Impact of COVID-19 Vaccination Rates on School Attendance in Connecticut

Tony Min (mint@rpi.edu), **Rensselaer Polytechnic Institute, Troy, NY, United States**

Abstract & Motivation

The COVID-19 pandemic has brought about unprecedented challenges in various sectors around the world, one of which including education. As schools grappled with implementing safety measures to stop the spread of the virus, one significant move was the mandates of masks. At the same time, vaccine companies aimed to mitigate the impact of the pandemic, which offered an avenue of going back to normal within schools. This study is to investigate the correlation between school attendance rates and vaccination rates in the state of Connecticut during the COVID-19 pandemic.

The motivation for this study stems from the observation of the dynamic changes in the world of education during the COVID-19 pandemic. With schools implementing protocols to keep individuals safe, it becomes crucial to understand the potential correlation between these measures and attendance rates. Specifically, the focus was on exploring whether the vaccination rates in the state of Connecticut had any discernible influence on school attendance.

Datasets

Dataset 1: COVID-19 Vaccination by Town and Race/Ethnicity

This dataset was sorted out by Towns in Connecticut and was further sorted by Race/Ethnicity. The purpose of this dataset was to understand the vaccination landscape at the local level.

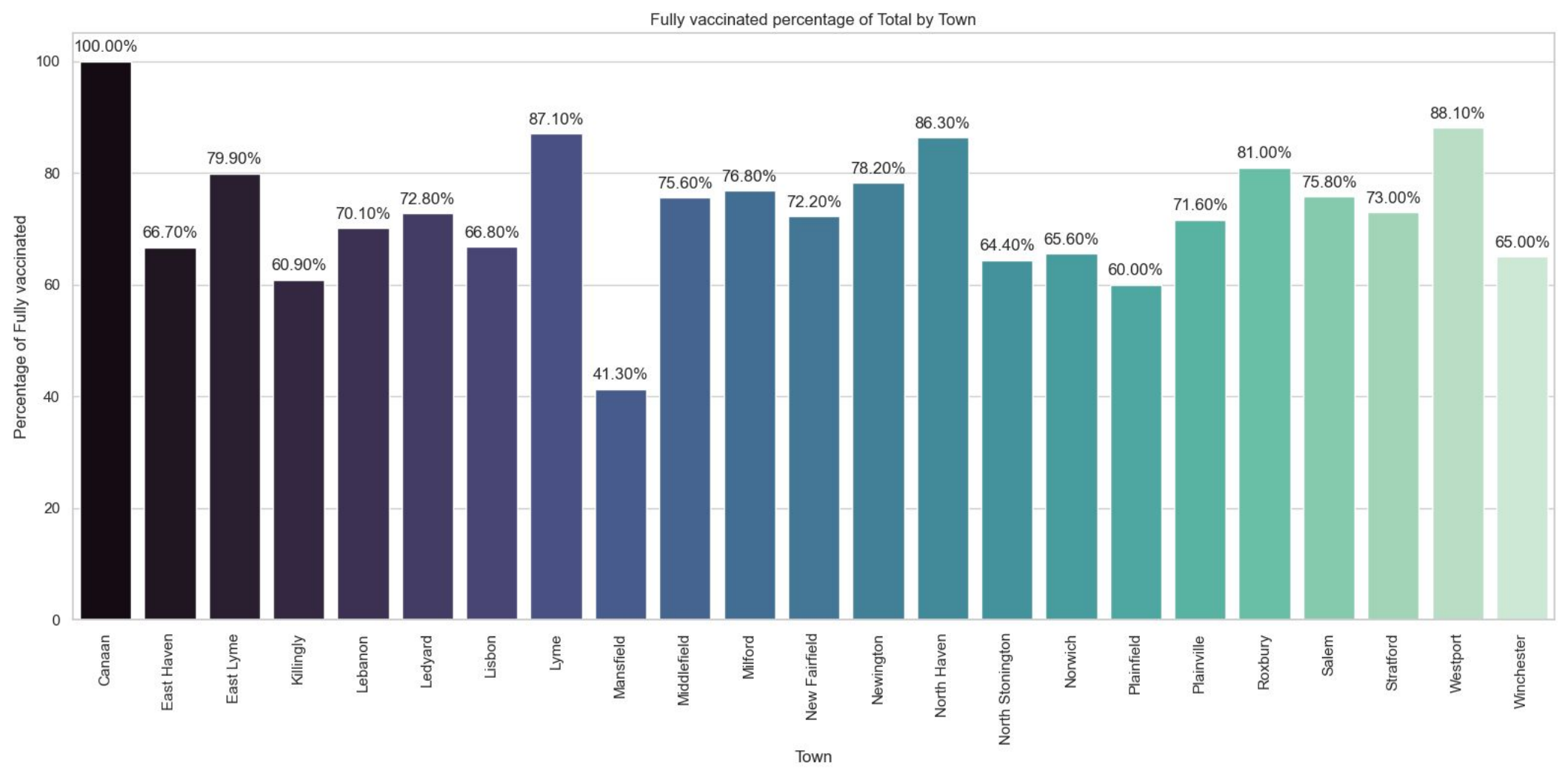
Dataset 2: School Attendance by Student Group and District, 2021-2022

This data set focused on school attendance rates during the 2021-2022 school year which was sorted out by school district and various student groups.

Column name	Description of data
Town name	Name of Town in CT
Vaccination status	“No Vaccine”, “At least one dose”, “Fully Vaccinated”
Race/ethnicity	Domestic Groups (White, Black, Hispanic, …)
Value	Vaccination percentage
Student Group	Same as Race/ethnicity
2020-2021 attendance rate	Percent of students attending school in the 2020-2021 school year
2019-2020 attendance rate	Percent of students attending school in the 2019-2020 school year
2021-2022 attendance rate	Percent of students attending school in the 2021-2022 school year (target feature)

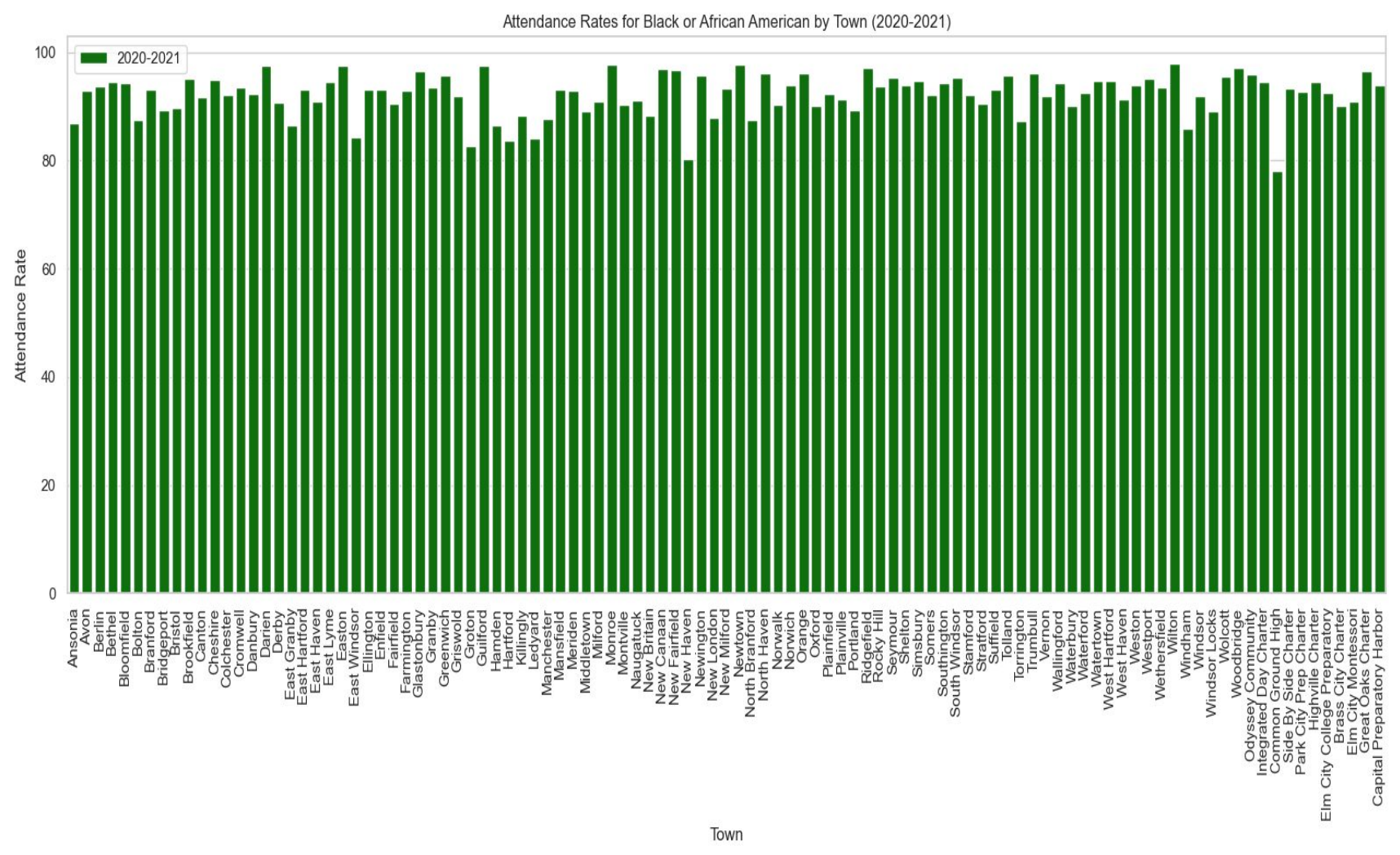
Exploratory Data Analysis

Below are some of the exploratory plots created to look at the data and the distribution of it.



Histogram of Fully Vaccinated percentage of 25 sample towns in CT

Histogram of Attendance Rates for Black or African Americans by Town in the 2020-2021 school year



It is important to understand the distribution of data points within the state of CT to get a better understanding of each of the towns. With it seemingly looks like a higher vaccination percentage you may have a high attendance rate as well. We also used the feature to feature engineer another option for Vaccine status and calculated the percent of the specific population that has not be vaccinated at all.

To clean up the data, I dropped unnecessary columns from the original separate datasets, examples of are, “District code”, “Reporting period”, and “Date update”. Another with dropping columns we also removed null values within datasets in order for our models to run on a full set of data.

Models

Five models were used: Linear Regression, Random Forest Regressor, Decision Tree Regressor, XGBoost regression, and Support Vector Regressor

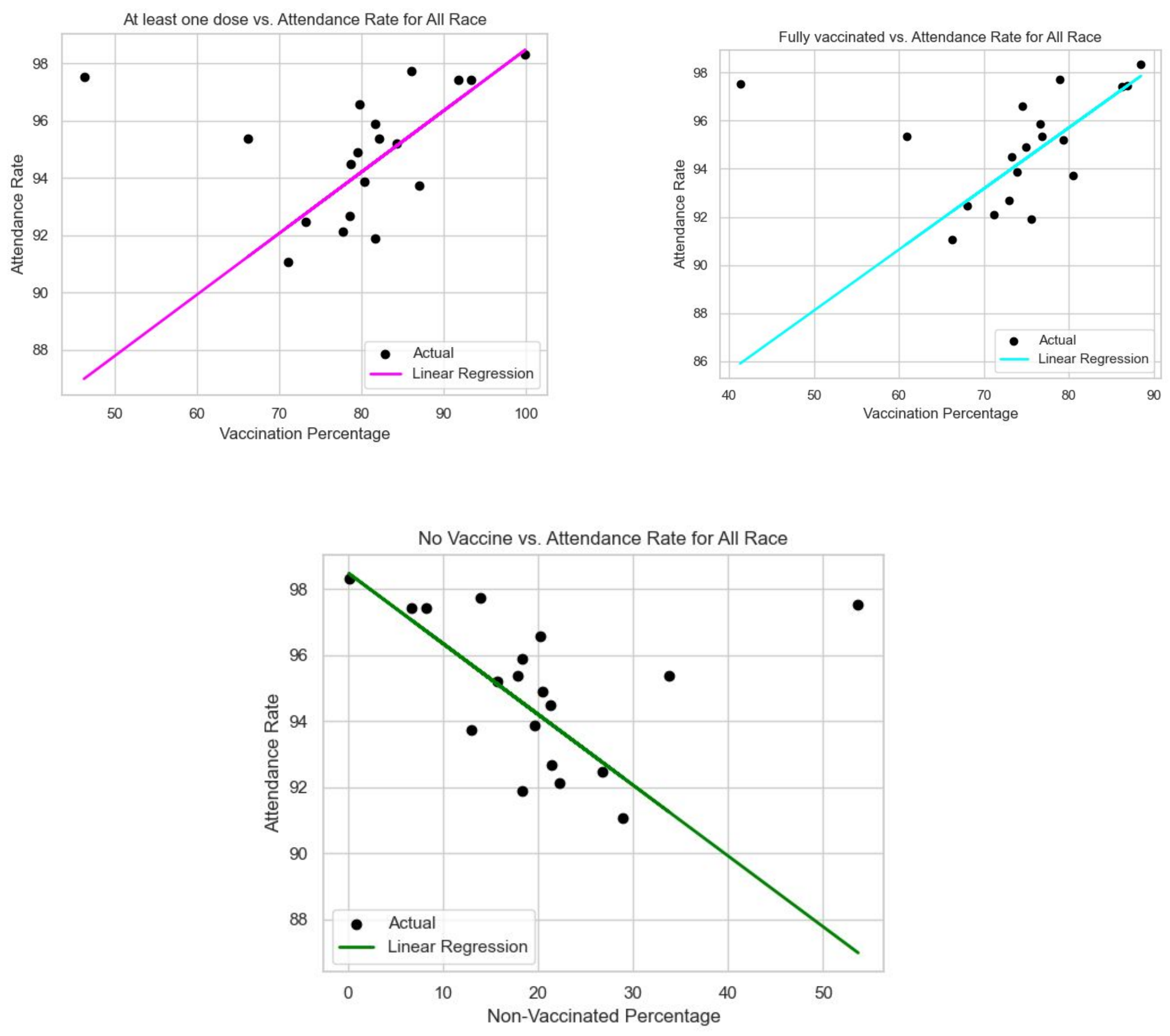
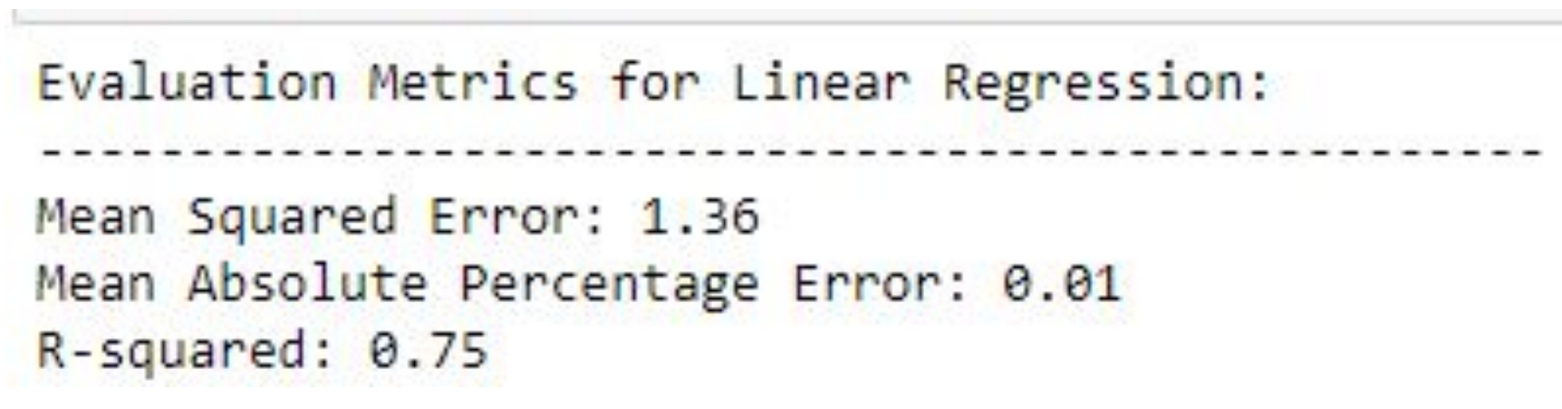


Figure 1: Linear Regression

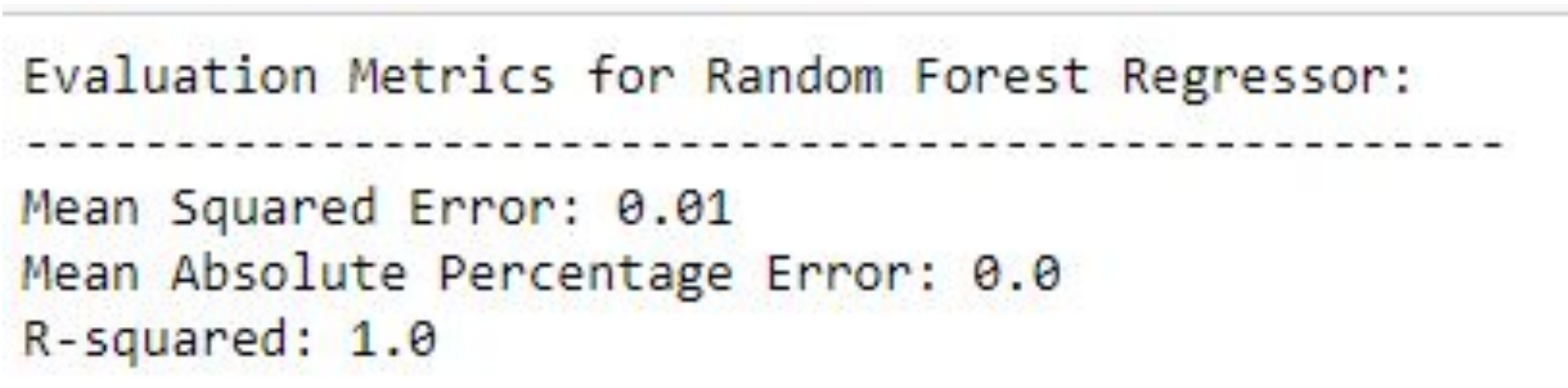
Model 1: Linear Regression

We see a positive correlation between vaccination percentage and attendance rate and then a negative correlation between non-vaccinated percentage and attendance rates as shown in figure 1.



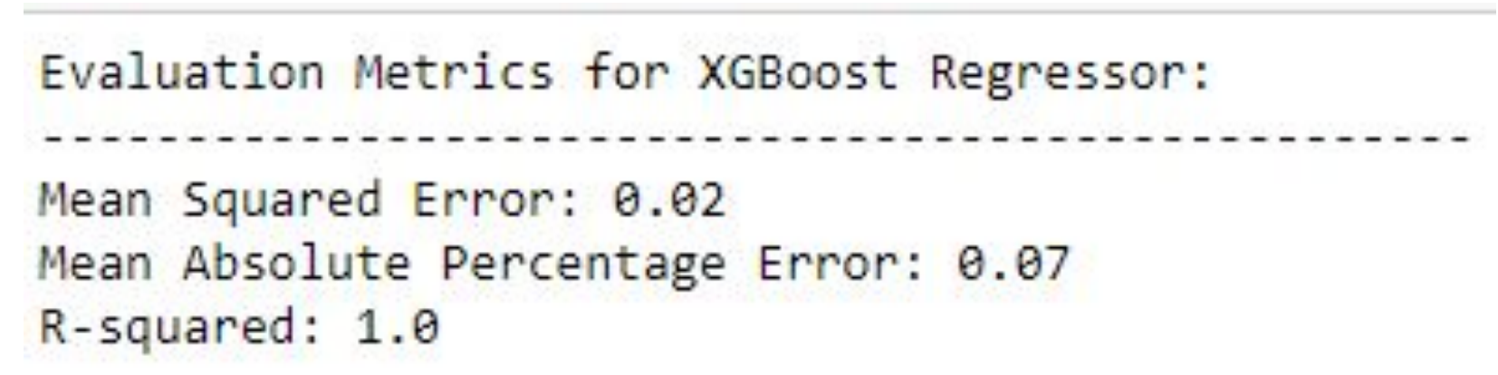
```
Evaluation Metrics for Linear Regression:
------------------------------------------------
Mean Squared Error: 1.36
Mean Absolute Percentage Error: 0.01
R-squared: 0.75
```

Model 2: Random Forest Regressor



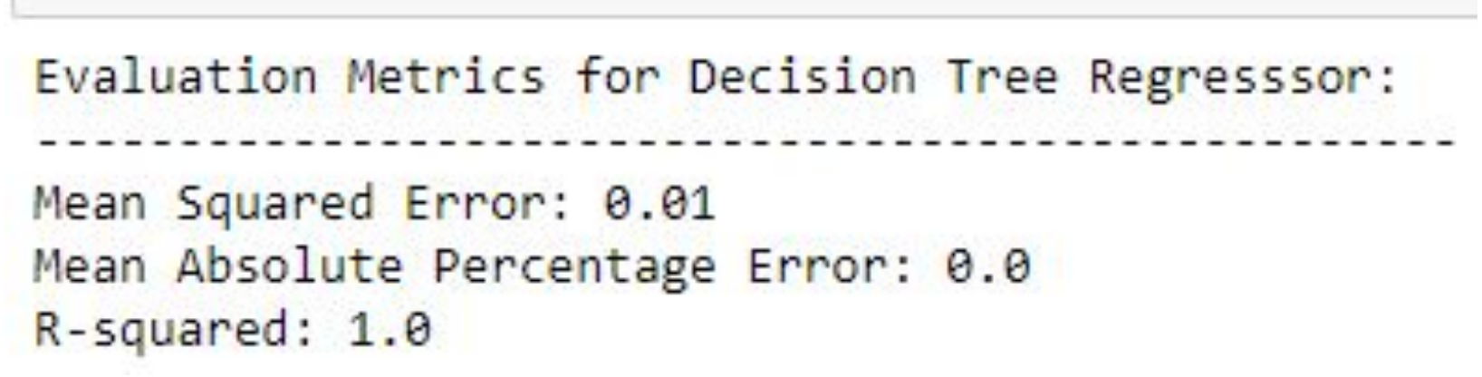
```
Evaluation Metrics for Random Forest Regressor:
--------------------------------------------------
Mean Squared Error: 0.01
Mean Absolute Percentage Error: 0.0
R-squared: 1.0
```

Model 4: XGBoost Regressor



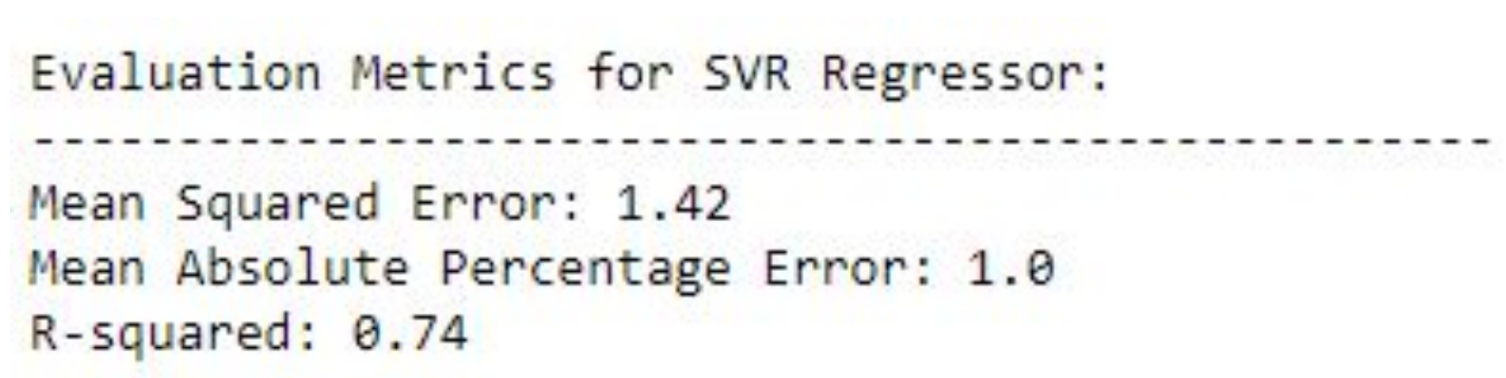
```
Evaluation Metrics for XGBoost Regressor:
------------------------------------------------
Mean Squared Error: 0.02
Mean Absolute Percentage Error: 0.07
R-squared: 1.0
```

Model 3: Decision Tree Regressor



```
Evaluation Metrics for Decision Tree Regressor:
------------------------------------------------
Mean Squared Error: 0.01
Mean Absolute Percentage Error: 0.0
R-squared: 1.0
```

Model 5: Support Vector Regressor



```
Evaluation Metrics for SVR Regressor:
------------------------------------------------
Mean Squared Error: 1.42
Mean Absolute Percentage Error: 1.0
R-squared: 0.74
```

Results

The 3 models achieved a perfect R-Squared score of 1.0 on our original merged dataset. Which demonstrated the ability to capture underlying patterns in the data and make accurate predictions. These 3 models, Random Forest, Decision Tree and XGBoost, are all tree based as well with Random Forest and XGBoost being ensemble learning methods. Also from our models we learned that for fully vaccinated individuals showed negative R-Squared scores meaning there are challenges in prediction attendance based on vaccination status. Another thing learned from the models is that it showed that the 3 tree based models had positive R-Squared score for White individuals indicating a stronger ability to predict for attendance rate for this group.

Glossary:

Ensemble learning Method – machine learning techniques that combine several base models in order to produce on optimal predictive model .

XGBoost - Extreme Gradient Boosting

Resources:

1. Pandas: https://pandas.pydata.org/getting_started.html
2. Scikit learn: https://scikit-learn.org/stable/install.html
3. Matplotlib: https://matplotlib.org/stable/tutorials/index
4. Dataset1: https://catalog.data.gov/dataset/covid-19-vaccination-by-town-and-race-ethnicity
5. Dataset2: https://catalog.data.gov/dataset/school-attendance-by-student-group-and-district-2021-2022
6. Jupyter Notebook: https://jupyter.org/
7. Seaborn: https://seaborn.pydata.org/installing.html
8.