**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 4th July 2024 |
| Team ID | 739794 |
| Project Title | Cost Prediction of Acquiring a Customer |
| Maximum Marks | 6 Marks |

**Data Exploration and Preprocessing Template**

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

|  |  |
| --- | --- |
| **Section** | **Description** |
| Data Overview | Dimensions: (60428, 40) |
| Univariate Analysis |  |
| Bivariate Analysis |  |
| Multivariate Analysis |  |
| Outliers and Anomalies | - |
| **Data Preprocessing Code Screenshots** | |
| Loading Data | df = pd.read\_csv('/content/media prediction and its cost.csv') |
| Handling Missing Data | df.isnull().any() |
| Data Transformation | from sklearn.preprocessing import LabelEncoder  def encoder(data, variable):  # Corrected parameter name to 'variable'      lb = LabelEncoder()      df[variable] = lb.fit\_transform(df[variable])      return lb |
| Feature Engineering | food\_category\_le = encoder(df,'food\_category')  brand\_name\_le = encoder(df,'brand\_name')  food\_department\_le = encoder(df,'food\_department')  food\_family\_le = encoder(df,'food\_family')  promotion\_name\_le = encoder(df,'promotion\_name')  store\_city\_le = encoder(df,'store\_city')  #unit\_per\_case\_le = encoder(df,'unit\_per\_case')  net\_weight\_le = encoder(df,'net\_weight')  sales\_le = encoder(df,'sales\_country')  martial\_le = encoder(df,'marital\_status') |
| Save Processed Data | import pickle  pickle.dump(rf,open('customers.pkl','wb'))  pickle.dump(food\_category\_le,open('food\_category\_le.pkl','wb'))  pickle.dump(brand\_name\_le,open('brand\_name\_le.pkl','wb'))  pickle.dump(promotion\_name\_le,open('promotion\_name\_le.pkl','wb'))  pickle.dump(store\_city\_le,open('store\_city\_le.pkl','wb')) |