Queries

Query1: Display all the regions (along with their counts) to which observations of this dataset belong.

Query2: Display all the information from the unemployment.csv file in descending order of Estimated unemployment rate(%) and store it in unemployment directory.

Query3: Display all the tuples having region name starts 'A' and has estimated employed population greater than 1000000.

Query4: Display all the tuples having Estimated Unemployment Rate(%) greater than 5.0 and belong to urban area.

Query5a: Display the region having maximum number of estimated employed population

Query5b: Display the region having minimum number of estimated employed population

Query6a: Display the region having maximum number of estimated labour participation rate(%).

Query6b: Display the region having minimum number of estimated labour participation rate(%)

Query7: Order the relation in descending order of estimated labour participation rate(%)

Query8: Display region name, estimated labour participation rate(%) and area of tuples whose estimated unemployment rate(%) is greater than 15%.

Query9: Display the number of observations recorded for each month.

Query10: Order the relation by estimated unemployment rate(%) in the month of feb 2020.

Query11: What is the average estimated unemployment rate (%) and average estimated employed for rural and urban.

Query12: Rank the relation with respet to estimated employed in the year of 2020.

Query13: What is the average estimated unemployment rate(%) and average estimated labour participation rate(%) for each year.

Query14: Generate 2% sample of this dataset and display and store it.

Query15: Rank the relation by decreasing estimated unemployment rate(%) and increasing region.

Query16: Split this dataset into 3 relations, one containing data with estimated unemployment rate(%) < 5.0, another with estimated unemployment rate(%) >= 5.0 and estimated unemployment rate(%) < 10.0, remaining in the third relation. Store all the files in the directory /unemploymentAnalysis of HDFS.

Query17: Display the region and estimated unemployment rate of records in which region name starts with ‘M’ in the month of july.

Query18: Split the relation in two halves, one with records belonging to rural area and other with records belonging to urban area. Also store these relations in the directory /unemploymentAnalysis of HDFS.

Query19: Rank the relation containing records which belong to rural area, with respect to decreasing estimated unemployment rate.

Query20: Display all the regions whose estimated employed population is in the range of 8000000 and 10000000 which belongs to urban area.

Query21a: Display the month having maximum estimated unemployment rate in 2020 in urban area.

Query21b: Display the month having minimum estimated unemployment rate in 2020 in urban area.

Query22: Order the relation alphabetically with respect to region names. Also the records belong to rural area.

Query23: Rank the rural relation with respect to decreasing estimated employed population.