Edgar Iván Rosas Gómez

BUSINESS INTELLIGENCE / BUSINESS ANALYST, LEAD BUSINESS CASE

[1] We just launched the credit card to market. As you might be aware, everyone was extremely busy planning and developing the product, but no one thought of coming up nor monitoring the key performance indicators of the business. 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?

KPIs for credit card business:

- Customer acquisition cost: this KPI is very important in every business and can be simply put as the cost of convincing a potential customer to acquire the credit card. Keeping this KPI as low as possible should be the goal.
- Card portfolio growth rate: this KPI measures how much the card portfolio is growing. To calculate this, we must determine the number of activated cards minus the total that have been closed, and then divide the total by the active cards.
- Cost of service per card: This KPI measure the cost of servicing every card in relation to the amount of profits.
- Average revenue per card: KPI used to measure the average revenue each card is generating to the business; this is determined by the total gross revenue divided by the total number of active cards.
- Average margin per card: KPI used to understand the overall state of cards and is determined by the customer acquisition cost, the cost of service per card and the average revenue generated by each card to get the measure of the amount each card is bringing in as profit on average.
- Income from the credit card business: this KPI is useful to determine what activities are important to generate income. In the credit card business, usually, there are three main criteria that are interchange fee, interest, and other fees.
- Protection from defaults: KPI that consider the risk of those customers that might default on their loans. It can be best considered that the involvement should occur by the 90 days overdue point.
- Charge-Offs: This KPI captures the point when the outstanding debt is not likely to be paid at all. Usually, this is after 120 days.
- Churn: KPI used to determine the rate at which customers are keeping using the products.

This KPIs main goal is to manage the risk is involved in the business and should be analyzed at least once a month and measured every day if possible.

[2] Dealing with diverse stakeholders is difficult. Where one might interpret a concept in a way, another one might differ from such interpretation. Let's take for example the concept 'dormant': some stakeholders might interpret the dormant customer as one that has not done any transactions in 6 months, where another one might say it takes only 4 months to reach this state. Propose a problem resolution strategy with the stakeholders. How would you deal with this issue? Which facts would you present?

The first step would be to define what dormant means to our company because the correct time frame would depend on the type of products or services we are offering.

To determine the correct time frame, we should look at our current customer data and look for factors such as customer purchase frequency and recency. Then, after established the first parameters, we should look at our customer activity to identify who have not interacted with the company recently. After having this data, we could break it down more and categorize and rank the customers to see helpful trends, such as which customers were repeat customers in the past but have since gone dormant, which customers are consistent, loyal customers and which customers are at risk of turning dormant.

After identifying the best time frame and the dormant customers, another important step would be trying to determine why a customer became dormant. We could use again customer data to analyze the profiles of dormant and look for patterns. Knowing who these customers are and why they became dormant would help us decide if we want to re-engage them again throw a target reactivation campaign or if we want to focus on retaining our active customers or new ones.

[3] It is a common practice to have many systems scattered all over: where one might be hosting the app, others might be hosting models needed for daily operations. This usually benefits usability over scalability. Nevertheless, data centralization is crucial for data exploitation. For simplicity, imagine there are 4 systems: - The first system hosts the app. It generates data that is stored in an internal database (ignore the database's architecture for now). Every time the user interacts with a screen, clicks a button, or opens the app, this is stored as an event. - The second system hosts the risk model. Every time a customer asks for a credit, the system retrieves the risk data from the credit bureau and evaluates whether the customer is prone to be a defaulter. - The third system hosts the customers information. Here, unrestricted information is hosted. This database contains the name, address, email, etc… - Finally, the fourth and last system hosts all the payments information, this means, all the information related to the usage of the credit card: swipes, payments, recurrent payments, credit line, etc… All systems share a unique identifier for all of our customers. This is the key that allows data to be joined on other databases. 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?

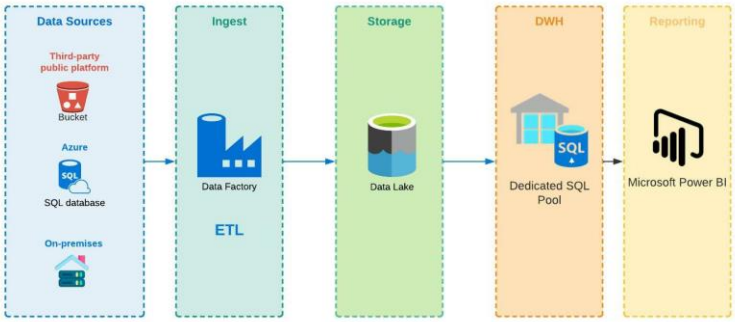
When working with a lot of data that can be raw, unorganized, or stored in a range multiple locations such as relational, non-relational and other stored systems there are now cloud tools that can centralized the ETL process to display charts to make better business decisions.

I would propose a tool like Azure Data Factory. ADF is a cloud based ETL and data integration service that enables you to crate data-driven workflows to orchestrate data movement and

transform data at scale. One of the main objectives of AZD is to be a single cloud service for data integration. It provides a set of tools and a management interface for all the data integration and all of data sources, wherever they are located: Azure, On-premises, or other third-party public platform. After all the ETL process has been done and the data has been stored in a SQL data base, we could connect to the DB from a visualization tool like power BI to display the analyzed information.

I proposed ADF because I have more experience working with this tool, but there are other trusted tools like AWS Glue of Amazon, or Google Cloud Data Fusion that can be a better fit depending on the business needs.

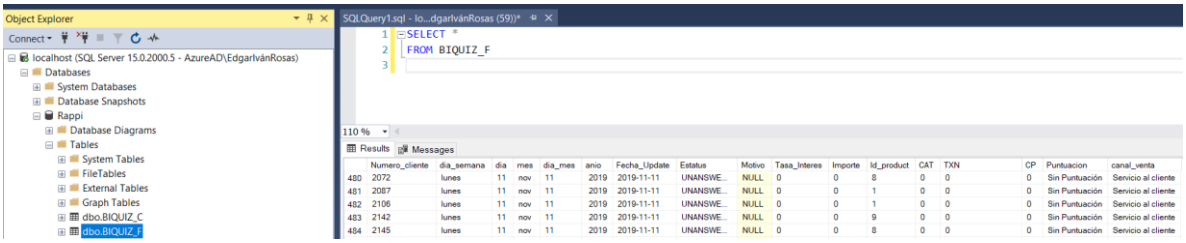
The architecture may look like the following image:



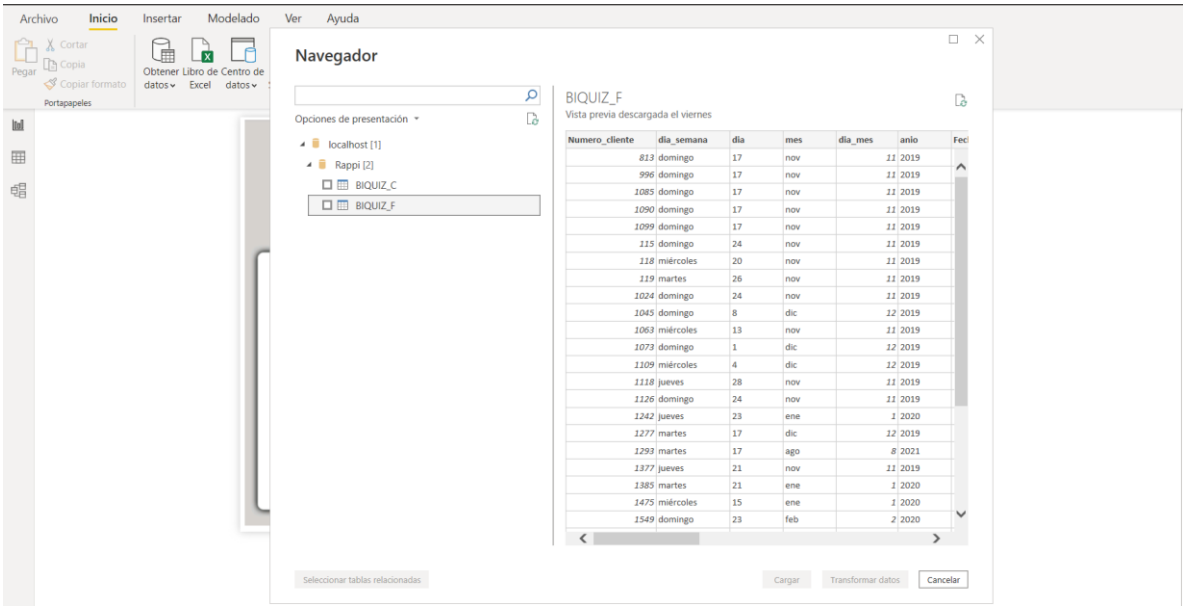
[4] Download the attached .csv file. Preferably upload it to a SQL db and query your way through the challenge. YOUR TASK IS TO exploit the information contained in the aforementioned file as you find fit. Some things to take into consideration:

[1] This database contains credit card information and transactions from multiple customers. Use your favorite data visualization tool / programming language to explore the data and present the results [R, Python, PowerBI, Spotfire, etc…].

Before starting the analysis, the CSV file of the challenge was uploaded to SQL Server Management Studio, the exploratory data analysis was started using T-SQL after some data was cleaned, mainly in the “status” column, where null values were replaced for "ACTIVE USER" and "UNANSWERED" depending on whether a user had made a transaction or did not respond to the communication.



The connection to the database was made through Power BI



Numero_cliente	dia_semana	dia	mes	dia_mes	anio
813	domingo	17	nov	11	2019
996	domingo	17	nov	11	2019
1085	domingo	17	nov	11	2019
1090	domingo	17	nov	11	2019
1099	domingo	17	nov	11	2019
115	domingo	24	nov	11	2019
118	miércoles	20	nov	11	2019
119	martes	26	nov	11	2019
1024	domingo	24	nov	11	2019
1045	domingo	8	dic	12	2019
1063	miércoles	13	nov	11	2019
1073	domingo	1	dic	12	2019
1109	miércoles	4	dic	12	2019
1118	jueves	28	nov	11	2019
1126	domingo	24	nov	11	2019
1242	jueves	23	ene	1	2020
1277	martes	17	dic	12	2019
1293	martes	17	ago	8	2021
1377	jueves	21	nov	11	2019
1385	martes	21	ene	1	2020
1475	miércoles	15	ene	1	2020
1549	domingo	23	feb	2	2020

[2] Display and plot the information you consider to be the most relevant for a Credit card business. You could consider the following departments: Operations, Growth (Marketing), Finance, Customer Service, and Product.

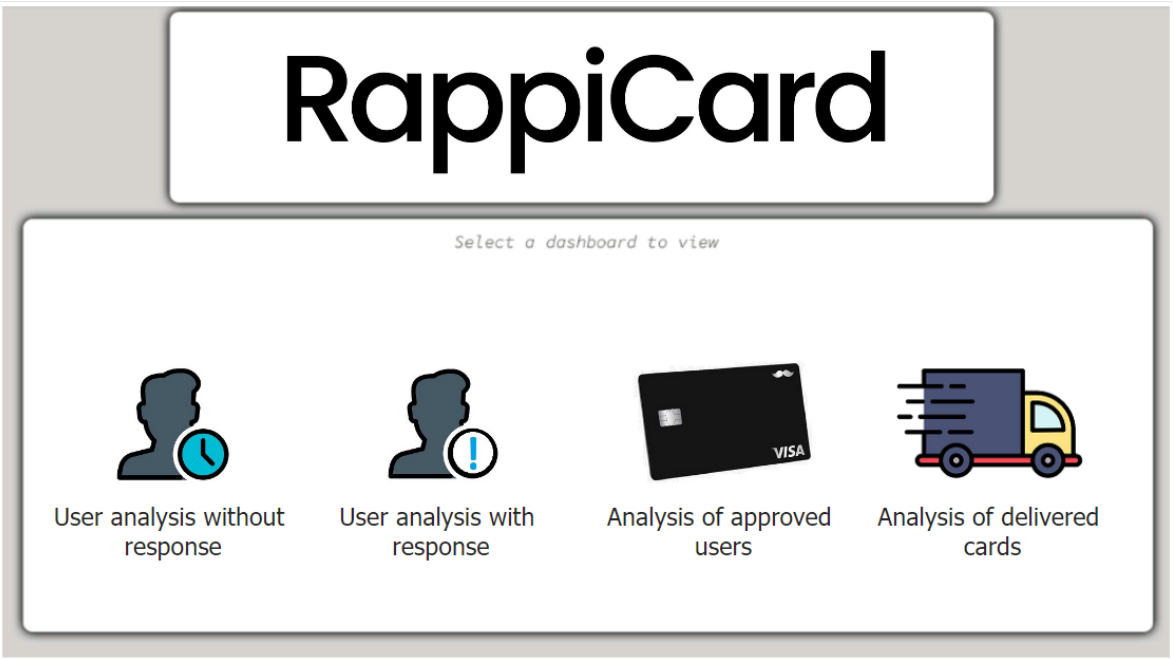
Then, the exploratory data analysis was completed to determine that the analysis could be divided into 4 parts:

User analysis without response: User analysis that didn’t respond to the communication vs the users that did respond.

User analysis with response: Analysis of the users who requested a credit card.

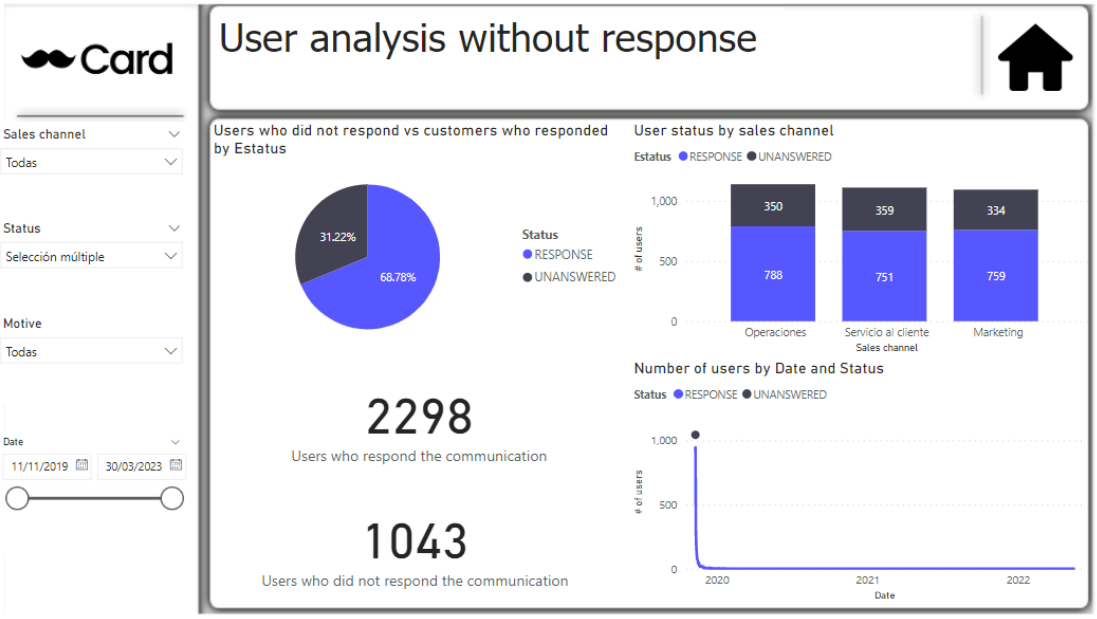
Analysis of approved users: Analysis of users who got approved a credit card.

Analysis of delivered cards: Analysis of users who requested a plastic credit card and already received it.



[3] Use your imagination to best describe the data with charts and tables. Select those key performance indicators you consider that drive the business. Present recommendations on those indicators that, to the best of your knowledge, might be low or could be boosted.

User analysis without response

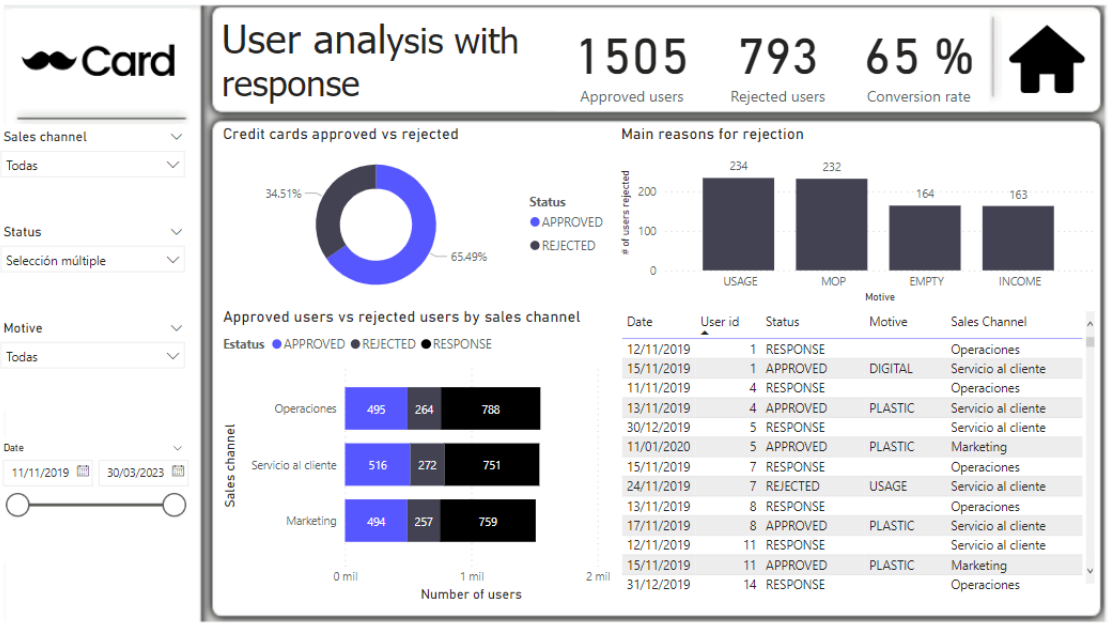


It is positive that of the total communications that were made, 68.78% correspond to users who did respond and showed interest in requesting a credit card.

Of the sales channels, the operations area presents the highest number of responses from users, and the marketing area the lowest number of responses, however, the difference does not seem to be significant between all the sales channels. It would be worth investigating in more depth the reason why the marketing area does not have a better answer, since it is understood that the main job of the area is to get new customers.

An attempt was made to visualize the time trend of user responses, their status and reasons, however, it was found that almost all user events occurred from November 2019 to early 2020 as shown in the line graph, this appears to be a dataset error so it might be worth validating the data.

User analysis with response

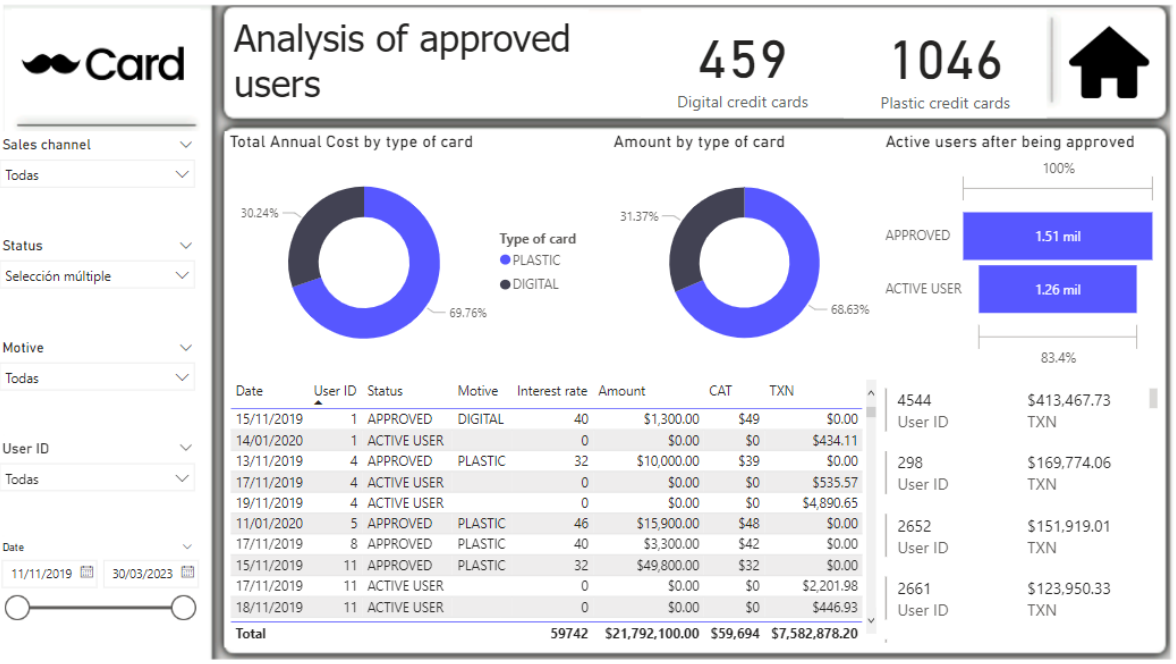


It is positive that of the total number of users who responded to the communication, 65.49% were approved and became credit card users with an overall 65% conversion rate.

At first glance, the customer service area has the highest number of approved cards, however, when calculating the conversion rate, we found that the marketing area has the highest percentage, although the difference with the other areas is 1%.

From the reasons for credit card rejection, we can see that the main reasons are "Usage" and "MOP". A strategy to approve more clients could be generated by further investigating each reason for rejection.

Analysis of approved users

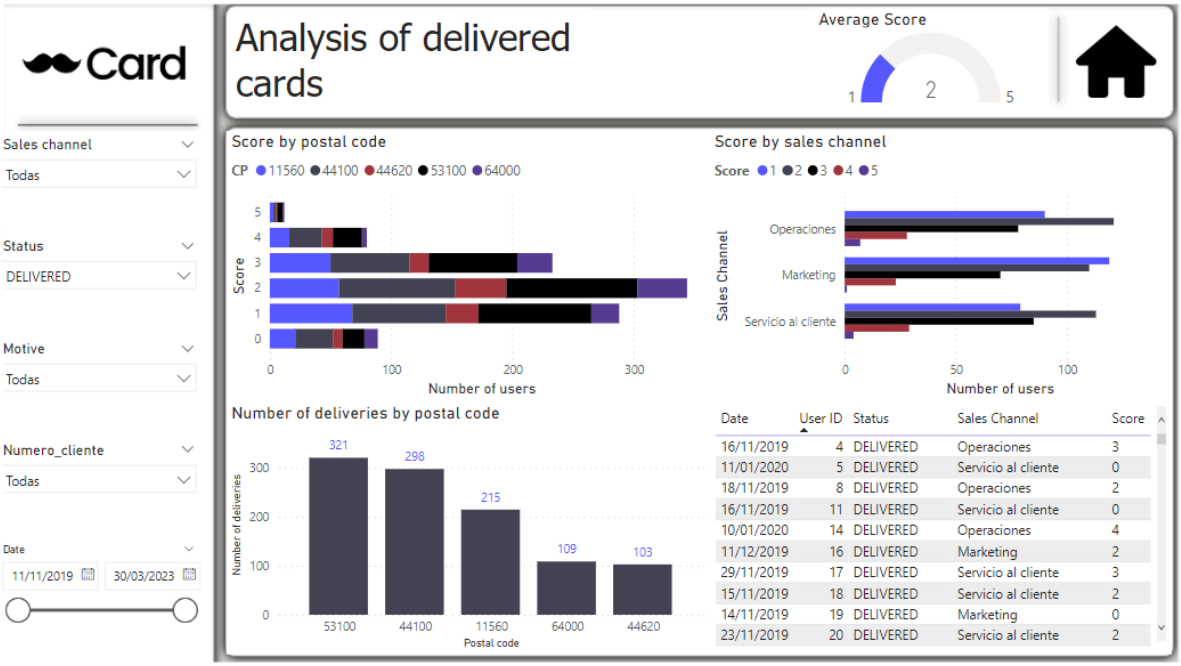


This board shows that, of the approved users, 69% correspond to users who requested a plastic card and the rest a digital card, so that the earnings from interest and the annual cost are from the credit card. plastic. It is worth continuing the analysis to find out if the users who received their credit card are satisfied with the delivery.

On the other hand, we can see that, of the total approved users, 83.4% are already active users, although it is a good percentage, a campaign could be carried out to encourage the remaining users to use the credit card.

Finally, loyalty programs could also be carried out for those users who have made the most transactions with the credit card.

Analysis of delivered cards



Analyzing the delivery score for plastic credit cards, we found that on average there is a score of 2 out of 5. Given that plastic cards represent an important part of the business, we can see that there is a large area of opportunity to improve the delivery.

The ratings given by users are mostly 2, 1 and 3 in the 5 different postal codes, with Naucalpan de Juárez being the postal code where most deliveries are made. As a city, Guadalajara represents the most important delivery point, so focusing on these places at the beginning can be a good strategy.

In general, the 3 sales channels that track delivery have the same distribution of scores and a similar number of deliveries, so working with all 3 areas is important.

[4] Think outside the box. If you feel that, extra information might be needed to support your arguments, include it in the folder: Power Point presentations, word documents, etc…

[5] Uploading your results to a git repo is desired but not mandatory.