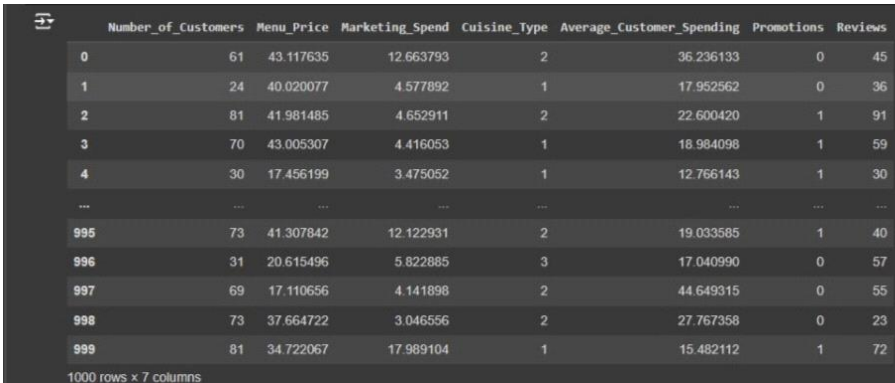


Data Collection and Preprocessing Phase

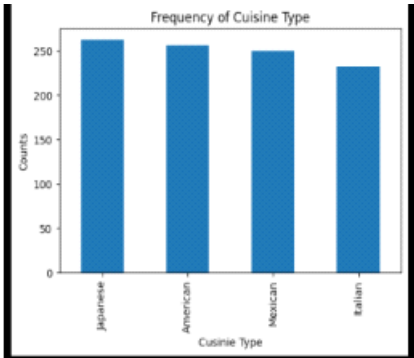
Date	15 July 2024
Team ID	739655
Project Title	Forecasting Feasts: A Culinary journey into Restaurant Revenue Prediction
Maximum Marks	6 Marks

Data Exploration and Preprocessing Template

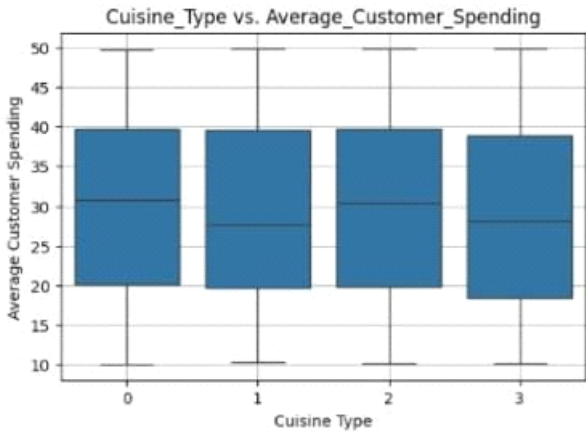
Data exploration and preprocessing involve a series of steps to understand the data and prepare it for analysis. This typically starts with loading the data and examining its structure, such as the number of rows and columns, data types, and summary statistics. Next, data cleaning is performed to handle 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><table><tr><th></th><th>Number_of_Customers</th><th>Menu_Price</th><th>Marketing_Spend</th><th>Cuisine_Type</th><th>Average_Customer_Spending</th><th>Promotions</th><th>Reviews</th></tr><tr><td>0</td><td>61</td><td>43.117635</td><td>12.663793</td><td>2</td><td>36.236133</td><td>0</td><td>45</td></tr><tr><td>1</td><td>24</td><td>40.020077</td><td>4.577892</td><td>1</td><td>17.952562</td><td>0</td><td>36</td></tr><tr><td>2</td><td>81</td><td>41.981485</td><td>4.652911</td><td>2</td><td>22.600420</td><td>1</td><td>91</td></tr><tr><td>3</td><td>70</td><td>43.005307</td><td>4.416053</td><td>1</td><td>18.984098</td><td>1</td><td>59</td></tr><tr><td>4</td><td>30</td><td>17.456199</td><td>3.475052</td><td>1</td><td>12.766143</td><td>1</td><td>30</td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr><tr><td>995</td><td>73</td><td>41.307842</td><td>12.122931</td><td>2</td><td>19.033585</td><td>1</td><td>40</td></tr><tr><td>996</td><td>31</td><td>20.615496</td><td>5.822885</td><td>3</td><td>17.040990</td><td>0</td><td>57</td></tr><tr><td>997</td><td>69</td><td>17.110656</td><td>4.141898</td><td>2</td><td>44.649315</td><td>0</td><td>55</td></tr><tr><td>998</td><td>73</td><td>37.664722</td><td>3.046556</td><td>2</td><td>27.767358</td><td>0</td><td>23</td></tr><tr><td>999</td><td>81</td><td>34.722067</td><td>17.989104</td><td>1</td><td>15.482112</td><td>1</td><td>72</td></tr></table><p>1000 rows x 7 columns</p></div></div>		Number_of_Customers	Menu_Price	Marketing_Spend	Cuisine_Type	Average_Customer_Spending	Promotions	Reviews	0	61	43.117635	12.663793	2	36.236133	0	45	1	24	40.020077	4.577892	1	17.952562	0	36	2	81	41.981485	4.652911	2	22.600420	1	91	3	70	43.005307	4.416053	1	18.984098	1	59	4	30	17.456199	3.475052	1	12.766143	1	30	995	73	41.307842	12.122931	2	19.033585	1	40	996	31	20.615496	5.822885	3	17.040990	0	57	997	69	17.110656	4.141898	2	44.649315	0	55	998	73	37.664722	3.046556	2	27.767358	0	23	999	81	34.722067	17.989104	1	15.482112	1	72
	Number_of_Customers	Menu_Price	Marketing_Spend	Cuisine_Type	Average_Customer_Spending	Promotions	Reviews																																																																																										
0	61	43.117635	12.663793	2	36.236133	0	45																																																																																										
1	24	40.020077	4.577892	1	17.952562	0	36																																																																																										
2	81	41.981485	4.652911	2	22.600420	1	91																																																																																										
3	70	43.005307	4.416053	1	18.984098	1	59																																																																																										
4	30	17.456199	3.475052	1	12.766143	1	30																																																																																										
...																																																																																										
995	73	41.307842	12.122931	2	19.033585	1	40																																																																																										
996	31	20.615496	5.822885	3	17.040990	0	57																																																																																										
997	69	17.110656	4.141898	2	44.649315	0	55																																																																																										
998	73	37.664722	3.046556	2	27.767358	0	23																																																																																										
999	81	34.722067	17.989104	1	15.482112	1	72																																																																																										

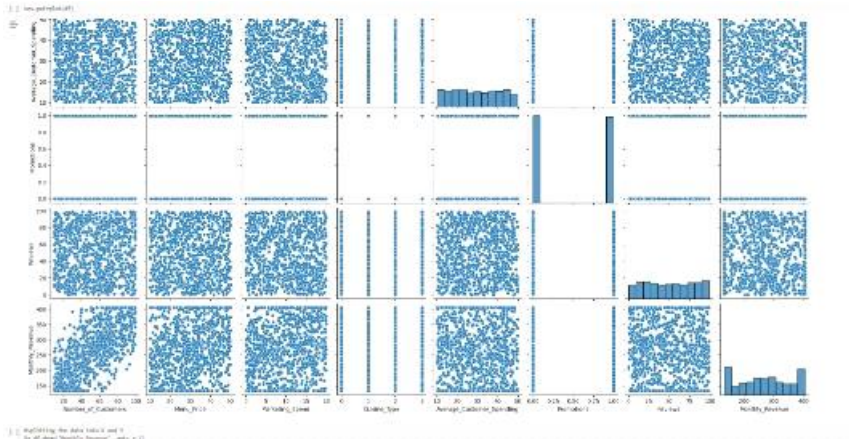
Univariate Analysis

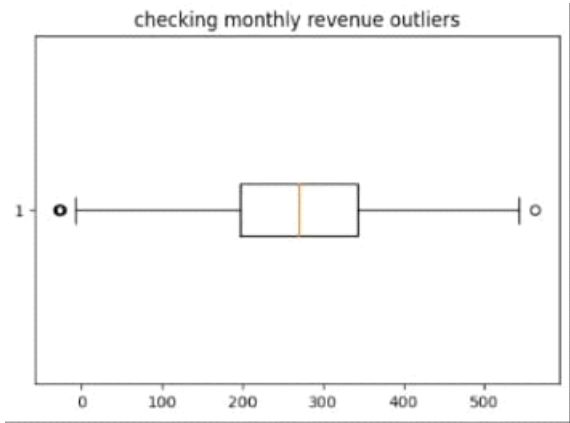


Bivariate Analysis



Multivariate Analysis



Outliers and Anomalies	 <p>A box plot titled "checking monthly revenue outliers". The x-axis represents monthly revenue, ranging from 0 to 500. The plot shows a median around 250, with the interquartile range (IQR) spanning from approximately 200 to 350. Whiskers extend from the box to the minimum value (around 50) and the maximum value (around 550). There is a single outlier point plotted at approximately 550, which is significantly higher than the rest of the data.</p>
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> <class 'pandas.core.frame.DataFrame'> 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	-