Name: Aditya Agashe
USC ID: 8589926108
GitHub: AdityaAgashe6

Final Project Report

1. Introduction

1) What is the name of your project?
The name of my project is The LFC Analyzer

2) Please write it as a research question and provide a short synopsis/description.
The purpose of this project is to examine Liverpool Football Club's various statistics across the various competitions they play in during the 2024 season. Additionally, we have a comparison between their 2022 and 2024 goal-scoring statistics. The scoring trends of each Liverpool player, the anticipated number of goals for Liverpool in each competition, and overall Liverpool trends can all be analyzed and comprehended with the help of these statistics. These analyses can be used by Liverpool's coaches and analysts to better improve the team through strategy development and team meetings before and after games.

3) What is the research question(s) that you are trying to answer?
Here are the Liverpool Football Club's goal-scoring trends for the 2024 season along with their comparison to the 2022 season. How would these insights inform strategic improvements for the 2024 season?

2. Data Collection
What type of data did you collect?

1. Specify exactly where data that you collected is coming from.
The data for this project is being collected from three different data sources. The first source is API (API Sports) from which I have used API Football. The second source is web scraping for which I have scraped data from FBREF.com. The third source is a SQL Database and it has a record of the Liverpool player stats in terms of goals scored and assists.

2. Describe the approach that you used for data collection.
For source 1, I selected API Football from the available API options and filtered the team as Liverpool which had an ID as 40. I generated the

secret API key required for fetching the data from the API. Using the requests library the data is fetched and is stored in a JSON file. For source 2, I used the BeautifulSoup library to get data from FBREF.com website. I further parsed the html to get the necessary information and added some basic data processing steps required to capture the data. For the third source, player information and statistics are stored in SQL databases using SQLite3. We created 2 tables for 2024 and 2022 seasons.

3. **How many different data sources did you use?**
   I used 3 data sources.

4. **How much data did you collect in total? How many samples?**
   The collected sample is specific to Liverpool Football Club for the year 2024 and 2022. The total number of collected samples is therefore 195 samples.

5. **Describe what has been changed from your original plan as well as the challenges that you encountered and resolved.**
   Initially, the plan focused only on Liverpool's 2024 season performance. However, the project now includes a comparative analysis of goal-scoring statistics between the 2022 and 2024 seasons to provide more context and insights into trends over time. Historical data from the 2022 season was harder to retrieve in formats compatible with the 2024 season data. This was later resolved.

3. **Data Cleaning**
   1) **Describe how and why you had to change the data before using it.**
      The data fetched from the API was in JSON format, which was unstructured. In order to analyze the data I had to structure and clean the data in order to derive useful insights.

   2) **What kind of format was the data in when you collected it?**
      The data fetched through API was in JSON format. Web scraping data was embedded in HTML, text etc. The SQL data was in the form of tables on the web.

   3) **What kind of format did you use for the processed data in the end?**

The data fetched through API was later processed to a csv file, the data fetched through web scraping was in the data frames format after processing and SQL data in the form of tables.

4) Did this process affect the data that you collected in any way, shape or form?
It enhanced the quality of data, making it easier to process and analyze the data.

4. Data Analysis
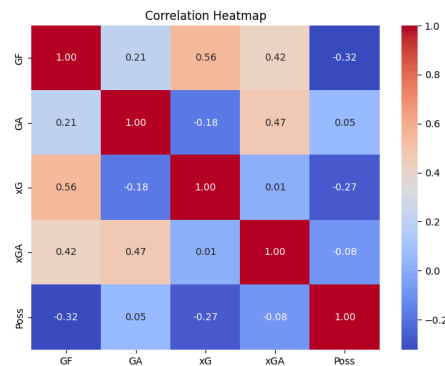**What kind of analysis and visualizations did you do?**
1) Which analysis technique(s) did you use? What are your findings?
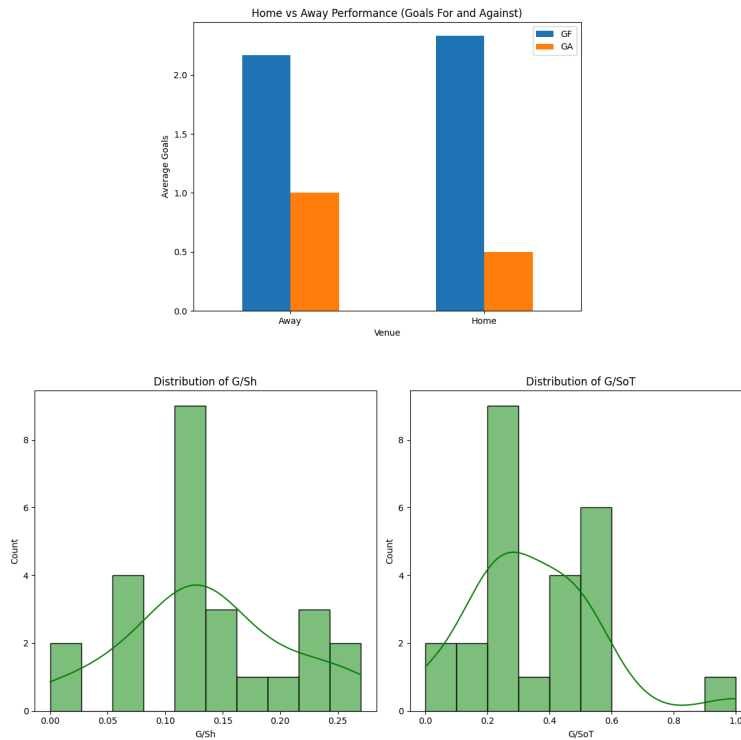The analysis techniques used are numerical columns such as POSS, GA, GF etc, correlation matrix for halftime and fulltime goals. Checking whether the number of matches with goals scored greater than 5 is threshold. Scoring trend analysis using grouping. Calculation of win loss ratio using count of values.
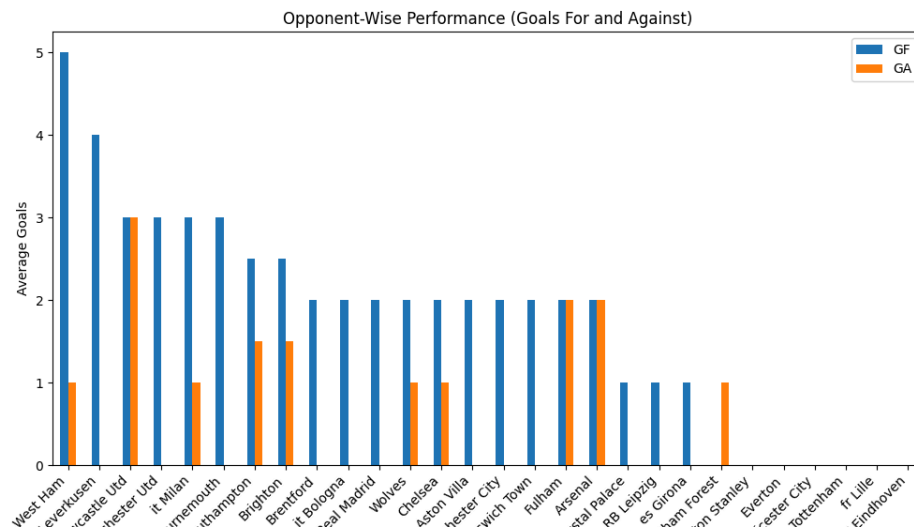
5. Visualizations
1) Describe the type of data visualizations that you used.
To illustrate the findings I used different bar graphs, histograms, box plots, pie charts and other diagrams.

Correlation Heatmap

	GF	GA	xG	xGA	Poss
GF	1.00	0.21	0.56	0.42	-0.32
GA	0.21	1.00	-0.18	0.47	0.05
xG	0.56	-0.18	1.00	0.01	-0.27
xGA	0.42	0.47	0.01	1.00	-0.08
Poss	-0.32	0.05	-0.27	-0.08	1.00

Home vs Away Performance (Goals For and Against)

Distribution of G/Sh

Distribution of G/SoT

2) Explain the setup and meaning of each element in the visualization.
Let's look at one such example


Opponent-Wise Performance (Goals For and Against)

In this chart, it tracks the Opponent Wise Performance. GF and GA stand for Goals For and Goals against. From this chart we can understand that in some matches goals were not scored by any of the teams. Understanding and finding the solution would be a crucial step for the team analysts.

6. Conclusion

1) Describe your observations and conclusion.

From this project, we can conclude that the findings would be a starting point for the team analysts to analyze the performance of Liverpool in the 2024 season, with regards to the goals scored, possession kept or penalties scored. So we can conclude that based on Figure 12, Liverpool score more than the average goals in the premier league whereas during the champions league they score less than the average.

Performance (GF) by Competition and Venue
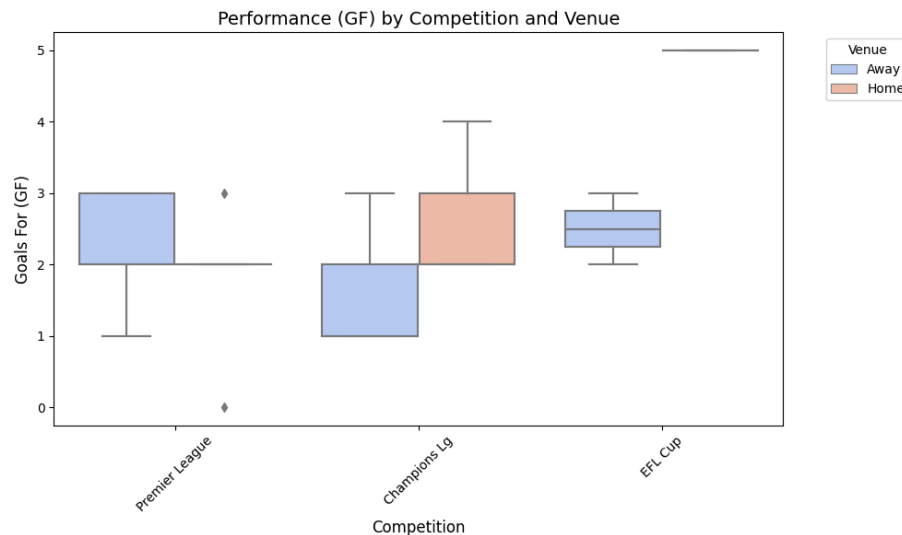
2) Describe the impact of your findings.

Analyzing and understanding the performance of Liverpool and helping Liverpool to better their game by learning from this analysis.

7. Future Work

1) Given more time, what would you do in order to further improve your project?

If given more time, I would love to have analyzed all teams competing in the Premier League so as to have a comparative analysis of all teams against Liverpool.

2) Would you use the same data sources next time? Why yes or why not?

Yes, and would search for more data sources for a more comprehensive analysis.