AUGUST 3, 2019

GROUP FINAL PROJECT
NFL INFORMATION AND ANALYSIS
INFS776

DANIEL CAREY, ANDREW MARSHALL, JUSTIN ROBINETTE, KARTHIK VANGURY

INTRODUCTION

For this project, we examined the NFL scores dataset from Kaggle, located here: https://www.kaggle.com/tobycrabtree/nfl-scores-and-betting-data#spreadspoke_scores.csv. From this dataset, we hoped to gather information about the NFL as a whole, individual NFL teams, and the influencers of the outcomes of NFL games. In doing so, we worked together to gather insights and produce this document that provides a summary of the above-mentioned items.

DATA

There are three distinct datasets provided from Kaggle through the link above. The first, titled **nfl_teams.csv** has the following columns:

- → **Team Name:**   The name of the NFL team
- → **Team Short:**   The "mascot" name of the team
- → **Team Id:**   A 3 letter abbreviation of the team
- → **Team Id Pfr:**   An alternative 3 letter abbreviation of the team
- → **Team Conference:**   The conference in which the team competes
- → **Team Division:**   The division within the conference in which the team competes
- → **Team Conference Pre2002:**   The conference in which the team competed before 2002
- → **Team Division Pre2002:**   The division within the conference in which the team competed before 2002

The second, titled **nfl_stadiums.csv** has the following columns:

- → **Stadium Name:**   This is the formal name of the stadium
- → **Stadium Location:**   This is the city in which the stadium resides
- → **Stadium Open:**   This is the year in which the stadium opened
- → **Stadium Closed:**   This is the year in which the stadium closed (if applicable)
- → **Stadium Type:**   This is a descriptive feature describing whether the stadium outdoor (no roof), indoor (roof), or retractable (roof that can can be opened)
- → **Stadium Address:**   This is the address of the stadium
- → **Stadium Weather Code:**   This is the code (zip code) used by the weather station for the stadium
- → **Stadium Weather Type:**   This is the type of weather category given to the stadium and can be either *warm, moderate, cold,* or *dome* based on the structure and location of the stadium
- → **Stadium Surface:**   This describes if the stadium field is *grass* or *field turf*
- → **Latitude:**   The latitudinal location of the stadium
- → **Longitude:**   The longitudinal location of the stadium
- → **Elevation:**   The number of feet above sea level that the stadium resides

The third, title **spreadspoke_scores.csv** has the following columns:

- → **Schedule Date:** The date in which the game was played
- → **Schedule Season:** The season in which the game was played
- → **Schedule Week:** The week within the season in which the game was played
- → **Schedule Playoff:** Whether or not the game was an NFL Playoff game
- → **Team Home:** Which team that was classified as the *Home Team*
- → **Score Home:** The number of points scored by the team that was the Home Team
- → **Score Away:** The number of points scored by the team that was the Away Team
- → **Team Away:** The team that was classified as the *Away Team*
- → **Team Favorite:** The *Team Id* of the team that was favored to win
- → **Spread Favorite:** The number of points by which the *Team Favorite* was favored to win
- → **Over Under Line:** The number of points that were predicted to be scored by both teams when aggregated
- → **Stadium:** The name of the stadium in which the game was played
- → **Stadium Neutral:** A binary variable denoting whether or not the game was was played in a stadium in which neither team was considered the *Home Team*
- → **Weather Temperature:** The temperature in which the game was played
- → **Weather Wind Mph:** The wind speed in which the game was played
- → **Weather Humidity:** The humidity in which the game was played
- → **Weather Detail:** Whether or not the game was played in a *DOME* (null indicates that the game was not played in a *DOME*)
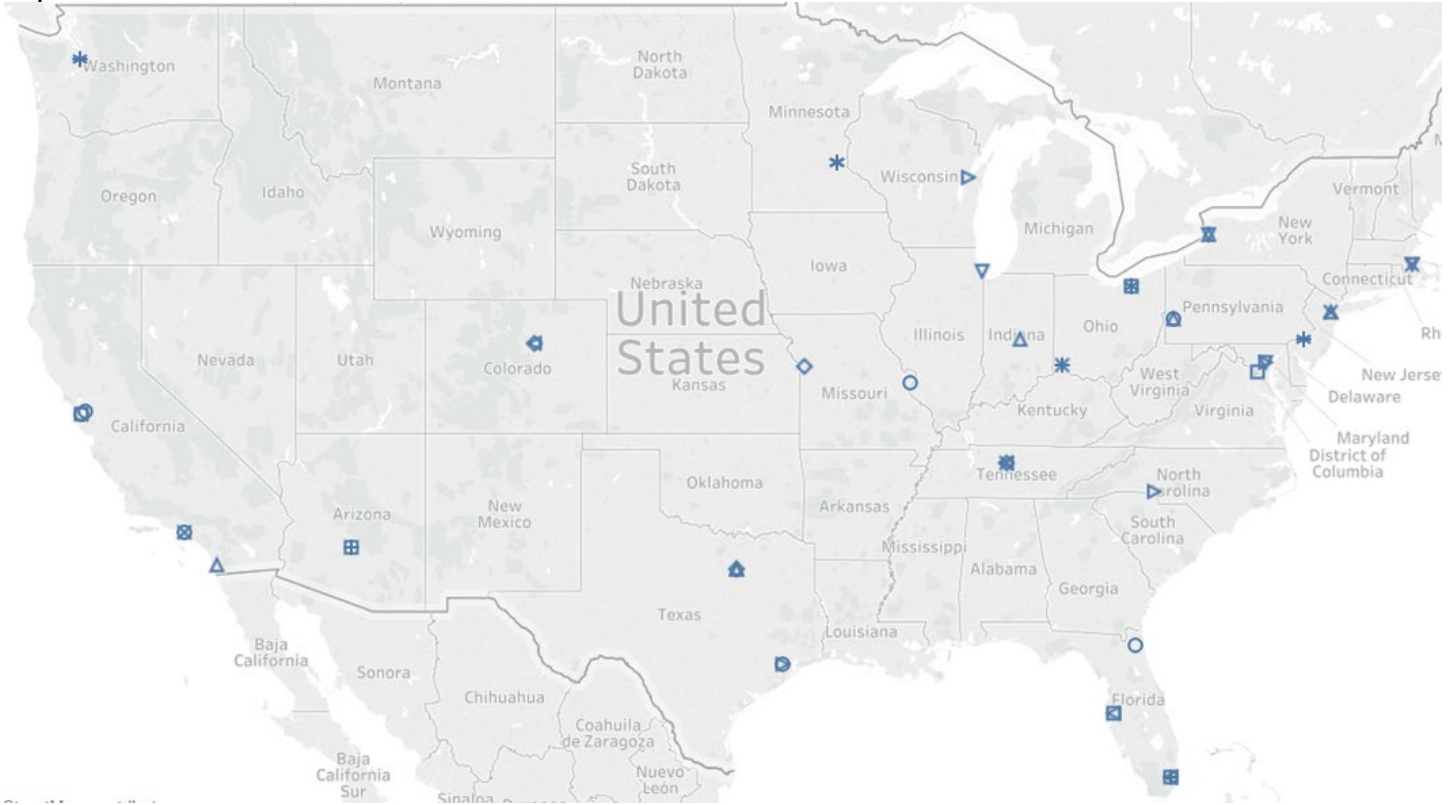
EXPLORATION

The first step in exploring these datasets was to understand the composition of the NFL. First, we examined the breakdown of Conference affiliation by the various teams that comprise the NFL. Below we can see the Conference to which each team belongs. The first thing that attracts our attention is that there are multiple teams listed with the same City or State name listed under *Team Name*. This indicates that there may have been teams that changed names, as we wouldn't expect that Baltimore, Maryland has two NFL team.

Teams by Conference

Team Name	NFC	AFC
Arizona Cardinals	✓	
Atlanta Falcons	✓	
Baltimore Colts		✓
Baltimore Ravens		✓
Boston Patriots		✓
Buffalo Bills		✓
Carolina Panthers	✓	
Chicago Bears	✓	
Cincinnati Bengals		✓
Cleveland Browns		✓
Dallas Cowboys	✓	
Denver Broncos		✓
Detroit Lions	✓	
Green Bay Packers	✓	
Houston Oilers		✓
Houston Texans		✓
Indianapolis Colts		✓
Jacksonville Jaguars		✓
Kansas City Chiefs		✓
Los Angeles Chargers		✓
Los Angeles Raiders		✓
Los Angeles Rams	✓	
Miami Dolphins		✓
Minnesota Vikings	✓	
New England Patriots		✓
New Orleans Saints	✓	
New York Giants	✓	
New York Jets	✓	
Oakland Raiders		✓
Philadelphia Eagles	✓	
Phoenix Cardinals	✓	
Pittsburgh Steelers		✓
San Diego Chargers		✓
San Francisco 49ers	✓	
Seattle Seahawks	✓	
St. Louis Cardinals	✓	
St. Louis Rams	✓	
Tampa Bay Buccane..	✓	
Tennessee Oilers		✓
Tennessee Titans		✓
Washington Redskins	✓	

First, to examine this further, we look at a nationwide map of the stadium locations. Here we can see that there are some cities, Los Angeles and Baltimore included, in which the location changed. We'll explore those below.

Looking further into the city redundancy, as described above, we see that some may be explained by changes in stadium locations. For example, below we see that the city of Baltimore has had two stadiums under different Stadium Names.