## **README – MCAS Score Analysis**

**Description**This project investigates performance on the Massachusetts Grade 10 MCAS (Massachusetts Comprehensive Assessment System) from 2019 to 2024 at the district level. The goal is to explore trends in average scaled scores and evaluate predictors of student achievement using regression and random forest models. The analysis incorporates both test results and enrollment-based demographic indicators such as income level, English Learner status, and racial/ethnic composition.

This is a personal project developed by Maekala Turner and is intended to demonstrate her ability to integrate real-world public education data into end-to-end modeling workflows.

**Setup Instructions**

1. Open the mcas\_score\_analysis.ipynb notebook in Jupyter or Google Colab.
2. Upload or connect the raw MCAS and enrollment CSV/XLSX files to the /data files folder.
3. Ensure required packages are installed:

| !pip install pandas matplotlib seaborn scikit-learn geopandas folium |
| --- |

1. Mount Google Drive and define the project directory paths.

**Key Components of the Notebook**

* MCAS scores loading and transformation (2019–2024, Grade 10 only)
* Enrollment demographics integration (matched by district and year)
* Missing value analysis and imputation (e.g., low income %, SGP)
* Exploratory visualizations: score trends, box plots, scatter plots
* Predictive modeling using Linear Regression and Random Forest

**How to Use**

*Example 1: View top-performing districts in Grade 10 Math in 2024*

| math\_2024 = mcas\_df[(mcas\_df['year'] == 2024) & (mcas\_df['subject'] == 'math')] math\_2024.sort\_values('avg\_scaled\_score', ascending=False).head() |
| --- |

*Example 2: Visualize average score vs. income percent*

| sns.scatterplot(data=model\_df, x="lowincome\_pct", y="avg\_scaled\_score") |
| --- |

*Example 3: Fit and evaluate a Random Forest model*

| rf = RandomForestRegressor() rf.fit(X\_train, y\_train) preds = rf.predict(X\_test) |
| --- |

**Current Status**

* All MCAS and enrollment data loaded and merged
* Data cleaning, EDA, and regression modeling complete
* Random Forest model fit on extended feature set
* Final dataset includes over 2,700 matched district-year-subject rows

**Next Steps**

* Continue refining models with additional contextual features
* Explore spatial/geographic insights using Folium or Tableau
* Finalize reporting with dashboards or presentation visuals