Data in the Beautiful Game

BAX 431 Data Visualization

Team Project: Visual Analysis

Overview

- Our Team
- Topic & Scope
- Data
- Challenges
- Key Findings
- Conclusion
- Visualization Demo
- Appendix

Our Team

- Arjun Remeshkumar Nair
- Daoxin (Alex) Wang
- Rahul Rajput
- Shaolong (Fred) Xue
- Vamsee Krishna Reddy Narahari

Project Topic

English Premier League (EPL)

Team Performances & Transfers

Project Context

- Football is a beautiful game. Outcomes are often unpredictable.
- EPL is arguably the most competitive league. Thus, the most unpredictable as well. It is also the most heavily invested.

Leicester City winning the 2015/16 season despite 5000:1 odds

Figure 7: net balances of transfers, big-5 leagues
€ million

	ENG	ITA	ESP	GER	FRA
2010	-263	-87	+23	-29	+18
2011	-311	-100	-104	-3	-17
2012	-338	+15	+46	-124	-59
2013	-568	-36	+65	-74	-141
2014	-508	-38	-6	-134	+31
2015	-720	-197	-156	+38	+100
2016	-1,038	+69	-19	-188	+151
2017	-767	-146	-62	-99	-209
2018	-1,213	-268	-220	+12	+333
2019	-762	-430	-367	-157	+152
Total	-6,488	-1,218	-800	-758	+359

https://football-observatory.com/IMG/sites/mr/mr47/en/

Project Context

- Can we find patterns among the uncertainty using data?
- Can patterns be turned into actionable insights for stakeholders?

Team & Player Level Stats
Transfer Activities
Seasonal Trends

Project Context

- Users: managers and club executives

- Goals:
  - Use insights to inform team line-up, tactics, transfer targets, etc.
  - Ultimately, teams can move higher up the standing table.

Data

- EPL Match Level Data:
  - https://www.kaggle.com/datasets/irkaal/english-premier-league-results
- EPL Player Level Statistics:
  - https://fbref.com/en/comps/9/stats/Premier-League-Stats#all_stats_standard
- EPL Transfer Data:
  - https://flatgithub.com/ewenme/transfers?filename=data%2Fpremier-league.csv&sha=1c9461c54fde878eb154ebdf631a96cc706fe8bc
- Team Level Data:
  - https://fbref.com/en/comps/9/2021-2022/stats/2021-2022-Premier-League-Stats

Challenges

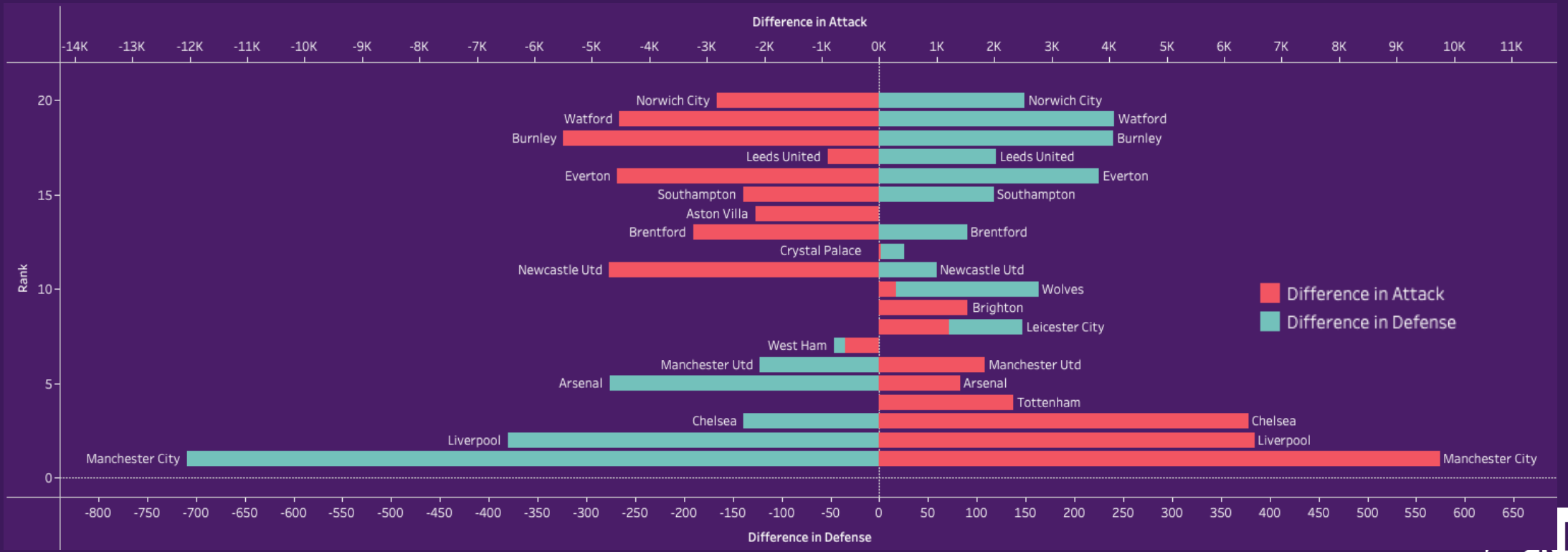
- Data Cleaning
- Data Standardization
  - For intuitive interpretation. E.g. “Away” to “Away Win”
  - Structuring
- Data Collation
  - Multiple sources

Key Findings – The Pandemic Effect

- Home team advantage completely disappeared during the 2020/21 season.

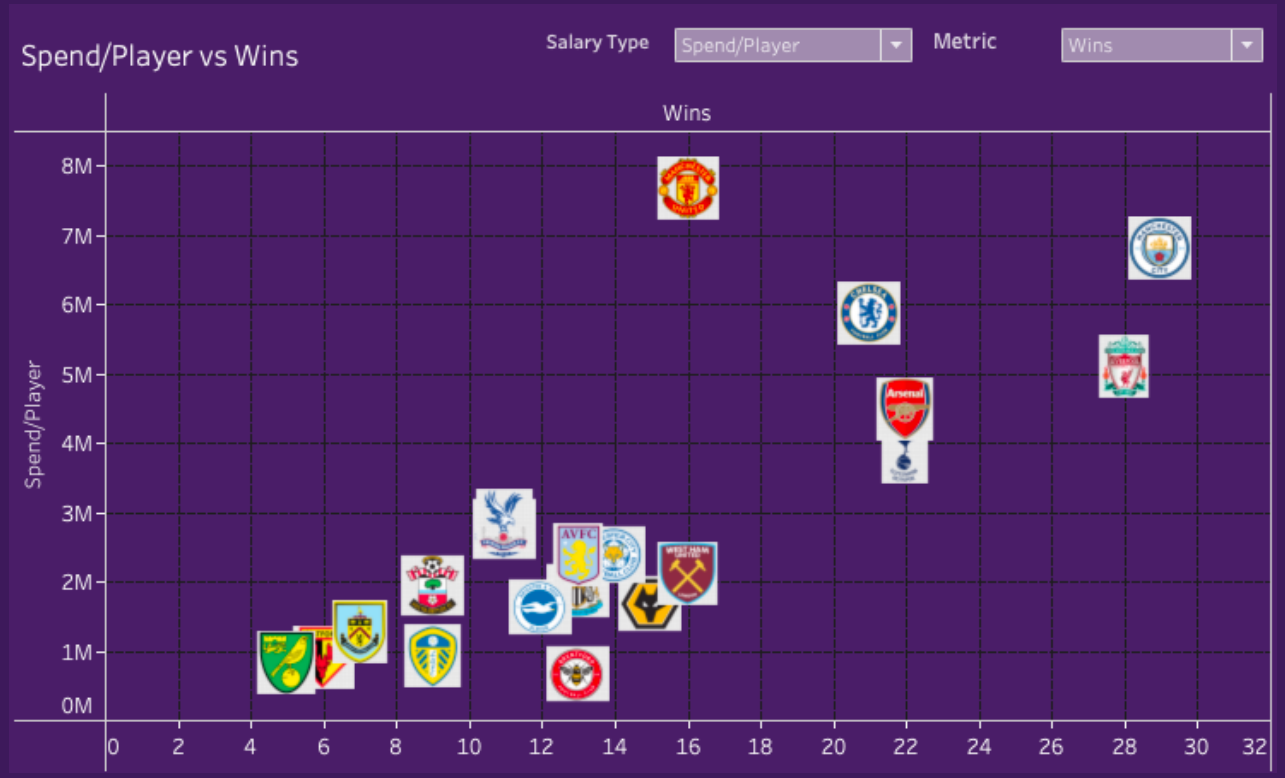
Key Findings – Is Sir Alex Ferguson Right?

- Top teams are better than average, more so at defending than at attacking.



Key Findings – Inflated Salary

- Player salaries are inflated based on team's perceived status instead of league performance



Main Conclusion

Executive Summary	• High-level overview on key metrics, across teams and seasons
Performance vs. Spending	• Analyze transfer spending against performance metrics
Performance Breakdown	• Home vs. away • Qualification and relegation • Half-Time vs. Full-Time

Premier League

Visualization Demo

- Let’s head to the Tableau dashboards

Appendix

- EDA Hypothesis
  - Home vs. Away
  - Attack vs. Defense
  - Positions in Transfers

Hypothesis 1: Home vs. Away

Teams playing at their home stadium have a higher chance of winning

- Description
  - For many fans mind, the home stadium advantage is the team's 12th player.
  - The passion of fans will affect the performance of home stadium team
- What we done
  - Created a table showing the number of wins for each season's home team and away team
  - Refined the table to Boxplot for easy to understand

Hypothesis 1: Home vs. Away

- Every season teams won more matches at their home ground.
- The 2020-21 season was an exception
  - Because of the COVID-19, no audiences allowed in the stadiums
- The hypothesis is correct

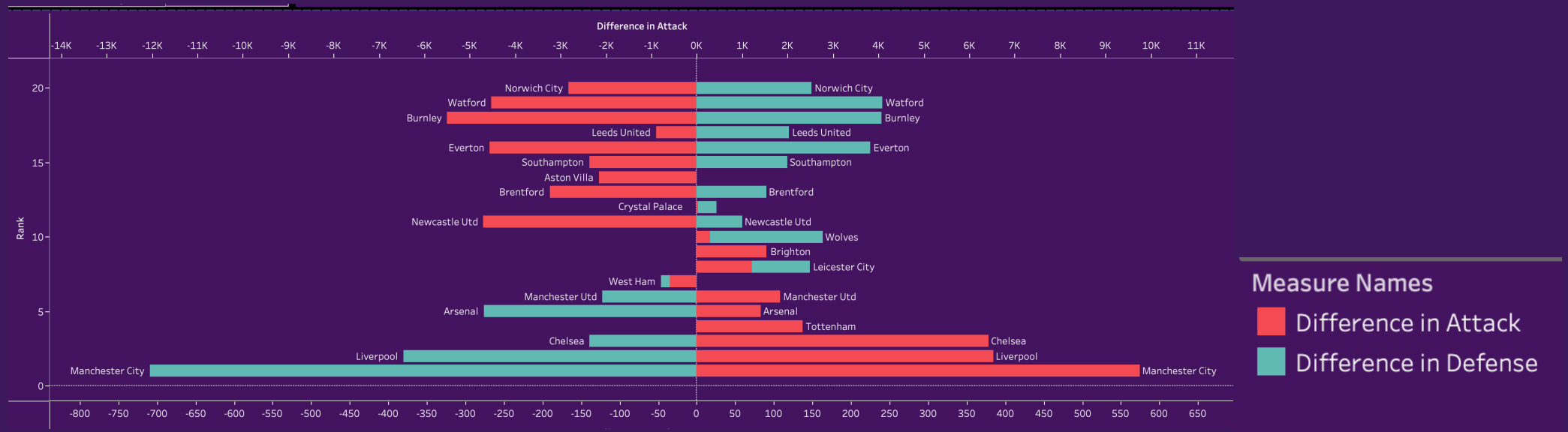
Hypothesis 2: Attack vs. Defense

- **Teams with better Defensive capabilities rank better in a season**

- "Attack wins you games. Defence wins you titles."

Hypothesis 2: Attack vs. Defense

- We can see the gap between top teams and bottom teams is larger in defense, than in attack.



Hypothesis 3: Positions in Transfers

Attacking players are more sought and expensive than defenders.

- Description:
  - We want to look at how transfer activities for different positions have evolved over time.
  - In recent years, attacking players top the charts with record transfer fees.

- What we've done:
  - We grouped players by positions (attacker, midfielder, and defender) and plotted the number of transfers by position over time.

Hypothesis 3: Positions in Transfers

- Defenders had consistently more transfer volumes over the years, contrary to our belief.

- However, attacker position tallied the highest transfer expenditure over the years.

- This indicates that clubs tend to buy defenders ‘in bulk’ while pursuing a few expensive attackers to supplement the squad.