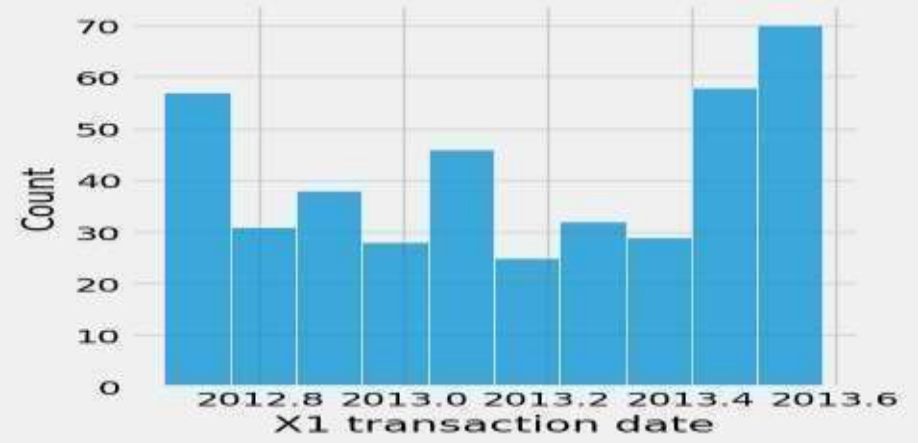
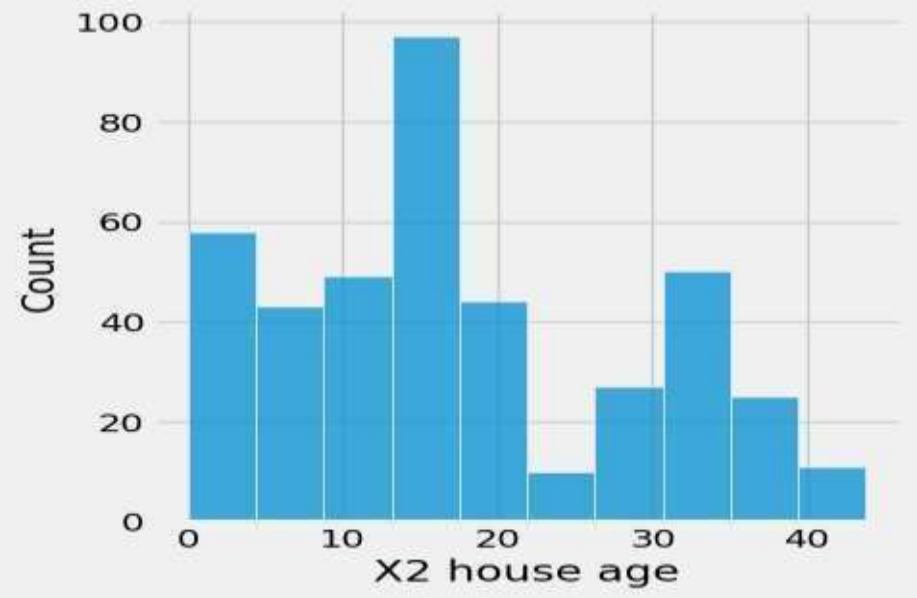
Data Collection and Preprocessing Phase

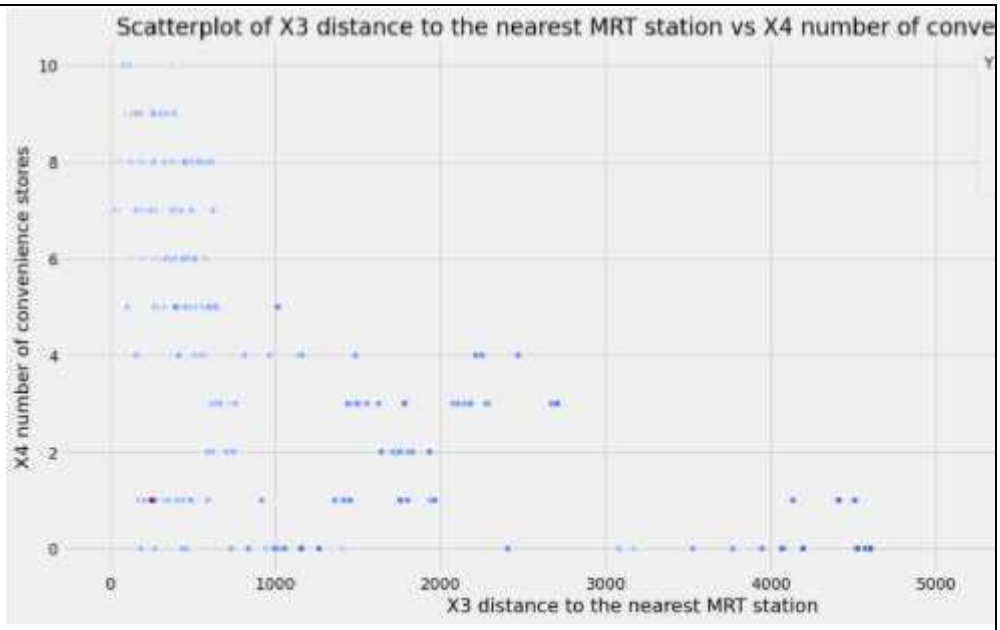
Date	8 July 2024
Team ID	739996
Project Title	Identification Of Methodology Used In Real Estate Property Valuation
Maximum Marks	6 Marks

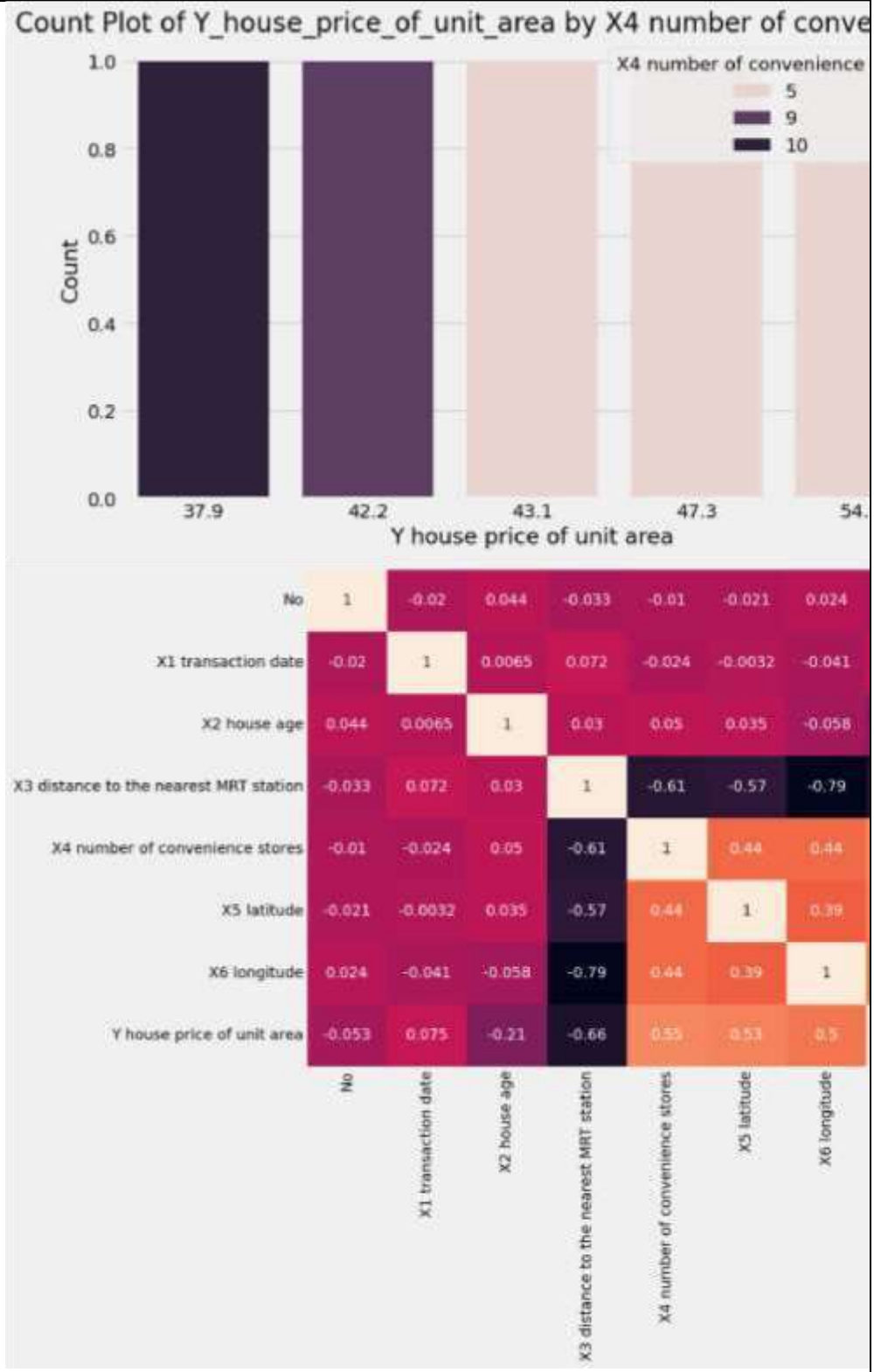
Data Exploration and Preprocessing Template

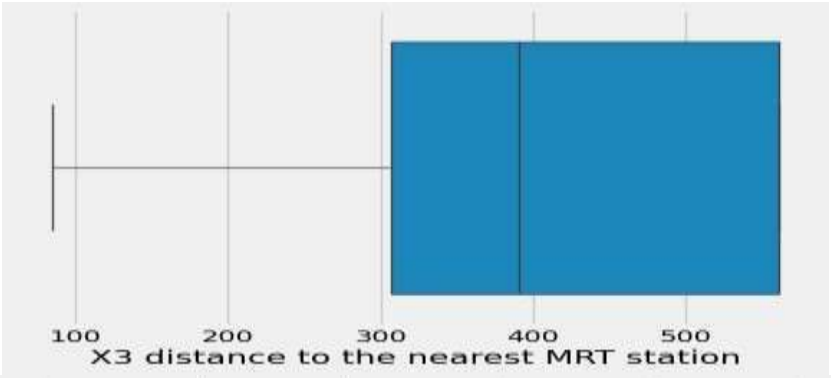
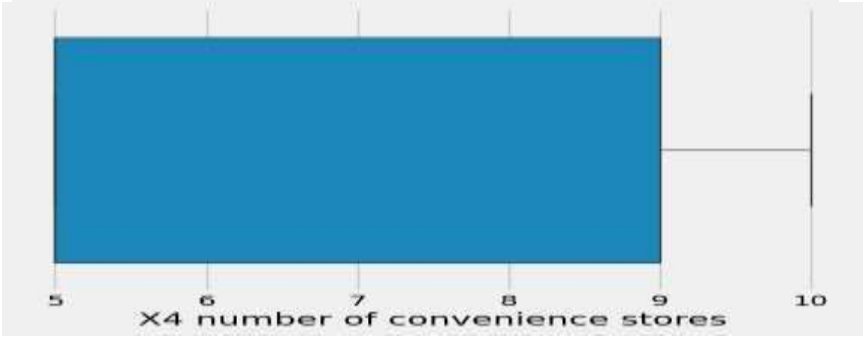
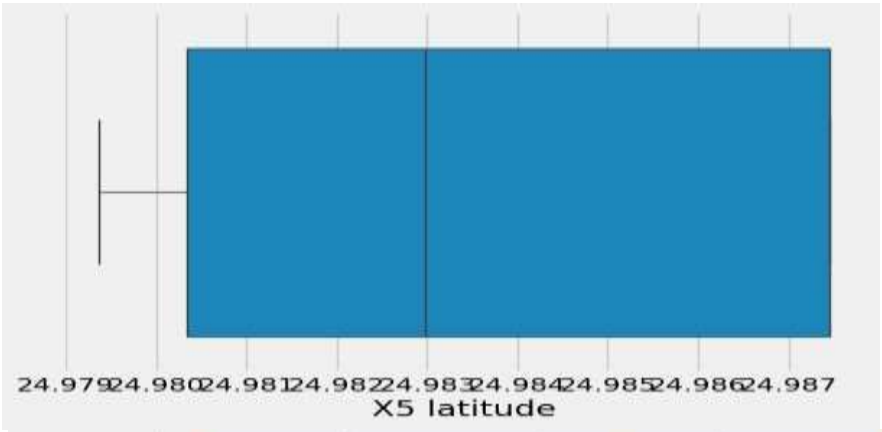
Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	Description
Data Overview	Dimension: 331rowsx 8columns Descriptive statistics: 

Univariate Analysis	
	

Bivariate Analysis	

Multivariate Analysis	

Handled Outliers and Anomalies	    

Data Preprocessing Code Screenshots	
Loading Data	
Finding &Handling Missing Data	
Data Transformation	-
Feature Engineering	Attached the code in final submission

Save Processed Data	

```
import pickle
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestRegressor
rf_model = RandomForestRegressor()
scaler = StandardScaler()
with open('price.pkl', 'wb') as f:
    pickle.dump(rf_model, f)
with open('scale.pkl', 'wb') as f:
    pickle.dump(scaler, f)
```

```
from google.colab import files
files.download('price.pkl')
```

```
from google.colab import files
files.download('scale.pkl')
```

```
from google.colab import files
files.download('/content/drive/MyDrive/ dataset/real estate valuation data set.csv')
```