- **Topic** : Credit Card Lead Prediction – Increase the credit card product customer base by identifying the most interested from the eligible base of existing customers.

- **Business Problem** : A Private Bank offers variety of products like Checking / Savings accounts, investment products, credit products etc. The bank intends to cross-sell products to its existing customer base via different channels like emails, telephone calls, recommendations on online banking interfaces like mobile apps etc. As a part of current business opportunity, bank plans on identifying the potential customers from the list of eligible customers.

- **Datasets** – https://www.kaggle.com/sajidhussain3/jobathon-may-2021-credit-card-lead-prediction
  https://github.com/priyalagarwal27/Credit-Card-Lead-Prediction/blob/main/dataset.zip

  There are two datasets – train and test. Following features are present in the train dataset
  - ID : Customer ID
  - Gender : Male / Female
  - Age : Age in years
  - Region Code : Represents the Region where the customer resides
  - Occupation : Occupation status e.g. Salaried, Self-employed, Entrepreneur, Other etc.
  - Channel Code : Method of connecting with the customers for credit card product offers
  - Vintage : Vintage for the Customers (in months)
  - Credit Product : If the customer has any active product (home loan, credit card etc.)
  - Avg Account Balance : Average account balance for the past 12 months
  - Is active : If the customer is active in last 3 months
  - Is lead : (Target variable) if the customer is interested

  The test dataset contains all of the above but the target variable, which would need to come out of the final predictions.

- **Methods** : I would look to perform statistical analysis with help of statistics like mean, mode, median, range, quartiles, outliers etc. Additionally, would also be looking performing Exploratory Data Analysis using visuals like scatter plots, box plots, bar charts etc. as applicable. Dropping redundant or unrequired data features can help with the analysis as well.

- **Ethical Considerations** : Ensuring the privacy of the customers whose data is being utilized for analysis is of utmost importance. Hence, will need to ensure that the dataset does not contain any personally identifiable information. The current dataset does not contain such information.

- **Challenges/Issues** : Ensuring that all the critical data elements hold valid data points is important. Currently, some of the data fields have missing information. So, these may not be very helpful with analysis and result in slightly reduced dataset for analysis

- **References** : The dataset reference also has a sample submission dataset to verify the results. https://www.kaggle.com/sajidhussain3/jobathon-may-2021-credit-card-lead-prediction