- **Yassine Sfaihi**
- www.linkedin.com/in/yassinesfaihi
- www.github.com/yassinesfaihi
- www.kaggle.com/yassinesfaihi

[Author] : Yassine Sfaihi

Customer Clustering with PCA, KMeans and Agglomerative for a Marketing Campaign

Introduction

This notebook focuses on exploring and preprocessing marketing campaign data using Python.
The data set contains information about customers who have received marketing emails and the response they made. The main aim of the notebook is to cluster the customers into similar groups based on their attributes and response to the marketing emails.

Tools used

- Pandas
- Numpy
- Missingno
- Seaborn
- Matplotlib
- Plotly
- Sklearn
    - StandardScaler
    - PCA
    - KMeans
    - Agglomerative Clustering
    - Silhouette Score
    - Davies-Bouldin Score

Data Preprocessing

- Handling missing values using Missingno library
- Cleaning and transforming data using Pandas and Numpy
- Standardizing the data using Sklearn's StandardScaler
- Applying PCA for dimensionality reduction

Exploratory Data Analysis

- Plotting histograms and density plots to visualize the distribution of data
- Using Seaborn's pairplot and heatmap to find correlations between features
- Plotting interactive graphs using Plotly

Customer Clustering

- **Yassine Sfaihi**
- www.linkedin.com/in/yassinesfaihi
- www.github.com/yassinesfaihi
- www.kaggle.com/yassinesfaihi

- Fitting KMeans model with the number of clusters set to 3
- Evaluating the performance of the KMeans model using the silhouette score
- Visualizing the clusters using 3D plots in Matplotlib
- Fitting Agglomerative Clustering and evaluating its performance using the Davies-Bouldin score

Clustering Visualization

- The clusters are visualized using 3D scatterplots.
- The scatterplots are interactive and can be rotated to see the clusters from different angles.

Conclusion

The notebook provides a comprehensive guide on how to explore and preprocess marketing campaign data and cluster the customers based on their attributes and response to the marketing emails. The tools used are Pandas, Numpy, Missingno, Seaborn, Matplotlib, Plotly, and Scikit-learn. The results of the notebook can be used for targeted marketing campaigns and for a better understanding of customer behavior.