complesso, sia per la pluralità di dimensioni che vi contribuiscono, sia per la difficoltà nella sua quantificazione. Per raggiungere questo scopo, in questa tesi i dati forniti dal Tribunale di Milano vengono esaminati mediante l'uso di tecniche di process mining: un insieme di metodi appartenenti al più vasto ambito di data science che permette di analizzare sistematicamente i dati generati dalla documentazione informatica dei processi. Queste tecniche permettono di rendere schematicamente l'andamento dei processi, di validarne la conformità rispetto a degli standard prefissati, di evidenziare le criticità e quindi permettere il miglioramento degli stessi. In complesso quindi offre la possibilità di monitorare il carico giudiziario e il suo svolgimento nel corso del tempo. Viene inoltre sviluppato e discusso un modello di apprendimento automatico con l'obiettivo di predire la durata della totale dei processi. In sintesi, questa tesi si propone di sviluppare una metodologia standardizzata che permetta il monitoraggio del carico giudiziario e delle performance della Sezione Immigrazione per poterne migliorare l'efficienza.

Parole chiave: giurisdizione, Tribunale, immigrazione, process mining, analisi dati

Contents

Abstract	i
Abstract in lingua italiana	iii
Contents	v
List of Figures	vii
List of Tables	ix
1 Introduction	1
1.1 Migration Phenomena	2
1.2 International Protection Appliance Iter	5
1.3 Thesis Objectives	9
1.4 Structure of the Thesis	9
2 Literature Review	11
2.1 Introduction to the Italian Juridical System	11
2.1.1 Digital Civil Trial	11
2.1.2 Italian Juridical System Terminology	13
2.1.3 A Process Timeline	17
2.2 Process Mining	17
2.2.1 Events Datalog	18
2.2.2 Process Mining Classes	19
2.2.3 Business Process Model and Notation	20
2.3 Analysis of Juridical Systems	21
3 Data Analysis	25
3.1 Data Science Methodology	25
3.1.1 Development Environment	25

3.1.2	Extraction from SICID	26
3.1.3	Dataset Selection Criteria	27
3.1.4	Objects File Adaption	29
3.1.5	Joining Data in a Comprehensive Data Structure	30
3.1.6	Chronological Ordering of Events	30
3.1.7	Initial Hearing Date Extraction	32
3.2	Analysis Objectives	33
3.3	Analysis One - Profiling of the Processes	33
3.3.1	Cases Characteristics	33
3.3.2	Judges Profiling	38
3.3.3	Cases Outcomes	40
3.3.4	Hearings Waitings Profiling	42
3.4	Analysis Two - Phases Analysis	45
3.4.1	Key Steps Isolation	45
3.4.2	Phases Duration	51
3.5	Analysis Three - Cases Duration Prediction	59
3.5.1	Factors Correlated with the Duration of the Processes	61
3.5.2	Machine Learning Models	69
4	Turin Immigration Section Comparison	73
5	Conclusions and Future Developments	77
	Bibliography	79
A	Appendix A	85
A.1	Processes schemes with average durations	85
A.2	Files columns description	86
A.3	All SICID columns description	88
A.4	Waitings distribution histograms	90
A.5	Judges Designations and Sentences	91

List of Figures

1.1	Number of new migrants in Italy by month [22]	4
1.2	Application requests in Europe increase by country of origin	4
1.3	Immigration Section processes objects	6
1.4	International protection appliance iter in Italy	8
2.1	Italian digital trial network architecture	12
2.2	Process mining as intersection between data science and process science [30]	18
2.3	Process mining as a data driven tool ¹	20
2.4	BPMN model example	21
3.1	Matlab user interface	26
3.2	First rows of the file "Stor.csv"	27
3.3	Family reunification object duplication	30
3.4	DataEV variable limited possibility to order events	31
3.5	New cases for every two weeks	34
3.6	Events per case distribution	35
3.7	Pending cases accumulation (also defined as backlog)	37
3.8	Rituality use over time	38
3.9	Designation of judges over time	39
3.10	Outcomes probability for each judge	40
3.11	Outcomes by date	41
3.12	Additional waiting over actual hearing date	44
3.13	Critical events isolation example	46
3.14	Temporal distribution of processes registrations	47
3.15	Time distribution of cases assignments for each judge	48
3.16	Total waiting for the hearing over hearing scheduling date	49
3.17	Temporal distributions of hearing dates.	50
3.18	Final judgments probability	51
3.19	Cases length over their beginning date	52
3.20	Number of started processes in blue of which completed in purple	53

3.21	Phases histograms. Each chart shows the duration of the phases over the number of cases	54
3.22	Total duration distribution	55
3.23	Immigration processes possible routes by states	56
3.24	First most frequent variant, 31.57% of all cases	57
3.25	Second most frequent variant, 10.88% of all cases	57
3.26	Third most frequent variant, 10.77% of all cases	57
3.27	Immigration processes possible routes by events	58
3.28	Decision tree illustrationan	59
3.29	Random forest composition	60
3.30	Case duration over starting date	61
3.31	Case duration over judge assignment delay	62
3.32	Case duration over case object	63
3.33	Cases duration over rituality	64
3.34	Cases duration over number of parallel ongoing processes	65
3.35	Cases duration over number of sub-processes	66
3.36	Cases duration over assigned judge	67
3.37	Case durations over judge assignment's clerk	68
3.38	Mean absolute error formula	69
3.39	Tree model graphical representation	70
3.40	Predicted duration over true duration. Random forest model.	71
4.1	Number of cases that reached a conclusion by semester	75
A.1	Apromore - Average label	85
A.2	SICID variables description Pt. 1	88
A.3	SICID variables description Pt. 2	89
A.4	Scheduled waiting distribution	90
A.5	Additional waiting days distribution	90
A.6	Total waiting days distribution	91
A.7	Number of cases for each judge and their outcomes	91

List of Tables

3.1	Number of cases by year	29
3.2	Objects and their frequency	36
3.3	Cases ritualities	37
3.4	Outcomes declinations by events	41
3.5	Processes phases	45
3.6	Median and average phases duration	54
3.7	Models performance parameters	71
4.1	Number of cases that reached a conclusion by semester	75

1 | Introduction

Immigration is a very relevant and current social phenomenon. It involves millions of people all over the world and poses challenges and opportunities for both the countries of origin and those of destination. Among the institutions responsible for managing and regulating migratory flows, the Courts play a fundamental role, called upon to rule on requests for asylum, citizenship, residence permits and expulsion of migrants.

The quality of juridical systems is a complex and multidimensional concept. Measuring it is therefore not a simple task. Some of the essential elements that are generally used to estimate the quality of justice are [14]: the accessibility of the justice system (eventually providing financial and linguistic support), the timeliness of the justice system referring to the reasonable duration of the processes, the fairness of the justice system meaning the level of impartiality and transparency of the juridical institutions, and the efficiency of the justice system which measures the use of the available resources, such as human, financial, material and technological.

The most problematic part of Italian Justice system is its timeliness. Italy suffers from some of the longest trial durations in Europe, especially in civil matters. In 2020 the average length of civil trials is 674 days for the first instance, more than twice the European average of 237 days. The second instance takes 1,026 days on average, almost six times longer than the European average of 177 days. The Court of Cassation requires 1,526 days on average, almost nine times longer than the European average of 172 days [15]. Reasonable durations of trials are important to respect the right to a fair trial and support the protection of the rights of citizens and businesses, which leads to greater trust in justice system. The duration of trials can vary based on the nature, complexity and type of cases, and can be influenced by external factors, such as legislation, organisation, resources and the process demand for justice.

The reduction of juridical times is an objective on which the juridical system has been strongly focused to improve in recent years. Among the objectives of the National Recovery and Resilience Plan there is also a 40% reduction by June 2026 in the duration of civil trials and a 25% reduction in the duration of criminal proceedings (measured on the

three levels of judgment and according to the disposition time indicator)[26].

The disposition time is one of the most used temporal indicators to measure trials durations. It is calculated as the number of resolved cases during a reporting period with the number of unresolved cases at the end of the same period. For the immigration processes, at the end of 2020 was around 10 times the maximum that the law permits (1,192 days against the maximum of 120 days law imposed) [7].

The disposition time is a convenient estimator, but it is based on the number of open cases and resolved case without measuring the effective length of trials, it is susceptible to the number of pending processes. Through the application of process mining, a family of methods that can be defined as "the intersection of data science and process science" [30], data provided by the Milan Court has been systematically analyzed measuring their effective durations and many other statistics.

This thesis aims to offer a methodology that can be used in the broader project in which is inserted ("Next Generation UPP: To improve the efficiency and performance of justice in the north-west of Italy¹) to analyze the Milan Court Immigration Section processes gaining useful insights that could help to improve their execution and management. This thesis offers also a comparison with the parallel work conducted by the Polytechnic of Turin and a discussion about the feasibility of a machine learning model with the aim to predict the cases duration a their beginnings.

The thesis is divided into four chapters. The first chapter provides a general overview of immigration in Italy and the role of the Immigration Section, as well as an introduction to the academic and legal context on which this thesis is based. The second chapter describes the research methodology and the analysis that have been performed. The third chapter presents a comparison with the work carried out by the Polytechnic of Turin. The fourth chapter presents the concluding remarks and possible future developments.

1.1. Migration Phenomena

Migration is the movement of people from one country or region to another, usually for economic, social, political, or environmental reasons. It can have positive and negative effects on both the host country and the country of origin. Some of the benefits of immigration are increased diversity, cultural exchange, innovation, and economic growth while the challenges include integration, discrimination, competition for resources, and social conflicts.

¹<https://www.nextgenerationupp.unito.it/>

There are many reasons that lead to migration [8]:

1. One of the primary motives behind migration is to find a better job prospects and better wages. Migration for economics motives can be voluntary, where individuals proactively are seeking out higher possibilities, and involuntary, where individuals are pressured to go away due to economic hardships.
2. Education is another significant factor that drives migration. Individuals often migrate in pursuit of better educational opportunities that are not available in their home countries. Access to quality education, specialized training, or higher education degrees can significantly impact an individual's future prospects and career opportunities.
3. Environmental factors, such as natural disasters, climate change, and environmental degradation can also be reasons behind migration. In such circumstances, people may be compelled to migrate to places that offer more sustainable living conditions and protect them from the adverse effects of environmental crises.
4. Family reunification is a powerful motivating factor for migration. Many individuals migrate to join their family members who are already living in another country.
5. Lastly, migration as a means to escape violence, conflict, and persecution. People who face persecution based on their ethnicity, religion, political beliefs, sexual orientation, or other aspects of their identity may seek refuge in countries where they can find safety, acceptance, and equal opportunities.

It is a currently intense phenomenon: data provided by the Department for Civil Liberties and Immigration, downloadable from the Ministry of Interior's website, show how the last ten years (and particularly the current year 2023) are among the most intense in terms of the number of immigrants received by Italy ever².

²<https://www.interno.gov.it/it/stampa-e-comunicazione/dati-e-statistiche>

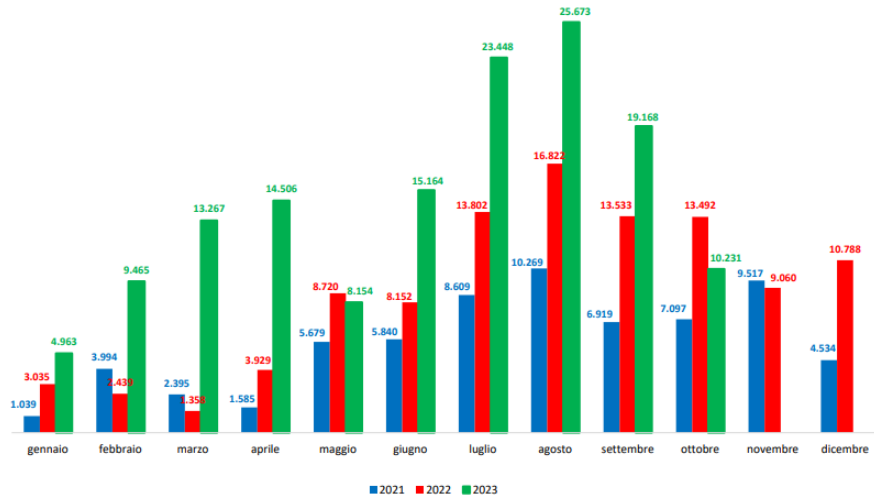


Figure 1.1: Number of new migrants in Italy by month [22]

The number of people requesting asylum in Europe increased subsequently, among them, the most prominent groups are those fleeing from the conflicts and crises in Syria and Turkey. Fig. 1.2 show the changes in the number of asylum applications by country of origin, showing the sharp increase of applicants from the different regions over one year time difference.

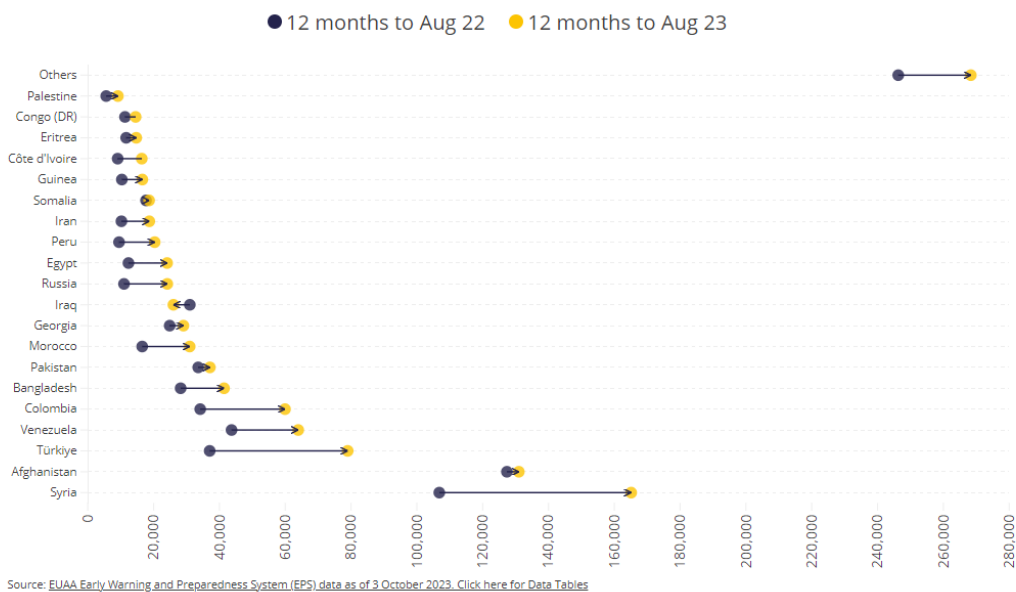


Figure 1.2: Application requests in Europe increase by country of origin

In 2020, four European countries accounted for the majority of asylum seekers in the

continent. Germany, France, Spain and Italy together received two thirds of all the asylum applications filed in Europe. The average acceptance rate, the percentage of applications that have been accepted, is 44%³.

1.2. International Protection Appliance Iter

The laws of EU countries set the economic autonomy of the immigrant as a necessary condition for having a residence permit and then citizenship. The immigrant is expelled if he or she does not demonstrate that has a regular job or someone who can provide him or her economic support, conditions to obtain a regular residence permit.

An exception to this principle concerns those who have a well-founded fear of being persecuted in their country of origin for reasons of race, religion, nationality, political opinion, membership of a particular social group and who cannot receive protection from their country of origin.

International law provides that in these cases the right to asylum, healthcare and first aid treatment is recognized. The law also lends itself to abuse, as it is difficult to prove the nationality of an illegal immigrant requesting political asylum, and whether he or she is actually entitled to it. The duty of discovering abuses lies with the States, expulsions must be justified on a case-by-case basis.

The processes of requesting international protection (asylum applications) account for more than 82% of the entire juridical workload faced by the Immigrant Section. Fig. 1.3 shows the number of processes (y axis) between the possible objects (the theme of the process), the highest bar corresponds to asylum requests. It is precisely this type of migrants whose process will now be illustrated, representing almost all of the workload dealt by the Immigration Section of the Court of Milan.

³<https://euaa.europa.eu/latest-asylum-trends-asylum>

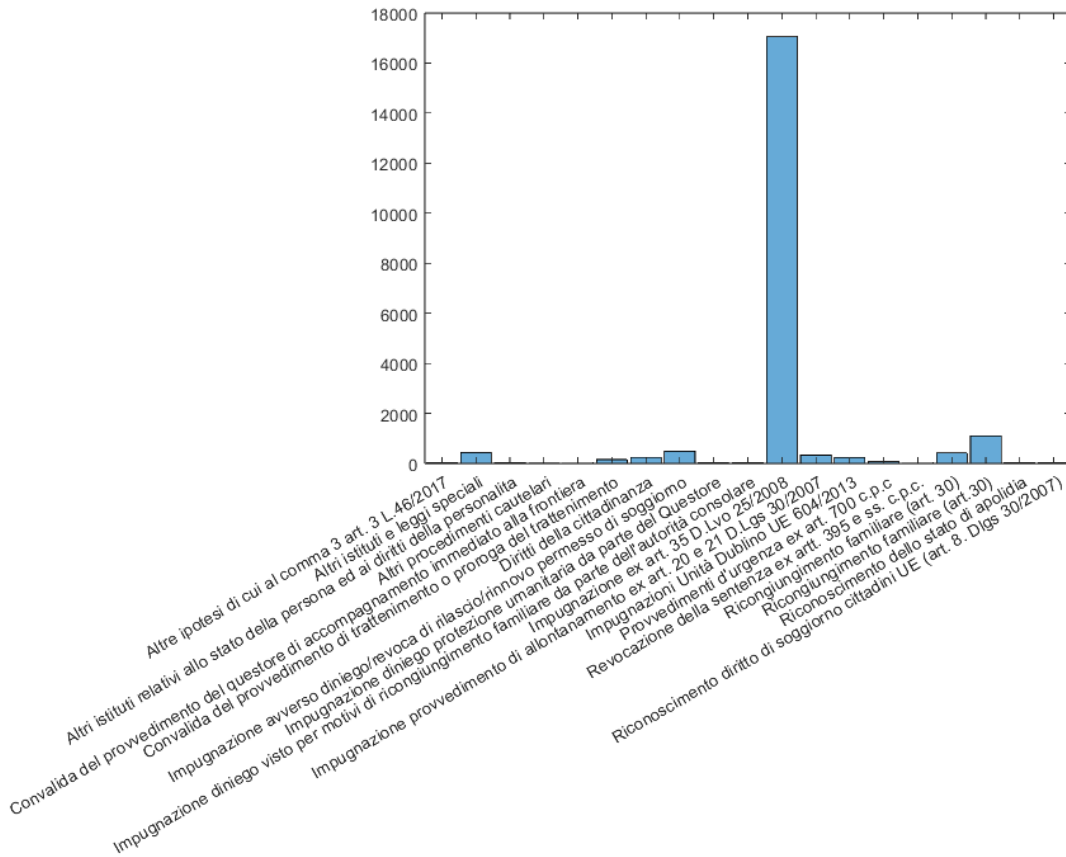


Figure 1.3: Immigration Section processes objects

First of all, the migrant must first express the desire to request international protection. It can be done at the Police Headquarters or the Border Police, as well as special spaces built in the most strategic places. The police will then record the migrant's data (especially the country of origin and family composition) to allow tracking and identification.

To carry out these procedures, the migrant can be held in a specific facilities for up to a maximum of one month. At the end of this period, if it has not yet been possible to ascertain the identity and country of origin, the migrant will be transferred to a Repatriation Permanence Centre, where he or she may be detained further for a maximum period of six months⁴.

The Dublin Regulation [16] establishes which country is responsible for examining an application for international protection based on multiple criteria: the presence of a family member in a country that applies the Regulation, the possession of an entry visa issued by one of the countries that apply the Regulation currently or in past, having crossed legally or illegally one of these countries. This last rule is often applied: the country responsible

⁴Interior Minister - Asylum Seekers in Italy guide

for evaluating the application for international protection is the first European country the migrant entered. There is also the possibility that the migrant decides to appeal to responsible country decision, in this case a case will start in the Immigration Section characterized by the object 'Appeals Dublin Unit UE 604/2013' which enjoy priority over other types of processes.

If according to the Dublin Regulation Italy has responsibility to evaluate the asylum request, the migrant must have an interview with a 'Territorial Commission for the Recognition of International Protection', institutions founded in 2002 [3]. Currently the Commissions are based in Milan, Gorizia, Rome, Foggia, Crotone, Syracuse, Trapani, Turin, Caserta and Bari, directed and coordinated by the National Commission. Each Commission is chaired by a prefect and is made up of an official serving at the Presidency of the Council of Ministers, an official from the diplomatic career, an official from the prefecture career serving at the Department for Civil Liberties and Immigration and a director of the Department of Public Safety. A representative of the United Nations High Commissioner for Refugees delegate in Italy participates in the meetings.

If the Commission does not grant international protection, the migrant has the right to appeal to the decision, effectively initiating a process in the Immigration Section of the appropriate territory. By law, the appeal must be notified, under penalty of inadmissibility, within one month from the outcome of the territorial commission decision⁵. Most of the processes of the Immigration Section in this thesis are therefore second grade processes since a first judgement has been already expressed.

In case the second international protection application (the appeal) is also rejected, the asylum seeker has the duty to repatriate. In 2022, 77,500 migrants were effectively repatriated, corresponding to 18.5% of all repatriations decisions issued during the year. The share of voluntary and forced returns was 35-65%, 75% of the total returns were assisted returns: the person received logistical and financial assistance⁶.

A graphical representation of all the possible routes that this iter is composed of are shown in Fig. 1.4. Specifically, the majority of the Milan Immigration Section processes origins from these steps: arrival in Italy, asylum request application, reject from the examination of the Territorial Institution, request of appeal.

⁵Link - Tribunale di Milano Protezione Internazionale e Immigrazione

⁶Link - European Commission overall figures of Immigrants in European Society

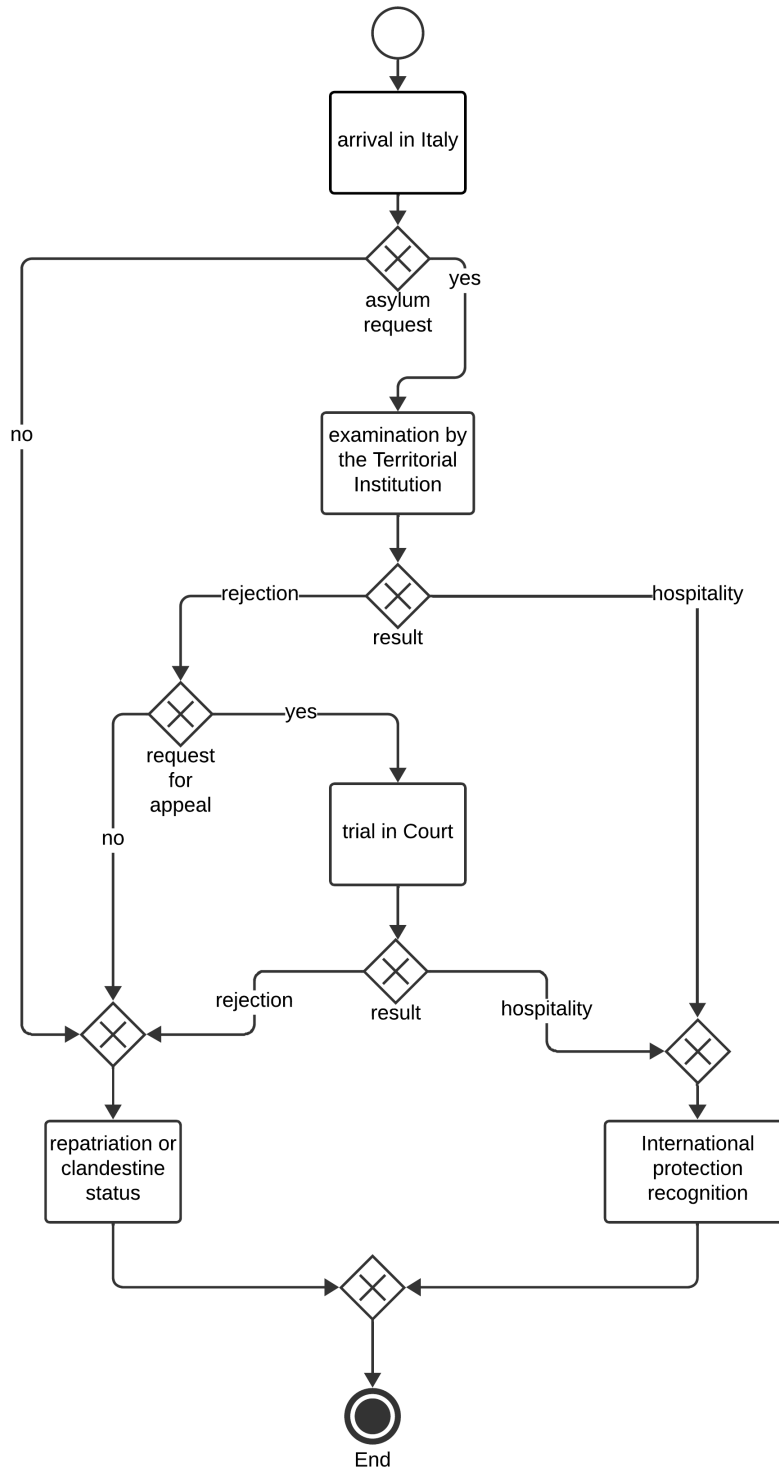


Figure 1.4: International protection appliance iter in Italy

1.3. Thesis Objectives

The objectives of this thesis are multiple: since this work is part of the broader project of the Italian Ministry of Justice: "Next Generation UPP: To improve the efficiency and performance of justice in the north-west of Italy"⁷, an objective is to progress in analysis of the entire justice system by developing working methods and documentation to help in the advancement of the broader objectives that the project aims for. Furthermore, this thesis aim to develop a systematic methodology that can be used to analyze the performance of the Immigration Section, offering the possibility to implement a monitoring tool and giving a clear picture of the workload that the Section faces. Moreover, machine learning models are also developed and discussed, with the aim of predicting the duration of processes at their beginning. A comparison of this thesis and the parallel work carried out by the Polytechnic of Turin is also provided.

1.4. Structure of the Thesis

Concluding the introduction with this section, the following part of the thesis is structured as follows: chapter two brings together the theoretical concepts necessary to contextualize the analyses, a section in particular is dedicated to the juridical related concepts this thesis requires. Chapter three presents a detailed analysis of the immigration processes at the Immigration Section of Milan that started from January 2 2018 to November 2 2022 and illustrates the methodology used for their execution, a section is dedicated to the discussion of machine learning models. Chapter four is a comparison between this thesis and a parallel work conducted by a team of the Polytechnic of Turin. Chapter five outlines the conclusions and possible future developments.

Furthermore, Appendix A contains additional documentation for a complete rendering of the work performed.

⁷<https://www.nextgenerationupp.unito.it/>

2 | Literature Review

This chapter aims to introduce the most important theoretical arguments and academic background used for the development of this work. Since the processes are of a juridical nature, an introduction to the legal theory used in this thesis is essential (Sec. 2.1), although the use of juridical terminology has been minimized, part of it is strictly necessary to contextualize the analyzes carried out. Sec. 2.2 aims to illustrate process mining, a branch of data science that deals with identifying, verifying and improving processes through data. Sec. 2.3 delves into the current state of academic articles published on the analysis of juridical systems with the aim of directing the work of this thesis in an optimal way from both an academic and pragmatic point of view.

2.1. Introduction to the Italian Juridical System

An overview of the Italian Juridical system is needed to contextualize processes analysis, this is especially true for the Immigration Section, which is a specialized department that needs specific knowledge. The following pages illustrate the main juridical concepts related to this thesis.

2.1.1. Digital Civil Trial

The electronic civil trial is a project of the Ministry of Justice, with the aim of improving the quality of the Italian juridical system, speeding up trials and reducing the costs of carrying them out [18]. It started in 1997 when the validity of a document through a digital signature was regulated, and its implementation occurred in several steps that led to a system of a rather complex architecture where individual parts were added progressively. Fig. 2.1 shows its composition in broad terms¹.

¹<https://www.tribunale.genova.it/pct.aspx?pnl=2>

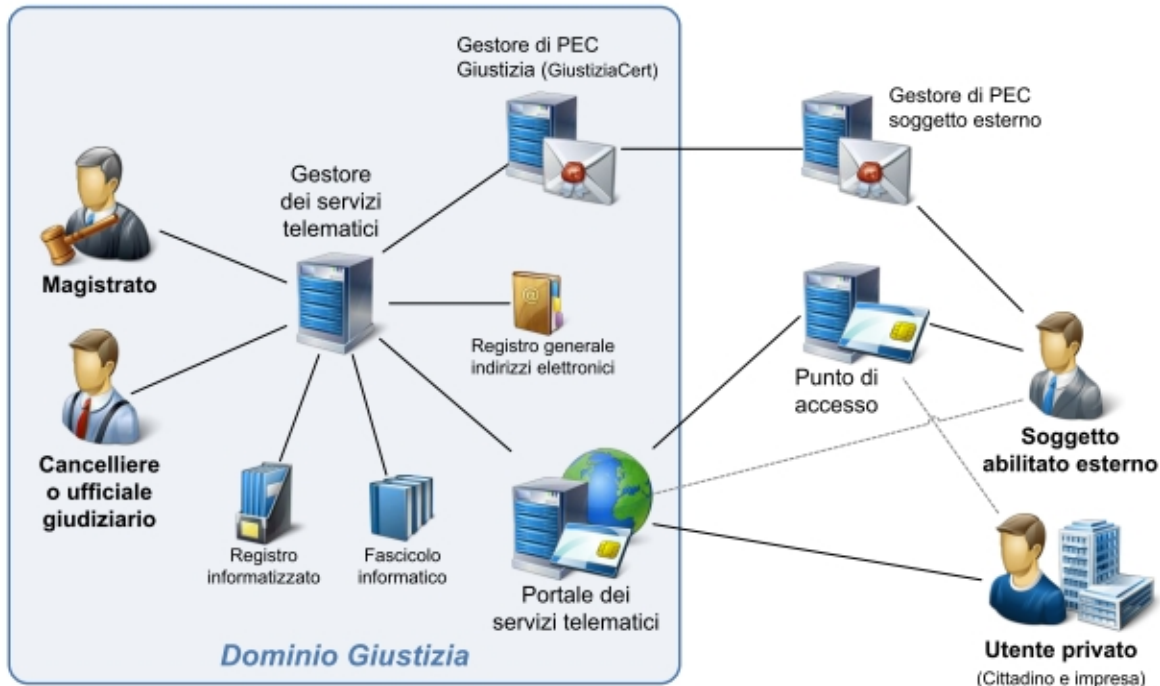


Figure 2.1: Italian digital trial network architecture

The architecture of the telematic process is mainly composed of the 'Justice Domain'. It is an IT network made up of hardware and software which has the main task of allowing the digitalisation of processes, their consultation and interaction by different actors [20]:

- Court clerks or advocates need to be able to work with processes, so they can keep track of what happened, store and examine juridical documents, and check the notifications that lawyers sent.
- Magistrates have a customized interface, which lets them see their pending cases and work with them in a suitable way.
- Lawyers must be able to send documents in a certified manner, by email or using the official online portal.
- External users such as companies, institutions or private individuals must be able to access certain data, so that they can use services such as consulting the status of processes.

With the Stability law of the year 2013 it is decreed that the deposition of many juridical documents is mandatory electronically since June 30 2014.

SICID stands for District Civil Litigation Information System (Sistema Informativo Con-

tenzioso Civile Distrettuale), it is the database management system that keeps track of the history of most of the Court's proceedings. Clerks and magistrates have direct access to it, being able to gradually add the events of each trial to the appropriate registers, possibly accompanied by electronically signed documents. It is from the SICID that the data on which this thesis is based were obtained, extracted through an SQL query by the competent office of the Court of Milan. The data set selected is constituted of 20,787 Immigration cases that started between January 2 2018 and November 2 2022.

States and Events

The Italian justice information system records the events that happen during the trials, creating a kind of 'diary' for each one. The SICID system provides the juridical system with 1,261 different possible events. Each event has a code and a description that explains their meaning. The history of the Immigration Section that will be used in the analyzes contains 264,633 events of 161 different distinct types. Some of the most important are for example: 'general register inscription' (20,787 events), 'monocratic hearing scheduling' (3,747 events), 'judge assignment' (21,206 events). The states of a process, on the other hand, are similar to the concept of an event, but with broader granularity. They keep track of the generic state that the process has reached.

The SICID imposes certain constraints regarding which events can be registered under which states, but it seems that these constraints are not always able to serve all possible cases, thus forcing chancellors to perform work-arounds that artificially change the state of the process in order to be able to insert the desired events. In the available data the events are cataloged in 43 different states. Some frequent examples are: 'assigned to a judge', 'waiting for hearing', 'waiting for collegial chamber sentence'.

2.1.2. Italian Juridical System Terminology

Collegiality and Monocraticy

The Monocratic Trial is characterized by the presence of a single magistrate with decision-making function. It contrasts with the collegial composition in which a plurality of judges participate.

Monocratic trials compared to collegial ones have the advantage of being generally faster and more efficient, since it avoids delays and conflicts that could arise from the deliberation and coordination of multiple judges. However, they have the disadvantage of being more prone to errors and biases, as they rely on the knowledge and judgment of a single judge,

who may not have all the relevant information or expertise or who may be influenced by personal or situational factors.

Process Object

Although the process object does not have an explicit definition in the civil procedure code, it is widely used and it is one of the fundamental parameter that the SICID uses to classify the cases. As a general meaning, the object of the trial (or object of the decision) can be defined as the content of the questions that the parties (actor, defendant or third parties) or the public minister ask the judge and on which with a suitable sentence is required². Specifically, the Immigration Section of the Milan Court deals with the following matters (in order of quantity of processes received):

1. 'Appeal pursuant to art. 35 Legislative Decree 25/2008': failure to recognize the right of residence on the national territory of citizens of other EU member states or their family members
2. 'Family reunification (art. 30)': regarding the denial of authorization for family reunification and residence permit for family reasons, as well as relating to other provisions of the administrative authority regarding the right to family unity
3. 'Appeal against denial/revocation of issue/renewal of residence permit': refers to the legal process by which an individual challenges a decision of the competent authorities regarding the issue, renewal or revocation of a residence permit in a country.
4. 'Other institutions and special laws': usually related to specific procedures or situations that require the intervention of a judge or a court.
5. 'Challenge of removal order pursuant to art. 20 and 21 Legislative Decree 30/2007': concerning the challenge of the expulsion order of citizens of other EU member states or their family members for reasons of public security.
6. 'Citizenship rights': concerns the set of rights and responsibilities that are associated with the status of citizen of a specific country or territory.
7. 'EU Dublin Unit Appeals 604/2013': appeals to the determination of which country should be responsible for examining the asylum request presented by a migrant.
8. 'Recognition of the state of a stateless person': for disputes regarding the assessment of the state of stateless person

²<https://www.brocardi.it/codice-di-procedura-civile/libro-secondo/titolo-i/capo-iii/art279.html>

9. 'Validation of the detention order or extension of detention': for proceedings for the validation of the provision with which the police commissioner orders the detention or extension of detention of the applicant for international protection

Types of Possible Granted Protections

The Italian state recognizes different types of protection³. The migrant can potentially be granted one of three main types of protection:

1. International Protection: can be granted to a person who requests asylum from a member state of the 1951 Geneva Convention. The convention defines who can benefit from it: "anyone, in the justified fear of being persecuted for his race, his religion, his citizenship, his membership of a particular social group or his political opinions, is outside the State of which he holds nationality and is unable or, due to such fear, does not wish to request the protection of that State" .
2. Subsidiary protection: it is another form of international protection, in this case provided for by the European Union and implemented by Italian law. A 2007 law defines the holder of subsidiary protection as a person in respect of whom there are well-founded reasons to believe that, if he or she returned to his country of origin, he or she would run a real risk of suffering serious harm.
3. Special protection: Introduced with the 2020 reform (law decree 130/2020), it establishes that: the rejection or expulsion of a person towards a State is also not permitted if there are well-founded reasons to believe that the removal from the territory national involves a violation of the right to respect for his private and family life, unless this is necessary for reasons.

These different types of protections are difficult to distinguish from a data perspective, as the events used to track them are different and overlapping without a clear distinction. For the purposes of this thesis they will be considered equivalent, except for the most granular analysis where events are displayed.

Definition of a case

The definition of a case is the achievement of a final judgement⁴. The completion of a process leads to its definition. A definitive outcome can take various forms, such as 'Accepted', 'Cancelled', 'Inadmissible' and 'Rejected'. These are the events that correspond

³<https://www.openpolis.it/parole/quali-sono-le-forme-di-protezione-per-gli-stranieri-in-italia/>

⁴<https://www.brocardi.it/codice-di-procedura-civile/libro-secondo/titolo-i/capo-iii/art279.html>

to a final judgement for the immigration related processes :

1. 'AT': Total acceptance
2. 'HF': partial (definitive) acceptance decree/order
3. 'DT': decree/order of total acceptance
4. 'DR': rejection decree/order
5. 'ES': Extinct process
6. 'IS': declared inadmissible
7. 'RPS': recognition of subsidiary protection
8. 'RPU': recognition of humanitarian protection
9. 'RSR': recognition of refugee status
10. '2T': rejection

These events have been selected through an interview to the Immigration Section clerk's office, and manual inspection of trials endings.

Rituality

In the juridical field, the rituality is the set of rules that regulates the conduct of a trial that the judge has the responsibility to account for. There are many ritualities that can be very complex to examine, they can change and depends on the nature of the processes. The immigration processes in the available database are classified by three main ritualities:

1. W1 - Voluntary First Degree Jurisdiction: indicates a type of process aimed not at resolving disputes but rather legal transactions, the conclusion of which requires the participatory intervention of a third-party judge.
2. IM - Chamber for Immigrants: it is a type of civil proceeding that does not take place in a public hearing, but in a reserved room of the Court, called the council chamber. They are not contentious in nature, that is, they do not concern a dispute between two or more parties, but deal with issues that concern the protection of rights.
3. 7O - Summary rituality of Cognition 702 bis cpc: it is a type of civil trial that takes place more quickly and simplified than the ordinary trial. (the Cartabia reform abolished this type of rituality on March 2023).

2.1.3. A Process Timeline

The typical progress of a process of the Immigration Section is presented here, conceived with the help of the Management Engineering Department of the Polytechnic of Milan. This procedure is valid specifically for the Immigration Section and more precisely for the Ex.Art appeal. 35 object (appeal to the denial of recognition of international protection), which represents by far the majority of the processes that the Section handles.

The one represented here is the ideal procedure while as the data shows there are actually more variations due to the individuality of the cases: If the Territorial Commission rejects the asylum request, the migrant has the right to appeal the decision before the Civil Court within 30 days. The migrant's lawyer who has expressed his desire to appeal the judgment of the local authority presents the appeal to the Court which enters the case in the general register. During the appeal, the migrant can remain in Italy.

The President of the Section assigns the case to a judge of the Section, according to criteria such as the number of cases already pending, the characteristics of the case, the judge's remaining time in the Section and the Section President's general experience. The competent judge has to set an hearing that is usually scheduled within three months (Sec. 3.3), although it can be delayed due to postponements. An interpreter may be requested.

Once the hearing took place, the decision-making phase begins. In this phase, the judge (or judges, in case of collegiality) sets aside some time to examine the case thoroughly and draft a sentence. The judge (or judges) has to review the evidence, the arguments, the legal precedents, and the applicable laws before reaching a conclusion and writing a sentence that explains the reasons and the outcome of the case.

The sentence is the final document that concludes the case and states the judge's decision. The sentence determines the future of the migrant who appealed the denial of their asylum request. If the appeal is accepted, the migrant is granted international protection, which comprehends the right to stay in the country and enjoy some benefits and services. Otherwise, if the appeal is rejected, the migrant is issued a repatriation order which means they have to leave the country.

2.2. Process Mining

The great amount of data made available by the Court of Milan has been analyzed with process mining techniques in order to approach it systematically. Process mining is the technique based on combining process optimization with data analysis, it was founded by computer scientist Wil Van Der Aalst who has published many works on the topic.

The paper [30] states that "process mining can be seen as the intersection of data science and process science" since combining process modeling and analysis with the event data that are available in the current information systems, new methods can be developed to address compliance checking and performance measurements.

Process mining allows organizations to take full advantage of the information stored in their systems and it also used to verify process compliance, detect bottlenecks and predict execution problems [28]. Process mining focuses operational processes, which are processes that involve repeated activities to produce or deliver products or services. These processes exist in all types of organizations and industries.

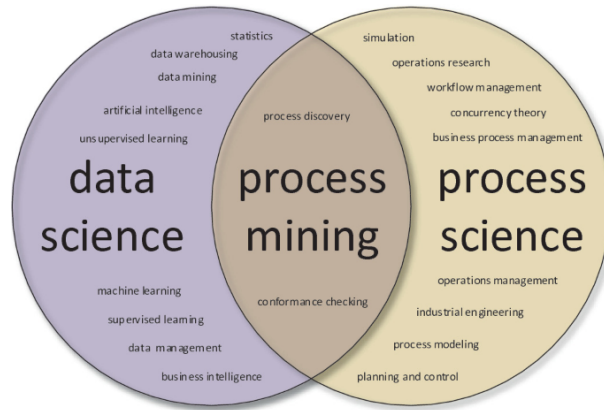


Figure 2.2: Process mining as intersection between data science and process science [30]

2.2.1. Events Datalog

Process mining methods are inherently linked to event data. Most process mining techniques require as input an event log, that typically refers to a chronological record of events, or activities, that have occurred in a system or process [17].

An event log can come from any data storage system that records an organization's activities along with timestamps of those activities, and can be represented in various formats, considering the information related to the various activities, that only in simplified representations can be considered of atomic nature and represented by an event ID alone. A task description (a textual description of the task performed) is also usually included, along with a timestamp of the task execution [29].

Starting from an event log, or a subset of it like a more specific workflow log, the process discovery proceeds towards its transformation into a process model. It must be noticed that, as in the case of most machine learning methods, the quality of the process model will depend on some measure on the quality of the event log it was derived from; particular

effort should be directed towards the identification of noisy entries and noisy patterns in the used event log in order to avoid further cascading negative effects on the model and its following uses [25].

2.2.2. Process Mining Classes

There are three main classes of process mining techniques. This classification is based on whether or not a prior model exists and how it is used. The three main classes are:

- **Process discovery:** it consists of transforming the event log into a process model, process discovery algorithms should be finite (terminates after a finite number of computational steps), definite (each computational step is unambiguous) and effective (each step can be performed correctly in a finite amount of time) and efficient (the fewer or faster computational steps, the better). However, the quality of such process discovery algorithms, as noticed by Professor Wil van der Aalst in [31], depends mostly on the input event log rather than the algorithm's internals.
- **Conformance checking:** methods that compare how a process is executed in reality with how it is supposed to be executed according to a model. Conformance checking methods can show how much they differ comparing a process model and an the behavior captured from an event log. These differences may be represented visually or textually as lists of natural language statements helping to discover inefficiencies highlighting them through data [24].
- **Performance analysis:** the goal is to enhance the process model with additional information about the performance of its activities, such as their duration, waiting time, and cost. The goal is not to check if the process follows the model, but to make the process better with respect to some parameters.

Process mining combines knowledge from computer science and management sciences to improve and manage operational processes. It compares event data and process models (manually created or automatically discovered) and aims to leverage event data in a meaningful way to provide insights, identify bottlenecks, anticipate problems, record policy violations, recommend countermeasures and optimize processes.

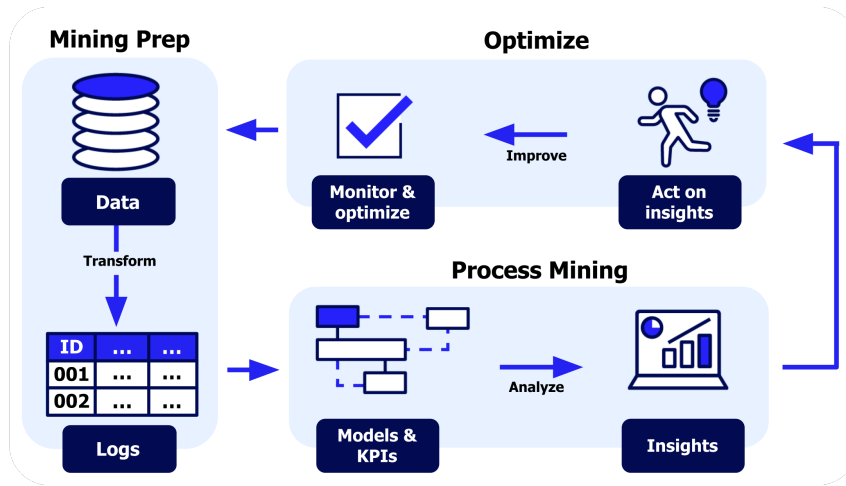


Figure 2.3: Process mining as a data driven tool¹

¹ <https://docs.appian.com/suite/help/23.3/pm-5.3/process-mining.html>

2.2.3. Business Process Model and Notation

A Business Process Model and Notation (BPMN) model is a graphical representation of a business process. It is a standard way to visualize the steps in a business process, as well as the flow of data and control between those steps. BPMN models can be used to document existing business processes, design new processes, or improve existing processes. A comprehensive study and evaluation of this model is presented in the paper [1]. BPMN models are used in a wide variety of industries, including healthcare, finance, manufacturing, and government.

BPMN models consist of four fundamental elements:

- **Flow objects:** represent the activities, events, and gateways that make up a business process. A task is a generic term for the work a company does and can be atomic or composite by sub tasks.
- **Connecting objects:** show how flow objects are connected to each other by mainly arrows that show the sequences of tasks and the possible paths a process can composed of.
- **Swim lanes:** swim lanes are a visual mechanism for graphical organization. They consist of boxes that divide tasks by roles or tasks categorization.
- **Artifacts:** represent what is used and produced by the business process linked to tasks.

BPMN models can be created using a variety of software tools, including both commercial

and open source options. Once a BPMN model is created, it can be used to communicate the business process to stakeholders, to analyze the process for potential improvements. In this thesis processes schemes are produced using the Apromore⁵ software, which offers an intuitive graphical interface and convenient automatically calculated statistics. It takes as input an event log where a timestamp column, a case ID column and an activity column have to be selected to render the data graphically.

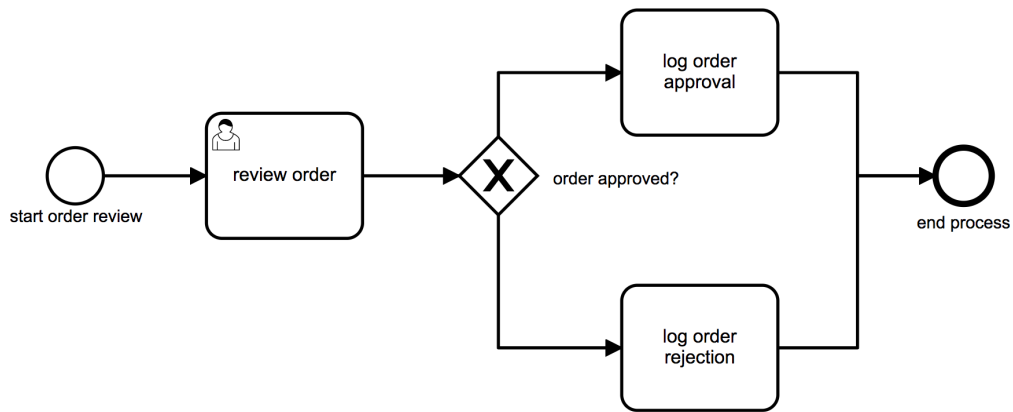


Figure 2.4: BPMN model example

2.3. Analysis of Juridical Systems

In the last two decades, there has been a growing interest in using event data to measure and improve the performance of business and institutional processes. By extracting and analyzing event data, organizations can identify the strengths and weaknesses of their processes, discover the root causes of problems, and implement solutions to enhance their quality and productivity.

Since this thesis is part of the larger project "Next Generation UPP: For the improvement of efficiency and of Justice Services in North-West Italy", the work already carried out has served as a solid starting point, part of which has not yet been published. The thesis [9] was useful as an introduction as it analyzed and measured the timing of the Appeal Court (the apparatus that deals with second degree appeals) proceedings and it also evaluates the median duration of the processes for each legal subject among other analysis.

The paper [27] introduces the utilization of process mining in an event log derived from a dataset encompassing business lawsuits from the Court of Justice of Sao Paulo, consid-

⁵<https://apomore.com/>

ered as one of the largest juridical department in the world. Its purpose is to conduct an analysis of juridical performance. The application of process mining analysis has facilitated the identification of the most recurrent activities and process bottlenecks. These insights, in turn, permitted to identify the underlying causes of inefficiencies. [10] and [11] are works that concern the definition of terminologies for the abstraction of a business process model.

The work carried out in [12] is very relevant: in collaboration with the Court of Livorno which provided data regarding the processes, the team, in addition to studying the legal constraints, created performance indicators (KPIs) capable of real time (week by week) to give a measure of the performance of various key actors (Clerks and Judges) responsible for the timing of the trials themselves. [13] develops an agent-based simulation prototype starting from legal process data, measuring and applying the KPIs developed by [12]. The Immigration Section, being a specialized Section, has very specific and change-prone constraints.

[19] suggests a process mining approach that aims to break down the waiting time observed in each activity transition within a process into various direct causes. It also seeks to examine the influence of each identified cause on the overall efficiency of the process.

Focusing on the prediction of the duration of processes, several works can be found with different approaches developed especially in recent years. In [33] using regression techniques, the time needed to close trials registered in Texas Courts over a considerable period of 17 years was estimated. An inverse correlation between Court congestion level and trial duration is detected.

In [32], trials in three different legal districts in Oregon were analyzed with the aim of comparing their performances according to the type of organization present within them, suggesting that greater autonomy and independence of the staff within the Court is correlated with greater speed in resolving disputes. Process analyzes have also been approached with different methods including text mining synergistically with the schematization of processes in [4].

In 2014, a study [21] was also developed which aims to detect how much the development of institutions in a certain territory attracts professionally qualified migration. A subdivision into different indices is used to measure the quality of institutional structures and therefore measure their impact on qualified migration.

This thesis finds space for work consisting in an in-depth analysis of the specialized Immigration Section which is characterized by processes of a very particular nature compared

to the processes of the rest of the Court from a legal point of view. Moreover it aims to develop a systematic methodology to analyze the processes which are broken down according to a scheme outlined from the study of the subject of immigration itself. Statistics regarding judges, hearings and outcomes are also calculated on events data, the most detailed information available in the system. Predictive machine learning models are also applied to the Immigration Section data, evaluating the importance of nine variables and therefore estimating the feasibility of a model that can anticipate the duration of the cases at their beginning.

3 | Data Analysis

Based on the data provided by the Milan Court, process mining analysis were carried out in conjunction with a necessary study of the legal context to which they belong. The objective is to investigate the performance and dimensions of the workload that the Immigration Section faces, as well as to outline the temporal duration of the processes in their totality and in their sub-phases.

These analysis are carried out using different software, the most used are Matlab and Apomore¹. Matlab, briefly shown in Sec. 3.1.1, is mainly used for data manipulation, which involves operations such as filtering, joining, and new variables computation. On the other hand, Apomore is used to create diagrams that help to graphically visualize the processes phase. It also allows the application of additional filters to data based on different criteria.

3.1. Data Science Methodology

The following sections present the methodology with which the data were processed before the analysis to allow the results to be contextualized in the most clear and accurate way possible. The sections are written, after a brief illustration of the software used, in the order in which the data were actually transformed.

3.1.1. Development Environment

The software used for data transformation is called Matlab², an acronym for MATrix LABoratory, created in 1970 as software designed for matrix analyses, distributed since 1984 by Mathworks. It is a platform designed for engineers and scientists that integrates computation, visualization and programming in an approachable environment.

The interface of MATLAB is composed of several panels that allow the user to perform different tasks. The three most important ones are shown in Fig. 3.1 and have the

¹<https://apomore.com>

²<https://it.mathworks.com/products/matlab.html>

following functions:

- The left panel is useful for exploring the data in use by opening them in calculation tables. This way, the user can inspect the values and properties of the variables and matrices that are involved in the computation.
- The middle panel allows the writing of algorithms, eventually divided in sections. The user can create, edit, and save scripts and functions that implement the desired logic and operations. The sections can be executed separately or all together, depending on the user's needs.
- The panel on the right is used for the execution of individual instructions useful for temporary tests preventing them from affecting the actual code. The user can type instructions in the command window and see the results there.

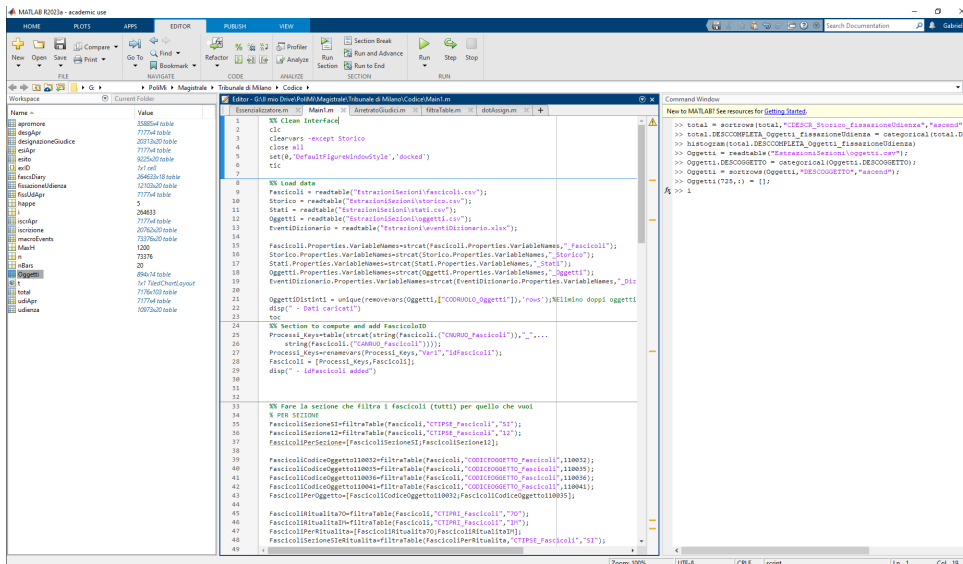


Figure 3.1: Matlab user interface

Matlab also provides several toolboxes: collections of specific functions for specific application areas. In Sec. 3.5 the 'Machine Learning' toolbox developed by the Mathworks company is used. It provides a graphical interface useful for inspecting the predictors and the accuracy of the built models.

3.1.2. Extraction from SICID

In order to obtain the data it was necessary to get in touch with the Milan Court's technical department which, through an SQL query, extracted all the data regarding the processes which started from January 1 2018 to November 2 2022: a three years and 10

months time window.

The result of the extraction has been stored as six CSV (Comma Separated Values) format files. 'Stor.csv' of 1.2GB size containing approximately five and a half million events referring to 444,861 processes whose parameters are contained by 'Fasc.csv'. Only a fraction of this data has an high information value (at least as far as the Immigration context is concerned).

These files appear as tables of different format based on what information of the processes the table keep record of. The 'fasc' part contains a list of all processes and their characteristics such as the starting date, the rituality under which the process has been carried out, a code that indicates which Section is its responsible. The 'stor' part instead records the events history for each process, a sort of 'diary' where every entry (row) corresponds to something that concerns a process. Difficulty in deducing the usefulness of the individual columns also lies in the ability to give meaning to their content: their titles are rather cryptic especially for those who have never studied jurisprudence.

Fig. 3.2 shows the first rows of 'stor'. A more detailed description of the variables can be found in Appendix A.2.

	1 Var1	2 NUMPRO	3 NUMPRV	4 CCDOEV	5 CTIPSE	6 NUMGIU	7 DATEV	8 CCODST	9 CNPARA	10 CDESCR	11 DATARE
1	0	1954485	21140067	'IA'		NaN	2017-01-10 00:00:00	'AS'		NaN 'ISCRITTO A...	2017-01-10 13:29:47
2	1	1954110	21133823	'IA'		NaN	2017-01-09 00:00:00	'AS'		NaN 'ISCRITTO A...	2017-01-09 11:29:22
3	2	1954650	21142815	'IA'		NaN	2017-01-11 00:00:00	'AS'		NaN 'ISCRITTO A...	2017-01-11 10:25:15
4	3	1954076	21133415	'IA'		NaN	2017-01-09 00:00:00	'AS'		NaN 'ISCRITTO A...	2017-01-09 10:52:32
5	4	1953593	21127515	'IA'		NaN	2017-01-04 00:00:00	'AS'		NaN 'ISCRITTO A...	2017-01-04 10:14:49
6	5	1954829	21145624	'IA'		NaN	2017-01-11 00:00:00	'AS'		NaN 'ISCRITTO A...	2017-01-11 15:56:04
7	6	1954838	21145720	'IA'		NaN	2017-01-11 00:00:00	'AS'		NaN 'ISCRITTO A...	2017-01-11 16:21:13
8	7	1954258	21137752	'IA'		NaN	2017-01-10 00:00:00	'AS'		NaN 'ISCRITTO A...	2017-01-10 09:43:22
9	8	1954698	21143501	'IA'		NaN	2017-01-11 00:00:00	'AS'		NaN 'ISCRITTO A...	2017-01-11 11:26:27
10	9	1954822	21145598	'IA'		NaN	2017-01-11 00:00:00	'AS'		NaN 'ISCRITTO A...	2017-01-11 15:49:55
11	10	1953817	21129953	'IA'		NaN	2017-01-05 00:00:00	'AS'		NaN 'ISCRITTO A...	2017-01-05 10:01:20
12	11	1954031	21133003	'IA'		NaN	2017-01-09 00:00:00	'AS'		NaN 'ISCRITTO A...	2017-01-09 10:17:19
13	12	1953546	21127275	'IA'		NaN	2017-01-04 00:00:00	'AS'		NaN 'ISCRITTO A...	2017-01-04 09:20:59
14	13	1954799	21145312	'IA'		NaN	2017-01-11 00:00:00	'AS'		NaN 'ISCRITTO A...	2017-01-11 15:05:13
15	14	1953891	21130628	'IA'		NaN	2017-01-05 00:00:00	'AS'		NaN 'ISCRITTO A...	2017-01-05 11:45:42
16	15	1954770	21144549	'IA'		NaN	2017-01-11 00:00:00	'AS'		NaN 'ISCRITTO A...	2017-01-11 12:49:41
17	16	1954832	21145635	'IA'		NaN	2017-01-11 00:00:00	'AS'		NaN 'ISCRITTO A...	2017-01-11 15:58:47

Figure 3.2: First rows of the file "Stor.csv"

3.1.3. Dataset Selection Criteria

From the extraction carried out by the Milan Court office, a database of data concerning all the trials was obtained. Since this work focuses in particular on the theme of immigration, a selection has been made to select the processes that concern it. As anticipated in 2.1, the processes are divided through different parameters which permitted to identify them.

Given the opportunity to directly observe the number of processes that can be viewed from the interface available to the Immigration Section Clerk's office, the numbers displayed has functioned as a safe reference for the immigration cases identification. They were also verified with the numbers published by the Superior Council of the Magistracy which, in a report published on 13 October 2021, reports an average of 5,005 annual cases arising in the three-year period 2018-2020 requesting international protection to the Milan Immigration Section [7].

To identify the correct processes, they have been filtered through the CTIPSE column, which reports the Section in which the last event was recorded for each process. Processes could also have been filtered through other criteria, including the IM rituality (Immigration rituality) or the most frequent object with code 110032, corresponding to the asylum requests processes. However, it was preferred to construct the data set as the totality of the processes faced by the Immigration Section for three main reasons:

- It is the only set of processes of which it is possible to verify with certainty the numerical correctness thanks to the possibility of comparing them with the numbers found by the Chancellors Office.
- The Immigration Section is a stand-alone and complete body. Studying all the processes that concern it means being able to offer fair statistics for the judges who belong to it.
- Selecting only a part of these processes would mean omitting some that may have a strong impact (for example those characterized by the Dublin Unit object) on the other cases in the same Section. Having as objective in the third analysis a machine learning model, the disposal of the entire context is very useful to be able to capture the dynamics within it.

The processes selected through the Section criteria are 24,148. An additional column was created consisting of the concatenation of the two attributes "CNURUO" and "CANRUO" to uniquely identify them. The processes are distributed as reported by Tab. 3.1.

Chancellery Office	Database (ctipse='si')	Year
3,168	3,175	2018
9,393	9,413	2019
3,208	3,200	2020
2,831	2,826	2021
3,918	3,141	2022

Table 3.1: Number of cases by year

The slight discrepancy between the number of cases detectable by the registry and those found in the database is attributable to the addition or elimination of the same in the period between the saving of the database and the detection which took place four months later. The largest discrepancy, i.e. that of 2022, it is explained by the fact that the available database does not cover all the 2022 year but up to November 2nd.

While a complete profiling of these processes is carried out in the first analysis of this section 3.3, a brief description is reported here: 20,762 cases have been isolated covering a time span of three years and 10 months, starting from 2 January 2018 until 2 November 2022. These trials were assigned to 40 distinct judge identification codes, more than 75% of the trials were assigned to 10 judges. 9,226 cases have reached completion, 1,749 cases are waiting for a sentence, 1,310 cases are waiting for an hearing, 8,208 cases have been assigned a judge but don't have e a scheduled hearing.

3.1.4. Objects File Adaption

The object of the process (or object of the decision) can be defined as the category in which the content of the process that the parties addressed to the judge. The file that contains the code and description of the juridical objects is of fundamental importance in order to catalog the processes. In order to be able to use it, however, it had to be transformed with two main operations:

1. Each juridical object has a unique code that identifies it. It is easily found in the process file (codiceoggetto column), but not in the file where the objects themselves are described. In the Objects file of the SICID system this code is in fact deconstructed into three columns: codruolo, codmateria, codoggetto. To be able to merge the two files, an additional column was created corresponding to their concatenation so that the object file also has the full object identifier.

2. In the Objects file, the entry 'Family reunification', which in our database corresponds to 354 processes has a duplicate differentiated only for an additional space character (Fig. 3.3). This leads to the duplication of the processes of this object when they are subjected to a Join operation. One of the two duplicates has been eliminated.

5
DESCCOMPLETA
Riconoscimento dello stato di apolidia
Ricongiungimento familiare (art.30)
Ricongiungimento familiare (art. 30)
Richiesta documentazione per la successione aperta in Italia in titoli del debito pubbl...
Richiesta documentazione per la successione aperta all'estero in titoli del debito pub...
Ricevimento e verbalizzazione di dichiarazione giurata

Figure 3.3: Family reunification object duplication

3.1.5. Joining Data in a Comprehensive Data Structure

Once the files have been properly transformed, they have been merged to a single one that contains all the data of all the processes that have been selected.

Through a Join operation (precisely an inner join: the union of the entries that exclusively have a reference in both the joined parts) the file of the processes filtered for Immigration Section and the history file was unified through the cases identification key.

Subsequently, the information contained in the other files was added via supplementary Innerjoins:

1. the 'States' file was merged using the state identifier code ('CCODST' variable).
2. the 'Objects' file through the ObjectCode column specially created by concatenating Codoggetto, Codruolo and Codmateria.
3. the 'Dictionary' file containing a description of all the events was added through the CCDOEV (event code column) of the history file. It should be noted that the 'Processes' file, which has a column of the same name but with different use.

The result is a single table made up of 264,633 rows (each corresponding to an event of a trial) and 92 columns that divide 24 million cells containing the total amount of data that the Immigration Section has registered in the Court's information system.

3.1.6. Chronological Ordering of Events

Arranging events chronologically is not as simple as it may initially seem: although having at disposal both the date of the event and the date on which this event is recorded, this

is not enough to trace the temporal order in which they actually occurred.

Without the advantage of any indications from the chancellery which has a graphical interface that automatically puts the events in chronological order, the numerous columns at our disposal have been empirically investigated: The variable DATAEV (event date) contained in the History file lacks the time stamp, therefore the ordering of the events that occurred on the same day is impossible to reconstruct. An example of this case is shown in Fig. 3.4 where the inscription on the general register appears to happen after the assignment to the Section, an impossible case as the registration on the general register is necessarily the first step.

1 idFascicoli	2 NUMPRV_Storico	3 DATAEV_Storico	4 CDESCR_Dizionario
'34065_2021'	26513214	2021-08-11 00:00:00	'ASSEGNAZIONE A SEZIONE'
'34065_2021'	26513208	2021-08-11 00:00:00	'ISCRIZIONE RUOLO GENERALE'
'34065_2021'	26521547	2021-08-17 00:00:00	'DESIGNAZIONE GIUDICE'
'34065_2021'	26606529	2021-09-20 00:00:00	'FISSAZIONE UDIENZA CAMERALE'

Figure 3.4: DataEV variable limited possibility to order events

The event recording date, being an automatic timestamp created by the informatic system has also the exact time (even including seconds) in which the event was recorded. However, it has been discovered that the registration date can be very different from the event date (and not in actual order of succession). The 'Var1' variable was also investigated, it is the only one that does not have an acronym of a legal term as a name. It contains progressive numbering, which however does not appear to be strictly related to the chronological order of events. It was useful to check that no event row was duplicated after the join operations.

The variable finally used is the NUMPRV variable: the name of the variable probably corresponds to 'Provision Number'. It has been used as a secondary variable to DATAEV: sorting as a priority DATAEV and then if multiple events of the same process occur on the same day using NUMPRV to extract the correct order. With this method the order of events is coherent and consistent.

The result is therefore a single code instruction, which first prioritizes the sorting by cases (through the case ID key) and then DATAEV and NUMPRV of the History file (two homonym ones are present in the Processes file which however must be ignored as they are only the last DATAEV and NUMPRV registered for that process).

3.1.7. Initial Hearing Date Extraction

To perform the analyzes that measure the delays of the hearings held in 3.3.4, it was necessary to isolate the dates set initially for the hearings.

For the date on which a hearing was actually held, the date on which the process begins the decision-making phase was used (events such as 'Reserve' or 'Remission to the collegial chamber'). To isolate the date on which the hearing was originally set, it was necessary to deal with the CDESCR column from the History file which contains personalized text for each event.

The CDESCR column for the event 'scheduled hearing' has three possible contents:

- 'MONOCRATIC INQUIRY HEARING SET FOR dd/MM/yyyy HH:mm'
- 'CHAMBER HEARING SCHEDULED ON dd/MM/yyyy HH:mm'
- 'APPEARANCE HEARING SCHEDULED ON dd/MM/yyyy HH:mm'

where dd/MM/yyyy HH:mm is the date on which the hearing is originally scheduled and which therefore allows a comparison with the date on which it is actually held. The difference between the due are due to postponements.

Code has been written that could isolate the date that extract from the text the characters between the last letter and the sixteenth letter preceding the last one. Finally, the string data type was converted to date type.

3.2. Analysis Objectives

To fully detect the performance of the Immigration Sections, three analysis have been executed with three different objectives:

- The first, 3.3 is a general profiling of the trials, judges, outcomes and hearings delays.
- The second, 3.4 is an in-depth analysis of the duration of the individual phases that compose the processes, as well as an illustration of the data that marks their succession.
- The third, 3.5 analyzes the factors that affect the duration of the processes and machine learning models are developed to predict the processes duration at their start.

3.3. Analysis One - Profiling of the Processes

The first analysis that is performed is a general profiling and statistics of the selected processes that are registered in the Immigration Section of the Milan Court. The purpose of this analysis is to provide a clear picture of the workload. Developing a methodology to profile processes characteristics, judges performance and hearing postponements is useful to gain an overview on how the entire Immigration Section works.

3.3.1. Cases Characteristics

The processes of the Immigration Section analyzed are 20,762, their beginning have been heavily concentrated in years 2018 and 2019 (13,178 cases, the 63% of all processes) due to a peak of new migrants in 2016 and 2017. In the next years (2020-2021-2022) it is reported a low amount of new arrivals characterized by a steady arising trend (Fig. 3.5). This trend has been confirmed by the Chancellors Office that reports an exponential increase of new cases after the period covered by the database available to this thesis.

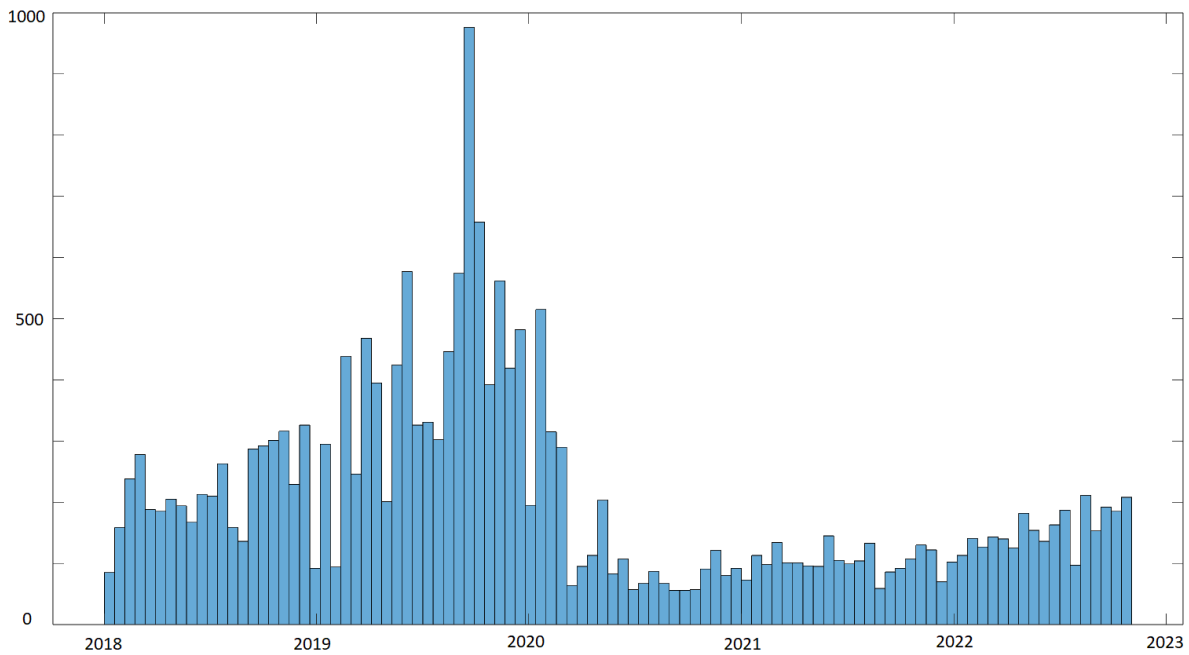


Figure 3.5: New cases for every two weeks

The dataset is composed by 264,633 events that divided by the 20,762 processes means around 13 events per process as average. As shown in Fig. 3.6, their distribution, however, is divided mainly into two parts: a first peak composed of the cases that have not exceeded three events (where usually the third event is the designation of the judge) and another second peak composed of the cases that actually progressed after the designation of the judge.

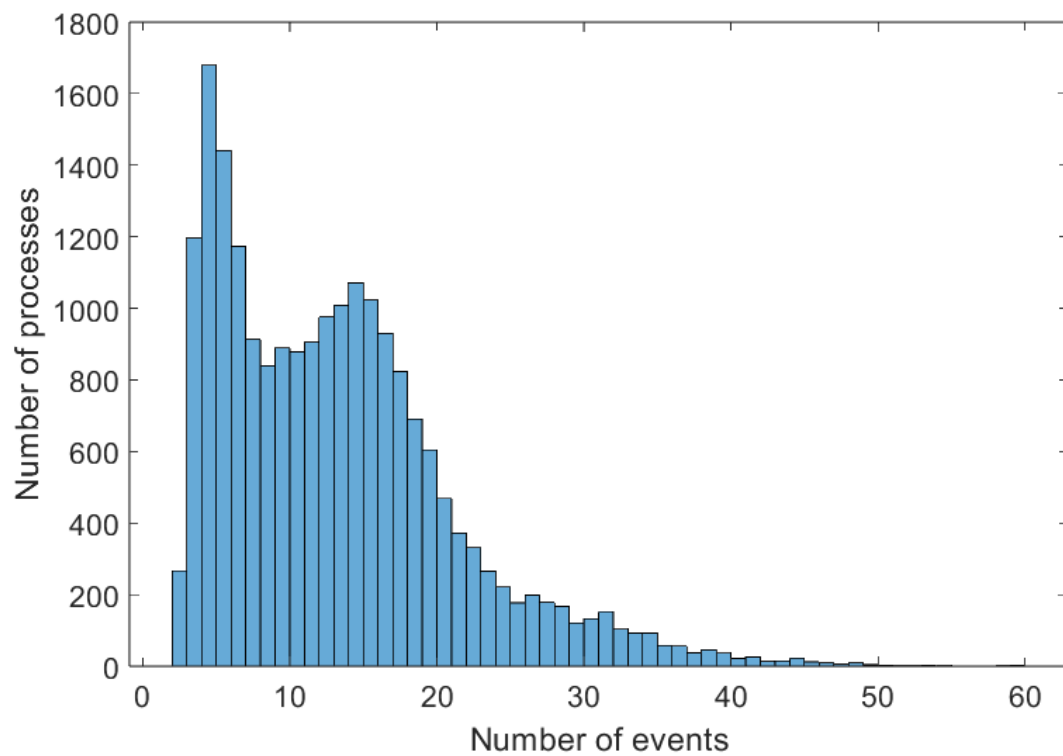


Figure 3.6: Events per case distribution

The cases that are registered in the Immigration Section of the Milan Court are classified 19 different distinct objects (Tab. 3.2), however, the most frequent object represents more than 80% of the cases. For each object, the number and the percentage of the cases that have reached a definition have been calculated and compared with the total number of cases for that object.

Object	# of cases	of which defined	% defined
Appeal pursuant to art. 35 Legislative Decree 25/200	17,049	7,041	42%
Family reunification (art. 30)	1,519	1,074	71%
Appeal against denial/revocation of issue/renewal of residence permit	492	229	47%
Other institutions and special laws	440	322	73%
Challenge of removal order pursuant to art. 20 and 21	334	205	61%
Citizenship rights	249	69	32%
EU Dublin Unit Appeals 604/2013	246	13	53%
Validation of detention order or extension of detention	155	112	46%
Emergency measures pursuant to art. 700 c.p.c	69	36	52%
Appeal for refusal of visa for reasons of family reunification by consular authority	38	16	42%
Other institutions relating to the status of the person and the rights of the personality	37	26	70%
Recognition of statelessness	32	18	56%
Other hypotheses referred to in paragraph 3 art. 3 L.46/2017	29	18	62%
Recognition of right of residence for EU citizens	28	9	32%
Appeal against refusal of humanitarian protection by the Police Commissioner	26	3	12%
Other precautionary proceedings	8	8	100%
Validation of the commissioner's provision for immediate accompaniment to the border	6	4	67%
Revocation of the sentence pursuant to articles. 395 et seq. c.p.c.	5	2	40%

Table 3.2: Objects and their frequency

In Fig. 3.7 the number of open cases (blue line), or backlog, has been obtained by subtracting the cumulative resolved cases (yellow line) from the cumulative new cases (closed), after the peak of new cases in 2019, the slope remained almost constant, effectively defining as many cases as there were from 2019 onwards. It is interesting how constant the number of processes resulting in an outcome per period of time is: approximately 2,500 processes per year; from this trend is quite feasible to be able to predict how many cases will be pending knowing how many new migrants are reaching Italy territory.

According to the judiciary report [7] at the end of 2017 the pending amount is approximately 2,700 cases, which was manually set as backlog starting amount.

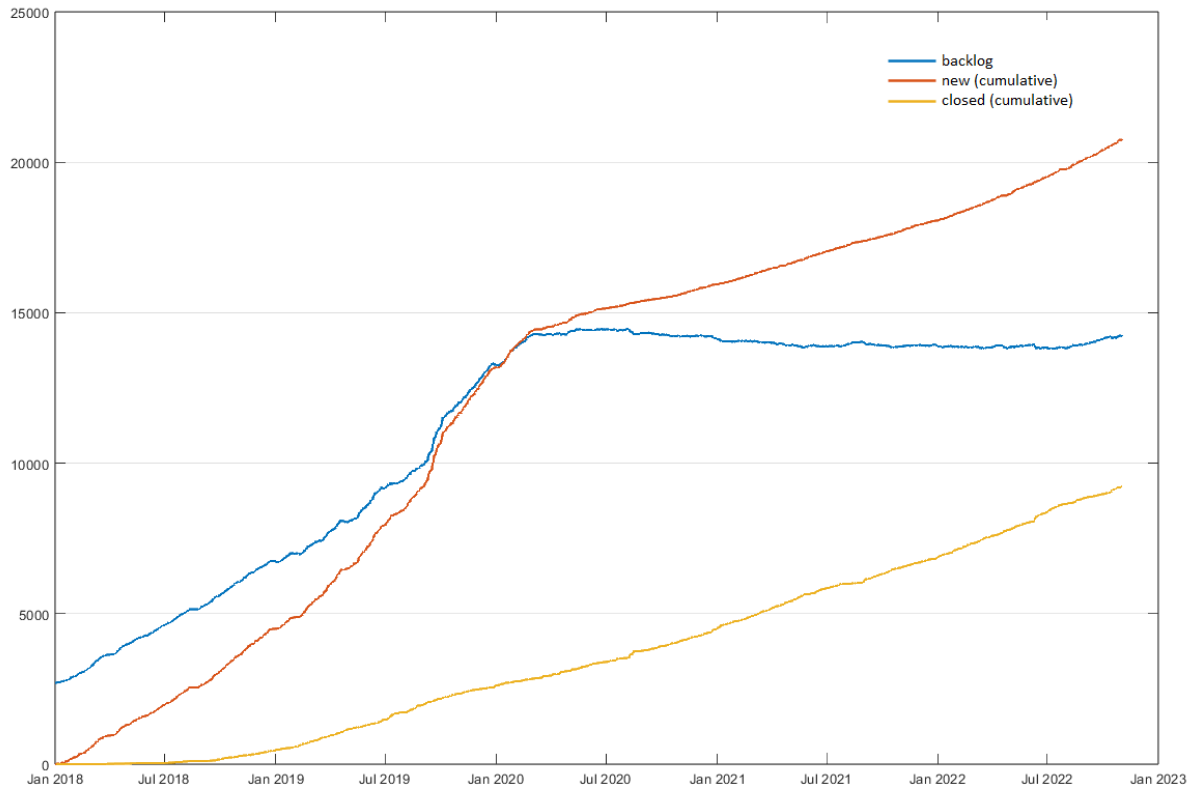


Figure 3.7: Pending cases accumulation (also defined as backlog)

The cases were classified in nine different ritualities, but three of them cover 99.3% of the cases, their characteristics are reported in the introductory Sec. 2.1. Fig. 3.8 clearly shows the disappearance of the W1 rituality (Voluntary First Degree Jurisdiction) at the beginning of 2019, when the rituality 'Chamber for Immigrants' rituality replaced it. The 7O rite ('Summary Cognition 702 bis') has been used consistently over the years, but it has been abrogated by the Cartabia reform starting from March 2023 .

Code	Rituality	Cases Quantity
IM	Chamber for Immigrants	13,350
W1	Summary Rite of Cognition	4,488
7O	Summary Cognition 702 bis	2,791

Table 3.3: Cases ritualities

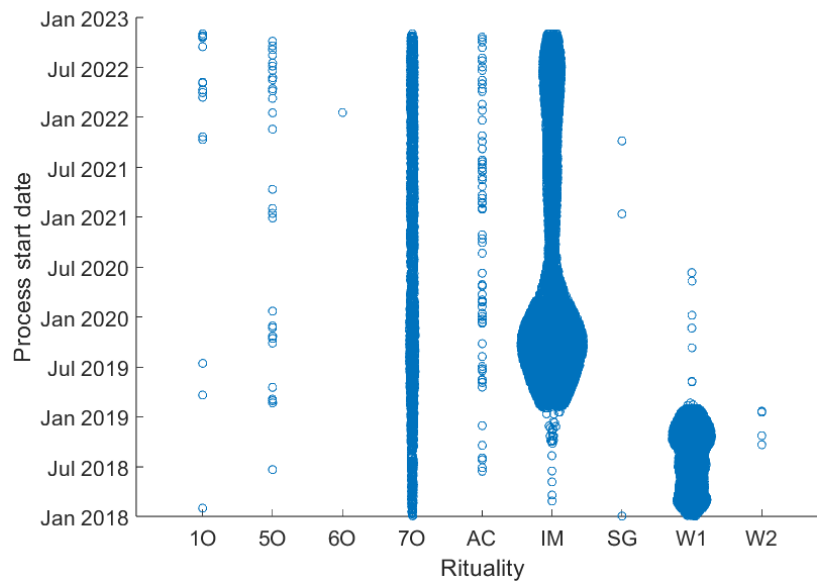


Figure 3.8: Rituality use over time

3.3.2. Judges Profiling

There are six judges assigned to the Immigration Section, according to the official website of the Court of Milan, however, analyzing the designations, 47 judges identifiers have been found. This is attributable to two main causes:

- Every three years a judge has the right to request transfer from one Section to another and is obliged to do so within 10 years [6], some of these are therefore attributable to judges who were part of the period for a limited time span. As Fig. 3.9 shows many have started or finished within the almost four years analyzed.
- Some of these codes actually correspond to the same judges. When a case from another Section is in fact transferred to another due to incompetence (i.e. when it is not suitable to be analyzed by the Section to which the case was submitted), the judge is assigned using another code even though the judge is the same.

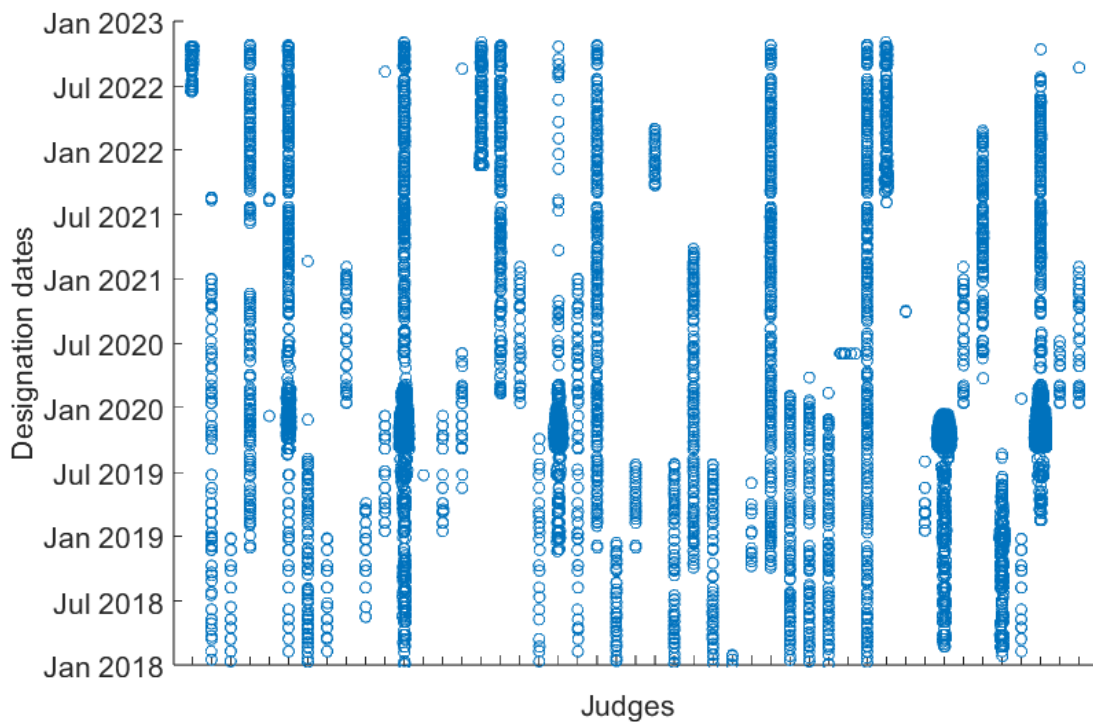


Figure 3.9: Designation of judges over time

The probability of judges to accept or deny an asylum requests has been also analyzed: while all the three judges with more assigned cases share the same asylum requests acceptance rate around 40%, the others judges manifest great disparity up to 98%. An interview with an expert would be necessary to contextualize completely these statistics. In Fig. 3.10 Only judges who have at least 100 cases defined to their identification code are shown.

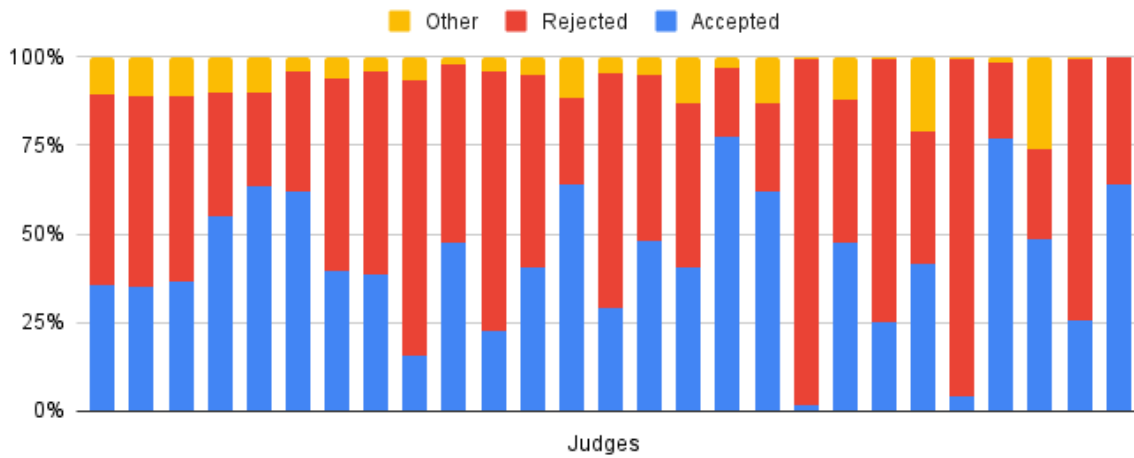


Figure 3.10: Outcomes probability for each judge

3.3.3. Cases Outcomes

As anticipated in Sec. 2.1 there are several possible outcomes regarding the processes of the Immigration Section. For this thesis objectives the outcomes can be grouped into four macro groups:

1. Acceptance of the request for international protection: a decree is issued granting the migrant the status of 'Refugee'. This outcome in particular tends to be made up of several ramifications, based on what type of protection is reserved for the asylum seeker.
2. Rejection of the request: it does not recognize to the migrant the refugee status that will receive a repatriation order.
3. Inadmissibility of the trial: it occurs when the process does not respect some forms established by law that lead to the closing of the process. Although it is not a defined as trial sentence (since a final judgement is not decided), is useful for our purposes to give an overall picture of how the total of the trials tend to conclude.
4. Extinction of the trial: it mainly occurs when the request for protection is withdrawn or the subject involved in the trial is missing.

9,226 processes have reached definition, selected through the events shown in Sec. 3.4. Their distribution in time is rather interesting, as Fig. 3.11 shows. The acceptance rate of protection requests has increased in recent years, compared to the past when the rejection rate was much higher. This may depend on various factors: the European regulations that define the criteria for granting international protection, the country of origin of the

migrants, which affects the reports that support or hinder the approval of the request (it is common that some countries of origin are more prevalent than others in certain periods).

The 'extinction' of the trial appears only lately due to its nature of being put in held before definitely declaring it closed.

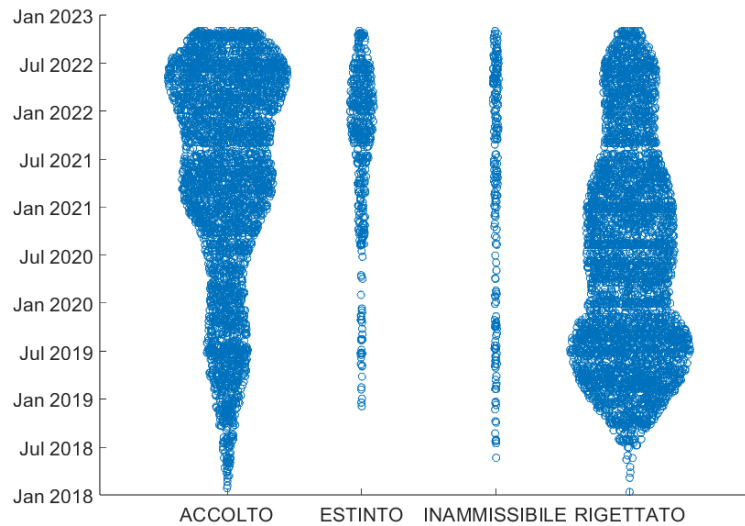


Figure 3.11: Outcomes by date

The final outcomes can also be broken down into different categories based on which events correspond to them. Especially for 'acceptance' there are multiple possibilities, some more advantageous (for example in terms of temporal validity) to the migrant than others. All the possible declinations are shown in Tab. 3.4.

Outcome	Event	#Cases	Percentage
'ACCEPTED'	'TOTAL ACCEPTANCE'	1009	10,94%
'ACCEPTED'	'DECREE/ORDER OF PARTIAL ACCEPTANCE (FINAL)'	660	7,15%
'ACCEPTED'	'DECREE/ORDER OF TOTAL ACCEPTANCE'	532	5,77%
'ACCEPTED'	'RECOGNITION OF SUBSIDIARY PROTECTION'	562	6,09%
'ACCEPTED'	'RECOGNITION OF HUMANITARIAN PROTECTION'	1137	12,32%
'ACCEPTED'	'RECOGNITION OF REFUGEE STATUS'	72	0,78%
'EXTINCT'	'EXTINCTION'	519	5,63%
'INADMISSIBLE'	'INADMISSIBILITY'	226	2,45%
'REJECTED'	'DECREE/ORDER OF REJECTION'	3839	41,61%
'REJECTED'	'REJECTION'	670	7,26%

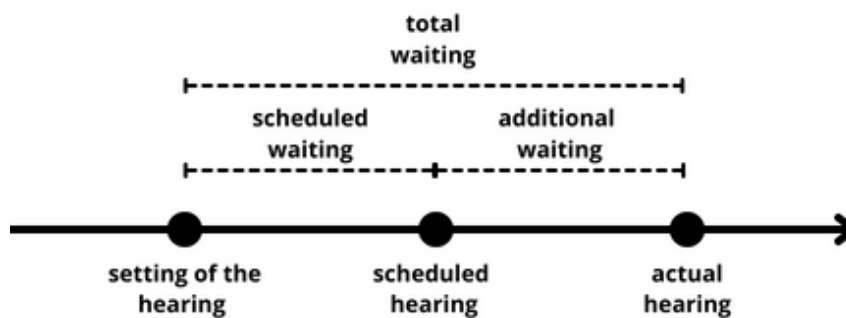
Table 3.4: Outcomes declinations by events

3.3.4. Hearings Waitings Profiling

Having isolated the three most important dates concerning a hearing in the preemptive transformations, namely:

1. The date of the hearing scheduling event
2. The date for which the hearing is initially scheduled
3. The date on which the hearing is actually held. If no postponements occur it coincides with the date on which the hearing was initially scheduled.

There are three waiting times that can be measured: the time distance originally set between the setting of the hearing and the hearing, the additional delay of the hearing due to postponements and the total wait between the setting of the hearing and the actual hearing date.



The reasons for postponing immigration hearings can be many and can concern both the accused and the competent authorities:

- Defendant's failure to appear: if the defendant does not appear at the hearing, the hearing may be adjourned to allow him or her to appear.
- Lack of an interpreter: if the migrant does not speak Italian, an interpreter must be present at the hearing. If the interpreter is not available, the hearing may be postponed.
- Lack of documentation: if the competent authorities do not have all the documentation necessary to proceed with the hearing, the hearing may be postponed to allow them to obtain the documentation.
- Legal activities in progress: if there is ongoing procedural activity that may impact the hearing, such as priority Dublin Unit trials, hearings may be postponed to allow space for them.

- Health emergency: in the event of a health emergency, the competent authorities may order the postponement of the hearings for safety reasons. During the Covid19 pandemic many hearings have been postponed.

Of the 7,177 cases that reached a conclusion, the wait in case the scheduled hearing date has been respected without delays are 4,951, corresponding to 69% of the total. More than 80% of the hearings are set within 100 days, and almost never more than a year apart. The average hearing is scheduled 66 days apart, and the median 51 days. The exceptions are borderline cases that could be also in part attributable to a typing error from the clerks, that by selecting the wrong year set the hearing date to more than 365 days apart. Confirmation of this can be found by analyzing the additional waitings of these cases: the majority of them are negative to an extension of a year or just under, due to fictitious 'postponements' aimed at correcting the hearing date.

Of 7,177 hearings, 2,353 hearings (31%) were instead postponed, suffering an average delay of 136 days and a median of 105 days, with a maximum of 1,008 days. From the statistics, hearings characterized by artificial corrections have been omitted from the calculations.

	#Cases	Average	Median
Scheduled Waiting	7177	66 days	51 days
Additional Waiting	2353	136 days	105 days
Total Waiting	7177	104 days	58 days

The histograms of the three waiting times are attached in Appendix A in order to give an exhaustive idea of their distribution (Fig. A.4 - Fig. A.5 - Fig. A.6). Fig. 3.12 shows the additional waiting compared to the date on which the process started. Cases that are placed on the X axis did not see their hearing postponed.

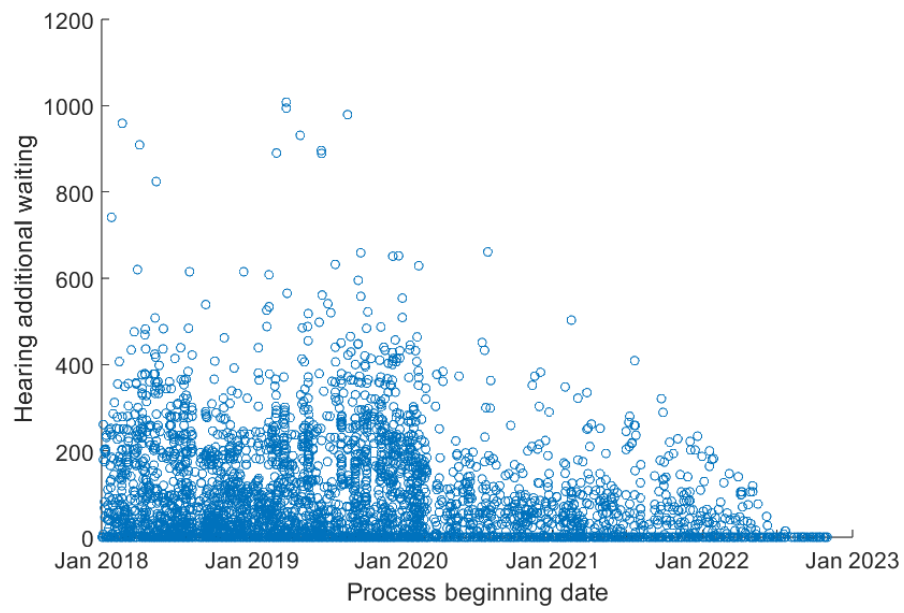


Figure 3.12: Additional waiting over actual hearing date

3.4. Analysis Two - Phases Analysis

This second analysis aims to divide the processes into four distinct phases (judge assignment, hearing scheduling, hearing waiting, decision phase) that represent the four most important parts of the processes. To divide the processes, five key steps were identified with the help of the Immigration Section's chancellors. Dividing the processes in phases permit to examine and understand better the processes and their progresses. These steps are designed for immigration processes, their validity are not verified for other categories of processes. A step can correspond to multiple types of events: as example, the "Hearing scheduled" step can be recorded with distinct types of events such as "Setting of monocratic hearing" or "Setting of chamber hearing".

Phase	Initial Step	Final Step
Judge assignment	Process registration	Judge designation
Hearing scheduling	Judge designation	Hearing scheduling
Hearing waiting	Hearing scheduling	Hearing
Decision phase	Hearing	Sentence

Table 3.5: Processes phases

Through the use of the software Apromore³ clear graphs have been obtained to show better how events corresponding to the steps are more or less frequent and how a specific route of events can influence the speed of each phase.

3.4.1. Key Steps Isolation

The key steps that delimit the phases have been identified as the followings:

1. Process registration: the process registration is the act by which the chancellor certifies the pending nature of a process at the Section to which is assigned. In practice, it is when the case is created in the computer system of the Milan Court and its juridical existence starts.
2. Judge designation: it is often the third event after the registration. It is the act with which the president of the Section designates the judge who will be competent for the progress of the process.
3. Hearing scheduling: is the establishment of a date, time and place in which the

³<https://apromore.com/>

hearing of the process will be held. It determines when the parties involved in the case, such as the litigants, their attorneys, and the designated judge, will meet to discuss the case and present evidence.

4. Hearing: during a hearing, the parties involved in the case, their attorneys, and the designated judge gather to discuss the case, present legal arguments, hear testimony from witnesses, and present evidence. Its actual date can differ from the scheduled day (sometimes more than six months) as shown in Sec. 3.3.4 .
5. Sentence: the sentence of a juridical proceeding is the final decision of the case by the judge. This represents the conclusion of the legal process.

Each step can correspond to multiple alternative events, being each step susceptible to multiple variations reflecting specific conditions (e.g., different process rituality) that have to be grouped as equivalents. In Fig. 3.13 an example of an ideal case where key steps have been highlighted in green.

10128_2019'	2019-03-01 00:00:00'	ISCRIZIONE RUOLO GENERALE'	ATTESA ASSEGNAZIONE SEZIONE'
10128_2019'	2019-03-01 00:00:00'	ASSEGNAZIONE A SEZIONE'	ATTESA DESIGNAZIONE GIUDICE'
10128_2019'	2019-04-04 00:00:00'	DESIGNAZIONE GIUDICE'	ASSEGNATO A GIUDICE'
10128_2019'	2019-07-22 00:00:00'	FISSAZIONE UDIENZA CAMERALE'	ATTESA ESITO UDIENZA CAMERALE'
10128_2019'	2019-10-15 00:00:00'	DEPOSITO ATTO NON CODIFICATO'	ATTESA ESITO UDIENZA CAMERALE'
10128_2019'	2019-10-15 00:00:00'	AMMISSIONE AL PATROCINIO A SPESE DELLO STATO'	ATTESA ESITO UDIENZA CAMERALE'
10128_2019'	2019-10-15 00:00:00'	DEPOSITO ATTO NON CODIFICATO'	ATTESA ESITO UDIENZA CAMERALE'
10128_2019'	2019-10-23 00:00:00'	RIMESIONE IN CAMERA DI CONSIGLIO'	ATTESA ESITO CAMERA DI CONSIGLIO'
10128_2019'	2020-02-12 00:00:00'	DESIGNAZIONE COLLEGIO'	ATTESA ESITO CAMERA DI CONSIGLIO'
10128_2019'	2020-05-04 00:00:00'	DECRETO/ORDINANZA DI ACCOGLIMENTO TOTALE'	ACCOLTO'
10128_2019'	2020-05-04 00:00:00'	ANNOTAZIONE'	ACCOLTO'
10128_2019'	2020-08-10 00:00:00'	ANNOTAZIONE'	ACCOLTO'
10128_2019'	2021-02-10 00:00:00'	ANNOTAZIONE'	ACCOLTO'
10128_2019'	2021-03-04 00:00:00'	ANNOTAZIONE'	ACCOLTO'

Figure 3.13: Critical events isolation example

Process Registration

The first key step, 'process registration', is undoubtedly the easiest step to identify: it is certainly (computer-imposed constraint in the Court's information system) the first event of all trials, recognizable also by the 'ccdoev' column with value equal to 'IA'. Of the initial 20,762 trials, 20,762 (exactly one per case) registrations in the general role are identified. The general role is the registry of all pending Court cases of a specific Section, which keeps track of them and monitors their status. The registration in the general register requires a petition addressed to the clerk⁴ that contains the essential data of the

⁴Link - Tribunale di Messina iscrizione a ruolo

Fig. 3.15 illustrates the temporal distribution of the assignments for each judge, which mirrors the processes registrations (Fig. 3.14) except for a peak at the end of 2020, where numerous new processes were initiated and then assigned after the summer break.

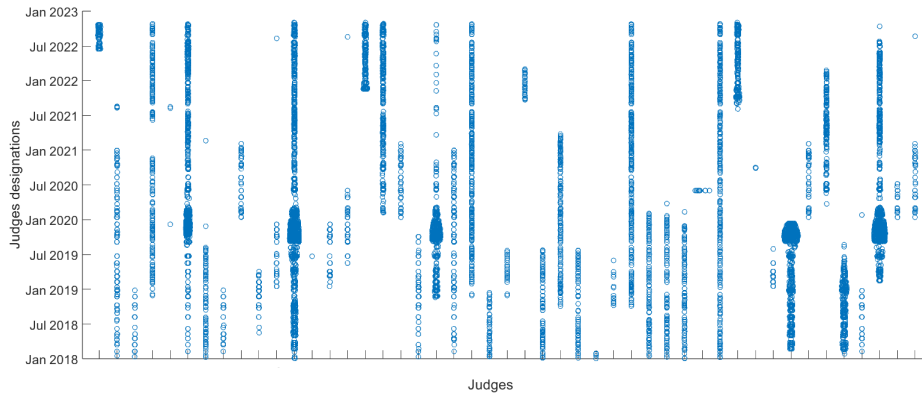


Figure 3.15: Time distribution of cases assignments for each judge

Hearing Scheduling

The third key step in the Immigration process is the hearing scheduling: once a judge is appointed, he or she has the responsibility to schedule an hearing based on availability and urgency. Most hearings are set within three months from the assignation as the graph Fig. 3.16 highlights, even if part of them are postponed after as illustrated in Sec. 3.3.4.

By imposing with Matlab code that the hearing has to be found the appointment of the judge to avoid recording 'false' events due to errors or the frequent work-arounds performed by the clerk's office 12,105 hearings scheduling are identified, 58% of the processes to which a judge has been assigned.

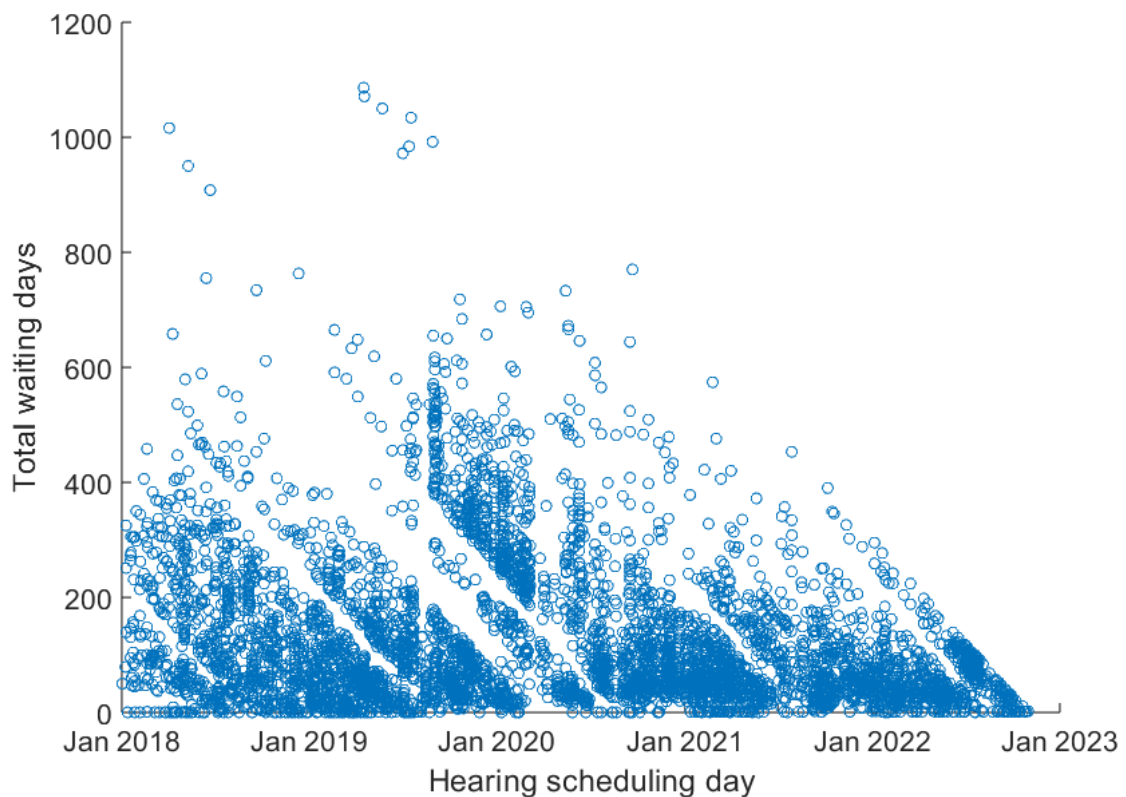


Figure 3.16: Total waiting for the hearing over hearing scheduling date

Hearing

An explicit event that records that the hearing has taken place does not exist. Instead according to the way the Court information system is designed an event that records the beginning of the decision phase will be used instead. The beginning of the decision phase occurs exactly with the hearing taking place. The decision phase is the phase where the judge (if the decision phase is monocratic) or the three judges (if the decision phase is collegial) examine the case and formulate their judgment. The ratios of monocratic decision phases and of collegial decision phase are respectively 41% and 59%.

Depending on the nature of the decision phase the event has two different codes: it corresponds if monocratic to 'RS' ('Reserve' - 'Riserva') or if collegial to 'RCC' ('Remission in Board' - 'Rimessione in Collegio'). Generally, the Court in monocratic composition decides crimes of moderate social alarm, such as theft, injuries, stolen goods. While the Court in collegial composition decides more severe crimes such as aggravated robbery, misappropriation, aggravated extortion, and abuse of office. Exactly 10,975 hearings (and thus beginnings of the decision phase) are found, 90% of the scheduled hearings have been

actually held and recorded in the database available.

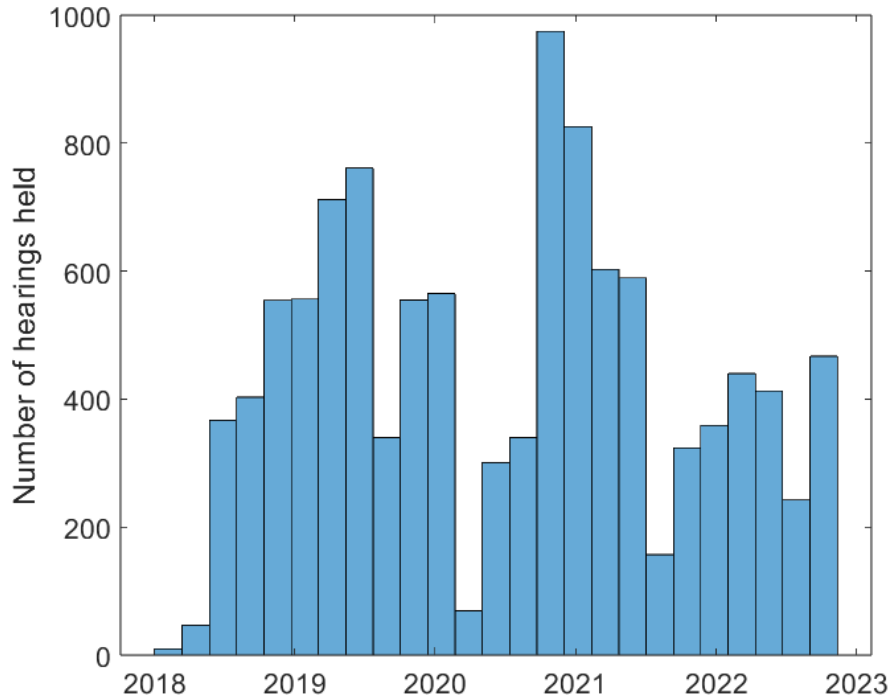


Figure 3.17: Temporal distributions of hearing dates.

Sentence

The last key step in the Immigration Processes is the sentences, the sentences are recorded with many different events. The events have been identified through an interview with the Immigration Section Chancellors Office and manual inspection of some processes.

9,226 of sentence events have been found, which are the 84% compared to the beginning of the decision phase (thus the majority of the processes that had an hearing received also an sentence) and 45% compared to the total enrollments.

In Fig. 3.18 the different events have been grouped into the four macro categories in which the sentences fall: 'accepted' (or 'accolto') if the migrant will have access to international protection, 'rejected' (or 'rigettato') if not. Extinguished ('estinto') is due to inactivity of the parties, and inadmissible ('inammissibile') if the process did not meet certain forms required by law.

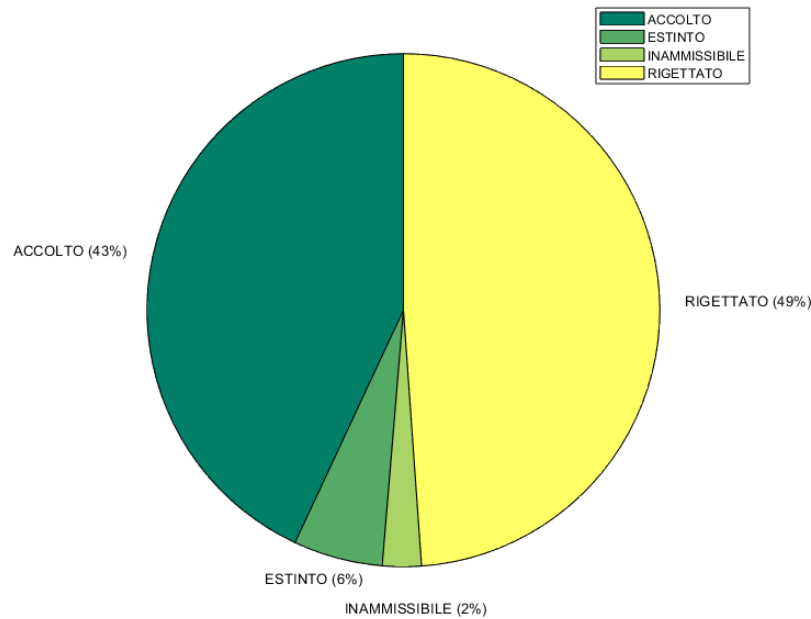


Figure 3.18: Final judgments probability

3.4.2. Phases Duration

Once identified and isolated the decisive moments in the immigration processes, it is now possible to proceed to analyze the duration of the phases between them. However, a premise is vitally important: the database available is limited in terms of time period covered and therefore a natural constraint arises that if left untreated would lead to a selection bias. Since immigration processes are characterized by duration that can reach two or even three years and more, most of the recent ones (reasonably from 2020 onwards) have not yet had the time needed to finish, or rather only the fastest ones have reached conclusion.

This statistical phenomenon is extremely visible by graphing the processes start date and their duration (Fig. 3.19). There is a clear natural boundary that the total durations cannot exceed: the date to which the database has been extracted minus the date of the start of the process. The closer to the database extraction date the process has started the less its maximum duration can be.

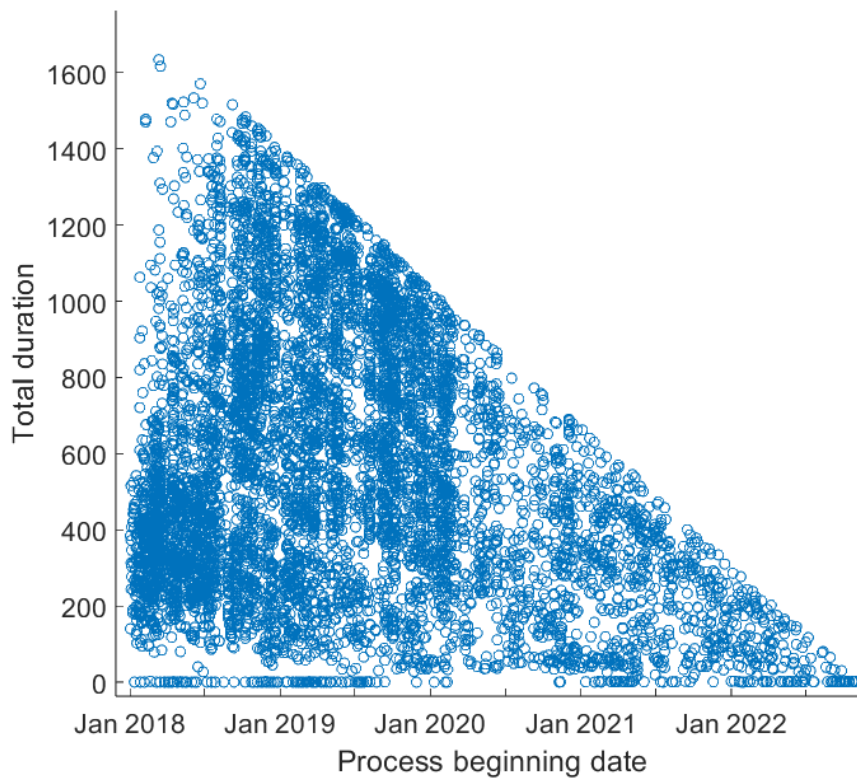


Figure 3.19: Cases length over their beginning date

To mitigate this dynamic, only processes that had years to finish have been selected. The trials that began in 2018 or 2019, which as Fig. 3.20 shows, are still the majority. They consists as the final dataset used for the phases analysis. It comprehends 7,177 trials, 78% of all trials that received a sentence.

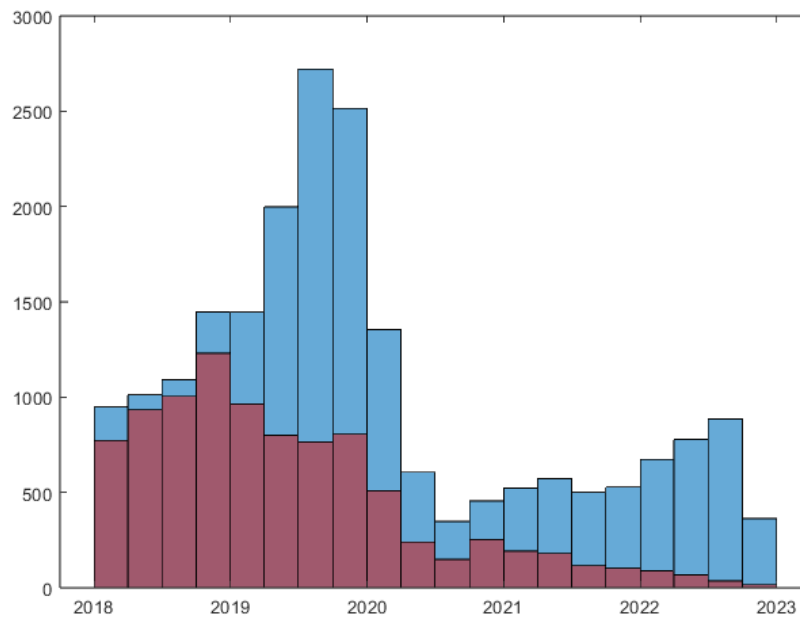


Figure 3.20: Number of started processes in blue of which completed in purple

Fig. 3.21 shows the histograms revealing the distribution of trial durations for each phase.

- The judge assignment phase is unquestionably the one that takes the least amount of time and is influenced by the smallest degree of variability due to the simplicity and standardization of the operation. Around half of the processes complete this phase in one month.
- The setting of the hearing has a high likelihood of being completed very quickly, but it also carries a significant possibility of taking more than a year to finalize. 42% of the processes complete this phase in three months.
- The waiting for the hearing even though it has some very large outliers that deviate from the average, is the second fastest and most consistent phase among all the phases involved in the process. For around two thirds of the processes the waiting of the hearing is less than three months.
- The decision phase is the one that lasts the longest and the most unpredictable, as it is the phase that depends the most on the specific characteristics of each trial and the discretion of the judge. It has an average duration of 246 days (8 months).

Phase	Median (days)	Average (days)
Judge Assignment	27	37.7
Hearing Scheduling	128	242.4
Hearing Waiting	57	107.4
Decision Phase	161	246.6
Total	373 (1.59 years)	628 (1.72 years)

Table 3.6: Median and average phases duration

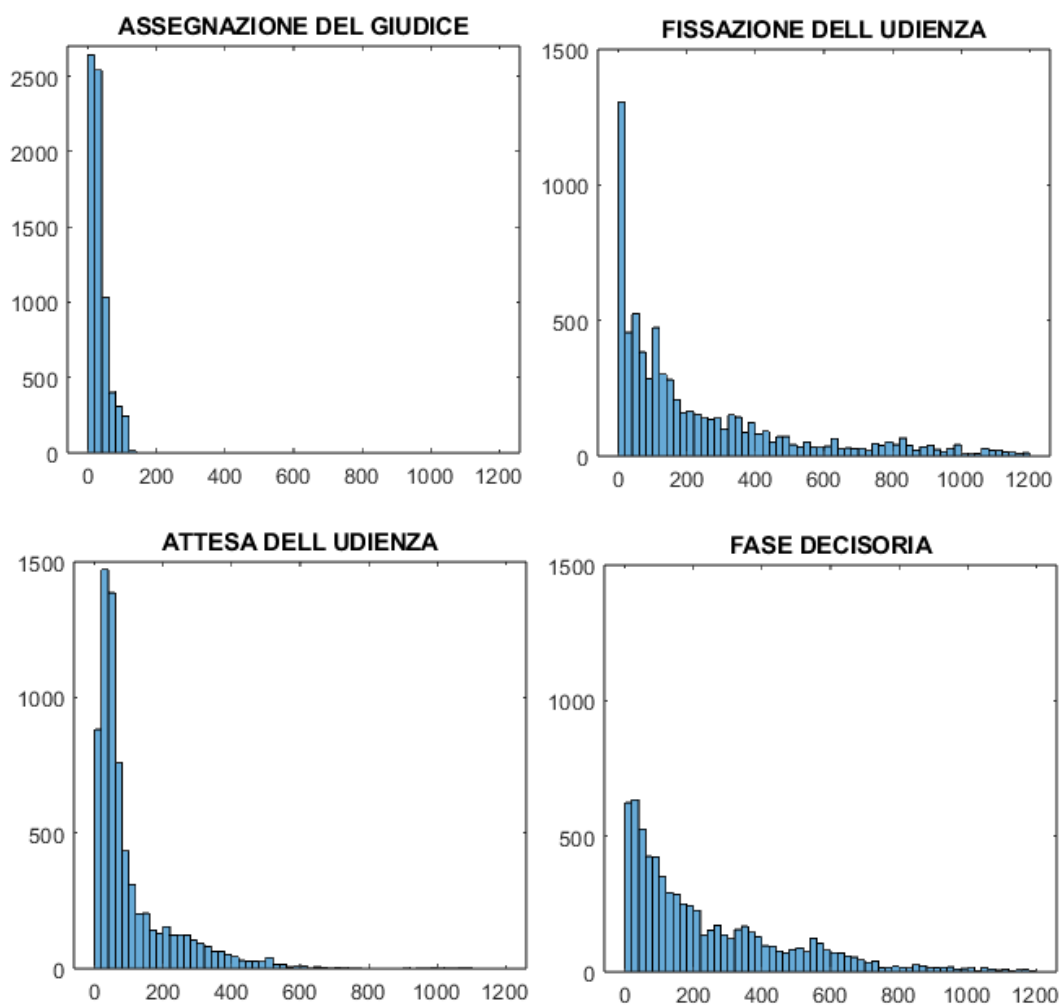


Figure 3.21: Phases histograms. Each chart shows the duration of the phases over the number of cases

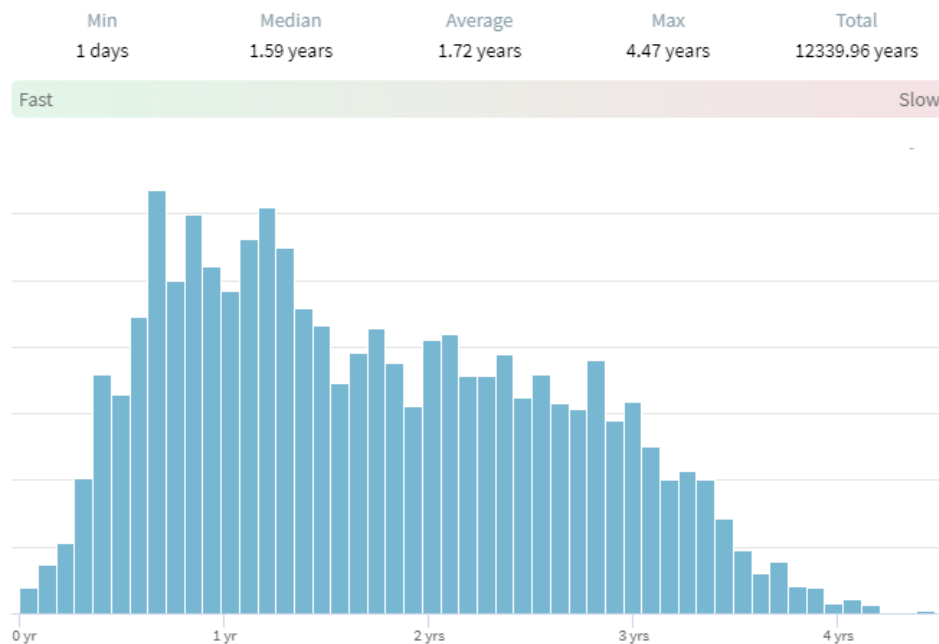


Figure 3.22: Total duration distribution

The dataset was then exported to csv format for analysis executed by Apromore software. The output shows a process scheme, where the arrows between the key steps indicate the probability of following that specific transition out of the total number of processes. The duration above each arrow is the median duration of the transition. The median can be considered more suitable than the average for representing the typical duration of trials, because the average can be distorted by a few outliers with very long durations. These outliers may depend on specific characteristics of some cases, that the Court may not be able to control.

Fig. 3.23 shows the different possible routes for an Immigration process, able to illustrate how different events can influence the duration of a phase. For example, after the judge designation, if the type of hearing is "Comparison hearing" is usually way faster than the other two types of hearings (monocratic and collegial) but it is usually also longer afterwards. It should be noted that this diagram does not include all the paths that trials can take, four rare variations have been excluded from the graphical representation for the benefit of a considerably better readability.

The same graph has been computed in terms of averages in appendix A (Fig. A.1) to make a comparison available; the difference between mean and median for some transitions is as much as two times as much.

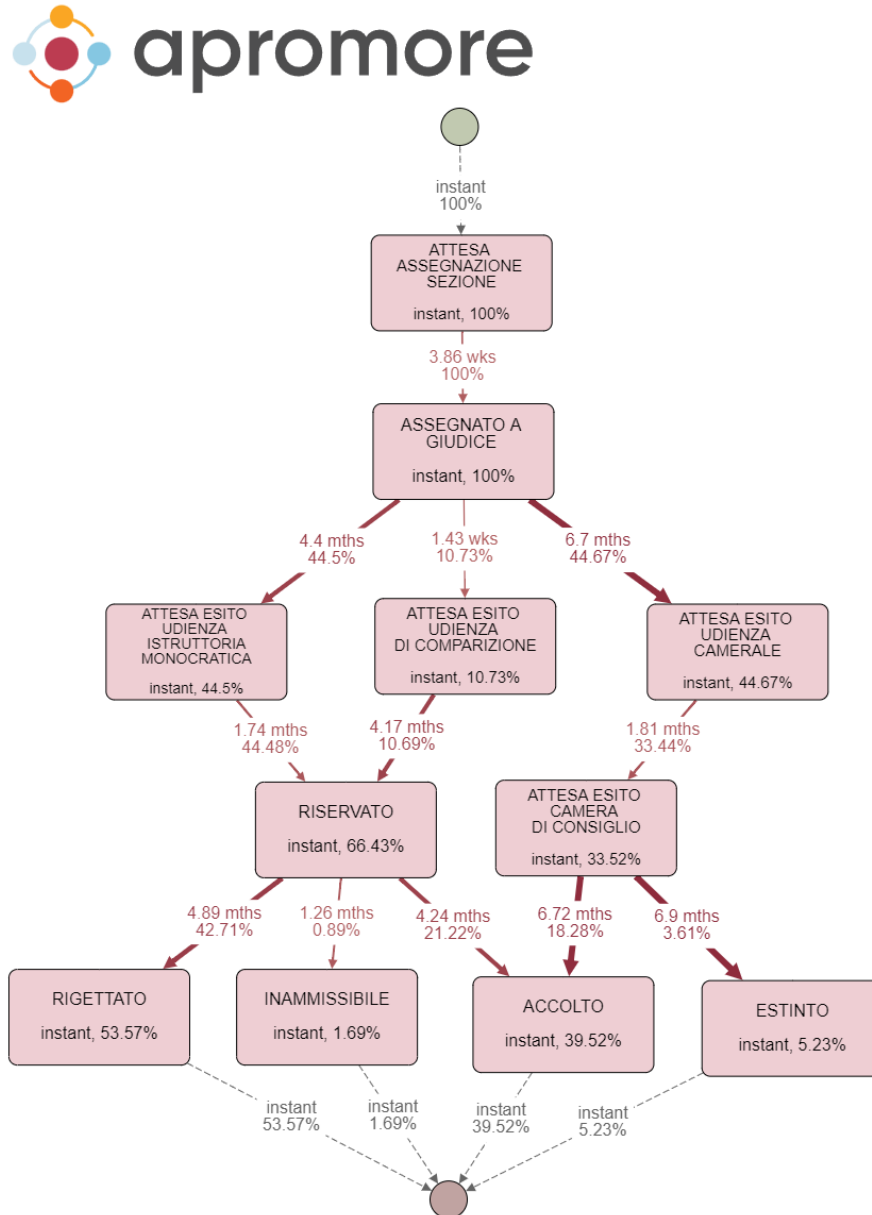


Figure 3.23: Immigration processes possible routes by states

The three most frequent route (or variant) has been also isolated and graphed, representing together in total 53.22% of all processes:

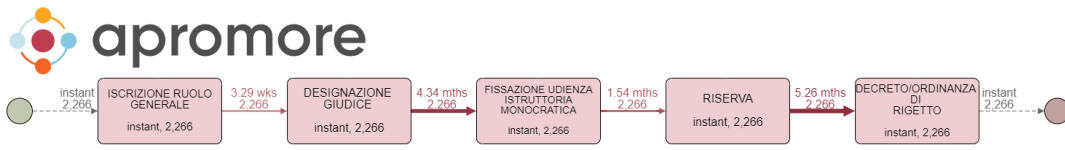


Figure 3.24: First most frequent variant, 31.57% of all cases

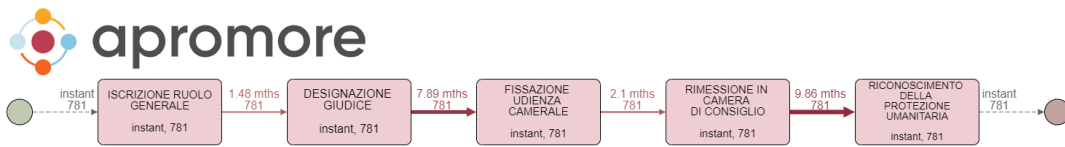


Figure 3.25: Second most frequent variant, 10.88% of all cases

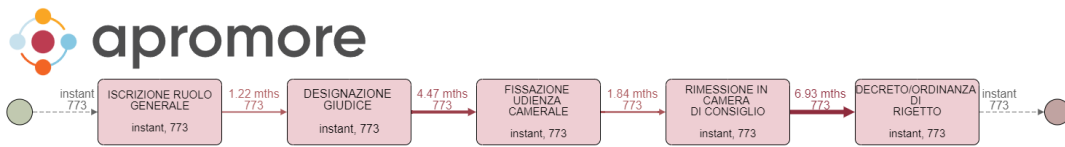


Figure 3.26: Third most frequent variant, 10.77% of all cases

Lastly, the entire chart by events (instead of macro states) has been also computed (Fig. 3.27), the most granular view it is possible to reach.

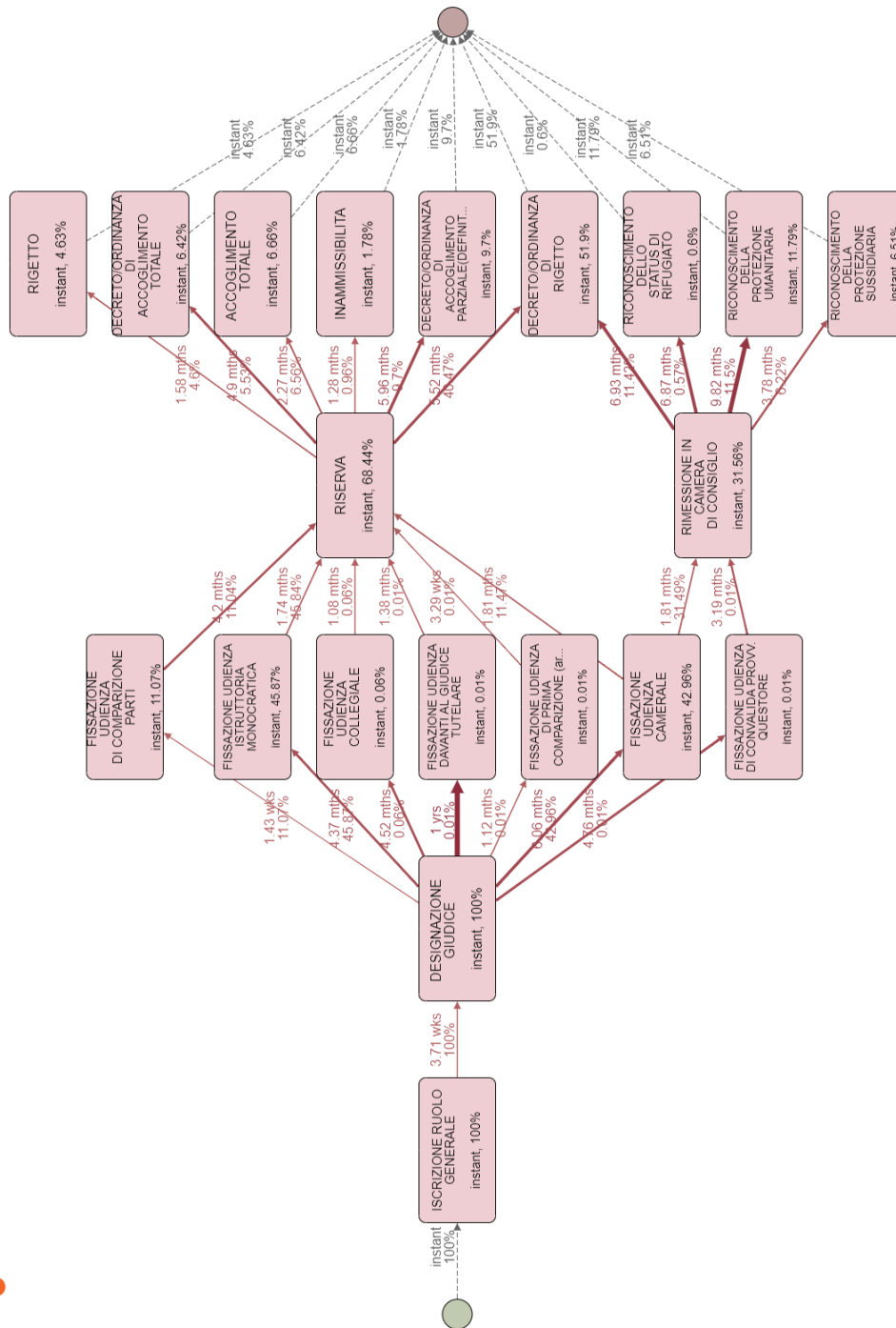


Figure 3.27: Immigration processes possible routes by events

3.5. Analysis Three - Cases Duration Prediction

Machine learning models constitute an important part of the broader field that can be defined as artificial intelligence. They are able to recognize patterns after having been trained on pre-existing data and therefore capable of making predictions on new data.

In this thesis, machine learning models are trained with the objective to estimate the total duration of the processes at their beginning. 10% of the processes have been selected in a heterogeneous manner in respect to the various parameters to use them as test cases, to verify that the built models work well not only with the data used for training but with data not seen before.

There are many different types of learning models, in this thesis the models will be built based on the decision tree and random forest algorithms. The decision tree evaluates the predictors based on conditions that can return a true or false result, this is performed n times in a consecutive and therefore increasingly granular manner, concluding in a final prediction. Some of the advantages of using decision trees are that their functioning is simple to understand and can be visualized graphically, input data can be both numerical or categorical and it requires low data preparation in respect to other forms of learning techniques [23].

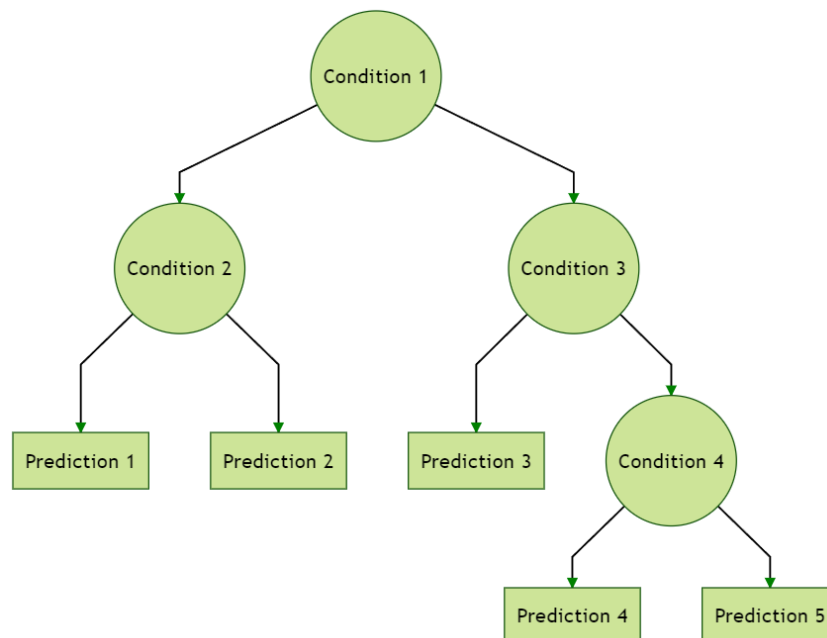


Figure 3.28: Decision tree illustrationan

<https://machinelearningtheory.org/docs/Random-Forest/tree/>

A random forest is a learning method that can be considered an evolution of decision trees. They consist of a set of decision trees, each of them trained with a different subset of the input data. Each subset is selected independently from the others. Once the set of trees has been trained, the output of the multiple trees can be combined by averaging them (in case of numerical data) or by a voting system (in case of categorical data). The final result is to be considered the final prediction. This type of learning model is generally more accurate than a single decision tree, although more expensive to execute [2].

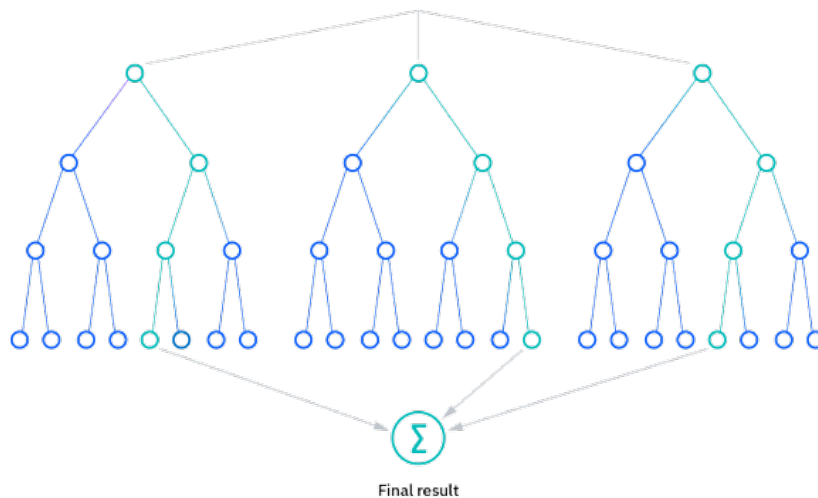


Figure 3.29: Random forest composition
<https://www.ibm.com/topics/random-forest/>

Having a considerable amount of data of the Immigration Section cases available, a machine learning model appears to be a viable direction. The objective is to predict the duration of the entire case, starting from very little information: the data that is available at the registration of the case and at the designation of the judge, the event that occurs reliably after the registration of the case in the register, as shown in Fig. 3.21. The models developed are to be used and intended only for the processes of the Immigration Section, since they are built based on parameters and values that are related to it only.

In Sec. 3.5.1, it is first clarified which predictors have been selected, and what their relationship is with the output variable (the total duration of the cases) through the graphs that show the relationships. In Sec. 3.5.2 models performance are measured and are discussed, with a mention also of how they can be improved.

The dataset used is the one produced by the phases analysis (Sec. 3.4) which comprehends the cases started in 2018 and 2019 that reached a final sentence.

3.5.1. Factors Correlated with the Duration of the Processes

In this section is shown how the different variables are correlated with the total duration of the cases. While many variables are simply 'extracted' from the original database, the number of parallel ongoing processes and the number of sub-processes that compose the cases have been computed respectively by numbering how many processes are going on when a process starts and how many sub processes are registered under the case identifier.

Case Starting Date

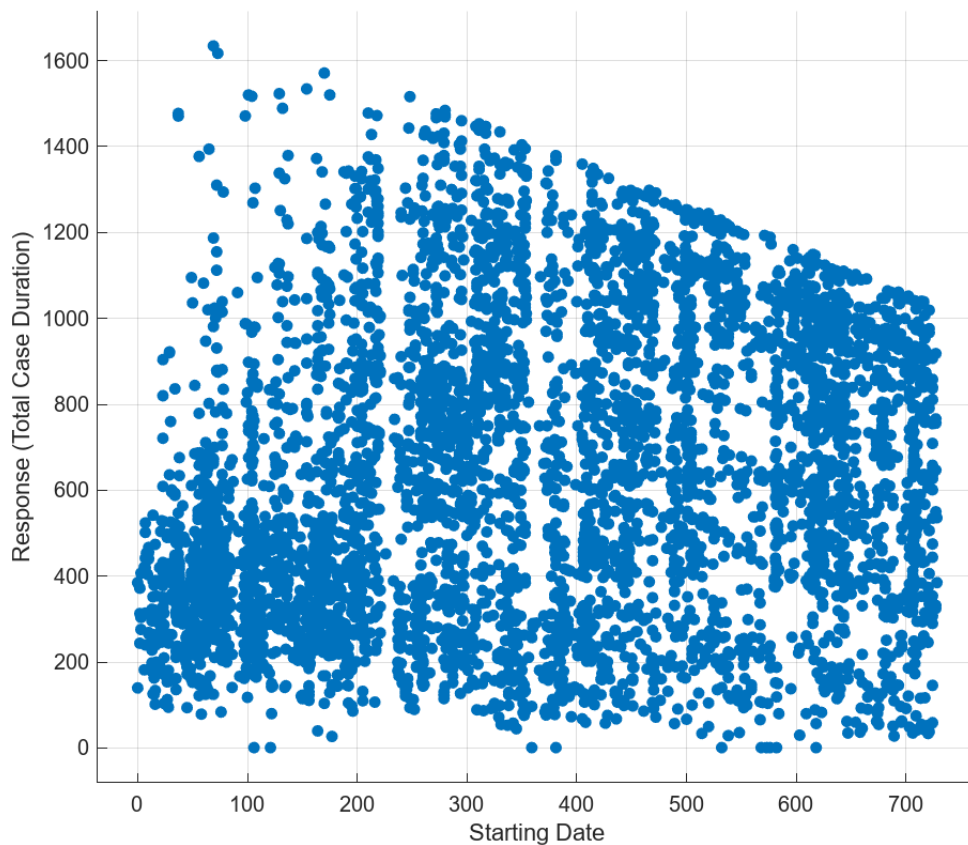


Figure 3.30: Case duration over starting date

The x axis represents the date by a numerical ordering, where zero is equal to the 2nd of January 2018 and 728 corresponds to 31st of December 2019. The maximum duration downward trend observed is caused by the relatively short duration of the time period covered by the database. Since the database ends on November 2 of 2022, the processes started in 2018-2019 and not concluded by that date are not present in this dataset, as the closing event is missing. The diagonal stratification is due to the coincidence of multiple processes closed on the same days. A concentration of cases started at the

beginning of 2018 seems to benefit of shorter durations, by inspecting the data manually no clear explanations have been found: the majority of these cases are assigned to three different judges, the rituality is ordinarily 'W1' and the object of most the processes are 'Appeal ex. art35' as expected. A possibility is that a group of similar cases (e.g. same provenance country) have been managed in parallel, the only common factor found is an high probability of rejection (77%).

Judge Assignment Delay

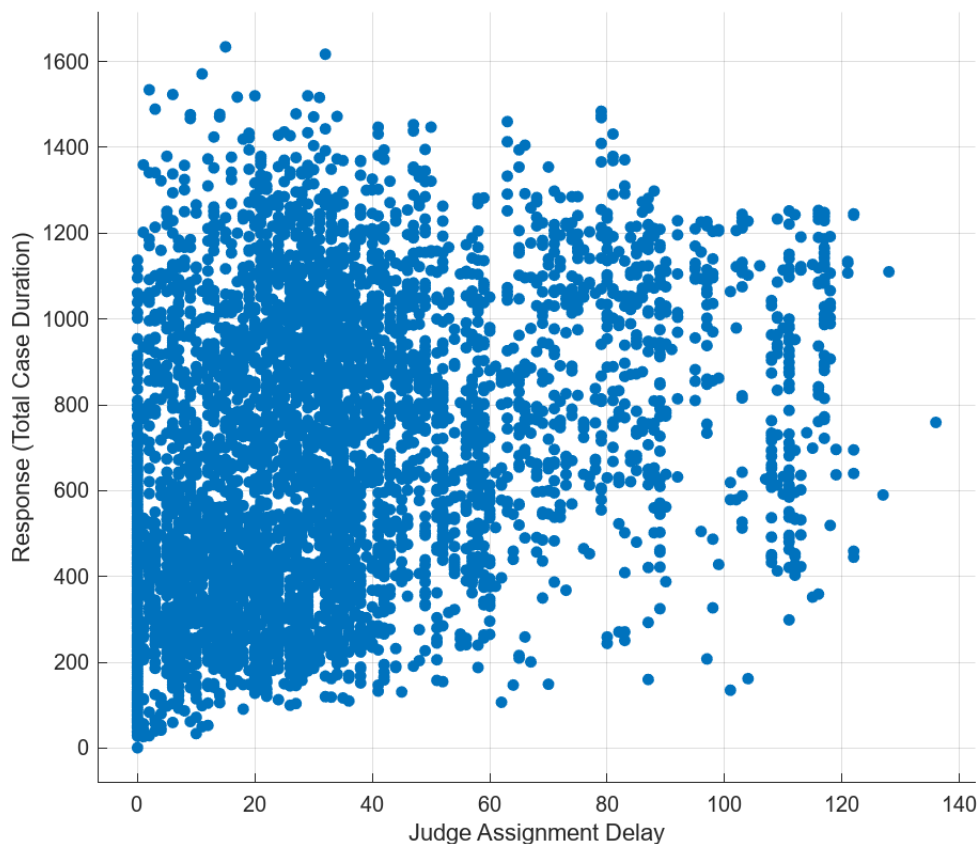


Figure 3.31: Case duration over judge assignment delay

The time that passes between the registration of the case and the assignment of the judge is not very indicative of the duration of the trials: even if this phase contributes to their length in absolute terms, it is the shortest and most stable and does not greatly influence the total time. However, if the judge's appointment takes more than 60 days, it is unlikely that the case will be closed in less than 600 days. In this direction this variable can help in predicting the total duration.

Process Object

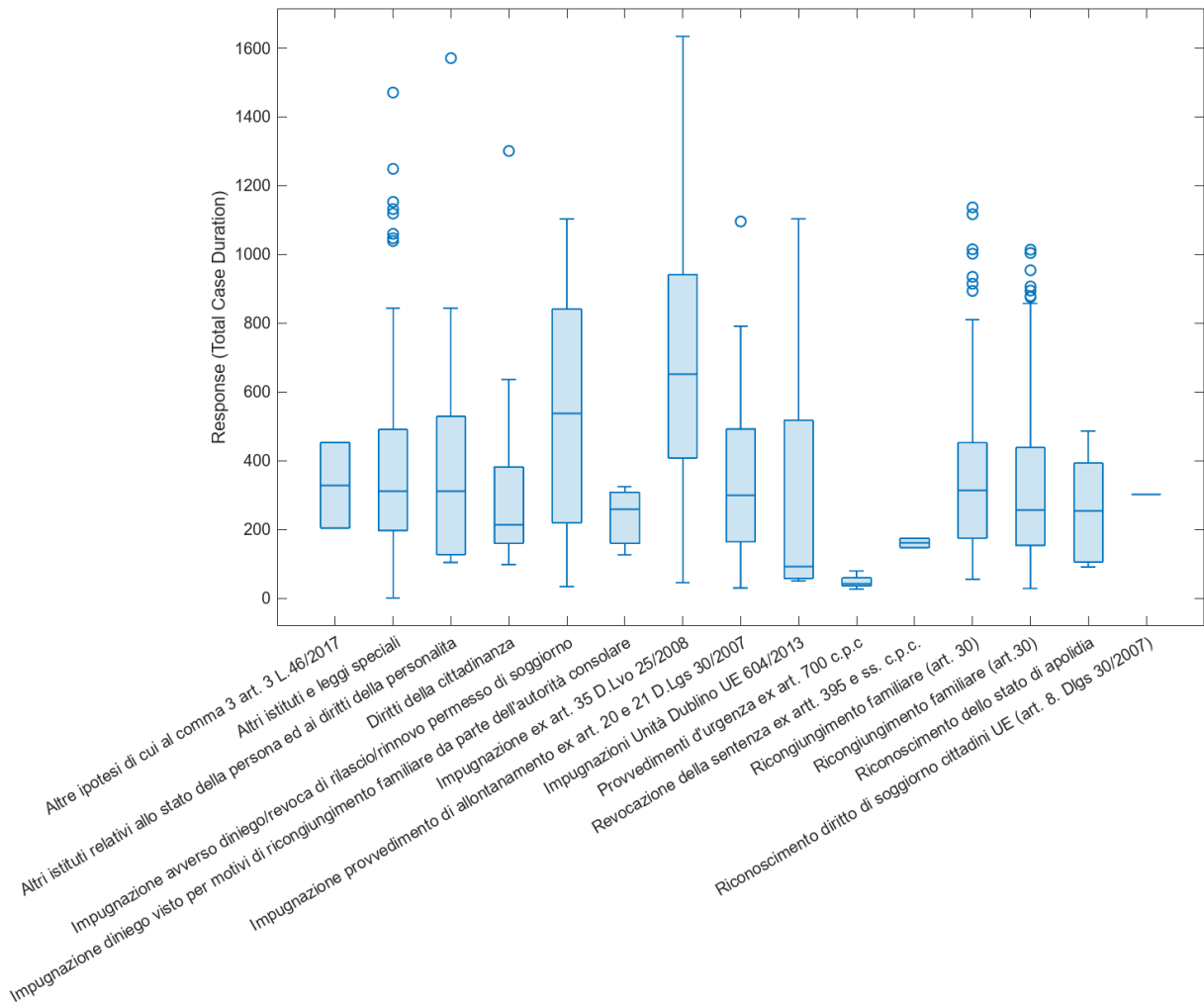


Figure 3.32: Case duration over case object

The appeal to art. 35 (Impugnazione ex. art 35), the most frequent object linked to the processes that the Immigration Section faces, and it characterized by wide variability. Although 50% of cases (second and third quartile) are between 408 and 941 days, the total range that the duration of these cases covers is larger than four years. On the other hand, however, the trials of other types of objects enjoy a narrower range: emergency measures ("Emergency measures pursuant to art. 700") cases last less than 100 days, while the cases with object 'revocation of the sentence' permanently lasts around 200 days. "Dublin Unit appeals" appeals processes within the lower two quartiles reach sentences in less than 91 days.

Process Rituality

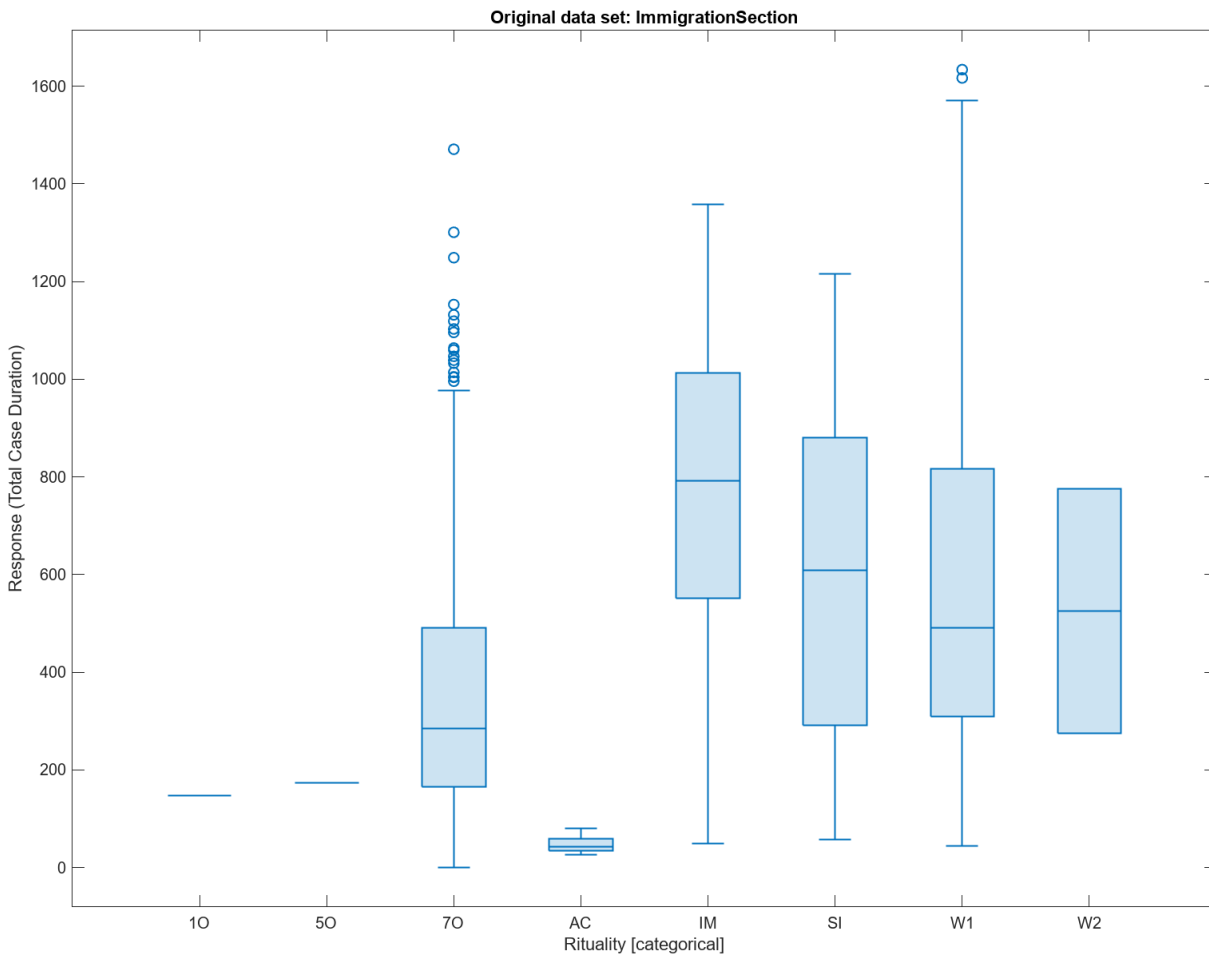


Figure 3.33: Cases duration over rituality

Rituals 70, IM and W1 represent the majority of cases analyzed (Tab. 3.3). Although limited, the belonging of a case to them helps to predict its duration. In particular, the 70, the simplified ritual, has a rather limited duration compared to the others. The 10, 50, AC rituals, despite having a very narrow duration range and therefore with a strong predictive value, unfortunately characterize insufficient cases to significantly improve the machine learning model.

Number of Parallel Ongoing Processes

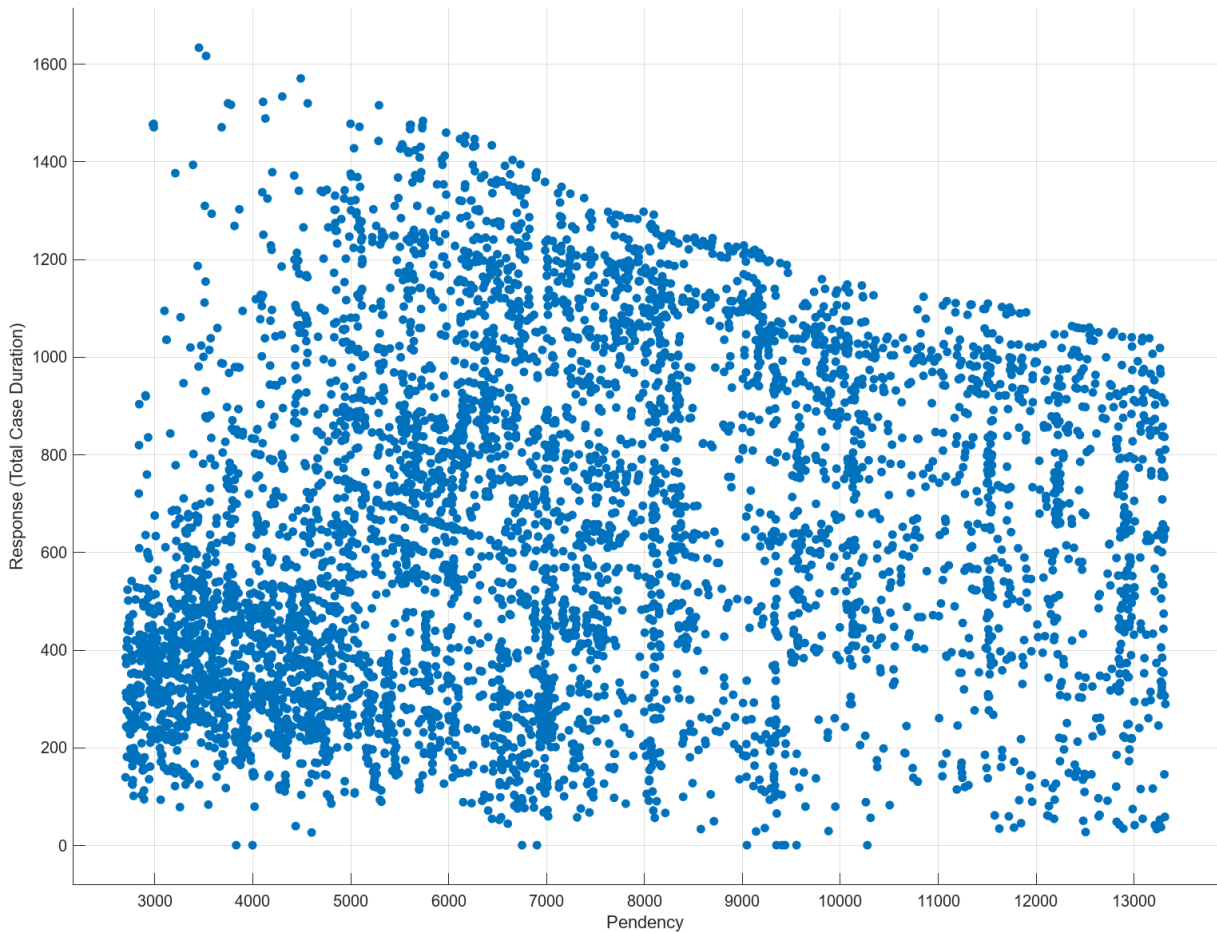


Figure 3.34: Cases duration over number of parallel ongoing processes

The number of parallel cases in progress that have not yet received a sentence on the date of registration of a case is strictly linked to its start date. Fig. 3.34 therefore imitates the trend of the "starting date" variable (Fig. 3.30). A relationship exists between an high number of open cases and a longer average duration, short trials become rare. It is not an original direct variable from the Court data but calculated after, since it has been hypothesized that a greater amount of new cases could lead to shorter durations as an attempt from the Court to keeping pace with the number of new cases.

Number of Sub-processes that Compose the Case

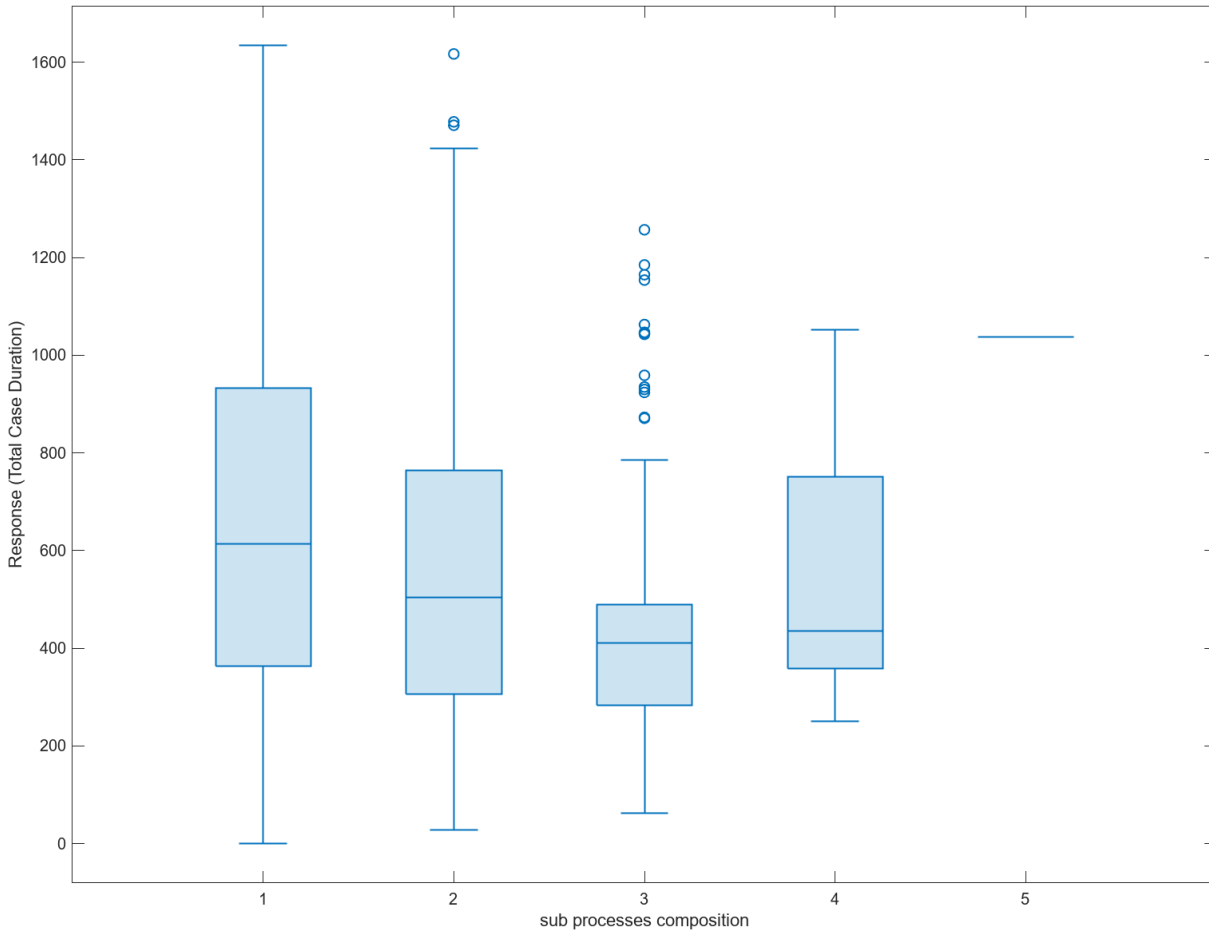


Figure 3.35: Cases duration over number of sub-processes

The sub-processes that make up the processes, a phenomenon that happens with specific juridical conditions (an example is a lawsuit initially against unknown persons and then against a specific person), have no clear correlation with the cases total duration. This variable has been computed by counting how many unique identifiers in the 'numpro' column have been associated with each case. Cases characterized by three sub-processes are faster than those with one or two sub-processes. Unfortunately the mechanisms with which the sub-processes are started and finished cannot be deduced from the available data and their juridical functioning is complex, thus making it difficult to explain the relationship found.

Assigned Judge

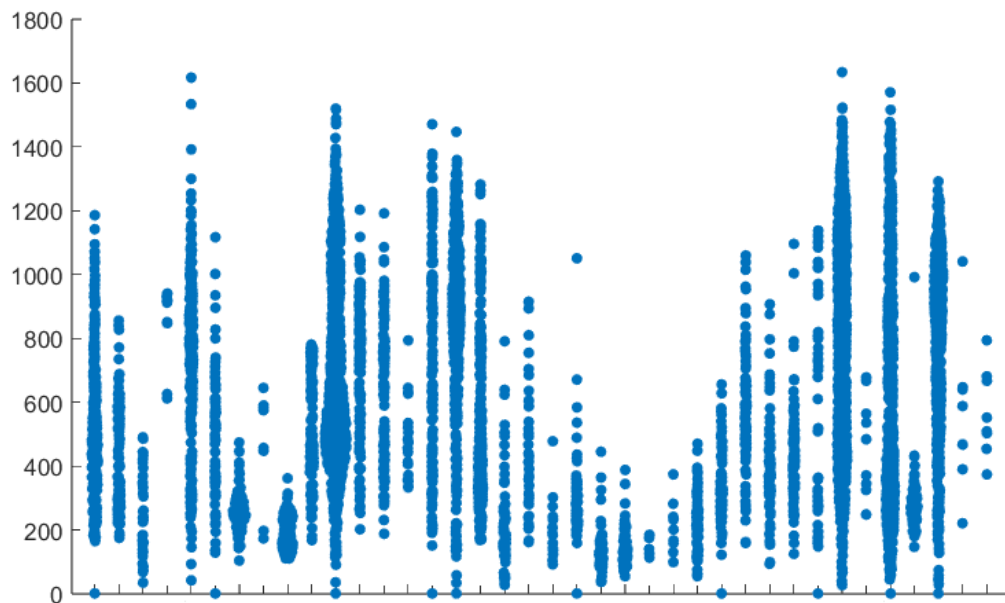


Figure 3.36: Cases duration over assigned judge

The judge assigned to the case has an interesting predictive value: judges who have fewer cases tend to show a smaller possible range of variability, whereas judges who have many cases tend to show a more homogeneous distribution of durations. The judge who is responsible for the most cases shows that a significant part of them are completed around one and half year from the start, while the remainder in the remaining two years. The assigned judge therefore represents a variable with considerable predictive capacity.

Judge Assignment's Clerk

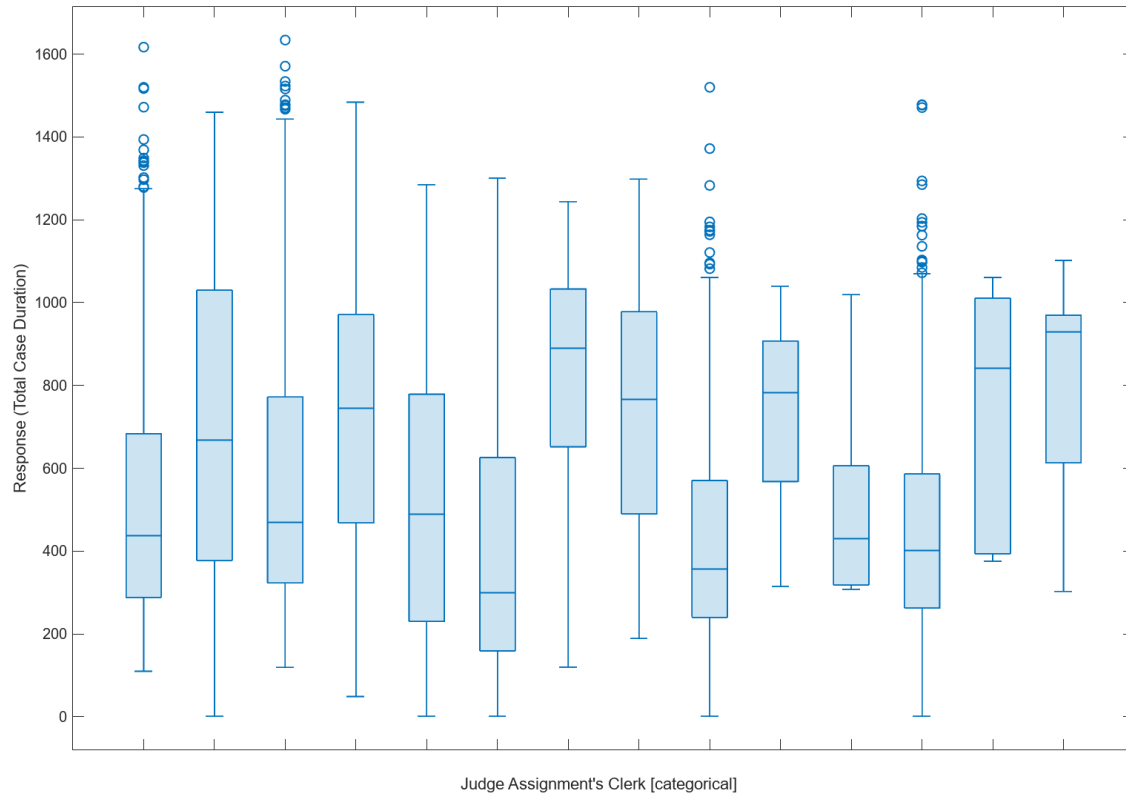


Figure 3.37: Case durations over judge assignment's clerk

It is not clear how the identification code of the clerk who records the judge designation event on the Court information system can be correlated with the total duration of the trials. This relationship was noticed accidentally. Plausible explanations may be that given registrars are more dependent on given judges rather than others, and that therefore the performance of the latter is reflected in the registrar who records the event. However, it may also be that certain clerks are more efficient than others and this has a direct impact on the total duration of trials.

Urgency Indicator

It was discovered that the predictor that would surely turn out to be the most important is missing in the Court's computer data: cases are in fact marked as urgent according to several dynamics, one of which for example is to be a "Dublin Unit Appeal" object process, which, however, represent too few cases to account for the total effect of urgent processes. Urgent cases are not only handled more quickly and thus benefit of a lower duration. Moreover, they also greatly affect those that are being carried out in parallel by

subtracting resources to them. It has been discovered that many hearing postponements are due to urgent cases priority over them. The application of the urgency status is done physically, by means of a stamp placed on the physical file containing the case documents. In the Court's information system, there is no marking of whether a case is urgent or not, nor is there any way to deduce this from the available data. This indicator would be extremely useful for analysis and prediction purposes, which would also allow the calculation of ongoing urgent cases parallel to other cases, another indicator that would surely be significant.

3.5.2. Machine Learning Models

Once the variables available at the moment of registration of the case at their beginning have been analysed, machine learning models were trained. To compare the different methods I preferred the optimization of the mean absolute error (MAE) rather than the root mean square error (RMSE) which is generally taken as a reference. I preferred the average absolute error as it is less affected by extreme outliers, which in the juridical context are due more to the individuality of the cases than to the factors that the Immigration Section can control, or that the available data can express. The mean absolute error is calculated by averaging the absolute difference between the predicted values and the actual values.

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - x_i|}{n} = \frac{\sum_{i=1}^n |e_i|}{n}$$

Figure 3.38: Mean absolute error formula

The model produced with a simple decision tree has an average absolute error of 167 days. The cases that are particularly difficult to predict are particularly long ones (more than 40 months), which the model tends to predict generally shorter in respect to their true duration. These are cases probably linked to the unpredictability of human behavior rather than a natural distribution of process durations. 167 days is a rather wide range, but if it is put into context considering that we find ourselves at the beginning of the process (very little information about the case) and that they can last more than four years, average errors of 167 days are enough to predict how many years the case will probably last. Fig. 3.39 shows the computed tree, being quite large it is impossible to graphically render the various conditions at the decision nodes, however an example of a result with its conditions can be shown.

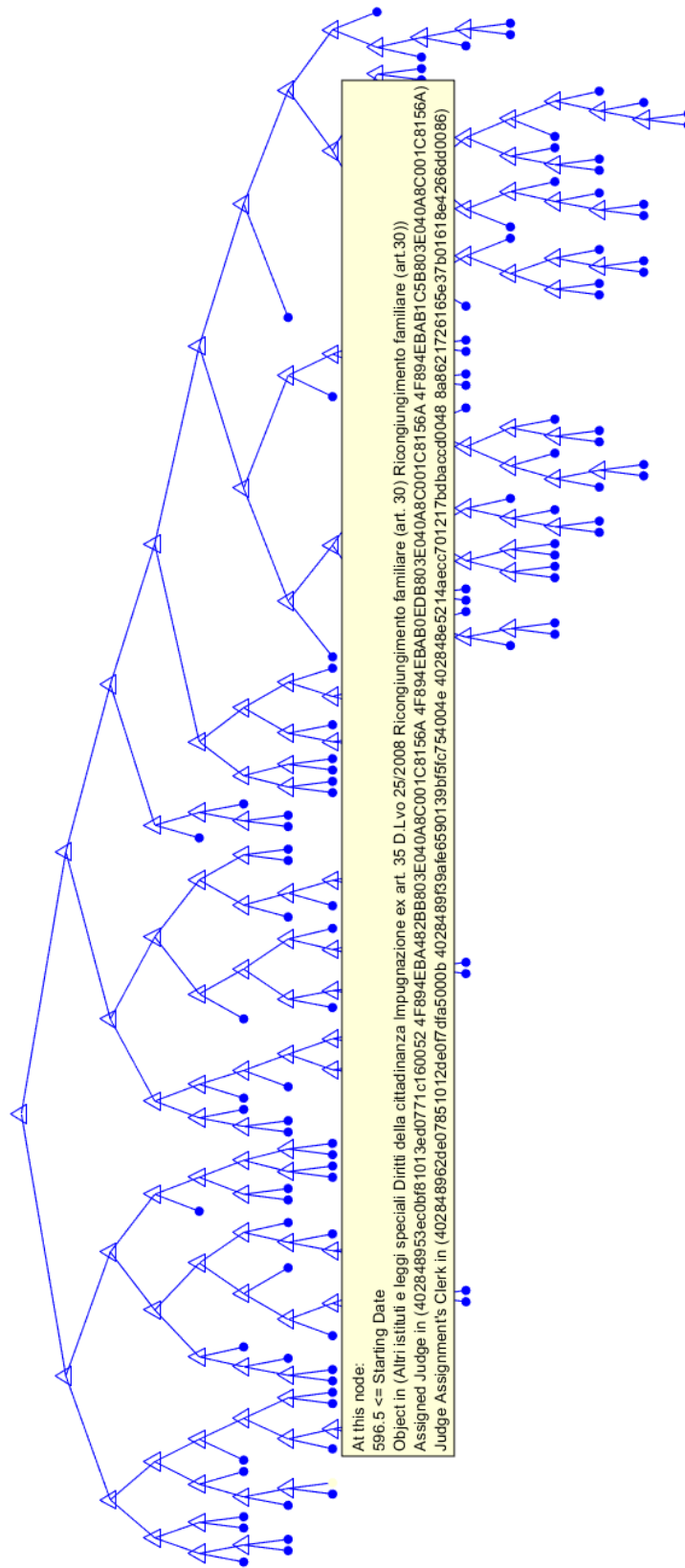


Figure 3.39: Tree model graphical representation

In addition to a decision tree, a Random Forest model was also trained, it is more computationally expensive but tends to offer more reliability in predicting cases outside of those used for training. Like the decision tree, there is greater difficulty predicting the duration of longer cases as Fig. 3.40 shows. However, the mean absolute error decreases to 159 days. As already express, 159 days (around five months) are a considerably wide span, but also a promising result that with the eventual digitalisation of the urgency mark that is applied to the physical cases files would probably achieve very interesting results. The implementation of a machine learning model could be useful as monitoring system of the Immigration Section to anticipate processes with critical durations. Tab. 3.7 shows the model performance results.

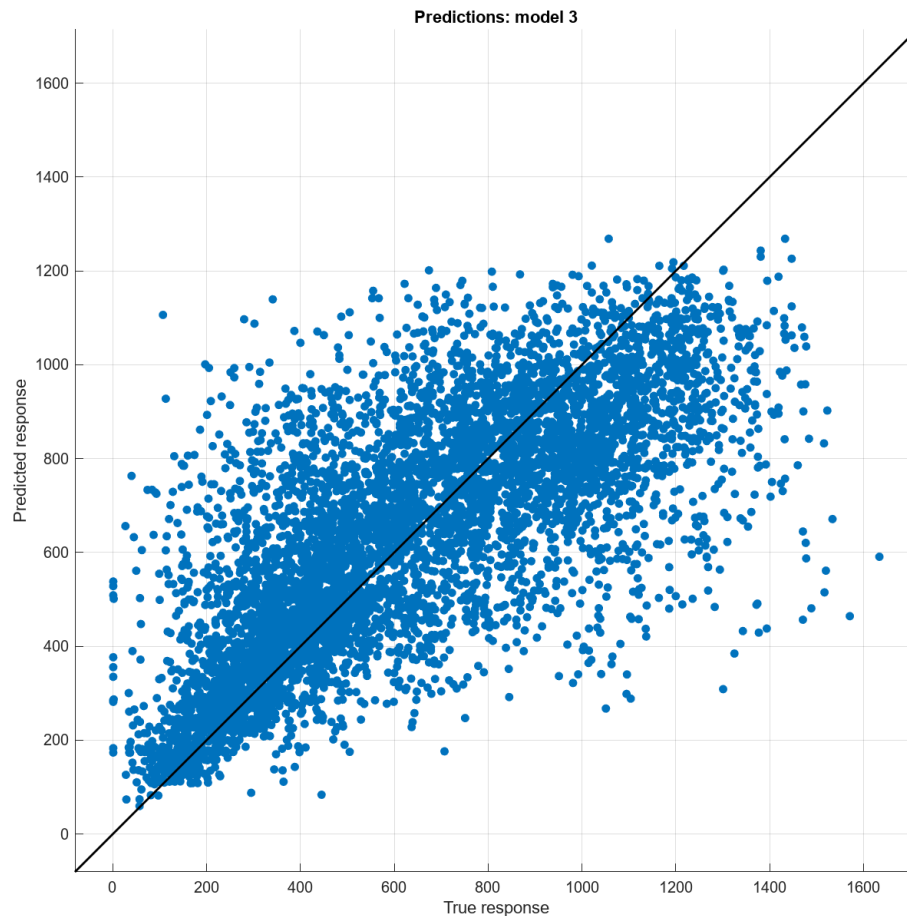


Figure 3.40: Predicted duration over true duration. Random forest model.

Model	MAE	RMSE	R-Squared	Cases average duration
Decision Tree	167.3 days	225.6 days	0.59	628 days
Random Forest	159.2 days	222.1 days	0.60	628 days

Table 3.7: Models performance parameters

4 | Turin Immigration Section Comparison

Since this thesis is part of the larger project "Next Generation UPP: a project to improve the performance of justice in north-west Italy"¹ it was possible to get in touch with a team from the Polytechnic of Turin that produced an unpublished manuscript entitled "Analysis of the Impact of UPP Employees in Improving the Performance of Judicial Offices: Study on the Labor and Immigration Sections of the Court of Turin". This chapter compares the two works carried out, highlighting similarities and differences, both in the approach and in the results achieved. The work carried out by the Polytechnic of Turin is focused on the analysis of the effect of the AUPP personnel (Process Office Employees) that has been hired since the legislative decree number 80 of 2021 [5] with the objective of supporting the Courts daily operations. To do so, the study reports many qualitative analysis (compared to the more inclined quantitative approach of this thesis) carried out with not only data but many interviews with the Sections personnel. Thanks to this transversal approach, many points of confrontation with this thesis emerge.

Number of pending cases

The number of cases that have not been concluded in both Sections (Turin and Milan) is very conspicuous. In this thesis, the number of pending cases of the Immigration Section of the Court of Milan was shown in Fig. 3.7, which in 2020 reaches a stable number of approximately 14,000 cases. In the work carried out in parallel on the Immigration Section of the Court of Turin it is reported that "For exogenous reasons it presents a significant backlog, contributing to 90% of the overall backlog of the Court". According to the report of the Superior Council of the Magistrature [7] in 2020 the Court of Turin has a backlog of pending cases of specific object "Impugnazione Art. 35" around 6,000 cases. It can therefore be reasonably estimated that the pending cases of the Court of Turin are approximately half that of the Court of Milan, while Turin Immigration Section has seven judges assigned to it, one more than the Milan Immigration Section at the date

¹<https://www.nextgenerationupp.unito.it>

of writing of this thesis².

Critical issues of the information system

The Court's information system is a step forward in the digitalisation of work in the field of jurisdiction, but has many points for improvement. For example, the team from the Polytechnic of Turin reports that: "The activities of the UPP staff have been significantly slowed down due to the lack of a field on SICID where the lawyer, when filing the appeal, can indicate the country of origin of the appellant. The implementation of a field indicating the country of origin would be very useful because based on the country of origin the cases can acquire more urgency (for example in 2022, cases involving migrants from Ukraine) and therefore be able to be filtered on the interfaces of the Courts to save a considerable amount of time and effort for staff who currently have to manually search for the desired cases." The work of this thesis fully agrees with what was found, as reported in Sec. 3.5, not only the digitalisation of the country of origin but also the digital equivalent of the "urgent" mark which is applied to the physical files of the cases would be extremely useful both from a practical point of view that the work of the Turin team highlighted and to achieve better analysis and prediction models. An interview with the Head of the Chancellors Office has highlighted that the development of the informative system should be done with close interactions with whom work with it, as it would help to design the software more compliant to the functioning of the jurisdiction dynamics.

Quantitative comparison

The team from the Polytechnic of Turin has also accessed the SICID processes data and executed two main quantitative analyses: one of them being the measurement of the number of cases concluded per semester. Not having available the entire methodology with which these results were calculated, the comparison is to be considered not completely truthful. Moreover, there are other parallel factors that could influence the results:

- There are various criteria to consider when a process ends: for example when it reach a sentence, but not all of the processes reach a sentence although being concluded) or when it reach a specific type of event or set of events such as archiving.
- The information systems can be used with different habits in different Courts, which for the same exact process could lead to two different data records.
- The original extraction from the Court Informative System can influence the final results.

²link - Sezione Specializzata per l'Immigrazione - Magistrati assegnati

Considering these factors, I simulated the analysis remaining as faithful as possible to the description provided by the team of the Polytechnic of Turin. Tab. 4.1 and Fig. 4.1 show the results of the first analysis.

	I Sem. 2019	II Sem. 2019	I Sem. 2020	II Sem. 2020	I Sem. 2021	II Sem. 2021	I Sem. 2022
Turin Imm. Sec.	1363	918	643	882	1059	1040	1421
Milan Imm. Sec	1025	1128	793	1125	1045	1472	869

Table 4.1: Number of cases that reached a conclusion by semester

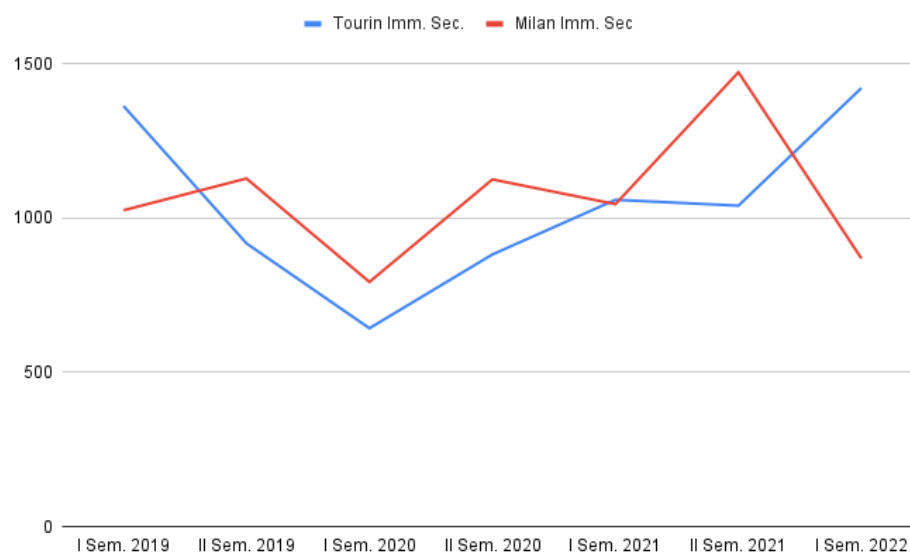


Figure 4.1: Number of cases that reached a conclusion by semester

The Fig 4.1 shows a certain similarity: in 2019 a low peak is registered in both sections due to the Covid19 pandemic. The sentences of cases started before 2018 cannot be fully counted as they are not present in the dataset since the processes started in 2017 or before.

Effect of the new Auxiliary Staff

With the recruitment announced by the Legislative Decree of 9 June 2021 [5], around 8,000 law graduates were introduced in the Courts to increase the speed and efficiency of the Sections. The work of the Polytechnic of Turin detected the impact of the new personnel to be overall positive and useful with an high degree of tasks specialization. In a meeting with the Head of the Chancellery Office of the Milan Immigration Section instead doubts were expressed about the desired beneficial effect: some of the work that the new staff produced was often not compliant with the standards and therefore required additional

work and that the Milan staff have been employed without specialized tasks assigned. As the team from the Polytechnic of Turin reports, "In the Immigration Sections, throughout the national territory, there is a greater problem linked to the training of UPP Employees than the training of the same figures who operate in the other Sections of a Judicial Office: in the latter, in fact, there is no need to explain the different legal institutions to law graduates, but only to show their tasks. In the Immigration Sections, however, the UPP staff must also be trained on the subject itself". From the comparison of the two different effects that the UPP employees have on the two Immigration Sections, it can be deduced that the usefulness of the inclusion of new personnel in the juridical system is strongly improved by the training that is dedicated to them first that helps to perform high specialization tasks.

5 | Conclusions and Future Developments

In this work, a methodology has been proposed to analyze the data that the Immigration Section of the Court of Milan generates and stores on the SICID information system. The developed methodology consists of replicable analyzes which, starting from the raw data extracted directly from the information system, achieves the measurement of processes phases duration useful for monitoring and therefore improving the efficiency of the Immigration Section. In addition, the methodology allows for a broad analysis of the cases, such as the frequency of the judge's sentences over time. The feasibility of machine learning models applied to predicting case duration is also studied.

Initially after giving a brief introduction to the legal context which this thesis confronts with, an illustration of the functioning of the international protection recognition system has been explained in order to contextualize the analyzes in the most complete way. A brief illustration of the field of process mining, a family of techniques that allows the analysis of processes through data, and the current related academic panorama are therefore consequently given. Therefore, a detailed methodology that describes how the data were transformed and prepared for the data analysis chapter: the cases were profiled, what are their juridical parameters, how many and how judges perform them, what are the outcomes and how the hearing postponements affect the total length. The processes were broken down into their sub-phases, thus analyzing their duration and how the different events that constitute them can influence their progress. The last analysis discusses how the various parameters that characterize the cases impact their total duration. Consequently, the feasibility of a machine learning model using these parameters was analyzed by comparing their performance with the mean absolute error estimator. The result is a model that can anticipate the excessive length of a process at its beginning.

This work leaves room for possible future developments, some of the most interesting are:

- Comparative analysis between multiple Sections: there are 17 Immigration Sections operating in Italy, the comparison through the data that is automatically recorded in

the systems could bring to light some interesting comparisons that can help mutual improvement.

- In-depth analysis of the performance of predictive models: through the expansion of the dataset (especially as a larger time window coverage) and the digitalisation of the state of urgency of the case could greatly improve the performance of the machine learning performance and therefore its utility.
- Data quality assessment: a methodology to measure the quality of data generated by legal processes could be developed, given the specificity of the legal field, a methodology to verify data compliance would be very useful to assess the data quality of a log, and consequently the value of the statistics derived from it.
- Urgent cases impact: with a digitization of the mark that is applied to the files that notifies their urgency, it would be possible to measure their impact on the processes carried out in parallel: how their hearings are rescheduled, how the decision phase is protracted, but also how the quality of the outcomes may vary.
- Agent based simulations: digitally simulate the agents that are part of or interact with the Immigration Section, so that it could be possible to verify how, for example, an increase in judges can affect the speed of the proceedings.
- Cartabia reform: from March 2023, the Cartabia reform abolished the rituality "70 - Summary rituality of Cognition 702 bis cpc" establishing the new rituality "art. 281 decies" as substituted. This new rituality consists as a simplified procedure of cognition which promises to be faster and efficient than the 70 rituality. A new data extraction from the Court's information system would allow access to the processes that began to benefit from it since March 2023.

Moreover, a natural extension of this work would be to enrich it with additional cases, extracting again the data that the Court's information system has available. It would be critical that such data not only contains more parameters, but cover a wider time span. Since immigration cases can last up to four years, a dataset that covers at least eight years would be very useful to improve the reliability of the statistics collected, and to measure how the management of the cases changes over the years.

Bibliography

- [1] G. Aagesen and J. Krogstie. BPMN 2.0 for Modeling Business Processes. *Handbook on Business Process Management 1: Introduction, Methods, and Information Systems*, 2015. URL https://link.springer.com/chapter/10.1007/978-3-642-45100-3_10.
- [2] J. Ali, R. Khan, N. Ahmad, and I. Maqsood. Random Forests and Decision Trees. *IJCSI International Journal of Computer Science*, 9(5):272, 2012. URL https://www.researchgate.net/publication/259235118_Random_Forests_and_Decision_Trees.
- [3] Alto Commissariato delle Nazioni Unite per i Rifugiati. Legge n. 189 del 30 luglio 2002 Modifica alla Normativa in Materia di Immigrazione e di Asilo. *Gazzetta Ufficiale n. 199 del 26 agosto 2002*, 2002. URL https://www.unhcr.org/it/wp-content/uploads/sites/97/2020/07/legge_bossi_fini.pdf.
- [4] D. R. Amancio. A complex network approach to stylometry. *PloS one*, 10, 06 2015. URL https://www.researchgate.net/publication/279633355_A_Complex_Network_Approach_to_Stylometry.
- [5] Consiglio dei Ministri. Decreto-Legge convertito con modificazioni dalla L. 6 agosto 2021. *Gazzetta Ufficiale della Repubblica Italiana*, 2021. URL <https://www.gazzettaufficiale.it/eli/gu/2021/06/09/136/sg/pdf>.
- [6] Consiglio Superiore della Magistratura. Regolamento in Materia di Permanenza nell’Incarico Presso lo Stesso Ufficio alla Luce della Modifica Introdotta dal Decreto Legislativo 160. *Delibera del 13 marzo 2008*, 2015. URL <https://www.csm.it/documents/21768/87316/Regolamento+in+materia+di+permanenza+ultradecennale+%28delibera+13+marzo+2008+e+succ.mod.+all%E2%80%9911+febbraio+2015%29/7cf31525-8174-4ab5-864b-0e47997f2714>.
- [7] Consiglio Superiore della Magistratura. Risoluzione sulle linee guida in materia di immigrazione, protezione internazionale e libera circolazione dei cittadini dell’unione europea. , 2021. URL <https://www.csm.it/documents/21768/>

121438/Risoluzione+sulle+linee+guida+in+materia+di+immigrazione%2C+protezione+internazionale+e+libera+circolazione+dei+cittadini+dell%E2%80%99unione+europea.+%28Delibera+di+Plenum+in+data+13+ottobre+2021%29.pdf/360d54e4-a06f-d7b0-7b2e-076185ae7515.

- [8] M. Czaika and C. Reinprecht. Migration Drivers: Why Do People Migrate. *Introduction to Migration Studies. An Interactive Guide to the Literatures on Migration and Diversity*, 2022. URL https://link.springer.com/chapter/10.1007/978-3-030-92377-8_3.
- [9] M. Del Treste. Temporal Analysis of Italian Civil Cases: the Appeal Court of Milan. *Master Thesis in Management Engineering - Polytechnic of Milan*, 2022. URL <https://hdl.handle.net/10589/197585>.
- [10] B. Di Martino, A. Esposito, S. Nacchia, and S. A. Maisto. Semantic Annotation of BPMN: Current Approaches and New Methodologies. *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services*, 12 2015. URL https://www.researchgate.net/publication/286930053_Semantic_annotation_of_BPMN_current_approaches_and_new_methodologies.
- [11] B. Di Martino, A. Esposito, S. Nacchia, and S. A. Maisto. A Semantic Model for Business Process Patterns to Support Cloud Deployment. *Computer Science - Research and Development*, 32, 07 2017. URL <https://link.springer.com/article/10.1007/s00450-016-0333-4>.
- [12] B. Di Martino, L. Colucci Cante, A. Esposito, P. Lupi, and M. Orlando. Supporting the Optimization of Temporal Key Performance Indicators of Italian Courts of Justice with OLAP Techniques. *Complex, Intelligent and Software Intensive Systems: Proceedings of the 15th International Conference on Complex, Intelligent and Software Intensive Systems (CISIS-2021)*, 2021. URL https://link.springer.com/chapter/10.1007/978-3-030-79725-6_65.
- [13] B. Di Martino, A. Esposito, and L. Colucci Cante. Multi Agents Simulation of Justice Trials to Support Control Management and Reduction of Civil Trials Duration. *Journal of Ambient Intelligence and Humanized Computing*, 2021. URL <https://link.springer.com/article/10.1007/s12652-021-03490-5>.
- [14] European Commission for the Efficiency of Justice. Measuring the Quality of Justice. *28th Plenary Meeting of the CEPEJ*, 2017. URL <https://edoc.coe.int/en/efficiency-of-justice/7500-measuring-the-quality-of-justice-guide.html>.

- [15] European Commission for the Efficiency of Justice. Country Profiles 2022 Evaluation cycle (2020 data). *European judicial systems CEPEJ Evaluation Report*, 2022. URL <https://rm.coe.int/cepej-fiche-pays-2020-22-e-web/1680a86276>.
- [16] European Parliament. Regulation (EU) No 604/2013 of the European Parliament and of the Council of 26 June 2013. *Official Journal of the European Union*, 2013. URL <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02013R0604-20130629>.
- [17] D. Fahland. Extracting and Pre-Processing Event Logs. *Eindhoven University of Technology*, 2022. URL https://www.researchgate.net/publication/365209104_Extracting_and_Pre-Processing_Event_Logs.
- [18] International Section of the New York State Bar Association. International Law Practicum. *International Section of the New York State Bar Association*, 2014. URL <https://nysba.org/NYSBA/Publications/Section%20Publications/International/International%20Law%20Practicum/PastIssues2000present/Autumn%202014/IntlPracticumAutumn14.pdf>.
- [19] K. Lashkevich, F. Milani, D. Chapela-Campa, I. Suvorau, and M. Dumas. Why am I Waiting? Data-Driven Analysis of Waiting Times in Business Processes. *International Conference on Advanced Information Systems Engineering*, 2023. URL https://link.springer.com/chapter/10.1007/978-3-031-34560-9_11.
- [20] Ministero della Giustizia. Servizi telematici della giustizia civile. *Dipartimento dell'Organizzazione Giudiziaria, del Personale e dei Servizi*, 2007. URL <http://www.lisug.it/IMG/tel.pdf>.
- [21] A. Nifo and G. Vecchione. Do Institutions Play a Role in Skilled Migration? The Case of Italy. *Regional Studies Volume 48*, 2014. URL <https://www.tandfonline.com/doi/abs/10.1080/00343404.2013.835799>.
- [22] G. Papavero. Sbarchi e Accoglienza di Migranti in Italia negli Anni 1997-2022. *Settore Statistica Fondazione ISMU*, 2023. URL http://www.libertaciviliimmigrazione.dlci.interno.gov.it/sites/default/files/allegati/cruscotto_statistico_giornaliero_31-10-2023.pdf.
- [23] L. Rokach and O. Maimon. Decision trees. *Data Mining and Knowledge Discovery Handbook*, 6:165–192, 01 2005. doi: 10.1007/0-387-25465-X_9. URL <https://link.springer.com/book/10.1007/978-0-387-09823-4>.
- [24] A. Rozinat and W. M. Van der Aalst. Conformance Testing: Measuring the Fit and

- Appropriateness of Event Logs and Process Models. *International Conference on Business Process Management*, 2005. URL https://link.springer.com/chapter/10.1007/11678564_15.
- [25] S. Shahzadi, U. Shahzad, W. Emam, Y. Tashkandy, and S. Iftikhar. Enhancement in Process Mining Model by Repairing Noisy Behavior in Event Log. *Anhui University of Science and Technology*, 2023. URL https://www.researchgate.net/publication/372412168_Enhancement_in_Process_Mining_Model_by_Repairing_Noisy_Behavior_in_Event_Log.
- [26] Uffici del Senato della Repubblica e della Camera dei deputati. Piano Nazionale di Ripresa e Resilienza. , 2021. URL <https://www.camera.it/temiap/allegati/2021/07/28/OCD177-5053.pdf>.
- [27] A. J. Unger, J. F. d. S. Neto, M. Fantinato, S. M. Peres, J. Trecenti, and R. Hirota. Process mining-enabled jurimetrics: Analysis of a brazilian court’s judicial performance in the business law processing. *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*, 2021. URL <https://dl.acm.org/doi/10.1145/3462757.3466137>.
- [28] W. van der Aalst. Process Modeling and Analysis. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, 2011. URL <https://link.springer.com/book/10.1007/978-3-642-19345-3>.
- [29] W. Van der Aalst, T. Weijters, and L. Maruster. Workflow Mining: Discovering Process Models from Event Logs. *IEEE Transactions on Knowledge and Data Engineering*, 16:1128 – 1142, 10 2004. doi: 10.1109/TKDE.2004.47. URL https://www.researchgate.net/publication/3297301_Workflow_Mining_Discovering_Process_Models_from_Event_Logs.
- [30] W. M. van der Aalst. Process Mining: A 360 Degree Overview . *Process Mining Handbook. Lecture Notes in Business Information Processing, vol 448*, 2022. URL https://link.springer.com/chapter/10.1007/978-3-031-08848-3_1.
- [31] J. M. E. Van der Werf, A. Polyvyanyy, B. R. van Wensveen, M. Brinkhuis, and H. A. Reijers. All that Glitters Is Not Gold: Towards Process Discovery Techniques with Guarantees. *International Conference on Advanced Information Systems Engineering*, 2021. URL https://www.researchgate.net/publication/352685407_All_that_Glitters_Is_Not_Gold_Towards_Process_Discovery_Techniques_with_Guarantees.
- [32] W. A. Walsh, T. Lippert, M. G. Edelson, and L. M. Jones. Length of Time to

Resolve Criminal Charges of Child Sexual Abuse: A Three-County Case Study. *Behavioral Sciences and the Law*, 33, 2015. URL <https://pubmed.ncbi.nlm.nih.gov/26206709/>.

- [33] J. D. Zhou. Determinants Of Delay In Litigation: Evidence and Theory. *American Law and Economics Association Annual Meetings*, 2008. URL <https://law.bepress.com/cgi/viewcontent.cgi?referer=&httpsredir=1&article=2240&context=alea>.

A | Appendix A

A.1. Processes schemes with average durations

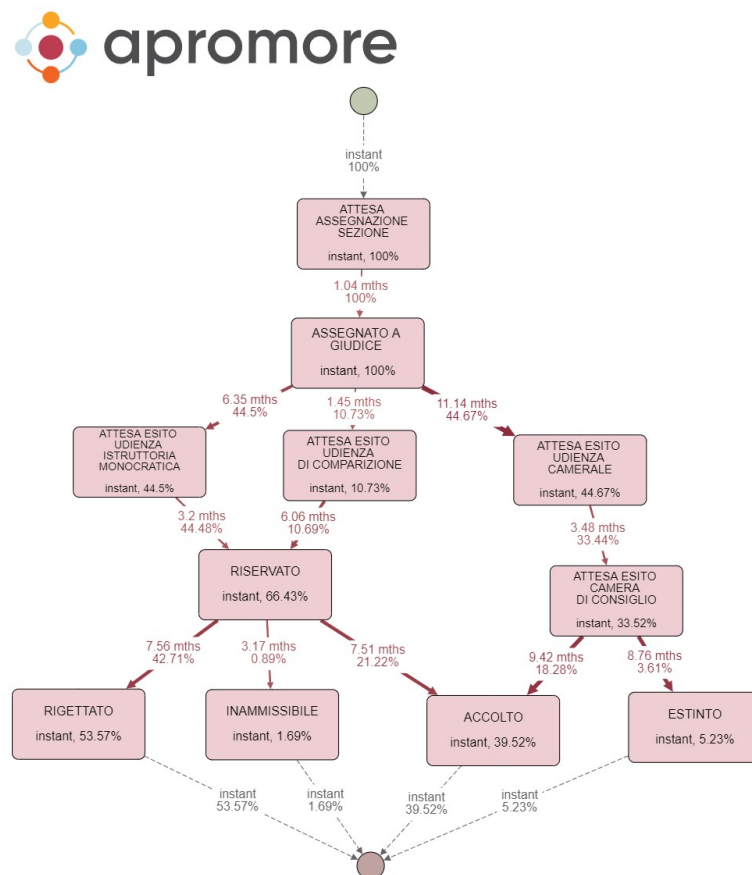


Figure A.1: Apromore - Average label

A.2. Files columns description

The six extracted files can be summarized as follows:

1. **processes** containing the general details of the processes categorized into 54 columns, the most important of which consist of:
 - (a) canruo and cnuruo: the first, "role year code" indicates the reference year for the start of the process, while the second "role number code" indicates a (non-progressive) number of registration in the system; the most important thing is that the concatenation of the two forms a unique key that identifies the process of which the process is part.
 - (b) numpro: uniquely identifies the sub-processes, thus allowing us to calculate how many sub-processes the processes are made up of.
 - (c) ctipse: section type code, identifies which section of the court the trial belongs to (the Immigrant Section corresponds to code 'SI'). In Milan there are nine sections: executions, family, business crisis and over-indebtedness, business court, protection, immigration, civil area, criminal area, work.
 - (d) object code: the object code is composed of three different codes (i.e. codruolo, codmateria and codoggetto) forming part of the "objects" file which therefore allows to identify the matter ('macro topic') of the processes.
 - (e) as written above there are 50 columns, some of them with data that could represent significant information (for example ctipri, nimpri which represent the rituality and the amount of the case), others containing null value for 99.98 percent of voices.
2. **history** containing all the events recorded for all the processes, among the 19 columns the most interesting are:
 - (a) numpro (process number): useful for subsequently merging this file with the others
 - (b) numprv (provision number): a unique identifier for each event
 - (c) dataev (event date): the date on which the event that the entry identifies actually occurred can differ greatly (even one year) from the date of recording of the event itself. the graph. there are also cases in which the event seems to have been recorded a year in advance, one can assume due to a simple typing error.

- (d) `ccdoev` (event code): contains the event that the line records, there are many (in the immigration section there are 260), about 20 of these however cover almost all of the recorded events. An 'event dictionary' file has been created which allows each event code to be associated with a description that explains what it refers to.
 - (e) `ccodst` (status code): indicates the state the process is in, it is relatively similar to the `ccdoev` column except for a coarser granularity.
 - (f) `datere`: (registration date): date and time in which a given entry is registered in the `sicid` system by the lawyers' or magistrates' registry, can differ greatly from the date in which the event actually occurred.
 - (g) `dataev` (event date): date on which the event actually occurred. There are also years of delay (and in one case even in advance). this is (at least for the majority of cases) due to corrections and workarounds carried out by the chancellery.
 - (h) `judge id`: contains the assigned judge for the trial
3. **objects, states**: contain a description respectively of the objects (uniquely connected to the object code), states (uniquely connected to the `ccodst`) and matter (macro objects categories)
 4. **eventistati**: contains a sort of mapping of all the possible transitions between the different events, a file that is quite complicated to approach and not very useful.
 5. **hearings**: contains the hearings held, which can be linked to the individual trials via various columns.

A.3. All SICID columns description

Titolo Colonna	File	Descrizione
CCDOEV'	storico	Codice identificativo dell'evento
CCODST'	storico	Codice identificativo dello stato
CDESCR'	storico	Descrizione dell'evento in formato specifico
CEVPAD'	storico	Codice evento precedente adiacente
CODUTE'	storico	Codice Utente, forse dell'utente che ha eseguito la registrazione
CRONOLOGIC	storico	Valori che si ripresentano inizialmente 7 volte in ordine progressivo fino a 10119, per il resto è NaN esattamente come IDTIPOATTOCRONOLOGICO
CSTAPR'	storico	Codice Stato Precedente
CTIPSE'	storico	Codice identificativo della sezione
DATAEV'	storico	Data in cui è avvenuto l'evento (senza orario)
DATARE'	storico	Data di registrazione della tupla, gli eventi non vengono necessariamente registrati in ordine cronologico
DATAULTIMA	storico	Data ultima modifica di qualcosa, il 90% di questi valori è uguale a DATARE
IDTIPOATTO	storico	90% circa è NaN, per il resto sembrano essere "ID TIPO ATTO CRONOLOGICO"
ISVISIBLE'	storico	Codice che identifica una scrittura manuale nel database da parte del tecnico
NUMGIU'	storico	Numero identificativo del giudice
NUMPRO'	storico	Numero del processo, un fascicolo può contenere più di un processo
NUMPRV'	storico	Numero provvedimento, usato per mettere in ordine cronologico gli eventi
Var1'	storico	Numero progressivo artificiale, utile per verificare non ci sia duplicazione degli eventi post join
CCODST'	stati	Codice identificativo dello stato, in join con CCODST di storico
CDESCR'	stati	Descrizione degli stati; c'è CDESCR anche in Storico che invece è la descrizione degli eventi
FKFASEPROCESSO'	stati	Fase k Fase processo
VISIBLE'	stati	14506 "0" per il resto tutto 1
CODGRUPPO'	oggetti	Codice gruppo
CODICE'	oggetti	Codice, formato da CodiceGruppo_CodiceMateria_Codiceoggetto; utilizzato per JOIN
CODMATERIA	oggetti	Codice materia
CODOGGETTO	oggetti	Codice oggetto
CODRUOLO'	oggetti	Per alcuni oggetti è diverso, non so cosa sia e lo elimino per evitare che il merge abbia duplicati di cui l'unica differenza è un codruolo differente
COLLEGIALI	oggetti	Indica se il codice materia è di tipo collegiale
DATAFINE'	oggetti	corrispettivo finale di datainizio, 90% finisce nel 2099 una 10% per cento finisce nel 2022/2023
DATAINIZIO	oggetti	per il 90% 1/1/1970, solo alcuni partono dal 2022 probabilmente per indire l'inizio di un nuovo oggetto
DESCCOMPLE	oggetti	Descrizione dell'oggetto di maggiore lunghezza rispetto a DESCOGGETT; da considerarsi pressoché equivalenti
DESCOGGETT	oggetti	Descrizione dell'oggetto
RITO'	oggetti	Rito a cui l'oggetto appartiene
RITUALITA'	oggetti	Ritualità a cui l'oggetto appartiene
RITUALITA2	oggetti	Sconosciuto

Figure A.2: SICID variables description Pt. 1

Titolo Colonna	File	Descrizione
ANNOGIUDIC	fascicoli	anno di assegnazione al giudice
BARCODE'	fascicoli	codice a barre associato al procedimento (in alcuni uffici al momento dell'iscrizione)
CANNSE'	fascicoli	anno sezionale
CANRUO'	fascicoli	Anno di ruolo, serve per creare la chiave identificativa del fascicolo
CATTOI'	fascicoli	natura dell'atto introduttivo (citazione, ricorso)
CCODGI'	fascicoli	grado del giudizio
CCODST'	fascicoli	Codice dell'ultimo stato raggiunto dal fascicolo
CCODUF'	fascicoli	codice ufficio
CEUENT'	fascicoli	codice ultimo evento
CLASS_ACT'	fascicoli	indica se il procedimento è di tipo Class-Action
CNTISTANZA	fascicoli	Codice numero tipo istanza
CNUMSE'	fascicoli	numero sezionale
CNURUO'	fascicoli	numero di ruolo, serve per creare chiave del fascicolo
CODICEOGGE	fascicoli	Codice oggetto, utilizzato per join con Codice di Oggetti
COLLEGIALE	fascicoli	Indica se il fascicolo in oggetto appartiene al rito monocratico o collegiale
COSTGI'	fascicoli	indica la parte che si è ricorrente in giudizio. (attore o controparte)
CSTAPR'	fascicoli	memorizza l'ultimo stato in cui si è ritrovato il processo
CSUBPR'	fascicoli	NULL' 20975 per il resto quasi tutti "1,0" codice sub procedimento cautelare
CTIDEF'	fascicoli	20204 NULL per il resto variegato Tipo definizione
CTIPPR'	fascicoli	tipo procedimento
CTIPRI'	fascicoli	Codice Tipo Rituallità
CTIPSE'	fascicoli	Codice tipo Sezione, indica la sezione in cui il fascicolo sta venendo svolto
CVELIN'	fascicoli	{'N' } 20802 , {'NULL'} 3364 ,{'S' }
DATAULTIMA	fascicoli	Data in cui e' stato effettuato l'ultimo aggiornamento delle informazioni
DDADEF'	fascicoli	20204 NULL per il resto date singole collegato con CTIDEF: Data definizione
DEVENT'	fascicoli	Data evento
DISCRI'	fascicoli	data di iscrizione a ruolo del procedimento
DORAUD'	fascicoli	23538 NULL per il resto date del 1970
DPROUD'	fascicoli	23538 NULL per il resto date variegate
ESEZIONEC	fascicoli	Indica se il procedimento è esente o meno dal pagamento del contributo unificato
IDREFASC'	fascicoli	identificatore della fase
IMPORTOCON	fascicoli	importo contributo unificato
NIMPOR'	fascicoli	valore della causa
NUMGIU'	fascicoli	Numero identificativo dell'ultimo giudice assegnato al processo, sostituibile da NUMGIU di Storico
NUMPRO'	fascicoli	Identificativo del numero del processo
NVALOR'	fascicoli	{'0' } 15884 {'NULL'} 8449

Figure A.3: SICID variables description Pt. 2

A.4. Waitings distribution histograms

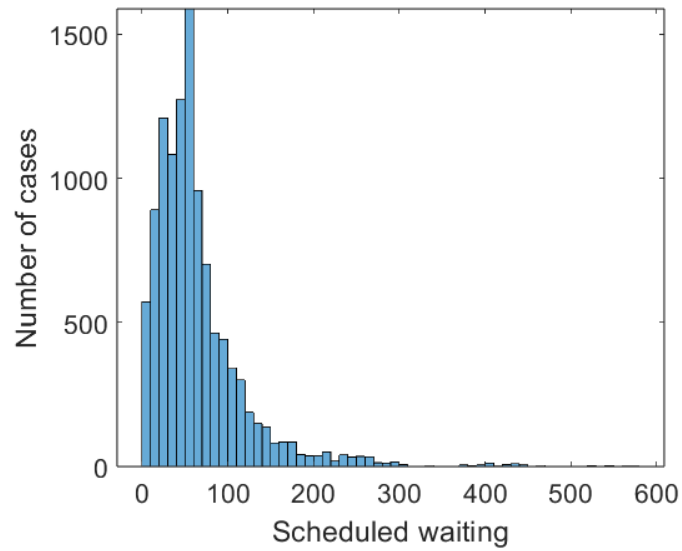


Figure A.4: Scheduled waiting distribution

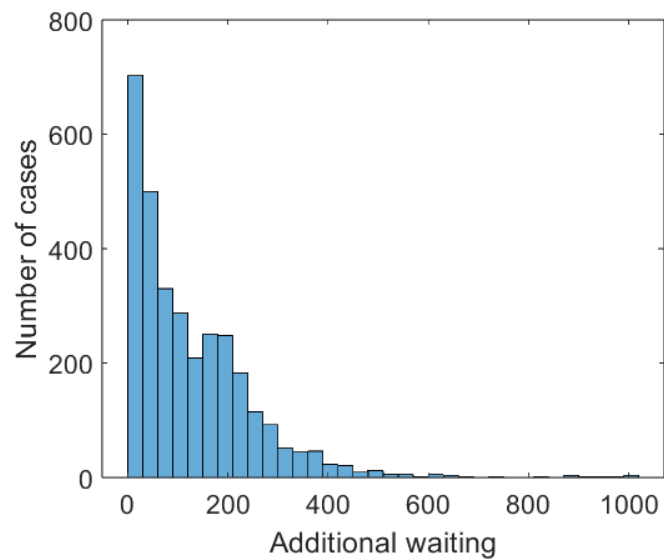


Figure A.5: Additional waiting days distribution

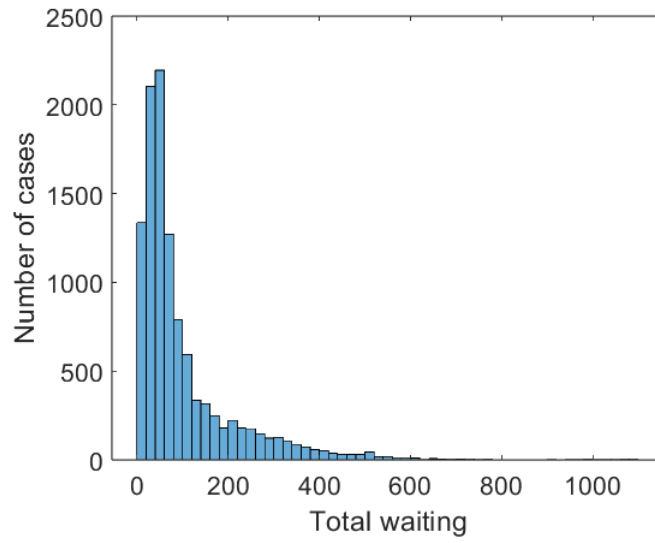


Figure A.6: Total waiting days distribution

A.5. Judges Designations and Sentences

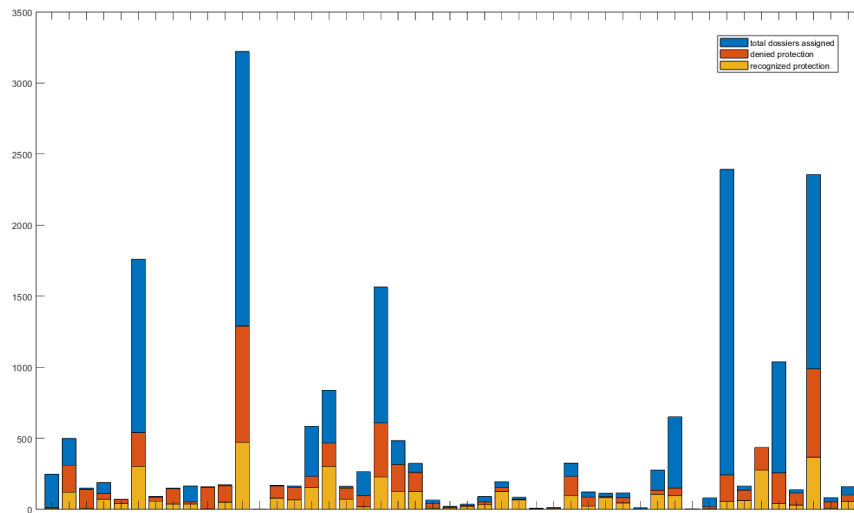


Figure A.7: Number of cases for each judge and their outcomes