**Link:** <https://samyfel.github.io/DS4200-FinalProject/index.html>

**\*\*Note:** Due to the static nature of github page hosting, the dynamic rendering of Flask cannot be hosted on there. To get the final project with model prediction, please run the following commands in Terminal:

git clone https://github.com/samyfel/DS4200-FinalProject.git

cd DS4200-FinalProject/flask

python app.py

**Figure 1 (Standard Deviation):** The goal with this visualization is to see what kind of information loss were to happen when removing outliers using standard deviations. To do this, a Kernel Density Estimator (KDE) plot was used to visualize the distribution.

* **Marks:**
  + **Density Curve (Area Plot):** The main distribution of the exam scores as a smoothed (KDE) plot
  + **Dashed Vertical Lines:** Represent the bounds of data used (3 standard deviations from the mean)
  + **Text Annotation:** Labels showing the exact values of bounds
* **Channels:**
  + **X-axis (Exam-Score):** Represents the exam score range
  + **Y-axis (Density):** Represents the approximate density of the scores relative to the scores within the dataset
  + **Color:** Red and blue, for the lower and upper bound respectively.

**Figure 2 (Distribution of Categorical Variables):** The goal with this visualization is to see how the distribution changes based on categorical variables. This can give us insights into which features we should select to train our model. To do this, a faceted KDE plot was used where each category's KDE distribution was plotted against each other.

* **Marks:**
  + **Density Curve (Area Plot):** The category distribution of the exam scores as a smoothed (KDE) plot
* **Channels:**
  + **X-axis (Exam-Score):** Represents the exam score range
  + **Y-axis (Density):** Represents the approximate density of the scores relative to the scores within the subgroup distribution
  + **Color:** To designate between the distributions of different categories

**Figure 3 (Scatterplot of Continuous Features with Low Unique Values):** The goal with this visualization was to show why these continuous features should be treated as categorical. To do this, a faceted scatterplot was used.

* **Marks:**
  + **Scatter Points:** Individual dots that represent data points
  + **Trend Line:** A fitted linear regression line on each subplot to indicate the general trend between variables
* **Channels:**
  + **X-axis (*Sleep Hours, Tutoring Sessions, and Physical Activity*):** The feature of study
  + **Y-axis (Exam Score):** The score obtained

**Figure 4 (Continuous Features):** The goal with this visualization is to show the relationship between continuous features and exam score. To do this, a faceted scatterplot was again used.

* **Marks:**
  + **Scatter Points:** Individual dots that represent data points
* **Channels:**
  + **X-axis (*Hours Studied, Attendance, and Previous Scores*):** The feature of study
  + **Y-axis (Exam Score):** The score obtained

**Figure 5 (Real vs Predicted Values):** The goal with this visualization was to understand the performance of the linear regression model, where expected and predicted values were plotted against each other in a scatterplot.

* **Marks:**
  + **Scatter Points:** The expected value against the predicted value
  + **Expected Trend:** The line that represents an ideal fit (perfect prediction)
* **Channels:**
  + **X-axis:** The expected values
  + **Y-axis (Exam Score):** The predicted values by the linear regression model.

**Figure 6 (Sleep Hours, Hours Studied vs Exam Scores):** The goal of this visaluzation was to understand the relationship between Sleep Hours and Hours Studied vs Exam Scores through a side by side Scatterplot.

* **Marks:**
  + **Scatter Points:** The number of either sleep hours or hours studied. They were also color coded based on Teacher Quality.
* **Channels:**
  + **X-axis:** The feature of the study
  + **Y-axis (Exam Score):** The score that the student received on the exam.

**Figure 6 (Attendence vs Exam Scores):** The goal of this visaluzation was to understand the relationship between Attendence and Exam Scores that can be segmented by access to resources .

* **Marks:**
  + **Scatter Points:** The number of classes attended thata are also color coded based on access to resources
* **Channels:**
  + **X-axis:** The feature of the study
  + **Y-axis (Exam Score):** The score that the student received on the exam.

**Figure 7 (Tutoring Hours vs Exam Scores):** The goal of this visaluzation was to understand the relationship between Tutoring Hoursand Exam Scores.

* **Marks:**
  + **Scatter Points:** The number of Tutoring Hours that the Student recieved
* **Channels:**
  + **X-axis:** The feature of the study (the number of tutoring hours)
  + **Y-axis (Exam Score):** The score that the student received on the exam.