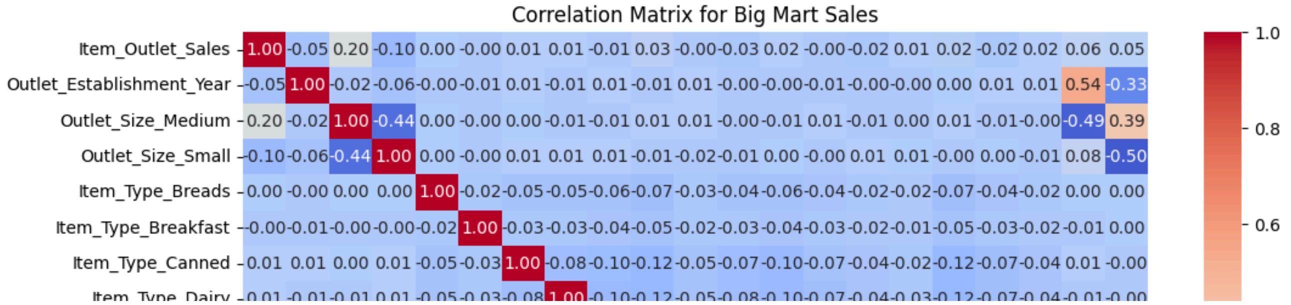
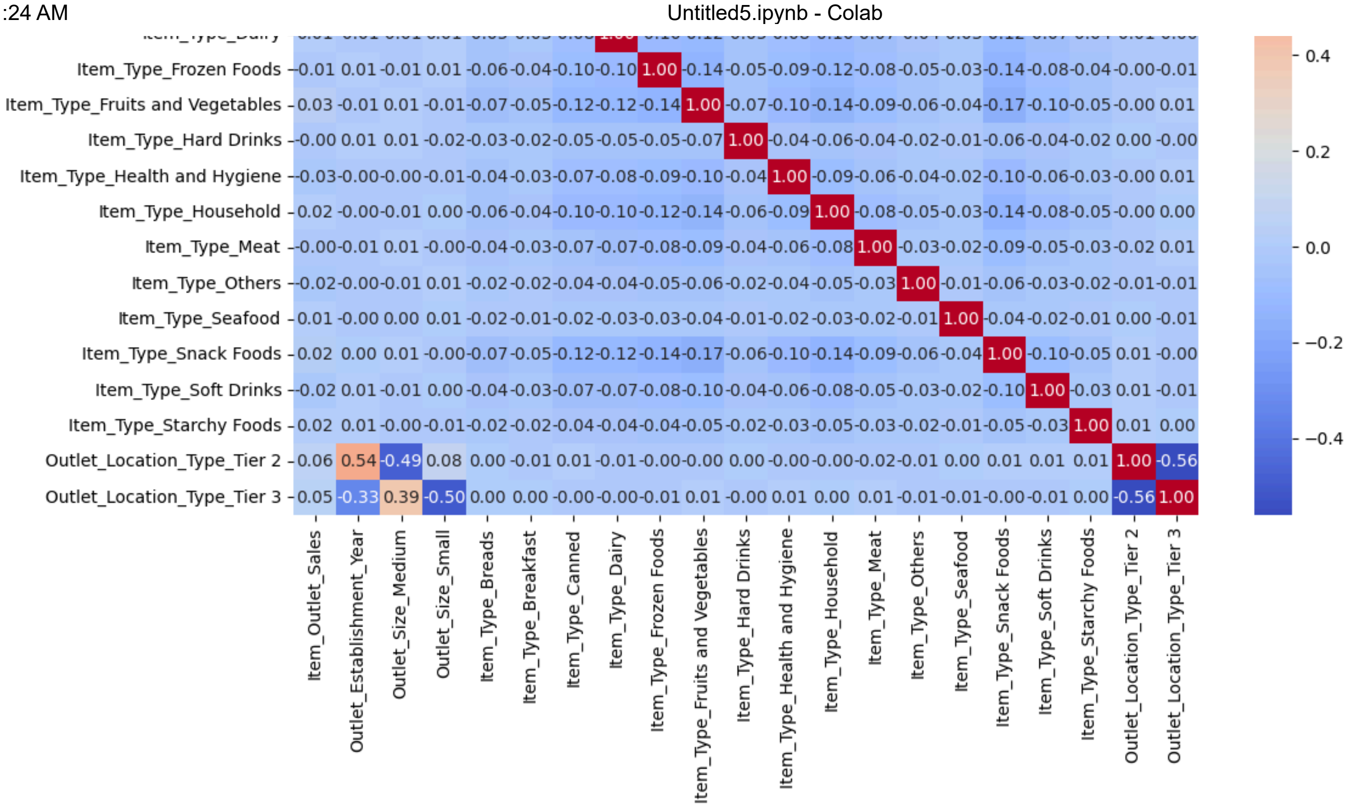
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Step 1: Load the Dataset
url = "https://github.com/akki8087/Big-Mart-Sales/raw/refs/heads/master/Train.csv"
data = pd.read_csv(url)

# Display basic information about the dataset
print("Dataset Overview:")
print(data.head())
print("\nSummary Statistics:")
print(data.describe())
print("\nDataset Info:")
print(data.info())

# Select relevant columns for the analysis
columns_of_interest = ['Item_Outlet_Sales', 'Outlet_Size', 'Item_Type', 'Outlet_Establishment_Year', 'Outlet_Location_Type']
data = data[columns_of_interest]

# Step 2: Compute the Correlation Matrix
# Convert categorical variables to numerical representations
encoded_data = pd.get_dummies(data, drop_first=True)
correlation_matrix = encoded_data.corr()

# Step 3: Visualize the Correlation Matrix
plt.figure(figsize=(12, 8))
sns.heatmap(correlation_matrix, annot=True, fmt=".2f", cmap="coolwarm", cbar=True)
plt.title("Correlation Matrix for Big Mart Sales")
plt.show()

# Step 4: Interpret the Results
# Identify strong correlations (absolute value > 0.5)
strong_correlations = correlation_matrix['Item_Outlet_Sales'][abs(correlation_matrix['Item_Outlet_Sales']) > 0.5]
print("\nStrong Correlations with Sales:")
print(strong_correlations)

# Observations based on the heatmap and correlations
print("\nObservations:")
print("1. Variables with a strong correlation to sales can be further explored.")
print("2. Additional analysis can include examining categorical features like Outlet_Size and Item_Type for patterns.")
```

```
Dataset Overview:
  Item_Identifier  Item_Weight Item_Fat_Content  Item_Visibility  \
0           FDA15         9.30          Low Fat         0.016047   
1           DRC01         5.92          Regular         0.019278   
2           FDN15        17.50          Low Fat         0.016760   
3           FDX07        19.20          Regular         0.000000   
4           NCD19         8.93          Low Fat         0.000000   

               Item_Type  Item_MRP Outlet_Identifier  \
0                  Dairy  249.8092            OUT049   
1            Soft Drinks   48.2692            OUT018   
2                   Meat  141.6180            OUT049   
3  Fruits and Vegetables  182.0950            OUT010   
4              Household   53.8614            OUT013   

   Outlet_Establishment_Year Outlet_Size Outlet_Location_Type  \
0                       1999      Medium               Tier 1   
1                       2009      Medium               Tier 3   
2                       1999      Medium               Tier 1   
3                       1998         NaN               Tier 3   
4                       1987        High               Tier 3   

         Outlet_Type  Item_Outlet_Sales  
0  Supermarket Type1          3735.1380  
1  Supermarket Type2           443.4228  
2  Supermarket Type1          2097.2700  
3      Grocery Store           732.3800  
4  Supermarket Type1           994.7052  

Summary Statistics:
       Item_Weight  Item_Visibility     Item_MRP  Outlet_Establishment_Year  \
count  7060.000000      8523.000000  8523.000000                8523.000000   
mean     12.857645         0.066132   140.992782                1997.831867   
std       4.643456         0.051598    62.275067                   8.371760   
min       4.555000         0.000000    31.290000                1985.000000   
25%       8.773750         0.026989    93.826500                1987.000000   
50%      12.600000         0.053931   143.012800                1999.000000   
75%      16.850000         0.094585   185.643700                2004.000000   
max      21.350000         0.328391   266.888400                2009.000000   

       Item_Outlet_Sales  
count        8523.000000  
mean         2181.288914  
std          1706.499616  
min            33.290000  
25%           834.247400  
50%          1794.331000  
75%          3101.296400  
max         13086.964800  

Dataset Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8523 entries, 0 to 8522
Data columns (total 12 columns):
 #   Column                     Non-Null Count  Dtype  
---  ------                     --------------  -----  
 0   Item_Identifier            8523 non-null   object 
 1   Item_Weight                7060 non-null   float64
 2   Item_Fat_Content           8523 non-null   object 
 3   Item_Visibility            8523 non-null   float64
 4   Item_Type                  8523 non-null   object 
 5   Item_MRP                   8523 non-null   float64
 6   Outlet_Identifier          8523 non-null   object 
 7   Outlet_Establishment_Year  8523 non-null   int64  
 8   Outlet_Size                6113 non-null   object 
 9   Outlet_Location_Type       8523 non-null   object 
 10  Outlet_Type                8523 non-null   object 
 11  Item_Outlet_Sales          8523 non-null   float64
dtypes: float64(4), int64(1), object(7)
memory usage: 799.2+ KB
None
```





```
Strong Correlations with Sales:
Item_Outlet_Sales    1.0
Name: Item_Outlet_Sales, dtype: float64
```