Group 3: Emmanuel Presley, Damarje Brown, Joshua Hale, Raheem Yusuff

Project Proposal

1. **What is your data? Why did you choose this data set?**
    a. NFL Stadium Attendance: https://www.kaggle.com/datasets/sujaykapadnis/nfl-stadium-attendance-dataset
    b. NFL Arrests: https://www.kaggle.com/datasets/washingtonpost/nfl-arrests
    c. We chose this data set because we all enjoy football and want to know the safest game to attend.

2. **Inspiration - what other public analysis has been done on this data set?**
    a. A vast amount of data has been researched as to the high number of arrests at NFL games that seem to be steadily increasing over time. These statistics are also being reported by major news agencies.
        i. https://bleacherreport.com/articles/921800-how-many-arrests-does-each-team-have-over-the-last-decade
        ii. https://www.sportsbookreview.com/news/most-dangerous-nfl-stadiums/
        iii. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2814990/

3. **What are some possible research questions?**
    a. Is there any correlation between the total attendance of an NFL game and the number of arrests on any given game day at that stadium?
    b. Which stadium is the most dangerous (which stadium has the most arrests)?
    c. Is there any correlation between score gap and total number of arrests on that game day?
    d. How strong is home field advantage (correlation between home attendance and home team wins)?
    e. What game day of the week has the most/least arrests per stadium?
    f. What game day of the week has the best attendance per stadium?
    g. What time of the day of games has the most arrests?
    h. Does stadium type (dome or no dome) have any relationship with arrests?

4. **What is something you might be able to predict with a regression?**
    a. Is there a correlation between attendance and total number of arrests?
    b. Is there any correlation between score gap and total number of arrests on that game day?
5. **What will be your color palette?**
    a. We will use the colors of each team.
6. **What are some potential visualizations you will make?**
    a. We will use: scatter plots, bar graphs, histograms, pie graph to visualize our data.
7. **Roles and Responsibilities.**
    a. Raheem:
        i. Is there any correlation between the total attendance of an NFL game and the number of arrests on any given game day at that stadium?

ii. Which stadium is the most dangerous (which stadium has the most arrests)?

b. Emmanuel:
    i. Is there any correlation between score gap and total number of arrests on that game day?
    ii. How strong is home field advantage (correlation between home attendance and home team wins)?

c. Joshua:
    i. What game day of the week has the most/least arrests per stadium?
    ii. What game day of the week has the best attendance per stadium?

d. DJ:
    i. What time of day games has the most arrests?
    ii. Does stadium type have any relationship with arrests?

Future analysis:

1. If we had more time, we could do an analysis comparison on temperature and number of arrests relationship.