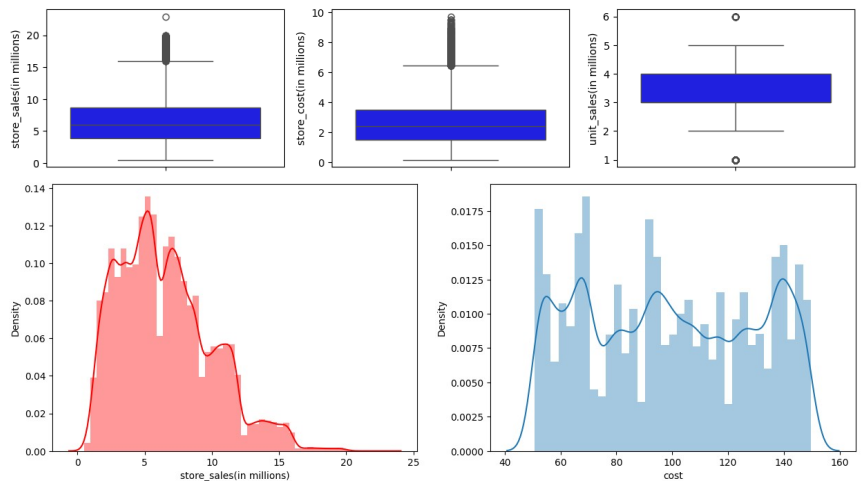


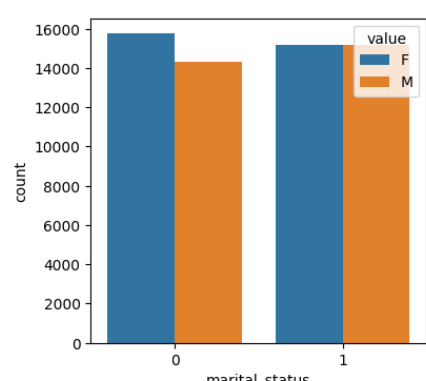
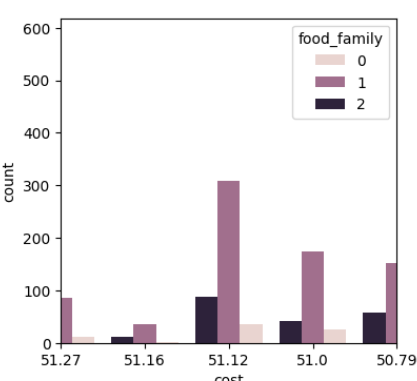
## Data Collection and Preprocessing Phase

Date	4 <sup>th</sup> 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	 <p>8 rows x 23 columns</p>
Bivariate Analysis	

	
<p>Multivariate Analysis</p>	
<p>Outliers and Anomalies</p>	<p>-</p>
<p>Data Preprocessing Code Screenshots</p>	
<p>Loading Data</p>	<pre>df = pd.read_csv('/content/media prediction and its cost.csv')</pre>
<p>Handling Missing Data</p>	<pre>df.isnull().any()</pre>
<p>Data Transformation</p>	<pre>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</pre>

Feature Engineering	<pre> 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') marital_le = encoder(df, 'marital_status') </pre>
Save Processed Data	<pre> import pickle pickle.dump(rf, open('customers.pkl', 'wb')) pickle.dump p(food_category_le, open('food_category_le.pkl', 'wb' )) pickle.dump p(brand_name_le, open('brand_name_le.pkl', 'wb')) pickle.dump p(promotion_name_le, open('promotion_name_le.pkl', 'w b')) pickle.dump p(store_city_le, open('store_city_le.pkl', 'wb')) </pre>