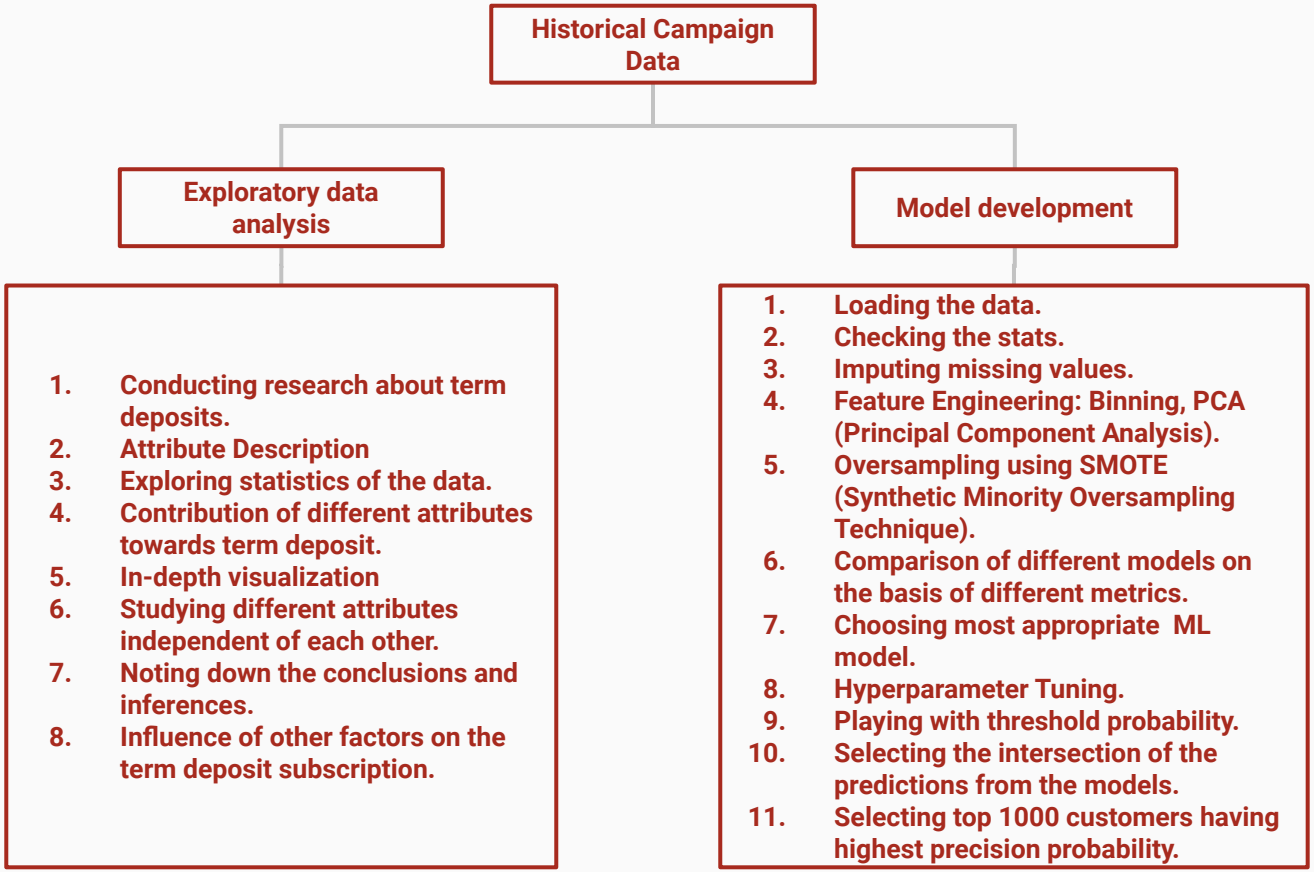
EXL EQ-22: Term Deposit Prediction Through Market Campaigns

Team Name: Cosmotic Hunters

Understanding of the Problem Statement

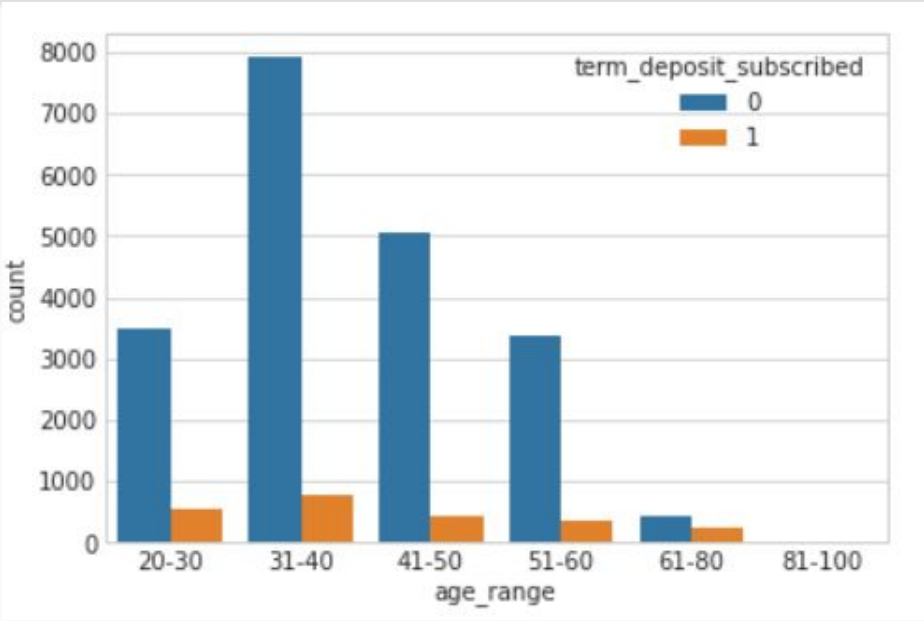
- Knowing that **term deposits** allow banks to **hold onto** a deposit for a specific amount of time, so banks can invest in **higher gain financial products** to make a profit. In addition, banks also hold better chance to persuade term deposit clients into **buying other products** such as **funds or insurance** to further increase their revenues.
- As a result, the bank would like to **identify existing clients** that have **higher chance** to **subscribe** for a term deposit and **focus marketing efforts on such clients**.
- The focus of **marketing efforts** is on the **demands of customers** and their total pleasure. Nonetheless, a variety of factors influence whether or not a **marketing campaign** will be **successful.**
- We have been provided with the **historical data** of the **past market campaigns**.
- Our task is to **study the data** and **derive insights** which will help the company to approach the potential existing customers to **open term deposit**.

Solution Design and Methodology



Data Pre-Processing and Treatment

1. **Imputing missing** values using ML models
2. Studying different attributes.
3. **Binning numerical features** such as age.
4. **Scaling** the attributes containing continuous values using **Standard Scaler**.
5. **Converting attributes** containing string names to numerical using **Label Encoder or One Hot Encoding** (whichever seems appropriate).
6. Selecting the best **oversampling methods** among **SMOTE, ADASYN, Random Over Sampler, etc.**
7. Performing **Principal Component Analysis (PCA)**
8. **Clustering** marital status and education features.
9. Mean Age is aproximately 41 years old.
10. The mean balance is 1,366.154 However, the Standard Deviation (std) is a high number that is 3035.795 so we can understand through this that the balance is heavily distributed across the dataset.



Results obtained after binning the age variable. The customers in the age range 31-40 are most likely to open a term deposit.

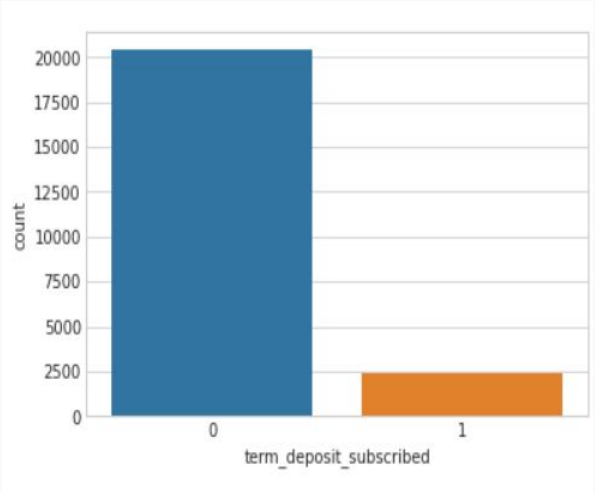
Data Visualization

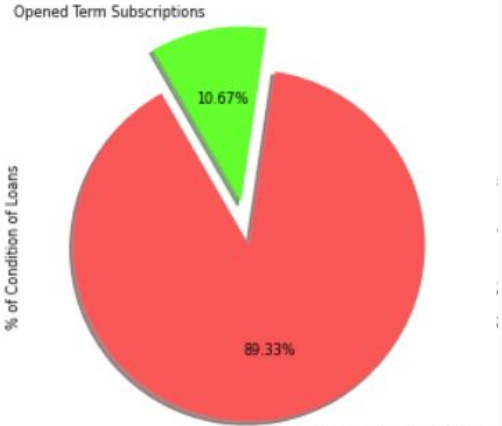
1. Our dataset is **imbalance** as concluded the bar chart. The count of customers who **opened term deposit (label 1) is around 2500** and those **who didn't opened (label 0) is around 21000.**

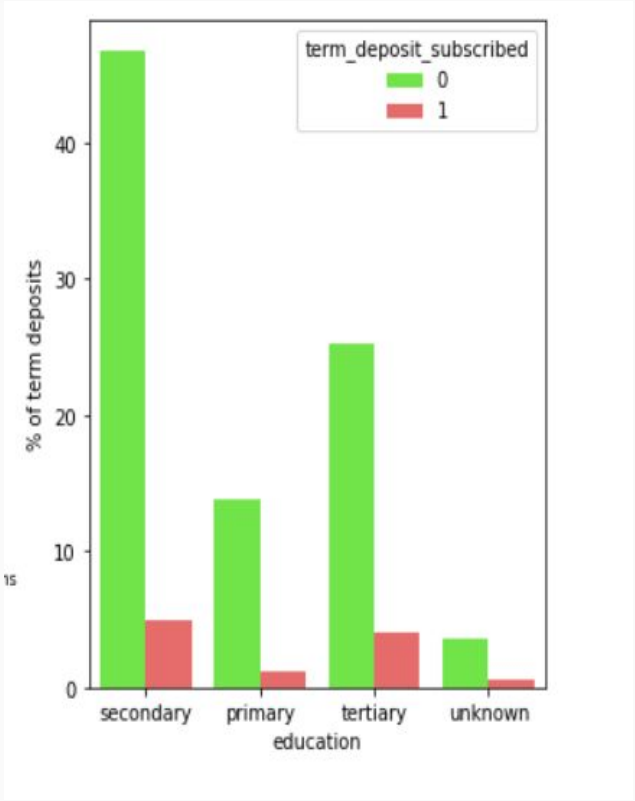
2. Customers having **secondary education level** are most **likely** to **open a term deposit**. Customers with **primary level** of education are most **unlikely** to open a term deposit.

3. Only **10.67%** of the customers who have **loans** are **likely** to **open** a **term deposit** whereas **89.33%** of **customers** having loans are **unlikely** to open a **term deposit**.

4. In the analysis we **didn't find** any **significant insights** for **marriage attribute** other than **most divorced individuals are broke**. No wonder since they have to **split financial assets!**

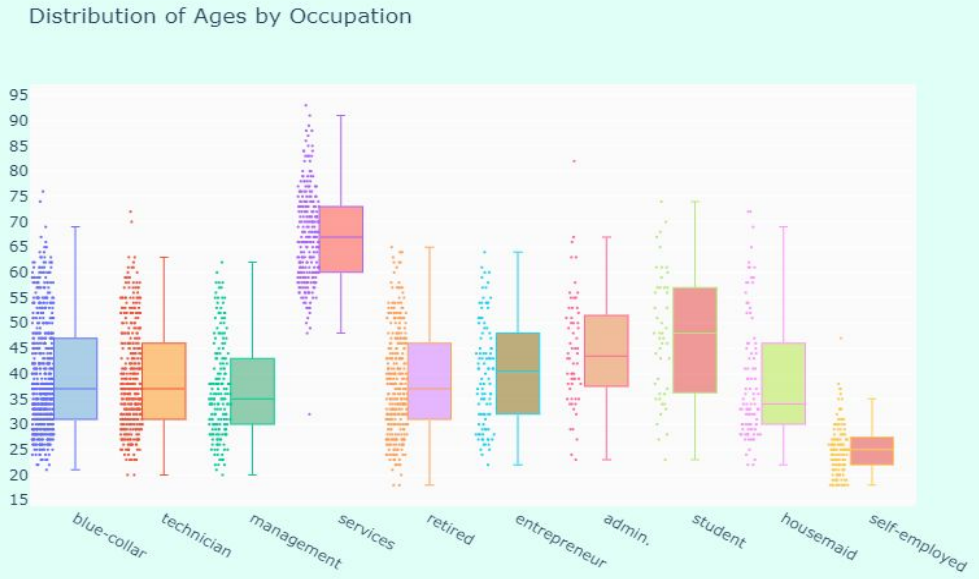






Data Visualization and Model Metrics

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
rf	Random Forest Classifier	0.9117	0.9681	0.9152	0.9085	0.9118	0.8234	0.8235	4.233
et	Extra Trees Classifier	0.8937	0.9576	0.9056	0.8842	0.8947	0.7874	0.7877	5.952
lightgbm	Light Gradient Boosting Machine	0.8861	0.9550	0.8496	0.9160	0.8815	0.7721	0.7742	0.452
dt	Decision Tree Classifier	0.8659	0.8659	0.8765	0.8578	0.8670	0.7318	0.7320	0.231
gbc	Gradient Boosting Classifier	0.8611	0.9270	0.8228	0.8905	0.8553	0.7222	0.7243	4.355
knn	K Neighbors Classifier	0.8502	0.9237	0.9226	0.8054	0.8600	0.7004	0.7080	2.653
ada	Ada Boost Classifier	0.8360	0.9045	0.7946	0.8658	0.8286	0.6719	0.6743	1.238
lr	Logistic Regression	0.7758	0.8494	0.7550	0.7869	0.7705	0.5515	0.5520	1.484
ridge	Ridge Classifier	0.7738	0.0000	0.7541	0.7842	0.7688	0.5475	0.5480	0.078
lda	Linear Discriminant Analysis	0.7738	0.8475	0.7541	0.7842	0.7688	0.5476	0.5480	0.299
svm	SVM - Linear Kernel	0.7479	0.0000	0.7822	0.7429	0.7545	0.4959	0.5065	0.640
nb	Naive Bayes	0.7399	0.8131	0.6865	0.7676	0.7247	0.4797	0.4824	0.041
qda	Quadratic Discriminant Analysis	0.5563	0.5562	0.4999	0.5619	0.4993	0.1124	0.1193	0.141



1. The table shows comparison of **different models** on the basis of different metrics along with the time taken (TT).

2. The **best model** according to **F1 Score** is **Random Forest Classifier.** The F1 Score is 91.18%. We also need to keep in mind the precision and recall as well.

1. The **customers** who are **services** based are more **likely** to **open** a **term deposit later in their life, after the age of 50.**

2. The **self-employed customers** are more likely to open a term deposit during **early age of their life, during the 20s**

Identifying the factors that contributed the most in Term Deposit account opening

1. The active loan status of the customer. If a customer doesn't have an active loan he is more likely to open a term deposit.
2. The education qualification of a customer contributes too towards the opening of a term deposit. The customers who have primary level of education are most unlikely to open a term deposit a compared to customers having secondary or tertiary level of education.
3. The next factor that comes into picture is occupation. The customers working in service sector are more likely to open a term deposit after the age of 50.
4. The customers who are self-employed are most likely to open a term deposit in early stage of their life, probably mid 20s.
5. Marital Status: The impact of a divorce has a significant impact on the balance of the individual.
6. Education: The level of education also has a significant impact on the amount of balance a prospect has.
7. Loans: Whether the prospect has a previous loan has a significant impact on the amount of balance he or she has.
8. Campaign Duration: We see that duration has a high correlation with term deposits meaning the higher the duration, the more likely it is for a client to open a term deposit.
9. Average Campaign Duration: The average campaign duration is 850, clients that were above this average were more likely to open a term deposit.
10. Hence, **the factors that influence the opening of term deposit are bank balance, active loan status, educational qualification, occupation, age, and campaign duration.**