Syracuse University

Memo

|  |  |
| --- | --- |
| To: | Dr. Debbie Landowski |
| From: | Patrick Bulger, Tim Hulak, Jordan Hyatt, Cal Wardell |
| Date: | 08/06/2021 |
| Re: | Project Proposal: Sports Analytics in the National Football League |

**Data:**

The data is a collection of play-by-play observations of a given play from the 2009 - 2018 NFL seasons. There are 449,371 observations of 255 variables. The data source is from Kaggle ([https://www.kaggle.com/maxhorowitz/nflplaybyplay2009to2016?select=NFL+Play+by+Play+2009-2018+%28v5%29.csv](https://www.kaggle.com/maxhorowitz/nflplaybyplay2009to2016?select=NFL+Play+by+Play+2009-2018+(v5).csv)). Given the specificity of the data, a lot of data cleaning and aggregation will be necessary.

**Objective**:

The main objective of the project is to focus on specific questions relating to performance. Each of the 4 team members will ask a question of the data and provide analysis on those questions. A report will be produced by the group with data descriptions, data cleaning annotations, and visualizations in order to provide a meaningful analysis and business recommendations.

**Questions:**

1. **Tim**: What are the effects of penalties on team performance?
2. **Jordan**: How conservative is a team with a lead (for example, running the ball with a lead to run out the clock)?
3. **Cal**: Winning percentage of any time quarterback that throws 40 or more passes a game
4. **Patrick**: Third down conversion rates and the impact it has on team success.

**Data Preparation And Plan:**

1. Load the data using python
2. Clean the data, including removing unneeded columns
3. Define functions to perform analysis to be used on each team’s data.
4. Explore the data for patterns and outliers
5. Analyze and visualize data based on research questions and provide conclusions for each.
6. Using web-scraping techniques, scrape web sources for news and information (such as Super Bowl Winners, season dates (since some seasons go into January of the following year), and data not included in the dataset such as a team’s division or conference).
7. Present business recommendation based on findings.