ICA1 - Data Cleaning and Visualization

Deadline: Sep 5, 2025, by 1:50 PM.

Objective: Clean and analyze the Melbourne real estate dataset by handling missing values and visualizing key attributes such as price and distance.

Dataset Description: The dataset includes columns: Rooms, Price, Method, Type, SellerG, Date, Distance, Regionname, Propertycount, Bedroom2, Bathroom, Car, Landsize, BuildingArea, CouncilArea.

Write a Python program to perform the following tasks:

1. **Read the melb_data.csv file** into a DataFrame.
2. **Check for missing values** in the DataFrame using:

```
missing_values = df.isnull().sum()

print(missing_values)
```

3. **Handle missing values** in the specified columns as follows:

   - Use the mode to fill missing values in the Car and CouncilArea columns.
   - Use the mean to fill missing values in the BuildingArea and median to fill missing values YearBuilt columns.

4. **Re-check for missing values** to ensure that all missing values have been addressed.
5. **Create the following visualizations**:

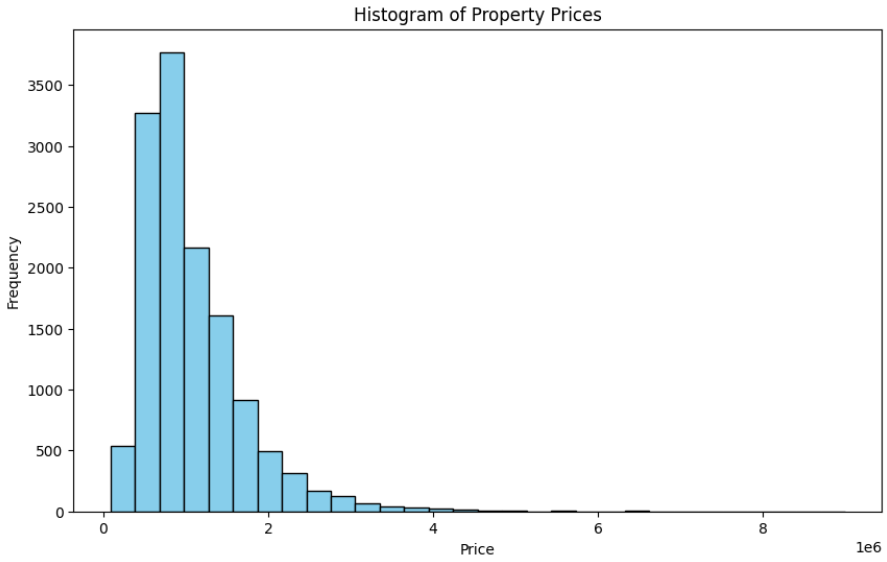
   - A histogram of the Price column.
   - A scatter plot of Price versus Distance.

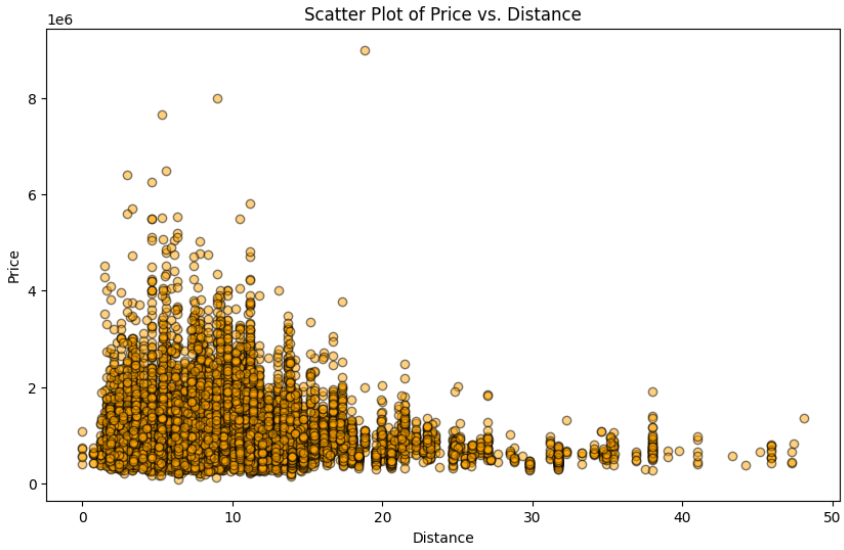
Expected Output:

```
Missing values before handling:
Suburb              0
Address             0
Rooms               0
Type                0
Price               0
Method              0
SellerG             0
Date                0
Distance            0
Postcode            0
Bedroom2            0
Bathroom            0
Car                62
Landsize            0
BuildingArea     6450
```

```
YearBuilt        5375
CouncilArea      1369
Lattitude           0
Longtitude          0
Regionname          0
Propertycount       0
dtype: int64


Missing values after handling:
Suburb           0
Address          0
Rooms            0
Type             0
Price            0
Method           0
SellerG          0
Date             0
Distance         0
Postcode         0
Bedroom2         0
Bathroom         0
Car              0
Landsize         0
BuildingArea     0
YearBuilt        0
CouncilArea      0
Lattitude        0
Longtitude       0
Regionname       0
Propertycount    0
   dtype: int64
```





Instructions to save your Colab notebook and submit it via Blackboard

1. **Run and Save Your Notebook**:
   - Click on the notebook name at the top and rename it to ICA1-YourGlobalID.ipynb.
   - Ensure that all cells in your notebook have been executed so that the expected outputs are visible.
   - Go to File > Save to save your progress.
2. **Download the Notebook**:
   - Go to File > Download > .ipynb to save the notebook to your computer.
3. **Submit via Blackboard**:
   - Log in to Blackboard, find the assignment link, upload the ICA1-YourGlobalID.ipynb file, and click Submit.