PROJECT PROPOSAL

INTRODUCTION TO DATA WAREHOUSE AND DATA MINING
1.
DATA WAREHOUSE: It is a relational database that stores historic operational data across an organization, for reporting, analysis, and research. The main idea of data warehousing is to store very huge volumes of data and optimize it to support complicated, multidimensional queries by data analysts, data engineers, business analysts, and data scientists who have a deep understanding of data. In data warehousing, a data processing pipeline involves collecting data, organizing it, transforming it into a standard structure, optimizing it for analysis, and processing it.

DATA MINING: It involves deriving insights from large/complex data sets. As its name metaphorically suggests, data mining is a process of thoroughly analyzing and processing huge quantities of data to discover important patterns, findings, or correlations that might be necessary for the business. In data mining, data processing pipeline involves analyzing data to discover hidden patterns, relationships, and insights.

ABOUT THE PROJECT DATA SET

- We have selected the **IPL Complete Dataset** data set from Kaggle.com for this project.
LINK: IPL Complete Dataset.
- This data is about **IPL Ball by a ball** from the year 2008 to 2020. There will be a total of 18 explanatory variables which describes almost every ball bowled each bowler, the player who faced the ball, how much was scored for that ball which helps in keeping track of the runs scored for that particular player for that particular over.
- There are a total of **18 dimensions(Columns)** in this dataset.
- And there are a total of **193469 rows** present in the dataset.

PROJECT STEPS & TIMELINES:

Below are the steps and estimated project completion timelines to develop this application.

STEPS:

1. Planning and Identifying Data Set
2. Data Cleansing
3. Data Transfer to MS Access
4. Dimensional Modeling
5. Data Transfer to SQL Server
6. Cube Development and Deployment in Visual Studio
7. Data Mining Queries and Reports

ESTIMATED PROJECT TIMELINES:

Steps	Start Date	End Date	Duration (in days)
Planning and Identifying Data Set (Draft)	3/9/2021	3/12/2021	4
Finalizing Data Set	3/26/2021	3/26/2021	1
Data Cleansing	3/29/2021	4/1/2021	4
Data Transfer to MS Access	4/2/2021	4/6/2021	5
Dimensional Modeling	4/7/2021	4/13/2021	7
Data Transfer to SQL Server	4/14/2021	4/20/2021	7
Cube Development and Deployment in Visual Studio	4/21/2021	4/27/2021	7
Data Mining Queries and Reports	4/28/2021	5/4/2021	7

Week 1	2	3	4	5	6	7	8	9

Planning and Identifying DataSet
Mar 9 ➤ Mar 12
4 days

Finalizing Data Set
Mar 26 ➤ Mar 26
1 day

Data Cleansing
Mar 29 ➤ Apr 1
4 days

Data Transfer to MS Access
Apr 2 ➤ Apr 6
3 days

Dimensional Modelling
Apr 7 ➤ Apr 13
5 days

Data Transfer to SQL Server
Apr 14 ➤ Apr 20
5 days

Cube Development and Deployment in Visual Studio
Apr 21 ➤ Apr 27
5 days

Data Mining Queries and Reports
Apr 28 ➤ May 4
5 days