Exploratory Data Analysis

Customer Segmentation – Case study

01/2022

Agenda

Executive Summary

Problem Statement

EDA

EDA Summary

Problem Statement

- XYZ is a private firm in US. Due to remarkable growth in the Cab Industry in last few years and multiple key players in the market, it is planning for an investment in Cab industry and as per their Go-to-Market(G2M) strategy they want to understand the market before taking final decision.
- Objective: Provide actionable insights to help XYZ firm to identifying the right company for making investment

EDA

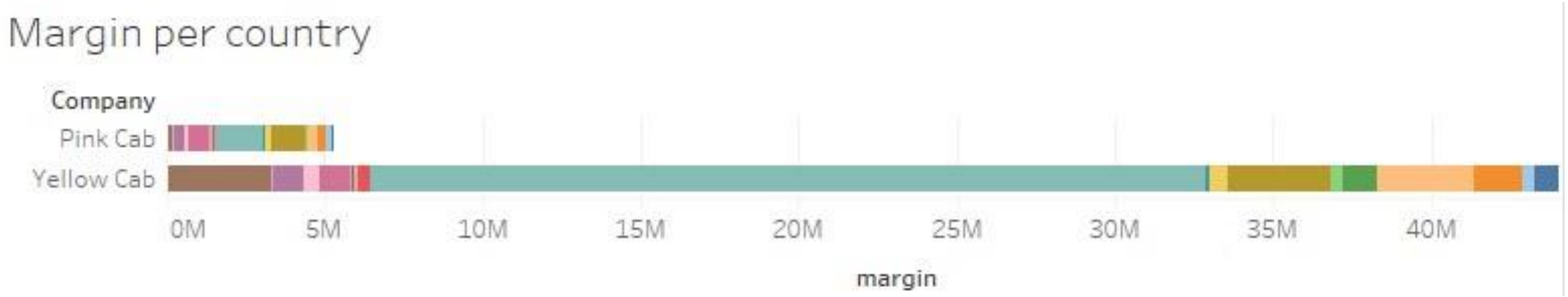
Source Data

- **Cab_Data.csv –** this file includes details of transaction for 2 cab companies
- **Customer_ID.csv** – this is a mapping table that contains a unique identifier which links the customer's demographic details
- **Transaction_ID.csv –** this is a mapping table that contains transaction to customer mapping and payment mode

- **City.csv –** this file contains list of US cities, their population and number of cab users

Profit analysis

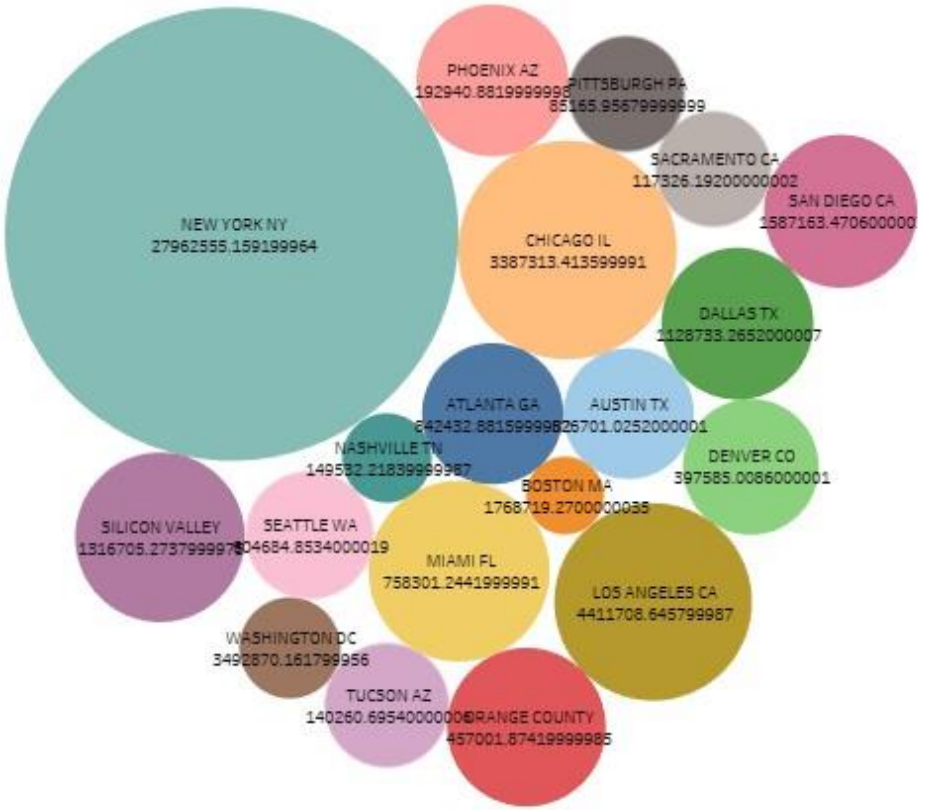
- Yellow Cab has the major profit



- The most profitable country for both is New York City

EDASummary

Does margin increase with a



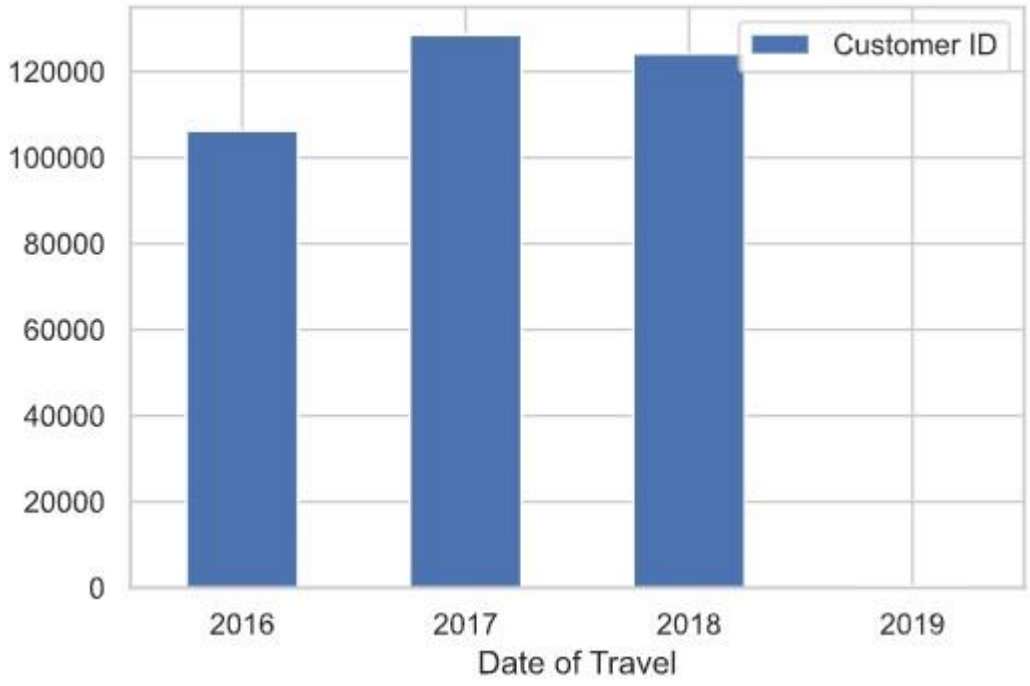
City	Population	margin
NEW YORK NY	8405837	2.796256e+07
CHICAGO IL	1955130	3.387313e+06
LOS ANGELES CA	1595037	4.411709e+06
MIAMI FL	1339155	7.583012e+05
SILICON VALLEY	1177609	1.316705e+06
ORANGE COUNTY	1030185	4.570019e+05
SAN DIEGO CA	959307	1.587163e+06
PHOENIX AZ	943999	1.929409e+05
DALLAS TX	942908	1.128733e+06
ATLANTA GA	814885	8.424329e+05
DENVER CO	754233	3.975850e+05
AUSTIN TX	698371	5.267010e+05
SEATTLE WA	671238	6.046849e+05
TUCSON AZ	631442	1.402607e+05
SACRAMENTO CA	545776	1.173262e+05

major population?

- We use qualitative analysis because of the nature of the fields (categorical and numerical), particularly descriptive statistics and note that it's not neccesary that margin increases with population, but yes with number of users.

Are number of cab users increasing with time advance?

- Using qualitative analysis, and a bar graph we can see that the number of user per year increase between 2017 and 2018 buy decrease in 2018, with regard to 2019 the data are insufficient.



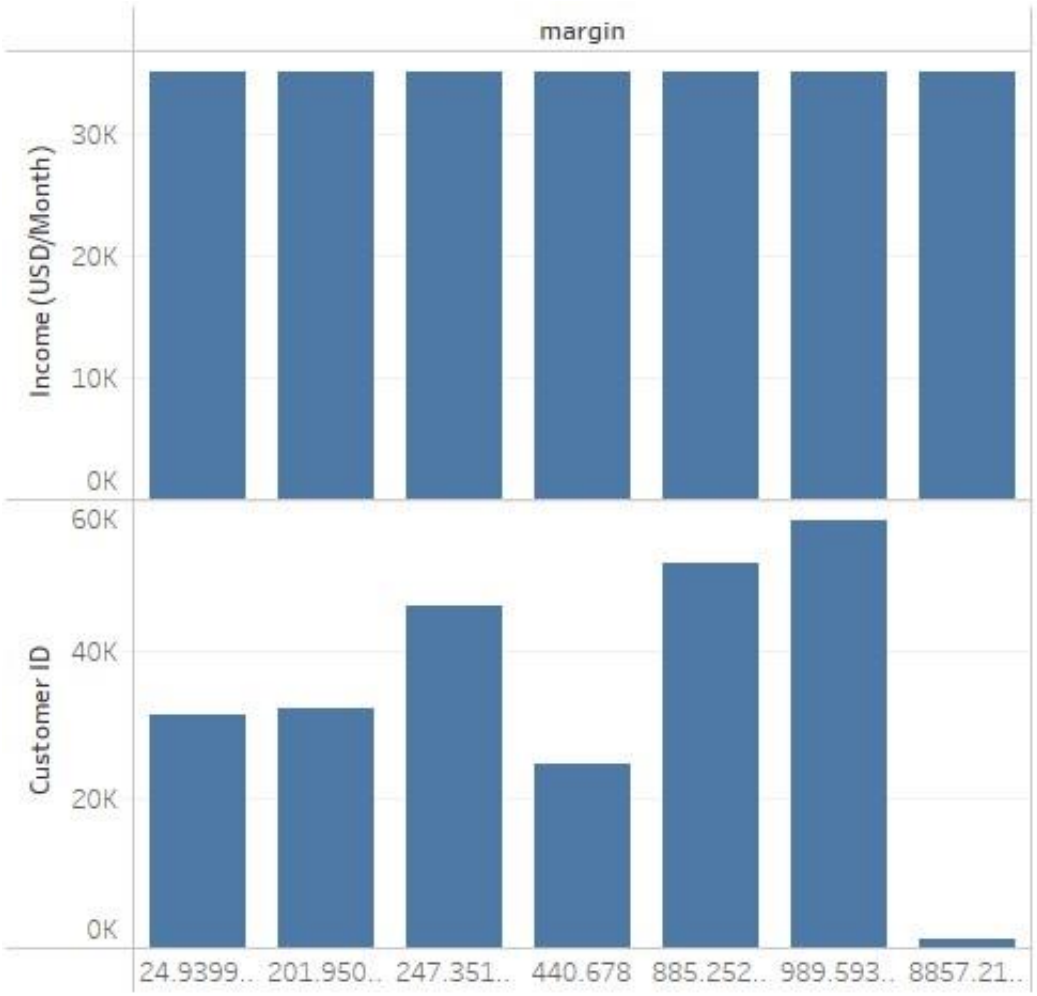
Is the most populated region the most profitable?

- It correspond to New York and yes it's the most profitable region for both companies

City	Population	margin
NEW YORK NY	8405837	2.796256e+07
CHICAGO IL	1955130	3.387313e+06
LOS ANGELES CA	1595037	4.411709e+06
MIAMI FL	1339155	7.583012e+05
SILICON VALLEY	1177609	1.316705e+06
ORANGE COUNTY	1030185	4.570019e+05
SAN DIEGO CA	959307	1.587163e+06
PHOENIX AZ	943999	1.929409e+05

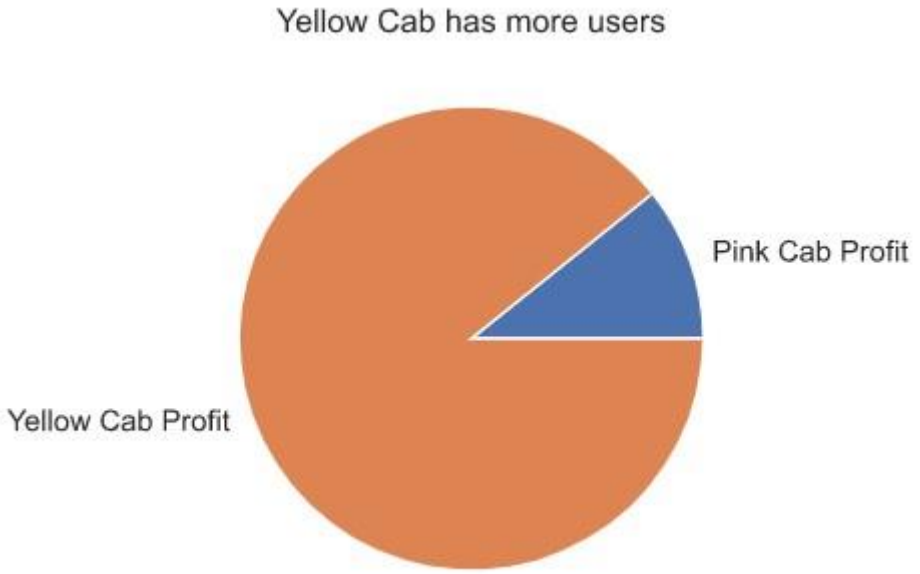
Is a user with major income related with more profit for the company?

- The answer was no, the income of the user doesn't related with margin. (Analysis similar to recent hypothesis



Is the company with more users the most profitable?

- Yes, The company most profitable is Yellow Cab and has the major number of users



Problem Statement

- XYZ bank wants to roll out Christmas with personalized offers to their customers. The group up process needs to be automated and can’t exceed 5 groups in total as a result.

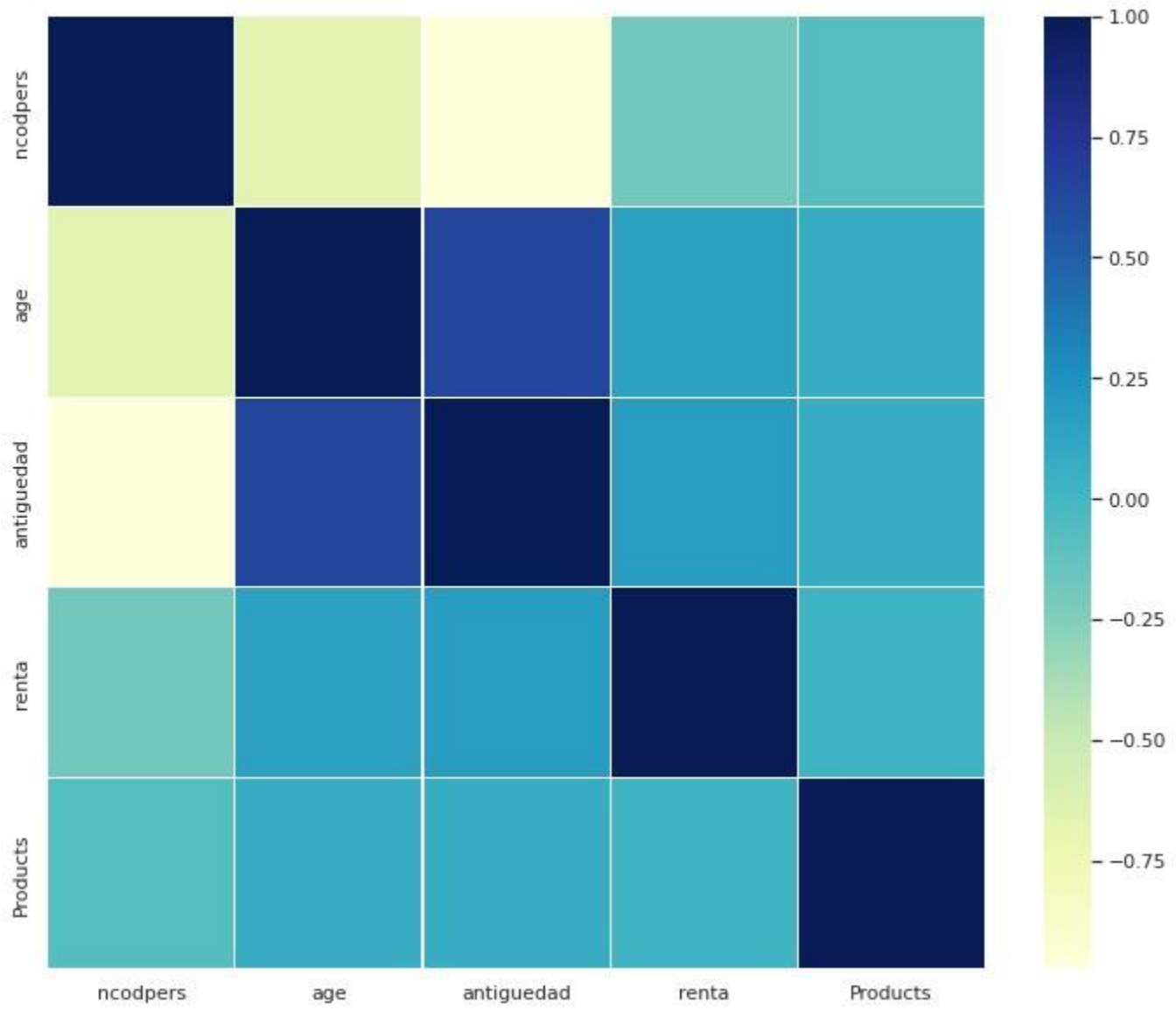
EDA

Source Data

- **Customer_set.csv –** this file includes details of transactions from different clients, each entry has taken when a client acquire a new financial product.

Selection of features

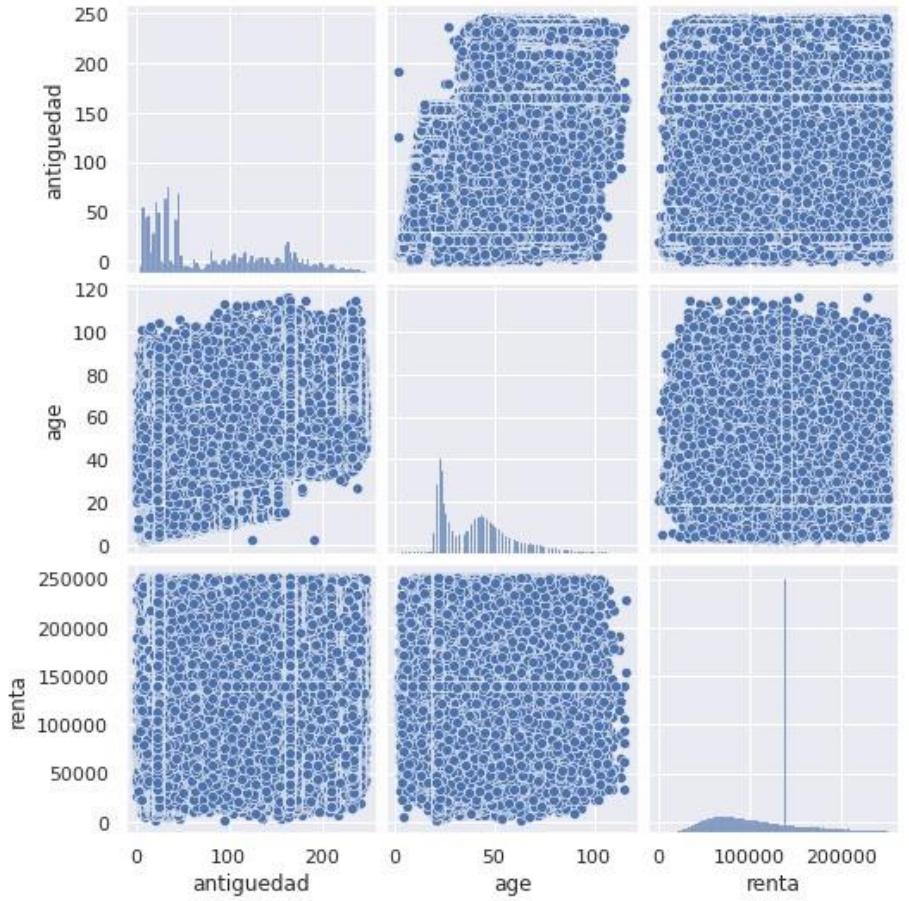
- In the graph we show a strong correlation from “antiguedad” and “age”, so we can only take one of this features
- We create a new feature called



“products” to express the total number of products that each client has.

Spread from numerical features

- We detect an strong dispertion from the numerical values, this can affect the model selected to get the classification of clients



Model Technique Proposed

Clustering – non supervised classification

- We suggest to utilize a classification algorithm as K-Means, because from high dispertion of data, DBSCAN and OPTICS can be wrong with the classification results. (the centroids can be too near one to the other)

Thank You