Home (http://hadoopgyaan.tk/)   About Me (http://hadoopgyaan.tk/about/)

f (https://www.facebook.com/Hadoop-Gyaan-237654456618006/?ref=bookmarks)

🐦 (https://twitter.com/HadoopGyaan/following)

▶ (https://www.youtube.com/channel/UCIGfpWblzZiV2yPkfSZ25gg)

g+ (https://plus.google.com/b/104550599737584656437/104550599737584656437/about?gmbpt=true&hl=en)

🔍

(http://hadoopgyaan.tk/)

HadoopGyaan (http://hadoopgyaan.tk/)

Home
(http://hadoopgyaan.tk/)

Hadoop Installation Gyaan
(http://hadoopgyaan.tk/category/hadoop-installation-gyaan/)

Hadoop Gyaan
(http://hadoopgyaan.tk/category/hadoop-gyaan/)

SQL Server Gyaan
(http://hadoopgyaan.tk/category/sql-server-gyaan/)

UDF's Gyaan
(http://hadoopgyaan.tk/category/hive-udfs-gyaan/)

Python Gyaan
(http://hadoopgyaan.tk/category/python/)

About Me
(http://hadoopgyaan.tk/about/)

New York City Bike Data Analysis MapReduce Case Study

Posted on August 31, 2016 (http://hadoopgyaan.tk/hadoop-gyaan/new-york-city-bike-data-analysis-mapreduce-case-study/)   by
HadoopGyaan (http://hadoopgyaan.tk/author/admin/)

Citi Bike, a public bicycle sharing system that serves parts of New York City, is the largest bike-sharing program in the United States. With the existing wealth of data pertaining to Citi Bike users in New York City, it is possible to identify, segment and categorize the customer domain based on several factors such as Age, Sex and Occupation which will help the company to identify their potential customers that can be targeted and also the weak customer domain which needs to be improved. The prime objective of this project is to identify long-run customers and potential audience that can be targeted to increase the company's customer base and drive more revenue.

Other deliverables of this project would be to identify hot-spot locations and peak-demand hours, which can be crucial information that can help the company to better, manage their business supply, which will also essentially help more customers. Identifying the hot-spots at peak hours would help the company to understand the market demand and thereby increase the bike availability at these hot-spots during peak hours will help in driving more revenue. Our design would rely largely upon existing Citi Bike trip histories data and Citi Bike Daily Ridership and Membership Data.

The Dataset:

- Trip Duration (seconds)
- Start Time and Date
- Stop Time and Date
- Start Station Name
- End Station Name
- Station ID
- Station Lat/Long

- Bike ID
- User Type (Customer = 24-hour pass or 7-day pass user; Subscriber = Annual Member)
- Gender (Zero=unknown; 1=male; 2=female)
- Year of Birth

Problem Statements :

- 1.Number of Subscribers and Customers
- 2.Number of Subscribers and Customers for each gender
- 3.Number of Subscribers and Customers for each gender in every age category
- 4.Average Trip Distance
- 5.Week Stats : Identify the day with most trips
- 6.Most Popular Stations : Identify the stations with most originating trips and destinations.

Downloads:

New York City Bike Data Analysis MR Codes (https://drive.google.com
/open?id=0B3iLhHNJWdlFcHdmOUxLbTB2X0E)

New York City Bike Data Analysis Dataset and Output (https://drive.google.com
/open?id=0B3iLhHNJWdlFOHFqbTF2d3lpcEE)

I hope this tutorial will surely help you. If you have any questions or problems please let me know.

Happy Hadooping with Patrick..

Share this:

🐦 (http://hadoopgyaan.tk/hadoop-gyaan/new-york-city-bike-data-analysis-mapreduce-case-study
/?share=twitter&nb=1)

f (http://hadoopgyaan.tk/hadoop-gyaan/new-york-city-bike-data-analysis-mapreduce-case-study
/?share=facebook&nb=1)

G+ (http://hadoopgyaan.tk/hadoop-gyaan/new-york-city-bike-data-analysis-mapreduce-case-study
/?share=google-plus-1&nb=1)

Related

Face Detection in Hadoop using
HIPI and OpenCV
(http://hadoopgyaan.tk/hadoop-
gyaan/face-detection-hadoop-
using-hipi-opencv/)

Weather Analysis using HBASE,
MapReduce and HDFS
(http://hadoopgyaan.tk/hadoop-
gyaan/apache-hadoop-weather-
analysis-using-hbase-

Movie Recommender
MapReduce Case Study
(http://hadoopgyaan.tk/hadoop-
gyaan/apache-hadoop-movie-
recommender-mapreduce-

September 1, 2016	mapreduce-and-hdfs/)	case-study/)
In "Hadoop Gyaan"	July 31, 2016	July 3, 2016
	In "Hadoop Gyaan"	In "Hadoop Gyaan"

Posted in Hadoop Gyaan (http://hadoopgyaan.tk/category/hadoop-gyaan/) Tagged bicycle (http://hadoopgyaan.tk/tag/bicycle/), big data (http://hadoopgyaan.tk/tag/big-data/), case study (http://hadoopgyaan.tk/tag/case-study/), city bike (http://hadoopgyaan.tk/tag/city-bike/), hadoop (http://hadoopgyaan.tk/tag/hadoop/), hadoop 2x (http://hadoopgyaan.tk/tag/hadoop-2x/), mapreduce (http://hadoopgyaan.tk/tag/mapreduce/), new york (http://hadoopgyaan.tk/tag/new-york/)

5 thoughts on "New York City Bike Data Analysis MapReduce Case Study"

Nancy **(http://hadoopgyaan.tk/hadoop-gyaan/new-york-city-bike-data-analysis-mapreduce-case-study/#comment-8)** *August 31, 2016*

hdfs@localhost:/opt/ecosystems/citibike$ $HADOOP_HOME/bin/hadoop jar genderAnalytic.jar genderAnalyticDriver /user/citibike/ /user/citout1
Exception in thread "main" java.lang.ClassNotFoundException: genderAnalyticDriver
at java.net.URLClassLoader$1.run(URLClassLoader.java:372)
at java.net.URLClassLoader$1.run(URLClassLoader.java:361)
at java.security.AccessController.doPrivileged(Native Method)
at java.net.URLClassLoader.findClass(URLClassLoader.java:360)
at java.lang.ClassLoader.loadClass(ClassLoader.java:424)
at java.lang.ClassLoader.loadClass(ClassLoader.java:357)
at java.lang.Class.forName0(Native Method)
at java.lang.Class.forName(Class.java:340)
at org.apache.hadoop.util.RunJar.run(RunJar.java:214)
at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
hdfs@localhost:/opt/ecosystems/citibike$

Reply (http://hadoopgyaan.tk/hadoop-gyaan/new-york-city-bike-data-analysis-mapreduce-case-study/?replytocom=8#respond)

Nancy **(http://hadoopgyaan.tk/hadoop-gyaan/new-york-city-bike-data-analysis-mapreduce-case-study/#comment-9)** *August 31, 2016*

getting above error while using the downloaded code..

Reply (http://hadoopgyaan.tk/hadoop-gyaan/new-york-city-bike-data-analysis-mapreduce-case-study/?replytocom=9#respond)

HadoopGyaan **(http://hadoopgyaan.tk/hadoop-gyaan/new-york-city-bike-data-analysis-mapreduce-case-study/#comment-10)** *September 1, 2016*

Hi Nancy,

Please mentioned the script which you are using to run CitiBike jar file. Like jar file name,input file and output file name as it's looks like compiling error Moreover check

out "genderAnalytic Driver" java code.

Reply (http://hadoopgyaan.tk/hadoop-gyaan/new-york-city-bike-data-analysis-mapreduce-case-study/?replytocom=10#respond)

Naitik Mehta (http://hadoopgyaan.tk/hadoop-gyaan/new-york-city-bike-data-analysis-mapreduce-case-study/#comment-25)   *January 19, 2017*

Hi,

I am receiving the below error during the Map phase itself. Could you please advise.:

Error: java.lang.NumberFormatException: For input string: "start_station_latitude"
at sun.misc.FloatingDecimal.readJavaFormatString(FloatingDecimal.java:1250)
at java.lang.Double.parseDouble(Double.java:540)
at citiBikeAnalytics.BikeAnalyticMapper.map(BikeAnalyticMapper.java:85)
at citiBikeAnalytics.BikeAnalyticMapper.map(BikeAnalyticMapper.java:1)
at org.apache.hadoop.mapreduce.Mapper.run(Mapper.java:145)
at org.apache.hadoop.mapred.MapTask.runNewMapper(MapTask.java:787)
at org.apache.hadoop.mapred.MapTask.run(MapTask.java:341)
at org.apache.hadoop.mapred.YarnChild$2.run(YarnChild.java:164)
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:415)
at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1693)
at org.apache.hadoop.mapred.YarnChild.main(YarnChild.java:158)

Reply (http://hadoopgyaan.tk/hadoop-gyaan/new-york-city-bike-data-analysis-mapreduce-case-study/?replytocom=25#respond)

Naitik Mehta (http://hadoopgyaan.tk/hadoop-gyaan/new-york-city-bike-data-analysis-mapreduce-case-study/#comment-26)   *January 19, 2017*

Silly one. Got it. Skipped the first line while parsing the input file. Works fine now. 😀

Reply (http://hadoopgyaan.tk/hadoop-gyaan/new-york-city-bike-data-analysis-mapreduce-case-study/?replytocom=26#respond)

Leave a Reply

Your email address will not be published. Required fields are marked *

Comment

Name *

Email *

Website

Post Comment

☐ **Notify me of follow-up comments by email.**

☐ **Notify me of new posts by email.**

Search ...

Recent Posts

> Movie Theatres SQL Server Case Study (http://hadoopgyaan.tk/sql-server-gyaan/movie-theatres-sql-server-case-study/)

> Computer Store SQL Server Case Study (http://hadoopgyaan.tk/sql-server-gyaan/the-computer-store-sql-server-case-study/)

> Beer Recommender Python Case Study (http://hadoopgyaan.tk/python/beer-recommender-using-python/)

> Excite Search Engine Log Pig Case Study (http://hadoopgyaan.tk/hadoop-gyaan/excite-search-engine-log-pig-case-study/)

> Face Detection in Hadoop using HIPI and OpenCV (http://hadoopgyaan.tk/hadoop-gyaan/face-detection-hadoop-using-hipi-opencv/)

Recent Comments

> HadoopGyaan on Excite Search Engine Log Pig Case Study (http://hadoopgyaan.tk/hadoop-gyaan/excite-search-engine-log-pig-case-study/#comment-55)

> borvestinkral (http://www.borvestinkral.com/) on Excite Search Engine Log Pig Case Study (http://hadoopgyaan.tk/hadoop-gyaan/excite-search-engine-log-pig-case-study/#comment-52)

> Teespring (http://ow.ly/AvWq30daL8Z) on How to Setup Hadoop 2.7.1 [Any Stable Version] on CentOS (http://hadoopgyaan.tk/hadoop-installation-gyaan/how-to-setup-hadoop-2-7-1-any-stable-version-on-centos-and-ubuntu/#comment-46)

> Ward (http://ow.ly/AvWq30daL8Z) on How to Setup Hadoop 2.7.1 [Any Stable Version] on CentOS (http://hadoopgyaan.tk/hadoop-installation-gyaan/how-to-setup-hadoop-2-7-1-any-stable-version-on-centos-and-ubuntu/#comment-45)

> marxist feminist dialectic shirt (http://ow.ly/lSuH30daJOW) on How to Setup Hadoop 2.7.1 [Any Stable Version] on CentOS (http://hadoopgyaan.tk/hadoop-installation-gyaan/how-to-setup-hadoop-2-7-1-any-stable-version-on-centos-and-ubuntu/#comment-44)

Archives

> July 2017 (http://hadoopgyaan.tk/2017/07/)

> September 2016 (http://hadoopgyaan.tk/2016/09/)

> August 2016 (http://hadoopgyaan.tk/2016/08/)

> July 2016 (http://hadoopgyaan.tk/2016/07/)

> June 2016 (http://hadoopgyaan.tk/2016/06/)

Categories

> Hadoop Gyaan (http://hadoopgyaan.tk/category/hadoop-gyaan/)

> Hadoop Installation Gyaan (http://hadoopgyaan.tk/category/hadoop-installation-gyaan/)

> Hive Gyaan (http://hadoopgyaan.tk/category/hive-gyaan/)

> Hive UDF's Gyaan (http://hadoopgyaan.tk/category/hive-udfs-gyaan-2/)

> PIG Gyaan (http://hadoopgyaan.tk/category/pig-gyaan/)

> Python Gyaan (http://hadoopgyaan.tk/category/python/)

> SQL Server Gyaan (http://hadoopgyaan.tk/category/sql-server-gyaan/)

Subscribe to Blog via Email

Enter your email address to subscribe to this blog and receive notifications of new posts by email.

Join 38 other subscribers

Email Address

Subscribe

Hadoop Gyaan
24 likes

Like Page

Be the first of your friends to like this

Contact Me Here

 (https://www.facebook.com
/hadoopgyaan/)
 (https://twitter.com/HadoopGyaan/)
 (https://instagram.com/hadoopgyaan/)
 (https://www.pinterest.com
/HadoopGyaan/)
 (https://www.linkedin.com
/in/Prateek%20harsh/)
 (https://www.youtube.com
/user/Hadoop%20Gyaan/)
 (https://plus.google.com
/u/0/+Hadoop%20Gyaan/)

Theme Designed by Prateek Harsh (http://hadoopgyaan.tk/). Copyright 2016 Hadoop Gyaan