This report contains the answers to the open-ended questions as well as an analysis on the dataset provided.

1. **What would be the Key Performance Indicators you would come up as the most important to monitor a credit card business? How often would you suggest such indicators must be monitored?**

Since the product was launched recently, the first key performance indicator that we need to look at is the response rate. This metric is calculated dividing the number of responses by the number of invitations sent. The overall response rate will give us an understanding on whether the customers are sending the application back to us. Also, this data provides important information about how effective the strategy is or if we need to modify it. The next key performance indicator to analyze is the rate of approval, this number is calculated dividing the number of approved accounts by the number of customers who applied to the card. With this metric we would be able to assess the process of acceptance. We could find insights on how rigorous the acceptance process is or find patterns on what the customers accepted have in common.

Furthermore, there are two important metrics that can be calculated with the total price of the campaign. The cost of each account and the cost of each active account. We calculate these metrics by dividing the total cost of the campaign by the number of new accounts and the second metric is obtained dividing by the number of active accounts. This key performance indicator is relevant because we can obtain information on new accounts that have not started using their card or have not activated it. In order to make revenue, we need to make sure that the customers will be actively using their new cards.

Finally, the opinion of the customer is highly valuable to provide a satisfactory service. It is important to have insights on a product that has been recently launched. Customers can give valuable feedback on things such as: the application process, the card delivery process, and first impressions of the service the customer has just acquired.

Due to the product being so new, these indicators must be monitored on a weekly basis. It is expected that there will be a lot of responses received, applications processed, and cards delivered every day. Therefore, tracking the weekly progress provides insights on how the product is performing and how the company is performing with this new service. We can find information on what day of the week we process the most applications, how long does it take to approve or reject a customer and what day of the week are most cards delivered.

2. **Propose a problem resolution strategy with the stakeholders. How would you deal with this issue? Which facts would you present?**

By definition, a customer that is traditionally considered 'dormant' has not done any transactions during a set period, usually between 4 to 6 months. This is the main issue that we have with disagreeing stakeholders right now. My solution would be to prove why this cannot be the only requirement to place a customer under the dormant category. In order to make a data-driven decision, we need to look at other aspects of the data that we have about our customers. Then, we will be able to find more fields in the data that contribute to a client being dormant. For example: how frequent does the customer use the card, number of transactions per month prior to the dormant period, average cost of the transactions, the amount of time in months or years since the customer uses our services. Consequently, we could find clients that make only two or three transactions per year, but these transactions can be higher than the average of all transactions, this would give the company a high revenue. We could also find clients that do not use their card because of an elevated interest rate, or the amount of credit granted to the customer is too low. These two reasons could make the customer use other services that are more beneficial to them.

The solution to this issue would be to present the stakeholders with more information about our customers. Showing them that there are other facts that also contribute to a customer being considered dormant. This will allow us to find other type of valuable information to gain insight on whether the customer is dormant based on usage trends and patterns. Based on our customer data, we could present recommendations from the analysis and be able to find at least two more characteristics that contribute to a client being dormant. The next part of the solution would be to come up with a plan to decrease the number of dormant clients, we want those clients to start using their credit card more often. It is a well-known fact that acquiring a new customer can cost five times more than retaining an existing customer. Since existing customers are familiar with our service, the success rate is higher when selling to an already existing customer. Based on the data we can also recommend a strategy to prevent clients becoming dormant by offering them a lower interest rate or increasing their credit amount. The issue of deciding how long until a customer is considered dormant should evolve to how can we prevent our existing customers to become dormant.

3. **What should we do to centralize the data in order to display it in charts for KPI monitoring? What would you propose the data governance strategy should be?**

The most important tool to perform successful business intelligence analytics is a Data Warehouse. This is one of the many ways data can be managed in a company. A Data Warehouse contains extensive amounts of historical data, it is specifically designed to perform queries and create analytics reports. Another important characteristic of a Data Warehouse is that it is designed to obtain the data from unrelated systems. Therefore, the data it contains comes from a variety of sources, such as data generated from the app, data retrieved from the credit bureau, a customer database, and the system that holds all the payments information.

The main characteristic of a Data Warehouse is that it is designed to organize a lot of data and centralize it. Since a Data Warehouse contains big amounts of historical data, it is considered the "single source of truth" of the data in a company. The Data Warehouse is composed of a Relational Database Management System (RDBMS) where the data is stored and managed. It also contains an Extraction, Transformation and Loading (ETL) workflow that prepares the data for analysis. Having a record of historical data allows the data warehouse to provide capabilities like reporting, statistical analysis and data mining. Furthermore, the Data Warehouse can connect to tools specialized for analysis that allow to create visualizations, which can be used to create reports and present data to business users.

The author Bill Inmon, the father of data warehousing, defines it as: "A data warehouse is a subject-oriented, integrated, time-variant, non-volatile collection of data in support of management decisions" (Building the Data Warehouse, 4th Edition, by W.H. Inmon). Moreover, he describes a Data Warehouse as "the central point of data integration for business intelligence and is the source of data for the data marts, delivering a common view of enterprise data" (Building the Data Warehouse, 2nd Edition, by W.H. Inmon).

Therefore, a data warehouse has four important benefits: Subject-oriented, from the business user's point of view, data about a specific area can be gathered. Integrated means that it can combine and cleanse data from different sources and organize it into relational tables. Time-variant gives each data value an associated time, meaning that we are capturing data as a snapshot in time. Lastly, a Data-Warehouse is non-volatile, not real time, so new data is a supplement rather than a replacement.

Lastly, since a data warehouse is designed to perform fast queries that deliver data for the user to specify the characteristics of the data needed to perform analysis. Then we must have an efficient data governance strategy where we have consistent names and logic for column names across all datasets, rely on relational database modeling, having clear star-schemas that provide an organized and reliable data source.

4. Data Analysis Report

Please use this link to look at the Git Repo, which contains the Data Analysis process in more detail, and it shows each step used to perform the calculations and create the graphs:
https://github.com/sasdiego5/Rappi_BI_Case

Even though it was recommended to upload the dataset to a SQL db and perform queries, I chose instead to create a Jupyter Notebook, and with the use of Python techniques for data analytics, I performed some data processing and analysis. Although I'm experienced with advanced SQL queries and relational databases, I believe that Python has more capabilities to perform data exploration and business intelligence analysis for a small project.

The dataset is composed of 12,755 records. These records contain information on customers that were sent an invitation to apply for a credit card. There may be more than one record per customer, depending on whether there was a response to the invitation. The STATUS column provides vital information on the type of customer record. With the help of the STATUS column, it was found that the dataset is composed of 29% records on customer transactions, 18% records on customers who responded, as well as 18% records that were checked on the risk model, 11% records on customers that were approved, 8% of records with a delivery status, 8% records of customers who did not respond, and 6% of records on rejected customers. See pie chart below for the type of records that compose the dataset.
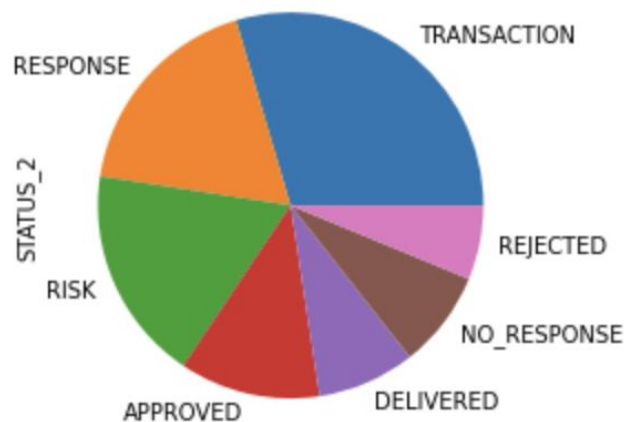
In order to have a more accurate analysis, I decided to delete 8% of the records (1042) that contained information on customers who did not respond, these rows were mostly null values across all fields, they provide no valuable information. Now we can be certain that each customer can have one or more records on transactions, and one record on: response, risk, rejected or approved and delivered.

Next, the first key performance indicator was calculated, out of 3,341invitations sent, 2,298 customers responded, this gives us a Response Rate of 68.78%, about one third of the customers did not respond to the invitation sent. Therefore, it is recommended that this number must increase, we want more customers to act and respond. Maybe the invitations sent to the customers could be more persuasive.

See Response vs No Response bar plot below:
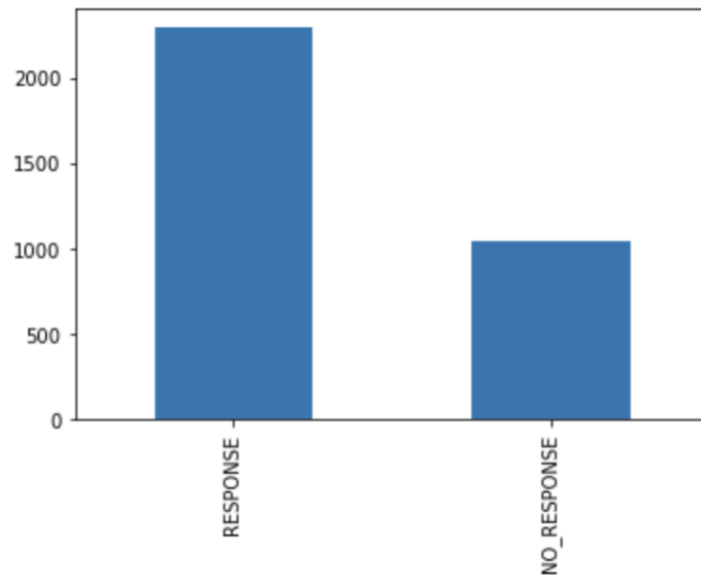
The following KPI we can look at is the Approval Rate, there were 2,298 applications received with 1,505 customers who were approved, this gives us a 65.49% Approval Rate. Since about one third of customers who apply are rejected, it seems like our risk model is somewhat rigorous when it comes to deciding if a customer is fit for a card or not. We can take a look at what do the rejected customers have in common and perform some changes based on those results.

See approval and rejection rates below:
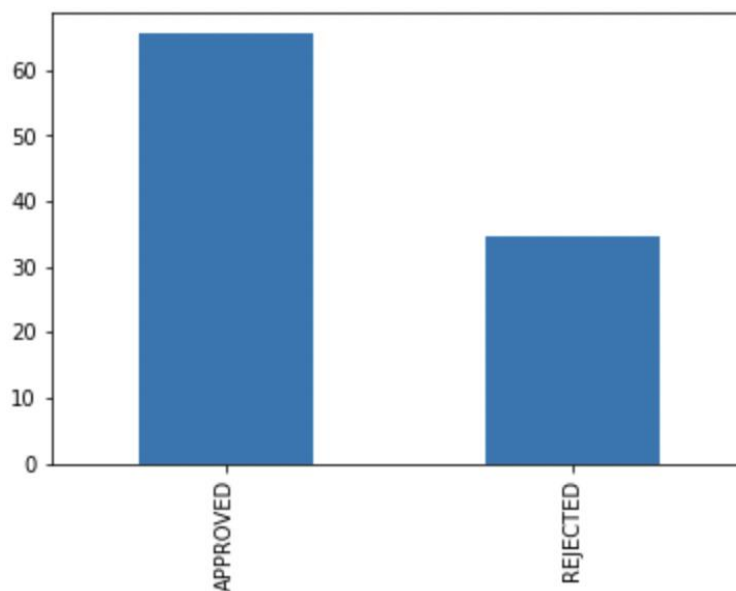
We can now do some analysis on Approved and Rejected customers. The data shows that the two most common motives to reject a customer were: MOP and Usage, more than half of the 793 customers rejected fall into one of these two categories.

As for the approved customers, they are given a Plastic or Digital card. We found that 70% of the customers are given a plastic card, while the remaining 30% obtained a digital card. Since we are a technology company, it would be ideal for us to have most customers using digital services. It could be an interesting campaign to encourage customers to use digital card services, this would require them to interact more with our app or website services and the expense of producing plastic cards could be reduced.

See bar plot below:
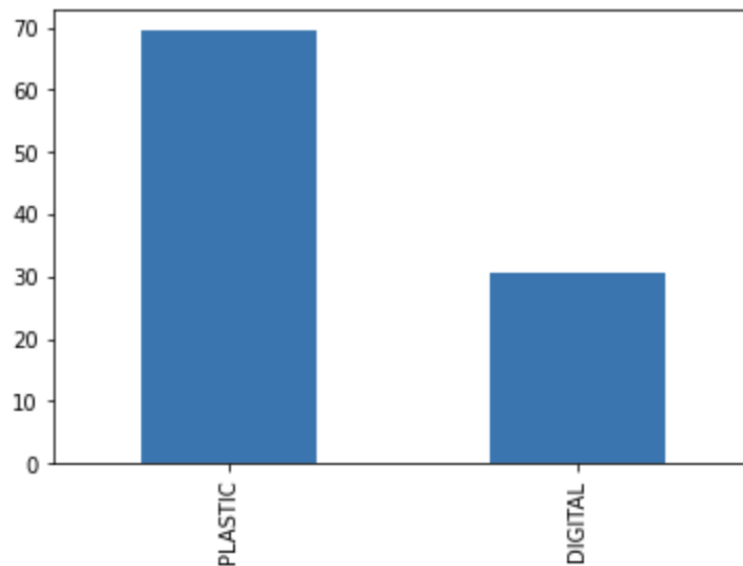
Next, I decided to explore some of the main differences between customers that possess a plastic or digital card. The numerical columns may give us some insight, we can look at average amount, average CAT or average interest rate. The table below shows that there aren't any major differences between the averages for both plastic and digital in the fields mentioned.

STATUS	MOTIVE	AMOUNT	CAT	INTEREST_RATE
APPROVED	DIGITAL	14891.29	39.32	39.40
	PLASTIC	14299.24	39.81	39.83

Since the average did not provide valuable information, we can look at the distribution of the values for the three numerical columns we looked at above. The values in the columns INTEREST_RATE and CAT were evenly distributed with values in between 30 and 50, therefore the average is a valid metric.

The AMOUNT column contains values between 0 and 108,300, where 77% of the values are between 0 and 21,740, therefore the average would not be a valid metric to perform analysis on this column.
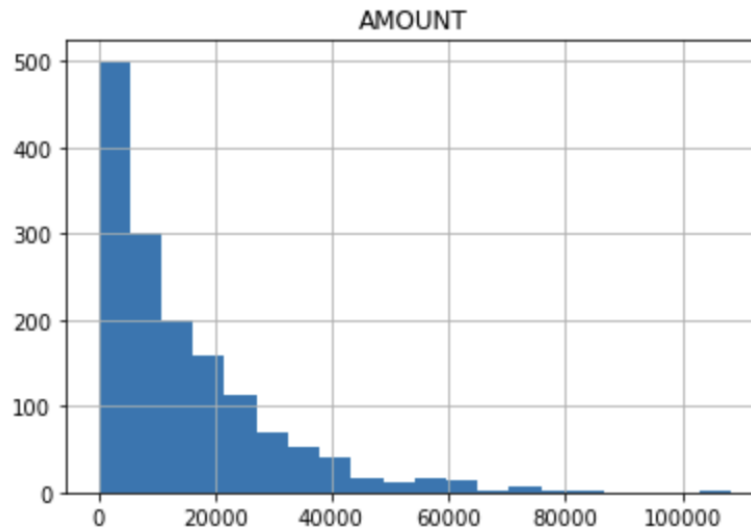
AMOUNT

Among the customers that were assigned an amount below 21,000, we found its distributed as follows:

35% have an amount below 4,100

23% have an amount between 4,100 and 8,300

17% have an amount between 8,300 and 12,500

12% have an amount between 12,500 and 16,700

11% have an amount between 16,700 and 21,000

It was also found that there were 11 customers who have an AMOUNT = 0, not a significant number that affects this metric, but it would be worth looking at these customers in case there was a mistake coming from the database.

The following step is to look at the TXN or Transactions column because we need to keep track of usage patterns among our customers. It was discovered earlier that 70% of the customers were given plastic cards. The dataset contains records on 3,773 transactions, since more customers possess plastic cards, more transactions should be made with plastic cards. It was found that 68.8% of the transactions are done with plastic cards. See bar plot below:
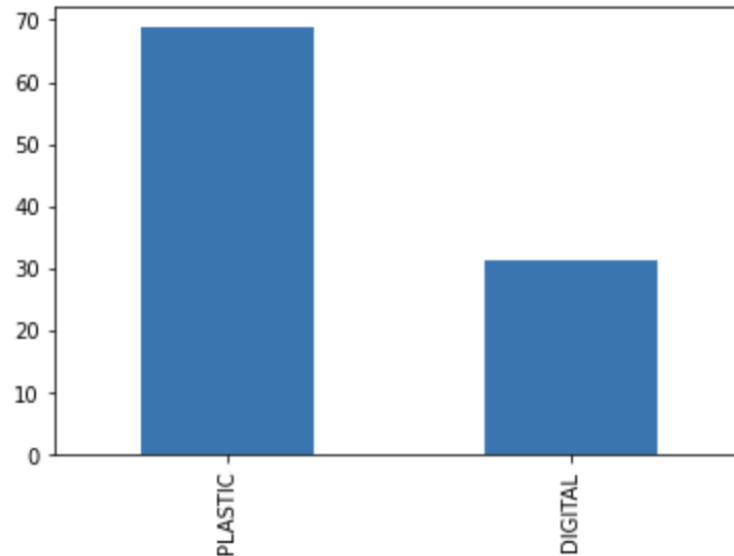

Lastly, I wanted to explore expense patterns in both digital card types. It was found that there are 1,178 digital transactions, which represent 31% of the transactions made, and there is an average expense amount of 2,022. Furthermore, the highest digital transaction is 110,631, followed by a few transactions with amounts between 68,000 and 98,000. See transaction amounts of digital card holders:
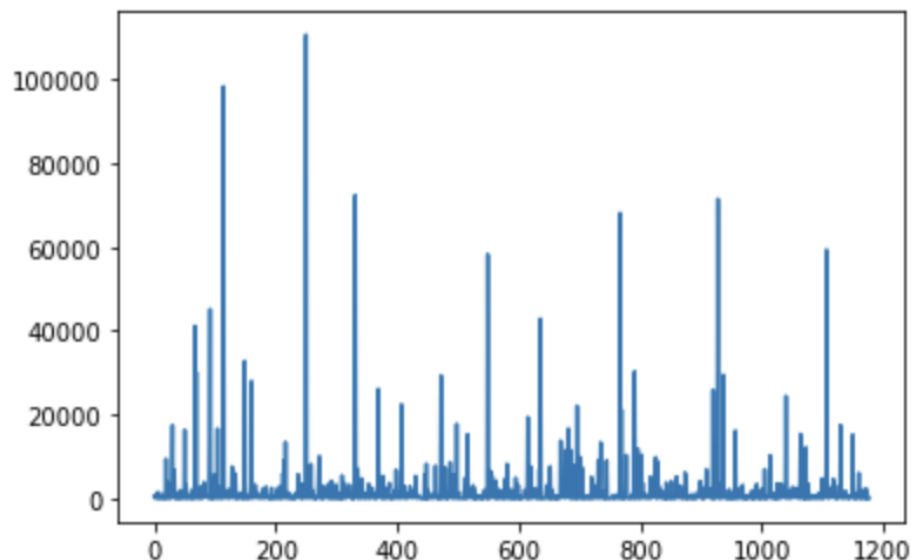
It was also found that 97% of digital card transactions are below 20,000. Inside that group of transactions below 20,000, the transactions under 1,943 represents 83% of the total. We can see that most transactions are lower than the average expense cost. Also, the amount of clients that make high-cost transactions with their digital card seems to be relatively high. It was found that among the customers that perform digital transactions higher than 20,000, which is around 1.7% of the digital transactions: there are 54% that spend between 20,000 and 39,000 and 22% make transactions between 57,000 and 74,000, followed by 13% with transactions between 39,000 and 57,000. It seems like there could be more incentives for customers to use their digital card frequently and for transactions of higher cost.

See bar plot that contains most given amounts to digital card customers:
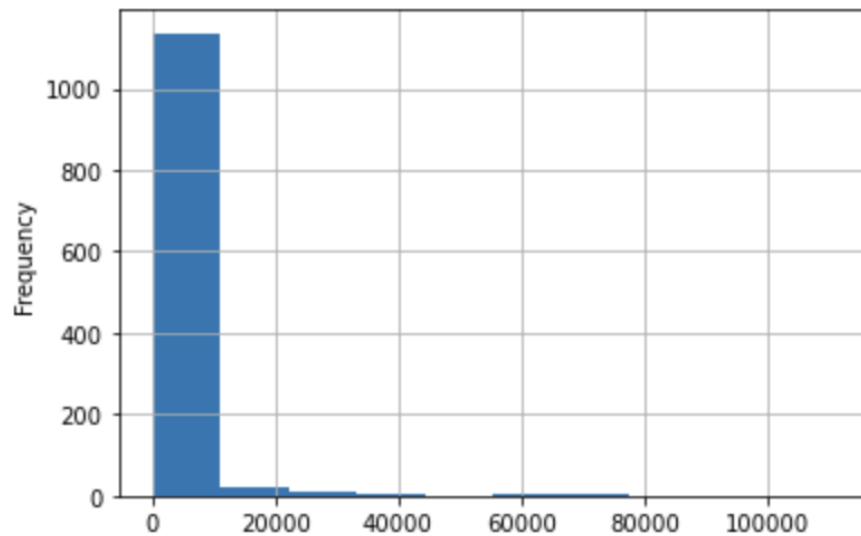
On the other hand, the average amount of plastic card transactions is 2,004, there is one outlier with a transaction of 412,498 followed by a few amounts in the 100,000 range. Therefore, these values greatly affect the average transaction cost for plastic cards. See transaction amounts of plastic card holders:
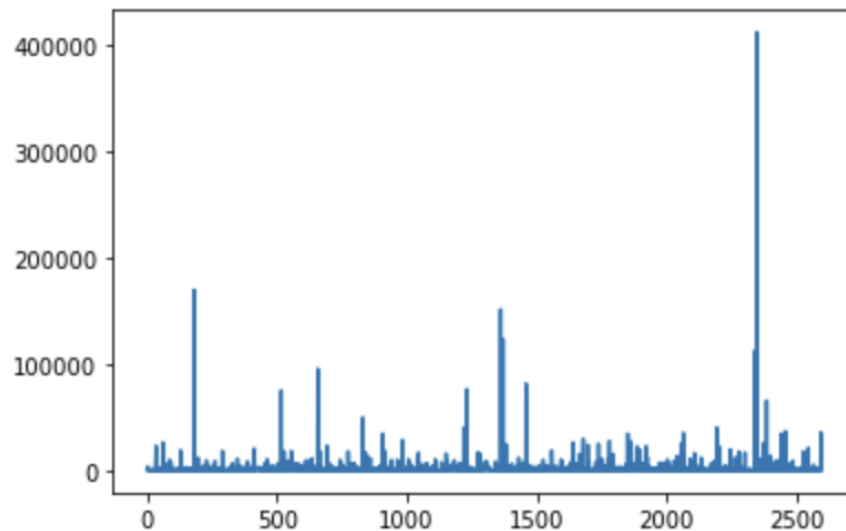
It was also observed that 98% of plastic card transactions are below 20,000. Inside these group of transactions, 75% are under 990. As for the high-cost transactions with plastic cards, customers tend to do very few of them, only 1.4% of the transactions are over 20,000. Inside this group of high-cost transactions, 74% of the transactions are between 20,000 and 59,000, and 12% of the transactions are between 59,000 and 98,000. See bar plot that contains most given amounts to plastic card customers:
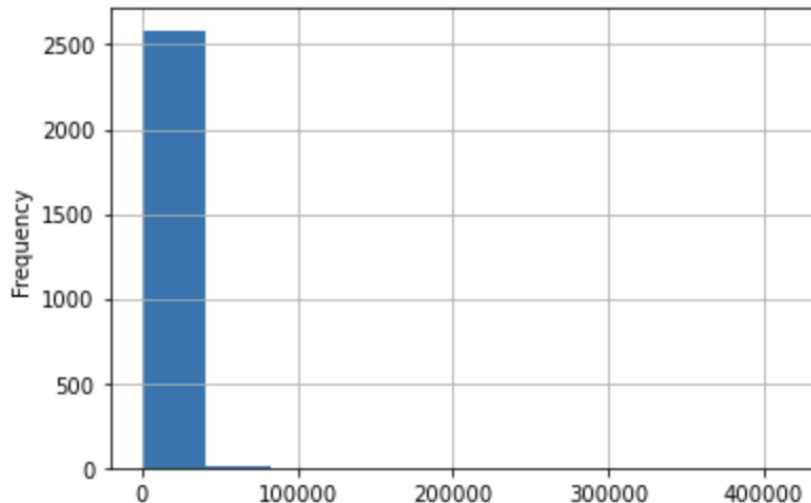
In general, most of the customers approved are granted an Amount below 21,660, and the insights mentioned above, support that most transactions costs are below 21,000. Also, more than half of the approved customers are given an amount below 8,000. On average plastic card holders tend to do more lower cost transactions than digital card holders. This only applies if we only look at customers that performed transactions under 20,000, which is the majority for both card types. The data presented above can also lead us to conclude that customers with digital cards tend to spend more on higher-cost transactions, than plastic card holders. The reason for this could be that everyone is used to use their plastic cards on their daily expenses, opposed to digital cards that can be used to buy more expensive items from a website or app, we see that the amount of digital card expenses tend to be higher. Additionally, lower cost transactions have a higher average price, we saw that for transactions under 20,000, there are 83% digital transactions below 1,943, compared to 75% of plastic transaction under 990, this is double the transaction price of digital compared to plastic.

Finally, the last key performance metric I explored is the customer satisfaction on the delivery of the card. It was found that customers are not very satisfied with the service. Only 1% of the customers rated the delivery with a 5, while more than half of the customers rated the delivery with a score of 1 or 2. And 83% of the scores are between 1 and 3. According, to customer's opinion, the delivery service of the card must be better.
See Bar plot below:
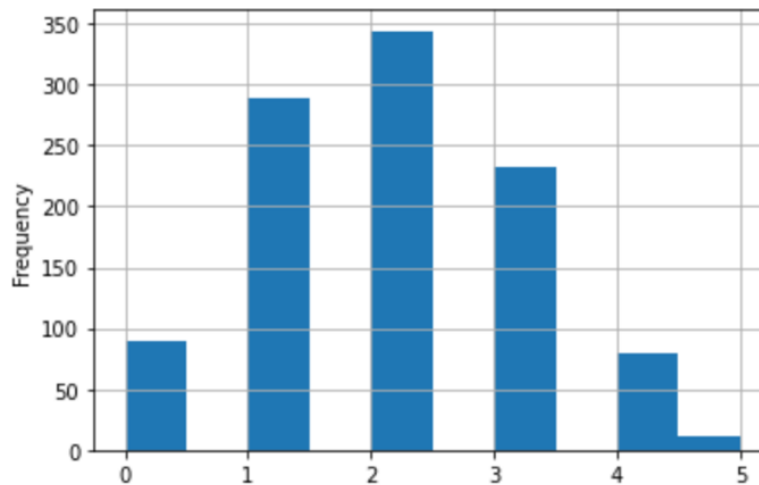
In conclusion, the company can provide a better delivery service of the card and improve customer satisfaction. We also found out that we need to improve the response rate to the applications sent to the customer, we need to receive more responses when these invitations are sent. The approval rate of the customers is fair, the risk model seems to be doing a good job, but we can still find what aspects to rejected customers have in common and maybe be more lenient in the acceptance process. As for the approved customers, we see that plastic card holders tend to use their card for lower cost expenses than digital card holders. This could be a great opportunity to create incentive programs for digital card users and encourage existing customers into changing to a digital cards program, since there is a tendency to spend more when performing digital transactions.