

Design decisions

I had initially decided to work with the human carbon food print dataset but when I started to analyse the dataset from different sources I was unable to completely understand and form an interesting story through the aspects of the dataset.

Then came across the home credit default risk data set home credit broadens the financial inclusion for the unbanked population by providing positive loan experience. The data set aims at using variety of telco and transactional information to ensure the clients repayment abilities. Doing so will ensure that the clients who are capable of repayment are not rejected of the loan and to make sure the loans are provide with a principal, maturity and repayment calendar.

Initially decided to work with tableau for our interactive graphs and dashboard. But when we did so we were unable to filter certain data and visualize it effectively enough to create a story and answer our initial questions about the data set.

Further decided should try to choose only certain attributes of the dataset and try and produce interactive graphs through python. During our research to plot this interactive visualization effectively we came across Cufflinks package, the library can be installed in python to interact with pandas charting to plot interactive graphs using plotly. Plotly is an online collaborative graphing tool which is accessible by the python API and the graphs produced have data accessible in them.

I found this quiet experimental than traditionally visualizing the data in tableau, we went ahead and decided we would work with the cufflink and plotly package within python and try to design the interactive graphs.

Contribution

My first task to learn the basics of cufflinks and plotly together, this took me about 10 – 12 hours to understand and learn.

Ravikeerthi Teliagi (10503966)

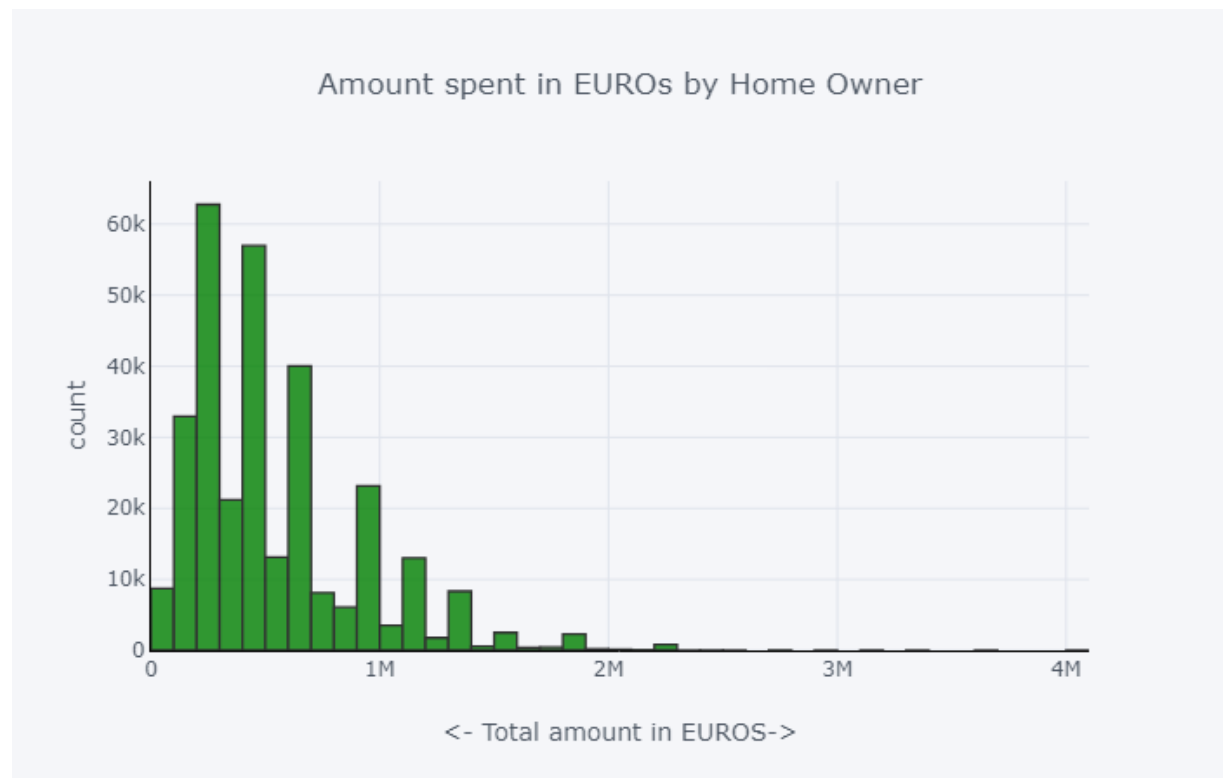
- Choosing dataset
- Choosing software and design techniques
- Determining the initial questions and story
- Cleaning dataset
- Analysing the necessary attributes for the questions
- Choosing types of graph representation
- Generating report

Model 1:

The following right skewed histogram denotes the amount spent by house loan owners on goods.

x-axis denotes the total amount spent in euros and y-axis denotes the count of the goods purchased by the owner.

From the below diagram, we can conclude that maximum goods are purchased around 300k-500k and the minimum goods are purchased above 4M.



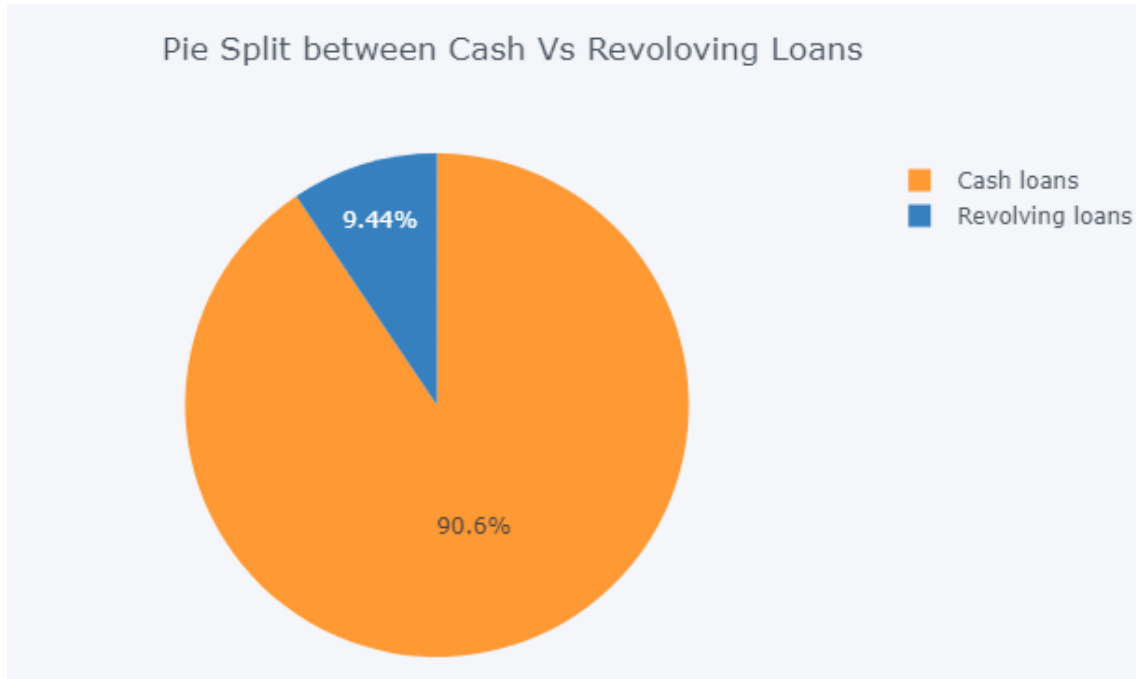
Model 2:

The following pie graph identifies the difference between contract type and the amount spent for the goods for that particular contract.

The two types of contracts are identified as,

- Cash Loans,
- Revolving Loans.

It is clear from the below diagram that the preferred choice is by cash loans.

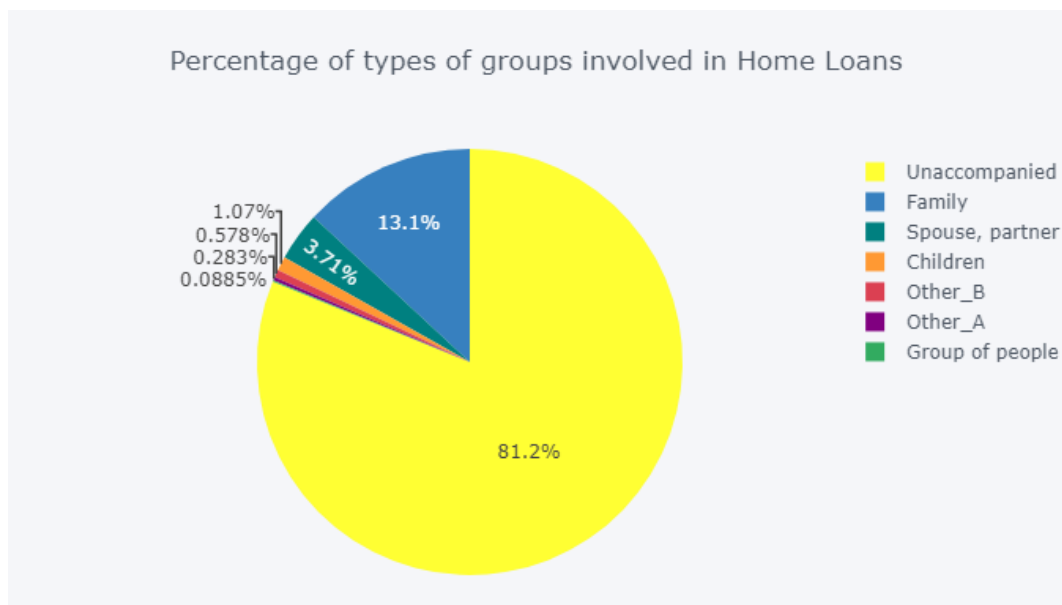


Model 3:

The following pie graph represents the amount of goods bought by the different groups of people involved in the home loans.

The different types of people involved are unaccompanied ones, family, spouse, partner, children and others. Unaccompanied group are the least trusted in terms of returning loan as compared to other category.

The maximum percentage is dominated by unaccompanied group where they do not provide their personal details such as income, housing and family which would lead to most fraud cases in bank loan.



Model 4:

The following line graph denotes the representation of goods purchased to build the homes based on the amount provided by the bank.

Both X-axis and Y-axis is represented in euros.

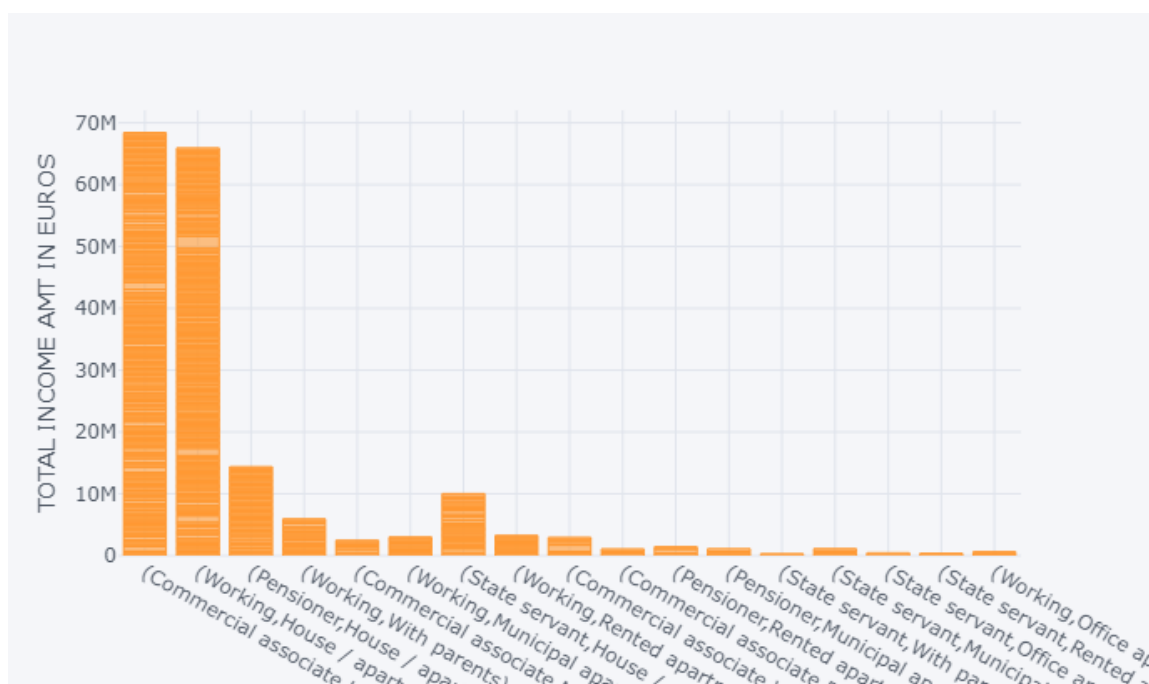
It is clear that amount credited is always higher than the amount spent on goods.



Model 5:

The following Stepped Bar chart represents the difference between income type and housing type against the total income. All these values are represented in Euros.

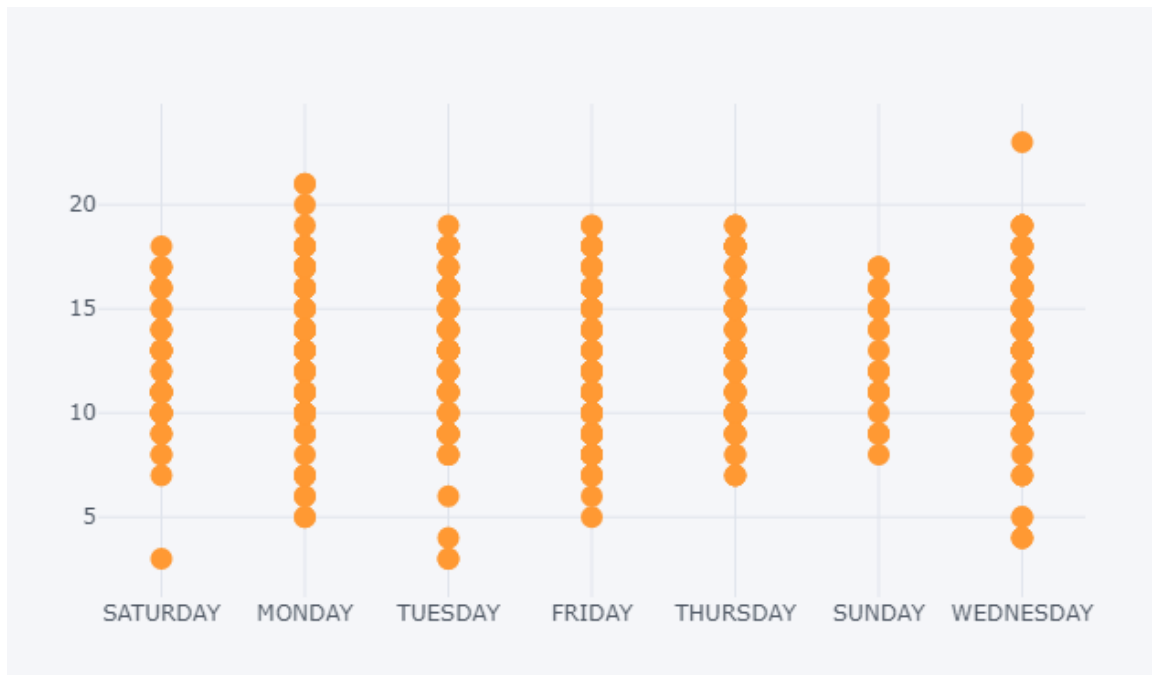
The Maximum income type is dominated by commercial working type and working class.



Model 6:

The following scatter plot denotes the number of hours in a week; the housing loan application was received.

From the values, we can conclude on Monday and Wednesday, most number of applications is received in a week. The graph also shows the mean time and mean number of application received for the day.



Model 7:

The below box plot represents the total amounts sanctioned for the different type of loans such as, Income total, credit, annuity and goods price.

From the below diagram we can conclude that annuity is lower and all others seems to be in par with each other.

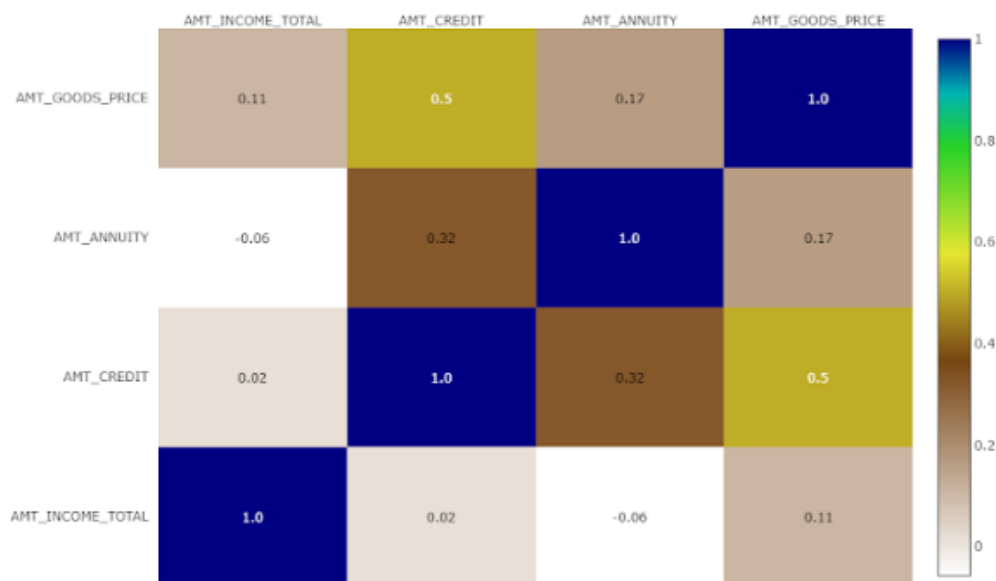


Model 8:

The following correlation plot compares the types of loans against each other.

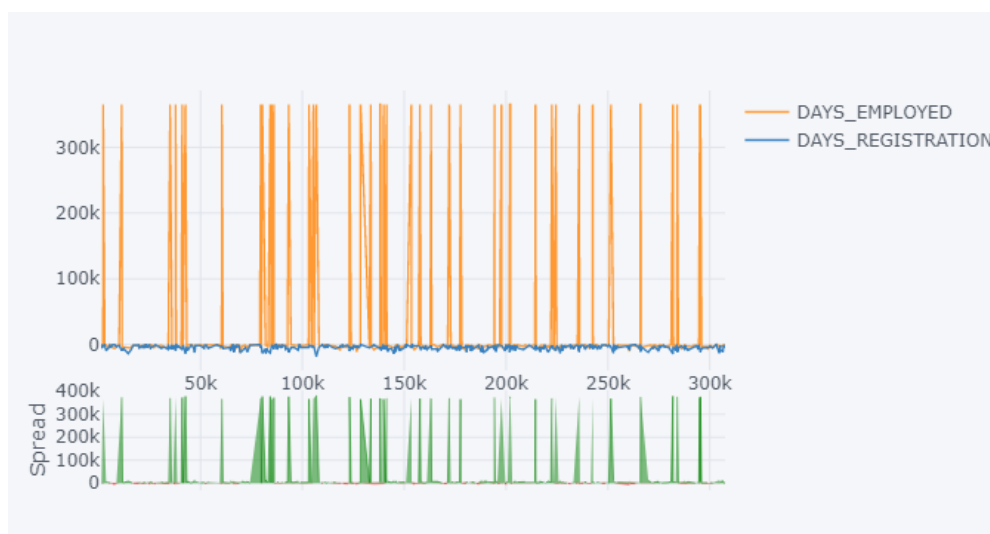
Amount spent on goods price is clearly higher when compared to others.

The darker coloured regions are highly correlated compared to the lighter ones.



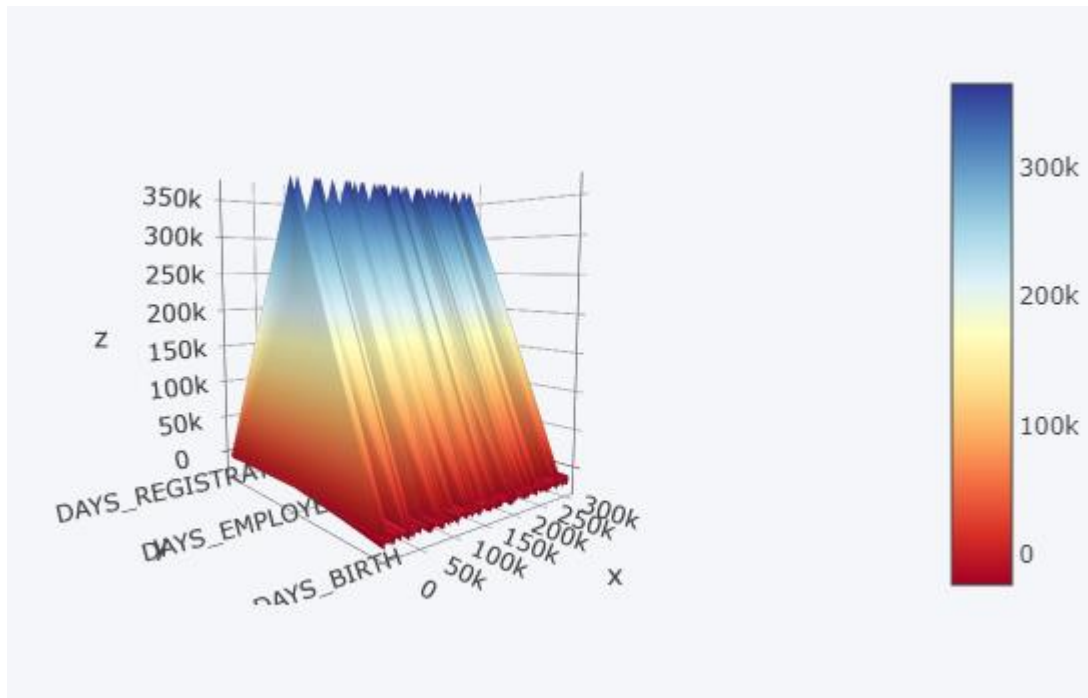
Model 9:

The following spread plot denotes the comparison against the Days employed against the days registered. The spread numbers are corresponded to the continuous values of the days employed and day registered with the bank.



Model 10:

3D interactive chart provides the three-dimensional view of three different fields such as Days employed, Days registered and Date of birth. As the heat map depicts there are shorter triangles in between the taller triangles, the representation of shorter triangles show the customers with all details who are capable of repaying their loans and the taller triangles represent the unaccompanied customers who most probably would not be eligible for a loan and will not be able to repay the loan.



Conclusion

Our initial question was to predict the number of customers that would be eligible for home loan with necessary credential and proper personal checks. Also we aimed at predicting the percentage of customer who were eligible for loan and would be able to repay it on time.

After analysing and visualizing the dataset of home credit we can conclude 18.8 percent of registered customer with the bank are eligible for loan and would be able to repay on time where as 81.2 percent of the registered customers are either without any personal details necessary or would commit fraud.