Virat Kohli Batting Performance – Exploratory Data Analysis Report

1. Introduction & Problem Statement

The objective of this project is to analyse Virat Kohli's batting performance using exploratory data analysis (EDA). The analysis aims to identify scoring patterns, consistency, and key factors that influence his run-making ability across different match situations.

2. Dataset Description

The dataset contains match-wise batting performance details including runs scored, balls faced, strike rate, boundaries, batting position, dismissal type, opposition, ground, and match date. Each row represents a single innings played by Virat Kohli.

3. Data Cleaning Summary

The dataset was inspected for missing values and duplicate records. No null values or duplicate rows were found. Date columns were converted to datetime format, and numerical columns were cast to appropriate data types to ensure accurate analysis.

4. Key Visualizations with Explanation

Various visualizations were used to understand performance trends. Univariate plots highlighted distributions of runs, strike rate, and boundary contribution. Bivariate plots explored relationships such as runs versus balls faced and runs versus strike rate. Multivariate visualizations and a correlation heatmap provided a consolidated view of performance drivers.

5. Insights & Findings

- Higher scores are strongly associated with spending more time at the crease.
- Boundary hitting supports scoring but is not the only contributor to high run totals.
- Strike rate varies based on match context and does not solely determine scoring success.
- Performance remains consistent across batting positions and varies sensibly across oppositions.

6. Conclusion

This exploratory data analysis shows that Virat Kohli's batting success is driven by consistency, adaptability, and effective innings construction rather than pure aggression. The project demonstrates how data analysis techniques can be applied to real-world sports data to extract meaningful and actionable insights.