Bank marketing

Mads, Casper &
Mathias

Bank Marketing Dataset

- Develop a model which predict if a customer is likely to make a deal with the bank

    - Of course, this is not guarantee, you still need to hire skilled salesmen in the bank

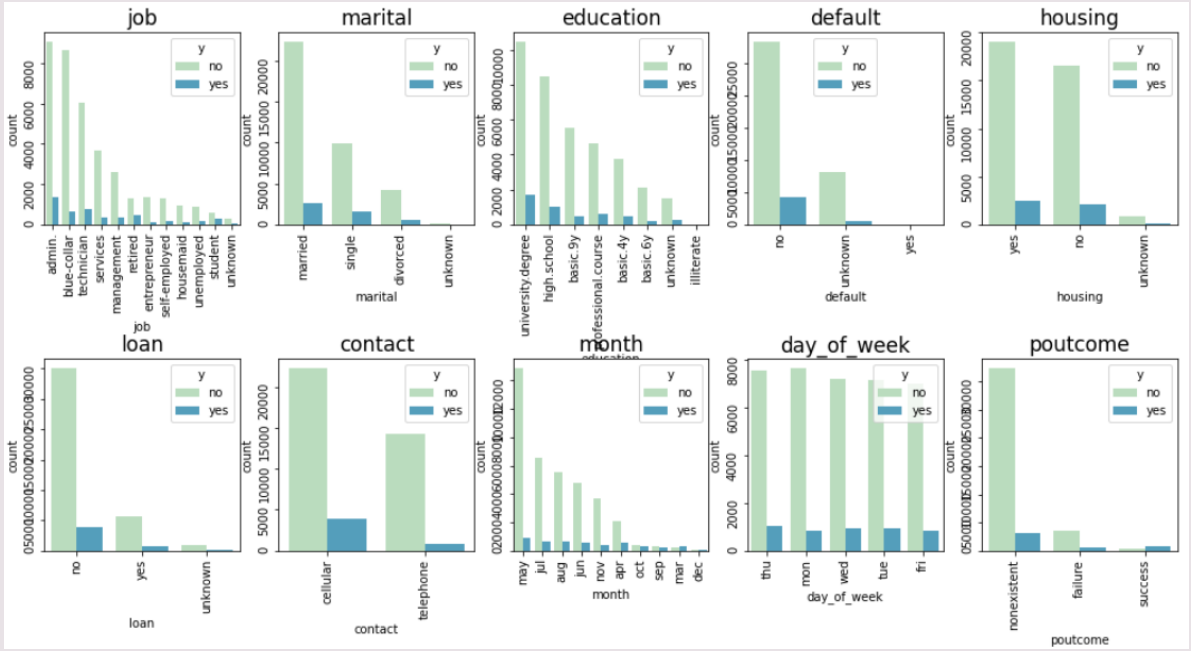
- Segment customers into groups based on demographic attributes

    - Cluster customers on demographic to see which customers are preferred to target

Streamlit link:

EDA Discussion

We explored the data through the proces of making different plots and discussing in the group.



Data Cleaning

- The data set have 4 subgroups

1. Bank Client information
2. Related with the contact of the current campaign
3. Other attributes
4. Social and economic context

Bank Client information is containing unkowns, which all have been dropped

Further has the duration of the call been dropped, since we cannot know this before a call is made

We are dropping the column default because it gives us little to no info, since only a handful has answered yes.

Feature Engineering

- Age has been binned into 4 groups, we presume that these ages are at roughly the same stages in their lives.
- We are using Onehot Encoding to create dummies on variables that have no inherent order such as job, marital and outcome of the previous market campaign
- LabelEncoder has been used to convert variables into numeric values, so it is compatible with ML
- Standard Scaler has been used on variables with an inherent order, e.g the variable age

SML

- Output variable is binary (either yes or no)
- Classification
- We test three classification model for SML

1. *Logistic Regression*
2. *XGB Classifier*
3. *Random Forest*

- We evaluate the model based on the accuracy of the models
- XGB has the next to highest accuracy, we decide to use this due to some error with the logistic regression, we couldn't solve
- The two model have similar result, the difference in the mean accuracy is about 1%

UML

- An attempt was made to make a customer clusterer based on UML working only with the data directly referring to the customer.

- To do this the relevant data was selected and then scaled/encoded based on the type of data.

- Kmeans and UMAP was then used once the data had been preprocessed to properly cluster the data.