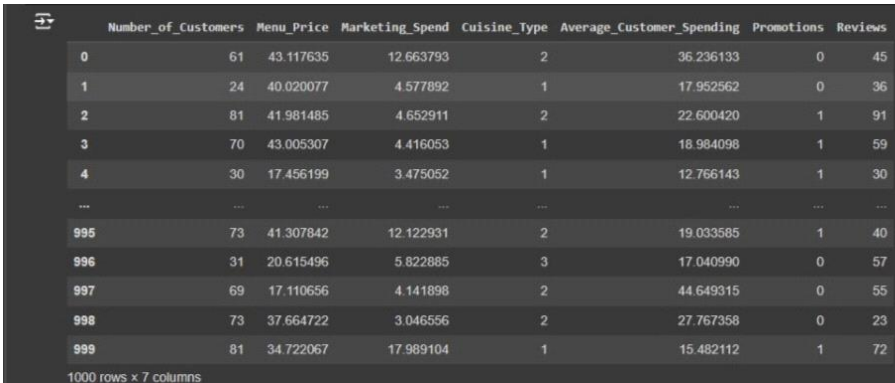


## Data Collection and Preprocessing Phase

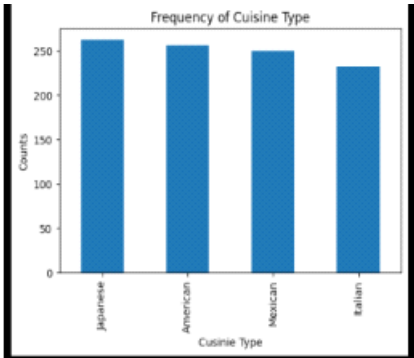
Date	15 July 2024
Team ID	739746
Project Title	Forecasting Feasts: A Culinary journey into Restaurant Revenue Prediction
Maximum Marks	6 Marks

### Data Exploration and Preprocessing Template

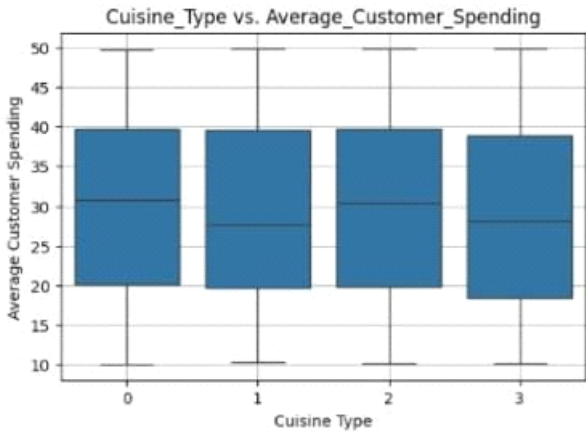
Data exploration and preprocessing involve a series of steps to understand the data and prepare it for analysis. This typically starts with loading the data and examining its structure, such as the number of rows and columns, data types, and summary statistics. Next, data cleaning is performed to handle missing values, remove duplicates, and correct errors. Feature engineering may be conducted to create new variables or transform existing ones. Finally, data is scaled or normalized if necessary, and categorical variables are encoded. Throughout this process, data visualization is often used to identify patterns and outliers.

Section	Description
Data Overview	<div><p>Dimension:</p><p>1000 rows x 7 columns</p><p>Descriptive statistics:</p><div></div></div>

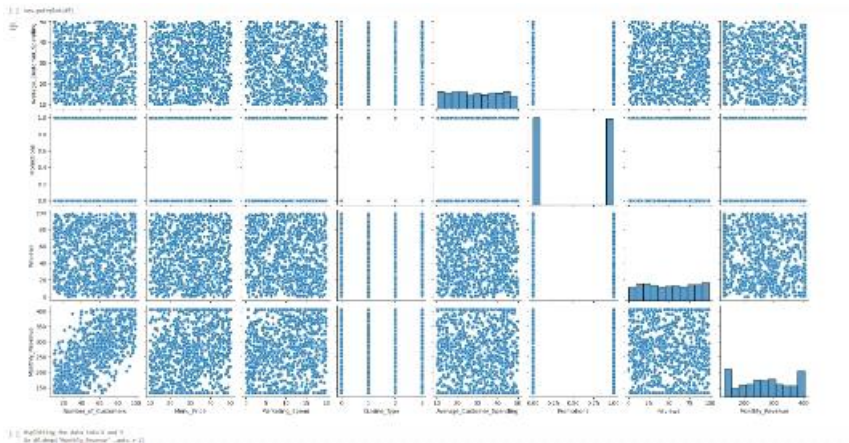
Univariate Analysis

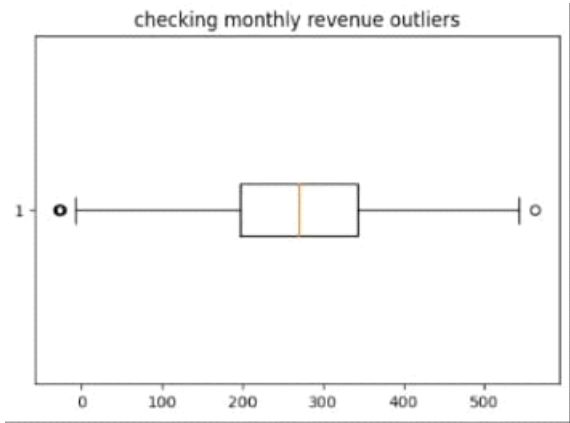


Bivariate Analysis



Multivariate Analysis



Outliers and Anomalies	
Data Preprocessing Code Screenshots	
Loading Data	<pre> #splitting into training and testing dataset from sklearn.model_selection import train_test_split  x_train,x_test,y_train,y_test = train_test_split(X, Y, test_size=0.2, random_state=30) </pre>
Handling Missing Data	<pre> df.shape (1000, 8) </pre> <pre> &lt;class 'pandas.core.frame.DataFrame'&gt; RangeIndex: 1000 entries, 0 to 999 Data columns (total 8 columns):  # Column          Non-Null Count  Dtype ---  --  0 Number_of_Customers  1000 non-null    int64  1 Menu_Price          1000 non-null    float64  2 Marketing_Spend     1000 non-null    float64  3 Cuisine_Type        1000 non-null    object  4 Average_Customer_Spending 1000 non-null    float64  5 Promotions          1000 non-null    int64  6 Reviews             1000 non-null    int64  7 Monthly_Revenue     1000 non-null    float64 dtypes: float64(4), int64(3), object(1) memory usage: 62.6+ KB </pre> <pre> Number_of_Customers    0 Menu_Price              0 Marketing_Spend         0 Cuisine_Type            0 Average_Customer_Spending 0 Promotions              0 Reviews                 0 Monthly_Revenue         0 dtype: int64 </pre>
Data Transformation	<pre> from sklearn.tree import DecisionTreeRegressor  decision_tree_model = DecisionTreeRegressor() decision_tree_model.fit(x_train,y_train) y_pred = decision_tree_model.predict(x_test)  mse_dt = mean_squared_error(y_test,y_pred) r2_dt =r2_score(y_test,y_pred)  print('MSE: ',mse_dt) print('R2 score: ',r2_dt) </pre>
Feature Engineering	-