Queries

Query1: Check for duplicate records in the relation, if any then override the relation with distinct records.

Query2: Display all the countries along with their occurences.

Query3: Display all the information from the Tests\_conducted.csv file in descending order of positive cases and store it in TestsAnalysis directory.

Query4: Display country name, positive cases and positive/tested% of those tuples whose positive/tested% is greater than 25%.

Query5: List all the countries whose positive cases are greater than 15000 and country name starts with 'A'.

Query6: How many tests have been conducted and out of them how many positive cases came according to this file.

Query7: Show the number of positive cases of australia in may 2020.

Query8a: Name the country having max positive cases.

Query8b: Name the country having minimum positive cases.

Query9a: Name the country having maximum %age.

Query9b: Name the country having minimum %age.

Query10: Display the number of countries having tested count greater than 700000.

Query11: Display the tuples whose country name starts with 'B and percentage > 2.0.

Query12: Order the relation with respect to positive cases in the month of may.

Query13: Display the number of tuples for every month.

Query14a: Display the month in which maximum positive cases were found, along with positive cases.

Query14b: Display the month in which minimum positive cases were found.

Query15: Order the relation in descending order of %age.

Query16: Rank the relation by decreasing %age and country name (dense rank).

Query17: Generate 1% sample of this dataset and display it and store it.

Query18: Split this dataset into 3 relations, one containing data with perc < 5, another with perc >= 5 and perc < 10, remaining in the third relation. Store all the files in the directory /TestsAnalysis of HDFS.

Query19: Display all the countries whose positive cases are in the range of 10000 and 20000.

Query20: Display the country and its percentage of positive cases of those records in which country name starts with 'U' in the month of july.

Query21: Display the month having maximum and minimum percentage.

Query22: Display the countries (whose survey has been conducted on 1st may and 31st may) along with percentage in the first day of month of may and in the last day of month of may.

Query23: Fetch the record of country starting with 'S' which has tested maximum persons.

Query24: Order the relation alphabetically with respect to country name.