Data Appendix: Fantasy Football Sentiment Analysis (2025-2026)

1. fantasypros_articles.csv + ffballers_articles.csv

Unit of Observation:

The unit of observation is a single article.

- Multiple players appear in a single article. During scraping, we used HTML class names to add markers before each individual player name. We could use this when we cleaned the csv to get individual player articles.
- We also stored the publish date of the article, which is used later to create a dataset based on week by week analyses
- COLUMNS:
    - url: link to the scraped article
    - title: title of the scraped article
    - publish_date: publish date of the scraped article
    - body_text: full body text, contains multiple analyses for a given week

2. text_dataset.json

Unit of Observation:

The unit of observation is a player-specific analysis, extracted from a subsection of an article.

- This dataset arranges our data article by article. Within each article, we organize player by player analyses based on the markers placed in the first two datasets.
- The two datasets from step 1 are merged together

```
[
  {
    "meta_title": "String (The headline of the article)",
    "meta_url": "String (URL to the full article)",
    "meta_date": "String (ISO 8601 Date/Time format, e.g., 'YYYY-MM-DD HH:MM:SS')",
    "intro_text": "String (The introductory paragraph of the article)",
    "players": [
      {
        "name": "String (Name of the player being analyzed)",
        "raw_header": "String (The full header text, often including position/team, e.g.,
'Trey McBride (TE - ARI)')",
        "type": "String (Category of the analysis, e.g., 'Standard', 'Sleeper', 'Bust')",
        "analysis": "String (The full text of the expert's write-up/advice for this player)"
      }

    ]
  }
]
```

3. fantasy_sentiment_scores_2025.csv

Unit of Observation:

The unit of observation is a unique player-week analysis.

- This dataset arranges our data week by week.
- For each row, we include the player, article metadata, and sentiment scores.
    - NOTE: some players have multiple entries for a single week, since they appeared in multiple articles
- We enrich each row with the sentiment score derived from the analysis in text_dataset.json
- COLUMNS:
    - week: the week of the 2025 NFL season (1-18)
    - player_name: name of the player
    - sentiment_compound: overall sentiment score
    - sentiment_pos: percentage of analysis with positive sentiment score
    - sentiment_neg: percentage of analysis with negative sentiment score
    - sentiment_neu: percentage of analysis with neutral sentiment score
    - word_count: number of words in the analysis, used to calculate sentiment score
    - article_date: date of the article containing the analysis
    - article_title: title of the article containing the analysis
    - article_url: link to the article containing the analysis

4. stats_dataset_2025_cleaned.csv

Unit of Observation:

The unit of observation is player-specific fantasy statistics, collected from a 2025 NFL fantasy statistics dataset.

- This dataset arranges player statistics by week for the 2025-26 NFL season
- Data includes week, position, team, player name, points scored, and week rank
- COLUMNS:
    - position: Player position (QB, WR, RB, TE, K)
    - week: the week of the 2025 NFL season (1-18)
    - PlayerName: name of the player
    - Team: player team (abbreviated, ex. NE, SEA, CHI, PHI)
    - Rank: the player's fantasy rank in the given week based on position
    - TotalPoints: the player's total fantasy points scored in the given week

5. fantasy_dataset_final.csv

Unit of Observation:

The unit of observation is a player-week event.

- This dataset merges the statistical data into the sentiment dataset for analysis use.
- Includes all columns from **stats_dataset_2025_cleaned.csv** and **fantasy_sentiment_scores_2025.csv** , merged based on player name and week

Descriptive Statistics & Plots: