Geospatial Analysis of NFL Game Outcomes and Environmental Factors

Data Source: nfldata/DATASETS.md at master · nflverse/nfldata (github.com)

- **Source**:

External; leesharpe on github repository

Who owns it? Inspired by nflscrpR and nflfastR open source NFL community

Original data collected is from ProFootballReference

- **Trustworthiness**?

Community Feedback: The package has received positive feedback from the NFL analytics community, indicating its reliability and usefulness.

Open Source: Being open source, the code is available for review on GitHub, allowing users to verify its accuracy and contribute to its improvement.

Regular Updates: The package is maintained and updated regularly, ensuring that it stays current with the latest NFL data.

Original Data Content:

Columns:

- game_id: The ID of the game as assigned by the NFL. Note that this value matches the game_id field in nflscrapR if you wish to join the data.
- alt_game_id: This is a more human-readable ID. It consists of: The season, an underscore, the two-digit week number, an underscore, the away team, an underscore, the home team.
- season: The year of the NFL season. This reperesents the whole season, so regular season games that happen in January as well as playoff games will occur in the year after this number.
- game_type: What type of game? One of the following values:
    - REG: a regular season game

  - WC: a wildcard playoff game
  - DIV: a divisional round playoff game
  - CON: a conference championship
  - `SB': a Super Bowl
- week: The week of the NFL season the game occurs in. This will be 1-17 for the regular season, 18 for wildcard playoff games, 19 for divisional playoff games, 20 for conference championships and 21 for Super Bowls.
- gameday: The date on which the game occurred.
- weekday: The day of the week on which the game occcured.
- gametime: The kickoff time of the game. This is represented in 24-hour time and the Eastern time zone, regardless of what time zone the game was being played in.
- away_team: The away team.
- away_score: The number of points the away team scored. Is NA for games which haven't yet been played.
- home_team: The home team. Note that this contains the designated home team for games which no team is playing at home such as Super Bowls or NFL International games.
- home_score: The number of points the home team scored. Is NA for games which haven't yet been played.
- location: Either Home if the home team is playing in their home stadium, or Neutral if the game is being played at a neutral location. This still shows as Home for games between the Giants and Jets even though they share the same home stadium.
- result: The number of points the home team scored minus the number of points the visiting team scored. Equals h_score - v_score. Is NA for games which haven't yet been played. Convenient for evaluating against the spread bets.
- total: The sum of each team's score in the game. Equals h_score + v_score. Is NA for games which haven't yet been played. Convenient for evaluating over/under total bets.
- gsis: The id of the game issued by the NFL Game Statistics & Information System.

- pfr: The id of the game issued by Pro Football Reference
- pff: The id of the game issued by Pro Football Focus
- espn: The id of the game issued by ESPN
- spread_line: The spread line for the game. A positive number means the home team was favored by that many points, a negative number means the away team was favored by that many points. This lines up with the result column.
- total_line: The total line for the game.
- roof: What was the status of the stadium's roof? Will be one of the following values:
    - outdoors: An outdoor stadium
    - open: Stadium has a retractable roof which was open
    - closed: Stadium has a retractable roof which was closed
    - dome: An indoor stadium
- surface: What type of ground the game was played on
- temp: The temperature at the stadium (for outdoors and open only)
- wind: The speed of the wind in miles/hour (for outdoors and open only)
- away_coach: Name of the head coach of the away team
- home_coach: Name of the head coach of the home team
- referee: Name of the game's referee (head official)
- stadium: Name of the stadium

Data Cleaning Procedures within Jupyter Notebook Link: NFL Final Project 2.zip

Data Content After Cleaning: nfl_spread_analysis.xlsx

'game_id':

'season':

'game_type':

'week':

'gameday':

'weekday':

'gametime':

'away_team':

'away_score':

'home_team':

'home_score':

'location':

'actual_spread': original is result but renamed

'actual_total': original column is total but renamed

'overtime':

'away_rest': away team rest time in days between next game

'home_rest': home team rest time in days between next game

'projected_spread': original column is spread_line but renamed

'projected_total': original column is total_line but renamed

'div_game':

'roof':

'surface':

'temp':

'wind':

'stadium_id':

'stadium':

'home_market': Home team U.S. market location

'home_name': Home team name

'away_market': Away team U.S. market location

'away_name': Away team name

'home_division': Home team regional U.S. geographic division

'away_division': Away team regional U.S. geographic division

Ethical Concerns and Limitations

1. **Data Privacy:**
    - Ensure that any personal data of players, officials, or fans is anonymized and protected.
    - Avoid using data that could lead to the identification of individuals without their consent.
2. **Bias and Fairness:**
    - Be aware of potential biases in the data, such as historical performance trends that may not reflect current team capabilities.
    - Ensure that the analysis does not unfairly favor or discriminate against any team.
3. **Data Quality:**
    - The accuracy and completeness of the data are crucial. Missing or incorrect data can lead to misleading conclusions.
    - Consider the source of the data and its reliability.
4. **Ethical Use of Findings:**

- Use the findings responsibly, especially when it comes to betting and gambling. Promote responsible gambling practices.
- Avoid using the analysis to manipulate or exploit betting markets unethically.

Analysis Questions

1. **Geospatial Trends**:
   - How do different teams perform in various locations and stadiums?
   - Are there any noticeable trends in team performance based on the game location (home vs. away)?
2. **Impact of Game Conditions**:
   - How does the type of roof (open, closed, retractable) affect game outcomes?
3. **Weather Impact**:
   - How do temperature, wind speed, and other weather conditions affect game outcomes and scores?
   - Are there specific weather conditions that favor home or away teams?
4. **Betting and Spread Analysis**:
   - How often does the team with the spread advantage win?
   - Is there a correlation between the spread line and the actual score difference?
   - How often do games exceed the total points line?
5. **Team Performance**:
   - Which teams have the highest average scores at home versus away?
   - How do different teams perform against the spread?