

Group 3 Project Proposal

Project Title: **“On the ground, or through the air”**
- *An objective end to football’s classic debate*

Team Members:

- Joseph Ayala
- Andrew Behrman
- Abdurrahman Darvesh
- Sean Findlay
- Michael Fox
- Michael Hankinson

Project Description: Our team will look to analyze historical NFL play-by-play and player data/trends to determine which offensive play strategy is more effective for a professional football team to win games: i) running plays or ii) passing plays

Preliminary Hypothesis: *“Running the football is a more effective means of i) winning football games overall, ii) improving average yards per offensive play, and iii) preventing turnovers during the game”*

Potential Research Questions to Answer:

- [BEHRMAN] Which type of offensive football play averages the most yardage per play?
- [BEHRMAN] Is running or passing a riskier means of advancing the ball down the field (i.e. is a team more likely to turn over the football via an interception or a fumble)?
- [FINDLAY] Are the league’s best teams tilted more towards a certain style of offense (i.e. 60% run, 40% pass)?
 - Do successful team’s strategies in certain “clutch situations” (i.e. once they get to the playoffs)?
- [DARVESH] Are there certain teams/players who are exceptions to the overall trends found in our dataset (for example, do the Patriots pass more and still win because they have Tom Brady)?
 - Picking 1 player from each position, which player had the more valuable season over the last 10 seasons?
- [FOX] Do certain weather conditions change the dynamics of a team’s strategy (i.e. does a team run more when it snows)?

- **[AYALA]** How much more effective is running the football for a team to maintain higher “time of possession?” Is this higher time of possession correlated to higher win percentages/chances of victory?
- **[HANKINSON]** Is winning percentage influenced by percentages of runs vs percentages of passes?

Potential Datasets to be Used:

- **Key Dataset - Kaggle:** “[Detailed NFL Play-by-Play Data \(2009-2018\)](#)”
 - **Features:** Home Team and Away Team, Yards per play, Play Type, Forced Fumble, Interception, Incomplete passes, Fumble out of bounds, fumbles lost, sacks, penalties caused by play type
 - 449k and 255 columns
- <https://www.footballoutsiders.com/stats/teamoff>
- <https://www.footballoutsiders.com/stat-analysis/2019/run-offense-number-backs-2018>
- <http://www.nfl.com/stats/statslab>
- <https://nextgenstats.nfl.com/> *Potential API Source*1
- Game Results: <https://www.pro-football-reference.com/boxscores/game-scores.htm>
- Weather: <https://www.ncdc.noaa.gov/cdo-web/webservices/v2>
- Stadiums/Teams: <https://www.kaggle.com/tobycrabbtree/nfl-scores-and-betting-data/downloads/nfl-scores-and-betting-data.zip/11>

Breakdown of Tasks:

- Tools/Infrastructure (git repo config): Mike Fox
- Presentation Slides: Andrew Behrman
- Statistics Guru/Mentor: Mike Hankinson
- Each Team Member will take one question to research and analyze and will perform this analysis end-to-end (data prep to final analysis). We will then bring results back together for summary conclusions and presentation.