



# CRICKET ANALYTICS



**Introduction**



**Analysis**



**Deployment**



**Conclusion**



# Introduction

- The project uses ball-by-ball data from all Indian Premier League matches played from its inception in 2008 until the 2022 season.
- It aims to analyze individual player performance at each ball for which different parameter are considered and utilized like Bowler Type, the Stadium capacity, pitch type, Ball type, Opposition player.
- The analysis is conducted on both SQL and NoSQL databases which can help us to perform a comparative study for the results achieved.

## Goal

To analyzed individual player performance based on different match conditions and demonstrate how the refined analytics can help teams make more sound off-the-field decisions.



## Motivation



# Motivation

- Cricket is one of the most popular games in the world played around 106-member states, which has 1.5 billion worldwide fans.
- We are trying to gain insights into how an individual player performs against a particular opposing team under given conditions.
- This analysis can be used further to decide which player to consider in critical match conditions in order to win the match.



# Introduction

- The project uses ball-by-ball data from all Indian Premier League matches played from its inception in 2008 until the 2022 season.
- It aims to analyze individual player performance at each ball for which different parameter are considered and utilized like Bowler Type, the Stadium capacity, pitch type, Ball type, Opposition player.
- The analysis is conducted on both SQL and NoSQL databases which can help us to perform a comparative study for the results achieved.

## Goal

To analyzed individual player performance based on different match conditions and demonstrate how the refined analytics can help teams make more sound off-the-field decisions.



## Motivation



# CRICKET ANALYTICS



**Introduction**



**Analysis**



**Deployment**



**Conclusion**



## Project Analysis

Demo

AWS



MYSQL



MongoDB



- The initial dataset is taken from kaggle, we conducted a rigorous cleaning and transformation of data.
- MySQL Workbench was used to set up the local DB.
- The conceptual data modeling was performed before setting up the database for which we create an ER digram including 23 different tables.
- The database was then setup using the designed schema.
- For the data import, we imported csv files into MySQL Workbench. We then scripted the Stored Procedures that were used to updated the derived columns in teams\_stats and players\_stats table.
- We have used SQL database to conduct our Preliminary Analysis.



## Project Analysis

Demo

AWS



MYSQL



MongoDB



- We have imported the cleaned data to MongoDB compass to perform analysis by using Aggregate Pipelines.
- With pipelines, the queries were designed to evaluate the effect of the parameters Bowling length, maximum margins and total wicket takers of all time.
- We have established a connection between Python and MongoDB using PyMongo, a built-in Python Driver that enables us to communicate with the MongoDB database.
- Following the connection, we have bulk imported the documents to the database in MongoDB Atlas for performing queries and for creating charts.
- To make some data-driven decisions, visualization has been used to compare different qualities.



## Project Analysis

Demo

AWS



MYSQL



MongoDB



## Amazon Web Services

- This project utilizes AWS S3 service as a data lake to store 23 csv files into S3 buckets.
- AWS Glue is used to perform ETL process, which ensures high quality data.
- Using AWS Redshift as a datawarehouse, we analysed our IPL\_Datase for hidden insights.

**Implementation  
of AWS**

# Implementation of AWS

- **S3:** A user may store and access any volume of data from anywhere on the internet with it , a cloud-based object storage services, utilized for website hosting, content distribution, data lakes, and big data analytics.
- **Glue:** Moving data between data stores is simple with the help of Glue It streamlines the procedure for getting data ready and adapted for analytics, machine learning, and other applications.
- **RedShift:** Amazon offers AWS Redshift, a cloud-based data warehousing service. With the aid of SQL-based tools, it is used for cost-effectively storing, querying, and analyzing massive amounts of data.
- **IAM roles:** IAM roles Make sure that only those with permission to access your AWS resources is possible through the creation and management of users, groups, and permissions.
- **Cloud watch:** It is a monitoring and logging solution. It is used to gather and monitor metrics, gather and watch over log files, and set resource use alarms. which enables system administrators and developers to learn more about the functionality and performance of their infrastructure and applications.



## Amazon Web Services

- This project utilizes AWS S3 service as a data lake to store 23 csv files into S3 buckets.
- AWS Glue is used to perform ETL process, which ensures high quality data.
- Using AWS Redshift as a datawarehouse, we analysed our IPL\_Datase for hidden insights.

**Implementation  
of AWS**



## Project Analysis

Demo

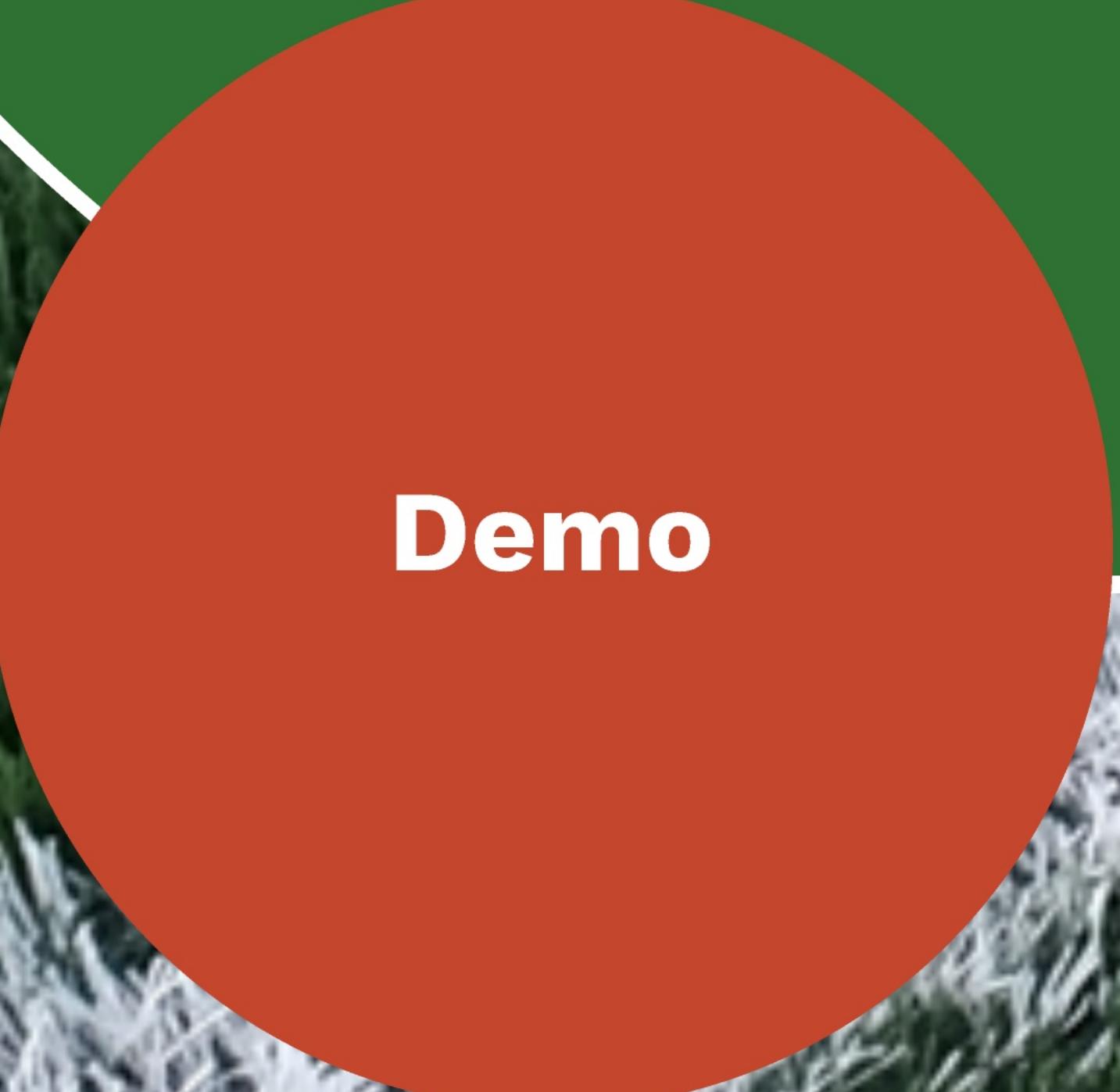
AWS



MYSQL



MongoDB



**Demo**



## Project Analysis

Demo

AWS



MYSQL



MongoDB



# CRICKET ANALYTICS



**Introduction**



**Analysis**



**Deployment**



**Conclusion**



# Visualization Dashboard

**Batsman Stats**

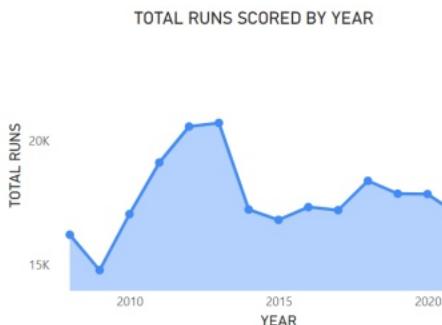
**Bowler Stats**

**Visualizations**

# Batsman Stats



BATSMEN STATS DASHBOARD



TOTAL INNINGS	<b>950</b>
STRIKE RATE	<b>126.41</b>
TOTAL FOURS	<b>34.58K</b>
TOTAL RUNS	<b>270K</b>
TOTAL BALLS FACED	<b>213.87K</b>
TOTAL SIXES	<b>10.49K</b>

SELECT BATSMEN:

SELECT STADIUM:

SELECT BOWLER:

SELECT OVER NUMBER RANGE:

SELECT YEAR RANGE

2008 2022

SELECT INNINGS:

1 2

SELECT BALL LENGTH:

SELECT BALL SPEED RANGE

70 155

BOWLING VARIATION

Fast	Leg Break Googly	Medium-Fast
Fast-Medium	Medium	Off Break
Leg Break	Medium_Slow	Slow

SELECT OPPosition:

Chennai Super Kings	Gujarat Lions	Kolkata Knight Riders
Deccan Chargers	Gujarat Titans	Lucknow Super Giants
Delhi Capitals	Kochi Tuskers Kerala	Mumbai Indians



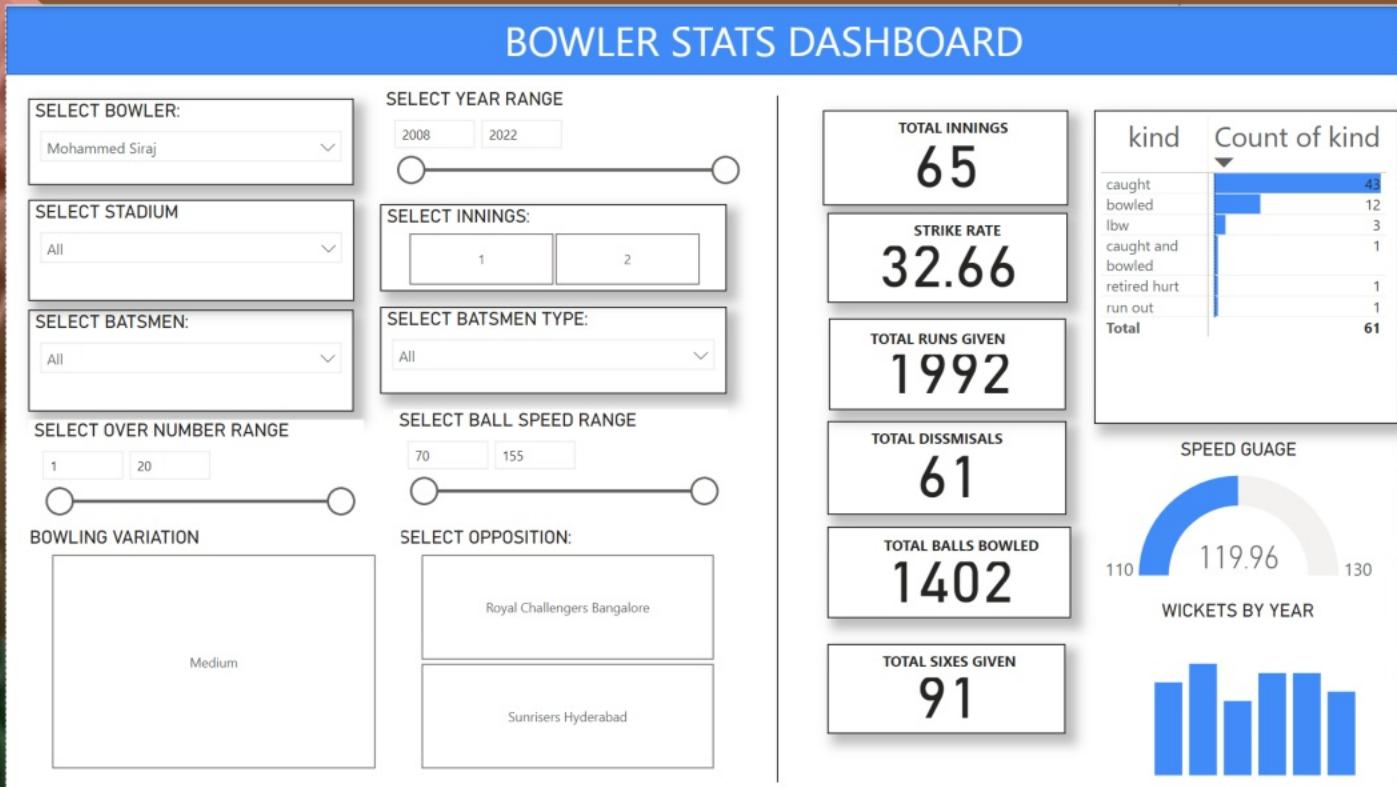
# Visualization Dashboard

**Batsman Stats**

**Bowler Stats**

**Visualizations**

# Bowler Stats





# Visualization Dashboard

**Batsman Stats**

**Bowler Stats**

**Visualizations**

# Visualizations Dashboard





# Visualization Dashboard

**Batsman Stats**

**Bowler Stats**

**Visualizations**



# CRICKET ANALYTICS



**Introduction**



**Analysis**



**Deployment**



**Conclusion**

# Conclusion



Our analysis has highlighted the significance of numerous crucial elements, including strike rates, economy rates, ball speed and bowling length. We have also identified additional variables that affect match outcomes, including the opposition, pitch, stadium and team makeup. We have discovered factors that have a substantial impact on individual and team performance by studying a range of cricketing statistics, which has given us invaluable insights into the game of cricket.



**Technical Difficulties**



**Significance to the Real World**



**Key Learnings**

# Technical Difficulty

- i) In contrast to relational databases, which allow us to link two tables, MongoDB does not allow this. The joining of the saved documents was an intricate task. subsequently is not impossible because we can manually code it, but it took a long time, which ultimately decreased performance.
- ii) Tracking players' performances across teams during IPL seasons has been difficult because it requires extensive data analysis. To track the movement of players between teams, we looked at information from a variety of sources, including official IPL records, team websites, and sports websites. The information was then manually entered into the necessary attributes.
- iii) During the AWS setup, creating jobs required changing the default data types that the system assigned and even while There were a lot of issues while creating VPC endpoints and assigning file paths while loading data into crawlers from buckets and assigning right IAM roles while running ETL jobs.

# Conclusion



Our analysis has highlighted the significance of numerous crucial elements, including strike rates, economy rates, ball speed and bowling length. We have also identified additional variables that affect match outcomes, including the opposition, pitch, stadium and team makeup. We have discovered factors that have a substantial impact on individual and team performance by studying a range of cricketing statistics, which has given us invaluable insights into the game of cricket.



**Technical Difficulties**



**Significance to the Real World**



**Key Learnings**



The brand value of IPL accounts up to 8.5 billion dollars at the end of the 2022 season. With each year passing the revenue it is generating is tremendous with substantial growth factor. Situations where insights are arduous to understand just by looking at the data, we can run queries those are provided in the report with its visualization, we can help admirer to understand game and augment their interest and engagement to the game, the more engaged the follower the more is the return on investment for the IPL brand.

# Conclusion



Our analysis has highlighted the significance of numerous crucial elements, including strike rates, economy rates, ball speed and bowling length. We have also identified additional variables that affect match outcomes, including the opposition, pitch, stadium and team makeup. We have discovered factors that have a substantial impact on individual and team performance by studying a range of cricketing statistics, which has given us invaluable insights into the game of cricket.



**Technical Difficulties**



**Significance to the Real World**



**Key Learnings**



# Key Learnings

- Implementing the course's concepts has involved experimenting with a variety of technologies, including PowerBI, Amazon Redshift and Glue, MongoDB Atlas, and various presentation tools like Prezi and Grammarly, all of which have an impact on our overall development.
- Implementing analytics on SQL, NoSQL and AWS made us understand the flexibility of AWS on computational power as and when required on executing complex queries and optimizing the performance with columnar storage.
- Following the Agile methodology, we made sure that our project was well-planned and that we were aware of any upcoming challenges and came up with appropriate action items in advance.
- We worked collaboratively as a team from the research to the project's development phase. As a result, we were more productive and imparted knowledge rapidly.

# Conclusion



Our analysis has highlighted the significance of numerous crucial elements, including strike rates, economy rates, ball speed and bowling length. We have also identified additional variables that affect match outcomes, including the opposition, pitch, stadium and team makeup. We have discovered factors that have a substantial impact on individual and team performance by studying a range of cricketing statistics, which has given us invaluable insights into the game of cricket.



**Technical Difficulties**



**Significance to the Real World**



**Key Learnings**



# CRICKET ANALYTICS



**Introduction**



**Analysis**



**Deployment**



**Conclusion**