



Cairo University

CMPS451 – Data Mining, Big Data and Data Analytics
Spring 2024

Project Proposal

Submitted to:

Prof. Lydia Wahid

Eng. Ahmed Yehia

Eng. Arwa Hammuda

Eng. Hanna Hatem

Eng. Youssef Mokhtar

Name	ID
Ammar Nasser	1190543
Anas Shreif	1190375
Omar Mohamed	1190477
Marwan Mostafa	1190281

Statement

The NBA Shot Logs dataset, found on Kaggle, includes detailed records of over 128,000 shots taken by NBA players throughout the 2014-2015 season. This project aims to conduct extensive data analysis on the dataset using Python and PySpark, with the objective of uncovering insights into player performance and team strategies.

Idea

Analyzing NBA Shot Logs using Python with PySpark.

Our Procedures

Data Preparation and Preprocessing	Prior to analysis, the NBA Shot Logs dataset requires cleaning and preprocessing, involving tasks such as eliminating null values, standardizing the data, and adjusting data types as needed.
Data Analysis	Prior to analysis, the NBA Shot Logs dataset requires cleaning and preprocessing, involving tasks such as eliminating null values, standardizing the data, and adjusting data types as needed.
Data Visualization	Once we've extracted the desired insights from the data, we'll proceed to visualize them using Python libraries. This visualization process aims to present the findings in a clear and compelling manner, facilitating effective communication of the results.

Expected Outcome

This project seeks to uncover insights that can help team's coaches make better decisions and improve team performance by understanding player behavior and effective shooting strategies for each player.

Dataset Link

[Click Here](#)