**Project README**

**Overview**

This project processes and analyzes data, computes metrics, and generates a scatterplot over New York City boundaries. It consists of three main scripts:

1. main.py - Core data processing and analysis pipeline
2. extra\_metrics.py - Additional performance and error metrics
3. nyc\_bound.py - Generates a scatterplot within NYC geographic boundaries

**Prerequisites**

* Python 3.7 or higher
* Required Python packages (install via pip):

*packages:* pandas, numpy, matplotlib, geopandas, shapely, scikit-learn

**Project Structure**

/your-project-folder

│

├── main.py

├── extra\_metrics.py

├── nyc\_bound.py

└── readme.txt

**Usage**

1. **Run the main pipeline**

python main.py

* + This script ingests raw data, performs cleaning, modeling, and saves primary outputs.

1. **Compute extra metrics**

python extra\_metrics.py

* + Calculates additional performance metrics and error analysis based on the outputs from main.py.

1. **Generate NYC scatterplot**

python nyc\_bound.py

* + Uses GeoPandas to load NYC boundary shapefile and overlays scatter data points.
  + Saves the figure as nyc\_scatterplot.png in the working directory.

**Notes**

* Ensure that all input data files are located in the paths expected by each script (check configuration within each script).