

Bank Marketing Analysis

Final Presentation

DS-GA 1007 Project

Introduction

Our project aims to apply Exploratory Data Analysis in a real world business scenario.

In the following case study, we will develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

Dataset

The Dataset provides us with microscopic details about each customer of the bank. This allows us to use this data in a wide array of scenarios. Some of these attributes include:

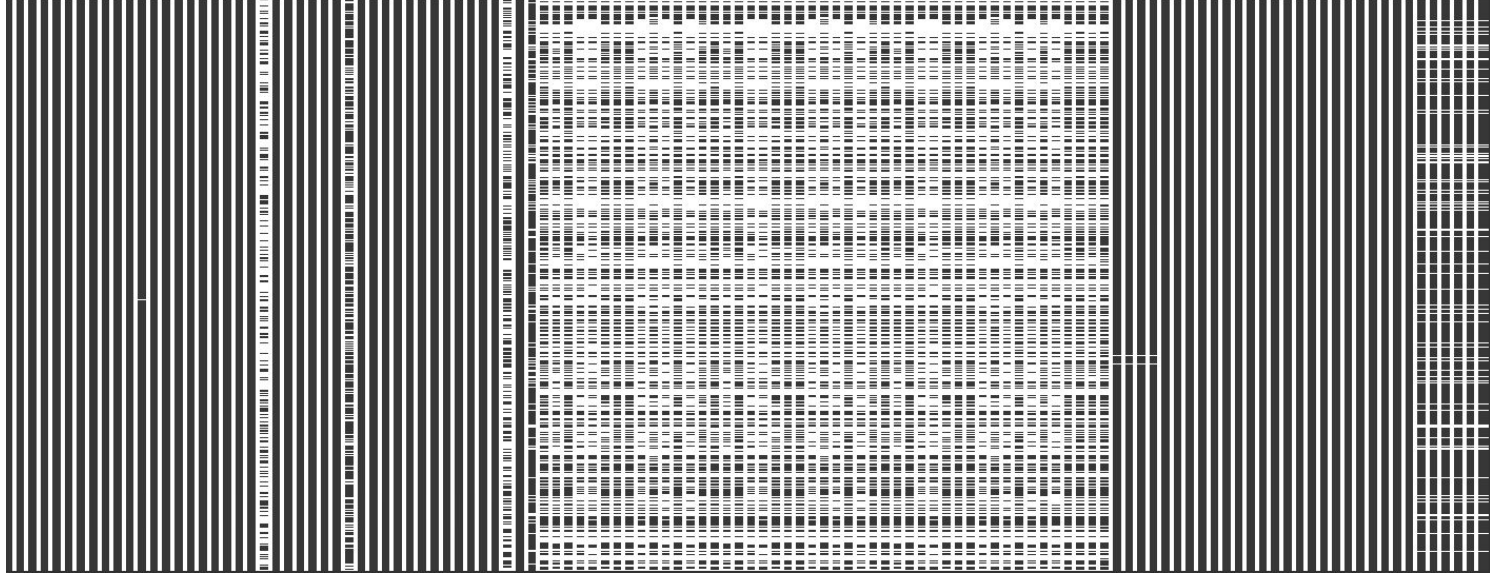
→ **Dimensions of the Dataset**

308,000 x 122



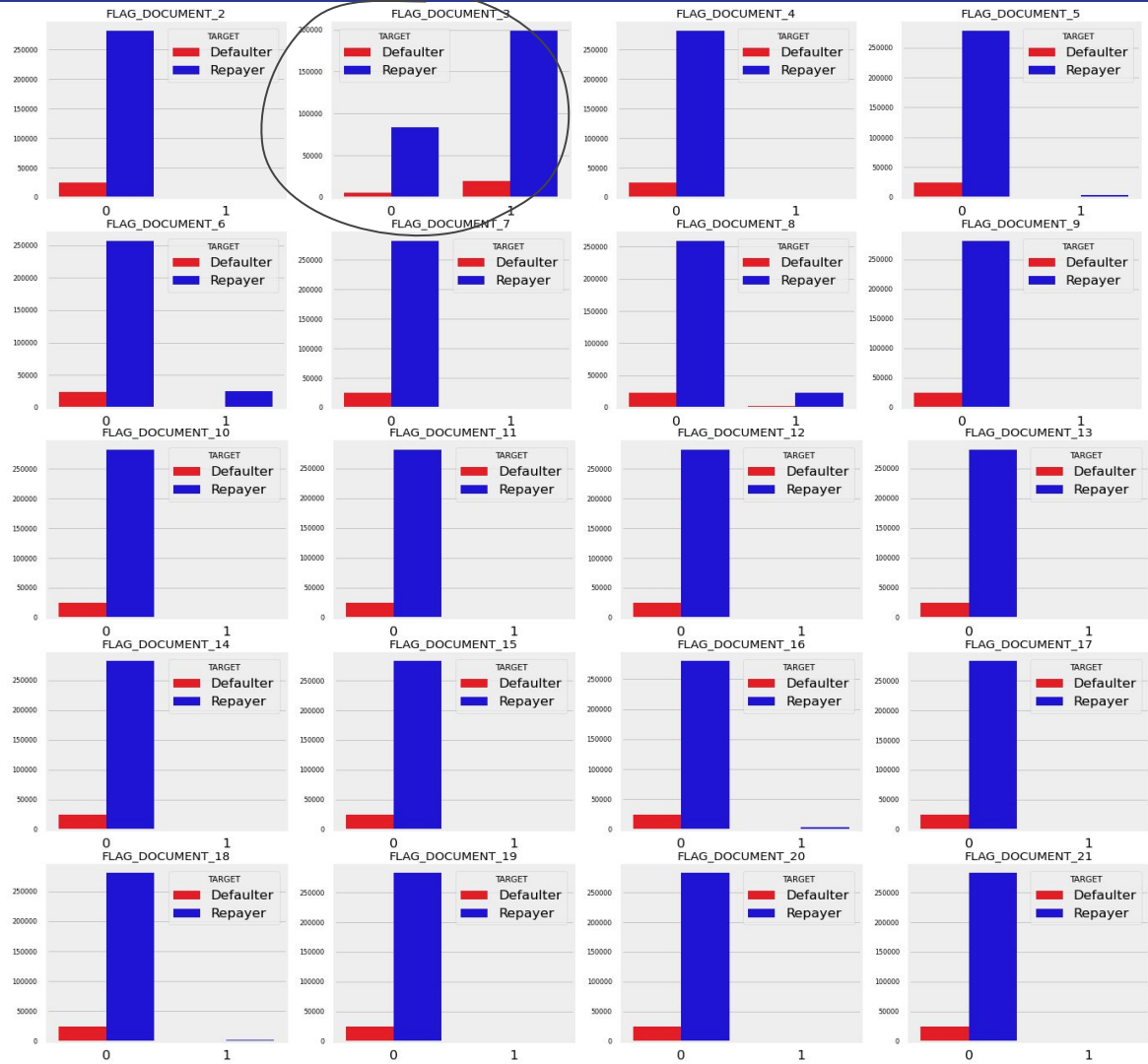
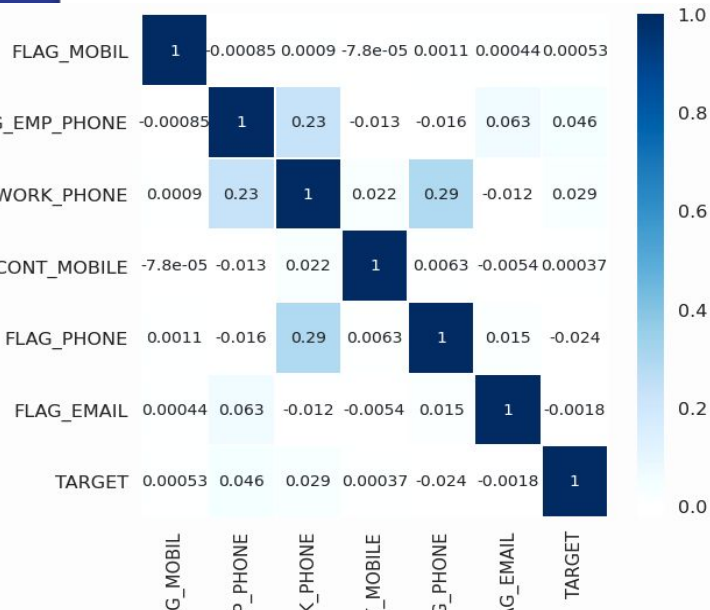
Dataset Preprocessing

Null Values



Correlations

- Document submission and Defaulting
- General Information and Defaulting



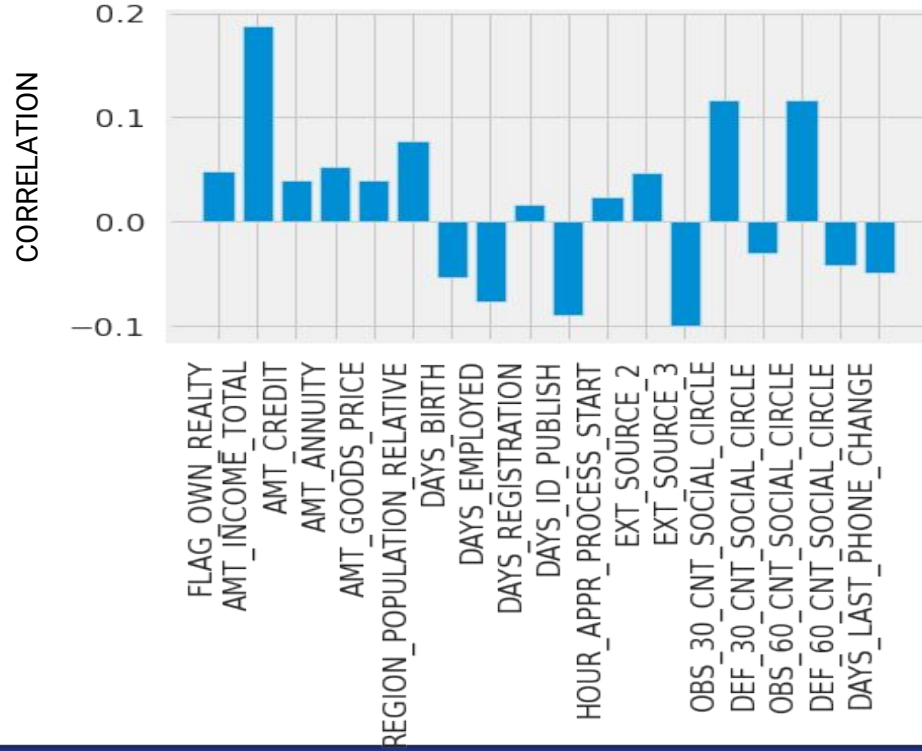


Dimensionality Reduction

Principal Component Analysis

- We perform Principal Component Analysis on our dataset to get the most relevant features thereby reducing dimensionality.
- We have picked 20 numerical features to reduce the dimensionality of our problem.
- We will be reducing 20 features to 9 PCA features.

Results from PCA





Data Analysis

We perform analysis on our data by making use of various plots :

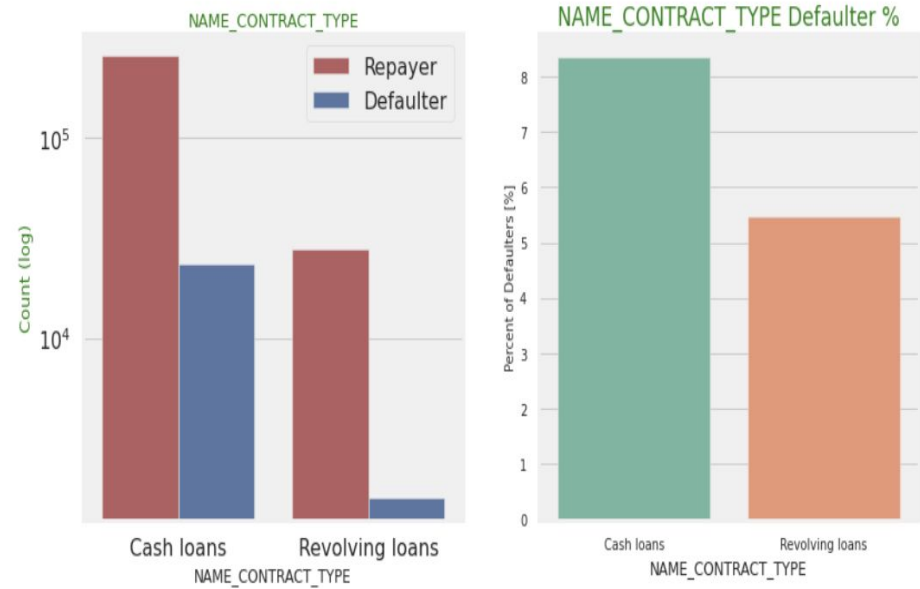


Fig 1

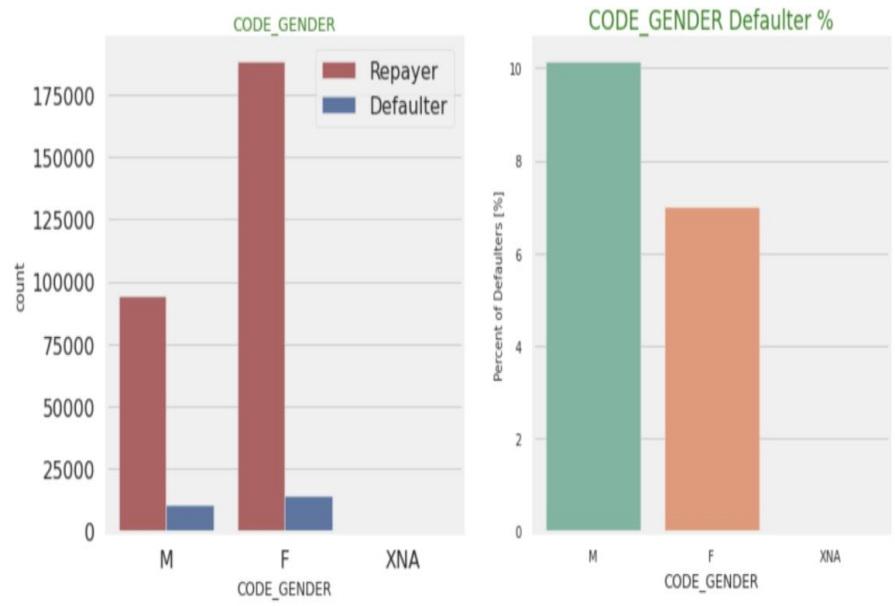
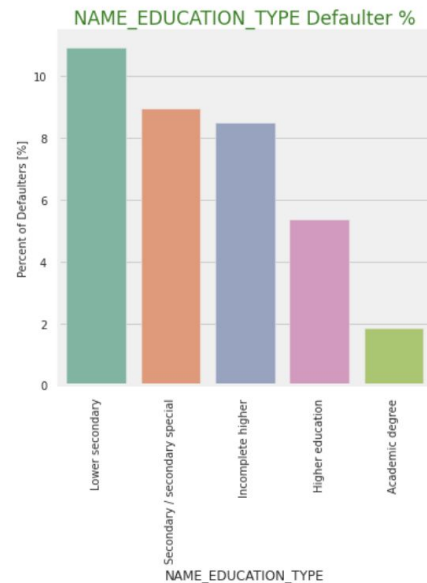
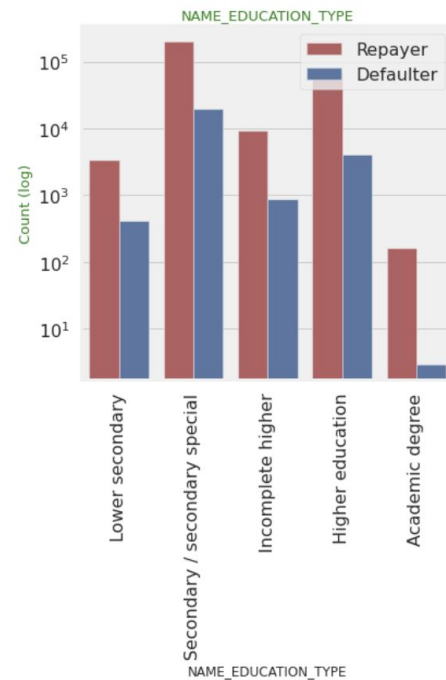
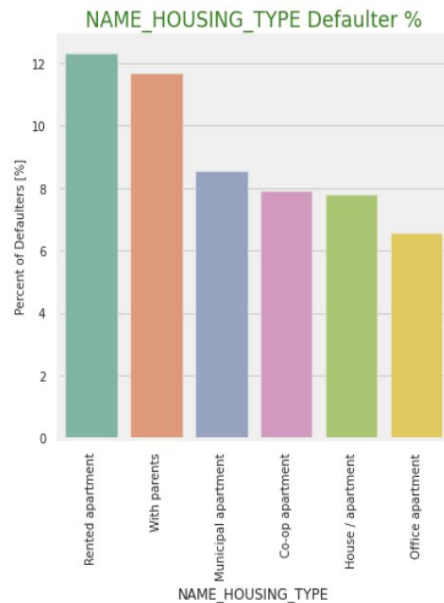
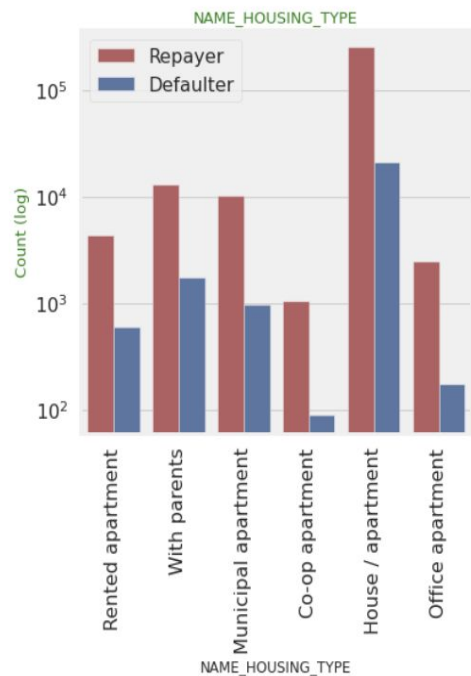


Fig 2



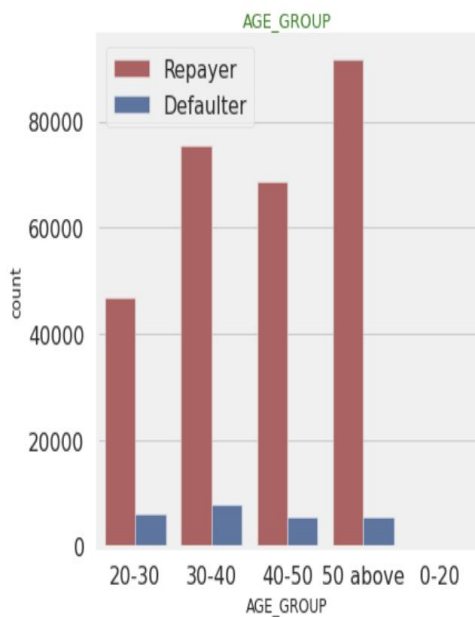


Fig 5

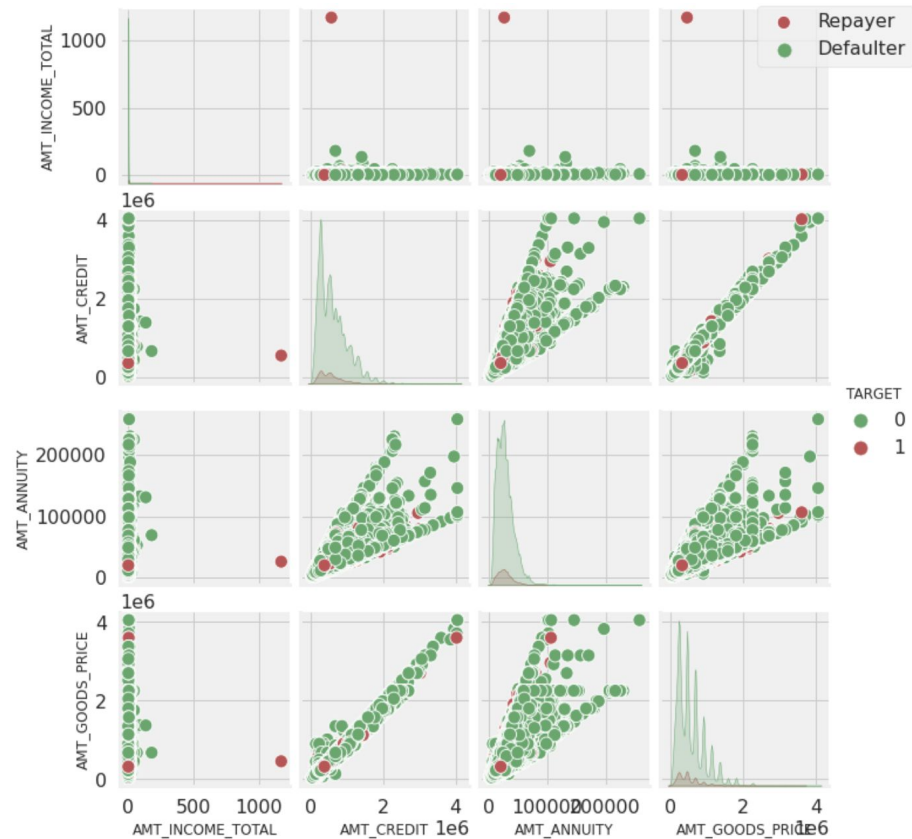
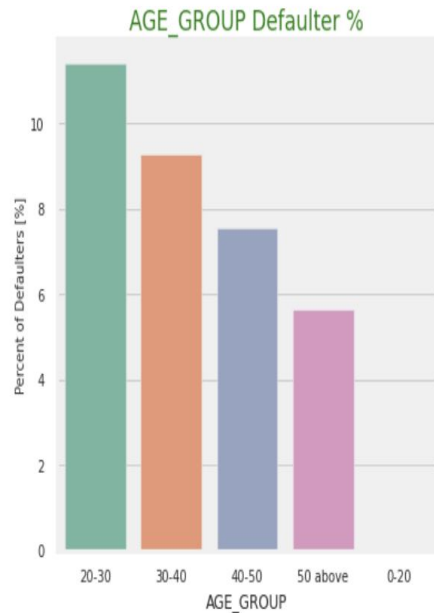
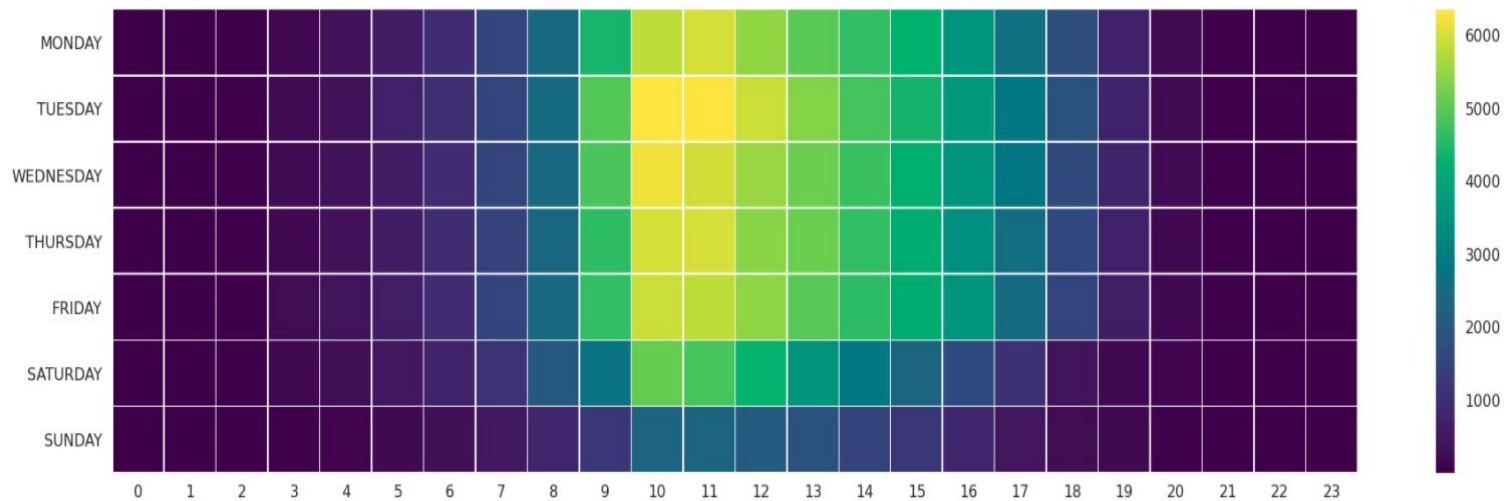


Fig 6

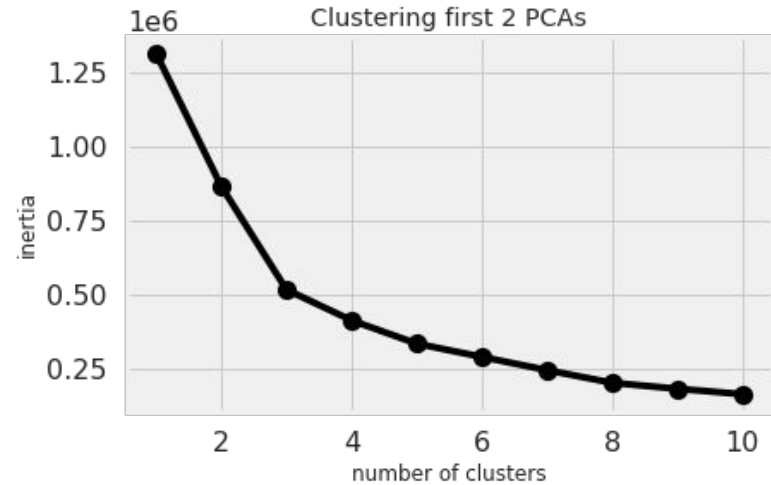
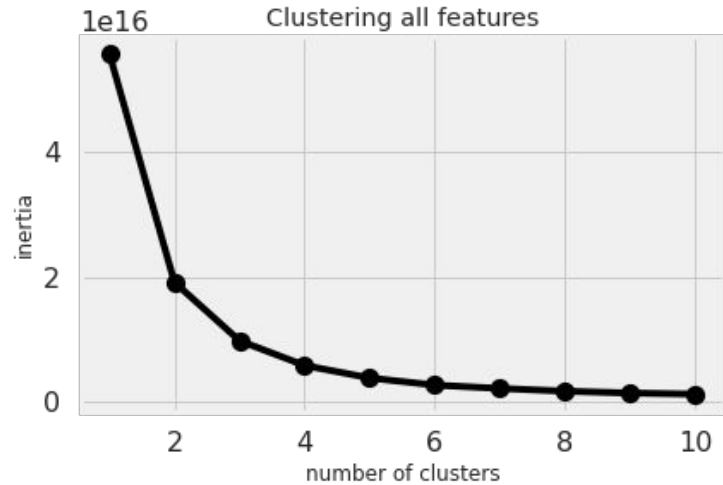
Covid Study

Customer foot traffic heat map



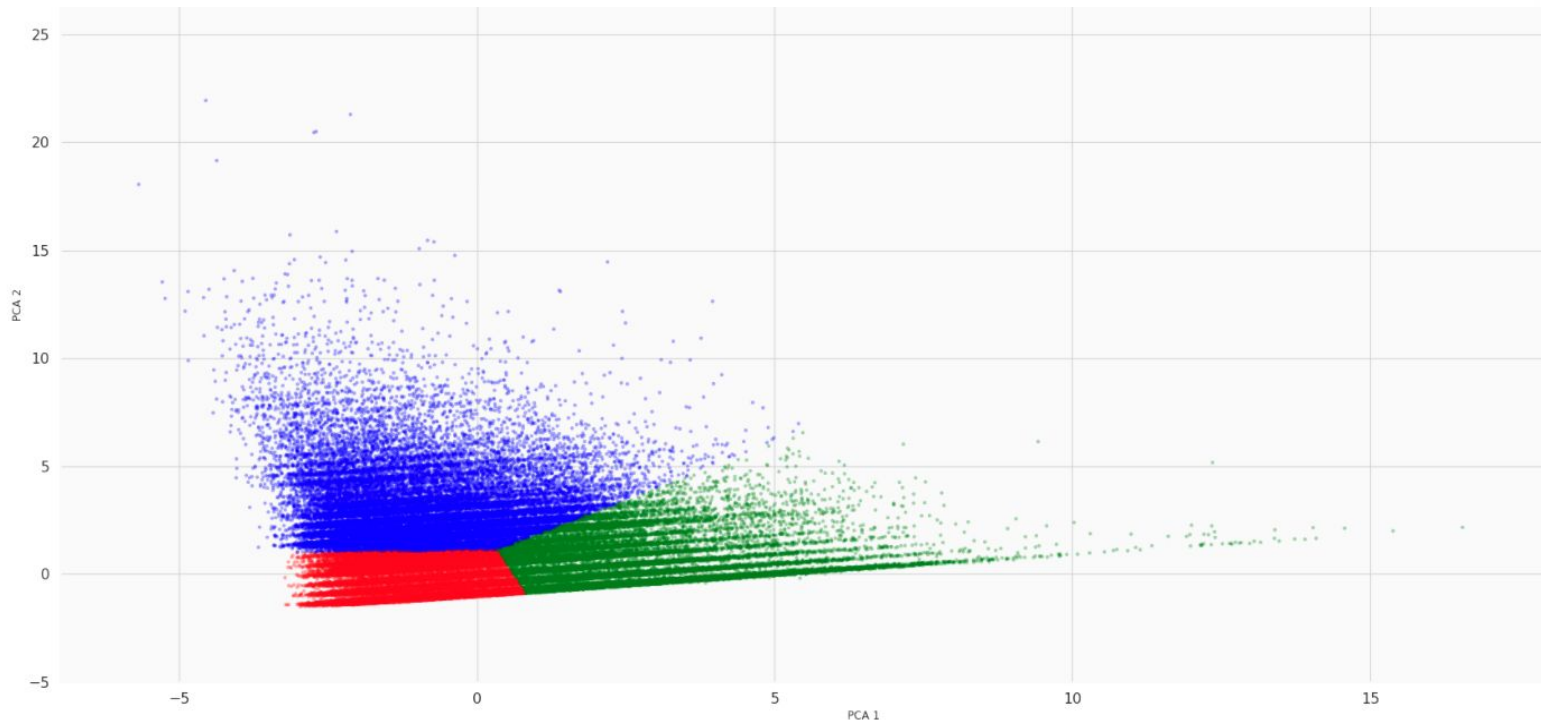
Clustering

K-means and KNN



K=3 optimal number of clusters in both cases.

K-means and KNN



You're Welcome, Banks! 🙌

DS-GA 1007 Project - Group 14

Akhand (aps646)

Mayank (md4570)

Aashvi (avs476)

Zechen (zw2746)

Anshu (am10263)