TP03: Full Data Analysis Project using Python & Power BI

kiram mohammed ali

1 Introduction

This report presents the complete workflow of TP03, which includes data loading, cleaning, exploratory data analysis (EDA), machine learning modeling, and dashboard creation using Power BI. Each step is supported with screenshots and code overviews extracted from the Jupyter Notebook.

2 Data Loading

The dataset is loaded using Python as follows:

```
import pandas as pd
df = pd.read_csv('your_data.csv')
```

2.1 Overview

The data is successfully imported and ready for cleaning.

2.2 Screenshot

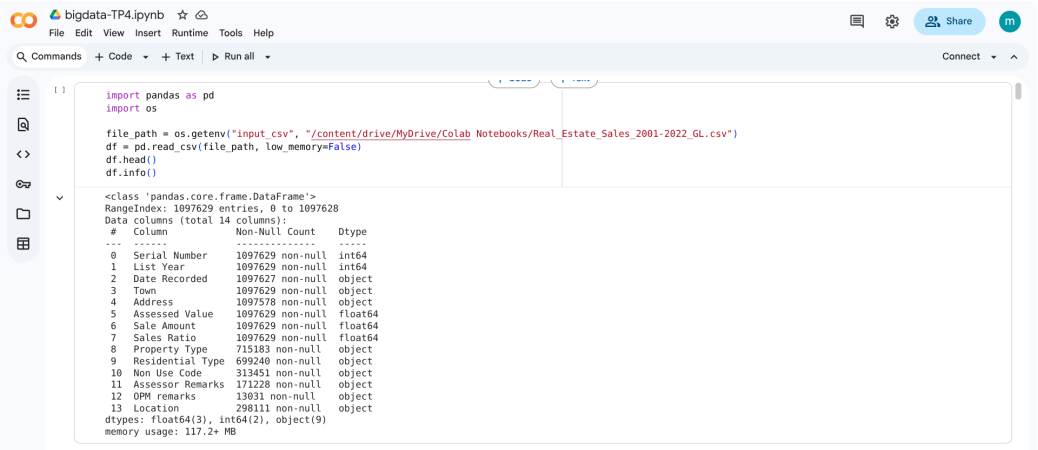


Figure 1: Data Import Preview (Replace with your screenshot)

3 Data Cleaning

The following operations were performed:

- Removing missing values
- Removing duplicates
- Converting date fields
- Handling outliers

3.1 Code Overview

```
df.dropna(inplace=True)
df.drop_duplicates(inplace=True)
df['Date'] = pd.to_datetime(df['Date'])
```

3.2 Screenshot

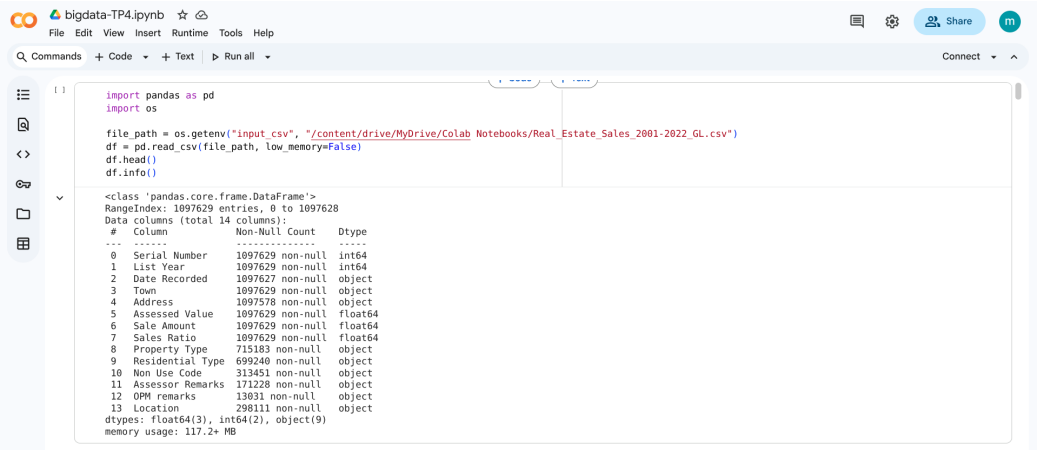


Figure 2: Cleaning Process Output (Replace with your screenshot)

4 Exploratory Data Analysis (EDA)

4.1 KPI Cards

- Total Sale Amount: 324M
- Number of Towns: 170
- Average Sales Ratio: 9.09

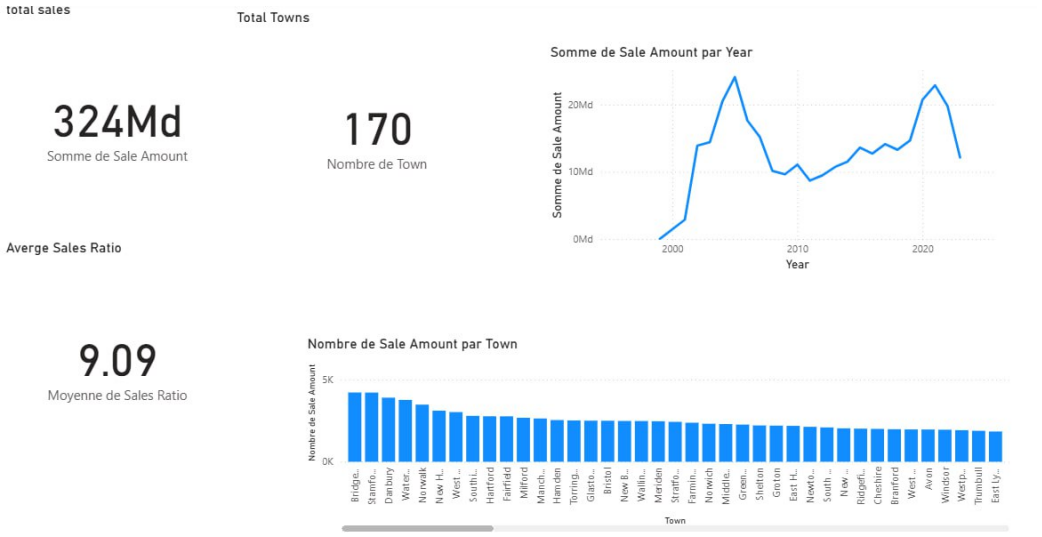


Figure 3: Power BI KPI Overview

The trend shows a rapid increase after 2000, a peak around 2018, and a noticeable decline in the most recent year.

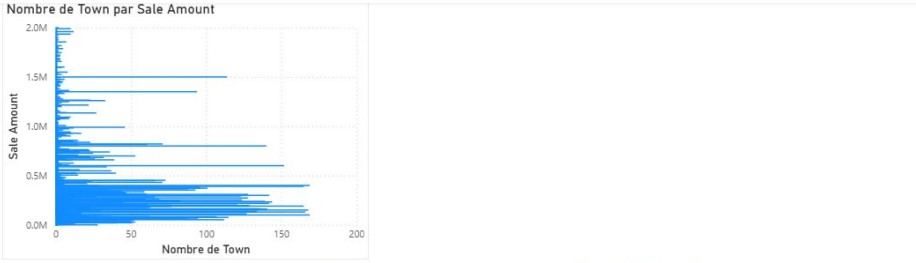
Bridgeport, Stamford, Danbury, Waterbury, and Norwalk appear as the top-performing towns.

Scatter Plot: Assessed Value vs Sale Amount

Figure 4: Assessed Value vs Sale Amount

Overview: A general positive correlation is visible, with significant variance and some extreme outliers.

Town-Level Summary



Town	Moyenne de Sale Amount	Moyenne de Assessed Value
Unknown	282450.00	66540.0(
Andover	225542.07	137619.8:
Ansonia	207046.09	128131.4:
Ashford	181353.74	111874.3:
Avon	422465.91	267199.6(
Barkhamsted	227958.14	139631.9(
Beacon Falls	251288.98	129858.0(
Berlin	267019.08	174953.4:
Bethany	342524.73	202480.5(
Bethel	336982.22	207646.8-
Bethlehem	304566.00	181634.6(
Bloomfield	247266.68	141569.3(
Bolton	264869.64	166569.0(
Bozrah	209738.73	127793.9-
Branford	294378.01	184049.0(
Total	**302544.68**	**188529.0(**

Figure 5: Summary Table of Town Metrics

Overview: Overall average Sale Amount = 302,544. Overall average Assessed Value = 188,529.

5 Machine Learning: Random Forest

5.1 Code Overview

```
from sklearn.ensemble import RandomForestRegressor
model = RandomForestRegressor()
model.fit(X_train, y_train)
pred = model.predict(X_test)
```

5.2 Error Distribution

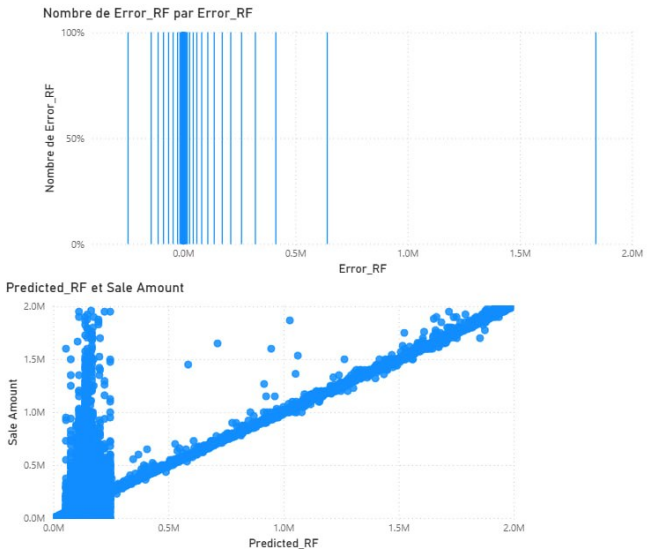


Figure 6: Random Forest Error Distribution

Overview: The model shows consistent errors, especially for high-value properties.

The RF model performs well at mid-range values but clearly underestimates luxury properties.

6 Power BI Dashboard Overview

All Power BI visuals are integrated in the previous sections. This dashboard enables interactive exploration of real estate trends.

7 Conclusion

This TP demonstrates the full analytical workflow: data loading, cleaning, exploration, modeling, and dashboard creation. Power BI and Python together provide a strong analytical stack for business insight generation.