Ca' Foscari University of Venice

CUSTOMER CHURN ANALYSIS: EXPLORING MUSEUM CARDHOLDER BEHAVIOR & RENEWAL TRENDS IN TURIN (2014)

PREPARED BY:

Eleonora Fiorentino	899000
Aldo Caumo	866626
Rasmita Shrestha	898598
Meron Kedir Hussen	898703
Dipta Roy	898395

CONTENTS

INTRODUCTION

In an era marked by increasing digitalization and evolving consumer preferences, the analysis of consumer behavior is pivotal for businesses and institutions seeking to optimize their offerings and enhance customer experiences. Museums have embraced membership cards and subscription models more and more to encourage patronage and interaction from art enthusiasts and the public with an interest in culture. One such instance is Turin's museum card program, which has had a notable increase in cardholders over time.

Churn is a common problem faced by businesses across various industries. It refers to the rate at which customers stop doing business with a company. High churn rates can be detrimental to a company's revenue and profitability. Therefore, it is essential for businesses to identify the factors that contribute to churn and take appropriate measures to retain customers.

The objective of this business report is to provide a comprehensive analysis of the data to the churn of customers of a museum cardholder in Turin. The paper aims to provide insights into the characteristics and behaviors of the customer who churn to understand consumer preferences and patterns in museum visits, to evaluate the economic impact of the card program on both the association and museums, and to provide actionable insights for enhancing digital marketing and customer retention strategies.

Therefore, it is crucial to identify the factors that contribute to churn and take appropriate measures to retain customers. This report is an attempt to provide a comprehensive analysis and provide insights into the customer's behavior and help identify the factors that contribute to churn enabling the concerned entities to make well-informed decisions that will improve membership programs, enhance visitor experiences, and encourage recurring business within Turin's thriving cultural scene.

DATASET DESCRIPTION AND EXPLORATIVE DATA ANALYSIS

The datasets under consideration provide a comprehensive view of the museum card program in Turin, serving as a valuable repository of consumer information and usage patterns. In the context of this report, we are presented with three datasets: an13.csv, in13.csv, and data1.csv. These datasets contain valuable information related to a museum card program in Turin, Italy, focusing on cardholders, card usage, and renewal data in 2014.

The an13.csv file provides demographic information about the cardholders, including their age, gender, and occupation. It also includes data on the price paid for the card, which goes entirely to the card association. This information is particularly valuable for businesses seeking to better understand their customer base and tailor their marketing strategies accordingly.

The in13.csv file describes the use of the card for each consumer, including the museum visited, the date of the visit, and the price they would have paid without the card. The file also specifies that the card-association pays each museum 50% of the real price for each visit. This information can be used to gain insights into customer behavior and preferences, as well as to identify patterns and trends in museum visitation.

Finally, the data1.csv file contains renewal data from 2014. This file includes information on whether the cardholder renewed their membership, as well as data on their age, gender, and occupation. This information is particularly valuable for businesses seeking to improve customer retention and loyalty, as it can be used to identify factors that may influence a customer's decision to renew their membership.

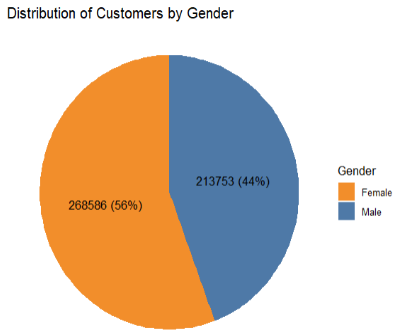
Explanatory Data Analysis

Our analytical approach commences with an initial explanatory analysis of the dataset, primarily focused on discerning the variables that exhibit the highest correlations with churn. This preliminary step aims to furnish a comprehensive understanding of the data from a customer-centric perspective. It involves the exploration of various demographic attributes pertaining to the customers, including their identity, occupation, churn behavior, and geographic location.

This approach enables us to shed light on several fundamental aspects: who the customers are, what they are engaged in, which individuals exhibit a propensity for churn, and where these customers are situated geographically. By delving into these variables, we aim to uncover patterns and insights that will be instrumental in the subsequent stages of our analysis, contributing to a more profound comprehension of customer behavior and facilitating data-driven decision-making processes.

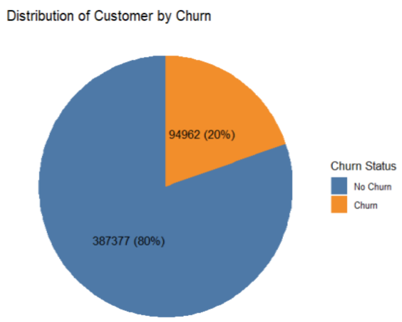
Distribution by Gender

The presented pie chart reveals a dataset comprising 268,568 individuals (56%) categorized as female and 213,753 individuals (44%) characterized as male. Notably, there appears to be a higher likelihood of females engaging in museum visits compared to their male counterparts. This observation is underscored by the higher proportion of females in the dataset, indicating a potentially distinct gender-related pattern in museum visitation behavior.



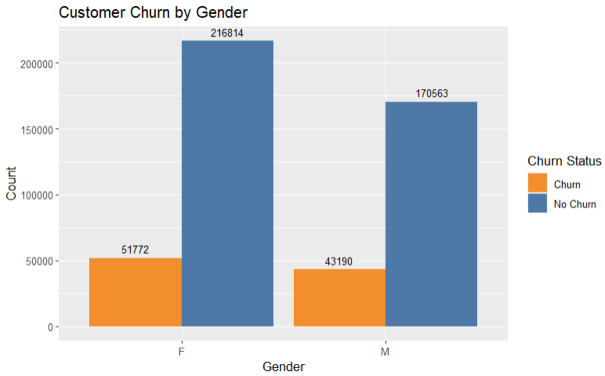
Distribution by Churn

The chart provided offers a marked representation of a visible contrast in the subscription renewal patterns among cardholders in 2014. Notably, a substantial majority, comprising 387,377 individuals, which constitutes 80% of the total, actively opted to extend the validity of their subscriptions. In contrast, a comparatively smaller portion of cardholders, specifically 94,962 individuals, representing 20% of the total, decided not to engage in the renewal process.



This evident distinction between the two groups highlights a significant divergence in subscription renewal behavior. It suggests that a substantial segment of cardholders is inclined towards maintaining their memberships, while a relatively smaller portion may be discontinuing their subscription, emphasizing the need for further exploration and tailored retention strategies.

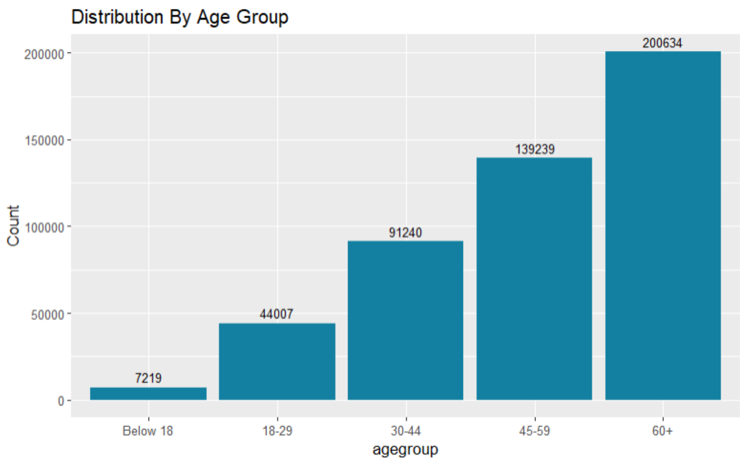
Customer Churn by Gender



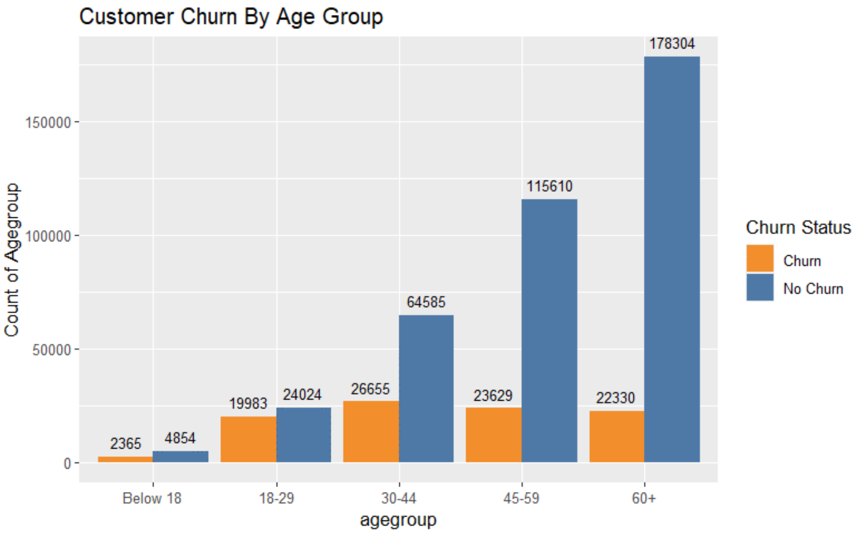
In the left above, we notice that the largest customer segment comprises non-churners, regardless of gender. Females are more prevalent than males in both the churners and non-churners categories. Interestingly, the male churners group exhibits the smallest frequency, possibly indicating factors contributing to churn. This data offers insights for identifying churn drivers and tailoring retention strategies based on gender and churn status.

Age Group

The graph illustrates a positive correlation between age and the customer count, with the number of customers increasing as the age group advances. This observation carries valuable implications for businesses. Analyzing consumer-level data and recognizing age groups with higher customer representation enables companies to craft precise marketing strategies tailored to these demographic segments.

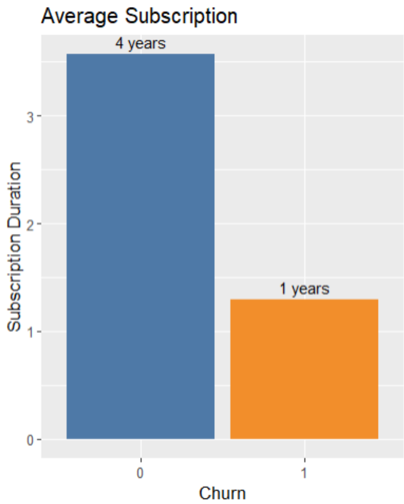


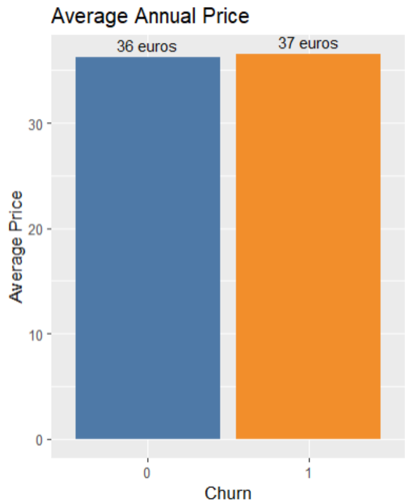
Churn by Age Group



The bar graph, depicting customer churn by age group, offers a clear visual representation of churn percentages within each age category. Notably, the 30-44 age group exhibits the highest churn rate, while the 60+ age group displays the lowest. These findings hold significance for businesses aiming to enhance customer retention and loyalty. By tailoring marketing strategies and retention initiatives to specific age segments, companies can better address the preferences and needs of their customer base. In essence, the bar graph delivers essential insights on the age-churn relationship, guiding strategic decision-making and fostering improved customer retention efforts.

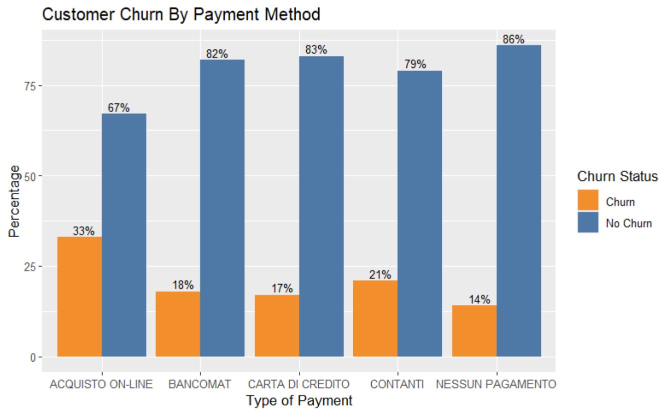
Subscription and Average Annual Price by Churn Status





One of the key findings from our analysis is the relationship between customer churn and subscription rates. We found that customers who did not churn had a higher subscription rate and average annual price compared to those who did churn. This suggests that customers who are more engaged and committed to the service are more likely to continue their subscription and pay a higher price. To illustrate this relationship, we have created a bar graph that displays the subscription and average annual price by churn status. The graph shows that customers who did not churn had a subscription rate of approximately 80% and an average annual price of approximately €36. In contrast, customers who did churn had a subscription rate of approximately 20% and an average annual price of approximately €37.

Churn by Payment Method



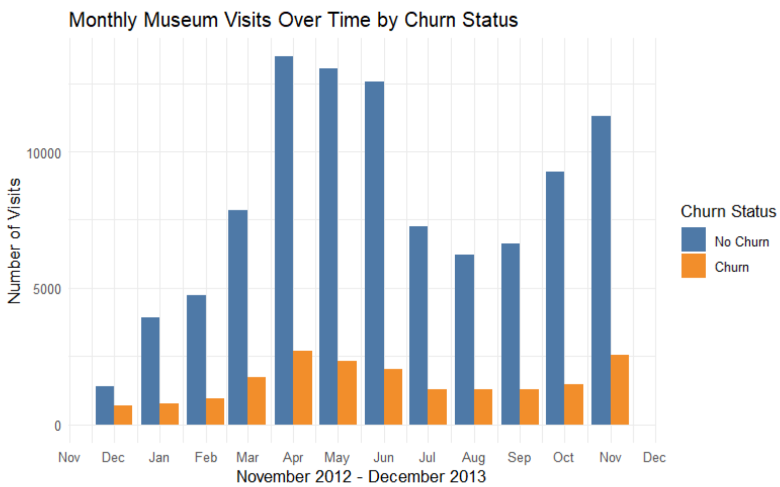
Observing the graph depicted below, it is evident that cash payment method registers the highest number of observations within both churn and non-churn statuses. Following closely is the usage of debit cards. However, when analyzing churn status within each payment method, a distinctive pattern emerges.

Remarkably, online payments exhibit the highest churn rate, with cash payments trailing as the second highest in terms of churn frequency. This insight sheds light on the impact of payment methods on churn status and can inform strategic decisions aimed at improving customer retention.

This finding suggests that the payment method can be an important factor in customer churn. For example, customers who used a payment method that was inconvenient or unreliable may have been more likely to churn compared to those who used a payment method that was easy and reliable since the customers will be distant to pay in cash where there are lots of payment options.

Museum Visits Over Time by Churn Status

As we can see from the graph below our analysis shows that customers who did not churn had a higher number of museum visits compared to those who churned for most months of the year. Even in the month of April 2013, the month with highest number of visits, the churners are responsible for less than 2500 visits.



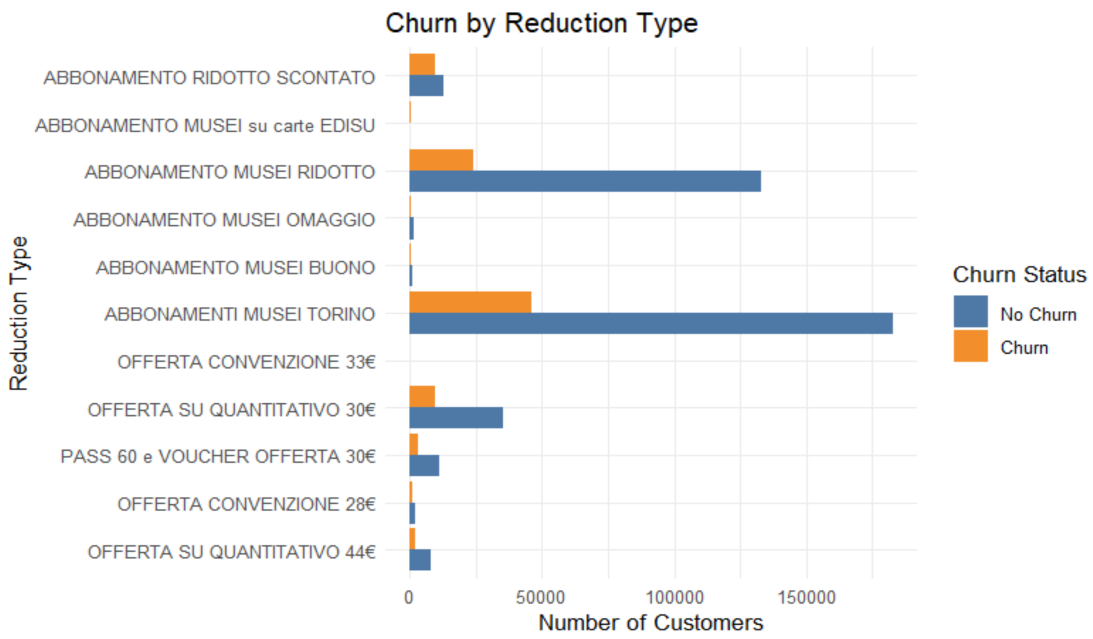
This finding suggests that there may be certain factors that influence customer churn and museum visits over time. Alternatively, customers who did not churn may have been more engaged with the museum during certain months, such as during special exhibits or events.

To improve customer retention and loyalty, businesses can use this information to develop targeted retention strategies. For example, businesses can focus on improving the customer experience during the months where the number of museum visits for customers who churned was higher. This can include offering special promotions or events to encourage customer engagement and reduce churn.

Churn by Reduction Type

Our analysis shows that customers who received a reduction from 'Abbonamento Musei Torino' have more customers who churn compared to the other reduction type. In fact, the high amount of non-churn customers has the same reduction type followed by "Abbonamento Musie Ridotto".

This finding suggests that offering higher discount promotions can be an effective strategy for reducing customer churn. For example, businesses can offer personalized promotions or discounts to customers who are at risk of churning, which can increase customer engagement and reduce churn.



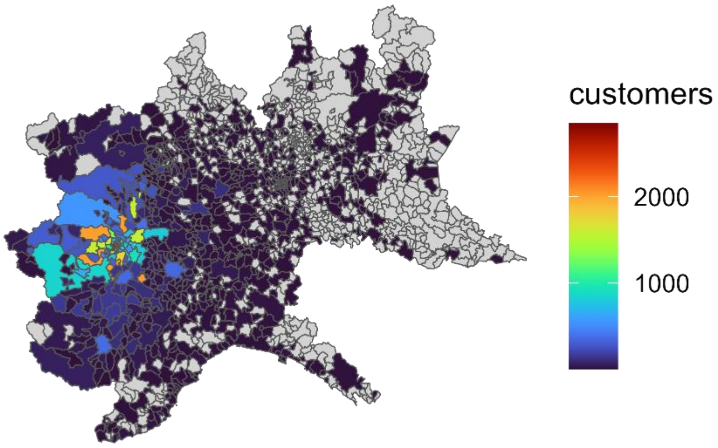
Based on this, we can provide valuable insights into the relationship between churn and discount promotions. By using this information to inform strategic decision-making, businesses can improve customer retention efforts and increase customer loyalty.

Distribution by Location

In this dedicated section of our analysis, we've taken a close look at our museum cardholders and where they're located. Using clear visual representations, we've mapped out where our cardholders live in Turin and the broader northwest Italy region.

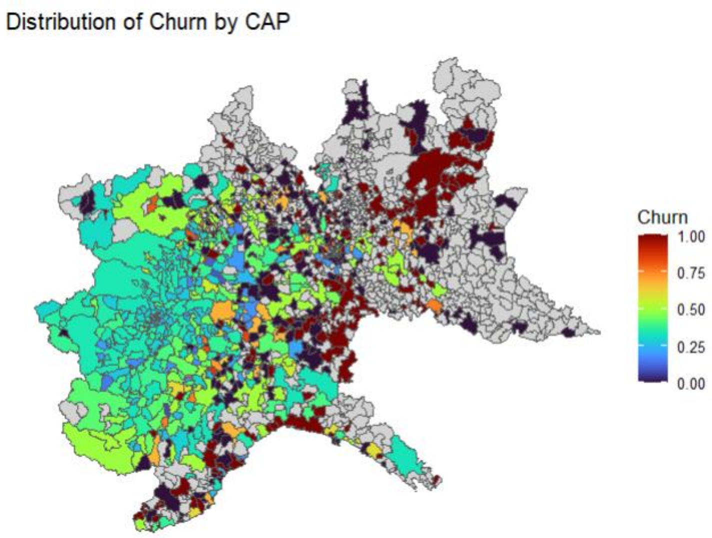
This geographic viewpoint gives us important insights into where our cardholders are situated. It goes beyond just knowing who they are and digs into the specifics of where they call home and where they visit.

The map shows a notable geographical distribution of customers, primarily concentrated in the northwestern region of Italy, with relatively fewer customers in the eastern areas. The city of Turin emerges as the focal point, boasting a significant concentration of cardholders, with each commune within the city hosting more than a thousand customers. This spatial insight underscores the prominence of Turin as a hub for museum cardholders and elucidates the regional variation in customer distribution.



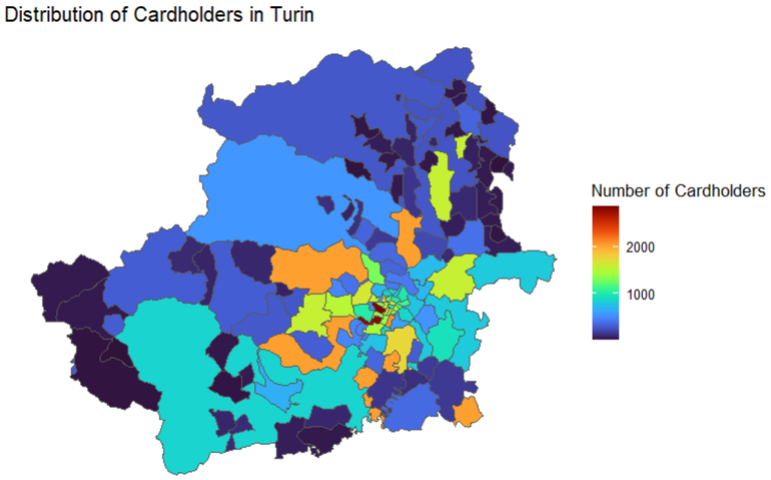
Distribution of Churn by Location

Now, a visible pattern emerges, indicating that a substantial proportion of customers who churn predominantly originate from the outer regions of northwest Italy, particularly those residing outside the city of Turin. This spatial insight underscores the significance of geographic factors in customer churn and highlights a potential area of focus for retention strategies.

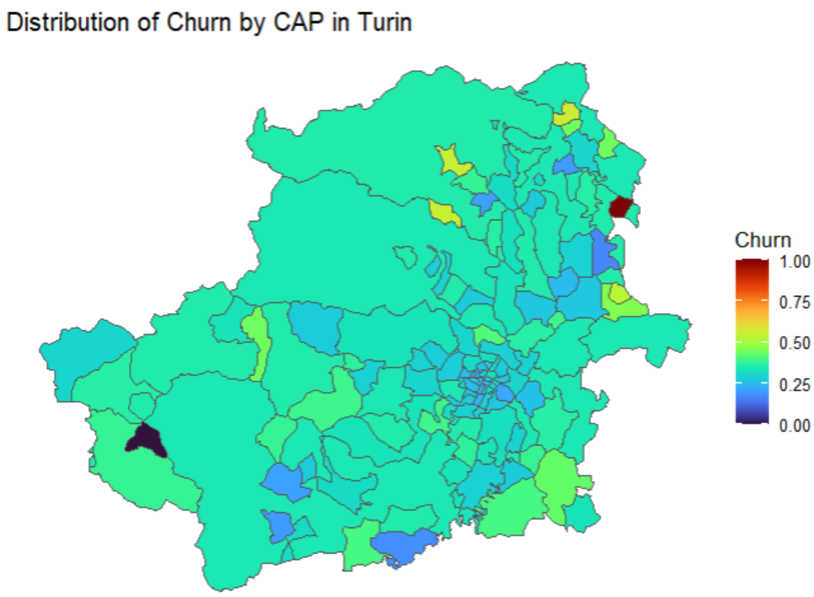


Distribution in Turin

In this refined map, a concentrated depiction of cardholder distribution within Turin comes to the forefront. Notably, communes situated in the central region of Turin showcase a robust cardholder presence, each boasting more than a thousand cardholders. In contrast, the northern part of Turin reveals a lower concentration of cardholders, with each commune having fewer than a thousand cardholders, as prominently illustrated in the map provided. This spatial insight underscores the varying intensity of cardholder presence within different areas of Turin, offering valuable context for targeted strategies and initiatives.



Distribution of Churn in Turin

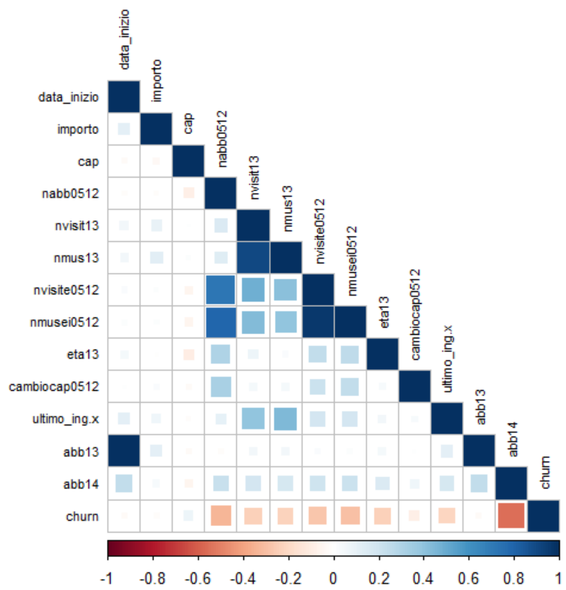


The presented map offers a focused perspective on the distribution of churn within Turin. Especially, a visible pattern emerges, revealing that the majority of communes exhibit lower churn rates. Only two communes, situated in the northeast and southwest regions of Turin, display notably higher churn rates, as can be clearly observed. This spatial insight underscores the variation in churn rates across different areas within Turin, providing valuable information for targeted analysis and retention strategies.

SECTION 2: VARIABLE SELECTION

Effective variable selection is a pivotal step in building predictive models. In our case, where the dataset encompasses both numerical and categorical variables, we employed a systematic approach to identify the most influential predictors.

For numerical variables, we initiated the process by using a correlation matrix to discern significant relationships. Subsequently, we applied advanced techniques like MARS and Boruta to further refine our selection, focusing on those variables with the highest predictive power.



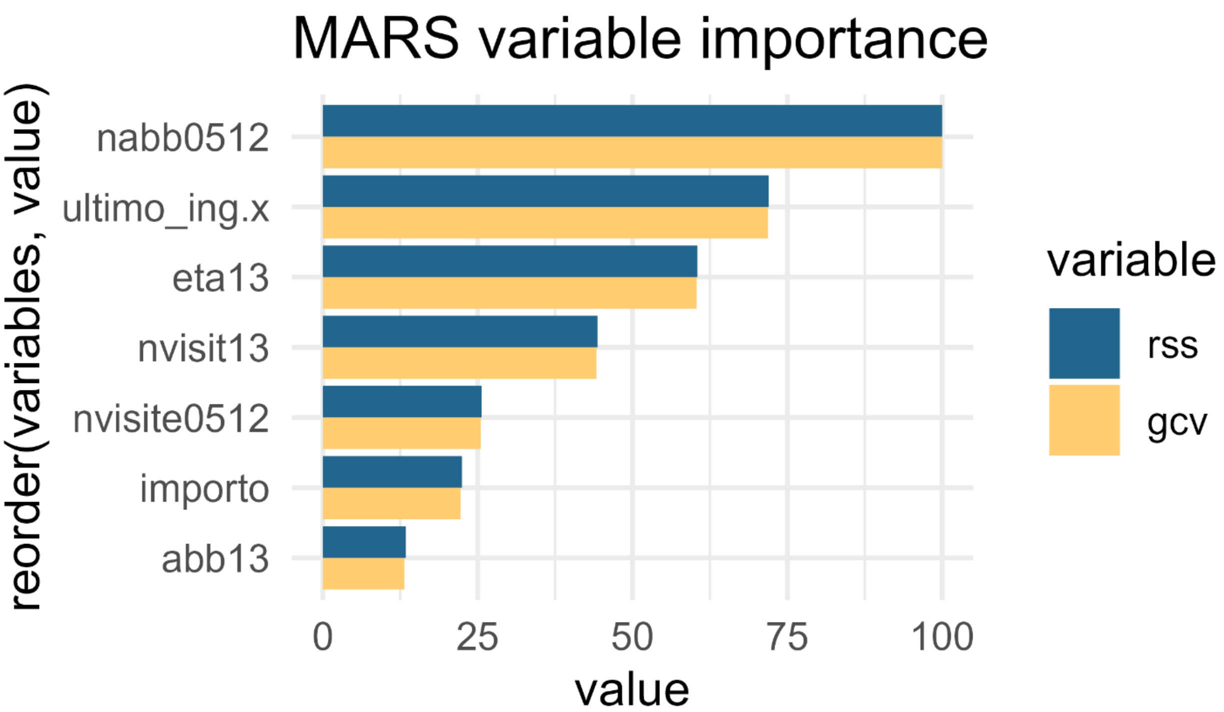
Concurrently, for categorical variables, we harnessed the power of Information Value and Weight of Evidence. This dual-pronged strategy allowed us to comprehensively explore the significance of our predictors.

We conducted a comprehensive correlation analysis among all numeric variables, with the aim of exploring potential relationships between these variables. The correlation matrix presented below provides insightful observations. It becomes apparent that the variable "churn" demonstrates a negative correlation with all other numeric variables, except for "importo," with which it does not exhibit any noticeable correlation. This analysis underscores the existence of distinct relationships between the churn status and various numeric attributes, shedding light on the influence of these factors on customer churn.

MARS

MARS models generated using the "earth::earth()" function in R come equipped with a built-in backward elimination feature selection routine. This routine assesses the impact of adding each predictor variable to the model by monitoring reductions in the Generalized Cross-Validation (GCV) estimate of error. The cumulative reduction in the GCV error estimate serves as the variable importance measure, denoted as 'gcv'. Importantly, MARS models perform automated feature selection during the pruning process, making it a dynamic process.

In this context, if a predictor variable is never utilized in any of the MARS basis functions in the final model, following pruning, it receives an importance value of zero. Additionally, an alternative approach is to monitor the change in Residual Sums of Squares (RSS) as terms are introduced ('rss'). It's worth noting that there's often little distinction between these two methods in practice.



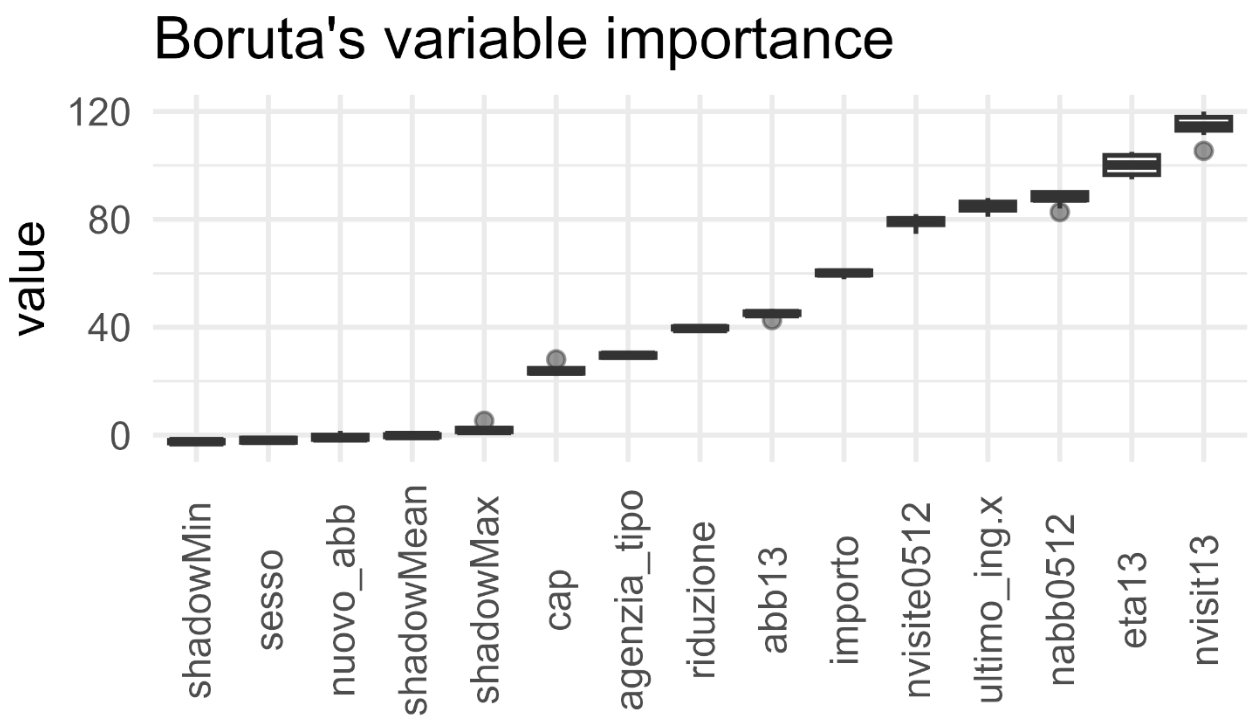
For our variable importance plot, we have chosen to use the GCV and RSS metrics as measures of variable importance. These choices are aligned with common practices and provide a balanced view of how individual predictor variables contribute to the overall predictive power of our MARS model.

From this figure we see that “nabb0512”, “ultimo_ing.x”, “eta13” and “nvisit13” are the four of the influential variables with importance value of more than 80.

Boruta

Boruta is a feature selection technique used in machine learning and data science to identify and select important variables from a set of predictors. It operates by creating a shadow dataset, shuffling predictor variables, and training a machine learning model on both the actual and shadow variables. The model evaluates the significance of each feature by comparing their variable importance scores with the highest score among their shadow counterparts. Features with higher importance scores than their shadow counterparts are considered significant and retained for analysis or modeling, while non-significant features are discarded. Boruta is particularly valuable for reducing overfitting and improving model performance when dealing with datasets containing many potential predictors, ensuring that only the most relevant variables are considered in the analysis.

It uses boxplots to show the distribution of features’ importance over Boruta run, using colors to mark final decision; it also draws boxplots for the importance of worst, average and best shadow in each iteration. As can be seen from the figure, “nvist13” feature has the highest importance followed by “eta13”.



Information Value and Weight of Evidence

Weight of Evidence (WoE) and Information Value (IV) are essential tools for assessing the predictive power of independent variables in the context of feature selection. WoE provides a comprehensive understanding of how specific classes or categories within an independent variable influence the likelihood of a particular outcome, such as 'churn' and 'non-churn' in customer retention analysis. By quantifying the distribution of these classes, WoE helps differentiate between categories that are more strongly associated with favorable or unfavorable outcomes, enabling the selection of the most impactful features.

Information Value (IV), on the other hand, serves as a global metric to evaluate the overall predictive strength of an independent variable. IV considers both the WoE and the proportion of cases within each category, providing a consolidated measure of the variable's ability to discriminate between outcomes. A high IV suggests that the variable contains valuable information for prediction, making it a prime candidate for feature selection.

VARS	IV	STRENGTH
RIDUZIONE	0.109	Highly Predictive
CAP	0.091	Somewhat Predictive
AGENZIA_TIPO	0.025	Not Predictive
SESSO	0.000	Not Predictive
NUOVO_ABB	0.000	Not Predictive

SECTION 3: MODELS AND PROFIT MAXIMIZATION

We now come to the crucial part of our analysis, where we will use statistical models to create a marketing and customer retention campaign that maximizes profits for the Turin Card company.

Previously, the examination of the datasets produced a few significant conclusions that provided important new information on customer behavior and the success of Turin's museum card program. The cardholders' demographics and preferences became more nuancedly understood, laying the groundwork for customized services and targeted marketing campaigns aimed at maximizing visitor engagement. Additionally, the analysis of the churn dynamics revealed important variables that affect cardholder retention, highlighting the significance of value-driven membership incentives and tailored experiences. The results of this study have important ramifications for managers of museums and other cultural organizations. They will help them devise strategies for keeping members engaged and improve visitor experiences so that they will be more likely to return to Turin's thriving cultural landscape over time.

The Models

1. Support Vector Machines (SVM)

The first model that we implemented is the Support Vector Machine (SVM). SVMs are a class of supervised machine learning models that have proven effective in a wide range of applications, including churn prediction. SVM excels at classification tasks, where the goal is to separate data into distinct categories based on their characteristics. It operates by identifying a decision boundary (known as a hyperplane) that maximizes the margin between different classes of data points.

SVM is particularly valuable for churn prediction because it can handle complex, high-dimensional data and is robust in scenarios with non-linear relationships. By finding the optimal hyperplane that maximizes the separation between customer groups (e.g., churners and non-churners), SVM can identify potential churners with high accuracy. This, in turn, enables businesses to take proactive measures to retain customers and reduce churn rates. SVM's versatility, adaptability to various data types, and ability to deal with both linear and non-linear relationships make it a formidable tool for predictive modeling and data-driven decision-making.

2. C5

The second model that we implemented is C5.0. It is a robust and popular machine learning method primarily designed for classification tasks. It specializes in constructing decision trees, which are graphical structures used to make decisions or predictions based on input features. C5.0 excels in data analysis and predictive modeling by creating highly interpretable and efficient decision trees.

The algorithm operates by recursively splitting the dataset into branches based on the most informative features. It selects the best attributes and thresholds to partition the data, optimizing for purity within each subgroup. C5.0 is particularly valuable in scenarios where understanding the decision-making process is essential, such as churn prediction, where it can identify key factors contributing to customer attrition.

C5.0's decision trees offer transparency, allowing users to follow and interpret the logic behind classification decisions. This makes it a favored tool for businesses aiming to uncover the drivers of churn, develop predictive models, and devise targeted retention strategies.

3. XGBOOST

Lastly, we implemented Extreme Gradient Boosting to predict the model. XGBoost is a state-of-the-art machine learning algorithm renowned for its exceptional predictive power. It belongs to the ensemble learning family, where multiple models are combined to form a more robust and accurate final model. What sets XGBoost apart is its ability to handle a wide range of data types, manage high-dimensional feature spaces, and excel in both classification and regression tasks.

XGBoost operates by iteratively training weak learners, typically decision trees, and improving upon their predictive performance in each iteration. It achieves this by optimizing a loss function, effectively minimizing prediction errors. The 'gradient boosting' approach, which forms the basis of XGBoost, places a strong emphasis on model interpretability and feature importance, making it invaluable for churn prediction.

Benefit function and profit maximization

Let's now get to the final part. In this section we will present the marketing campaign and the benefit function that we implemented to maximize the profits. At the end, we will present the best model and threshold and our results and comments.

1. Marketing campaign

The Turin's card society is planning a direct marketing campaign with a budget of 5.000 euro. This campaign consists of two possible marketing choices: the first is to call the customers we think might churn in order to make them an appealing offer, the second is to contact them through e-mail. Those two options have different costs and different response rates:

- The phone call costs 1 euro (line and operator) and has a response rate (based on past campaigns) of 35%.
- The e-mail costs 0,15 cents (emailing and operator) and has a response rate of 15%.

2. Benefit function

We planned a promotional campaign oriented specifically at the customers predicted to churn. The profitability of the campaign is measured in 'benefit', which is some function of profitability. In fact, the scope of our promotional campaign is to achieve the highest benefit value. Subsequently, we have identified costs and benefits associated with each category of customers. The promotional campaign is highly effective at reducing churn but also a very expensive one. So, it is important that the promo is presented ONLY to customers predicted to have churn-intent. There is a positive net benefit if churn-intent customers respond positively to the churn but there is also a cost (negative benefit) if churn-non-intent customers that are presented the promo sign up to the promo.

Assumptions:

- 'tp' represents the number of customers that were correctly predicted as intending to churn
- 'tn' represents the number of customers that were correctly predicted as not intending to churn
- 'fn' represents the number of customers that were incorrectly predicted as not intending to churn but actually did churn
- 'fp' represents the number of customers that were incorrectly predicted as intending to churn but never churned or intended to either

- Each churn-intent customer that is predicted to churn but responds positively to the campaign and ends up not churning subsequently contributes to gain in benefit of 25€ (tp.bnft) over the next year.
- Each churn-intent customer that eventually churns contributes to a loss of benefit of - 60€ (CLV) due lost revenue over the next 3 years (the explanation of why we count 3 years will be provided in the next paragraph about CLV) and marketing costs to procure that customer back in the future
- Each churn-non-intent customer predicted incorrectly to have churn-intent that signs up to the promotional campaign contributes to a loss of benefit of -10€ (fp.cost) relative to the positive benefit that they would have contributed anyway if they had not been presented the promo
- Each churn-non-intent customer will continue to contribute a net positive benefit of 35€ over the next year
- 70% of churn-intent customers presented the promo will respond positively while the rest will churn (attrition rate: 30%)
- 100% of churn-intent customers not presented the promo will churn
- Assumption: 100% of churn-non-intent customers presented the promo will sign up to the promo
- 100% of churn-non-intent customers will continue to be customers

These assumptions are captured in the 'benefit' function. The function takes the confusion matrix output (tp, fn, fp, tn) as input and returns the computed benefit based on the assumptions.

3. CLV calculation

In order to calculate the Customer Lifetime Value for our benefit function, we calculated 3 different values: first of all, we discovered how much in mean a customer would spend in one year in museum's tickets if he wouldn't have the card. We discovered that it would be around 30 euros. So, the price for the association to pay in mean for each customer annually is about 15 euros (note that the card-association pays each museum 50% of the real price for each visit).

After calculating the costs, we focused on the revenues: we calculated in mean how much a customer spends for an annual subscription, and we discovered that it's around 35 euros.

Finally, we get the Customer Value by taking the revenues and subtracting the costs, obtaining a CV equal to 20 euros per year.

Anyway, this value is not considering the 'lifetime value' of a customer. To solve this issue, we calculated the average age of the clients and subtract it to the average lifetime in Italy to get the mean remaining lifetime for a customer. This way the CLV would results in around 760 euros.

However, we decided to include in our function not the CLV, but rather an annual CV multiplied by the years passed between this marketing campaign and the next one, which will be in 3 years. That's why we considered an indicative value of 60 euros per customer as CLV.

4. Attrition rate

The attrition rate is a metric used to measure customers lost over a specific period of time.

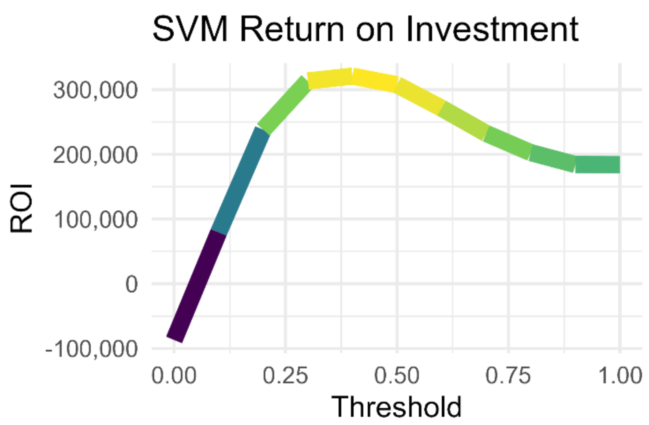
In our case, we calculated the attrition rate as the number of people who churned in 2014 divided by the total number of people who had a subscription during year 2013. We obtained this way an attrition rate equal to 0.3 (30% of customers churned the program during that year).

Model's Results

SVM results

CUTOFF	CALL COSTS	MAIL COSTS	CALL REVENUES	MAIL REVENUES	ROI	REMAINING BUDGET
0	-15063	-2259.45	-63775.125	-23136.075	-86911.2	-12322.45
0.1	-12536	-1880.4	22906.5	56412.1	79318.6	-9416.4
0.2	-9181	-1377.15	106916.25	131113.1	238029.35	-5558.15
0.3	-6419	-962.85	148327.5	165005.65	313333.15	-2381.85
0.4	-4339	-650.85	154842.125	165986.775	**320828.9**	**10.15**
0.5	-2886	-432.9	150108.625	157436.225	307544.85	1681.1
0.6	-1676	-251.4	133571.5	137786.1	271357.6	3072.6
0.7	-762	-114.3	115217.5	117111.2	232328.7	4123.7
0.8	-263	-39.45	100988.25	101636.8	202625.05	4697.55
0.9	-2	-0.3	92122.75	92127.45	184250.2	4997.7
1	0	0	92010	92010	184020	5000

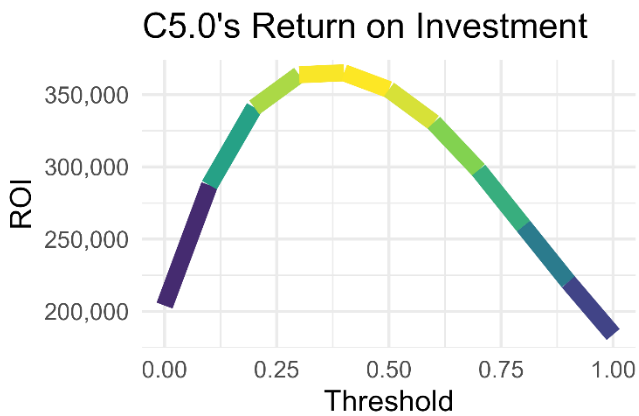
The analysis of the results indicates that, for the SVM model, the optimal threshold is determined to be 0.4. At this threshold, the model yields a Return on Investment (ROI) of €514,040, with a remaining budget of €363. This specific threshold value demonstrates the point at which the model's performance is maximized, resulting in a favorable financial outcome.



C5.0

CUTOFF	CALL COSTS	MAIL COSTS	CALL REVENUE	MAIL REVENUE	ROI	REMAINING BUDGET
0	-10725	-1608.75	87825.625	116204.375	204030	-7333.75
0.1	-8986	-1347.9	131876.25	155427.35	287303.6	-5333.9
0.2	-7313	-1096.95	161047.875	180023.425	341071.3	-3409.95
0.3	-5918	-887.7	174253	189471.3	363724.3	-1805.7
0.4	-4705	-705.75	176530	188529.25	365059.25	-410.75
0.45	-4216	-632.40	176287.00	186994.60	**363281.6**	**151.60**
0.5	-3579	-536.85	172206.625	181251.275	353457.9	884.15
0.6	-2576	-386.4	162066	168511.6	330577.6	2037.6
0.7	-1746	-261.9	146721.75	151052.85	297774.6	2992.1
0.8	-1034	-155.1	128250.5	130795.4	259045.9	3810.9
0.9	-464	-69.6	109635.125	110770.025	220405.15	4466.4
1	0	0	92010	92010	184020	5000

In our second model, C5.0, the optimal threshold is identified as 0.45. At this threshold, the model achieves a Return on Investment (ROI) of €569,483, with a remaining budget of €151. This threshold value is indicative of the point where the model's performance is maximized, resulting in a favorable financial outcome.

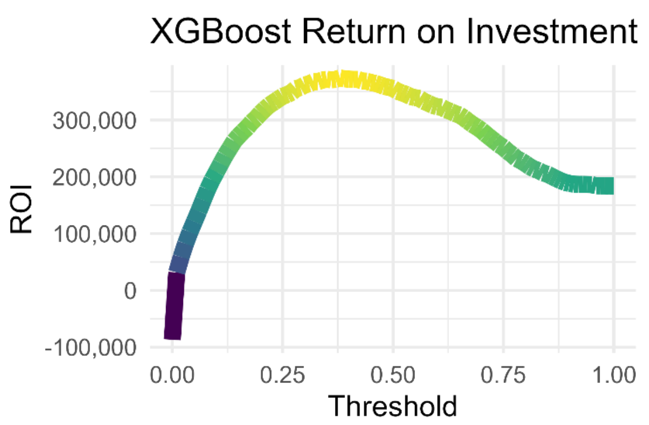


XGboost

CUTOFF	CALL COSTS	MAIL COSTS	CALL REVENUE	MAIL REVENUE	ROI	REMAINING BUDGET
0	-15063	-2259.45	-63775.125	-23136.075	-86911.2	-12322.45
0.1	-10855	-1628.25	89977.125	118688.375	208665.5	-7483.25
0.2	-8678	-1301.7	144713.375	167383.175	312096.55	-4979.7
0.3	-6699	-1004.85	172261.875	189555.525	361817.4	-2703.85
0.4	-4874	-731.1	179729.625	192159.025	371888.65	-605.1
0.44	-4213	-631.95	178802.375	189488.93	**368291.30**	**155.05**
0.5	-3404	-510.6	173558.375	182133.275	355691.65	1085.4
0.6	-2307	-346.05	158885.5	164636.45	323521.95	2346.95
0.7	-1352	-202.8	139370.625	142698.325	282068.95	3445.2
0.8	-476	-71.4	110982.75	112142.35	223125.1	4452.6
0.9	-42	-6.3	94090.125	94190.325	188280.45	4951.7
1	0	0	92010	92010	184020	5000

Finally, for the XGBOOST model the best threshold is 0.44, which is associated to a ROI of 576060€ and a remaining budget of 155€.

This model has turned out to be the best one at maximizing profits for our specific situation, making us gather around 60k € more than the SVM model and around 7k € more than the C5.0, without exceeding the budget.



CONCLUSION

The thorough examination of the Turin museum card program data has yielded priceless insights into customer behavior, use trends, and churn dynamics. These insights highlight the importance of customized membership strategies and individualized experiences in promoting recurring business. The in-depth analysis of the cards' demographics and preferences has brought attention to the variety of the visitor population and underscored the necessity of focused engagement programs that accommodate the wide range of interests and preferences of museum visitors.

Additionally, the examination of the churn dynamics has highlighted the crucial role that customer satisfaction and perceived value play in influencing cardholder retention. This has advocated for ongoing efforts to improve the overall visitor experience and maximize membership benefits to guarantee sustained program participation.

The study's conclusions have broad interesting results for managers of the Turin card program, helping them to develop data-driven plans that maximize long-term profits and participation. The churn managment issue is a complex task but it should be address by every society in order to keep high performances.