Portfolio Milestone Class of 2022

Andrew Morcos
SUID: 625203065
https://github.com/AMorcos8/MADSPortfolio

Table of Contents

Andrew Morcos
SUID: 625203065
https://github.com/AMorcos8/MADSPortfolio

Introduction

Throughout the coursework at Syracuse University's Master of Science in Applied Data Science, students are equipped with valuable hard and soft skills to be a well rounded in the Data Science field. Students gain the ability to collect, validate, analyze data and thus develop meaningful insights from multiple data sources. Courses such as Database Administration & Database Management (IST 659), Introduction to Data Science (IST 687), and Data Analysis & Decision Making (MBC 638). Various tools were used in these courses that are imperative to the development of understanding in the Data Science field such as Microsoft Access, SQL Server Management Studio, Python, R, Excel, and Microsoft Visual Studio. The Master's of Science in Applied Data Science program has seven main learning objectives that displayed what outcomes a student should fulfill by the end of the program; various project from the former courses have been successful at addressing these 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.

Andrew Morcos
SUID: 625203065
https://github.com/AMorcos8/MADSPortfolio

IST 659: Database Administration & Database Management

Project Description

Throughout the semester, Dr. Chad Harper had the class develop a database for music streaming. This music database contained information such as the artist, song and the metadata of the songs. Each week, a different layer was added to the database to develop an understanding of how to build a database on SQL Server starting with the foundation. Part of the foundational understanding include the development of conceptual and logical models, which help organize the relationships of various data points to one another.

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 **(Figure 1)** while the population of data was performed in Microsoft Access **(Figure 2)**.

Andrew Morcos
SUID: 625203065
https://github.com/AMorcos8/MADSPortfolio

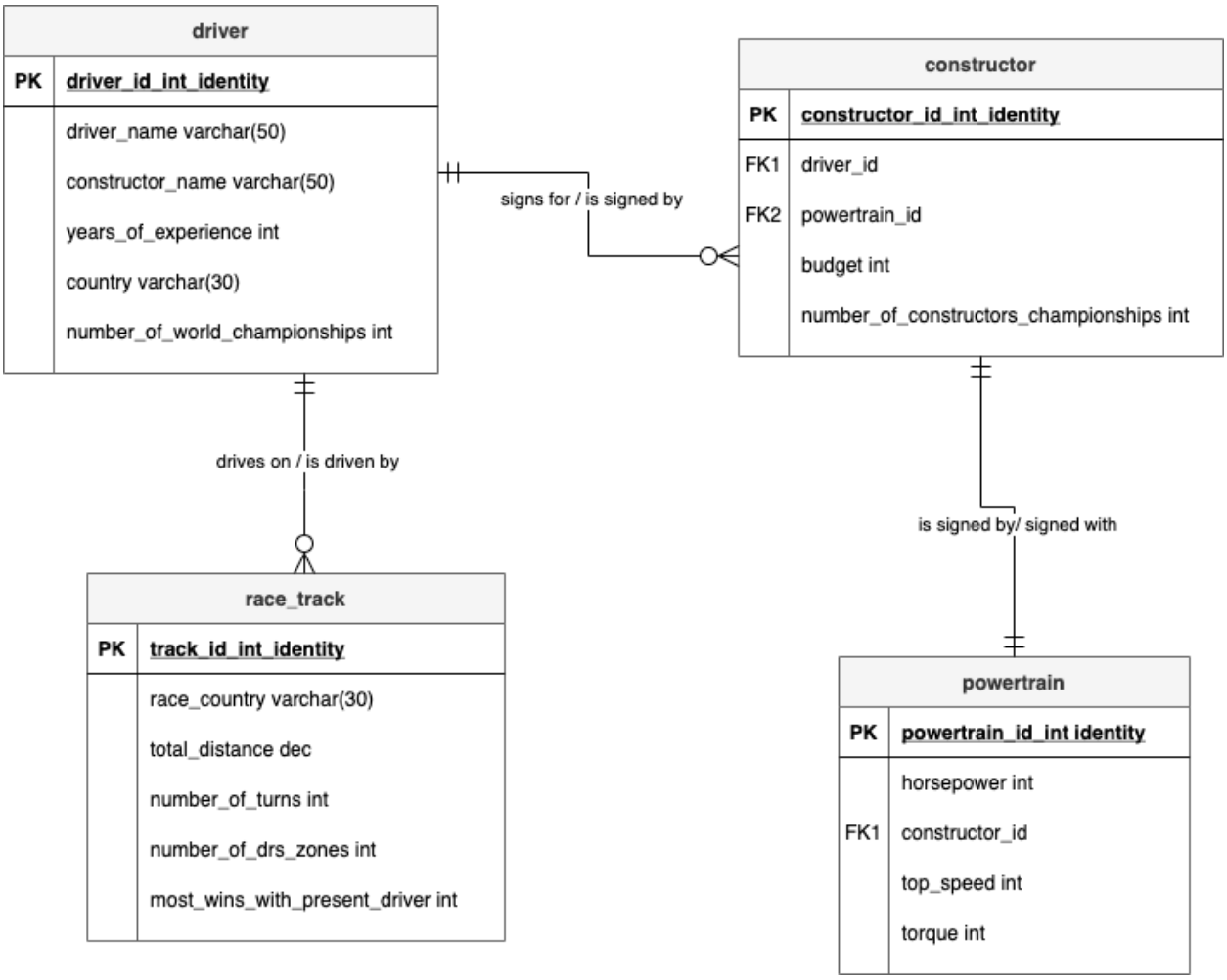


Figure 1: Physical Model (Morcos, IST 659, 2021)

Andrew Morcos
SUID: 625203065
https://github.com/AMorcos8/MADSPortfolio

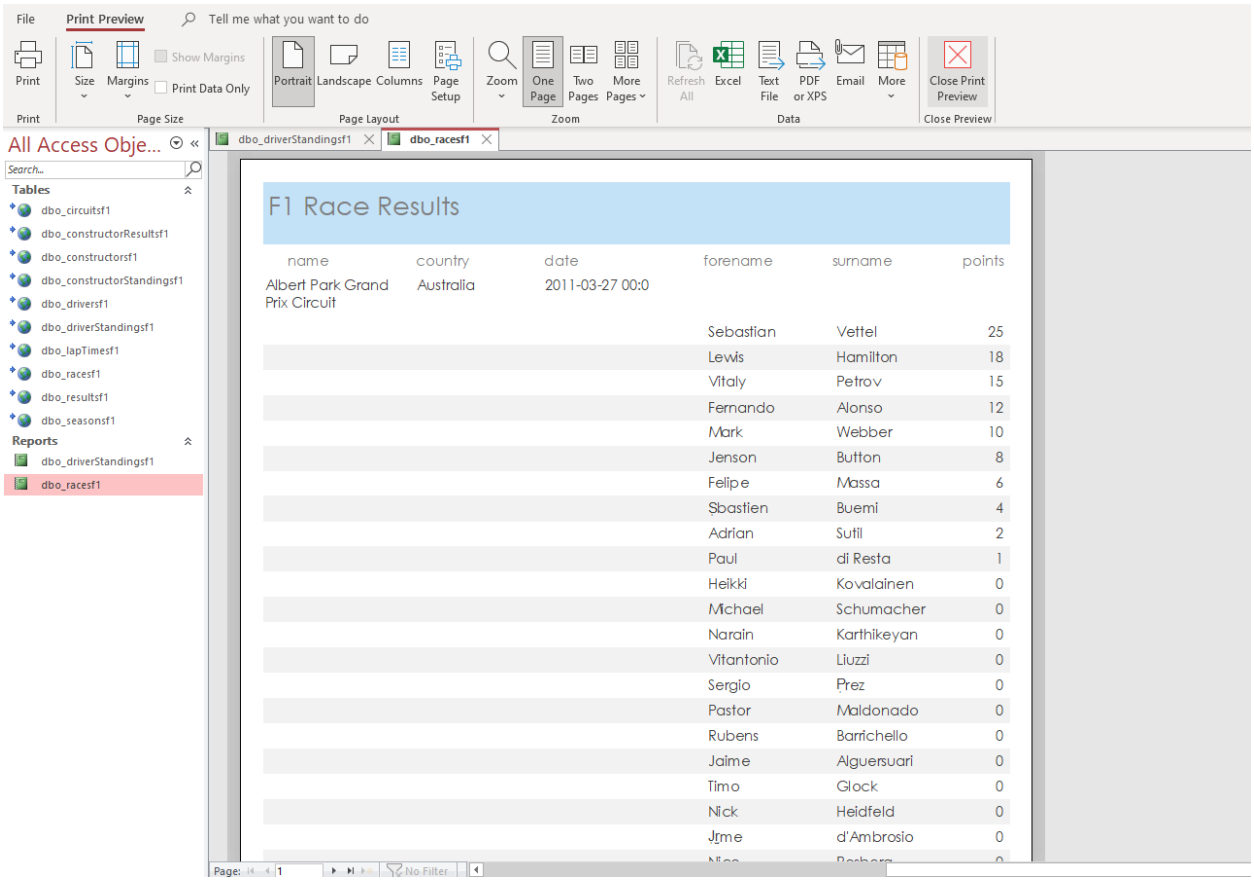


Figure 2: 2011 F1 Race Results (Morcos, IST 659, 2021)

Andrew Morcos
SUID: 625203065
https://github.com/AMorcos8/MADSPortfolio

Reflection & Learning Objectives

The Data Administration and Database Management course laid down the foundational concepts of database management and building relational databases from the ground up. Courses later on in the program such as Data Warehousing, helped dive deep into data administration and data management by attacking business problems and coming up with solutions. This specific project contributed to the ability to find and deliver insights in data analysis by giving the framework to that data.

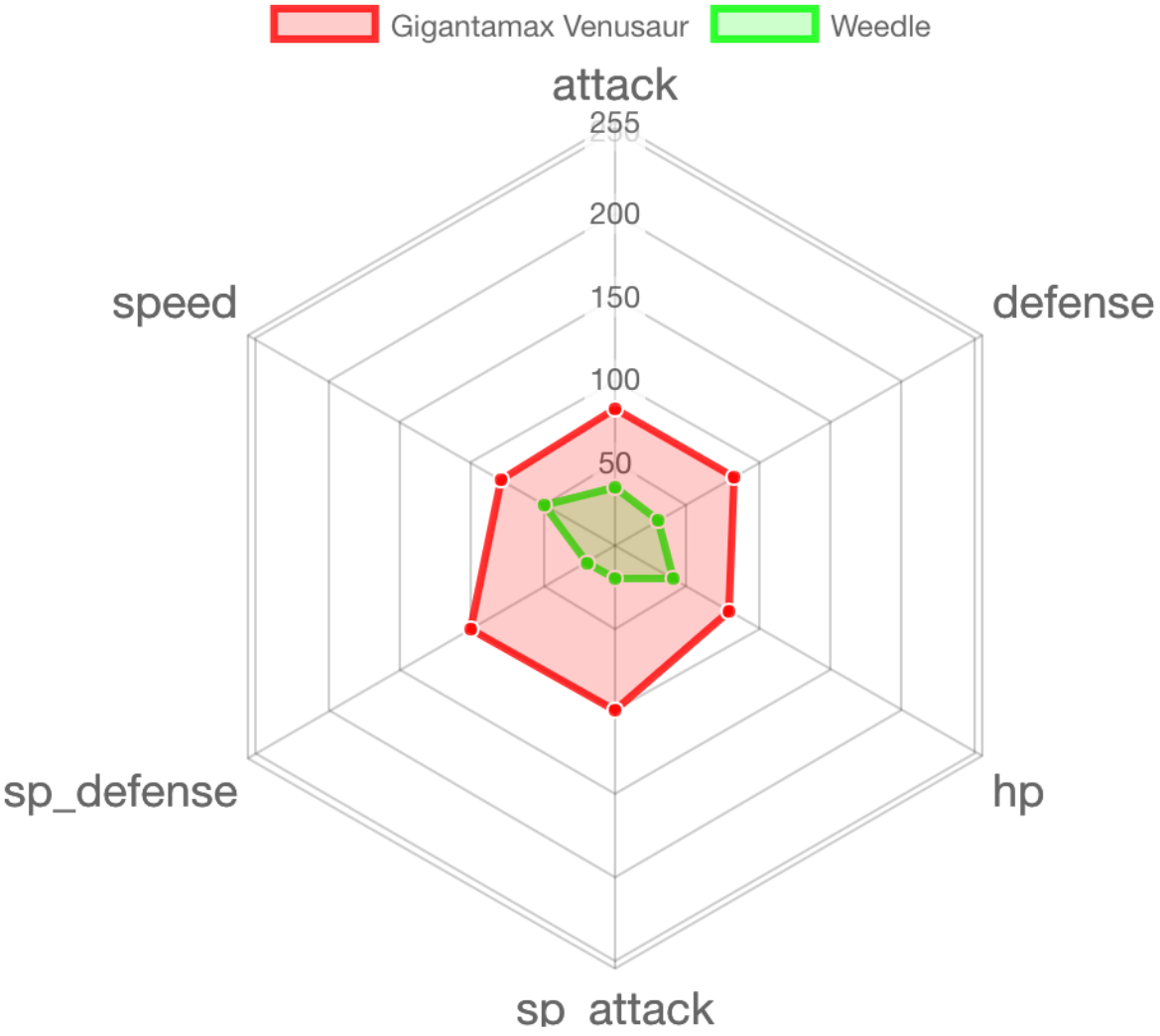
The learning objectives achieved through this project and this course was by giving the ability to collect data, manage data and data mine all while identifying patterns via statistical analysis. These skills were then leveraged to further answer business problems later in the program.

IST 687: Introduction to Data Science

Project Description

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 **(Figure 3)** while a Random Forest was used for machine learning **(Figure 4)**.

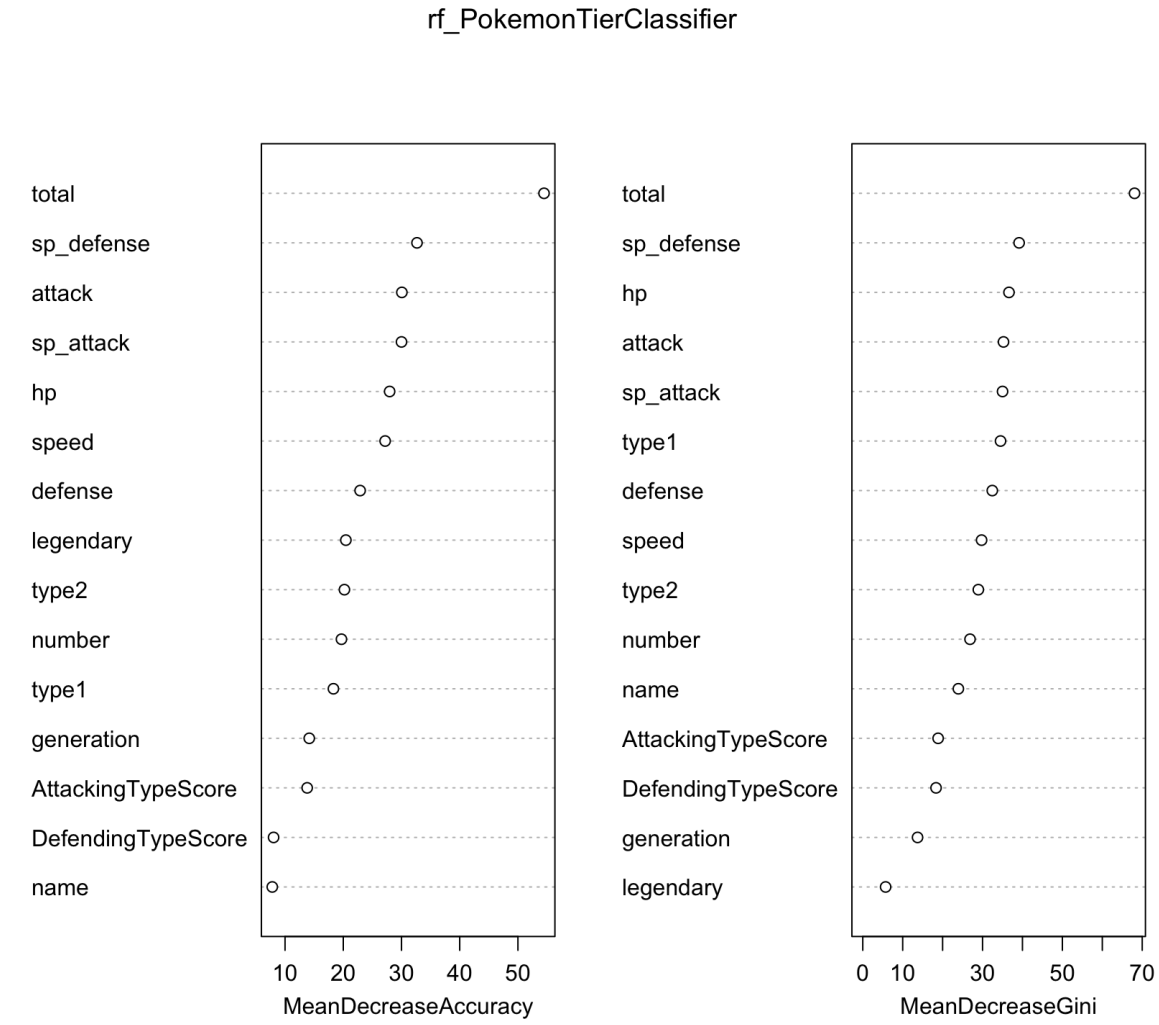


(Figure 3) Pokémon Head-to-Head Matchup (Morcos, IST 687, 2021)

Andrew Morcos
SUID: 625203065
https://github.com/AMorcos8/MADSPortfolio

```
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
```



(Figure 4) Pokémon Random Forest (Morcos, IST 687, 2021)

Andrew Morcos
SUID: 625203065
https://github.com/AMorcos8/MADSPortfolio

Reflection & Learning Objectives

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 & Decision Making

Project Description

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 **(Figure 5)**.

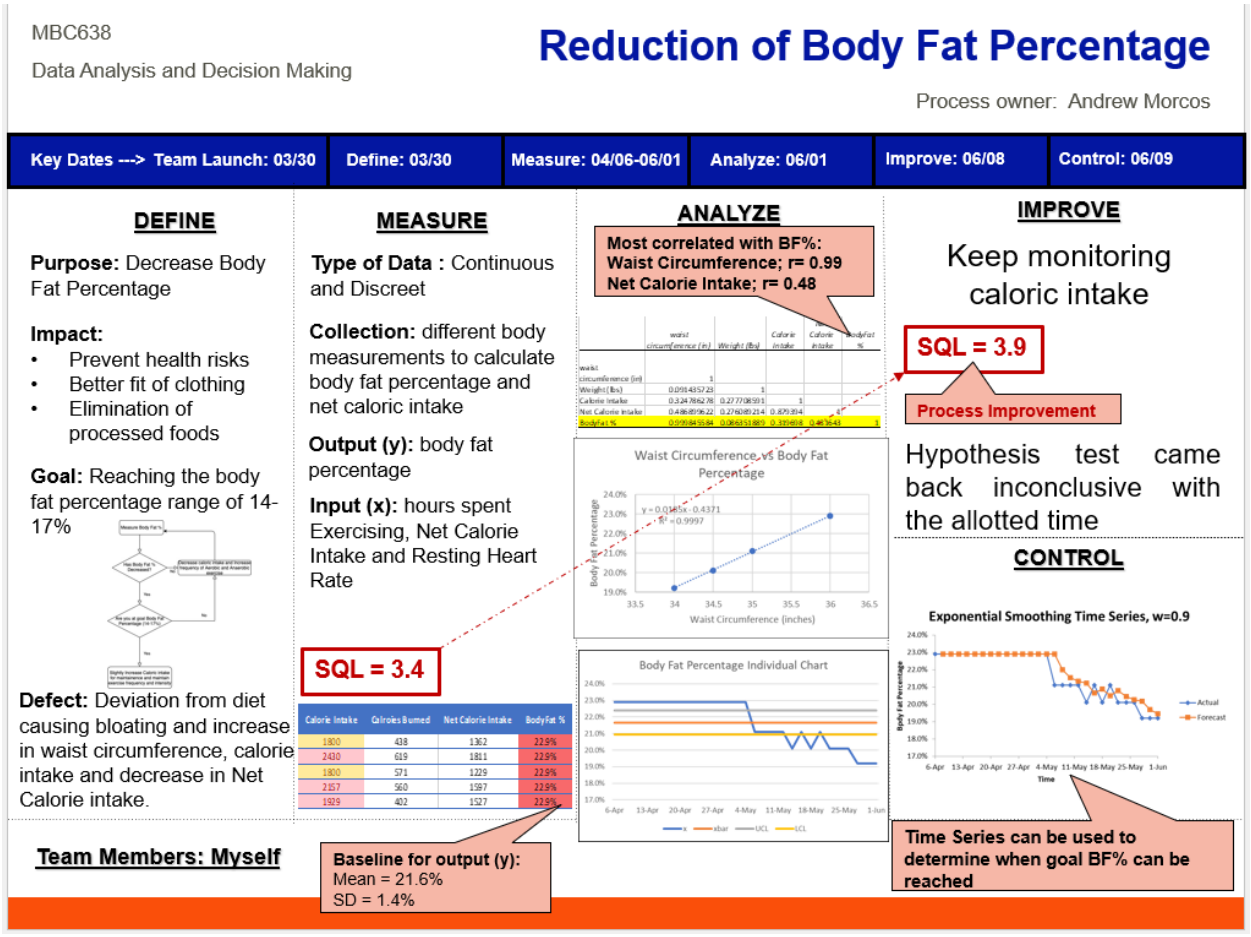


Figure 5 Process Improvement Storyboard (Morcos, MBC 638, 2021)

Reflection & Learning Objectives

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.

Andrew Morcos
SUID: 625203065
https://github.com/AMorcos8/MADSPortfolio

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 (Morcos, IST 659; IST 687; MBC 638). Various skills in data visualizations were used in conjunction with analytical and machine learning techniques to ultimately answer business questions. The primary focus on predictive analysis equips students with a skills to become an asset for various businesses when looking at the job market.

Communication skills were challenged and developed, which was displayed in the proper presentation of these projects to faculty in a coherent and concise manner. The ethical dimensions of data science practice was also implemented in these applications by discerning only relevant data and the consideration of privacy to the users.

In conclusion, Syracuse University Master's in Applied Data Science program equips students with the necessary skills to implement the learning objectives in the workforce. The skills attained helped student tackle multiple data structures, multiple styles of analysis, and multiple visualizations. The ethical standards in the program gives students the awareness of data integrity and the proper practice and protocol with data handling. 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.

References

Syracuse University MADS GitHub. (2022). GitHub; Andrew Morcos. https://github.com/AMorcos8/MADSPortfolio