

Read Me

Please find commands required for executing each of the Exercise listed below. Please note that the command includes complete format along with the get command required for looking at the output. Also, Copy of an execution output is present in docs folder.

Exercise 1 Command : `hadoop jar YelpStatistics-jar-with-dependencies.jar
yelp.statistics.BusinessesFromPaloAlto_Q1 "/yelpdatafall/business/business.csv"
/user/kpw150030/A4_o1;
hdfs dfs -get /user/kpw150030/A4_o1/part-r-00000 Q1_output.txt`

Exercise 2 Command : `hadoop jar YelpStatistics-jar-with-dependencies.jar
yelp.statistics.ColocatedBusinessesFromNY_Q2 "/yelpdatafall/business/business.csv"
/user/kpw150030/A4_02 ;hdfs dfs -get /user/kpw150030/A4_02/part-r-00000
Q2_output.txt;view Q2_output.txt`

Exercise 3 Command : `hadoop jar YelpStatistics-jar-with-dependencies.jar
yelp.statistics.TopBusinessDensityZipCodes_Q3 "/yelpdatafall/business/business.csv"
/user/kpw150030/A4_00003/ ;hdfs dfs -get /user/kpw150030/A4_00003/part-r-00000
Q3_output.txt;view Q3_output.txt`

Exercise 4 Command : `hadoop jar YelpStatistics-jar-with-dependencies.jar
yelp.statistics.TweetDownloader "/user/kpw150030/A4_ip1/" "/user/kpw150030/A4_04/"
"NYSE";hdfs dfs -get /user/kpw150030/A4_04/part-r-00000 Q4_output.txt;view Q4_output.txt`

Exercise 5 Command : `hadoop jar YelpStatistics-jar-with-dependencies.jar
yelp.statistics.Top10BusinessesAverageRating_Q5 "/yelpdatafall/review/review.csv"
/user/kpw150030/A4_05 ;hdfs dfs -get /user/kpw150030/A4_05/part-r-00000
Q5_output.txt;view Q5_output.txt`

Exercise 6 Command : `hadoop jar YelpStatistics-jar-with-dependencies.jar
yelp.statistics.Bottom10BusinessesAverageRating_Q6 "/yelpdatafall/review/review.csv"
/user/kpw150030/A4_06 ;hdfs dfs -get /user/kpw150030/A4_06/part-r-00000
Q6_output.txt;view Q6_output.txt`