**Final Project Part-3**

**VINAYAKA S GADAG**

**Summary statistics and schema:**

The report provides an analysis of F1 Race by Race (1950-2017). The dataset has been reduced to 3500 records and 19 attributes for the simplicity using python and pandas. The uploaded dataset has total of 5868 nodes, 9 labels, 26227 relationships and 12 relationship types in the database. Namely, nodes are GrapndPrix, Driver, CircuitName, Result, Points, Rank, Constructor, Status, and Laps. Relationship types are, HAS\_CIRCUITS, CURRENT\_STATUS, RESULT, DRIVES, PARTICIPATES\_IN, HAS\_POINTS, RANKED\_AT, PART\_OF, COMPLETED\_LAPS, CONSTRUCTOR\_RANKED\_AT, CONSTRUCTOR\_POINTS, CONSTRUCTOR\_COMPLETED\_LAPS (See Figure 1 and Figure 2). The schema was used to query useful information and statistics about the F1 race, see below example queries and results for more information.

**Exploratory data description and queries:**

Query, ‘Get the number of completed grand prix for Hamilton’. This gives the insights on the number of times Hamilton participated and completed the grand prix. The ‘Status’ node used with ‘WHERE’ conditional statement to fetch the finished grand prix and get the count of participation by driver using ‘Driver’ node property and its relationship ‘PARTICIPATES\_IN’ and return the count in descending order by using COUNT aggregate function and ORDER\_BY clause. (See Figure 3)

Query, ‘Return all grand prix organized in 2000 or greater and count number of wins by Hamilton in each grand prix’. The question gives statistics on driver ‘Hamilton’ with the number of grand prix wins organized in 2000 or greater. This same question will give us another interesting result in the next queries. The node ‘Driver’ is used with ‘WHERE’ condition to filter out the drivers by using property ‘driverWin’ and ‘name’. Based on this fetch the grand prix by using node ‘GrapndPrix’ and ‘WHERE’ condition to set the year greater than equal to 2000 using operator (>=), COUNT aggregate function, and ‘AND’ Boolean operator. (See Figure 4)

Query, ‘Return the average number of laps by drivers in all grand prix’. The relationship type ‘COMPLETED\_LAPS’ used between the nodes ‘GranpndPrix’ and ‘driver’ to fetch the average numbers of laps using aggregate function avg() for each driver. (See Figure 5)

The interesting fact I came across while exploring the data was that the Hamilton’s all time highest wins from the F1 race’s inception. He has won total of 17(highest in the cleaned dataset and he has the second highest wins overall with full dataset) grand prix. The above analysis is done using COUNT aggregate function and ORDER BY clause. Fetching the total number career wins by Hamilton using node ‘Driver’ with property ‘driverWin’ set to TRUE with the help of relationship type ‘PARTICIPATES\_IN’ to match with the grand prix where Hamilton is participating in and return the count in descending order by using ORDER BY clause. (See Figure 6 and Figure 7)

Query, ‘Return all the distinct circuits in the grand prix’. For this question, fetch and return the distinct circuits in the grand prix using DISTINCT operator in conjunction with aggregate function COLLECT. The collect function returns the single aggregated list containing values. (See Figure 8)

**REFERENCES**

* <https://neo4j.com/docs/cypher-manual/current/>

**APPENDIX**

**A screenshot of a computer

Description automatically generated with medium confidence**

Figure Neo4j Schema

A screenshot of a computer

Description automatically generated with medium confidence

Figure Neo4j Schema2

A screenshot of a computer

Description automatically generated with medium confidence

Figure Fetch GP and Count

A screenshot of a computer

Description automatically generated with medium confidence

Figure Fetch GP wins for Hamilton organized in 2000 or later

A screenshot of a computer

Description automatically generated with medium confidence

Figure Fetch average number laps completed by driver

A screenshot of a computer

Description automatically generated

Figure Highest wins by Hamilton

A screenshot of a computer

Description automatically generated with medium confidence

Figure Graph Visualization for highest wins by Hamilton

A screenshot of a computer

Description automatically generated with medium confidence

Figure GP Circuits