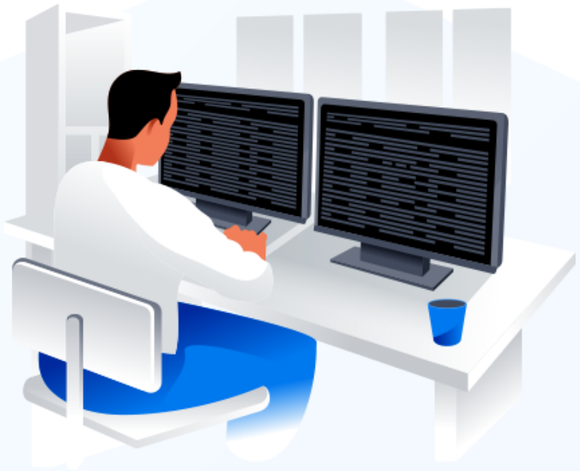
Capstone Session 7



Machine Learning for Modeling

Machine Learning for Strategic Decision-Making

The intelligence provided by Aura will help customers make decisions for their omnichannel marketing and customer acquisition programs.

Aura’s early adopter customers

Bike Rentals

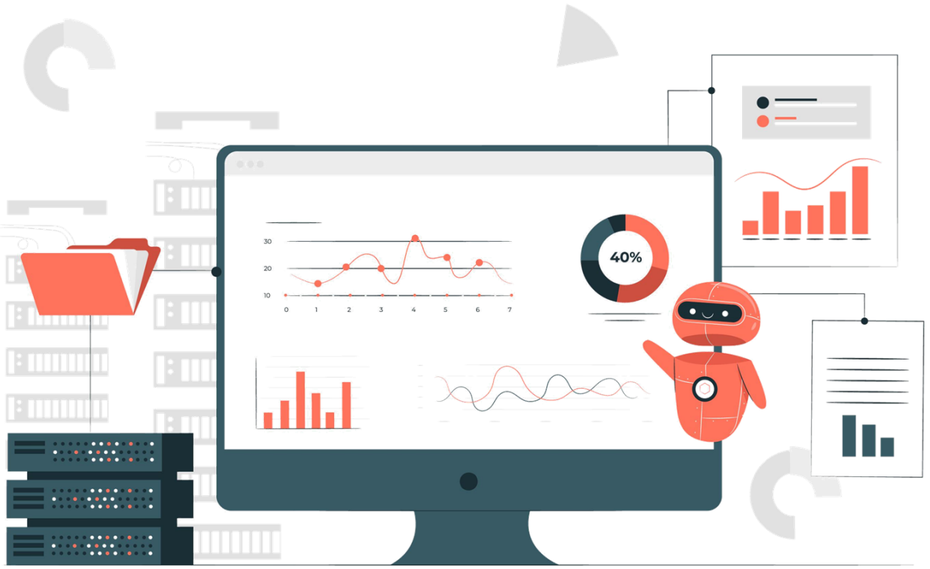
Realtor Association

Credit Card Agency

Startups

Project Statement

Aura must do the following:



- Predict bike-sharing demand
- Classify incomes
- Cluster credit card users
- Build a recommendation engine

Session 7 : Dataset Description

CC GENERAL.csv

Variable	Description	Variable	Description
CUSTID	Identification of Credit Card holder (Categorical)	BALANCE	Balance amount left in their account to make purchases
BALANCEFREQUENCY	The frequency of balance getting updated; a score between 0 and 1	PURCHASES	Number of purchases made from the account
ONEOFFPURCHASES	The maximum purchase amount in one-go	PURCHASESINSTALLMENTSFREQUENCY	The frequency of purchases in installments
CASHADVANCEFREQUENCY	The frequency of cash in advance being paid	CASHADVANCETRX	The number of transactions made with **Cash in Advance**
INSTALLMENTSPURCHASES	Amount of purchase done in installment	CASHADVANCE	Cash in advance given by the user

Session 7 : Dataset Description

CC GENERAL.csv

Variable	Description	Variable	Description
ONEOFFPURCHASESFREQUENCY	The frequency of purchases in one-go	PURCHASESFREQUENCY	The frequency of purchases being made; a score between 0 and 1
PURCHASESTRX	Number of purchase transactions made	TENURE	Tenure of credit card service for the user
PAYMENTS	Payment done by the user	CREDITLIMIT	Limit of Credit Card for user
PRCFULLPAYMENT	Percentage of full payment paid by the user	MINIMUM_PAYMENTS	Minimum number of payments made by the user

Session 7: Clustering Credit Card Users with PCA and K-means

Task: Cluster the credit card users into different groups to find any meaningful patterns. Use Principal Component Analysis (PCA) to reduce the dimension of the feature space and then use the K-means algorithm to find clusters. Import relevant Python libraries.

- Load dataset (CC GENERAL.csv, the name of the file when downloaded from Kaggle.)
- Check for null values and handle those values.
- Perform feature scaling using StandardScaler.
- Perform PCA with all the columns and plot number of components vs. PCA cumulative explained variance. From the plot, identify the number of components required to cover 85% of the variance.
- Perform PCA with 2 principal components with the aim of visualizing clustering.
- Find the 2 columns which give the most covariances.
- Interpret the results of PCA by looking at the covariance matrix (use get_covariance() method of PCA).

Session 7: Clustering Credit Card Users with PCA and K-means

- Perform K Means Clustering on the 2 component PCA transformed data with clusters ranging from 2 to 11 and plot the K Means inertia against the number of clusters (Elbow Method). From the elbow plot, identify the ideal required number of clusters.
- Perform K Means Clustering on the 2 component PCA transformed data with the ideal number of clusters found in the sixth bullet point.
- Visualize the clusters on a scatter plot between 1st PCA and 2nd PCA component giving different colors to each cluster.

Thank You