

Smart Rent BKK

DADS5001 DATA ANALYTICS AND DATA SCIENCE TOOLS AND PROGRAMMING

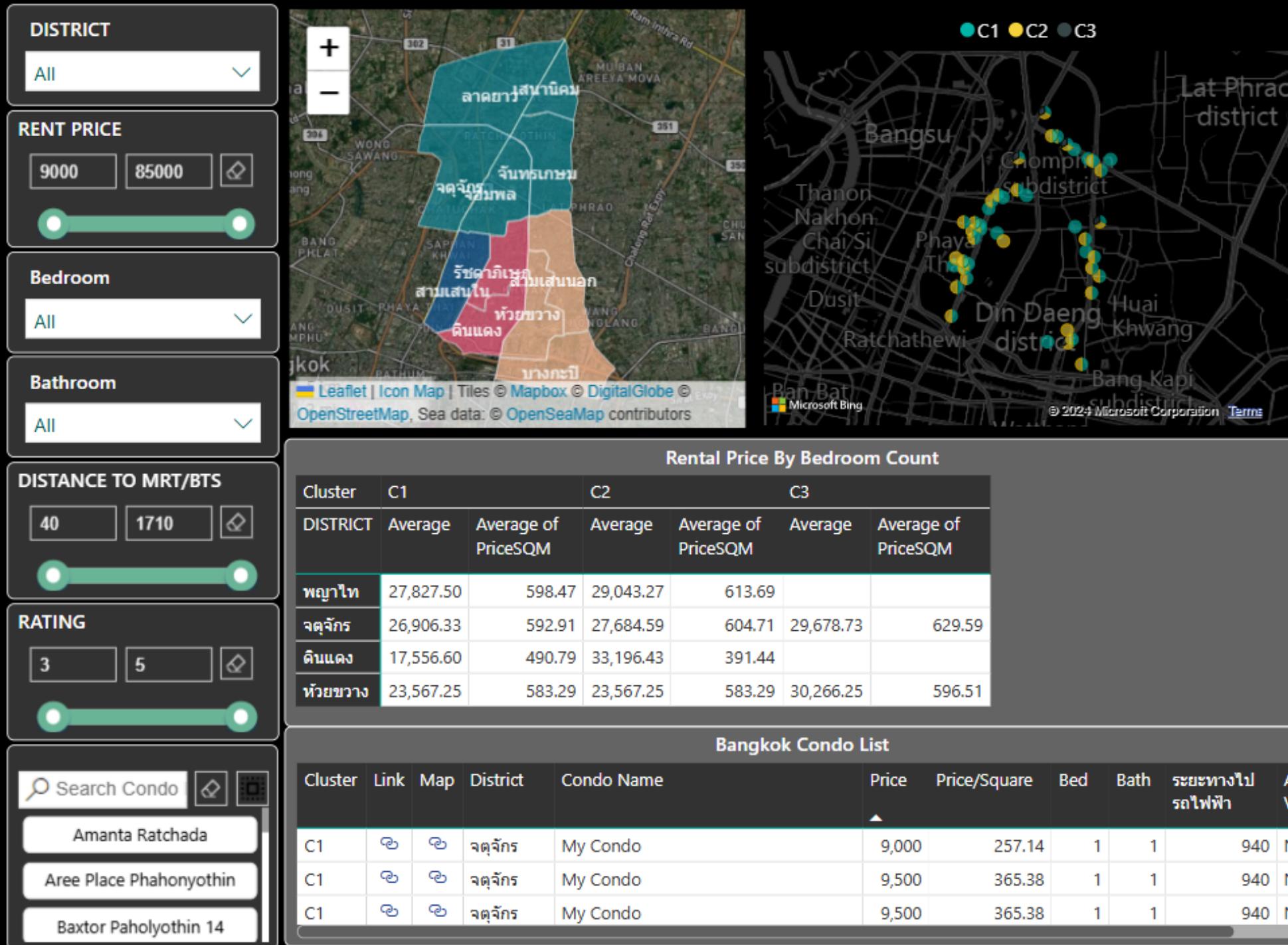
TEAM MEMBERS

Mr. Kitisak Namtae	ID: 6710422007
Mr. Kantapong Charusiri	ID: 6710422015
Miss Pitchaporn Nimdum	ID: 6710422016
Miss Pornchanok Tuntikulwattanakit	ID: 6710422023
Mr. Suparerk Jankam	ID: 6710422028

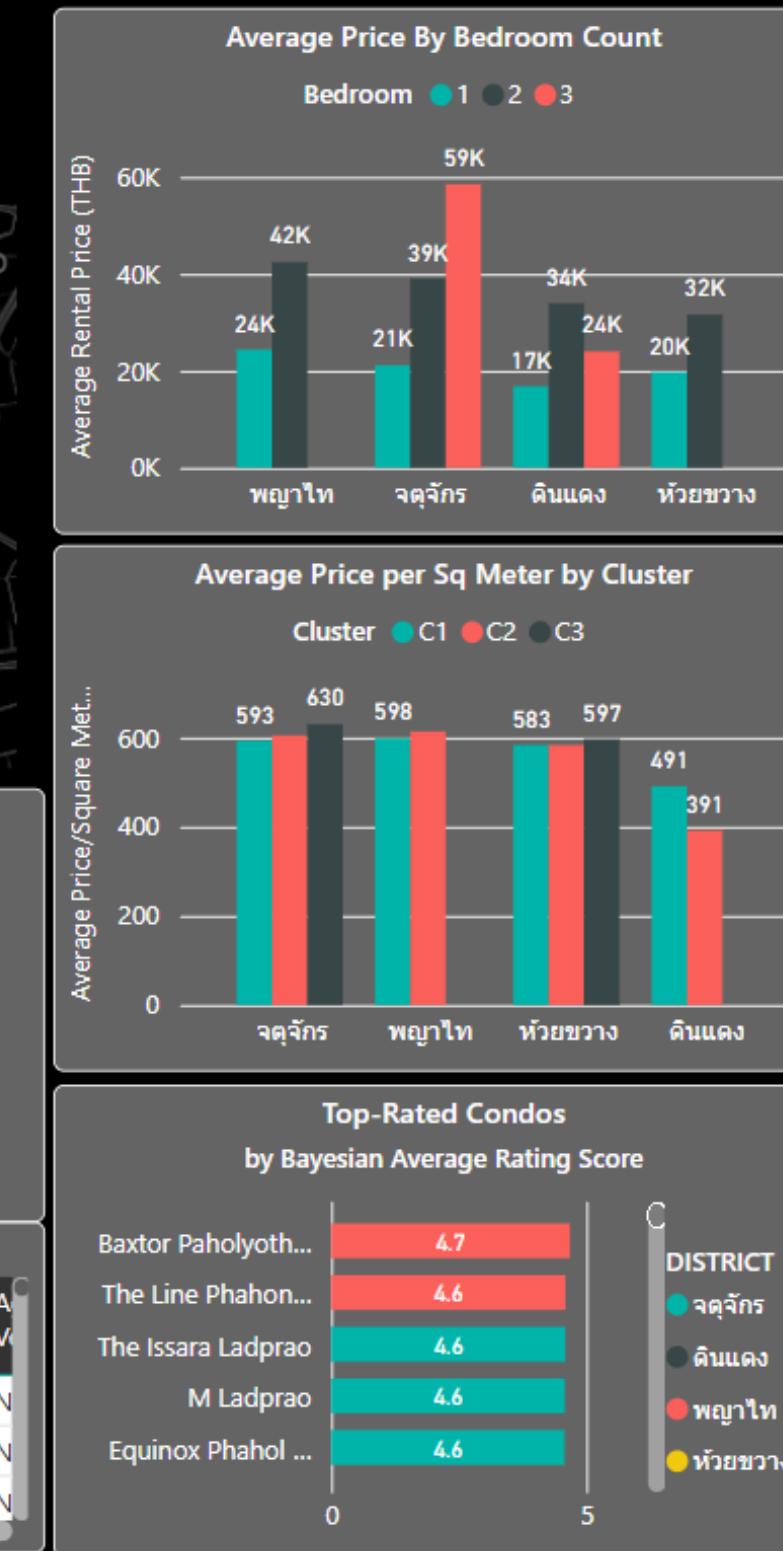
Objective

This project enhances a Bangkok condo rental search engine for renters in **Chatuchak, Din Daeng, Phaya Thai, and Huai Khwang** by integrating intelligent price prediction and personalized filtering.

Bangkok Condo Rental Data: Exploratory Data Analysis (EDA)



Previous Project from DADS5002



PART 1: DATA PREPARATION

Raw Data : Web scraping from DDproperty combined with Google Maps data.

rent_cd_id	rent_cd_address	rent_cd_price	rent_cd_bed	rent_cd_bath	rent_cd_floorarea
11556700	ซอยลาดพร้าว 27 ถนนลาดพร้าว, จันทร์เกษม, จตุจักร, กรุงเทพ	9,000	1		1 35 ตร.ม.
10979613	ซอยลาดพร้าว 27 ถนนลาดพร้าว, จันทร์เกษม, จตุจักร, กรุงเทพ	9,500			26 ตร.ม.
11479296	22 ซอยพหลโยธิน 18/1 ถนนพหลโยธิน, จอมพล, จตุจักร, กรุงเทพ	9,500	1		1 32 ตร.ม.
11539315	ซอยลาดพร้าว 27 ถนนลาดพร้าว, จันทร์เกษม, จตุจักร, กรุงเทพ	9,500	1		1 26 ตร.ม.
11540821	วิภาวดีรังสิต, จอมพล, จตุจักร, กรุงเทพ	9,500	1		1 32 ตร.ม.
11540880	ซอยลาดพร้าว 27 ถนนลาดพร้าว, จันทร์เกษม, จตุจักร, กรุงเทพ	9,500	1		1 26 ตร.ม.
11017517	วิภาวดีรังสิต, จอมพล, จตุจักร, กรุงเทพ	10,000	1		1 32 ตร.ม.
11541510	ซอยรัชดาภิเษก 20 ถนนรัชดาภิเษก, สามเสนนอ, ห้วยขวาง, กรุงเทพ	10,000	1		1 31 ตร.ม.
11567495	ถนนประชาอุทิศ, ห้วยขวาง, ห้วยขวาง, กรุงเทพ	10,000	1		1 23 ตร.ม.
7755351	ซอยวิภาวดีรังสิต 3 ถนนวิภาวดีรังสิต, จอมพล, จตุจักร, กรุงเทพ	11,000			25 ตร.ม.
7969117	ซอยวิภาวดีรังสิต 3 ถนนวิภาวดีรังสิต, จอมพล, จตุจักร, กรุงเทพ	11,000			25 ตร.ม.
10808406	1038 พหลโยธิน 18, จอมพล, จตุจักร, กรุงเทพ	11,000	1		1 31 ตร.ม.
10980007	ซอยวิภาวดีรังสิต 3 ถนนวิภาวดีรังสิต, จอมพล, จตุจักร, กรุงเทพ	11,000			25 ตร.ม.

PART 1: DATA PREPARATION

Data Cleansing: Apply KNN to impute missing values for 4 features: rent_cd_price, rent_cd_bed, rent_cd_bath, and rent_cd_floorarea.

rent_cd_id	rent_cd_address	rent_cd_price	rent_cd_bed	rent_cd_bath	rent_cd_floorarea
11556700	ซอยลาดพร้าว 27 ถนนลาดพร้าว, จันทร์เกย์ม, จตุจักร, กรุงเทพ	9000	1	1	35
11539315	ซอยลาดพร้าว 27 ถนนลาดพร้าว, จันทร์เกย์ม, จตุจักร, กรุงเทพ	9500	1	1	26
10979613	ซอยลาดพร้าว 27 ถนนลาดพร้าว, จันทร์เกย์ม, จตุจักร, กรุงเทพ	9500	1	1	26
11540880	ซอยลาดพร้าว 27 ถนนลาดพร้าว, จันทร์เกย์ม, จตุจักร, กรุงเทพ	9500	1	1	26
11540821	วิภาวดีรังสิต, จอมพล, จตุจักร, กรุงเทพ	9500	1	1	32
11479296	22 ซอยพหลโยธิน 18/1 ถนนพหลโยธิน, จอมพล, จตุจักร, กรุงเทพ	9500	1	1	32
11567495	ถนนประชาอุทิศ, ห้วยขวาง, ห้วยขวาง, กรุงเทพ	10000	1	1	23
11541510	ซอยรัชดาภิเษก 20 ถนนรัชดาภิเษก, สามเสนนอก, ห้วยขวาง, กรุงเทพ	10000	1	1	31
11017517	วิภาวดีรังสิต, จอมพล, จตุจักร, กรุงเทพ	10000	1	1	32
11565979	ถนนประชาอุทิศ, ห้วยขวาง, ห้วยขวาง, กรุงเทพ	11000	1	1	22
11564843	ถนนประชาอุทิศ, ห้วยขวาง, ห้วยขวาง, กรุงเทพ	11000	1	1	24
10980007	ซอยวิภาวดีรังสิต 3 ถนนวิภาวดีรังสิต, จอมพล, จตุจักร, กรุงเทพ	11000	1	1	25
11540893	ซอยวิภาวดีรังสิต 3 ถนนวิภาวดีรังสิต, จอมพล, จตุจักร, กรุงเทพ	11000	1	1	25
11543587	ซอยวิภาวดีรังสิต 3 ถนนวิภาวดีรังสิต, จอมพล, จตุจักร, กรุงเทพ	11000	1	1	25
11547112	ซอยวิภาวดีรังสิต 3 ถนนวิภาวดีรังสิต, จอมพล, จตุจักร, กรุงเทพ	11000	1	1	25
11547848	ซอยวิภาวดีรังสิต 3 ถนนวิภาวดีรังสิต, จอมพล, จตุจักร, กรุงเทพ	11000	1	1	25

PART 1: DATA PREPARATION

Create New Feature: price_per_sqm was created, which proved to be a highly important feature for Model.

$$\text{price_per_sqm} = \frac{\text{rent_cd_price}}{\text{rent_cd_floorarea}}$$

	rent_cd_price	rent_cd_bed	rent_cd_bath	rent_cd_floorarea	rent_cd_features_time	star	starnum_list	near_rail_meter	price_per_sqm
0	60000.0	3	3	102	1.0	4.6	94	50.0	588.235294
1	70000.0	3	3	102	1.0	4.6	94	50.0	686.274510
2	55000.0	3	3	102	1.0	4.6	94	50.0	539.215686
3	32000.0	2	2	103	8.0	4.3	38	680.0	310.679612
5	35000.0	2	2	112	12.0	4.0	5	1010.0	312.500000

PART 1: DATA PREPARATION

Feature Reduction: Apply PCA to transformed features into 5 principal components

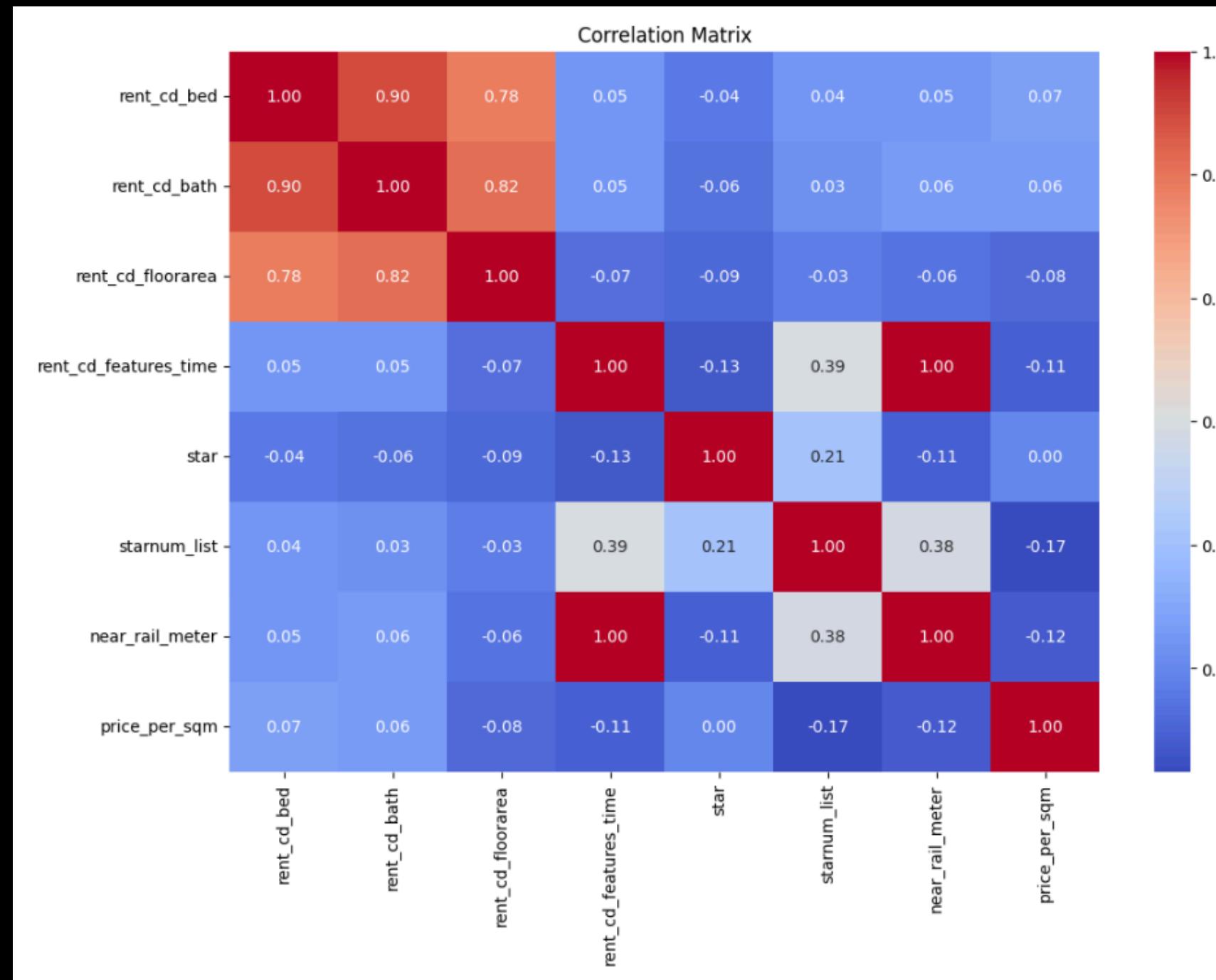
	rent_cd_price	rent_cd_bed	rent_cd_bath	rent_cd_floorarea	rent_cd_features_time	star	starnum_list	near_rail_meter
PC1	0.087599	0.576289	0.584012	0.551417	0.069790	-0.058788	0.038318	0.072746
PC2	-0.040369	-0.036425	-0.035779	-0.116366	0.643861	-0.050498	0.391540	0.641725
PC3	0.286531	0.035849	0.008501	-0.011059	-0.107455	0.818631	0.474327	-0.098644
PC4	0.942042	-0.064406	-0.073511	-0.028410	0.090138	-0.162646	-0.243368	0.091606
PC5	0.141026	-0.055699	-0.044296	0.036218	-0.233863	-0.541353	0.748935	-0.255103

Top 3 Contributing Features per PC (explain ~90% of the variance) :

- PC1: rent_cd_bath, rent_cd_bed, floorarea
- PC2: near_rail_meter, features_time, starnum_list
- PC3: Strongest influence from star rating

PART 1: DATA PREPARATION

Correlation Analysis: Calculates and visualizes the correlation matrix of the features using a heatmap.



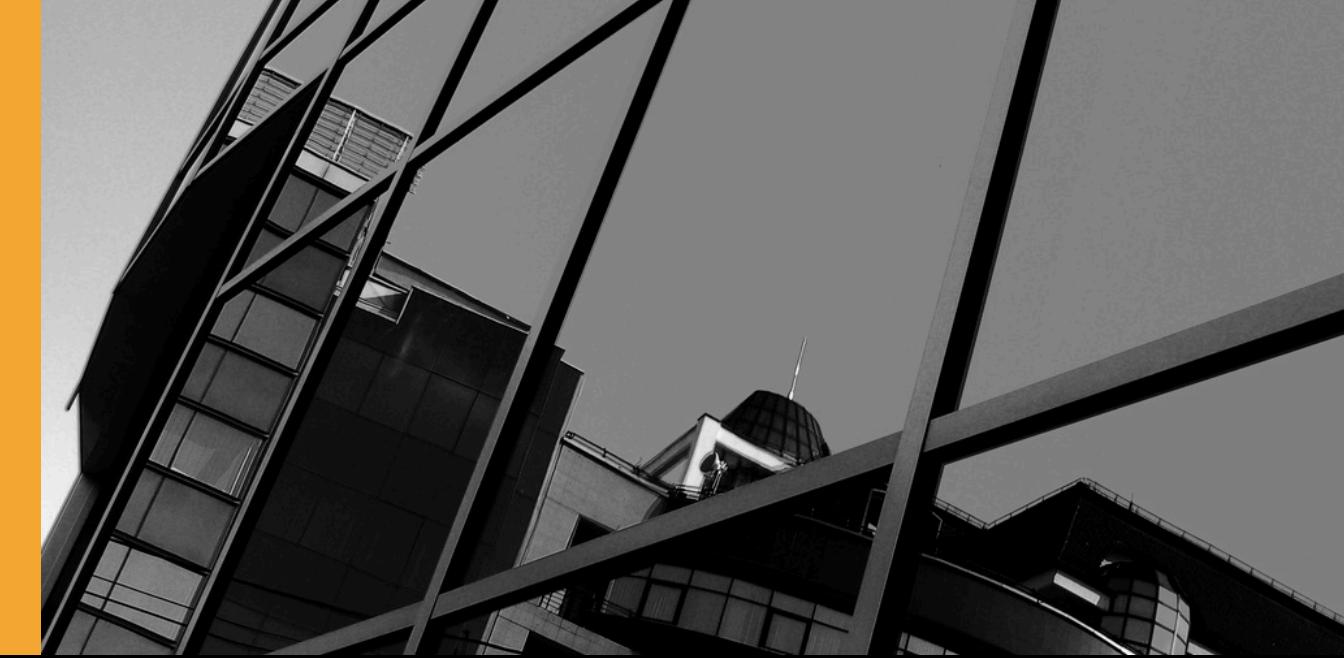
The features dropped due to high correlation (>0.80) are:

- 'rent_cd_floorarea'
- 'near_rail_time'.

PART 2

MACHINE LEARNING MODEL

: CLASSIFICATION



✓ GridSearchCV

Uses GridSearchCV to find the best hyperparameters for max_depth, min_samples_leaf, and criterion

```
Fitting 4 folds for each of 48 candidates, totalling 192 fits
Best Parameters: {'criterion': 'entropy', 'max_depth': 10, 'min_samples_leaf': 1}

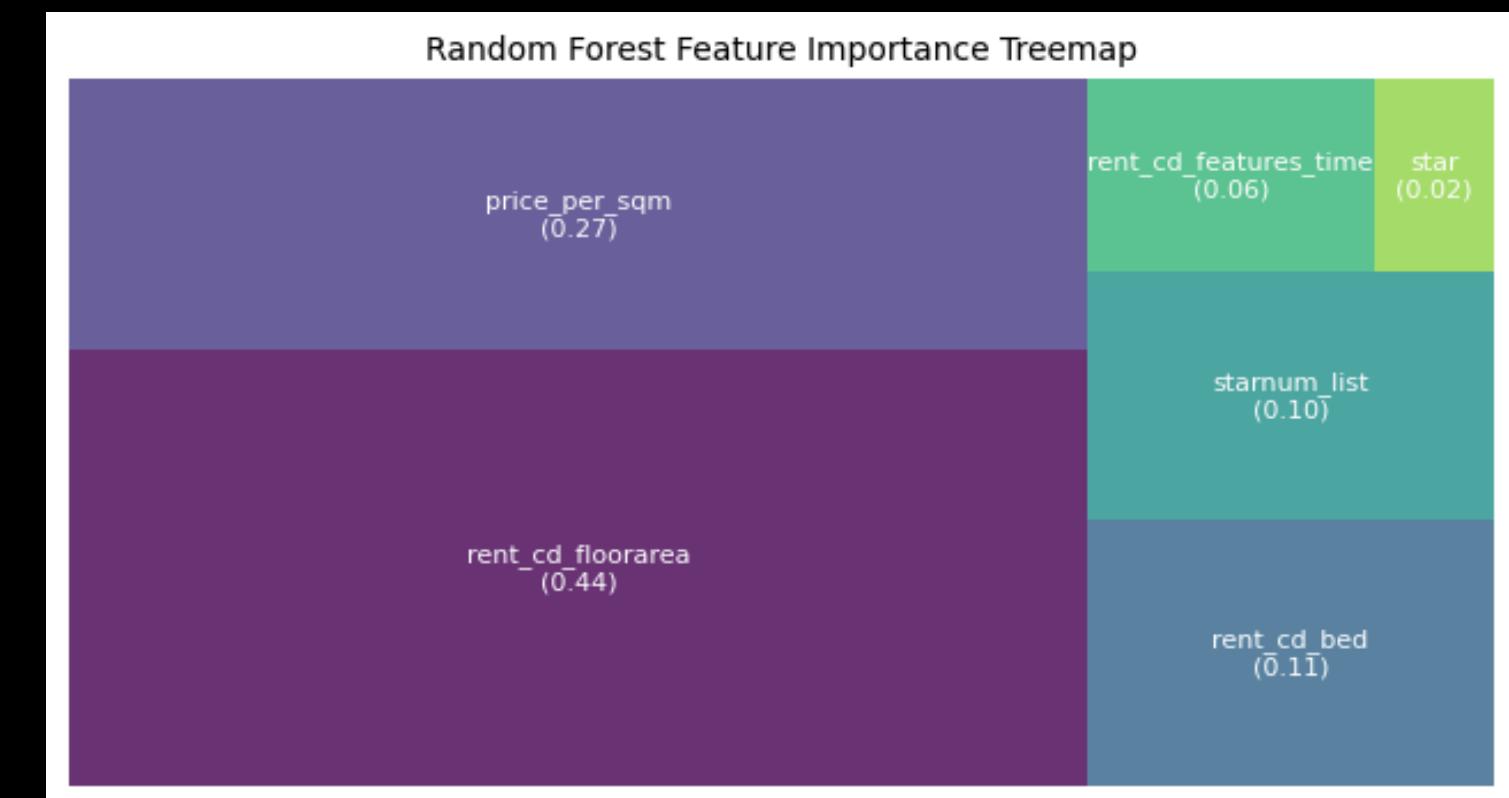
Accuracy with Best Model: 0.9883
```

✓ Standardization

scale numerical features so that they have a mean of 0 and a standard deviation of 1

✓ Feature Importance

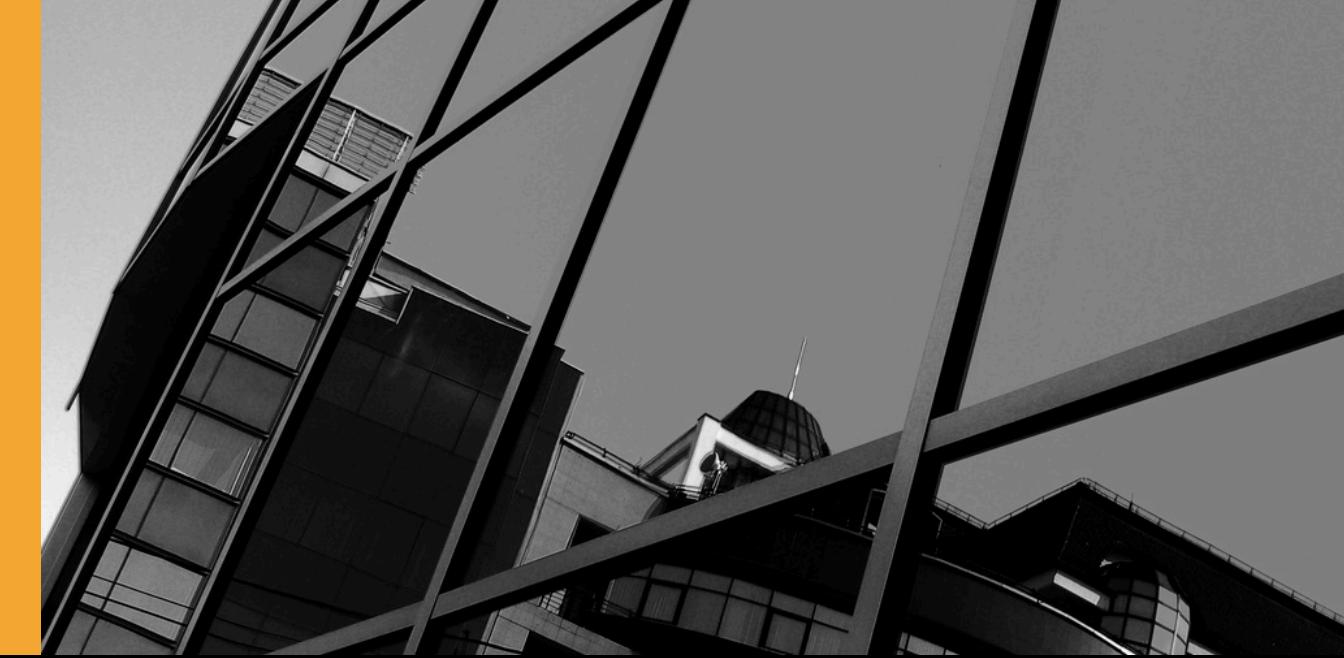
indicates the contribution of each feature to the model's predictions. The 'star' feature was dropped due to its lower importance.



PART 2

MACHINE LEARNING MODEL

: CLASSIFICATION



✓ Result : Train Model with Machine Learning Technique

Model : Decision Tree Classification

--- Model Evaluation (Decision Tree) ---

Model Accuracy: 0.9844

Classification Report:

	precision	recall	f1-score	support
0	0.00	0.00	0.00	1
1	0.98	0.98	0.98	94
2	0.99	0.99	0.99	161
accuracy			0.98	256
macro avg	0.66	0.66	0.66	256
weighted avg	0.98	0.98	0.98	256

Model : Random Forest Classification

--- Model Evaluation (Random Forest) ---

Model Accuracy: 0.9883

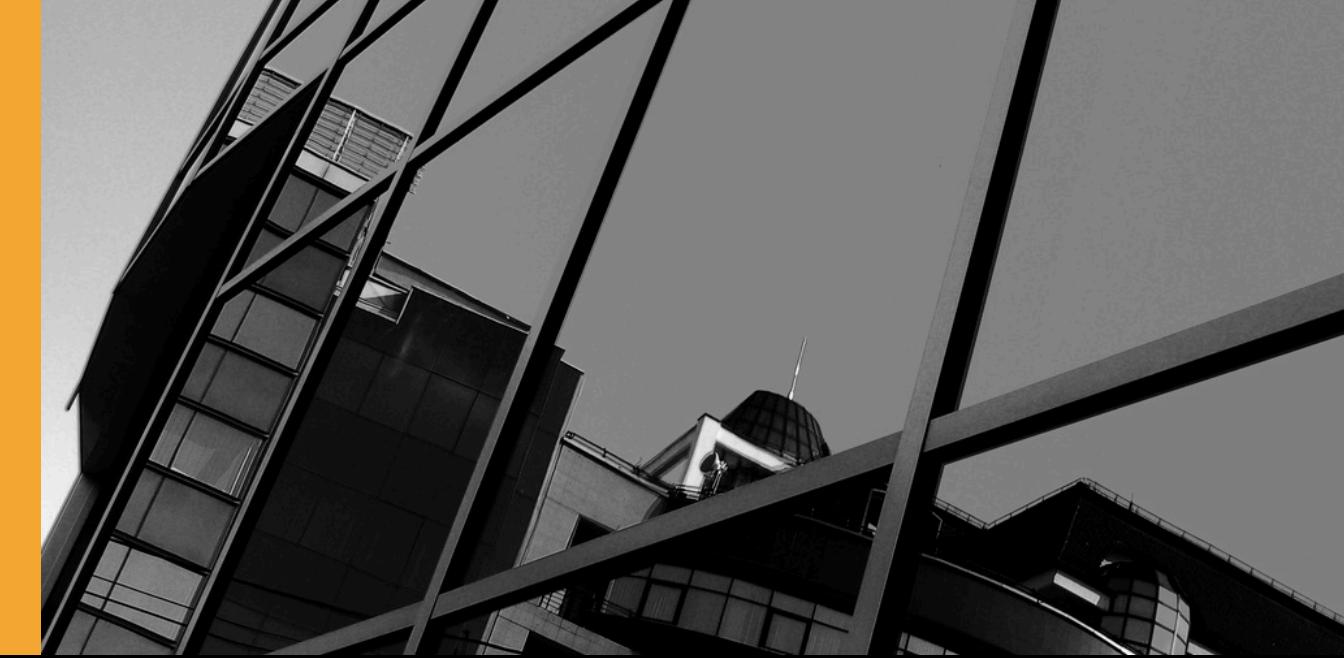
Classification Report:

	precision	recall	f1-score	support
0	0.00	0.00	0.00	1
1	0.97	1.00	0.98	94
2	1.00	0.99	0.99	161
accuracy				256
macro avg	0.66	0.66	0.66	256
weighted avg	0.98	0.99	0.99	256

PART 2

MACHINE LEARNING MODEL

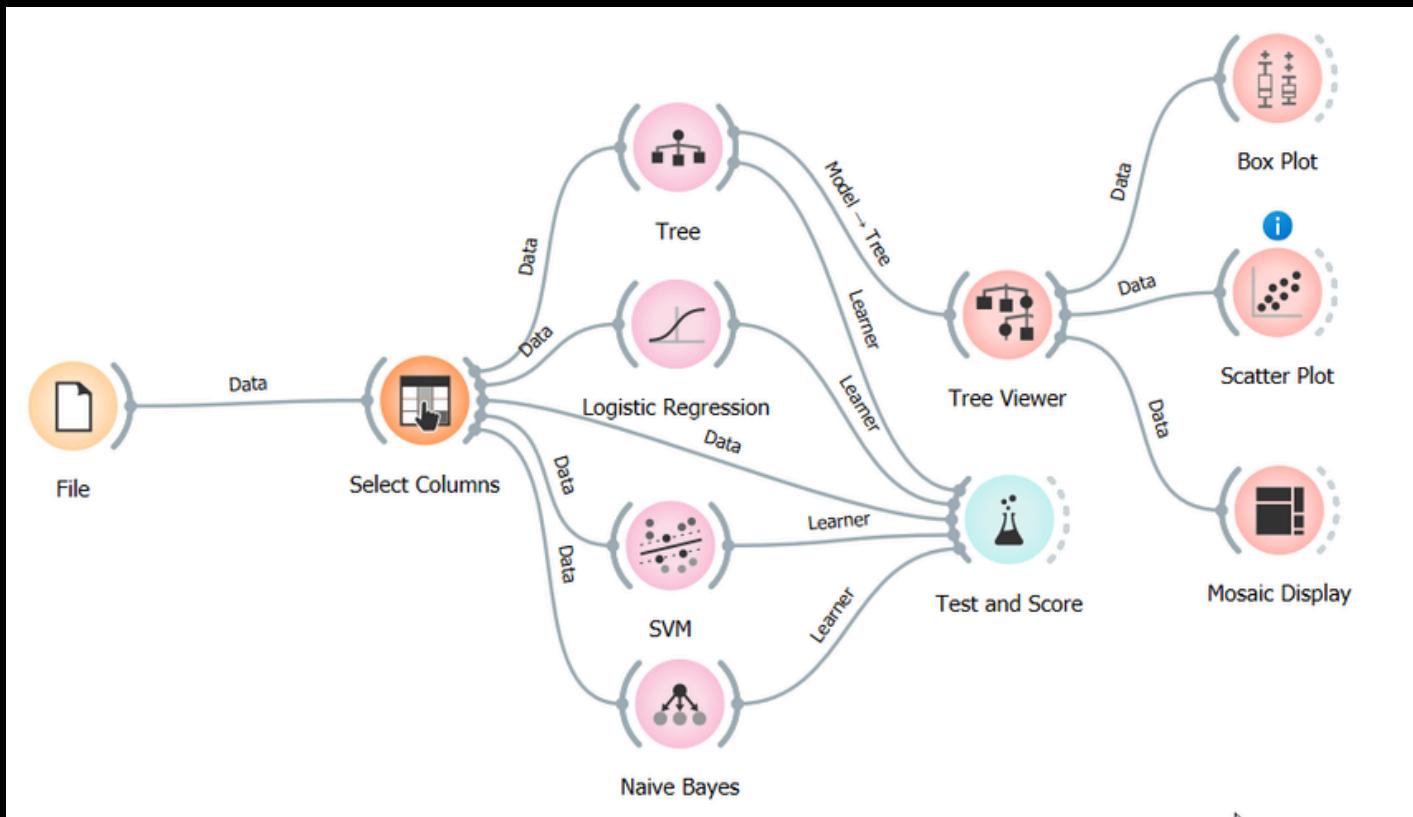
: CLASSIFICATION



✓ Compare to training model from previous project.



Model : Decision Tree Classification via *Orange Data Mining*



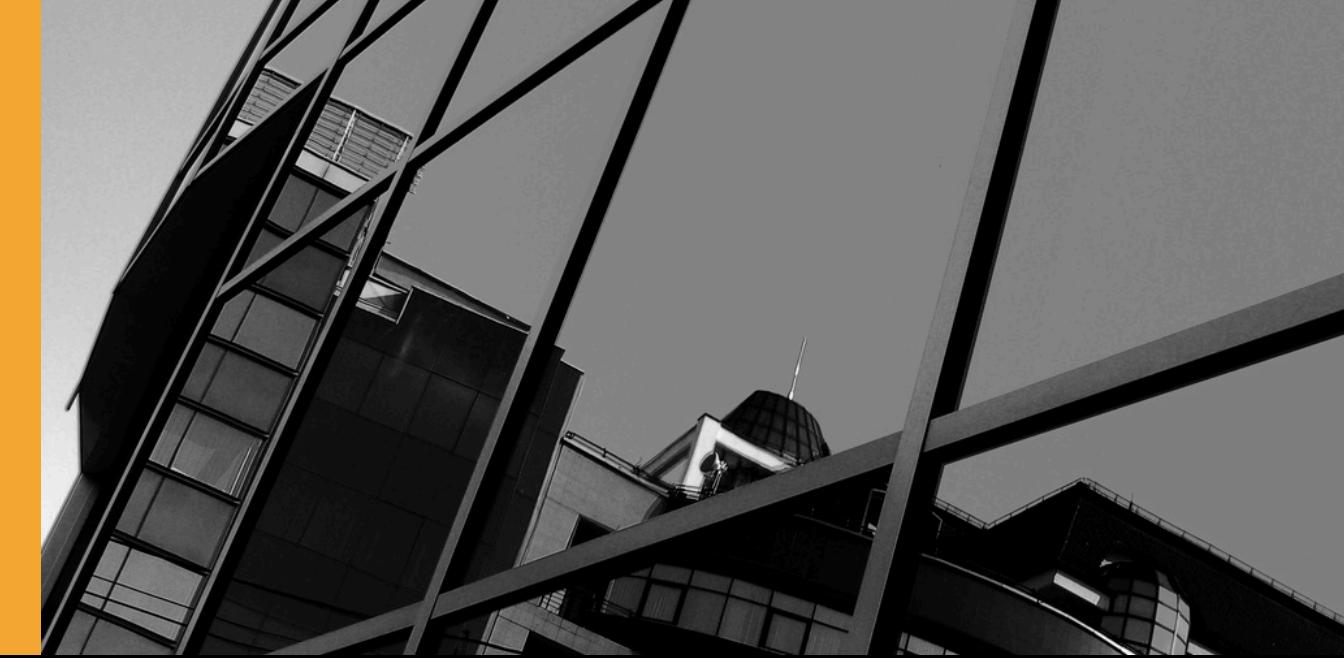
Model	AUC	CA	F1	Prec	Recall	MCC
Tree	0.919	0.862	0.860	0.860	0.862	0.718
Logistic Regression	0.873	0.797	0.797	0.810	0.797	0.608
Naive Bayes	0.865	0.770	0.767	0.814	0.770	0.589
SVM	0.807	0.695	0.688	0.764	0.695	0.469

Decision Tree: มีค่าความแม่นยำสูงสุด (AUC = 0.919, Accuracy = 0.862)
ทำงานได้ดีกว่า Model อื่น ๆ

PART 2

MACHINE LEARNING MODEL

: PRICE PREDICTION



✓ Random Forest Regressor

To predict monthly condo rental price (THB) based on user input (bedrooms, bathrooms, floor area, and proximity to rail)

Training & Prediction:

- Trained model on selected features
- Used `.predict()` on scaled user input to estimate condo price

Example output:

Predicted Price for your preferences: 13,049 THB

PART 3

STREAMLIT



Smart Rent BKK App

Page 1: Homepage

Page 2: Descriptive Analysis

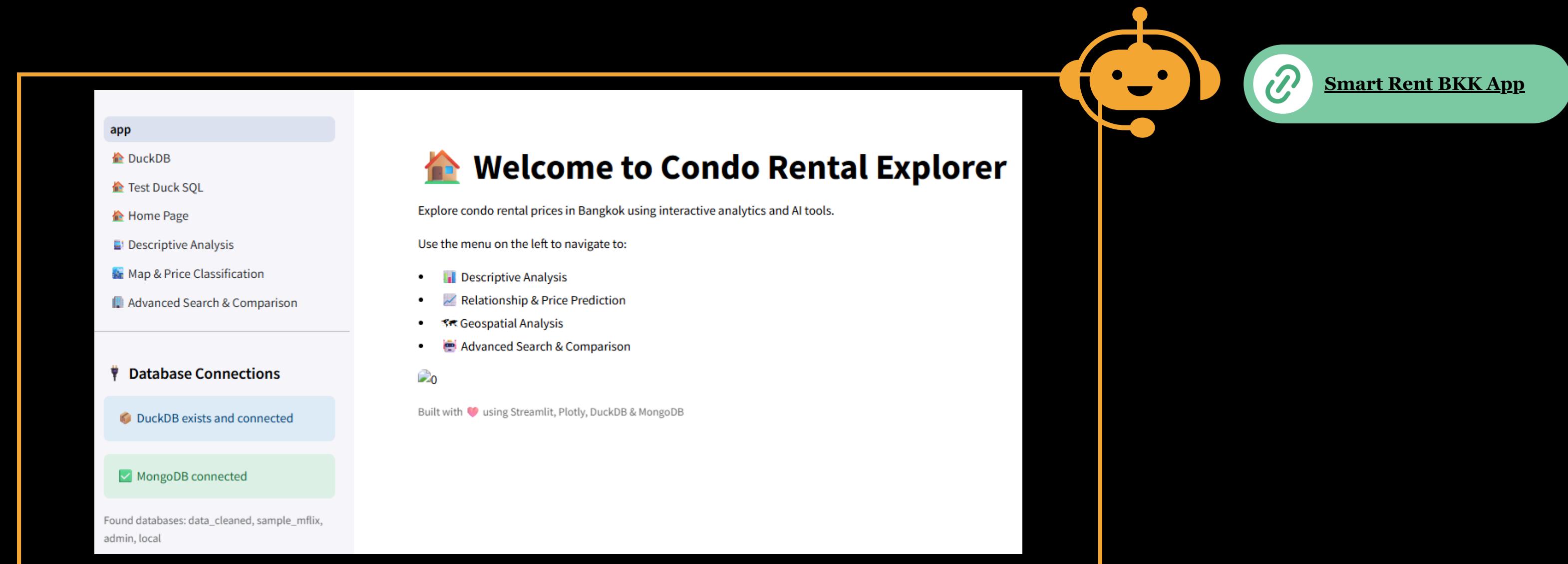
Page 3: Relationship Analysis & Predictive

Page 4: Bangkok Condo Map & Classification

Page 5: Advanced Search & Comparison

IMPROVEMENT FROM PREVIOUS PROJECT

- ✓ Insight Analysis: Relationship Between Rental Price and Distance to public transportation stations, Comparative Condo analysis, Average rental rate by public transportation, Interactive Condo Filtering via Map
- ✓ Price Prediction Model: Machine learning-based rental price estimator using key features (e.g., size, distance, rooms)
- ✓ AI Chatbot Assistant: Conversational agent designed to answer user queries and recommend condos based on preferences





Thank You