Finding Fantastic Feats of Football in the NFL

Alex Plazas
Raul Ramos
Benjamin Price

Questions

Which aspects of a play have the greatest impact on play outcome?

Can players be classified into levels of achieving tiers based on their performance in their plays?

Which players perform the highest when faced with high-pressure situations, such as on 4th down or in the red zone?

Tools Used

Data Cleaning:

- Python: Numpy, Pandas, Scikit Learn
- Microsoft Excel

Visualization

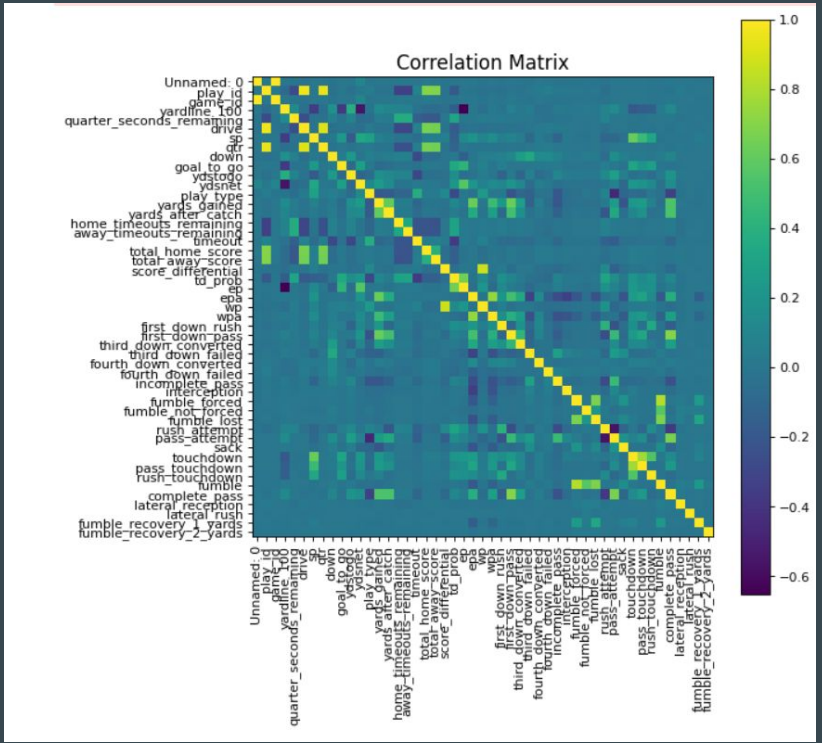
- Python: Matplotlib, Seaborn
- Tableau

Data Mining:

- Python: Numpy, Pandas, Scikit Learn, statsmodels
- Jupyter Hub

Data Preparation

- Reduced attributes from 253 to ~50 based on questions of interest
- Set nulls for integers to 0, set nulls for strings to “NA”
- Enumerated categorical attributes such as Play Type
- Feature Selected for 10 most important features correlated with EPA

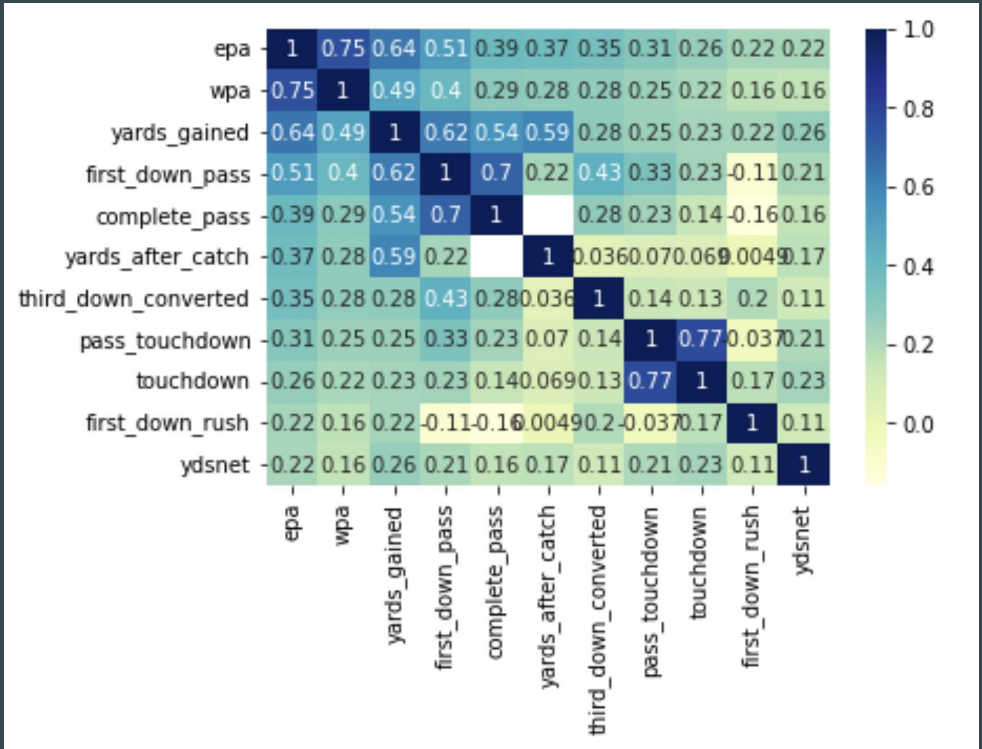


Linear Regression Applied

EPA as preferred attribute

Compare against top 10 attributes

Heatmap shows correlation

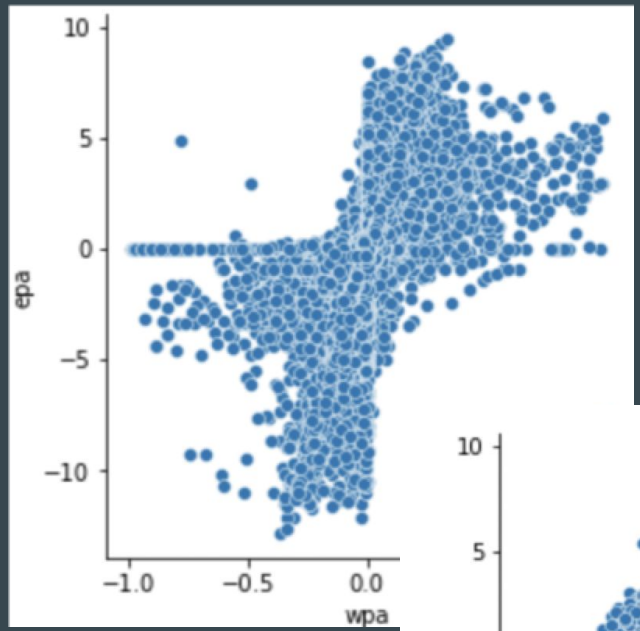


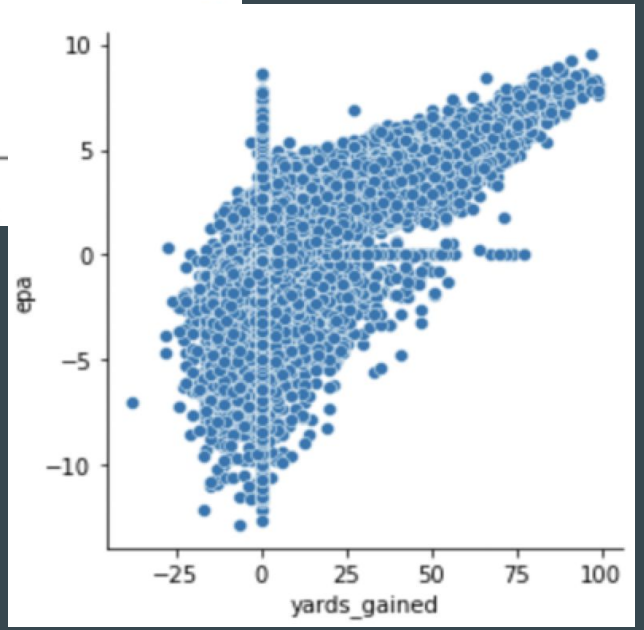
Linear Regression Applied (cont.)

Second attribute WPA

Third attribute yards gained

Scatterplots positive correlations





Linear Regression Applied (cont.)

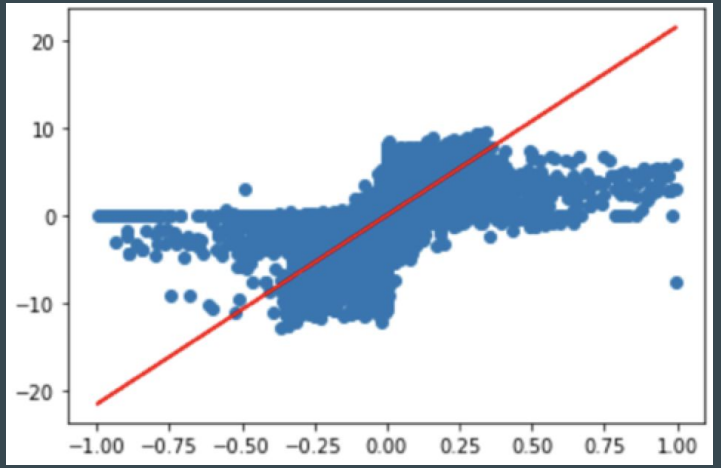
Years 2009-2018 divided into 10 segments

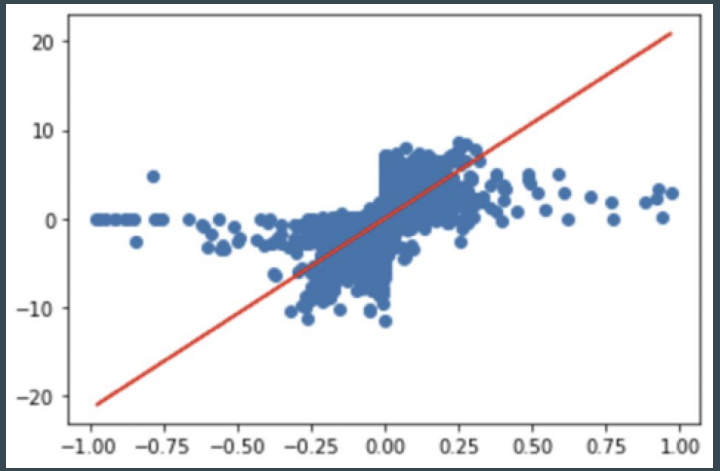
Train set (top) - 9/10

Test set (bottom) - 1/10

Positive correlation

-Talented players
-Winning games



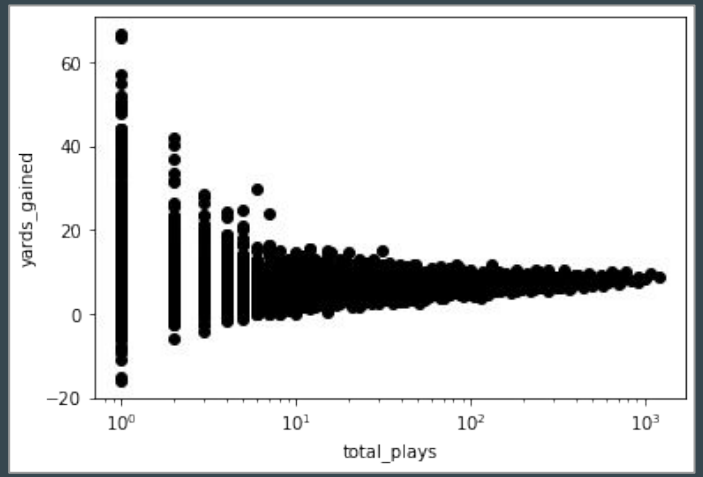


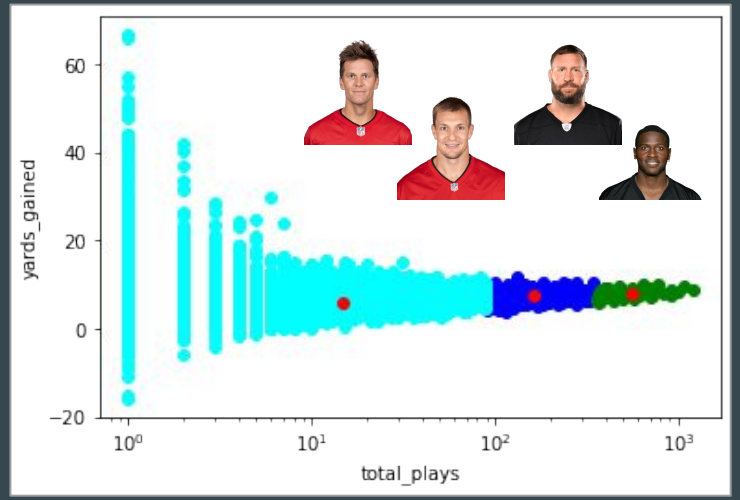
Clustering/Outlier Analysis Applied

- Main questions identified prior to analysis
  - “what do quarterback/receiver duos average for yards per completion versus the duo's number of attempts?”
  - “what running backs have completed ‘big plays’ (runs of 25 yards or more) per average yards a rush?”
  - “how do quarterbacks perform in the key situations of 4th down passes and ‘red zone’ passes?”
- Data filtered along key attributes per question
  - Dictionary created per player/player(s)
- K-Means Clustering implemented
  - Straightforward algorithm
  - Relatively fast / O(n)

Clustering/Outlier Analysis Applied (cont.)

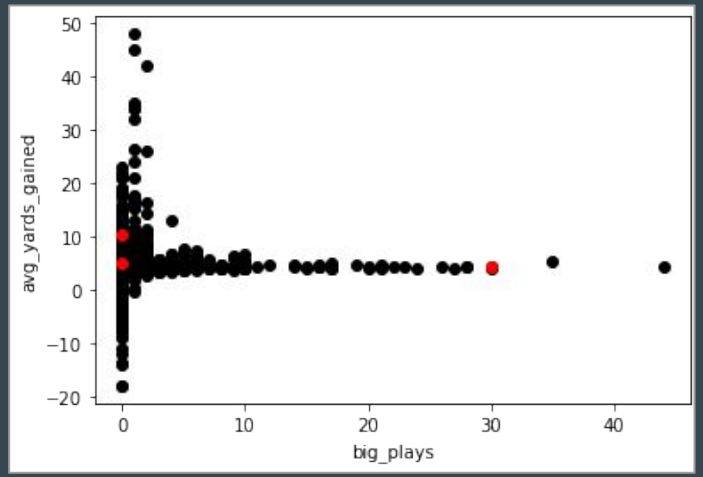
Quarterback/Receiver Duos:

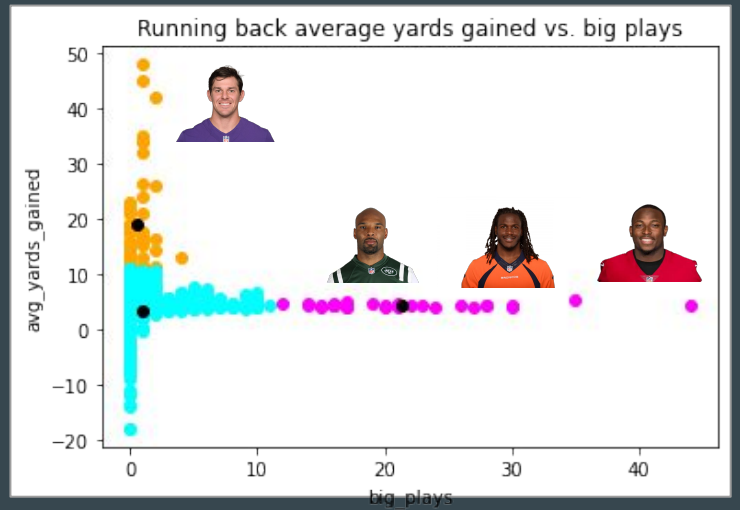




Clustering/Outlier Analysis Applied (cont.)

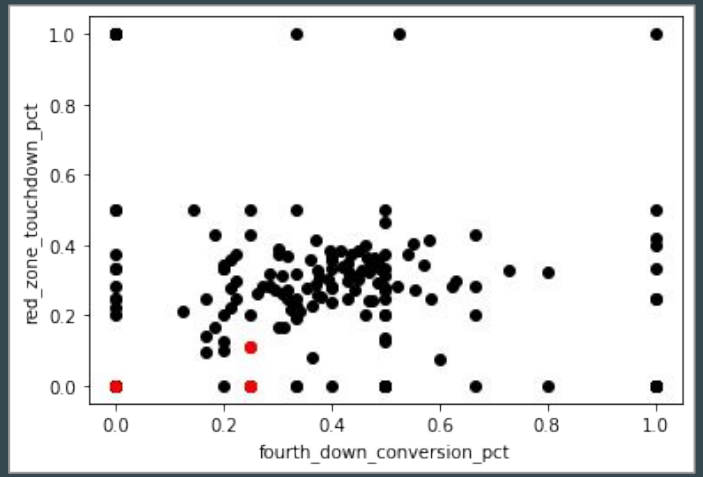
Running Back Explosive Plays:

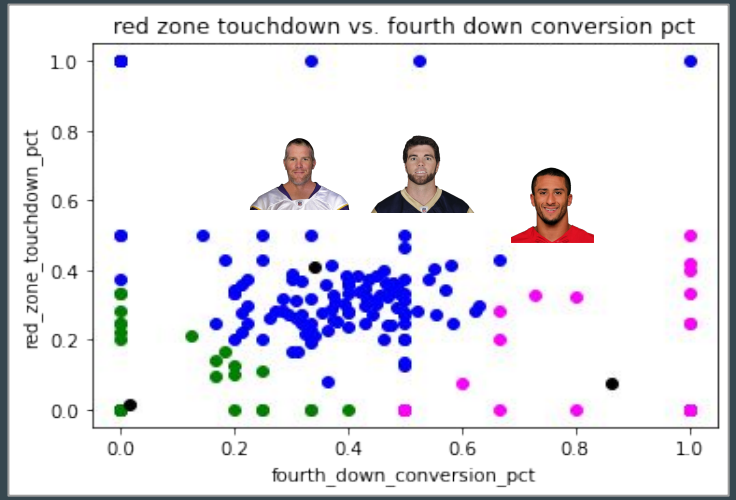




Clustering/Outlier Analysis Applied (cont.)

Crunch Time Quarterbacks:





Knowledge Gained

- Regression Analysis:
  - Talented players positive effect on team performance
  - Attack is the best form of defense
  - Teams should invest in talent to win games
  - Market eventually levels out
- Clustering/Outlier Analysis:
  - Quarterback/Wide Receiver Duos
    - Duos in the right-most cluster are Hall-of-Fame candidates
  - Explosive Running Backs (>25 yard runs)
    - Outer cluster comprised of all-time great running backs
  - High-Pressure Quarterbacks
    - Successful quarterbacks (Brett Favre, Peyton Manning, Colin Kaepernick) occupy the center cluster

How Knowledge Can Be Applied

- NFL Coaches/Support Staff
  - Crafting gameplans to give their team an edge
  - Identifying key plays and players
- NFL General Managers/Owners
  - Deciding which players to sign/give contracts to
  - Evaluating how their team is doing compared to the rest of the league
- Sports Writers
  - Better sense of who is outperforming the competition
  - Votes for end-of-season awards and the Hall-of-Fame
- Football/Sports Bettors
  - Picking teams based on expected point spreads
  - Strategies that involve outlier stat categories
- General Sports Fans
  - Take model implementation and use it in different contexts

Thank you!