



Sports Statistics Analysis

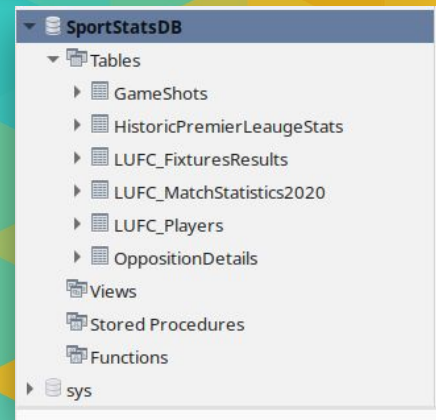
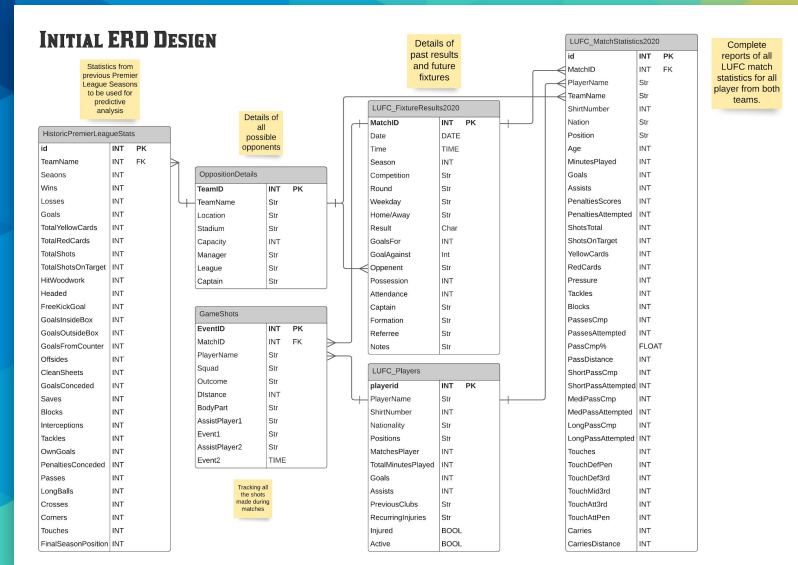
Andrew Ashdown
December 2020

Project Outline

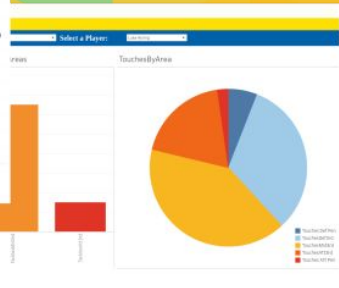
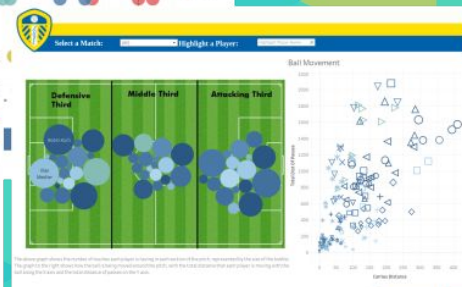
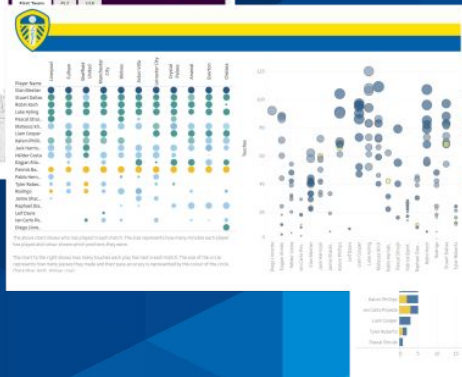


Data Collection and Database

- Data Scrapping
 - FBREF.com
- Kaggle Datasets
 - User:Zaeemnalla
- ERD Design
- MySQL



Data Visualisation

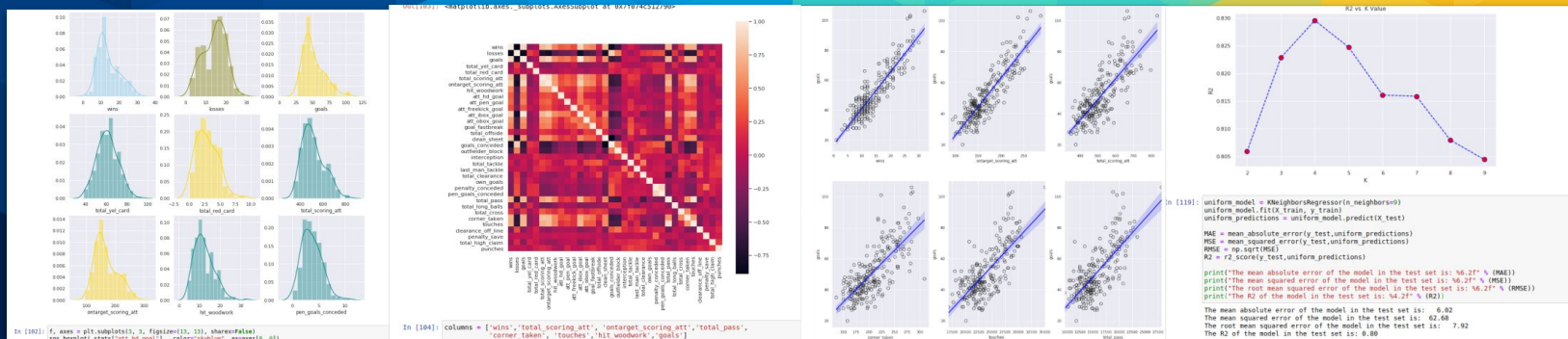


[Click here for to be taken to the Tableau Dashboard.](#)

Statistical Analysis

Can we accurately predict the number of goals that will be score in a season?

- Selecting features
- Different Regression Models
- Scikit-Learn Linear Regressor - 88% accuracy



Hypothesis Testing

With the introduction of VAR, will we see a significant increase in the amount of penalties awarded this season?

- Null Hypothesis Testing
- Confidence Level: 95%
 - T value: 2.09
 - Z score: 3.18
- Confidence Level: 99%
 - T value: 2.86
 - Z score: 3.18

$$t = \frac{(\bar{X} - \mu_0)}{\hat{\sigma} / \sqrt{n}}$$

where:

\bar{X} is the sample mean

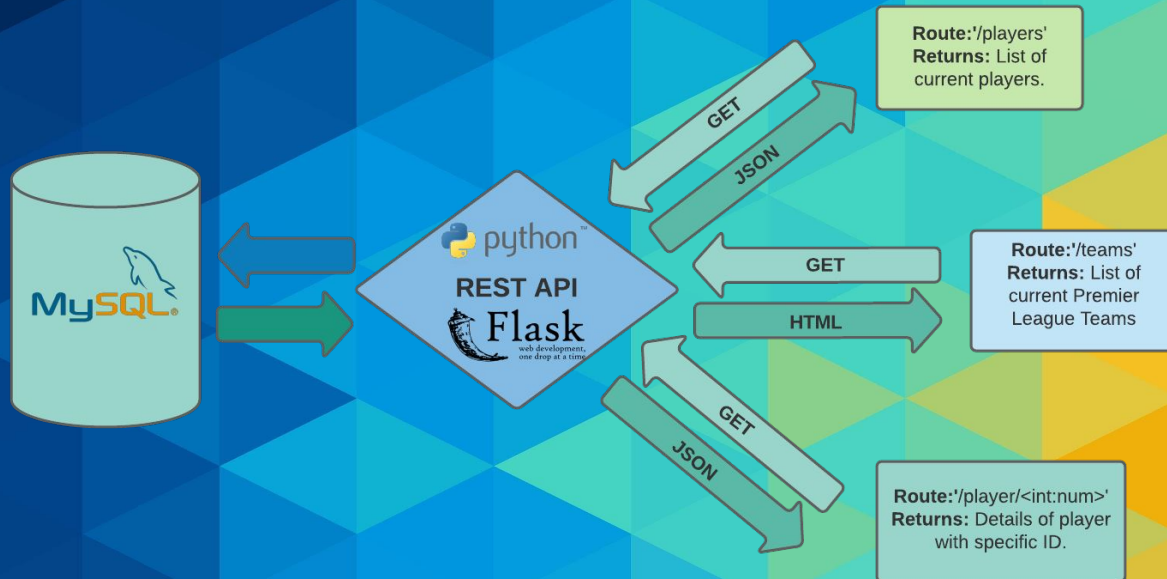
μ_0 is the population mean

$\hat{\sigma}$ is the sample standard deviation

n is the number of measures in our sample

Delivering Insight

- Create a tool to deliver insight to a user
 - Built Restful API in Python, using the Flask library.
 - Connect to MySQL, queries database, delivers results.



Review

- Created a number of tools that deliver useful insight
 - Will continue to develop further
- Answered several statistical questions
- Learnt lots along the way