

Master's in Applied Data Science Portfolio Draft

Yashaswi Pandey

2023-2025

Introduction

The Master's in Applied Data Science program at Syracuse University is designed to equip graduates with the tools and skills to provide enterprise solutions in the fields of data collection, management, analysis and decision making.

The learning objectives of the program are: -

1. Collect, store, and access data by identifying and leveraging applicable technologies
2. Create actionable insight across a range of contexts (e.g. societal, business, political), using data and the full data science life cycle
3. Apply visualization and predictive models to help generate actionable insight
4. Use programming languages such as R and Python to support the generation of actionable insight
5. Communicate insights gained via visualization and analytics to a broad range of audiences (including project sponsors and technical team leads)

6. Apply ethics in the development, use and evaluation of data and predictive models (e.g., fairness, bias, transparency, privacy)

This report will use some examples from different courses I completed that demonstrate achievements of the above learning goals.

IST 652 – Scripting for Data Analysis

1. Collect, Store, and Access Data:

The project collected and processed data using the Statsbomby Python library and mplsoccer wrapper to access open datasets. Match data, player lineups, and event data were collected, stored in data frames, and converted into CSVs for analysis.

2. Create Actionable Insights Across Contexts:

By analyzing Lionel Messi's performance through different phases of his career, the project provided insights into his evolution as a player. For example, it identified changes in Messi's positioning and passing patterns across different managerial eras, which could help scouts and coaches in real-world player analysis.

3. Use Programming Languages (R and Python):

Python was extensively used for data processing, analysis, and visualization using libraries like pandas, matplotlib.

4. Communicate Insights to Broad Audiences:

Complex and sparse data was broken down to visualizations and

tables to provide insights that are easily communicable from the results of the analysis of the data.

5. Ethics in Data Science:

Data used in the project is open source thus provides some level of transparency.

IST 719 – Information Visualization

1. Collect, Store, and Access Data:

The project utilizes datasets from StatsBomb, comprising 132,832 event records and 520 match records. The data was cleaned to focus solely on matches and events involving Messi.

2. Apply Visualization and Predictive Models:

Visualizations such as heatmaps, bar charts, and line graphs show Messi's scoring and assist patterns, and shot percentages across different seasons. This helps identify trends, such as his performance and consistency.

3. Use of Programming Languages (R):

The analysis employs R libraries like ggplot2, tidyverse, and StatsBombR for data manipulation and visualization.

4. Communicate Insights Effectively

The use of graphs and heatmaps makes complex data accessible to both technical and non-technical audiences, including fans, analysts, and sports enthusiasts.

5. Ethics in Data Science:

Data used in the project is open source thus provides some level of transparency.

IST 737 – Visual Analytic Dashboards

1. Collect, Store, and Access Data

The project utilizes datasets from StatsBomb. Match data, player lineups, and event data were collected, stored in data frames, and converted into CSVs for analysis.

2. Create Actionable Insight

The analysis uncovers meaningful insights into player and team performance across different tournaments. For example, identifying trends like scoring patterns and win rates

3. Apply Visualization

Tableau gave a lot of Visualisation tools.

4. Communicate Insights Effectively

Interactive dashboards in Tableau make the insights accessible

Github link for data - [YashaswiPandey/IST_782-Portfolio](https://github.com/YashaswiPandey/IST_782-Portfolio)

