**FINAL PROJECT – PART-1**

**Analyzing the Dataset Using Neo4j**

**Vinayaka S Gadag**

Data Description:

- Dataset containing Race by Race and Driver by Driver result from 1950 to 2017 (Manav, 2021).

- The dataset is created by Manav on Kaggle (<https://www.kaggle.com/manabendrarout/f1-dataset-race-by-race-19502017/metadata>).

- Size of the data is 5.43 MB.

- Dataset has 39 columns and 23777 rows

Model description

- I have selected the below nodes for my analysis, and this may vary during in-depth analysis.

- Nodes, Relationships, and Properties:

* + Driver: This node has a property of a driver and I can create a relationship with other nodes using relationships (PARTICIPATES\_IN, FROM\_COUNTRY, PART\_OF, DRIVER\_ACCUMULATED\_POINTS, DRIVES, RESULT\_ID, POSITIONED\_AT). This will give me an analysis of who is participating in which grand Prix and who won the race. Also has properties to define driver information like name, date of birth, and nationality.
  + Grand Prix: This node has properties of the organized racing events. I can create a relationship between other nodes using relationships (HAS\_CIRCUITS, HAPPENED\_IN, ORIGINS\_FROM) to analyze the results on who is participating in the racing event, who won, and country. It also has properties like GP name, year, GP date, country.
  + Circuit: This node has a list of circuits in the particular grand Prix. Using this we can analyze the results like who won in which circuit. It has the following properties and relationships.
    - Properties – circuit name, circuit location, and circuit country
    - Relationships – ORGANIZED\_IN, ORGANIZED\_AT
  + Year: This node will help in getting the list of grand Prix and circuits organized in the particular year. It has property, Year.
  + Country: This node will help in segregating the list of GPs, Circuits, and Drivers belonging to the particular country and will also help in fetching the result of the race by country. It has a property country.
  + Rank: This node will help in fetching the rank of a driver and a car maker. It has property rank.
  + Result: This node will give the result of each grand Prix related to the Grand Prix and driver. It has a unique property result\_id.
  + Car: This node will help in analyzing which car belongs to whom, manufactured by, and at last which carmaker won the Grand Prix. It has the following relationship and property.
    - Relationship: CONSTRUCTOR\_ACCUMULATED\_POINTS
    - Property: constructor nationality and name.
  + Points: This node will help in analyzing how many points are collected by each driver, car maker, and winner of the Grand Prix. It has property, points.

Data Schema:

Diagram

Description automatically generated

**REFERENCES**

Kaggle - <https://www.kaggle.com/manabendrarout/f1-dataset-race-by-race-19502017/metadata>