**Data world Exercise**

The following project aims to understand the working of data world and how to use its different features. The platform is easy to understand and navigate. The are multiple features available according to different needs of the users and organization. The UI is also consistent across all the screens, which makes the job easier. The most important feature that I liked was the seamless data integration with various connection managers.

Advantages: The most important thing is a friendly UI with consistency and intuitiveness. There is SQL support with help of which one can easily create queries and work on the specific parts of data as per the requirement. There is also an option for interactive chart building on the whole data as well as the queries.

Disadvantages: - It was taking a few minutes of time to load the data while doing insights. As the chart builder was an extension it might have added to the response time, that thing can be worked upon. It is also difficult to find customized help as it currently must be documented as a project.

*Fig 1: - Project Summary*

Graphical user interface, text, application

Description automatically generated

The various connection managers include Athena, Snowflake, MySQL, Oracle and many more. Here we have integrated our project with Excel for simplicity purpose. The dataset consists of football data, including the footballers age, nationality, club, wage, preferred foot and much more. With respect to the two methods learned in class, i.e., row-based and column-based formats. Column based is useful when performing queries that require only a subset of columns for analysis and row based is useful when you require access to most or all the columns and rows in the dataset. Here, we have used CSV format which is basically a tabular data using plain text. It is a row-based file format, which indicates that each row in the file is a row in the table. The csv files are splitable when it is in raw or uncompressed format.

*Fig 2:- Data Dictionaries*

Graphical user interface, application, Teams

Description automatically generated

*Fig 3: - Data Connection*

Graphical user interface, application, table, Excel

Description automatically generated

There are two queries that are run on the data:

1. Age >31
2. Preferred Foot is Right

*Fig 4: - Query for Preferred Foot = Right*

Graphical user interface, application, table

Description automatically generated

There are two insights provided, one on the entire dataset and one on the selected query:

1. Average Age of Footballers at different field positions
2. International Reputation vs Potential (Filter applied for Query 2 above)

*Fig 5: - Insight for International Reputation vs Potential for right foot players*

Table

Description automatically generated with medium confidence