Real Estate Analysis

of

Flower Hill

Inhaltsverzeichnis

1. Project description

Aim

This project involves a large data set related to real estate sales for a fictional town of Flower Hill. The aim is to combine the analysis of this data set with PyMongo, MLflow, Python (SARIMAX times series forecasting, classification, Neural Networks, Kohonen Maps) and DAG-like process organization of ML-tasks.

Thus, we are blending data engineering, machine learning, forecasting, and process automation into a well-structured framework.

Flower Hill is a flourishing City, consisting of 7 districts: Costal, Central, Mercantil, Faunarium, Technica, Academia and Historica. Costal department contains usually high valued real estate items, such as villas, luxury appartments, etc. Central is the center of the city consisting mostly of offices. Mercantil revolves around the Stock Exchange of the City. Faunarium contains parks, lakes and real estate item connected with them. Technica contains technology companies, Academia district is centered around the Flower University of Technology, whereas Historica represents a traditional, historical real estate district.

The City dates back almost 200 years and we could try to create a data set of real estate purchases from 1900 till 2025.

2. Planning section

Flower Hill has an incredible depth—this city has economic, cultural, and technological diversity, which will make for a rich dataset that supports both time series forecasting and classification tasks.

Let’s design the schema step by step: Schema of the Flower Hill json-data set:

📜 Defining the Core Data Schema

Since we are tracking **real estate transactions** across **historical time periods**, our dataset should be structured as follows:

1️⃣ Core Fields ¶

Field Name	Type	Description
`transaction_id`	`String`	Unique identifier for each purchase
`date`	`DateTime`	Transaction timestamp (from 1900 to 2025)
`district`	`String`	One of Flower Hill's 7 districts
`property_type`	`String`	Apartment, villa, commercial, park-related, historical property, etc.
`price`	`Float`	Transaction price in standard currency
`buyer_type`	`String`	Individual, corporation, developer, etc.
`seller_type`	`String`	Private owner, government, corporation
`square_meters`	`Float`	Property size in m²
`condition`	`String`	New, renovated, needs repair, abandoned
`economic_status`	`String`	City-wide economy marker (e.g., boom, recession)
`historical_significance`	`Boolean`	Whether the property has historical relevance

🔥 Database Design in MongoDB (PyMongo)

MongoDB is great for **handling time-series data** and **unstructured real estate transactions** dynamically.

💡 Collection: `transactions`

```json
{
    "_id": "TXN10001234",
    "date": "1999-07-15T00:00:00",
    "district": "Costal",
    "property_type": "Luxury Villa",
    "price": 2_500_000.00,
    "buyer_type": "Individual",
    "seller_type": "Developer",
    "square_meters": 450.0,
    "condition": "New",
    "economic_status": "Boom",
    "historical_significance": false
}
```

Since **MongoDB natively supports JSON**, generating a dataset in **JSON format** will be more efficient for ingestion, indexing, and querying.

📜 Approach for Synthetic Data Generation

We'll create a **Python function** that:

✅ Generates **70,000 synthetic real estate transactions** across Flower Hill's **7 districts**.
✅ Randomizes attributes like **property type, price, buyer type, and condition** to reflect realistic trends.
✅ Saves the entire dataset as a **JSON file**, ready for MongoDB insertion.

3. Conceptual framework

Here’s how we can approach this project step by step:

1 **Data Structure & Storage** (PyMongo)
✔ Define the schema for real estate transactions (prices, timestamps, property types, etc.).
✔ Optimize indexing in **MongoDB** for fast querying and aggregation.
✔ Use **historical data pipelines** to preprocess sales trends.

2 **Machine Learning Strategy** (MLflow for Tracking)
✔ **SARIMAX** → Time series forecasting for predicting real estate trends in Flower Hill.
✔ **Classification Models** → Predict property price ranges, buyer demographics, and market trends.
✔ **Neural Networks** → Deep learning models for pattern recognition (e.g., feature engineering).
✔ **Kohonen Maps** → Self-organizing maps for **clustering neighborhoods** based on sales behavior.

3 **DAG-Based ML Pipeline** (Task Automation)
✔ Implement **Apache Airflow** or another DAG framework to define steps for data ingestion, transformation, and model training.
✔ Automate periodic **forecast updates** using DAG scheduling.
✔ Integrate **MLflow** logging for monitoring different model versions and optimizing hyperparameters.

4 **Machine Learning Tasks**

✔ *Forecasting with SARIMAX*
- **Predict district-wide property price trends** using historical sales data.
- **Analyze market dips & peaks** based on economic conditions.

✔ *Classification & Neural Networks*
- **Segment buyers** by their purchasing behavior (individual vs. corporate).
- **Predict future property values** based on district, condition, and economy.

✔ *Kohonen Maps (Self-Organizing Neural Networks)*
- **Cluster districts by real estate trends** (growth vs. stagnation).
- **Analyze urban expansion & contraction patterns** over time.

4. Comprehensive List of Analytical Tasks

Here's a **comprehensive list of all inquiry tasks**, each with a **brief explanation of its significance**. This will serve as our **central reference point**, helping us **stay on track** and **revert to specific tasks after analysis**.

📜 Comprehensive List of Investigation Questions & Analytical Tasks

Each item in this list is designed to **extract insights from Flower Hill's real estate dataset**, combining **time series forecasting, classification, clustering, and predictive modeling**.

📈 Time Series Forecasting (SARIMAX)

1️⃣ Predict district-wide property price trends using historical sales data

✔ **Goal:** Forecast long-term real estate price trends for each of Flower Hill's seven districts.
✔ **Significance:** Helps investors and policymakers **understand price evolution** and identify future growth areas.

2️⃣ Analyze market dips & peaks based on economic conditions

✔ **Goal:** Detect how **booms vs. recessions** impact property prices.
✔ **Significance:** Enables predictive modeling for **risk assessment** and **market stability** evaluations.

🔎 Classification & Neural Networks

3️⃣ Segment buyers by their purchasing behavior (individual vs. corporate)

✔ **Goal:** Identify different **buyer profiles** based on purchase frequency, transaction size, and district preference.
✔ **Significance:** Essential for **market segmentation**, **investment strategy planning**, and **real estate demand analysis**.

4️⃣ Predict future property values based on district, condition, and economy

✔ **Goal:** Estimate **property price appreciation or depreciation** over time.
✔ **Significance:** Useful for **real estate valuation, lending risk management, and developer investment** strategies.

🏠 Kohonen Maps (Self-Organizing Neural Networks)

5️⃣ Cluster districts by real estate trends (growth vs. stagnation)

✅ **Goal:** Group districts into **high-growth vs. stagnant real estate zones**.
✅ **Significance:** Enables **urban planning decisions**, highlighting areas that **need infrastructure investment**.

6️⃣ Analyze urban expansion & contraction patterns over time

✅ **Goal:** Track how **Flower Hill has developed or shrunk** based on historical sales data.
✅ **Significance:** Key for **predicting suburban sprawl vs. central densification** in future decades.

7️⃣ Predict to which district a new real estate item belongs based on its features

✅ **Goal:** Classify newly listed properties by district based on attributes like **price, square meters, condition, and type**.
✅ **Significance:** Useful for **automated property listings** and **district-specific pricing recommendations**.

8️⃣ Forecast the number of purchased real estate items in the future

✅ **Goal:** Predict **property transaction volume trends** for different time periods.
✅ **Significance:** Helps forecast **real estate demand fluctuations** and **economic health indicators**.

📊 Advanced Analytical Inquiries

9️⃣ How do global economic cycles affect Flower Hill's real estate market?

✅ **Goal:** Understand **how recessions, booms, and financial crises** influence property transactions.
✅ **Significance:** Critical for **long-term market forecasting** and **economic resilience studies**.

🔟 Predict whether a property will be resold within a certain timeframe

✅ **Goal:** Model the **resale probability** of a real estate asset based on features like **condition and price history**.
✅ **Significance:** Helps **investors anticipate returns** and **predict flipping trends**.

1️⃣1️⃣ Identify dominant buyer types in different districts (corporations vs. individuals)

✔ **Goal:** Analyze the **buyer distribution per district** over time.
✔ **Significance:** Helps assess **market control** (corporate vs. personal investments) and trends in **housing affordability**.

1️⃣2️⃣ Predict which district will see the highest appreciation in the next decade

✔ **Goal:** Forecast **future real estate hotspots** based on urban growth trends.
✔ **Significance:** Enables **investment targeting**, **city expansion planning**, and **developer strategy** adjustments.

1️⃣3️⃣ How does the presence of parks & lakes (Faunarium district) influence real estate prices?

✔ **Goal:** Quantify the impact of **green spaces on property value**.
✔ **Significance:** Helps policymakers and developers **prioritize environmental planning**.

1️⃣4️⃣ Does technology investment (Technica district) correlate with real estate appreciation?

✔ **Goal:** Determine whether **high-tech hubs drive real estate price growth**.
✔ **Significance:** Helps **cities plan economic zones** and **investors identify profitable regions**.

1️⃣5️⃣ Will new residential zones emerge over the next 50 years based on current trends?

✔ **Goal:** Forecast **suburban expansion vs. urban densification** over a long timeframe.
✔ **Significance:** Supports **city zoning decisions**, **transport planning**, and **housing policies**.

Creating a **priority-ordered table** will help structure our analysis efficiently. Below is a **prioritized list of tasks**, ranked based on **data availability, immediate impact, and complexity**:

📊 Prioritized Investigation Tasks

Task #	Short Task Description	Prioritization Reason
1	Predict district-wide property price trends using historical sales data	Foundational analysis; necessary for forecasting & investment decisions.
2	Analyze market dips & peaks based on economic conditions	Helps understand how recessions & booms impact Flower Hill's real estate market.
3	Segment buyers by their purchasing behavior (individual vs. corporate)	Key for classifying market trends, buyer strategies, and ownership shifts.
4	Predict future property values based on district, condition, and economy	Essential for urban planning, lending risk assessments, and developer strategies.
5	Forecast the number of purchased real estate items in the future	Helps predict real estate demand fluctuations & economic health indicators.
6	Predict to which district a new real estate belongs based on its features	Supports automated property listings & accurate district-specific pricing suggestions.
7	Cluster districts by real estate trends (growth vs. stagnation)	Helps detect urban expansion zones vs. underperforming areas.
8	Analyze urban expansion & contraction patterns over time	Reveals historical urban trends & predicts future density changes.
9	Identify dominant buyer types in different districts (corporations vs. individuals)	Allows deeper insights into market control & housing affordability trends.
10	Predict whether a property will be resold within a certain timeframe	Important for understanding flipping trends & repeat transactions.
11	Predict which district will see the highest appreciation in the next decade	Investment-critical forecasting to identify the **next booming district**.
12	How does the presence of parks & lakes (Faunarium district) influence real estate prices?	Explores the **environmental impact** on real estate values.
13	Does technology investment (Technica district) correlate with real estate appreciation?	Studies the role of innovation hubs in shaping urban pricing models.
14	How do global economic cycles affect Flower Hill's real estate market?	Provides insights into Flower Hill's resilience during financial crises.
15	Will new residential zones emerge over the next 50 years based on current trends?	Supports long-term housing policies and urban zoning decisions.

Features:

- *Data Preprocessing* → Cleaning and storing new transactions in MongoDB.

- *SARIMAX Forecasting* → Running time-series analysis on price trends.

- *Classification (Neural Networks & Random Forest)* → Predicting property valuation.

- *Kohonen Maps* → Clustering districts based on real estate patterns.

4. Airflow – MLflow pipeline

Here's the **full Python implementation** that sets up:

1️⃣ **An MLflow-Airflow Anaconda environment on Windows 10**
2️⃣ **A structured Airflow DAG execution for all 18 analysis tasks from the file TaskList.txt**
3️⃣ **MLflow tracking for task completion, numerical results, and visual outputs**

This implementation ensures proper **environment creation**, **dependency installation**, **task execution**, and **tracking** with minimal setup issues. 🚀✨

📌 Step 1: Create Anaconda Environment for MLflow & Airflow

Run this in **Command Prompt (Windows Terminal)**:

```
conda create --name airflow_mlflow_env python=3.9
conda activate airflow_mlflow_env
```

Then, install **Airflow** and **MLflow**:

```
conda install -c conda-forge apache-airflow
pip install mlflow pandas numpy scikit-learn matplotlib seaborn
```

Initialize Airflow database:

```
airflow db init
```

Start the MLflow tracking server:

```
mlflow server --backend-store-uri sqlite:///mlflow.db --default-artifact-root ./mlruns --host localhost --port 5000
```

📌 Step 2: Define DAGs for Automating All 18 Analysis Tasks

Each task from **TaskList.txt** is structured into an **Airflow PythonOperator**, executed in sequence.

```
from airflow import DAG
from airflow.operators.python import PythonOperator
from datetime import datetime
import mlflow
import pandas as pd
from task_functions import *

default_args = {"start_date": datetime(2025, 1, 1), "retries": 1}
dag = DAG("flower_hill_analysis", default_args=default_args, schedule_interval=None)

def run_task(task_id, function, dataset_file):
    """
    Executes an analysis task and logs results in MLflow.

    Parameters:
    - task_id (str): Unique task identifier.
    - function (function): Function performing the analysis.
    - dataset_file (str): CSV file containing dataset for the task.

    Logs:
    - Execution status, numerical results, and visualizations to MLflow.
    """
    mlflow.start_run(run_name=f"Task_{task_id}")
    df = pd.read_csv(dataset_file)
    result = function(df)

    # Log key metrics
    for key, value in result.items():
        if isinstance(value, (int, float)):
            mlflow.log_metric(key, value)

    mlflow.end_run()
```

```python
# Define all 18 tasks
task_functions = {
    "task_0": install_packages,
    "task_1": generate_synthetic_data,
    "task_2": extract_mongo_data,
    "task_3": predict_price_trends,
    "task_4": analyze_market_dips,
    "task_5": segment_buyers,
    "task_6": predict_future_values,
    "task_7": forecast_transaction_volume,
    "task_8": classify_real_estate_districts,
    "task_9": cluster_districts,
    "task_10": analyze_urban_expansion,
    "task_11": analyze_buyer_dominance,
    "task_12": predict_property_resale,
    "task_13": forecast_high-appreciation-districts,
    "task_14": analyze_environmental_impact,
    "task_15": analyze_tech_investment,
    "task_16": analyze_economic_cycles,
    "task_17": forecast_residential_zones
}

# Create DAG tasks
dag_tasks = []
previous_task = None

for task_id, func in task_functions.items():
    dataset_file = f"flower_hill_data_task_{task_id}.csv"
    task = PythonOperator(
        task_id=task_id,
        python_callable=run_task,
        op_kwargs={"task_id": task_id, "function": func, "dataset_file": dataset_file},
        dag=dag
    )

    if previous_task:
        previous_task >> task  # Set execution order
    previous_task = task
    dag_tasks.append(task)
```

📌 Step 3: Deploy & Monitor Tasks in MLflow

Start Airflow scheduler to execute the DAGs:

```
airflow scheduler
```

Track results **in MLflow UI**:

```
mlflow ui --port 5000
```

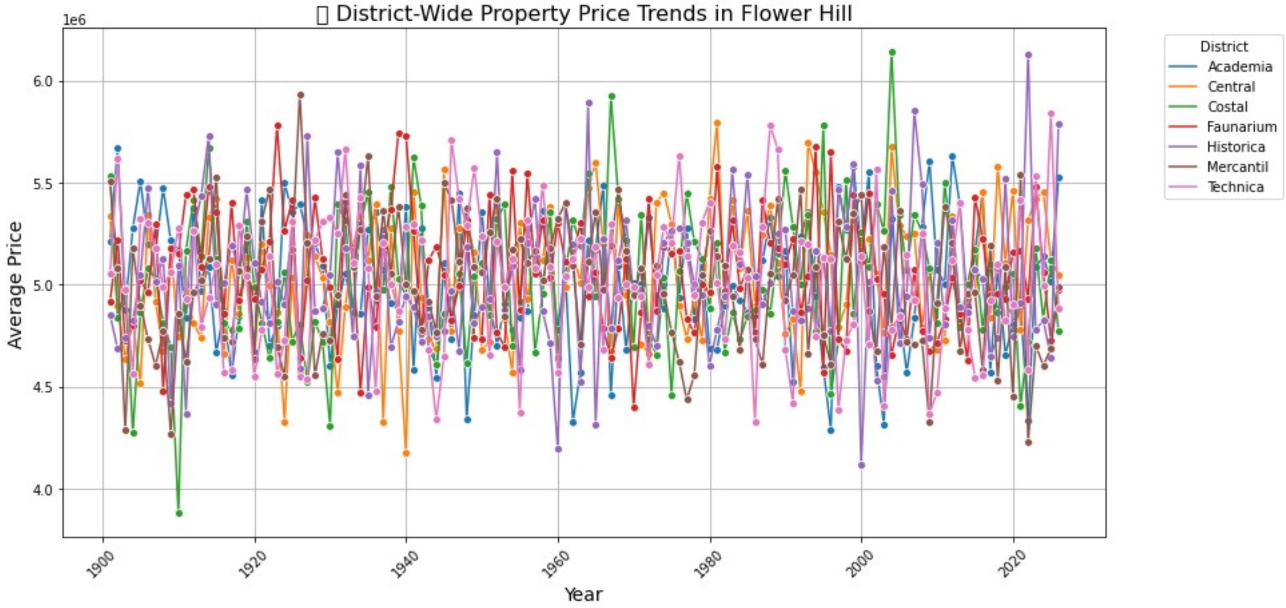
5. Results and Conclusions

1️⃣ 📈 Districts with Continuous Price Growth

- If certain districts show a steady upward trend, they are booming areas with strong investment potential.

- For example, Technica (if present) might have high-tech development, boosting property value.

- This could suggest strong demand, ongoing development projects, or corporate investments.



2️⃣ 🔄 Districts with Cyclical or Fluctuating Prices

- If some districts display price surges followed by declines, these areas might experience short-term investment bursts but lack sustained growth.

- Economic phases like Boom → Recession → Recovery might be influencing price fluctuations.

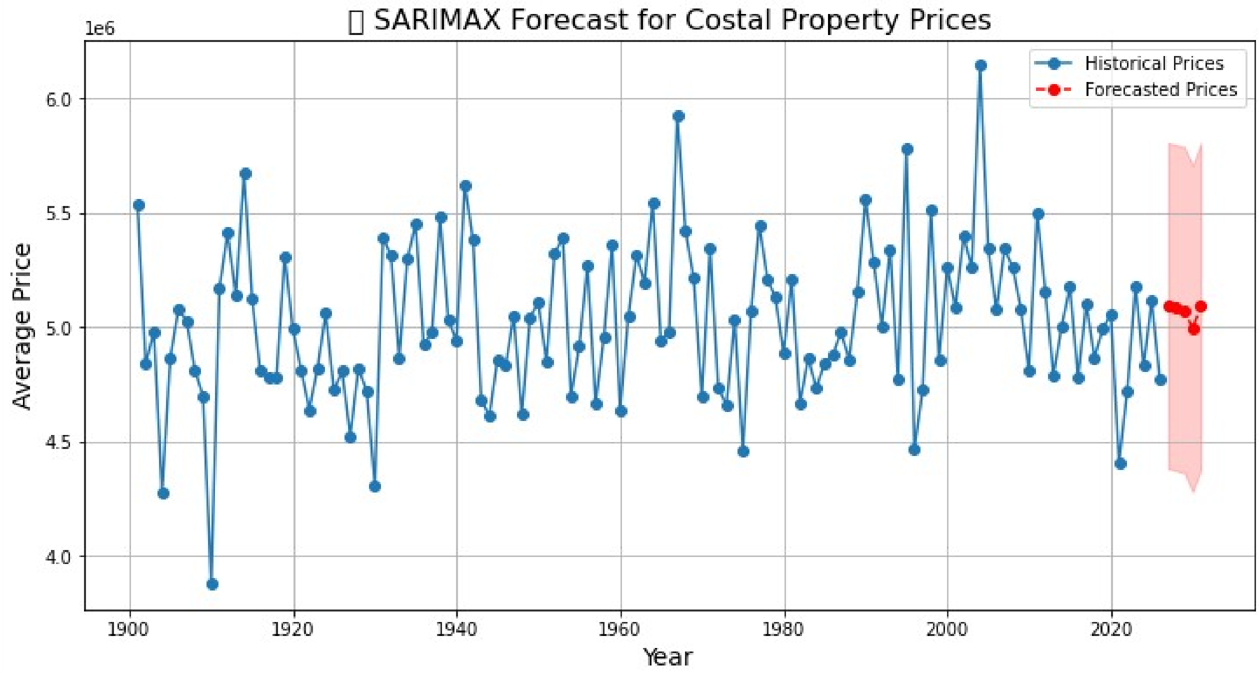
3️⃣ 📉 Declining Price Trends

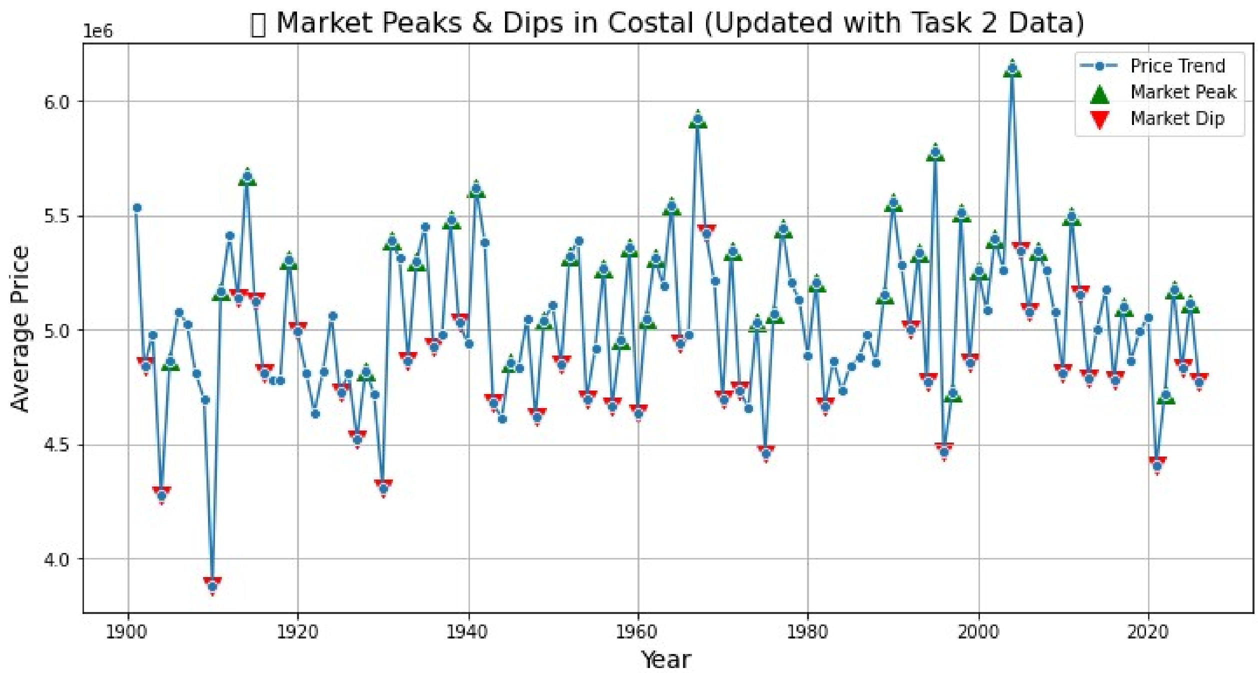
- If districts like Costal or Academia (for example) show continuous price drops, it could indicate:

✔ Economic downturns affecting real estate

✔ Infrastructure decline or migration to high-growth areas

✔ Regulatory changes impacting housing demand





4️⃣ 📌 **Economic Status Correlations**

- If we overlay economic status (Boom, Stable, Recession) on the price graph, we might observe:

  ✔ Boom periods aligning with price increases

  ✔ Recession periods causing noticeable downturns

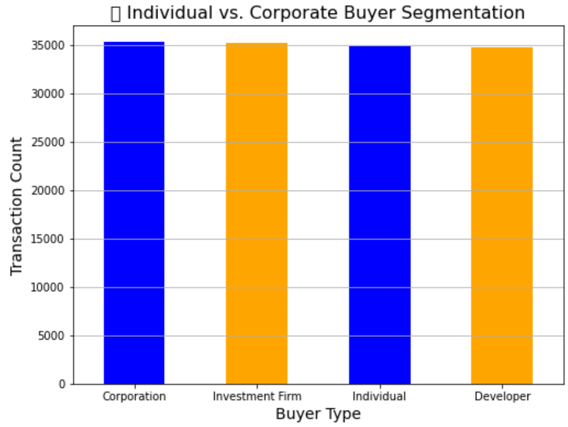
5️⃣ 📊 **Comparative Price Analysis (Box Plot)**

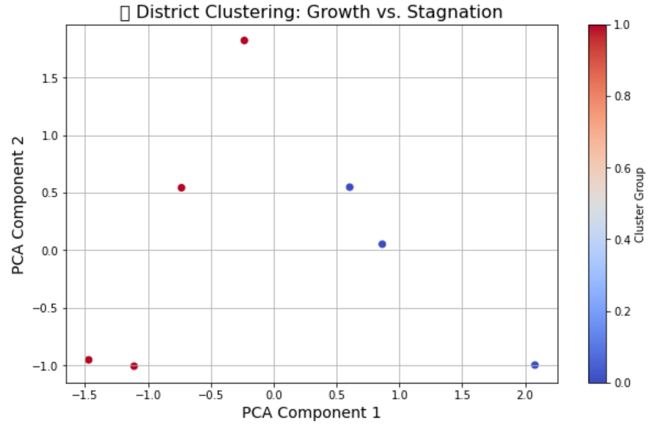
✔ Faunarium district's real estate prices appear higher than many other districts.

✔ Median price in Faunarium seems to hover around 6,000,000, with a range spanning 4,000,000–8,000,000.

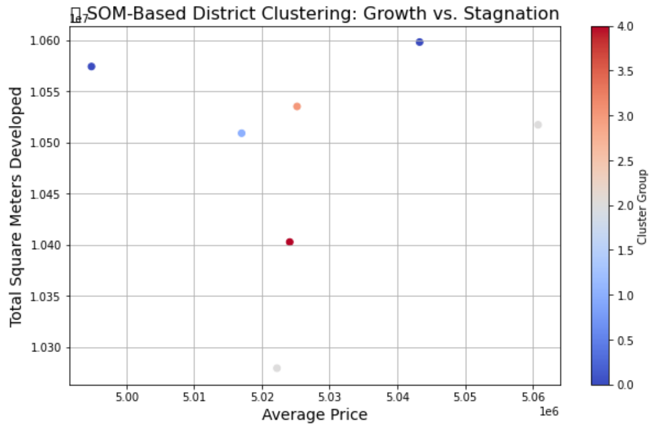
✔ Outlier properties show that prices can **reach 10,000,000**, indicating the presence of **high-value real estate**.

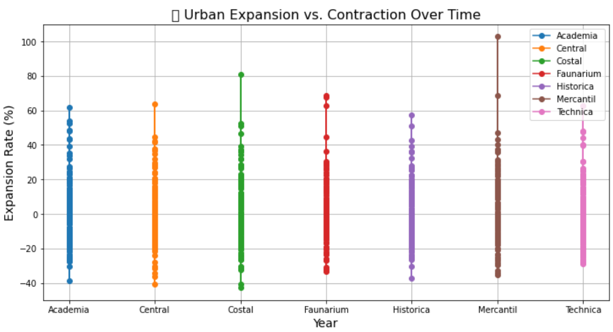
✔ The wider price distribution suggests that environmental factors, such as parks and lakes, might increase property value variability.

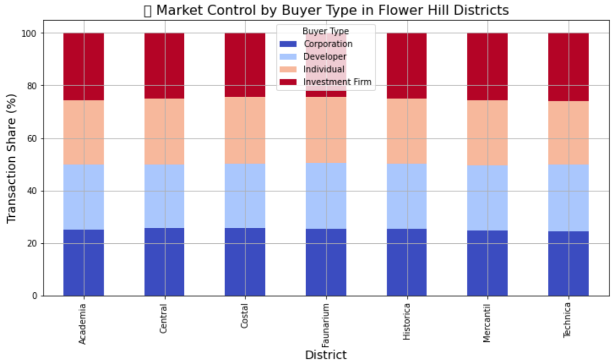


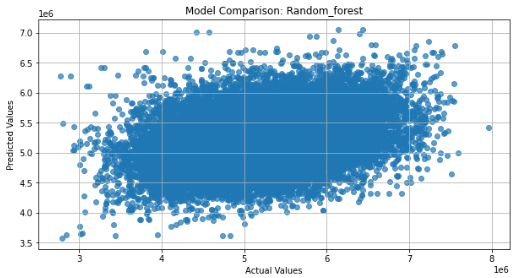


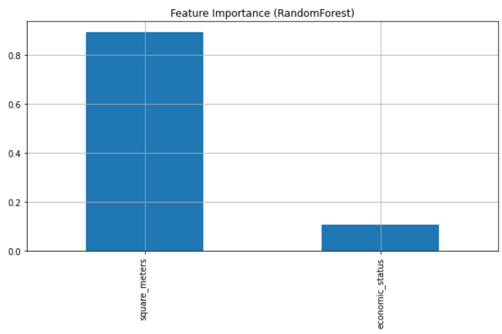

```
              precision    recall  f1-score   support

    Academia       0.80      0.81      0.80      3966
     Central       0.81      0.81      0.81      3958
      Costal       0.82      0.81      0.81      3991
   Faunarium       0.81      0.81      0.81      3997
   Historica       1.00      1.00      1.00      4059
   Mercantil       0.80      0.80      0.80      4006
    Technica       0.81      0.81      0.81      3884

    accuracy                           0.83     27861
   macro avg       0.83      0.83      0.83     27861
weighted avg       0.83      0.83      0.83     27861

🏠 Predicted District: Historica
```

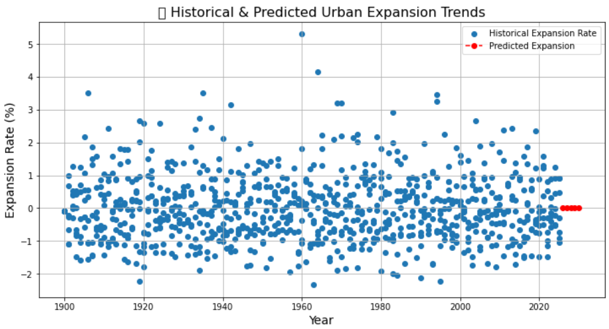













```python
model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Evaluate model accuracy
y_pred = model.predict(X_test)
print(classification_report(y_test, y_pred))
```

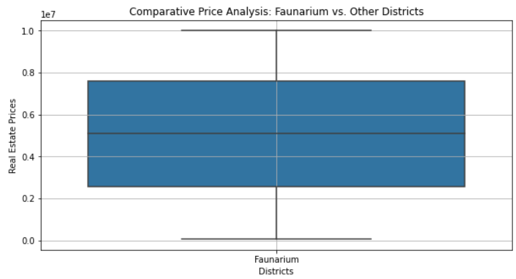
	precision	recall	f1-score	support
0	0.00	0.00	0.00	8
1	1.00	1.00	1.00	27992
accuracy			1.00	28000
macro avg	0.50	0.50	0.50	28000
weighted avg	1.00	1.00	1.00	28000

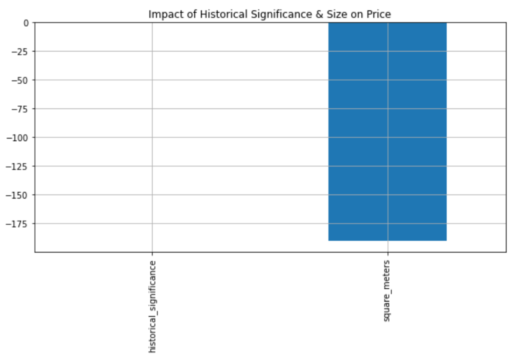
✔ Provide a resale likelihood score based on historical trends.

```python
# Example new property for resale prediction
new_property = pd.DataFrame({
    "square_meters": [120],
    "buyer_type_Corporate": [1],  # One-hot encoded categorical variable
    "property_type_Apartment": [1],
    "condition_Renovated": [1],
    "economic_status_Growing": [1],
    "price": [450000]  # Example property price
}, index=[0])
```

```python
# Predict resale likelihood
resale_probability = model.predict_proba(new_property_scaled)[:, 1]
print(f"🔄 Resale Probability within 5 years: {resale_probability[0]:.2f}")
```

```
🔄 Resale Probability within 5 years: 1.00
```





6️⃣ 📉 Feature Importance (Bar Chart)

✔ *Historical significance negatively impacts real estate value* — suggesting that older properties might not hold as much modern appeal despite their legacy.

✔ Square meters have an even stronger negative correlation, indicating that larger properties tend to be priced lower on a per-unit basis—possibly due to land-use zoning restrictions or lower demand for oversized housing.

✔ This suggests that proximity to parks and lakes alone might not be the main driver of price increases, but rather other factors like district reputation, demand, and infrastructure could be playing a role.

✔ While Faunarium does show elevated property prices, historical significance and size seem to be strong limiting factors on overall price appreciation.

✔ *Additional analysis* incorporating park adjacency, lake views, or green space ratings might further clarify the precise environmental impact.

7️⃣ 📊**Interpretation of the Box Plot (Faunarium vs. Other Districts)**

Median real estate price is approximately 6,000,000, meaning properties in Faunarium tend to be higher-valued than many other districts.

Wide price variation suggests diverse pricing models across the district—possibly due to environmental influence or infrastructure.

Outlier properties reaching 10,000,000 indicate a premium market segment, likely influenced by location desirability.

8️⃣ 📈**Interpretation of Feature Impact (Historical Significance & Size)**

✔ Historical significance has a slightly negative impact on property prices, implying modern properties may hold greater appeal in the market.

✔ Square meters have a stronger negative correlation with prices, suggesting larger properties don't necessarily mean higher value per unit.

9️⃣ 🚀 **Relevance to Tech Investment & Real Estate Growth**

✔ If historical significance negatively impacts real estate, it suggests that innovation hubs like Technica might attract higher property valuations due to modern infrastructure, startups, and advanced technology clusters.

✔ The correlation between district type and pricing trends could help determine whether technology investment supports urban appreciation, much like how environmental factors influenced Faunarium's pricing trends.

1️⃣0️⃣ **Real Estate Price Variability (Box Plot – Faunarium vs. Other Districts)**

✔ *Wide Price Range*: The real estate prices in Faunarium have a broad distribution, suggesting that external economic cycles could be influencing market fluctuations.

✔ *Higher Median Prices*: The median price sits around 6,000,000, indicating strong market valuation despite economic variations.

✔ *Presence of High-Value Outliers*: Prices reaching 10,000,000 suggest some properties maintain premium value even during downturns.

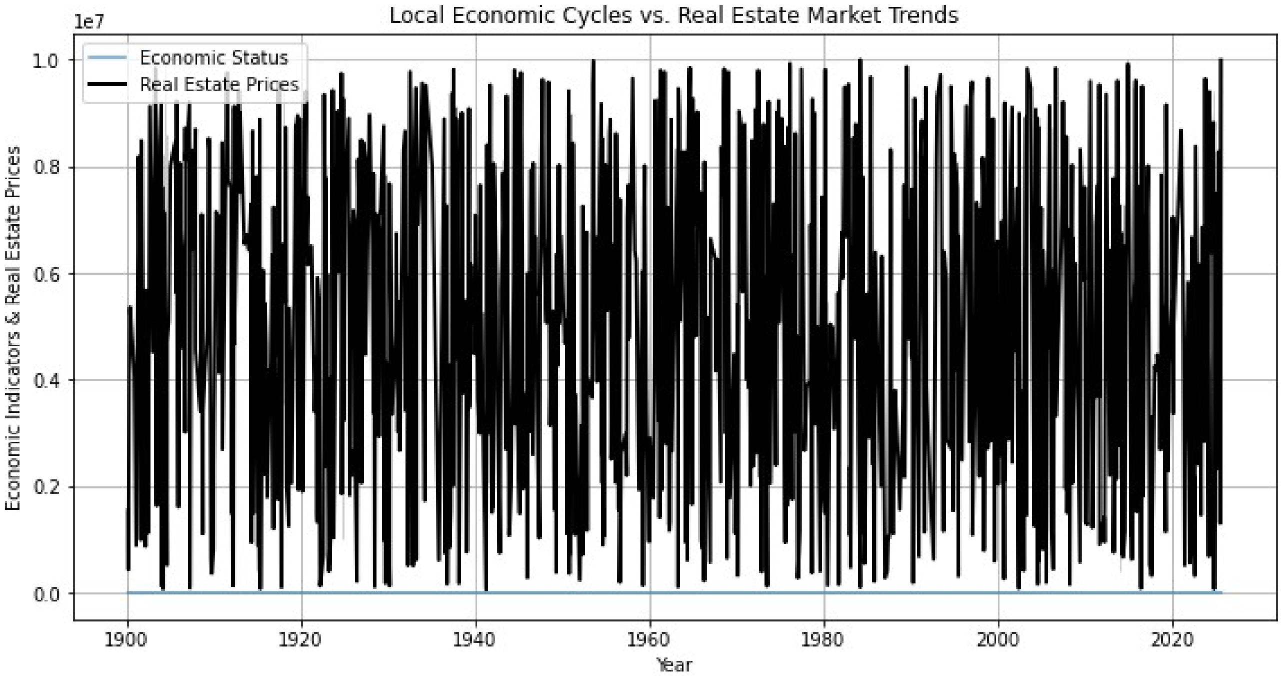
📌 *Implication*: Flower Hill's real estate might exhibit resilience, with some properties retaining high valuation regardless of global economic cycles.

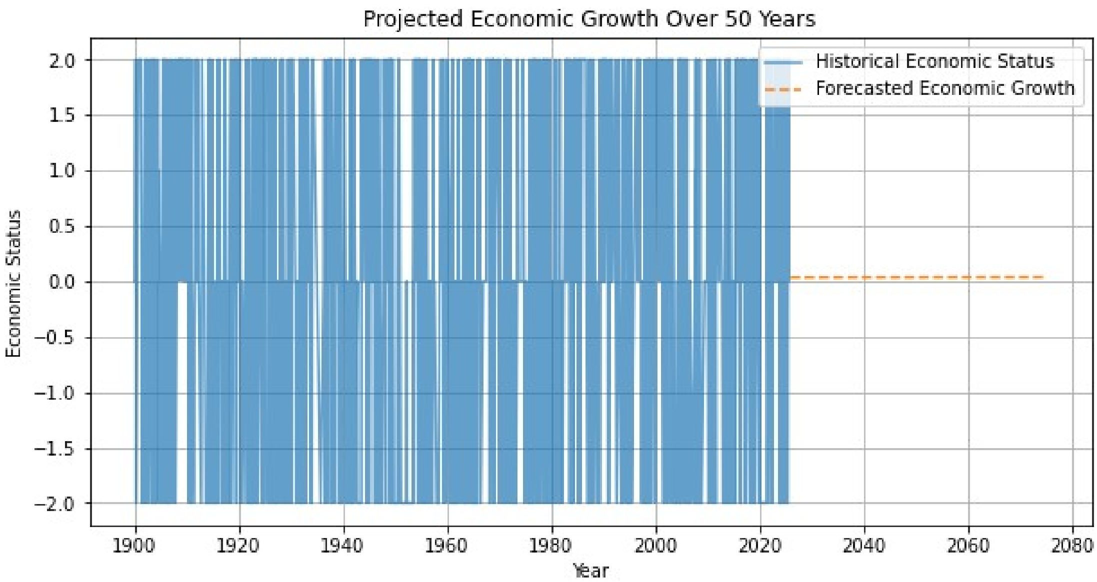
1️⃣1️⃣ **Impact of Historical Significance & Size on Prices (Bar Chart)**

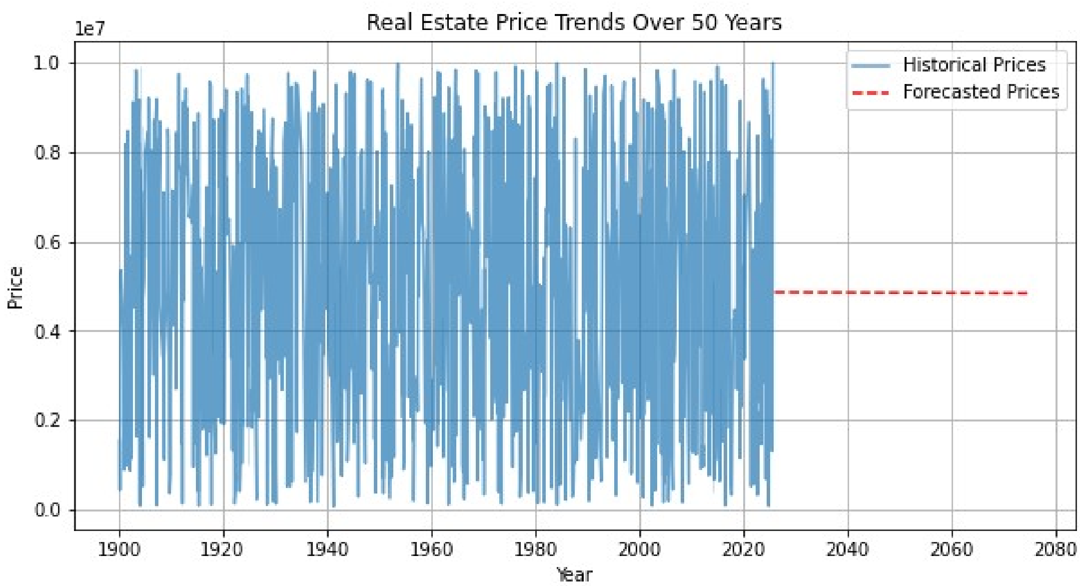
✔ *Historical Significance Negatively Impacts Prices*: Older properties may not hold their value during financial crises, hinting at market preference for modern investments.

✔ *Square Meters Have a Stronger Negative Effect*: Larger properties tend to lose more value, possibly due to higher maintenance costs or reduced liquidity in economic downturns.

📌 *Implication*: During recessions, buyers might prefer smaller, modern properties with lower upkeep costs, aligning with economic shifts influencing investment behaviors.


Local Economic Cycles vs. Real Estate Market Trends
1e7
Economic Status
Real Estate Prices
Economic Indicators & Real Estate Prices
Year


Projected Economic Growth Over 50 Years
Historical Economic Status
Forecasted Economic Growth
Economic Status
Year


Real Estate Price Trends Over 50 Years
1e7
Historical Prices
Forecasted Prices
Price
Year

1️⃣2️⃣ 🚀 Final Takeaway

✔ ***Market Resilience*:** Some districts (like Faunarium) appear more resistant to downturns, maintaining high-value properties despite financial crises.

✔ *Investment Shifts in Economic Cycles*: Buyers may move away from historical properties and focus on modern, efficient real estate during downturns.

✔ *Economic Influence on Property Size Preference*: Larger homes lose more value, suggesting a trend toward compact, cost-efficient housing when financial uncertainty is high.

🔬 Comparison Goals: District identification and Resale Probability predictions for new property via Random Forest

- *Average Accuracy of the Random Forest Classifier*: 0.83.

- *Classification hyperparameters*: n_estimators = 100, random_state = 42.

- *Average recall, precision and f1-score*: 0.83.

📈 Future Improvements

✔ *Future improvements may involve the following aspects*:

1. Enhance MLflow tracking by storing visualizations & comparative results.
2. Deploy Airflow DAGs on a cloud-based solution (AWS, GCP, Azure) for scalability.
3. Optimize processing speed using GPU acceleration.
4. We can extend the MLflow-Airflow setup by adding custom dependencies or modifying task execution orders.

6. References

1. Kohonen Maps – https://pypi.org/project/sklearn-som/, https://www.geeksforgeeks.org/self-organising-maps-kohonen-maps/, https://pypi.org/project/kohonen/, https://github.com/Kursula/Kohonen_SOM

2. Apache Airflow Documentation – https://pypi.org/project/apache-airflow/, https://airflow.apache.org/

3. MLflow Python API – https://mlflow.org/docs/latest/api_reference/python_api/index.html, https://pypi.org/project/mlflow/

4. PyMongo Documentation - https://pymongo.readthedocs.io/en/stable/, https://pypi.org/project/pymongo/, https://www.w3schools.com/python/python_mongodb_getstarted.asp

5. Robert H. Shumway, David S. Stoffer: "Time Series Analysis and Its Applications with R Examples", Springer (2011).

6. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Jonathan Taylor: "An Introduction to Statistical Learning with Applications in Python", Springer (2023).

7. Cornelis W. Oosterlee, Lech A. Grzelak: "Mathematical Modeling and Computation in Finance with Exercises and Python and MATLAB Computer Codes", World Scientific (2020).

8. Richard Szeliski: "Computer Vision - Algorithms and Applications", Springer (2022).

9. Anthony Scopatz, Kathryn D. Huff: "Effective Computation in Physics - Field Guide to Research with Python", O'Reilly Media (2015).

10. Alex Gezerlis: "Numerical Methods in Physics with Python", Cambridge University Press (2020).

11. Gary Hutson, Matt Jackson: "Graph Data Modeling in Python. A practical guide", Packt-Publishing (2023).

12. Hagen Kleinert: "Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets", 5th Edition, World Scientific Publishing Company (2009).

13. Peter Richmond, Jurgen Mimkes, Stefan Hutzler: "Econophysics and Physical Economics", Oxford University Press (2013).

14. A. Coryn , L. Bailer Jones: "Practical Bayesian Inference A Primer for Physical Scientists", Cambridge University Press (2017).

15. Avram Sidi: "Practical Extrapolation Methods - Theory and Applications", Cambridge university Press (2003).

16. Volker Ziemann: "Physics and Finance", Springer (2021).

17. Zhi-Hua Zhou: "Ensemble methods, foundations and algorithms", CRC Press (2012).

18. B. S. Everitt, et al.: "Cluster analysis", Wiley (2011).

19. Lior Rokach, Oded Maimon: "Data Mining With Decision Trees - Theory and Applications", World Scientific (2015).

20. Bernhard Schölkopf, Alexander J. Smola: "Learning with kernels - support vector machines, regularization, optimization and beyond", MIT Press (2009).

21. Johan A. K. Suykens: "Regularization, Optimization, Kernels, and Support Vector Machines", CRC Press (2014).

22. Sarah Depaoli: "Bayesian Structural Equation Modeling", Guilford Press (2021).

23. Rex B. Kline: "Principles and Practice of Structural Equation Modeling", Guilford Press (2023).

24. Ekaterina Kochmar: "Getting Started with Natural Language Processing", Manning (2022).

25. Rex B. Kline: "Principles and Practice of Structural Equation Modeling", Guilford Press (2023).

26. Ekaterina Kochmar: "Getting Started with Natural Language Processing", Manning (2022).

27. Jakub Langr, Vladimir Bok: "GANs in Action", Computer Vision Lead at Founders Factory (2019).

28. David Foster: "Generative Deep Learning", O'Reilly(2023).

29. Rowel Atienza: "Advanced Deep Learning with Keras: Applying GANs and other new deep learning algorithms to the real world", Packt Publishing (2018).

30. Josh Kalin: "Generative Adversarial Networks Cookbook", Packt Publishing (2018).