Predicting Draft Round for College Football Skill Positions

By Chris Kourkoulakos
Brown University
October 24, 2024
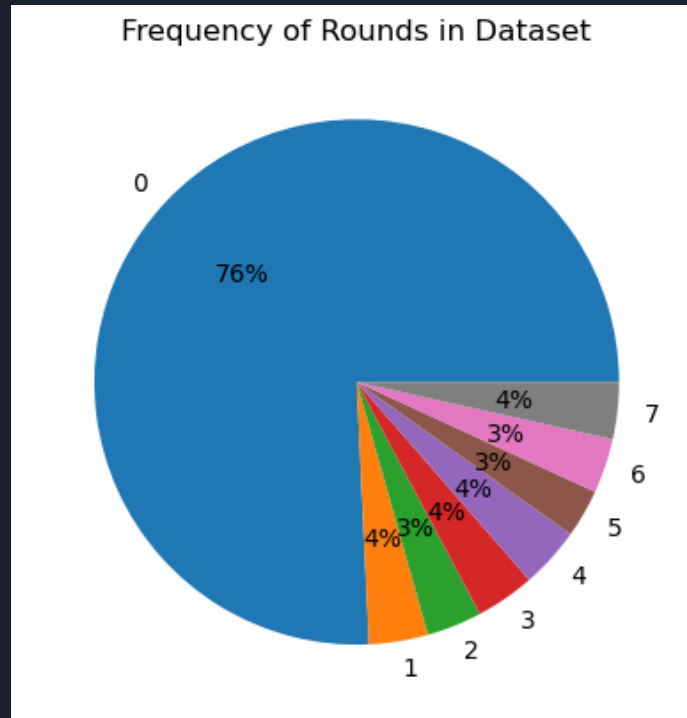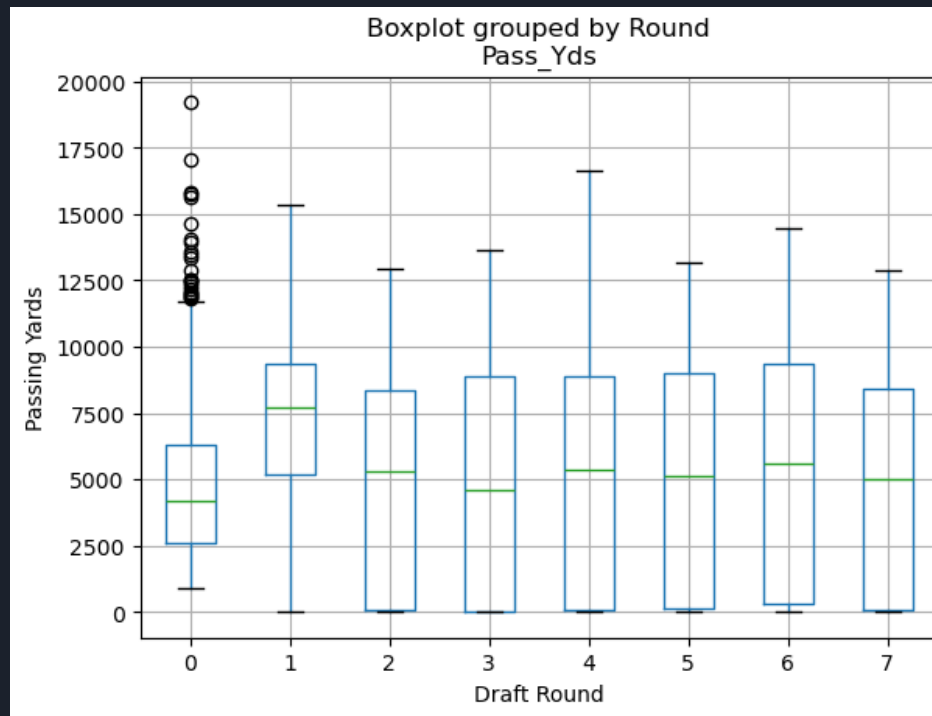https://github.com/ckourkou/DATA-1030-Final-Project.git

Intro

- Given college football player stats, predict what round players were drafted (or undrafted)
  - Important for scouting, draft strategy, trades, roster optimization
- Data from Stathead - partner of Pro Football Reference
  - Six separate data sets merged into one
- Data collected on all college football players dating back to 1956
  - Major passing, rushing, receiving stats from games played
  - Filtered year to 1993
- Classification problem - 8 groups (7 rounds plus undrafted group)
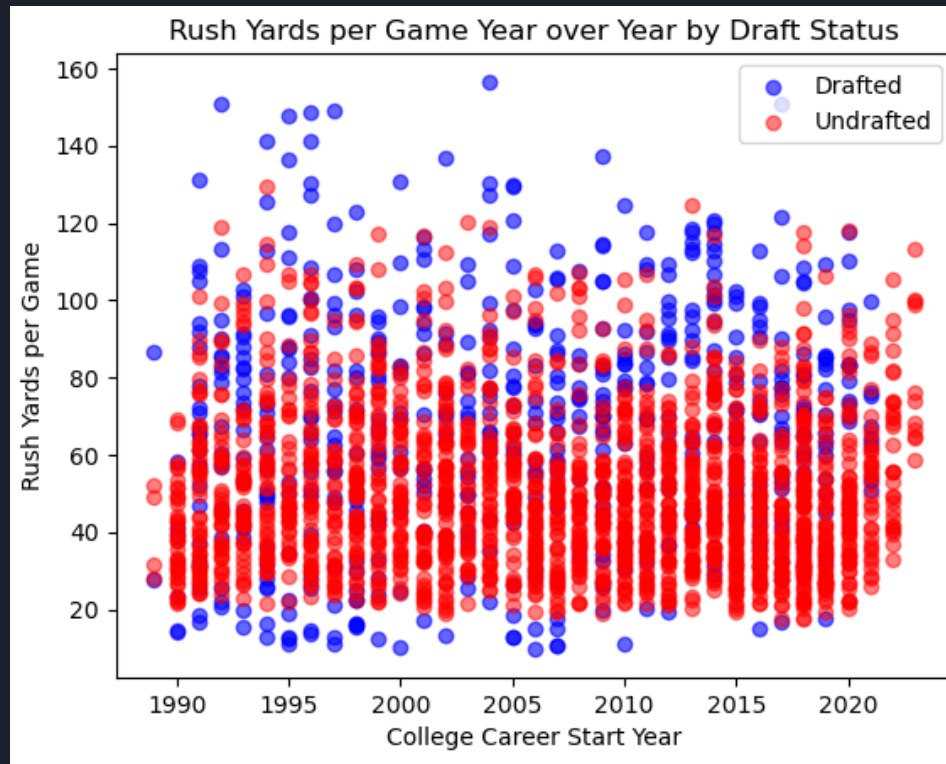- Challenge: missing data

EDA

Frequency of Rounds in Dataset

EDA

Boxplot grouped by Round
Pass_Yds

EDA

Rush Yards per Game Year over Year by Draft Status

Splitting and Preprocessing

- Many missing values
  - Mostly due to merging, some missing by nature
  - Position filled in by max of passing, rushing, receiving yards
  - Football metrics replaced with 0
- Stratified split because of disproportionate amount of undrafted players
- One-hot encoded features
  - Position, Years (From and To)
- Multilabelbinarizer features
  - Team
- MinMax scaled features
  - Completion %, TD %, Int %
- Standard Scaled features
  - Everything else
- 29 columns in training data prior to preprocessing, 244 after
  - 5295 rows

```
fraction of missing values in features:
Pass_TD       0.785042
Cmp           0.785042
Pass_Att      0.785042
Inc           0.785042
Cmp%          0.785042
Pass_Yds      0.785042
Int           0.785042
TD%           0.785042
Int%          0.785042
Rate          0.785042
Pass_Y/A      0.785042
Pass_AY/A     0.785042
Pass_Y/C      0.785042
Pass_Y/G      0.785042
Pos           0.010085
Rush_Yds      0.662550
Rush_Att      0.662550
Rush_Y/A      0.662550
Rush_TD       0.662550
Rush_Y/G      0.662550
Rec_Yds       0.422436
Rec           0.422436
Y/R           0.422436
Rec_TD        0.422436
Rec_Y/G       0.422436
```

Questions?