



Arsenal FC :Dawn of a New Era

Submitted in Partial Fulfillment of Requirements
for the Degree of
Bachelor of Data Science

By

Sujal Suresh Shetty

Seat No: 31013022031

Guide

Ms. Minal Dive



S K Somaiya College

Somaiya Vidyavihar University

Vidyavihar, Mumbai - 400 077

2022-25

Somaiya Vidyavihar University

S K Somaiya College

Certificate

This is to certify that the project report on dissertation entitled _____ is
bonafide record of the dissertation work done by _____
in the year 2024-25 under the guidance of _____,
Department of Information Technology and Computer Science in partial
fulfillment of requirement for the Bachelor of Data Science degree of Somaiya
Vidyavihar University.

Guide / Co-Guide

Programme Coordinator

Head of Department

Director

Date:

Place: Mumbai-77

Bachelor of Data Science

2022-25 Batch

Somaiya Vidyavihar University

S K Somaiya College

Certificate of Approval of Examiners

This is to certify that the project report on dissertation entitled _____ is
bonafide record of the dissertation work done by _____ in
partial fulfillment of requirement for the Bachelor of Data Science degree of
Somaiya Vidyavihar University.

External Examiner /Expert

Internal Examiner/ Guide

Date:

Place: Mumbai-77

Somaiya Vidyavihar University

S K Somaiya College

DECLARATION

I declare that this written report submission represents the work done based on my and / or others' ideas with adequately cited and referenced the original source. I also declare that I have adhered to all principles of academic honesty and integrity as I have not misinterpreted or fabricated or falsified any idea/data/fact/source/original work/ matter in my submission.

I understand that any violation of the above will be cause for disciplinary action by the college and may evoke the penal action from the sources which have not been properly cited or from whom proper permission is not sought.

Signature of the Student

Name of the Student

Seat No.

Date:

Place: Mumbai-77

Abstract

This project presents a comprehensive analysis of Arsenal FC's 2018-2020 seasons, focusing on player and team performance metrics using data analytics and visualization techniques. The main objective of this study is to identify key performance indicators (KPIs), compare player contributions, and analyze season-over-season trends. Data from the 2018-19 and 2019-20 seasons were collected and visualized using Tableau, with additional statistical analysis performed through MySQL and Python-based linear regression models.

Literature survey conducted on football data analytics shows that significant research has been done in understanding player performance metrics, predicting match outcomes, and improving team strategies. The sports industry, particularly football, has seen a rise in the use of data analytics to enhance both individual and team performance. This growing field provides valuable insights into decision-making for team management and performance optimization.

In this report, an exhaustive study on Arsenal FC's performance over two seasons is presented. The analysis focuses on key metrics such as goals, assists, tackles, passes, and defensive contributions, both at the individual player level and across the team. Furthermore, predictive models were developed to forecast player performance based on historical data, helping to understand critical factors influencing success on the field. The findings offer a data-driven perspective on football performance analysis, contributing to the evolving field of sports analytics.

Key words: Football analytics, player performance, data visualization, Tableau, linear regression, KPIs, season analysis, Arsenal FC, predictive modeling, sports statistics.

Contents

List of Figures			
List of Tables			
Nomenclature			
1	Introduction		
	1.1	Overview	6
	1.2	Background	6
	1.3	Scope of the project	7
	1.4	Technology	8
2	Literature Survey		
	2.1	How football clubs use data for performance evaluation	11
	2.2	Use of Tableau in sports analytics	12
	2.3	Predictive modeling in football	13
	2.4	Aim	14
3	Mathematical Background and fundamental concepts		
	3.1	Key elements	16
	3.2	Statistical Methods	20
4	Implementation of Issue 1 (2018-19 Season Analysis)		
	4.1	Introduction to the 2018-19 Season Data	28
	4.2	Tableau Visualization for 2018-19 Season	31
5	Implementation of Issue 2 (Predictive Analysis)		
	5.1	Data Preparation	44
	5.2	Importing files to sql	49

	5.3	Data modification	56
	5.4	Data Export	63
	5.5	Predictive Analysis	66
	5.6	Comparative Analysis	108
6	Conclusion		
	6.1	Key Takeaways	118
	6.2	Future Work	119
References			121
Appendix			124

List of Figures

Sr. No.	Name of the Figure	Page No.
3.2.1	Scatter plot for linear regression	21
3.2.2	Graph for linear regression	22
3.2.3	Train test split procedure	24

List of Tables

Sr. No.	Name of the Table	Page No.
3.2.1	Change seen in a dataframe/set due to Train test split method	26

Nomenclature

Y_i	Dependent variable
β_0	<i>Intercept</i>
β_1	Slope
X_i	Independent variable.

ε_i	Residual or Random Error
MSE	Mean Squared Error