

IRS-ML Fall 2019



WinaGodwin A., Suma C., and Samuel T.

Team Members

- WinaGodwin Anyanwu
 - 4th Year Computer Science Major
 - Programming Experience: Java, C, Android, Python, SQL, C#
- Suma Cherkadi
 - Second Year Computer Science Major
 - Programming Experience: Java, Python/Sci-Kit Learn, Scala, Javascript
- Samuel Toh
 - Third Year Computer Science Major
 - Programming Experience: Java, Python, Django, SQL

Introduction

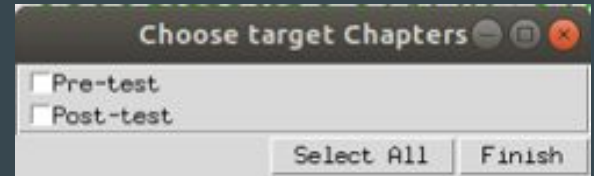
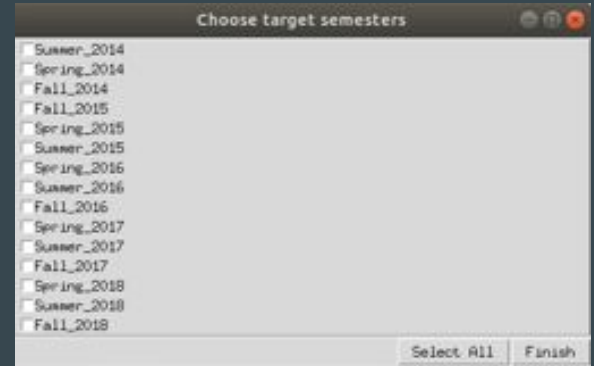
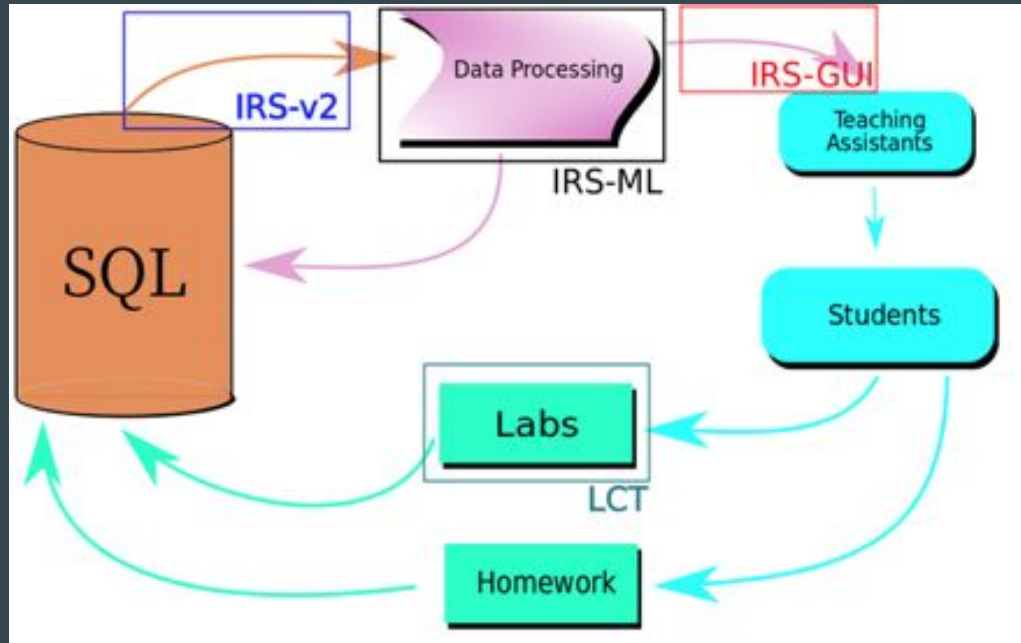
What is IRS?

- The Intelligent Response System aims to create a user interface that leverages the the ITS database to provide feedback to improve student performance.

What does IRS-ML do?

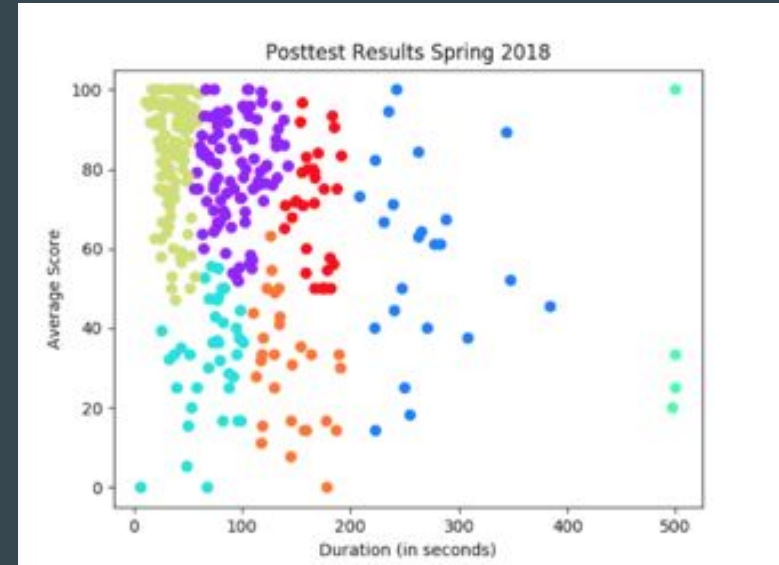
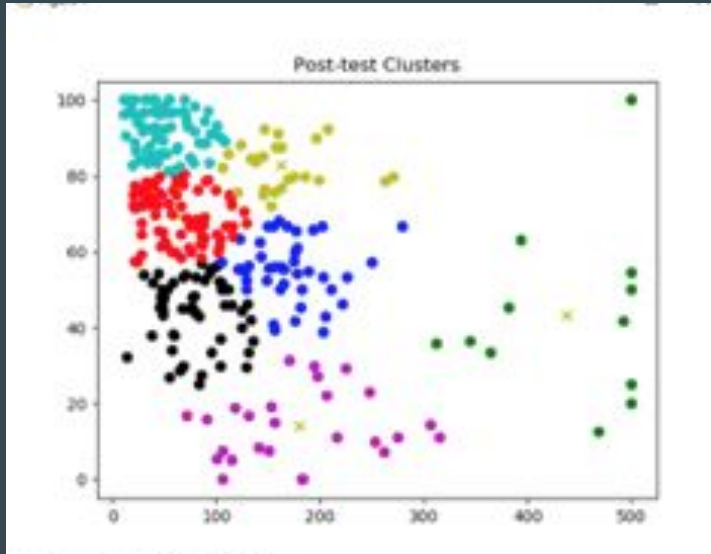
- IRS-ML aims to perform data analytics and use Machine Learning algorithms to provide the personalized feedback.

Last Semester



Began processing data and improved data selection implementation

Last Semester



Performed K-Means and Agglomerative Clustering Algorithms

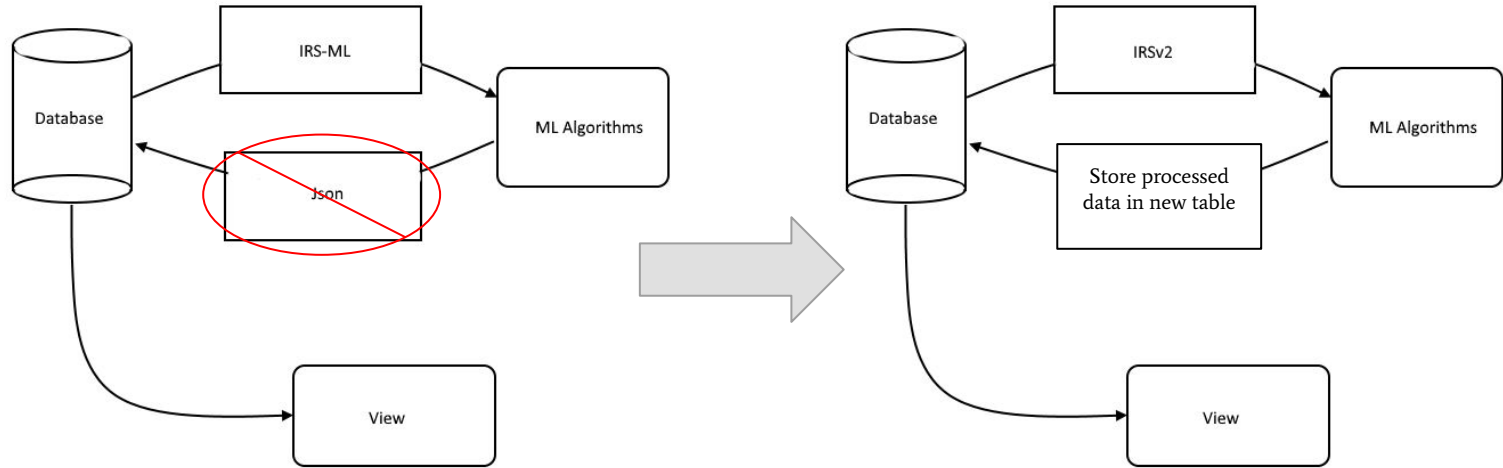
This Semester - Goals

- Write previous semesters work back into SQL database
- Analyze data to provide personalized feedback for students
- Analyze data using predictive algorithms

Project Details - Writing Back

- Writing back previous semester's work to the database in tables
- Can create tables of averaged data for different combinations of data (semester averages, yearly averages, global averages, etc.)
- Eliminates intermediate step of generating JSON files

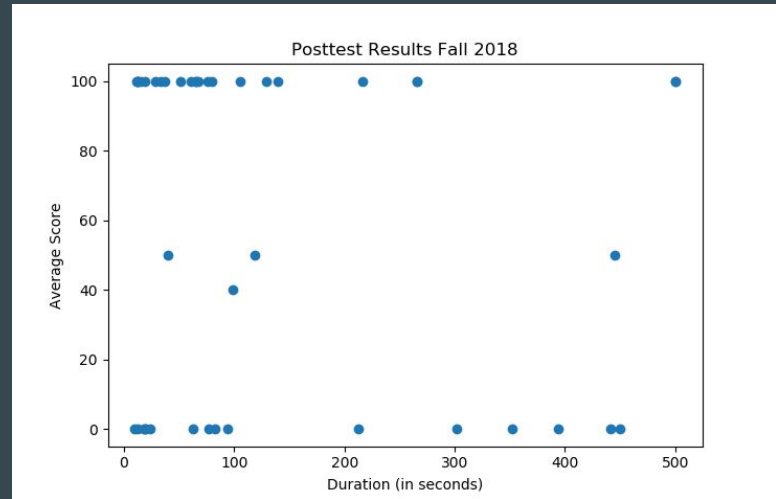
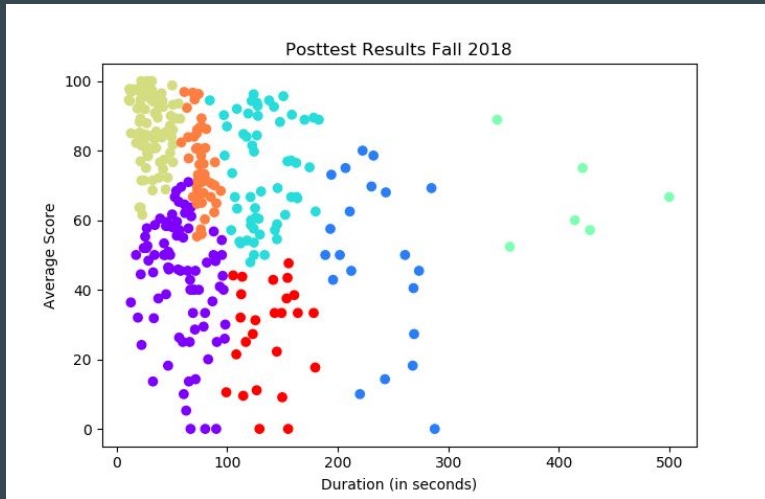
Results - Writing Back

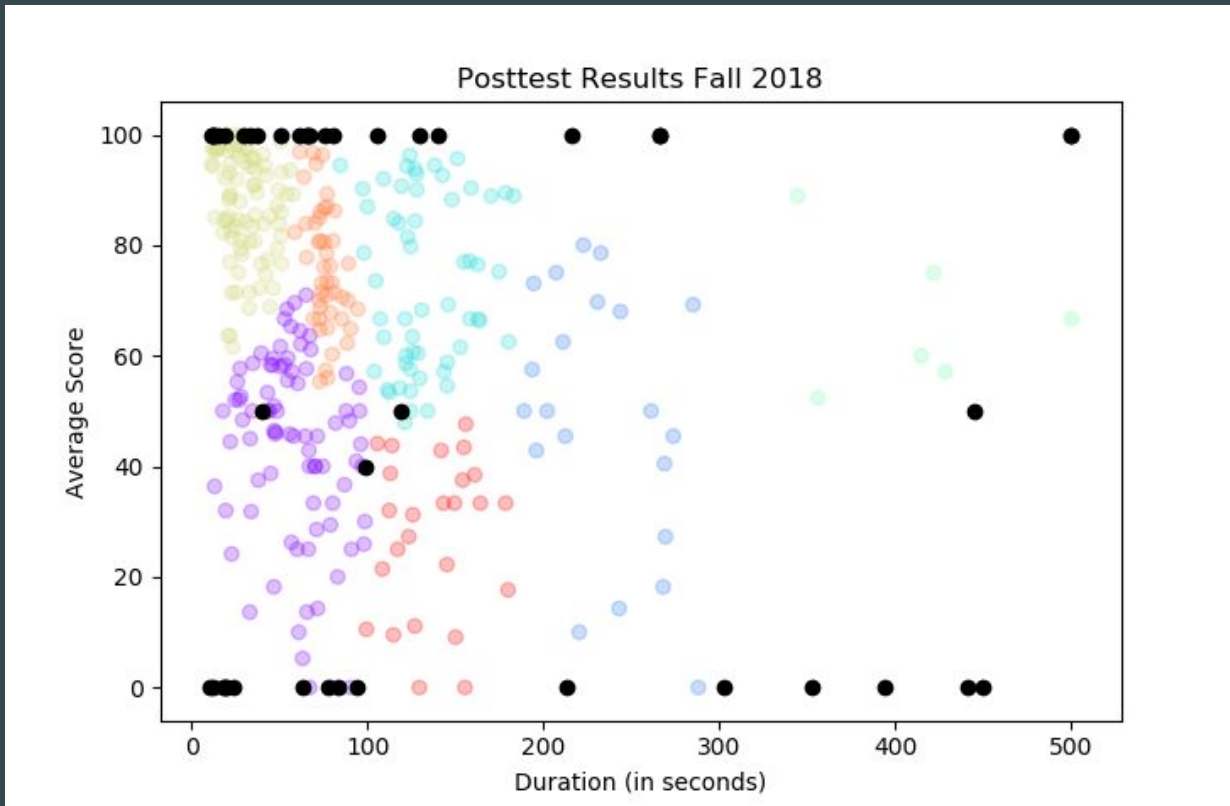


Personalized Feedback - Initial Approach

Goal: Create predictions based on data to provide personalized feedback for user

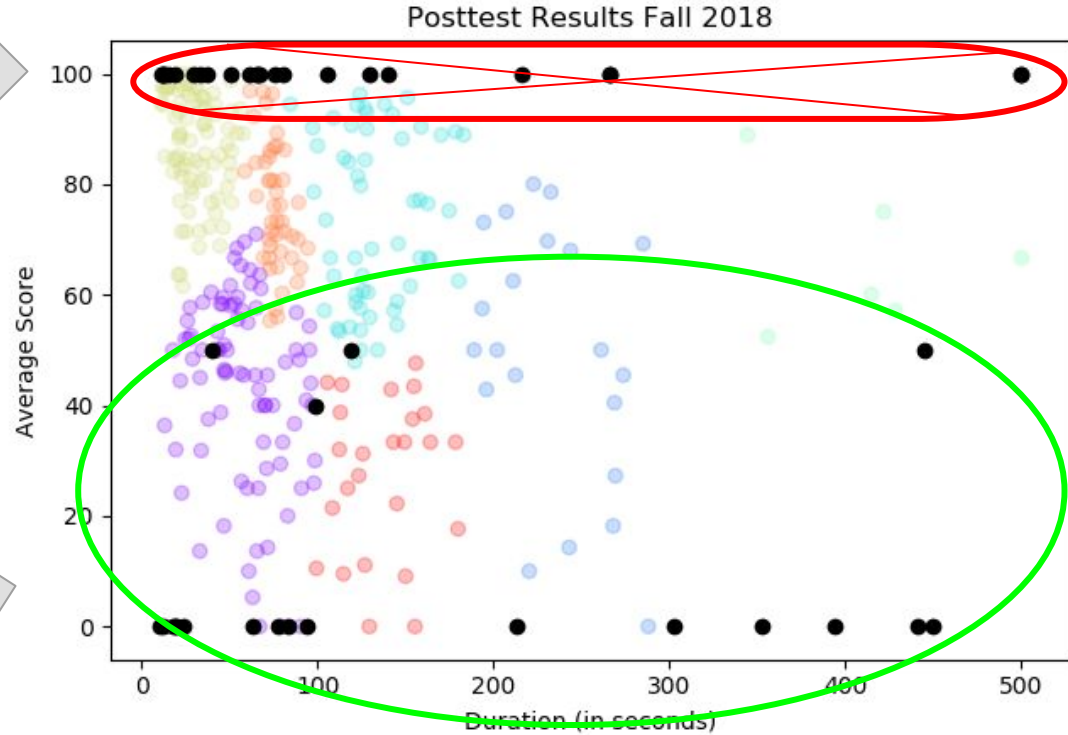
Approach: Recreate Clusters for Fall 2018 Data and Plot Against Student





Challenge: Although the superimposed image gives interesting insight, a user is not able to act upon what this image tells them.

Questions the student did well on



Questions the student should study

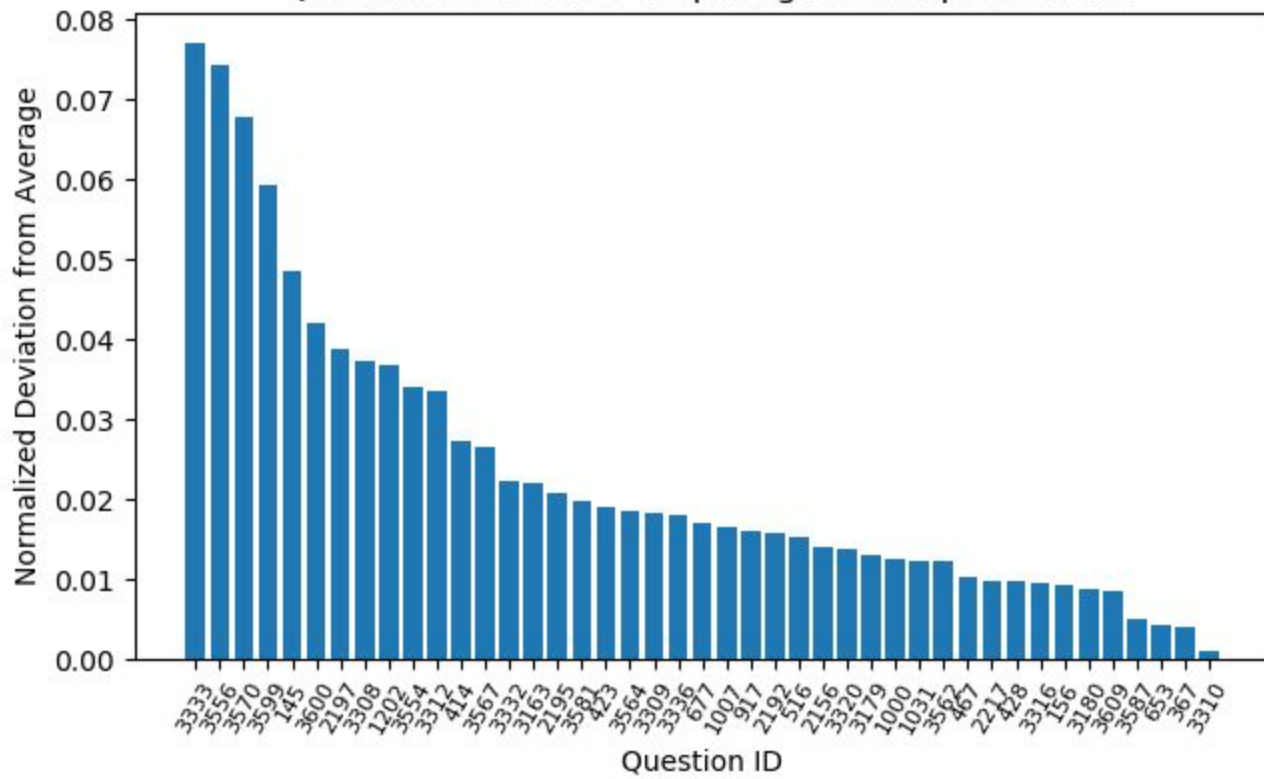
Challenge: Although the superimposed image gives interesting insight, a user is not able to act upon what this image tells them.

Personalized Feedback - Results

Final Image:

- Sorting the questions the user got wrong to provide them with a study scheme.
- Provides student with specific question id ranked in order of most important to study.
- The goal from studying these questions is to make the students reach/exceed the average performance.
- The questions are sorted using euclidean distance between class averages and individual performance
- $Distance = \sqrt{(Avg_Score - Actual_Score)^2 + (Avg_Duration - Actual_Duration)^2}$

Questions Sorted for Requiring Most Improvement



Personalized Feedback - Final Step

Use the averaged data directly from the database and eliminate the intermediate JSON file step.

```
(base) suma@suma-VirtualBox:~/ITS/IRS-ML/src$ python main.py
Enter the name of the database to be accessed: itsDB
Enter your SQL password: sumacherkadi
Rating data is insufficient
Cannot create rating json due to insufficient data
(base) suma@suma-VirtualBox:~/ITS/IRS-ML/src$ python main2.py
Enter the name of the database to be accessed: itsDB
Enter your SQL password: sumacherkadi
Rating data is insufficient
Cannot create rating json due to insufficient data
(base) suma@suma-VirtualBox:~/ITS/IRS-ML/src$ python plotPerformanceDifference.
py
```



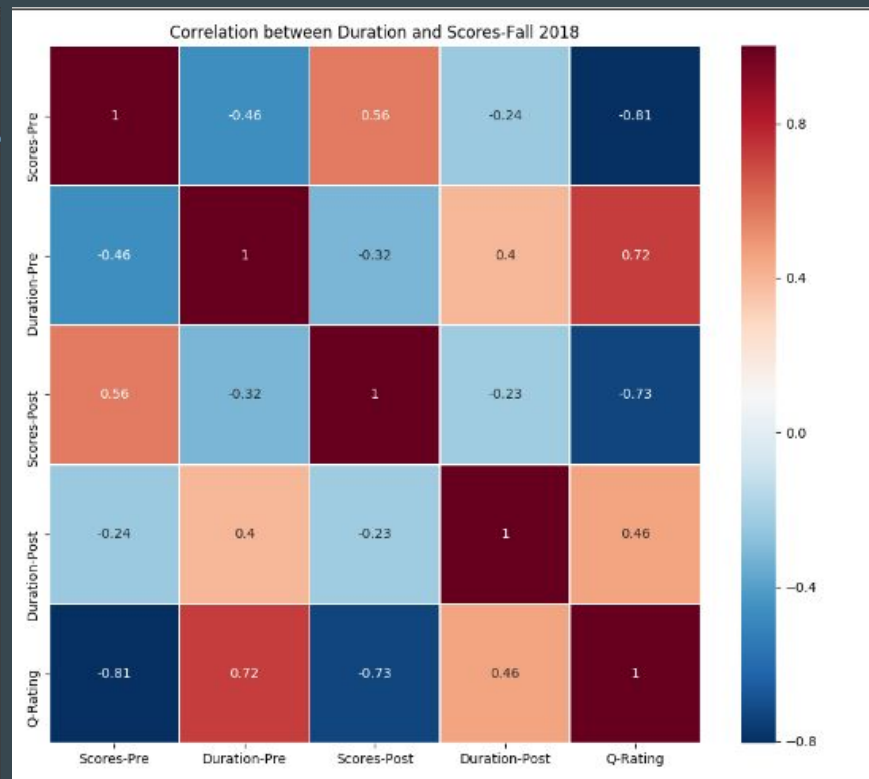
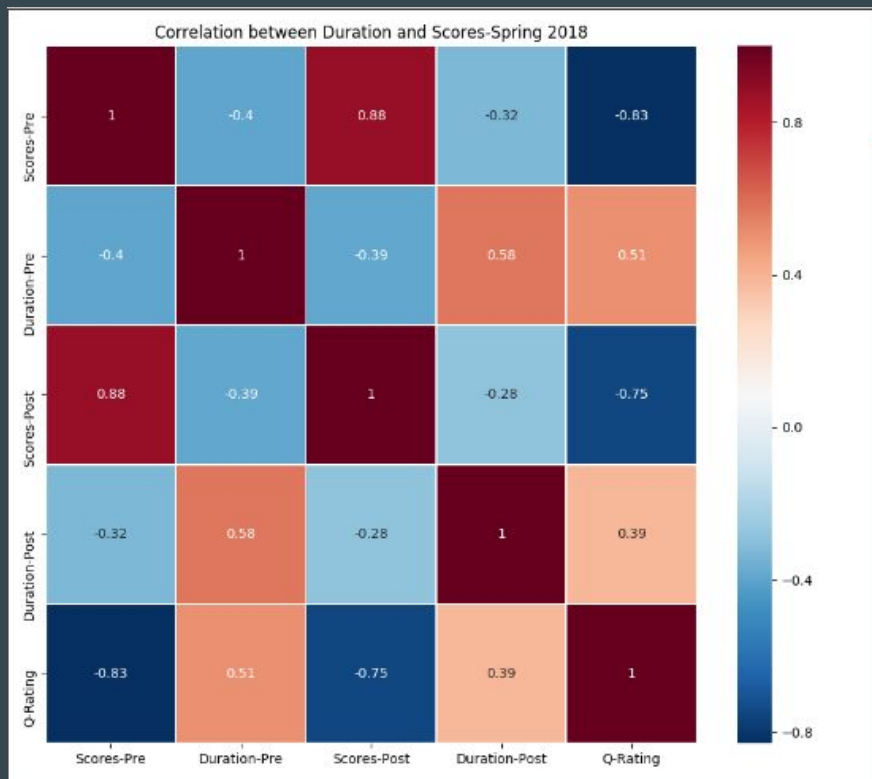
```
(base) suma@suma-VirtualBox:~/ITS/IRS-ML/src$ python plotPerformanceDifferenceW
ithDB2.py
Enter the name of the database to be accessed: itsDB
Enter your SQL password: sumacherkadi
```

Project Details - Predictions Using Correlation

- Correlation between average score, duration and rating for each question
- Predict students' future performance based on the correlation of their past performance using scores and durations taken to complete the tests.
- Replacing missing data with np.nan

```
cols = ["Scores-Pre", "Duration-Pre", "Scores-Post", "Duration-Post", "Q-Rating"]  
df[cols] = df[cols].replace({0: np.nan, 1: np.nan})
```

Results - Correlation Matrix





Biggest Challenges

- Installing Ubuntu
- Choosing exactly what to predict
- Displaying analysis in user-friendly manner
- Getting data directly from database
- Getting useful data that doesn't skew results

Next Steps

- Writeback:
 - Store plots in database as well
 - Figure out how to get the information from plots back to users
- Personalized FeedBack:
 - Make graph interactive to reduce crowding
 - Sort tags of questions to provide student with topics to study
 - Generalize algorithm to apply for any student (add selection menu for student and for data averages)
- Prediction:
 - Create predictions of performance based on correlation matrix