



**POLITECNICO
MILANO 1863**

**SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE**

EXECUTIVE SUMMARY OF THE THESIS

SmartRetail Predictive Analytics: Full-Stack Digital Twin and Behavioral Predictive Modeling

LAUREA MAGISTRALE IN MECHANICAL ENGINEERING

Author: LEONARDO DAMIANI

Advisor: PROF. MATTHIAS KLUMPP

Academic year: 2025-2026

1. Introduction

The rapid digital transformation of the retail sector has shifted consumer expectations and increased the need for analytical frameworks capable of interpreting shopping behavior in a detailed and measurable way. Traditional retail analysis often relies on indirect indicators such as sales figures or post-purchase feedback, which reveal the final outcome of a shopping process but not the sequence of decisions that led to it. This thesis investigates consumer behavior through the development of a digital twin-based retail environment. The proposed framework, named *Picklist*, is a web-based simulation platform that reproduces a shopping experience in a controlled virtual setting. Users can browse products, inspect their characteristics, and assemble a cart while their interactions are recorded at a fine level of granularity.

The main objective of the work is to connect digital interaction data with interpretable predictive modeling in order to better understand how consumers evaluate products, navigate categories, and make purchase decisions. More specifically, the thesis combines behavioral logging, survey data, clustering, and scenario-based analysis to bridge consumer psychology and engineering-oriented retail analysis.

2. Methodology

The experimental platform was designed as a full-stack digital twin of a retail shopping environment. *Picklist* allows participants to interact with a virtual product shelf, move across categories, add or remove products from the cart, and receive real-time feedback through a budget-aware interface. The system records low-level behavioral signals such as hover duration, clicks, category changes, cart updates, and overshoot events when the declared budget is exceeded.

A bilingual questionnaire was developed in parallel to complement the interaction logs with self-reported preferences and shopping perceptions. The survey was administered in Italian and English and included sections on general respondent information, product-choice criteria, and shelf perception. This allowed the study to compare declared attitudes with observed behavior and to analyze possible differences across language groups.

The empirical dataset includes 103 unique participants and 13,801 logged interaction events. The resulting data were processed through a structured analytical pipeline that combined descriptive statistics, behavioral modeling, clustering, and policy simulation. This framework made it possible to move from raw interaction

traces to interpretable behavioral profiles and to compare alternative retail strategies in a controlled experimental setting.

3. Results

The exploratory analysis showed that the proposed interface successfully captured meaningful differences in shopping behavior across sessions. Hover-time distributions revealed that most product interactions concentrated around a moderate attention interval, while higher-priced, more discounted, or scarcer products tended to generate longer or more variable evaluation times. These patterns suggest that price pressure, promotional signals, and perceived availability all influence the pace of decision-making.

Before discussing the main behavioral outcomes, it is useful to summarize the scale of the experimental dataset collected through the Picklist platform. The study combines participant-level information with detailed interaction logs, providing a structured basis for both descriptive and predictive analysis.

Summary of sample size and interaction logs

Metric	Value
Number of participants	103
Number of logged events	13,801

Table 1: Summary of sample size and interaction logs collected in the Picklist experiment.

In addition to interaction logs, the questionnaire provided useful insight into the criteria participants considered most relevant during product selection. This helps connect observed behavior with declared preferences and improves the interpretability of the experimental results.

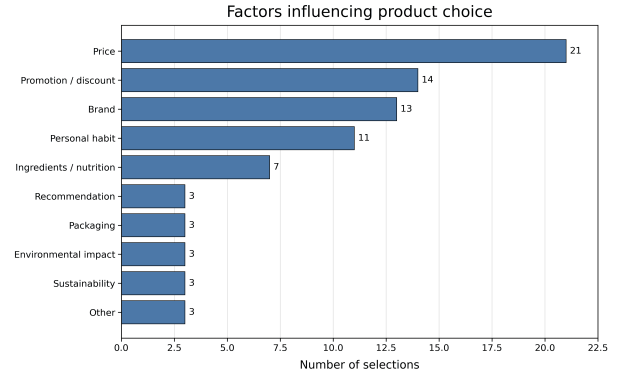


Figure 1: Factors influencing product choice.

The questionnaire results complement the interaction logs by showing that participants rely on a combination of practical and perceptual criteria when selecting products. Price remains the dominant factor, but promotion, brand, and personal habit also play a substantial role. This confirms that product choice in the simulated environment is driven by multiple dimensions rather than by a single attribute alone.

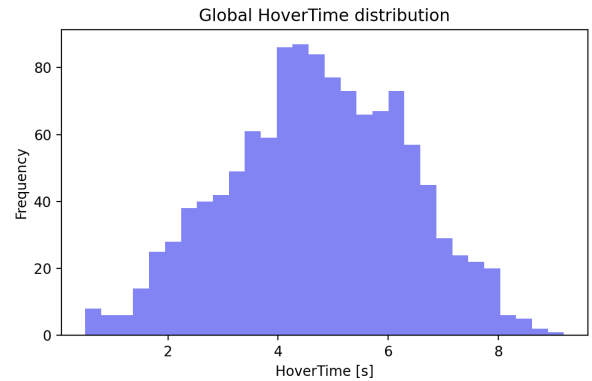


Figure 2: Global distribution of hover time across all product interactions.

The behavioral indicators extracted from the sessions make it possible to describe how users interacted with the platform at a more granular level. Variables such as shopping time, clicks, hover events, cart updates, and category changes offer a quantitative representation of attention, exploration, and purchase-oriented behavior.

Descriptive statistics for behavioral indicators

Variable	Mean
TotalTime (s)	2231.76
n_clicks	15.62
n_hovers	30.67
cart_total	32.47
n_categories	8.62

Table 2: Descriptive statistics for the main behavioral indicators.

Clustering analysis identified three main behavioral profiles. One group was characterized by longer hover times and lower purchase probability, suggesting a more cautious and evaluative behavior. A second group showed stronger responsiveness to discounts, with moderate hover times and higher conversion tendency. A third group occupied a balanced position, reflecting more exploratory behavior. These clusters provided a structured interpretation of the heterogeneity observed in the sessions.

Behavioral profiles identified by clustering

ID	Hover	Purchase	Profile
0	High	Low	Cautious
1	Moderate	High	Discount-driven
2	Moderate	Moderate	Balanced

Table 3: Cluster profiles based on behavioral features.

Clustering analysis identified three main behavioral profiles. One group was characterized by longer hover times and lower purchase probability, suggesting a more cautious and evaluative behavior. A second group showed stronger responsiveness to discounts, with moderate hover times and higher conversion tendency. A third group occupied a balanced position, reflecting more exploratory behavior. These clusters provided a structured interpretation of the heterogeneity observed in the sessions.

Cluster profiles and key attributes (1/2)

ID	HoverTime	PickProb	Price	Discount
0	6.118	0.445	2.249	14.872
1	4.322	0.678	1.614	23.539
2	3.343	0.528	2.390	22.574

Table 4: First part of the compact summary of the behavioral clusters identified in the thesis.

Cluster profiles and key attributes (2/2)

ID	BrandStr	Pack.	Scarcity
0	0.285	0.526	0.556
1	0.638	0.667	0.532
2	0.594	0.270	0.407

Table 5: Second part of the compact summary of the behavioral clusters identified in the thesis.

The cluster-based analysis helps interpret this heterogeneity more clearly. The identified groups correspond to different shopping styles, ranging from cautious and low-conversion behavior to a more discount-sensitive and conversion-oriented profile. This segmentation provides a compact but informative representation of the behavioral structure observed in the sample.

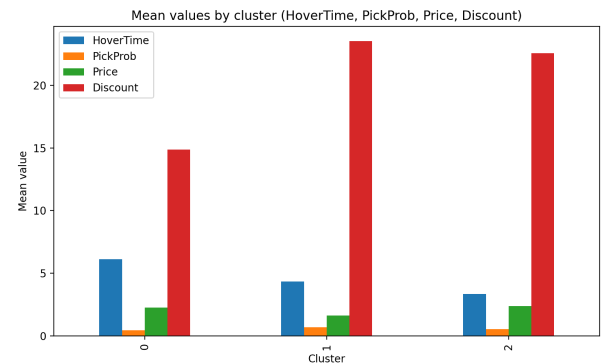


Figure 3: Mean values of hover time, purchase probability, price, and discount across the three identified clusters.

The cluster-level comparison confirms that the observed sessions cannot be reduced to a single average shopping pattern. Instead, the platform reveals distinct combinations of attention, purchase tendency, price level, and discount sensi-

tivity, which support a more nuanced interpretation of consumer behavior.

The policy simulation compared a uniform strategy with a cluster-aware personalized strategy. The personalized configuration produced a higher predicted conversion rate and slightly reduced the average hover time, indicating that adapting product strategy to behavioral profiles can improve both efficiency and conversion potential.

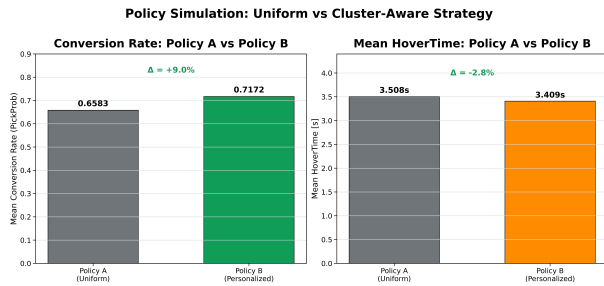


Figure 4: Comparison between uniform and cluster-aware strategies in terms of predicted conversion rate and hover time.

These findings suggest that the proposed framework is not limited to retrospective analysis, but can also support prospective evaluation of retail strategies. By simulating alternative policies on the basis of behavioral segments, the digital twin becomes a practical tool for controlled experimentation before real-world implementation.

Overall, the results support the usefulness of a digital twin approach for studying retail interactions and for testing alternative marketing configurations before physical deployment.

From a practical perspective, the results suggest that the Picklist framework can support scenario testing and early-stage decision support for retail strategies. At the same time, the study remains limited by the use of a web-based experimental setting, by the sample composition, and by the reliance on self-reported survey data alongside behavioral logs. Future developments could extend the platform with richer interaction modalities and a broader participant base to improve external validity and behavioral resolution.

Taken together, these results indicate that the combination of behavioral logging, clustering, and policy simulation offers a coherent and reproducible way to study consumer decision-making. Rather than treating shopping behav-

ior as a uniform process, the analysis shows that it can be decomposed into interpretable components that are useful both for academic understanding and for practical retail decision-making.

In practice, these data can be used to identify distinct customer segments, detect hesitation or promotional sensitivity, and support the design of more targeted retail interventions.

4. Conclusions

This thesis shows that consumer shopping behavior can be studied effectively by combining behavioral logs, survey responses, and synthetic simulation within a digital retail environment. The Picklist platform provided a controlled and scalable setting for observing how users interact with products, budgets, and categories, while also enabling the extraction of predictive and interpretable behavioral patterns.

From a methodological perspective, the work demonstrates that digital twin technology can be adapted beyond industrial systems and applied to retail research as a decision-support tool. From a practical perspective, the framework offers a way to evaluate product arrangement and marketing policies before implementing them in physical stores. In practice, the data can be used to identify distinct customer segments, detect hesitation or promotional sensitivity, and support more targeted retail interventions. From a managerial perspective, the framework provides a decision-support layer that helps retailers compare alternative policies before implementation and allocate attention, promotion, and product placement more effectively. Future developments could extend the platform to additional devices, richer interaction metrics, and more advanced personalization strategies, further strengthening the connection between behavioral data and evidence-based retail decision-making.