Analytics Case Study – Jaylhane Veloso Nunes

Link from the presentation of the process and result: https://bit.ly/analytics-case-netflix-jaylhane

Link to the dashboard: Dashboard Netflix (shinyapps.io)

You are a new analyst at Netflix and have been given the following SQL dataset on your first day on the job. The dataset consists of 2 tables, as follows: a customer table and a usage table. Definitions for the fields follow. Datasets can be found at the end of this document.

- **Table Definitions**
    - **Customer**
    - Customer_ID: unique identifier per customer
    - Customer_Name: name of the customer
    - Plan: the Netflix usage plan the customer is on
    - Signup_Date: the date a customer started their free trial
    - First_Charge_Date: the date the customer converted their free trial to a subscription
    - Cancel_Date: the date a customer canceled their subscription
    - Channel: the marketing channel the customer came through to start their free trial
    - **Usage**
    - Customer_ID: unique identifier per customer
    - Movie_Name: name of the movie being viewed
    - Movie_Genre: genre of the movie being viewed
    - Movie_Length: length of the movie being viewed (hours)
    - Start_Time: the timestamp a customer starts watching a movie
    - End_Time: the timestamp a customer stops watching a movie

Question #1:

Based on the data at your disposal, what standard metrics would you propose tracking on an ongoing basis to answer the following questions?

- **VP Marketing**
    - How do we measure customer growth?

        We can measure the quantity of signed-up customers through time, analyzing and comparing, for example, the channels and plans that bring more customers.

        In that scenario, we should quantify in time the variation and increase for signup and ongoing customers (that has been charged and doesn't cancel his signatures).

        How do we have an ongoing basis, it's possible to measure this increment by calculating:

        - *The moving average (consists in calculate the arithmetic average from most recent **r** observations):*

$$M_t = \frac{Y_t + Y_{t-1} + \cdots + Y_{t-r+1}}{r}$$

*In case that M_t it's the average from the last **r** observations and ignores the past values.*

This approach's necessary because the metric varies smoothly in time and so the past observations need to be replaced.

- *The proportional variation:*

$$P = \frac{Y_n}{Y_{n-1}} - 1$$

Considering:

n = the length of available data,

Y_n= it's the value of ongoing customers on the actual time of data,

Y_{n-1} = it's the value measured in the time before Y_n .

It's important to follow and observe the evolution from the proportional variation of the ongoing customers because if we have a negative variation means that are more cancellations than ongoing customers, and if the variation it's equal to zero we can conclude that the number of ongoing customers it's equal to cancelation.

Even though both this scenario is not desirable, focusing on the customers' growth, it's preferable deal with a zero variation than a negative one.

- What attributes are important for increased growth?

For an increased quality growth, it's important to measure the customers that were converted (the ones who presents the first charge in card) and know from which channel they came and what plan they signed, because are these customers turn into revenue. Plus, knowing these customers will also allow us to develop products with the Director of Product thinking in retain these customers.

Also, it's equally important to analyze the not converted and canceled, because with this information we acquire suggestions from channels and plans that require our attention in future marketing campaigns (demanding more or less investment of money, time, and staff's energy). Therefore, we can evaluate:

- *Proportions of customers that pass from trial to actual signature.*
- *Proportions of giving up.*
- *Proportions of canceled.*

- *Calculate the above proportions of customers per channel, plan, and movie genre (separated).*
- *Average time from signature until canceling.*

- **Director of Product**
  - What type of usage creates sticky customers (customers who retain at high rates)?

    To measure the type of usage that creates the sticky customers we can analyze the type of plan that keeps the greatest number of ongoing customers because it means that this product has the more reliable customers.

    Besides that, we can measure what type of genres it's most frequent between the ongoing customers, to invest more in this kind of product.
  - How can we tell if our customers are happy?

    To know if our customers are happy we can measure:
    - *The average number of times in a week they have accessed Netflix.*
    - *Average days between Netflix access.*
    - *Average time watching movies on Netflix (general and for movie genre).*

- **Product Engineering Manager**
  - What additional fields would you like to instrument in the product and add to the data?

    Thinking about the retention rate, customer satisfaction, and aggregating data, I would like to suggest the following instrument in the product:

    - *About possible data we already have, I would like to see:*

    1. *Recommendation Rate (There may have any pattern in the recommendations between canceled and ongoing?).*
    2. *Information about the terms searched at Netflix, to identify interests mainly from the sign-up customers in the intention of retaining them.*

    - *New suggestion:*

    3. *A specific field on Netflix that allows the customers to write reviews from the series and movies. With this data, we could apply techniques from textual analysis like sentimental analysis to know better how our customers feel about our products.*

Question #2

Visualizations are frequently used to convey complex data in an intuitive manner. What are some best practices for creating visualizations in your opinion and why?

Some best practices for creating visualizations it's attached to thinking of simple solutions that reduce the reader's interpretive cost, such as:

- *Avoid saturation*
- *Pay attention to alignment and spacing*
- *Analyze the need for frame and grid in the graph (this makes the graph a bit "suffocated" and usually can be dispensed)*
- *Use shapes, colors, and tones that help's the reader understands the graph*
- *Use clear and simple fonts*
- *Highlight key and relevant information, but without polluting the graph*
- *Choose suitable reference measures*
- *Thinking about the logic of the order of the axes to favor the information, while facilitating the reader's interpretation.*

These practices help to think about the context and usefulness of the information, thinking about alternative ways and visualization possibilities before generating the visualization itself

Describe your approach to creating self-service interactive dashboards to allow the business to access this information. What information and interactions would you focus on? How much information would you make available?

1. I usually start thinking about the problem and context involved, elaborating relevant guiding questions.

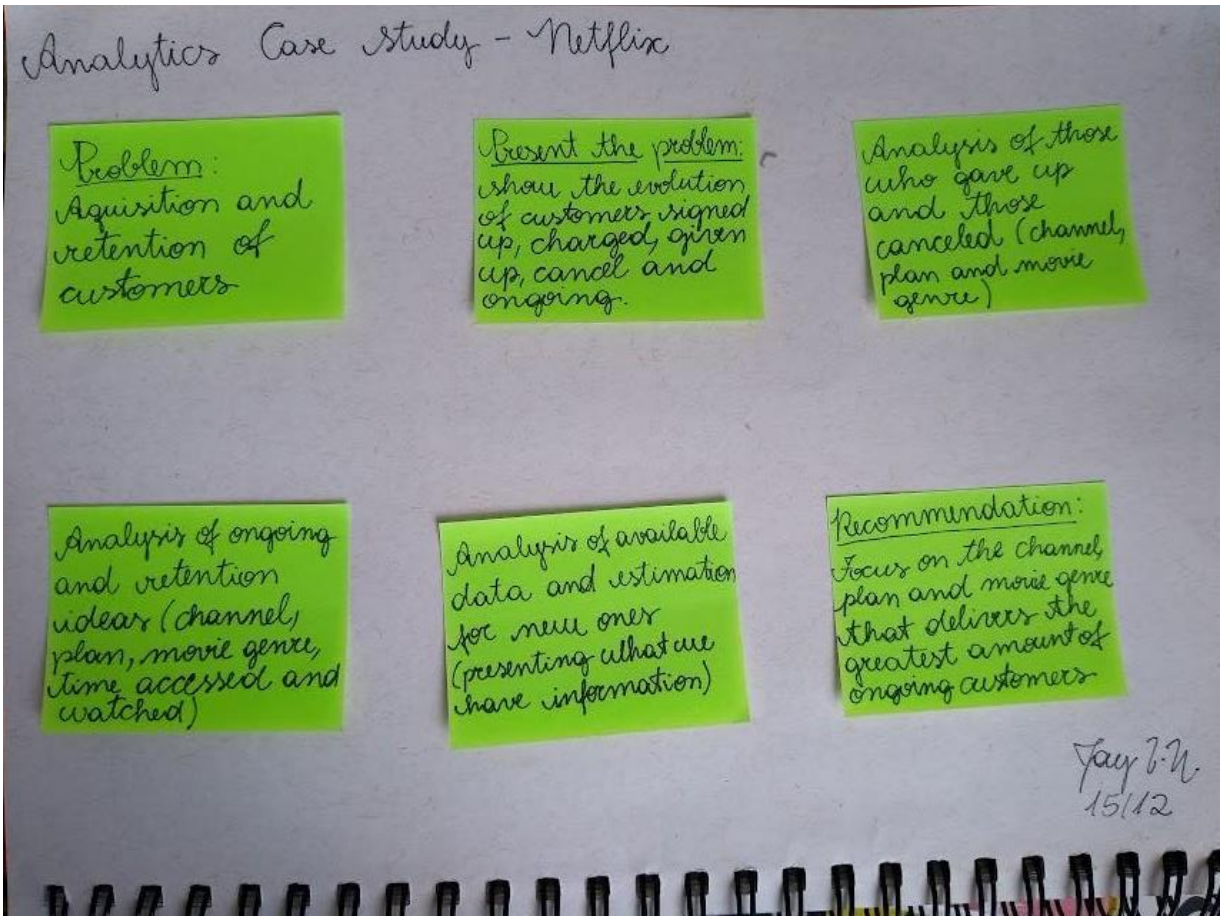
2. After that I elaborate a storyboard organizing the ideas to:

a) establish what is the main problem/idea to be investigated

b) create a flow inspired in respond the guiding questions

c) organize how to communicate my information

Example:

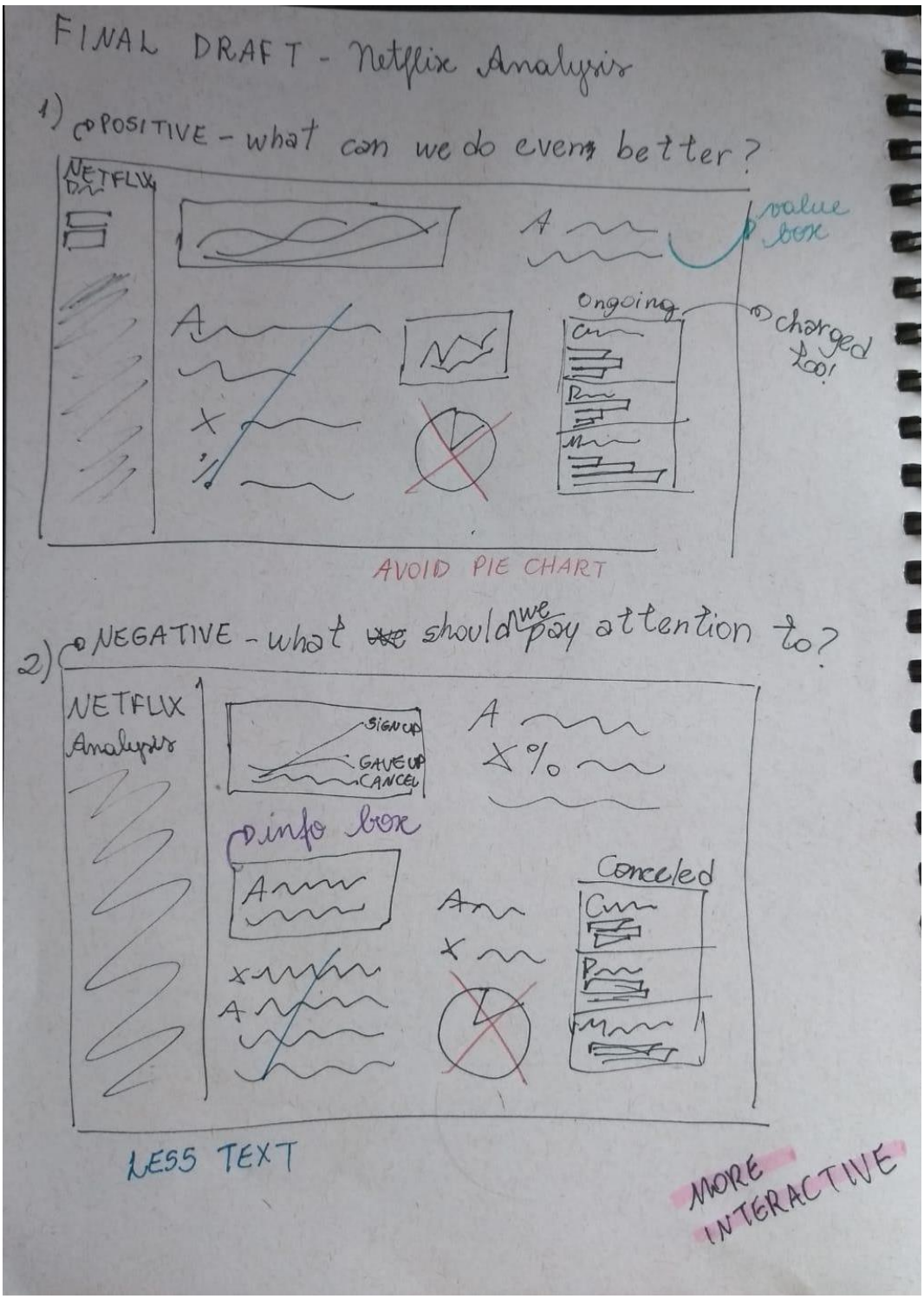


3. In the sequence I elaborate indicators and measures that answer the guiding questions and business needs. At this moment I also think about how much and how the information will be present on the dashboard:

a) I think about the layout and how to present the information to an easy understanding for the user. For example: what graphics, options to customize the visualization, etc. For this, the metrics selected must make sense in a general and specific way, or the customization it's more harmful than useful.

4. I prepare a visual draft of the result; it can be done on paper by hand for later facility programming or development in specific software. This step is also important to review how much information will be available on the screen. In a dashboard, I usually set available 5 measurements/graphs and use a maximum of 4 panels, but the ideal it's 1 or 2 panels.

Example:



Question #3:

Netflix launched original content series on the 27th of December.

- Describe the steps you would take and stakeholders you would work with to ensure tracking of this new feature

I worked together with the VP Marketing and Social Media Team.

With the VP Marketing, I would ask for a report or data from the channels that we have publicity to measure the numbers of clicks, times that trailer was watched, and a mean from how long the trailer was watched (mainly with the channel that brings more customers). These metrics will allow us to measure the movement around the launch and possibly estimate the number of new signed up. I also like to follow how many times the trailer was watched inside Netflix, for how long and searches.

With the Social Media Team, I would analyze the quantity and quality of comments in social media, like Twitter, for example, I would do a scraping in Twitter looking for comments mentioning the show and the feelings about it.

- Usage data is only available for acquisition through 26th Dec – what early reads can you get on the success of the original content launch on the 27th?

With the available data, I would like to realize clustering technics to identify groups that consume similar content in the platform with the one in the launch. With this information I would like to analyze the conversion rate, the guiding question here could be: how many signed ups customers turn into customers charged? and for what channel do they arrive on the platform? With this data, we could estimate a conversion rate between the possible new signed up.

This data in reunion with the information about channels and social media brings a notion about enforcements we could make to costumers' growth. For example, if there's a lot of negative comments about the trailer, a new trailer could turn the opinion about the show.

If we focus the new trailer on the channel that it's responsible for most of our ongoing customers, this possibly brings more customers with similar retaining behavior.

For the same reason, I would like to see the rates for canceled customers with more than one month. Even if these customers aren't our ideal, with a one-month signature we have a one-month fee. We will not spend big efforts to attract potential customers that will cancel, but it's a kind of revenue and we can't ignore it.