**Project Name**: Bank Customer Segmentation

Prepared By: Harshit Bhalsod(21IM60R13), Industrial and Systems Engineering Department, IIT Kharagpur

**What Is Customer Segmentation?**

Any company or organization has large number of customers having different traits related to company like age, income, values, lifestyle, and more. Customer segmentation is the process of dividing a customer dataset into specific groups based on shared traits. This is very useful method to identify potential customers of company and further company can use this method to market their products and/or services to specific customers. For example, through customer segmentation, banks can get to know their customers on a more personal level and offer them more tailored products and services.

**What are the business benefits of customer segmentation in banking?**

Customer segmentation is often a top marketing priority for banks, and for good reason. Because segmentation solutions help banks group customers by behaviour, they can offer more tailored products and services. Moreover, by better understanding customer preferences, marketers can maximize up-selling opportunities and encourage customers to explore related services. Other ways that segmentation can help banks:

* Decide on the right kind of promotional content to send each customer
* Select the right marketing channels for the right audience
* Identify new and profitable segments and launch innovative products and services

**Objective:**

* Perform Clustering / Segmentation on the dataset and identify popular customer groups along with their definitions/rules
* Perform transaction-related analysis to identify interesting trends that can be used by a bank to improve / optimize their user experiences

**GitHub Repository link:** [**https://github.com/hdb1603/bank-customers-segmentation**](https://github.com/hdb1603/bank-customers-segmentation)

**Steps involved in implementing customer segmentation using Python language:**

1. Importing Libraries:
   1. NumPy – to perform numerical operations on large number of numerical rows and columns
   2. Pandas – store dataset in data frame
   3. Seaborn and matplotlib – to prepare bar plots, histograms, line plots etc.
   4. Scikit-learn – to import clustering algorithms, encode data, standardize data features scales, data pre-processing modules, model performance measures, dimensionality reduction modules
   5. Wget – to retrieve data from URL
2. Feature Exploration and Pre-processing:
   1. Loading dataset, checking distributions of features of data.
   2. Gender counts, converting birthdate into age, dropping Customer ID, Transaction ID etc.
   3. Treating Nan records, removing outliers
3. Exploratory Data Analysis
   1. Gender wise Transaction amount, Gender wise account balance, Month wise spending for both genders, Day wise spending.
   2. Overall month wise transaction amount and month wise account balance
   3. Location wise transaction behaviour analysis
   4. Age wise Spending behaviour and Age wise Transaction behaviour
   5. Checking for multi-collinearity
4. Feature Scaling
   1. Creating data frame having same features of original data but standardized scale
   2. Transform right skewed Age, Account Balance and Transaction data into normal form
5. Unsupervised Machine Learning
   1. Fitting K-means Clustering, Silhouette score
   2. Checking for other number of clusters with respective Sum Squared Distance and selecting optimal cluster numbers
   3. Plotting scattered plot with optimal cluster labels
   4. Rechecking number of clusters
   5. Hierarchical clustering with minmax scaled data using Agglomerative Clustering.
   6. Principal Component Analysis and labelling clusters on original features.
6. Findings
   1. Clustering algorithm has been able to cluster customers in 2 group.
   2. The first group is having dynamic customers having low account balance and frequently expend less cash on transactions barring a few who perform large-value transactions.
   3. The second group includes the more conservative and money saving-minded people who, in spite of having really high account balances, spend the least amount of money in transactions, thereby judiciously maintaining their savings accounts.