**JOB-A-THON MAY-2021**

**Conclusions & Approach:**

**EDA:**

Exploratory Data Analysis is always the step of any machine learning lifecycle. It involves different steps like Univariate Analysis, Bi-Variate analysis, checking for missing values, outliers, missing value imputation etc. Steps followed and some salient observations were as follows,

1. **Univariate Analysis:**
2. Only Credit\_Product variable had missing values.
3. It was found that the Response variable was not balanced. Around 76% of the respondents did not show higher intent towards a recommended credit card and only around 24% showed interest towards the credit card offered.
4. It was found that most of the customers who were reached out to were Self Employed and a very few were Entrepreneurs.
5. Majority of the customers did not have any existing credit product with the bank.
6. It was also found that majority of the customers were not active in the past 3 months.
7. It was also observed that the Avg\_Account\_Balance variable was positively skewed and the distribution appeared to be a Log Normal Distribution due to presence of positive outliers
8. Most of the customers were below 35 and there was a sizeable population of customers between 40 and 55.
9. **Bi-Variate Analysis:**
10. During the Bi-Variate analysis, it was found that if the customer was an entrepreneur, he or she had a higher chance of showing higher intent towards the credit card. This is contrary to all other variables.
11. It was also found that higher the Avg\_Account\_Balance, higher is the chance that the customer will subscribe to the offer.
12. It was observed that the Average Age of the customers who subscribed to the offer was higher than customers who did not subscribe to the offer.
13. It was also found that customers who subscribed to the offer had longer relationship with the bank.
14. **Correlation:**

It was found most of the numerical dependent variables were not correlated with the independent variable, the correlation was not strong enough in either direction. Only the Vintage variable had a better correlation compared to the other variables. It was also found that Vintage and Age were correlated but the correlation was not strong enough. Below is the diagram representing the correlation,

Chart, treemap chart

Description automatically generated

1. **Missing Value Treatment & Feature Generation:**
2. Credit\_Product Column contained around ~12% missing values and they were replaced with ‘Yes’.
3. Following feature Generation techniques were tried, but were not found to be useful, Binning of Numerical columns like Age, Vintage etc., creating interaction features between categorical variables. Grouped features were generated using Avg\_Account\_Balance and categorical variables.

**Modelling:**

Modelling always comes second to feature engineering and exploratory data analysis, having said that, it plays an important role in the entire predictive modelling purpose.

1. Deep Learning was used with the help of FastAi library.
2. Model took around 30 min for training and testing
3. Ended up with an roc\_auc score of 0.872.

**Room for Improvement:**

There is always room for improvement in the field of machine learning.

1. Generating more features based on domain knowledge always helps the cause.