FURMAN COMPUTER SCIENCE

DEVELOPING STRATEGIES TO EVALUATE FOOTBALL PLAYS AND TO PROVIDE PREDICTIVE SCHEMATIC COMBINATIONS

KENDALL THOMAS

ABSTRACT

Football has long been a sport with a high volume of followship and a group that is passionate and loyal. This sport is a key income stream for high school, colleges and on a professional level. This income flow is heavily influenced by winning games. Having a way to identify key variables that affect wins and drives efficiency would valuable assets for a coach when strategizing game plans for a given season.

INTRODUCTION

Analytics is a growing area in the information technology sector. Getting data and knowing how to interpret it is key in making decisions as well as being more predictive in decisions that are made. My aim was to blend the two: take data from the game of football, specifically on the offensive side of play and use it to identify trends, areas of success and possibly derive combinations of plays that are more inclined to drive future success.

This report looks at a key areas of growing focus in our society today: football analytics and specifically if components could be identified that lead to more efficiency and ultimately wins.

PIPELINE
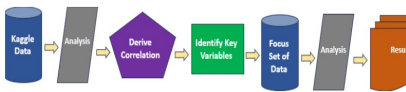
Kaggle data - statistical data was downloaded for review
Analysis - this data was investigated manually and then analytically for trends
Derive Correlations – correlation report generated
Identify key variables – key variables with higher correlations were identified for the next set of studies (to be applied on the **focus set**)
Analytics – data investigated to see if the behaviors held true or if others were observed
Results - Final observations will then be delivered as result sets

DATA

- Data across aspects such as Offensive plays, Offensive yards, Offensive TDs, PATs, 3rd down conversions, etc. were collected for FBS Teams from Kaggle. The first image shows one of the extensive sets of data files that were loaded as baseline data for observation.
- The next set of research involved the collection of data from ESPN (2nd and 3rd images) which would be a smaller more focused target set of data. This dataset was sliced into various aspects, including views based on one of the highest correlative factors identified in Step 1.
- Data then interpreted for any trends, influencing patterns or way to identify a more predictive approach
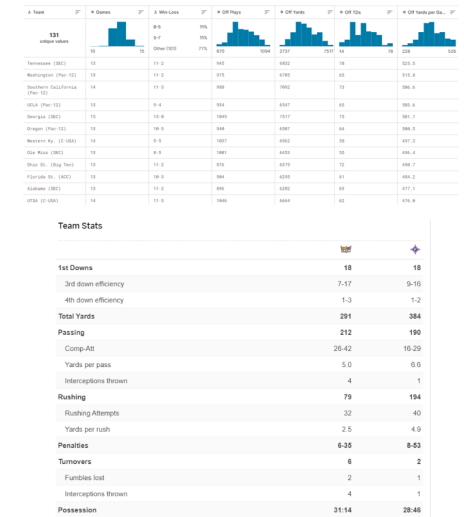
My projects are at:
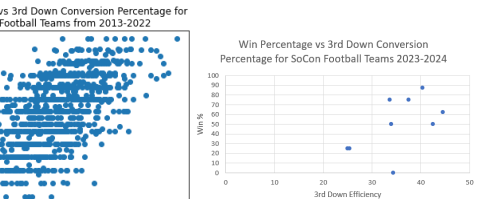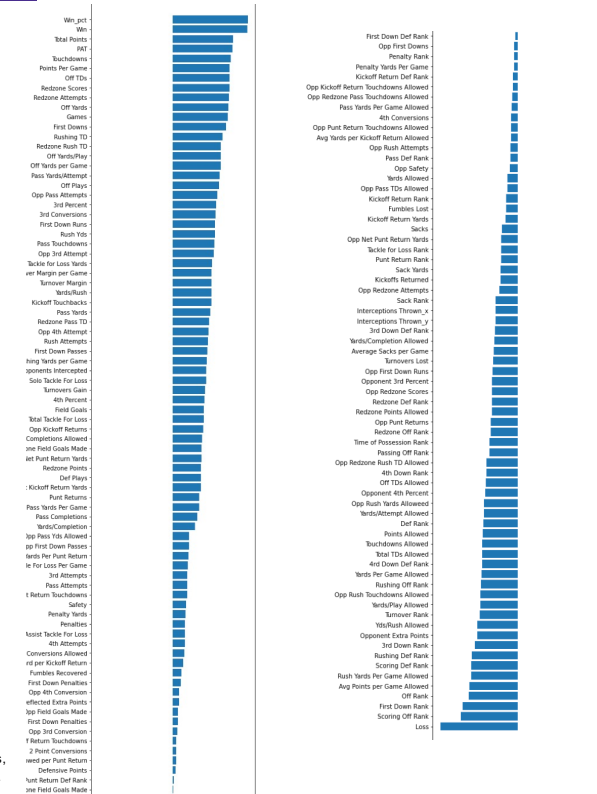httpshttp://github.com/Kendalltxii/CSC475/

CONCLUSION

There was a high and consistent correlation between 3rd down efficiency and win %. This was taken to a more detailed look within a target conference and proved consistent across team performing in the top 50%-win category. Future work could include a lower level of details to the type of 3rd down plays to see if the behaviour could become more predictive.

RESULTS

Win Percentage vs 3rd Down Conversion Percentage for College Football Teams from 2013-2022

Win Percentage vs 3rd Down Conversion Percentage for SoCon Football Teams 2023-2024

- Correlation analysis over a 20-year range for FBS schools (left image)
- Analytic reporting of 3rd down efficiency when applied to the largest data set (top image)
- Analytic reporting of the same variable being applied to a smaller more focused dataset (bottom image)
- It was confirmed that the 3rd down efficiency was consistently a contributor to wins
- Analytics identified other variables, but they didn't have a consistent trend to report