Building a Platform for NFL Data Insights

Anish Ari, Abhinav Arun, Tyler Pomposelli, Brad Powell, Yitong Qian (Mentor: Jeremy Abramson)

Introduction/Motivations:

The realm of data in sports, particularly within the NFL, is experiencing exponential growth. NFL teams have progressively incorporated data-driven decision-making to enhance their chances of success, prompting fans to develop a heightened fascination with understanding and analyzing these very statistics. Although various online platforms such as Stathead offer access to essential NFL statistics they often suffer from intricate navigation and limited accessibility, creating a need for a more user-friendly way of accessing data.

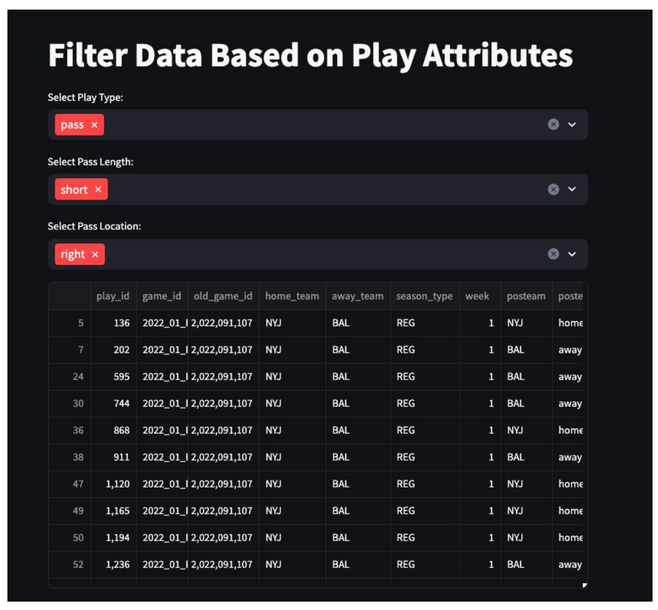
Aims:

Build a versatile NFL data processing dashboard, similar to what many of the professional teams use internally, to provide relevant data analysis, visualizations, and insights to interested individuals

Methods:

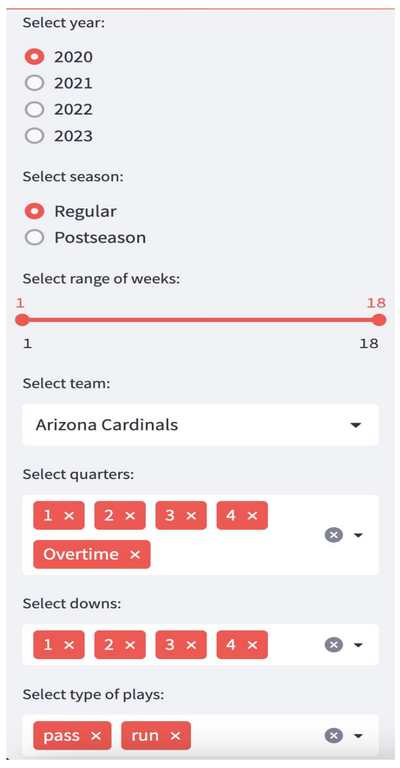
1. Collect NFL statistics from open source in csv format
2. Use library like pandas to process the file
3. Use streamlit library as a dashboard to display the data

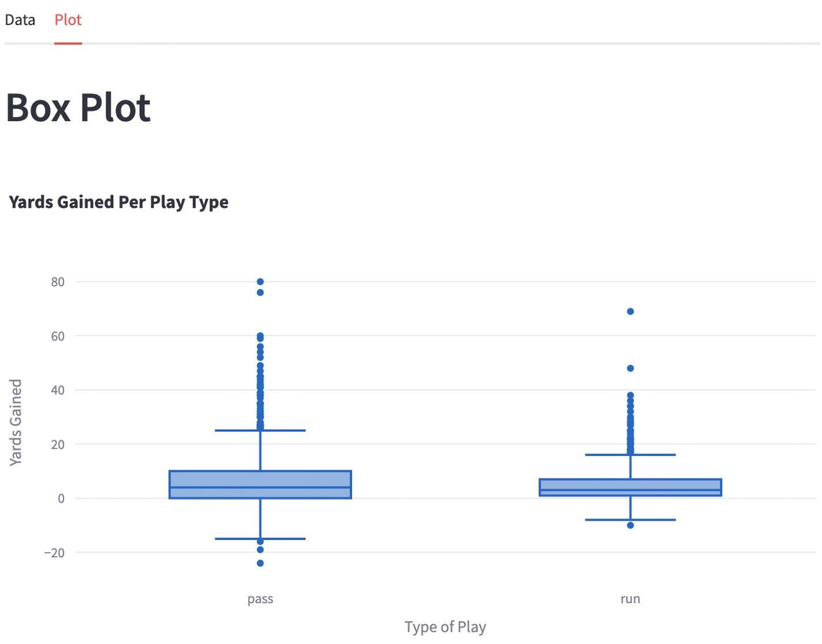
```python
import streamlit as st
import pandas as pd
import numpy as np

# Load the data
data = pd.read_csv('play_by_play_2022.csv')

def filter_data():
    # Selections
    play_types = st.multiselect('Select Play Type:', ['pass', 'run'], default=['pass'])

    # Dynamic selection based on play type
    selections = {
        'play_types': play_types
    }

    if 'pass' in play_types:
        pass_lengths = st.multiselect('Select Pass Length:', data['pass_length'].dropna().unique())
        pass_locations = st.multiselect('Select Pass Location:', data['pass_location'].dropna().unique())
        selections['pass_length'] = pass_lengths
        selections['pass_location'] = pass_locations

    if 'run' in play_types:
        run_locations = st.multiselect('Select Run Location:', data['run_location'].dropna().unique())
        run_gaps = st.multiselect('Select Run Gap:', data['run_gap'].dropna().unique())
        selections['run_location'] = run_locations
        selections['run_gap'] = run_gaps

    masks = {}
```



Data or Results:

- Analyzed NFL play-by-play data
- Dataset contains a total of 372 columns
- Dashboard filters data based on user-selected inputs (i.e. widgets)





season	season_type	week	game_date	home_team	away_team	qtr	down	play_type	yards_gained
2020	Regular	1	2020-09-13	San Francisco 49ers	Arizona Cardinals	1	1	pass	5
2020	Regular	1	2020-09-13	San Francisco 49ers	Arizona Cardinals	1	1	run	14
2020	Regular	1	2020-09-13	San Francisco 49ers	Arizona Cardinals	1	1	run	2
2020	Regular	1	2020-09-13	San Francisco 49ers	Arizona Cardinals	1	2	run	-6
2020	Regular	1	2020-09-13	San Francisco 49ers	Arizona Cardinals	1	3	pass	11

Next Steps:

- Continue adding selection widgets
- Add additional graphs and analysis
- Integrate with LLM for English-language queries