PARKING LOT OCCUPANCY ANALYSIS

RESOURCES:

Parking Lot Data Analytics

This repository showcases an end-to-end analysis of parking lot data using advanced machine learning techniques and dynamic pricing models. The project integrates data preprocessing, visualization, and performance evaluation of models such as Linear Regression, Random Forest, and K-Means clustering.

Features

- **Data Preprocessing**:
    - One-hot encoding for categorical variables.
    - Extraction of numerical features (e.g., total work hours).
    - Handling and filtering of raw data.
- **Visualization**:
    - Occupancy rate analysis by district.
    - Clustering of parking lots using K-Means.
    - Comparison of model performance metrics.
- **Machine Learning Models**:
    - Linear Regression for occupancy rate prediction.
    - Random Forest Regressor for enhanced predictions.
    - K-Means clustering for grouping parking lots based on capacity and occupancy rate.
- **Dynamic Pricing**:

- Pricing model based on occupancy rate, peak hours, and location popularity.

Requirements

Install the required Python libraries using the following command:

pip install pandas matplotlib scikit-learn folium numpy

Data Source

The data used in this project is sourced from the ISPAK Parking API, which provides detailed information about parking lots in Istanbul, including capacity, occupancy, free parking time, and more.

Usage

1. Clone this repository:
2. git clone https://github.com/your-username/parking-lot-data-analytics.git

cd parking-lot-data-analytics

3. Install the required dependencies:

pip install -r requirements.txt

4. Open the notebook:

jupyter notebook ispark_da.ipynb

5. Run the cells to explore the analysis, visualizations, and model performance.

Directory Structure

parking-lot-data-analytics/

|

├── data/

| ├── example_data.csv # Example parking lot dataset

```
|
├── notebooks/
|   ├── ispark_da.ipynb    # Main analysis notebook
|
├── src/
|   ├── dynamic_pricing.py # Dynamic pricing implementation
|   ├── model_training.py  # Machine learning models
|
├── README.md             # Project documentation
```

Highlights

Data Visualization

- **Occupancy Rate by District**
  Analyzes parking lot occupancy rates across districts.
- **Parking Lot Clusters (K-Means)**
  Visualizes clustering of parking lots based on total capacity and occupancy rate.

Model Performance Comparison

Model	MAE (Mean Absolute Error)	R^2 Score
Linear Regression	5.32	0.75
Random Forest	3.21	0.89

Future Enhancements

- Include external data sources like weather and real-time traffic for richer analysis.

- Expand the dynamic pricing model to account for seasonal trends and historical patterns.
- Integrate live APIs for real-time parking lot updates and predictions.

GITHUB PROJECT FILE

https://github.com/busrayatlav/parking-lot-data-analytics.git

ispark_da.ipynb

THE DATA SCIENCE PROCESS FOR PARKING LOT DATA ANALYTICS PROJECT CAN BE BROKEN DOWN INTO THE FOLLOWING KEY STEPS:

1. Defining Research Goals

This is the foundation of the project, where you determine the objectives and expected outcomes.

Key Research Questions for Project:

- What factors influence parking lot occupancy rates in Istanbul?
- How can we predict future occupancy rates based on historical data?
- Can we group parking lots based on their capacity and usage patterns?
- How can dynamic pricing be optimized based on occupancy levels, time of day, and location?

The end goal is to develop machine learning models that provide occupancy predictions and implement an effective dynamic pricing strategy.

2. Retrieving Data

Data retrieval involves collecting raw data from relevant sources.

Data Source for Project:

- **ISPAK Parking API**: This provides real-time and historical data on parking lots, including:
    - Parking lot name
    - Capacity
    - Current occupancy
    - Free parking time
    - Location details (district, coordinates)
    - Pricing details

Steps to Retrieve Data:

1. Use Python's **requests** library to pull data from the ISPAK API.
2. Store the retrieved data in CSV/JSON format for further analysis.
3. Ensure regular updates if working with real-time data.

3. Data Preparation

Before applying machine learning models, the raw data must be cleaned and structured.

Key Steps:

- **Handling Missing Values**: Fill or remove missing occupancy values.
- **Data Transformation**:
    - Convert categorical data (e.g., districts) into numerical form using **one-hot encoding**.

    - Extract numerical features (e.g., **total work hours**, peak time usage).
- **Feature Engineering**:
    - Create new features such as **average daily occupancy rate** or **weekend vs. weekday trends**.
- **Filtering Data**:
    - Remove parking lots with unreliable or incomplete data.
    - Exclude extreme outliers in occupancy rates.

4. Data Exploration

At this stage, you analyze patterns and relationships within the data.

Techniques Used in Project:

- **Descriptive Statistics**:
    - Compute mean, median, standard deviation of occupancy rates.
- **Visualization Techniques**:
    - **Occupancy Rate by District**: Bar plots to compare parking demand across districts.
    - **Time-Series Analysis**: Line charts to study occupancy trends over time.
    - **Correlation Matrix**: Understand relationships between features.
    - **K-Means Clustering**: Group parking lots based on occupancy behavior.

Key Insights:

- Identify peak usage times.
- Determine which districts have consistently high or low occupancy rates.

- Find parking lots with similar patterns for **pricing optimization**.

5. Model Building

This involves training machine learning models to predict occupancy rates and cluster parking lots.

Models Used in Your Project:

(A) Occupancy Prediction

1. **Linear Regression**:
    - Predicts occupancy rate based on features like time of day, capacity, and location.
    - Simple and interpretable but may not capture complex relationships.
2. **Random Forest Regressor**:
    - More powerful than Linear Regression.
    - Can handle non-linear relationships and interactions between features.

(B) Clustering for Parking Lot Segmentation

3. **K-Means Clustering**:
    - Groups parking lots based on capacity and occupancy rate.
    - Helps in designing targeted pricing strategies for different clusters.

(C) Dynamic Pricing Strategy

4. **Rule-Based Pricing Model**:
    - Adjusts prices dynamically based on:
        - Occupancy rate thresholds.
        - Peak hours vs. off-peak demand.
        - Location popularity.

Final Steps: Model Evaluation & Deployment

- **Model Evaluation**: Compare models using:
  - **Mean Absolute Error (MAE)**
  - **R^2 Score**
- **Model Deployment**:
  - If real-time predictions are needed, deploy the model using a **Flask API**.
  - Integrate with live ISPAK API for **dynamic pricing updates**.

Conclusion

By following this **data science workflow** project ensures a structured and data-driven approach to:

- Understanding parking lot occupancy.
- Predicting future demand.
- Implementing a **dynamic pricing model** for revenue optimization.