**README**

**Overview**

This project explores the optimization of Electric Vehicle (EV) charging station infrastructure in Connecticut using data science methodologies. It includes two primary components:

1. **Report Document** A comprehensive analysis of EV infrastructure challenges, objectives, and clustering methodologies.
2. **Jupyter Notebook (EV\_vehicle.ipynb)**: The Python-based implementation of the analysis, including data preprocessing, clustering, and visualization.

**Project Objectives**

1. Analyze the current EV charging station infrastructure in Connecticut.
2. Identify underserved regions using clustering techniques.
3. Provide actionable recommendations for infrastructure improvement.

**Structure of the Files**

**Report**

This document provides the theoretical framework and analysis:

* **Introduction**: Explains the background, problem statement, and objectives.
* **Literature Review**: Discusses relevant studies on EV infrastructure optimization.
* **Methodology**: Details preprocessing, clustering techniques (K-Means, DBSCAN, Agglomerative Hierarchical Clustering), and visualization methods.
* **Results and Discussion**: Analyzes clustering results, including insights into infrastructure gaps and recommendations.
* **Conclusion and Future Work**: Summarizes findings and suggests further research areas.

**Jupyter Notebook (EV\_vehicle.ipynb)**

The notebook contains the Python implementation:

* **Data Preprocessing**:
  + Handling missing data.
  + Geospatial data extraction and transformation.
  + Standardizing features for clustering.
* **Clustering Methods**:
  + **K-Means Clustering**: Finds optimal clusters using the Elbow Method.
  + **DBSCAN**: Identifies density-based clusters with parameter optimization.
  + **Agglomerative Hierarchical Clustering (AHC)**: Groups cities based on infrastructure characteristics.
* **Visualization**:
  + 2D PCA scatter plots to display cluster distributions.
  + Geospatial maps using Folium to visualize EV infrastructure density and gaps.
* **Statistical Analysis**:
  + Provides cluster-specific summaries for actionable insights.

**Key Findings**

1. **Cluster Characteristics**:
   * **Underserved Cities**: Minimal chargers and station counts.
   * **Well-Equipped Cities**: Balanced charger types and sufficient station density.
   * **Specialized Cities**: Heavy reliance on DC Fast Chargers for highway travel.
2. **Recommendations**:
   * Prioritize underserved areas with Level 2 and DC Fast Chargers.
   * Strengthen infrastructure in high-density regions for future EV adoption.

**Requirements**

**Dependencies**

* Python packages: pandas, matplotlib, folium, scikit-learn, kneed, seaborn, shapely.
* Dataset: Electric\_Vehicle\_Charging\_Stations.csv (/content/drive/MyDrive/EV dataset/ directory).

**Installation**

1. Install required libraries using pip:

bash

Copy code

pip install pandas matplotlib folium scikit-learn kneed seaborn shapely

1. Place the dataset in the expected directory.

**Running the Analysis**

1. Open the EV\_vehicle.ipynb file in Jupyter Notebook or Google Colab.
2. Execute the cells sequentially to preprocess data, run clustering, and visualize results.

**Future Work**

1. Use advanced machine learning models like Gaussian Mixture Models or deep learning for improved predictions.
2. Integrate renewable energy and smart grid data for sustainable infrastructure planning.
3. Expand the study globally for broader insights.

**Contact**

For questions or feedback, please contact the project author through the University of Hertfordshire.