REAL ESTATE ANALYSIS AND PREDICTION

Using Python for Data Insights

Katherine Ponce | feb.2024

Introduction

- **Objective:** Analyze real estate pricing trends and predict resale prices using Python.
- **Key Steps:** Data cleaning, exploratory analysis, correlation, and predictive modeling.
- **Outcome:** Insights into the relationship between floor area and resale price.

Data Summary

Total Records: 1000

Key Statistics:

- Average resale price: **$477,080.11**
- Minimum resale price: **$151,040.77**
- Maximum resale price: **$799,838.52**
- Average floor area: **118.34 sqm**

Observations:

Price Trends by Location:

- Rotterdam: **$464,913.87**
- Amsterdam: **$466,964.69**
- Utrecht: **$471,604.78**
- Eindhoven: **$484,483.66**
- The Hague: **$496,454.02**

Premium locations have significantly higher resale prices.

DATA PREPARATION

- **Dataset:** `real_estate_data.csv`
- **Columns:** Floor Area (sqm), Resale Price, Property Type, Town, Lease Year.

- **Data Cleaning Steps:**
  - **Handled missing values** using median imputation.
  - **Checked for duplicate rows** and removed if necessary.
  - **One-hot encoded categorical variables** (Town, Property Type).

Exploratory Data Analysis

Descriptive Statistics:

- Summary of resale prices (Mean, Min, Max, Std Dev).
- Identified potential outliers using boxplots.

Visualizations:

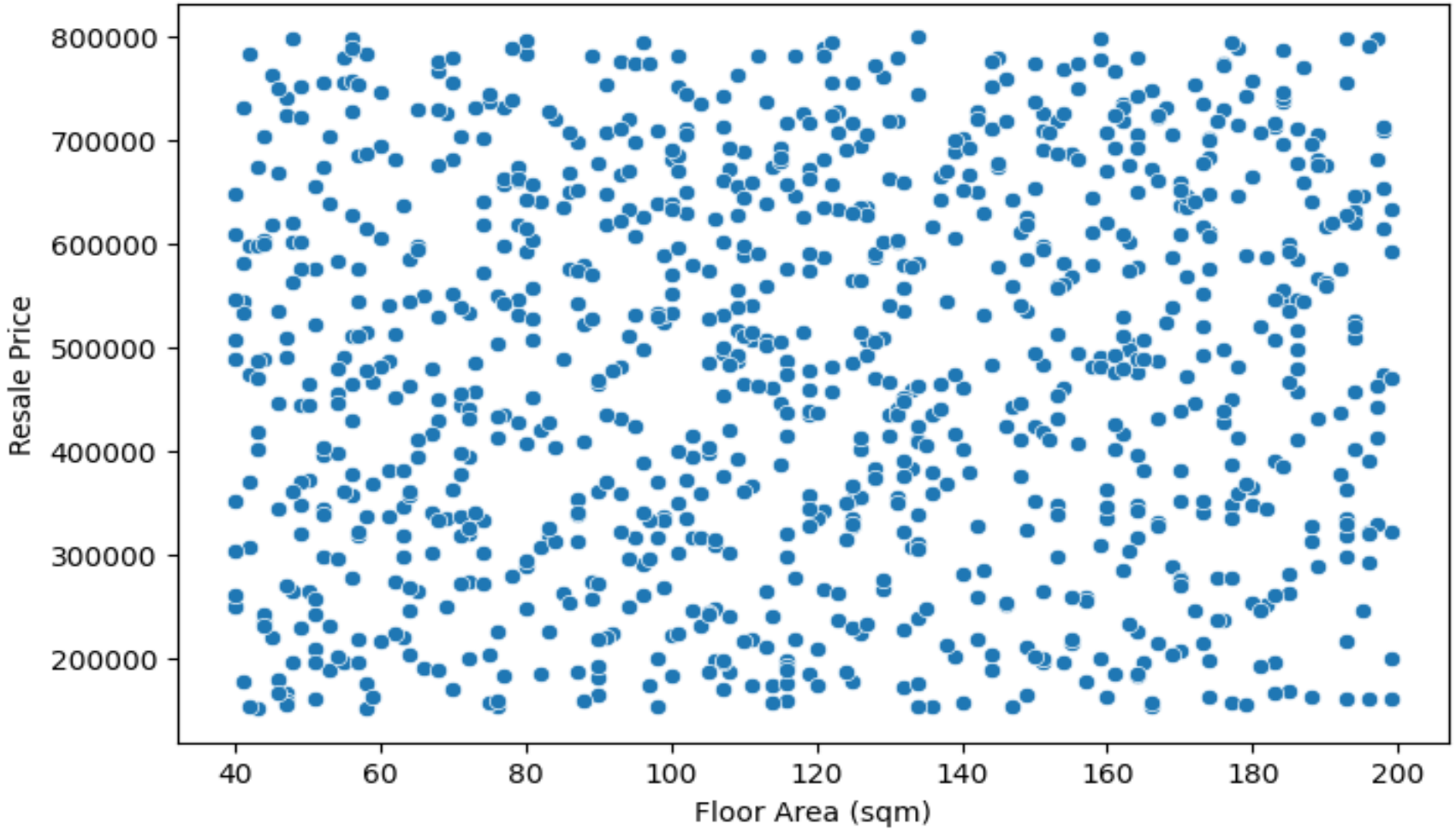
- Scatter plot of Floor_Area_Sqm vs. Resale_Price.
- Histogram: Distribution of resale prices.

Descriptive Statistics

	Lease_Commence_Year	Resale_Price	Floor_Area_Sqm	Mortgage_Rate
count	1000.000000	1000.000000	1000.000000	1000.000000
mean	1985.640000	477080.113300	118.345000	2.983130
std	20.923149	185964.592557	45.569821	0.836509
min	1950.000000	151040.770000	40.000000	1.500000
25%	1968.000000	322762.407500	79.750000	2.290000
50%	1985.000000	480647.165000	118.000000	2.980000
75%	2004.000000	641345.302500	159.000000	3.662500
max	2022.000000	799838.520000	199.000000	4.490000

Scatter plot: Floor_Area_Sqm vs. Resale_Price

This helps visualize how resale price changes with floor area



Correlation Analysis

Pearson Correlation Coefficient:

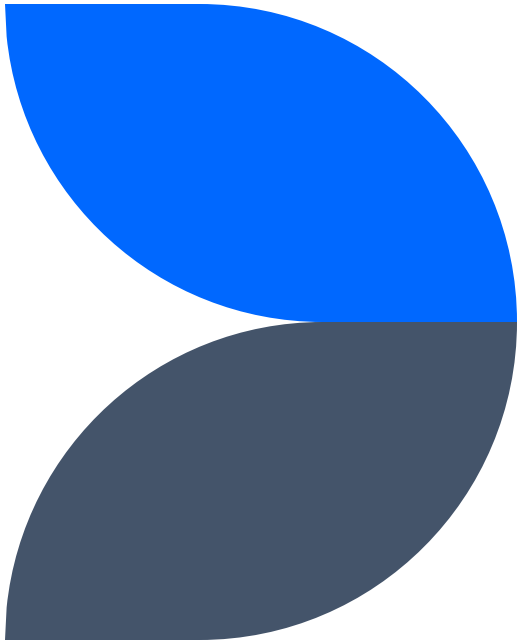
- Measures the relationship between floor area and resale price.
- **Result:** A correlation coefficient of `0.0706` suggests a weak positive correlation.

Interpretation:

- Larger properties **tend to have higher prices**, but other factors also play a role.
- Location, property type, and economic factors should be explored further.

Linear Regression

Model & Interpretation



- **Model:** Predicting resale price based on floor area.
- **Equation:**

```
Resale_Price = 443006.57
+ 287.92 *
Floor_Area_Sqm
```

Interpretation:

- **Each additional sqm increases the resale price by ~$288.**
- The model provides a **baseline** prediction for property valuation.

Regression Model

Performance & Insights

Regression Line Visualization:

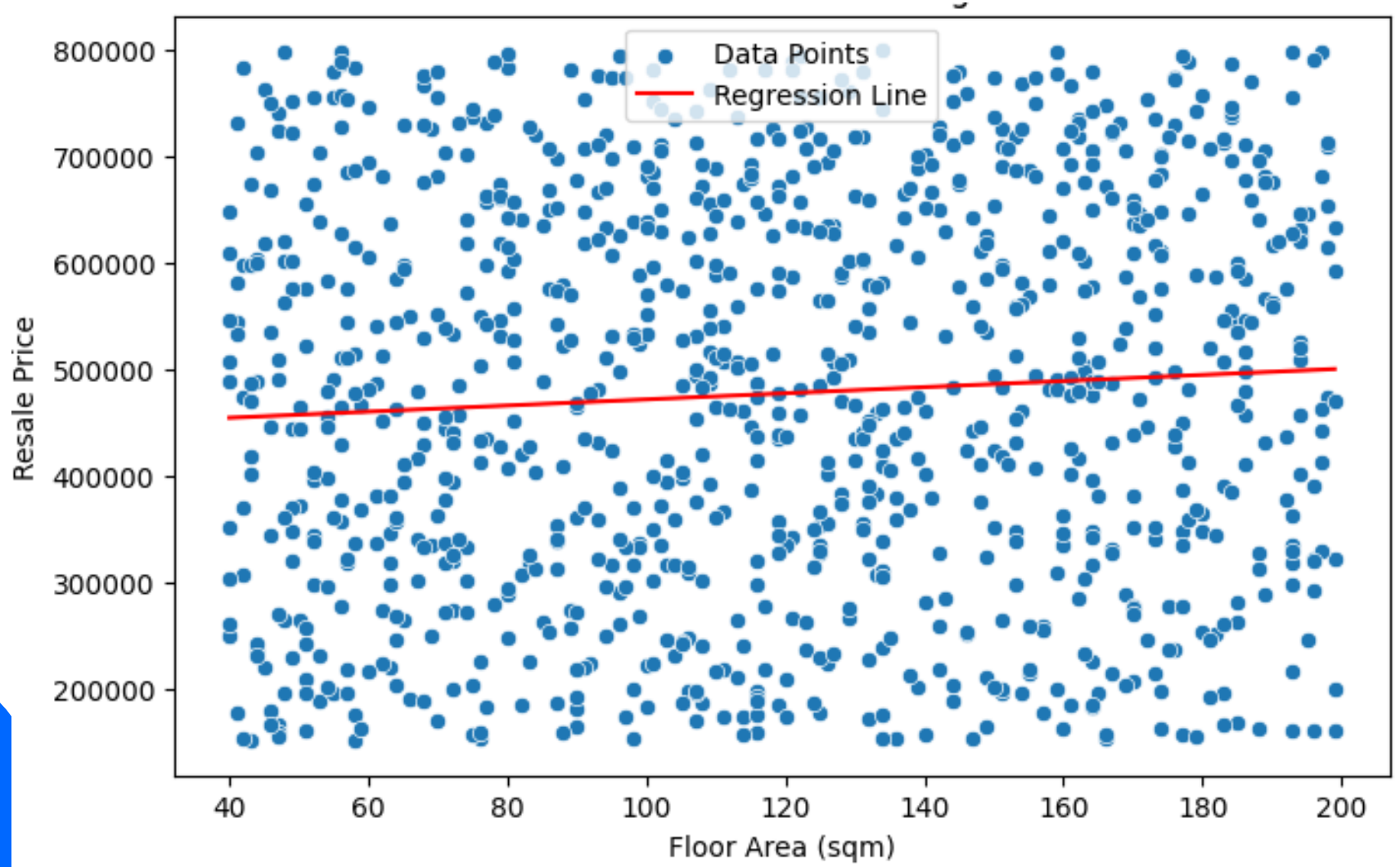
- **Blue dots:** Actual resale prices.
- **Red line:** Predicted prices based on regression model.

Observations:

- If points closely follow the line, the model is **accurate**.
- If widely scattered, other factors **affect price predictions**.

Regression Line Interpretation

This shows the general trend the model has learned: larger floor areas generally have higher prices.



Business Use Cases:

- **Assisting buyers** in making data-driven investment decisions.
- **Helping real estate agencies** predict pricing trends.

Business Impact & Key Takeaways

Key Takeaways:

- **Floor area influences price but is not the only determinant.**
- Other factors like location, mortgage rates, and demand impact pricing.

Next Steps & Recommendations

Future Work:	Call to Action:
❑ **Incorporate additional variables** like mortgage rates and neighborhood scores. ❑ **Use advanced models** (Random Forest, XGBoost) to improve accuracy.	❑ Expand analysis to **other real estate markets.** ❑ Use insights to **optimize property pricing strategies.**

Conclusion

- Python-powered analysis provides valuable insights into real estate pricing.
- Data-driven decision-making is crucial for accurate property valuation.
- Further refinements can improve prediction accuracy and bussines impact.

Thank you

Any Questions?

Katherine **P**once