**Project Check-In**

**NFL Team Performance vs Weather**

**Andy Roose, Dave Marcheschi, Jack Gray**

1. **Introduction**

Weather has always been a major factor in NFL games because of the effect it can have on the outcome of a game, but which weather elements have the most impact? This information is important for teams to know as they head into games that are not going to have the best weather. Our hypothesis is that extreme temperatures, increased precipitation, and higher windspeeds would increase turnovers and decrease both total points and total yards gained for a given game. Based upon prior knowledge of the NFL we think precipitation would have the largest impact on these variables. We are also curious if games played in domes typically have higher points scored and total yards since the element of weather is taken out of the equation in these games. We believe that domed stadiums would likely have more points and yards because of this. We also want to see how teams who are located in colder climates play in the cold compare to teams who are located in more moderate climates. It is often believed that teams in colder climates have an advantage over teams from moderate climates in cold games. We believe coaches and players could use the information we learn in order to know how to better prepare for a game. For example, if it is known that it will rain for a certain game than the team could use data to know that there will be a higher likelihood of turnovers so it will be important for the team to focus on minimizing turnovers during the contest.

In our project we used data obtained from Kaggle and Pro-Football Reference to see if we can answer out analysis questions. We joined both datasets by using horizontal integration in R Studio.

**2. Data**

This project uses two sources of data: *Weather Data – Kaggle;* This dataset contains weather data from the 2000-2020 NFL seasons. *Pro-Football Reference;* This data set was scraped from pro-football-reference.com and contains game result data from the 2000-2020 NFL seasons including turnovers, winning team, yards, and points.

<https://www.pro-football-reference.com/years/2000/games.htm> <https://www.kaggle.com/datasets/tombliss/weather-data?resource=download>

*Scaping/Integrating Process:*

We first started by scraping data from the Pro Football Reference website (NFL data frame). The data consisted of NFL games stats throughout each week of the season. We then crawled through 20 pages to collect data from the years 2000-2020. Each column name was changed for ease of understanding and integration. One column was added using the data from the ‘Location’ column giving us the home team for each game. A few columns and rows were removed due to them being unnecessary for our analysis questions. Our final data frame for the NFL data was 5,593 observations of 13 variables.

The next set of data we used was from Kaggle, which had data on both NFL stadiums and the weather during each hour of each game played. Using pivot tables, we summarized the data into sums and averages as the original dataset had weather data by hour instead of for the entire game. Like the NFL data frame, column names were changed for ease of understanding and integration, and unnecessary columns and rows were removed. The final data frame (weather) consisted of 5,583 observations of 10 variables.

Horizontal integration was very simple after cleaning both data frames. We merged the data based on the date and home team to connect the date and location of each game with the weather. The weather data frame contained information on games played internationally, where the NFL data frame does not. Those rows were lost in the integration of the two data frames. Our final, merged data frame (NFL2) consists of 5,346 observations of 21 variables.

*Data Cleaning:*

Changing Categorical Values to Binary: For the Roof Type column we changed the instance indoor to a 1 and the instance Outdoor to a 0 to make analysis easier.

Removing Unnecessary Columns: We removed the Day Of Week, Time, and Location columns as they were not necessary to our analysis. We also removed Average Pressure and Average Humidity because we felt these data points do not impact games in a strong way.

Our final data set has 23 columns and 5347 rows.

*Data Dictionary:*



**3. Analysis**

*3.1 Do teams located in colder climates play better in the cold compared to teams located in more moderate Climates?*

To determine whether NFL teams from colder climates perform better in cold weather games compared to teams from warmer climates, an analysis of various performance metrics was conducted. The data points considered were average points, average yards, average turnovers, and win percentages.

To analyze the performance statistics of NFL teams in different climates, we established a binary variable in our Excel spreadsheet. We determined which NFL teams originated from colder climates by consulting a map and subsequently created a reference table. This table allowed us to utilize the VLOOKUP function to assign a binary variable to each game's winner, based on their climate origin. To define what constitutes a cold NFL game for our study, we set a temperature threshold of 30 degrees Fahrenheit or lower. This temperature criterion was then used to segment the data accordingly.

The analysis started by comparing average points scored in cold weather games. Teams from colder climates scored an average of 28.6 points, slightly higher than the 26.22 points scored by teams from warmer climates. This suggests a minor advantage in scoring for teams used to colder environments, potentially due to greater familiarity and preparation for such conditions.

Further exploration into average yards gained showed a negligible difference between the two groups, with colder climate teams achieving 358.78 yards and warmer climate teams achieving 358.32 yards. This indicates that the ability to gain yardage is consistently maintained across teams, regardless of their typical climate exposure.

When examining turnovers, both groups again showed similar performance with an average of 1.2 turnovers for colder climate teams and 1.15 for warmer climate teams. The close figures imply that cold weather does not significantly affect the ball security.

However, a significant variance was observed in the win percentages. Teams from colder climates exhibited a dramatically higher win rate of 82.28% in comparison to the 17.72% for teams from warmer climates. This stark contrast highlights a considerable advantage for teams habituated to colder conditions, likely due to their regular exposure and adaptation to such environments, which could enhance their overall game strategy and execution in cold weather.

A screenshot of a graph

Description automatically generated

*3.2 Does inclement weather affect how teams perform during games by a significant amount?*

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Average\_of\_WindSpeed** | **Sum\_of\_Precipitation** | **Average\_of\_Temperature** |
| **Total\_PTS** | -0.111490605 | -0.05980349 | -0.002075731 |
| **Total\_To** | -0.004564124 | 0.01454379 | -0.022272761 |
| **Total\_Yards** | -0.111885648 | -0.03509638 | 0.025704802 |

*Table 1 Correlation Coefficient Summary*

For this portion of the analysis, I wanted to dive into the effect that the weather elements of wind, precipitation, and temperature have on the turnovers, yards, and points scored in games. I began by looking at the correlation coefficients between all these variables. There ended up not being any strong correlations between any of the variables, which was surprising to our group. Windspeed had a negative correlation with both total points and total yards of -.11. Although both instances had the strongest correlations between any of the variables it was still negligible. The most surprising result from analyzing the correlation coefficients was that precipitation did not have a significant negative correlation with points scored because this was the variable we initially thought would have the largest impact on games.

Our first thought was that the weak correlations are due to the fact that most NFL games have pretty moderate weather so there may not be enough games in our dataset with poor weather to establish a strong relationship. To dive into this further we created histograms for the weather variables. The histograms can be viewed in the appendix, and they all show that like we assumed, there is a low number of games that have either high precipitation, extreme temperatures, or high winds. To better visualize the results, I created scatterplots using the different variables and found out that total points and precipitation displayed the best indication of a trend. The scatterplot can be seen below in figure 1. The scatterplot shows a downward trend in accordance with more precipitation, so we dove into this further.

A graph with blue dots

Description automatically generated

Figure 1 Scatter Plot of Total Points by Precipitation

After looking at the initial analysis we decided to focus on total points as the target as we believe that this variable plays the biggest role in the outcome of a given game, and it was affected the most by weather when conducting the analysis that follows. The scatter plot showed that increased precipitation resulted in fewer points scored so we created a new column that categorized precipitation according to the amount of precipitation for a given game. We created a bar plot showing the Total Points scored by precipitation category which is shown below in figure 2. The bar chart shows a clear decrease in total points as precipitation increases. The Average Total points are 44 for no precipitation and 32 for heavy precipitation meaning there is a 12 point decrease in points between the two groups

A graph of different colored rectangular shapes

Description automatically generated

Figure 2 Average Total Points by Precipitation Category

We also looked at the effects wind speed and temperature have on total points by creating similar groupings and creating bar plots. The definitions for the groupings can be seen in the data dictionary and the bar plots can be seen below in figure 3 and 4. When looking at figure 4 based on then categories we created that temperature has little effect on total points. However, figure 4 shows a decrease in total points between low and high windspeed.

*A graph of different colored rectangles

Description automatically generated*A graph of different colors

Description automatically generated

Figure 4 Average Total Points by Wind Speed Category

Figure 4 Average Total Points by Temp. Category

Although there was little correlation between the weather variables and game metrics the graphs we created show that when grouping the weather variables high wind and precipitation decrease the total points scored on average.

*3.3 Do teams with more turnovers typically lose more?*

Through our data, we wanted to answer the question on if teams that turn the ball over more also lose more. To begin this analysis, we created a new data frame containing the total amount of turnovers each NFL team has had in the past 20 seasons, the number of losses each team has had in the past 20 seasons, and the average number of turnovers per loss for each team (turnovers/losses).

A screenshot of a graph

Description automatically generated

With this data frame we then organized the teams from highest average turnovers to the lowest average turnovers. The Arizona Cardinals had the highest average number of turnovers per loss at 2.43. The Kansas City Chiefs had the lowest at 1.83.

To further our analysis, we used a scatterplot with a smooth line to visualize the relationship between turnovers and losses. Through statistical analysis and the graph, it is clear that there is a strong positive relationship betweent the two variables (0.98 correlation). We calculated the formula for the smooth line using linear regression and added it to the graph. Through this analysis we can conclude that teams that turn the ball over more typically lose more.

A graph with a red line and black dots

Description automatically generated

**4. Conclusion**

In this project, we analyzed NFL games across 20 seasons and found how different variables might affect outcomes and statistics. The following answers were found for our analysis questions presented in the proposal.

1. *Do teams located in colder climates play better in the cold compared to teams located in more moderate Climates?*
2. *Does inclement weather affect how teams perform during games by a significant amount?*

In our analysis we discovered that there is low correlation between the weather elements we looked at and the game metrics we chose. However when grouping the weather elements by severity we discovered that heavy rain and high windspeeds decrease the amount of points scored in a given game.

1. *Do teams with more turnovers typically lose more?*

The analysis conducted on turnovers and losses in NFL games reveals a strong positive correlation of 0.98. This indicates that as the number of turnovers increases, so does the likelihood of a team losing a game. However, we must acknowledge limitations. Although the correlation is strong, correlation does not imply causation. Many other factors could influence game results. In the future, it would be interesting to dive deeper and explore the impact of turnovers in specific game scenarios (turnovers in the red zone or turnovers leading to points for the opposing team) which could provide deeper insights into their significance.

This project has a few limitations including not having game data past 2000 and only having turnover, points, and yards for each game. The dataset also does not have many games where there was extreme weather which limited our analysis. Future work could include looking at a larger range of games and looking at more statistics within games.

**Appendix**

A graph of a graph

Description automatically generatedA graph with blue bars

Description automatically generatedHistograms for Weather Elements

A graph of a temperature

Description automatically generated