Syracuse University: Masters of Science in Applied Data Science

Portfolio Milestone

By: Andrew Morcos

SUID: 625203065

https://github.com/AMorcos8/MADSPortfolio




Introduction

In Syracuse University's Masters of Applied Data Science program, students gain the ability to collect, validate, analyze data and thus develop meaningful insights from multiple data sources.

Various projects were created in courses which display the skills that were developed including but not limited to:

- IST 659: Database Administration and Database Management
- IST 687: Introduction to Data Science
- MBC 638: Data Analysis and Decision Making

The Applied Data Science Program has seven Learning Objectives:

1. Describe a broad overview of the major practice areas in data science.
2. Collect and organize data.
3. Identify patterns in data via visualization, statistical analysis, and data mining.
4. Develop alternative strategies based on the data.
5. Develop a plan of action to implement the business decisions derived from the analyses.
6. Demonstrate communication skills regarding data and its analysis for relevant professionals in their organization.
7. Synthesize the ethical dimensions of data science practice.

IST 659 Database Administration and Database Management

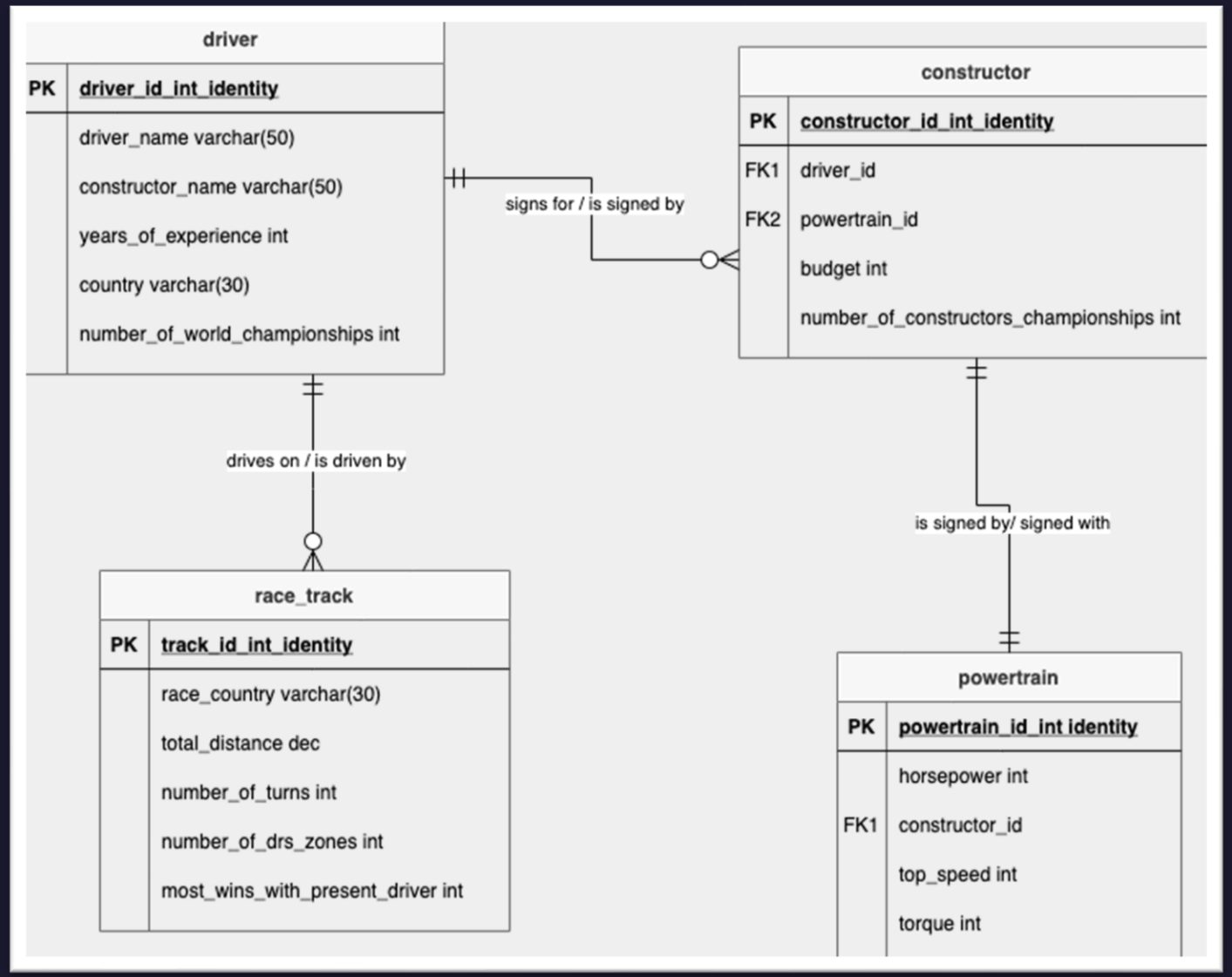
Formula 1 Database

IST 659: Database Administration and Database Management

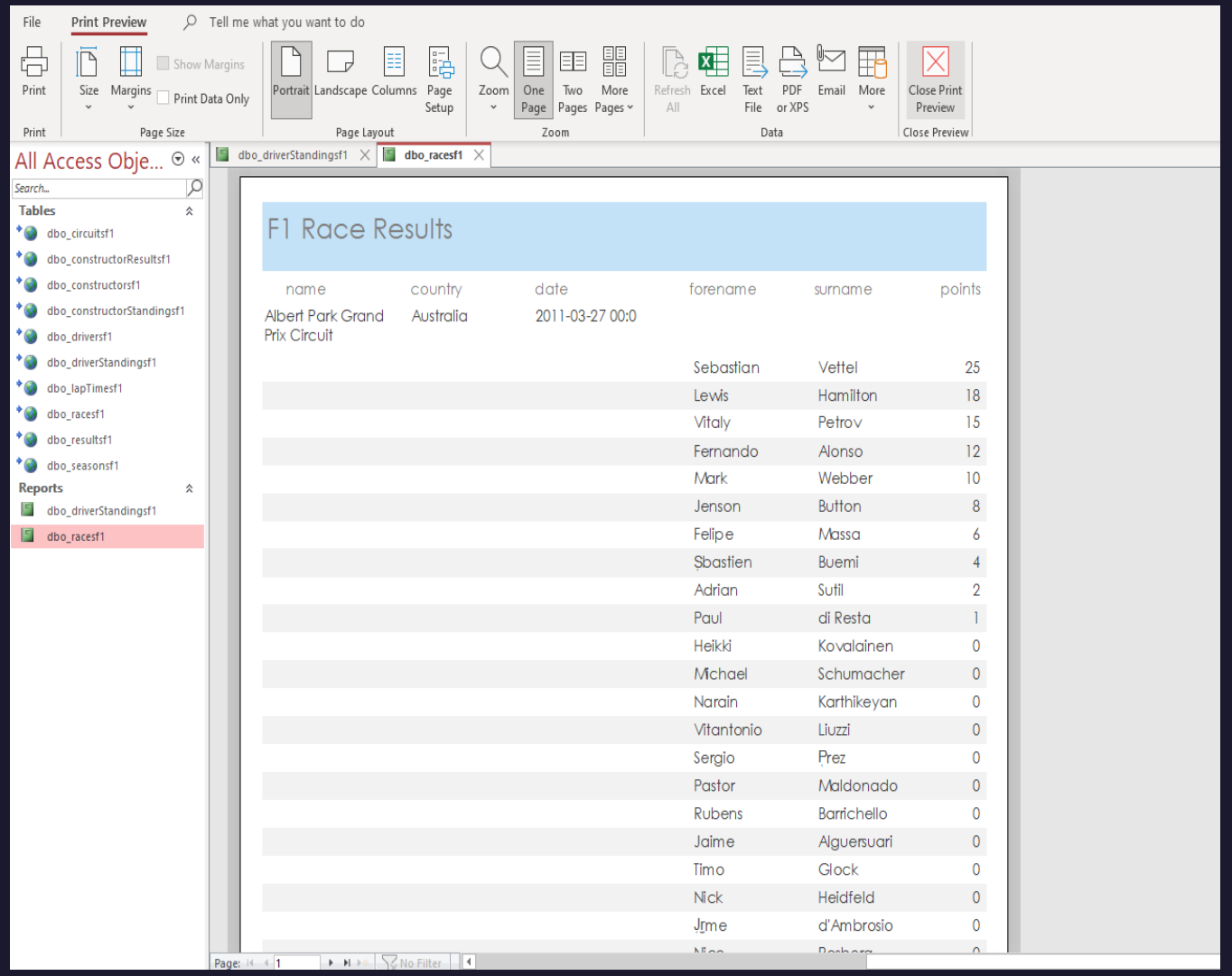
Introduction:

- The final project that was developed was a dataset from Kaggle, which was the Formula 1 dataset. This was a large dataset spanning over 50 years and included race results, season results, race times and many more, which was good exposure for data cleaning and organization. Using the skills that were developed with the music database, the Formula 1 database was developed extensively and later could be used for analysis. The conceptual, logical and physical models were created using SQL Server Management Studio while the population of data was performed in Microsoft Access.

Physical Model



Formula 1 Race Results in 2011



IST 659: Database Administration and Database Management

Reflection:

- This specific project contributed to the ability to find and deliver insights in data analysis by giving the framework to that data.
- The learning objectives achieved through this project and this course was by giving the ability to collect data, manage data and data mine all while identifying patterns via statistical analysis. These skills were then leveraged to further answer business problems later in the program.

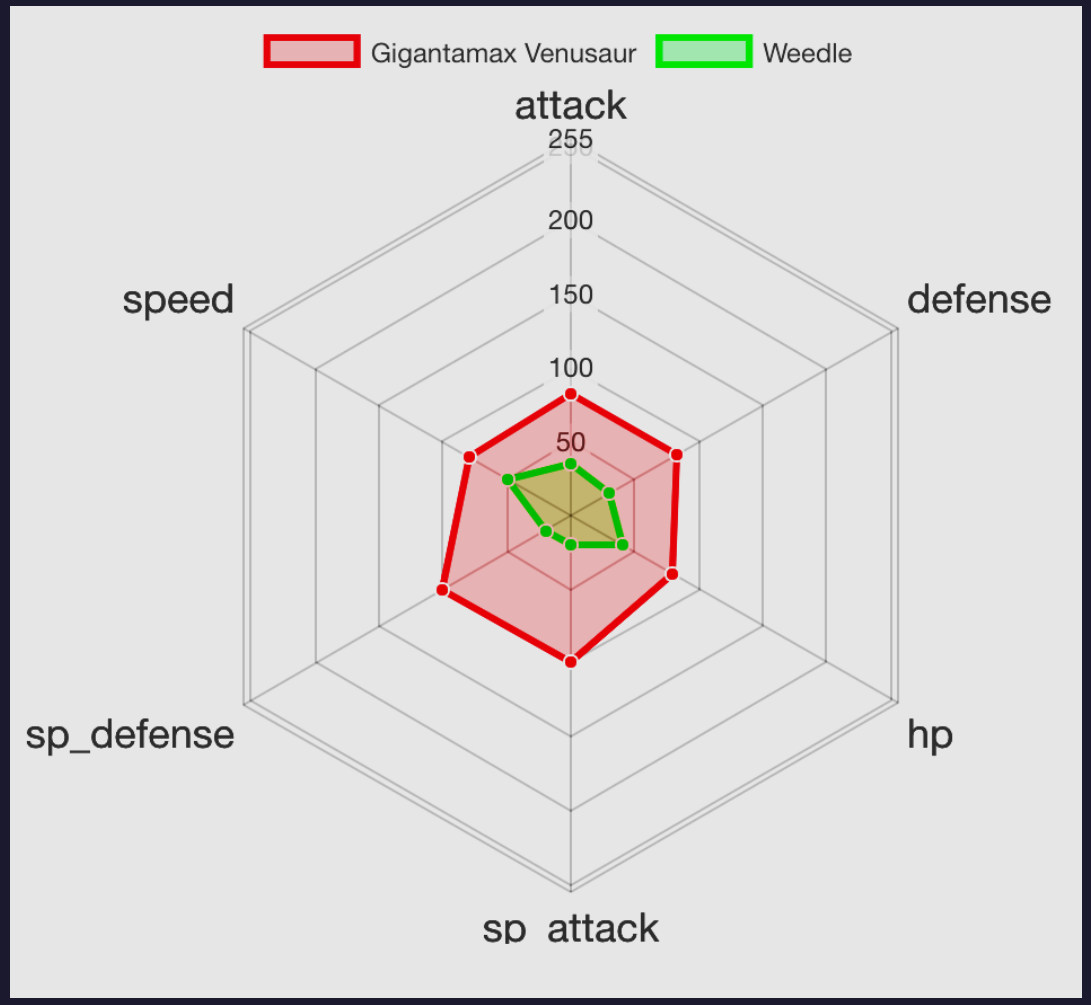
IST 678 Introduction to Data Science

Pokémon Matchup Analysis

IST 687: Introduction to Data Science

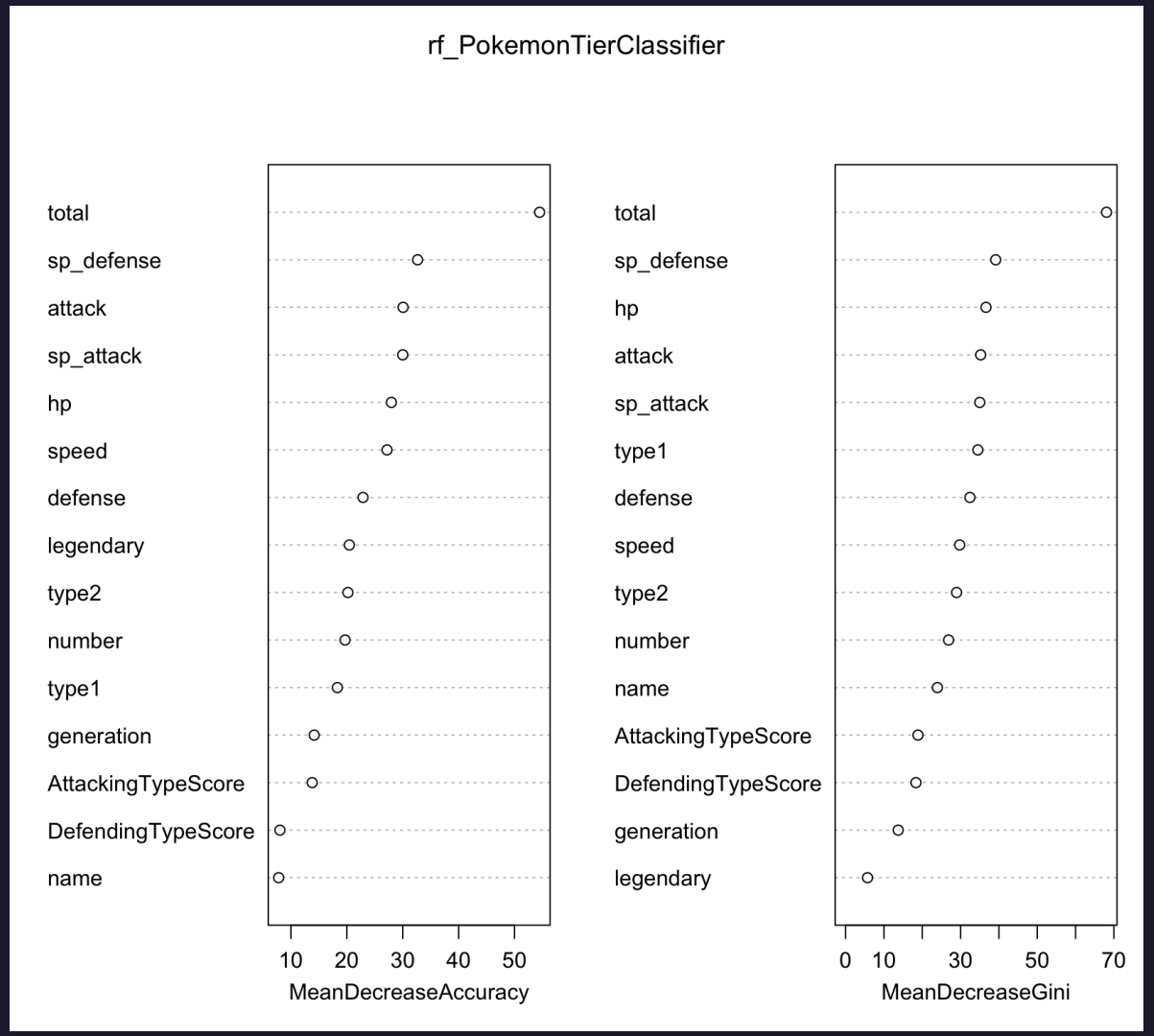
Introduction:

- In this course under the guidance of Dr. Santerre, students were tasked with choosing a dataset and developing an analysis using R.
- Throughout the course of the semester, students learned to use various tools in R to leverage various analytical tools to answer a question.
- This course provided a hands-on introductory experience to data science. The concepts explored were statistical analysis, information visualization, text mining and machine learning.
- The final project was a team project in which R was leveraged for statistical analysis and data visualization by looking at a massive Pokémon dataset. Machine learning was used in predicting head-to-head matchups between Pokémon. There are various attributes and factors that contribute to the outcome of a battle and the team wanted to tackle those factors and attributes and provided various visualizations while a Random Forest was used for machine learning.


Gigantamax Venusaur
Weedle
attack
255
200
150
100
50
defense
hp
sp_attack
sp_defense
speed

```
Call:
 randomForest(formula = tier ~ ., data = train, ntree = 1000,      mtry = 2, importance = TRUE)
               Type of random forest: classification
                     Number of trees: 1000
No. of variables tried at each split: 2

        OOB estimate of  error rate: 38.31%
Confusion matrix:
   1  2  3 4 5  6   7 class.error
1 13  4  1 0 0  2   2   0.4090909
2  1 16 13 5 0  8   7   0.6800000
3  1 12 19 3 4 17   6   0.6935484
4  0  3 10 3 3 19   0   0.9210526
5  0  0  9 3 1 31   5   0.9795918
6  1  0  5 1 4 86  29   0.3174603
7  2  1  0 0 0 19 234   0.0859375
```



IST 687: Introduction to Data Science

Reflection:

- Introduction to Data Science was the perfect way to dive into data science. The skills gained throughout the semester provided the perfect framework to analyze a large dataset.
- Ultimately, the project was successful in fulfilling learning objectives.
- The learning objectives achieved through this project were identifying patterns in data via visualization, statistical analysis, and data mining as well as developing strategies based on that data.
- The random forest displayed the skills for developing strategies because it gave an avenue to further explore different matchups besides attributes such as generation and Pokémon type.

MBC 638 Data Analysis and Decision Making

Business Process Improvement:
Reduction of Body Fat Percentage

MBC 638: Data Analysis and Decision Making

Introduction:

- This course, offered by Syracuse University Whitman School of Management, had a different approach than the other “heavy-coding” courses. Majority of this course was performed using Microsoft Excel.
- The project goal is to find a business problem and provide action insights on that problem. The specific business problem was the “Reduction of Body Fat Percentage” and using statistical analysis to measure and gauge any changes needed to improve the business problem.
- The identification of the business problem had to be defined first by presenting a “Process Map”, Operation definitions, identifying the data and the data collection which was personal data. Sigma Quality Level and the identification of error were used to measure the data and its accuracy.
- Finally, Simple Linear Regression and Moving Ranges were used to analyze the data and later to improve the problem by developing a new Sigma Quality Level and Hypothesis test. All of these were implemented in a high level Process Improvement Storyboard

MBC638

Data Analysis and Decision Making

Reduction of Body Fat Percentage

Process owner: Andrew Morcos

Key Dates ---> Team Launch: 03/30	Define: 03/30	Measure: 04/06-06/01	Analyze: 06/01	Improve: 06/08	Control: 06/09

DEFINE

Purpose: Decrease Body Fat Percentage

Impact:
- Prevent health risks
- Better fit of clothing
- Elimination of processed foods

Goal: Reaching the body fat percentage range of 14-17%

Defect: Deviation from diet causing bloating and increase in waist circumference, calorie intake and decrease in Net Calorie intake.

MEASURE

Type of Data : Continuous and Discreet

Collection: different body measurements to calculate body fat percentage and net caloric intake

Output (y): body fat percentage

Input (x): hours spent Exercising, Net Calorie Intake and Resting Heart Rate

SQL = 3.4

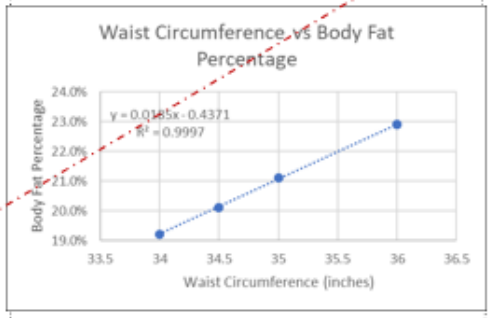
Calorie Intake	Calroies Burned	Net Calorie Intake	BodyFat %
1800	438	1362	22.9%
2430	619	1811	22.9%
1800	571	1229	22.9%
2157	560	1597	22.9%
1929	402	1527	22.9%

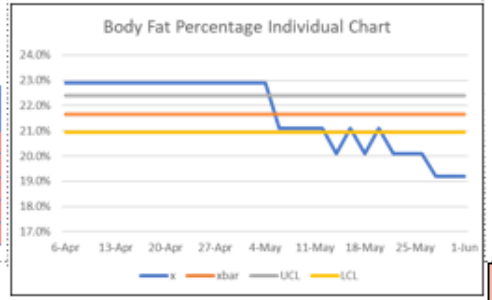
Baseline for output (y):
Mean = 21.6%
SD = 1.4%

ANALYZE

Most correlated with BF%:
Waist Circumference; r= 0.99
Net Calorie Intake; r= 0.48

	waist circumference (in)	Weight (lbs)	Calorie Intake	Net Calorie Intake	BodyFat %
waist circumference (in)	1				
Weight (lbs)	0.091435723	1			
Calorie Intake	0.324786278	0.277708593	1		
Net Calorie Intake	0.486809622	0.276089214	0.879394	1	
BodyFat %	0.999884584	0.086351889	0.339808	0.483643	1





IMPROVE

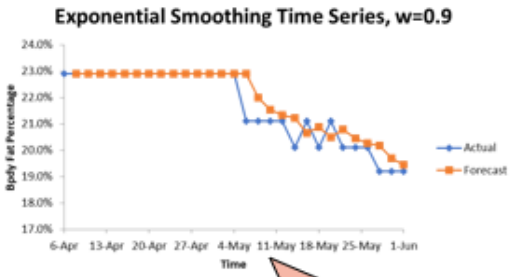
Keep monitoring caloric intake

SQL = 3.9

Process Improvement

Hypothesis test came back inconclusive with the allotted time

CONTROL



Time Series can be used to determine when goal BF% can be reached

Team Members: Myself

MBC 638: Data Analysis and Decision Making

Reflection:

- There were critical learning objectives achieved that were imperative to the learning process in this program. All aspects of statistical knowledge were challenged in this project that provided the tools needed for a student's career.
- Through the heavy statistical analysis performed and the development of a plan of action to implement the business decisions derived from the analyses, this course was successful in fulfilling the learning objectives and providing students with the statistical, analytical and critical thinking needed to answer business questions.

Conclusion

- This comprehensive portfolio successfully displays the achievement of each learning objective and proper practices in the Master's of Applied Data Science program.
- Data was collected, organized and mined with the ultimate execution of data visualization and statistical analysis across all courses by using various tools and skills such as Microsoft Excel, Microsoft Access, SQL Server Management Studio and R
- The primary focus on predictive analysis equips students with the skills to become an asset for various businesses when looking at the job market.
- The various skills acquired in the Applied Data Science program develops students to be a well-rounded Data Scientist/Data Engineer and ultimately fulfill business needs and answer business problems.

Thank You