Bicycle Sharing Demand

Data Exploration and Transformation

1) Application Development for Model Generation -refer file **"BIKE_AVL_PREDICTION.ipynb"**

I have taken input data as Train.csv, and performed various Transformations and cleaning of data, using Google Collab I have done this process in python, for better visual representation of Graphs , to understand the nature of data, and developed ML using Linear Regression, K-Nearest Neighbour, Decision Tree Regression and finally compared Actual and Predicted value, and got good Predicted results.
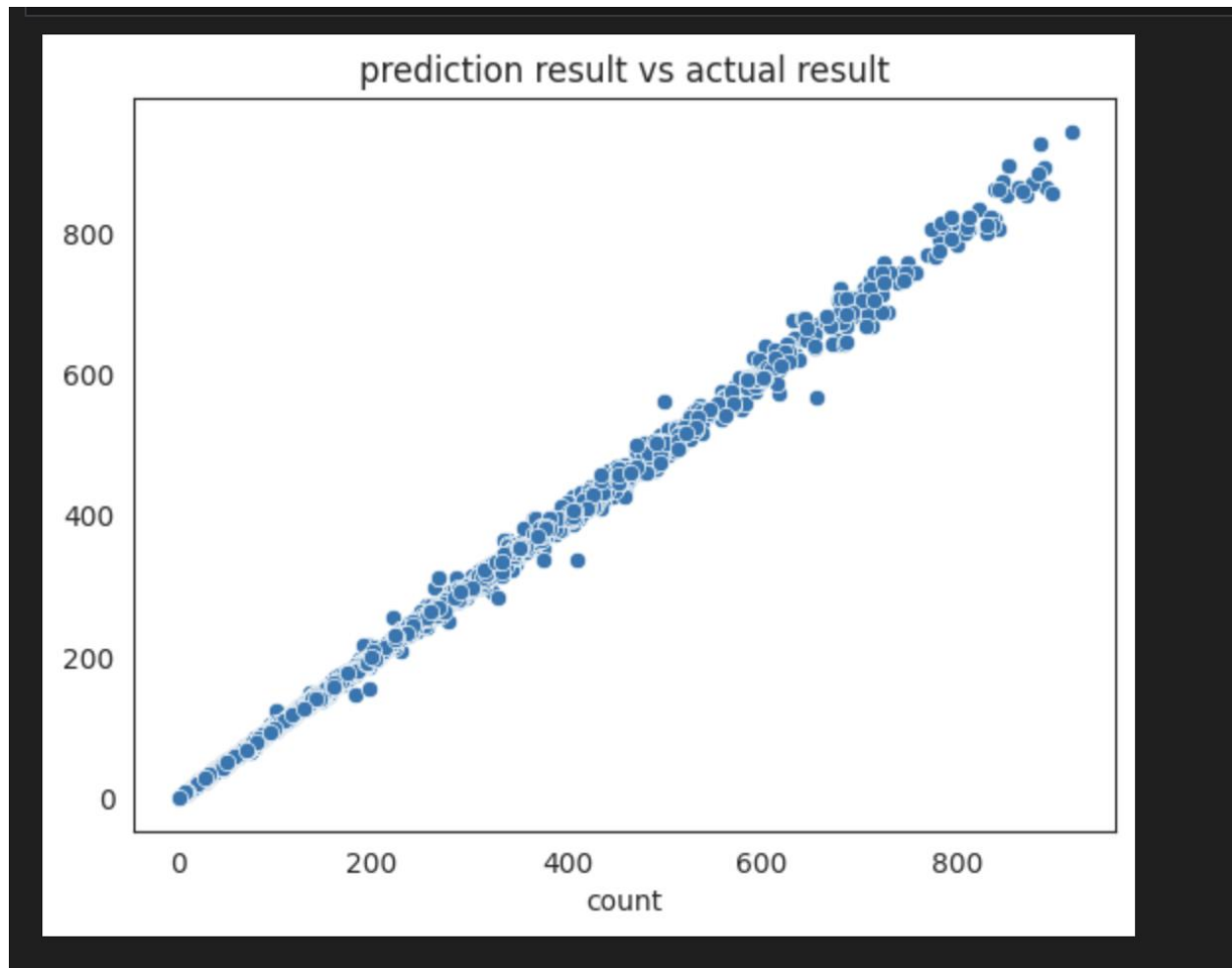
2) Application Development for Demand Prediction – refer file - **"PYSPARK_ML_MODEL.ipynb"**

I have performed same cleaning and Transformation of data in Pyspark as well and have taken input data from HDFS path

1. Load the persisted model.
2. Predict bike demand

Done using Pyspark and loaded the model to local file , attached the model file as well which I have used for the live stream data processing ML model

Refer: **"Saved_model"**

3) Application for Streaming Data – refer file - "**Running_Spark_live.py**"

Using flume I have conected the source Telnet and the sink, like data i have feeded via telnet so the fumeagen running in back end will push the data to the sink
Input data is given in json format,

FLUME_AGENT:
Setting up flume to push data into spark flume sink

```
Last login: Sun Aug 11 11:19:58 on ttys000
(base) arunkumar@Aruns-MBP ~ %   ~/flume-ng/bin/flume-ng agent --conf ~/flume-ng/conf -f ~/Downloads/KafkaNC.conf -Dflume.root.logger=DEBUG,console -n KafkaNC
curl telnet://localhost:44444
Warning: JAVA_HOME is not set!
Info: Including Hadoop libraries found via (/Users/arunkumar/hadoop-3.3.0/bin/hadoop) for HDFS access
Info: Excluding /Users/arunkumar/hadoop-3.3.0/share/hadoop/common/lib/slf4j-api-1.7.25.jar from classpath
Info: Excluding /Users/arunkumar/hadoop-3.3.0/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar from classpath
+ exec /Users/arunkumar/jdk-11.0.24.jdk/Contents/Home/bin/java -Xmx20m -Dflume.root.logger=DEBUG,console -cp '/Users/arunkumar/flume-ng/conf:/Users/arunkumar/flume-ng/lib/*:/Use
rs/arunkumar/hadoop-3.3.0/etc/hadoop:/Users/arunkumar/hadoop-3.3.0/share/hadoop/common/lib/accessors-smart-1.2.jar:/Users/arunkumar/hadoop-3.3.0/share/hadoop/common/lib/animal-s
niffer-annotations-1.17.jar:/Users/arunkumar/hadoop-3.3.0/share/hadoop/common/lib/asm-5.0.4.jar:/Users/arunkumar/hadoop-3.3.0/share/hadoop/common/lib/audience-annotations-0.5.0.
jar:/Users/arunkumar/hadoop-3.3.0/share/hadoop/common/lib/avro-1.7.7.jar:/Users/arunkumar/hadoop-3.3.0/share/hadoop/common/lib/checker-qual-2.5.2.jar:/Users/arunkumar/hadoop-3.3
.0/share/hadoop/common/lib/commons-beanutils-1.9.4.jar:/Users/arunkumar/hadoop-3.3.0/share/hadoop/common/lib/commons-cli-1.2.jar:/Users/arunkumar/hadoop-3.3.0/share/hadoop/commo
n/lib/commons-codec-1.11.jar:/Users/arunkumar/hadoop-3.3.0/share/hadoop/common/lib/commons-collections-3.2.2.jar:/Users/arunkumar/hadoop-3.3.0/share/hadoop/common/lib/commons-co
mpress-1.19.jar:/Users/arunkumar/hadoop-3.3.0/share/hadoop/common/lib/commons-configuration2-2.1.1.jar:/Users/arunkumar/hadoop-3.3.0/share/hadoop/common/lib/commons-daemon-1.0.1
3.jar:/Users/arunkumar/hadoop-3.3.0/share/hadoop/common/lib/commons-io-2.5.jar:/Users/arunkumar/hadoop-3.3.0/share/hadoop/common/lib/commons-lang3-3.7.jar:/Users/arunkumar/hadoo
p-3.3.0/share/hadoop/common/lib/commons-logging-1.1.3.jar:/Users/arunkumar/hadoop-3.3.0/share/hadoop/common/lib/commons-math3-3.1.1.jar:/Users/arunkumar/hadoop-3.3.0/share/hadoo
p/common/lib/commons-net-3.6.jar:/Users/arunkumar/hadoop-3.3.0/share/hadoop/common/lib/commons-text-1.4.jar:/Users/arunkumar/hadoop-3.3.0/share/hadoop/common/lib/curator-client-
```

TELNET INPUT : nc localhost 44444

```
{'datetime':'2011-07-01 09:00:00','season':'3','holiday':'0','workingday':'1','weather':'1','temp':'30.34','atemp':'32.575','humidity':'35','windspeed':'11.0014','casual':'58','
registered':'202','count':'260'}
{'datetime':'2011-07-01 10:00:00','season':'3','holiday':'0','workingday':'1','weather':'1','temp':'31.16','atemp':'33.335','humidity':'33','windspeed':'11.0014','casual':'61','
registered':'109','count':'170'}
{'datetime':'2011-07-01 11:00:00','season':'3','holiday':'0','workingday':'1','weather':'1','temp':'32.8','atemp':'34.85','humidity':'26','windspeed':'0','casual':'68','register
ed':'146','count':'214'}
{'datetime':'2011-07-01 12:00:00','season':'3','holiday':'0','workingday':'1','weather':'1','temp':'32.8','atemp':'34.85','humidity':'26','windspeed':'0','casual':'83','register
ed':'180','count':'263'}OK
OK
OK
```

Sample input:

```
--------------------Providing sample inputs to TELNET
{'datetime':'2011-06-19 13:00:00','season':'2','holiday':'0','workingday':'0','weather':'2','temp':'30.34','atemp':'33.335','humidity':'51','windspeed':'0','casual':'149','registered':'214','count':'363'}
{'datetime':'2011-06-19 14:00:00','season':'2','holiday':'0','workingday':'0','weather':'3','temp':'30.34','atemp':'34.09','humidity':'58','windspeed':'7.0015','casual':'124','registered':'193','count':'317'}
{'datetime':'2011-06-19 15:00:00','season':'2','holiday':'0','workingday':'0','weather':'2','temp':'30.34','atemp':'34.09','humidity':'55','windspeed':'7.0015','casual':'146','registered':'216','count':'362'}
{'datetime':'2011-06-19 16:00:00','season':'2','holiday':'0','workingday':'0','weather':'2','temp':'31.16','atemp':'34.09','humidity':'48','windspeed':'0','casual':'121','registered':'257','count':'378'}
{'datetime':'2011-06-19 17:00:00','season':'2','holiday':'0','workingday':'0','weather':'1','temp':'30.34','atemp':'33.335','humidity':'51','windspeed':'0','casual':'159','registered':'238','count':'397'}
{'datetime':'2011-06-19 18:00:00','season':'2','holiday':'0','workingday':'0','weather':'1','temp':'30.34','atemp':'34.09','humidity':'58','windspeed':'6.0032','casual':'78','registered':'218','count':'296'}
{'datetime':'2011-06-19 19:00:00','season':'2','holiday':'0','workingday':'0','weather':'1','temp':'29.52','atemp':'34.09','humidity':'62','windspeed':'8.9981','casual':'94','registered':'217','count':'311'}
{'datetime':'2011-06-19 20:00:00','season':'2','holiday':'0','workingday':'0','weather':'1','temp':'28.7','atemp':'32.575','humidity':'70','windspeed':'12.998','casual':'86','registered':'146','count':'232'}
{'datetime':'2011-06-19 21:00:00','season':'2','holiday':'0','workingday':'0','weather':'1','temp':'27.88','atemp':'31.82','humidity':'79','windspeed':'8.9981','casual':'57','registered':'110','count':'167'}
{'datetime':'2011-06-19 22:00:00','season':'2','holiday':'0','workingday':'0','weather':'1','temp':'27.06','atemp':'30.305','humidity':'78','windspeed':'15.0013','casual':'38','registered':'114','count':'152'}
{'datetime':'2011-06-19 23:00:00','season':'2','holiday':'0','workingday':'0','weather':'1','temp':'27.06','atemp':'30.305','humidity':'78','windspeed':'12.998','casual':'14','registered':'78','count':'92'}
{'datetime':'2011-07-01 00:00:00','season':'3','holiday':'0','workingday':'1','weather':'1','temp':'27.06','atemp':'31.06','humidity':'50','windspeed':'0','casual':'20','registered':'48','count':'68'}
{'datetime':'2011-07-01 01:00:00','season':'3','holiday':'0','workingday':'1','weather':'1','temp':'27.06','atemp':'31.06','humidity':'50','windspeed':'0','casual':'15','registered':'16','count':'31'}
{'datetime':'2011-07-01 02:00:00','season':'3','holiday':'0','workingday':'1','weather':'1','temp':'25.42','atemp':'30.305','humidity':'69','windspeed':'0','casual':'6','registered':'7','count':'13'}
{'datetime':'2011-07-01 03:00:00','season':'3','holiday':'0','workingday':'1','weather':'1','temp':'26.24','atemp':'31.06','humidity':'53','windspeed':'0','casual':'5','registered':'6','count':'11'}
{'datetime':'2011-07-01 04:00:00','season':'3','holiday':'0','workingday':'1','weather':'1','temp':'25.42','atemp':'31.06','humidity':'57','windspeed':'0','casual':'1','registered':'5','count':'6'}
{'datetime':'2011-07-01 05:00:00','season':'3','holiday':'0','workingday':'1','weather':'1','temp':'24.6','atemp':'30.305','humidity':'64','windspeed':'8.9981','casual':'3','registered':'27','count':'30'}
{'datetime':'2011-07-01 06:00:00','season':'3','holiday':'0','workingday':'1','weather':'1','temp':'25.42','atemp':'30.305','humidity':'61','windspeed':'8.9981','casual':'11','registered':'97','count':'108'}
{'datetime':'2011-07-01 07:00:00','season':'3','holiday':'0','workingday':'1','weather':'1','temp':'27.06','atemp':'31.06','humidity':'54','windspeed':'7.0015','casual':'25','registered':'218','count':'243'}
{'datetime':'2011-07-01 08:00:00','season':'3','holiday':'0','workingday':'1','weather':'1','temp':'28.7','atemp':'31.82','humidity':'42','windspeed':'11.0014','casual':'39','registered':'453','count':'492'}
{'datetime':'2011-07-01 09:00:00','season':'3','holiday':'0','workingday':'1','weather':'1','temp':'30.34','atemp':'33.335','humidity':'35','windspeed':'11.0014','casual':'58','registered':'202','count':'260'}
{'datetime':'2011-07-01 10:00:00','season':'3','holiday':'0','workingday':'1','weather':'1','temp':'31.16','atemp':'33.335','humidity':'33','windspeed':'11.0014','casual':'61','registered':'109','count':'170'}
{'datetime':'2011-07-01 11:00:00','season':'3','holiday':'0','workingday':'1','weather':'1','temp':'32.8','atemp':'34.85','humidity':'26','windspeed':'0','casual':'68','registered':'146','count':'214'}
{'datetime':'2011-07-01 12:00:00','season':'3','holiday':'0','workingday':'1','weather':'1','temp':'32.8','atemp':'34.85','humidity':'26','windspeed':'0','casual':'83','registered':'180','count':'263'}
```

Sink is running
Configured spark streaming to pull data from spark flume sink using receivers and
predicted the demand using model and persist the result to RDBMS.

```
(base) arunkumar@Aruns-MBP ~ % $KAFKA_HOME/bin/kafka-server-start.sh $KAFKA_HOME/config/server.properties
[2024-08-11 11:24:45,845] INFO Registered kafka:type=kafka.Log4jController MBean (kafka.utils.Log4jControllerRegistration$)
[2024-08-11 11:24:46,054] INFO Setting -D jdk.tls.rejectClientInitiatedRenegotiation=true to disable client-initiated TLS renegotiation (org.apache.zookeeper.common.X509Util)
[2024-08-11 11:24:46,114] INFO Registered signal handlers for TERM, INT, HUP (org.apache.kafka.common.utils.LoggingSignalHandler)
[2024-08-11 11:24:46,115] INFO starting (kafka.server.KafkaServer)
[2024-08-11 11:24:46,116] INFO Connecting to zookeeper on localhost:2181 (kafka.server.KafkaServer)
[2024-08-11 11:24:46,125] INFO [ZooKeeperClient Kafka server] Initializing a new session to localhost:2181. (kafka.zookeeper.ZooKeeperClient)
[2024-08-11 11:24:46,141] INFO Client environment:zookeeper.version=3.8.4-9316c2a7a97e1666d8f4593f34dd6fc36ecc436c, built on 2024-02-12 22:16 UTC (org.apache.zookeeper.ZooKeeper
)
[2024-08-11 11:24:46,141] INFO Client environment:host.name=aruns-mbp (org.apache.zookeeper.ZooKeeper)
[2024-08-11 11:24:46,141] INFO Client environment:java.version=11.0.24 (org.apache.zookeeper.ZooKeeper)
[2024-08-11 11:24:46,142] INFO Client environment:java.vendor=Oracle Corporation (org.apache.zookeeper.ZooKeeper)
[2024-08-11 11:24:46,142] INFO Client environment:java.home=/Users/arunkumar/jdk-11.0.24.jdk/Contents/Home (org.apache.zookeeper.ZooKeeper)
[2024-08-11 11:24:46,142] INFO Client environment:java.class.path=/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/activation-1.1.1.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../l
ibs/aopalliance-repackaged-2.6.1.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/argparse4j-0.7.0.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/audience-annotations-0.12
.0.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/caffeine-2.9.3.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/checker-qual-3.19.0.jar:/Users/arunkumar/kafka_2.13-3.7.1
//bin/../libs/commons-beanutils-1.9.4.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/commons-cli-1.4.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/commons-collections-3
.2.2.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/commons-digester-2.1.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/commons-io-2.11.0.jar:/Users/arunkumar/kafka_2.13
-3.7.1/bin/../libs/commons-lang3-3.8.1.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/commons-logging-1.2.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/commons-validat
or-1.7.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/connect-api-3.7.1.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/connect-basic-auth-extension-3.7.1.jar:/Users/arun
kumar/kafka_2.13-3.7.1/bin/../libs/connect-json-3.7.1.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/connect-mirror-3.7.1.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs
/connect-mirror-client-3.7.1.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/connect-runtime-3.7.1.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/connect-transforms-3.7.1
.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/error_prone_annotations-2.10.0.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/hk2-api-2.6.1.jar:/Users/arunkumar/kafka_2.
13-3.7.1/bin/../libs/hk2-locator-2.6.1.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/hk2-utils-2.6.1.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/jackson-annotations
-2.16.2.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/jackson-core-2.16.2.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/jackson-databind-2.16.2.jar:/Users/arunkumar/ka
fka_2.13-3.7.1/bin/../libs/jackson-dataformat-csv-2.16.2.jar:/Users/arunkumar/kafka_2.13-3.7.1/bin/../libs/jackson-datatype-jdk8-2.16.2.jar:/Users/arunkumar/kafka_2.13-3.7.1//
```

Running the Spark Streaming:

```
>>>
>>> query = predictions.writeStream.foreachBatch(write_to_mysql).outputMode("append").start()
24/08/11 12:10:19 WARN ResolveWriteToStream: Temporary checkpoint location created which is deleted normally when the query didn't fail: /private/var/folders/4m/z36lvyp57r37vs_g
dh88c2mc0000gn/T/temporary-83f3f3d4-5e19-42ff-9fca-f40789a5277f. If it's required to delete it under any circumstances, please set spark.sql.streaming.forceDeleteTempCheckpointL
ocation to true. Important to know deleting temp checkpoint folder is best effort.
24/08/11 12:10:19 WARN ResolveWriteToStream: spark.sql.adaptive.enabled is not supported in streaming DataFrames/Datasets and will be disabled.
>>> 24/08/11 12:10:19 WARN AdminClientConfig: These configurations '[key.deserializer, value.deserializer, enable.auto.commit, max.poll.records, auto.offset.reset]' were supplie
d but are not used yet.
```

Finaly Predicted data is been loaded to the MYSQL DATABASE:
Pushing messages from flume to test the application. Application processed and persisted the result to RDBMS

Local instance 3306 - Warning - not supported

Query 2

```
1   show databases;
2   use customers;
3   show tables;
4
5   CREATE TABLE `edureka_spark` (
6       `epoch_id` INT AUTO_INCREMENT,
7       `label` DOUBLE,
8       `prediction` DOUBLE DEFAULT NULL
9   );
10  select label as ACTUAL_VALUE,prediction as PREDICTED_VALUE from edureka_spark;
11
12
13
14
```

100%   79:10

Result Grid   Filter Rows: Search   Export:

ACTUAL_VALUE	PREDICTED_VALUE
170	171.56264302067007
108	111.39134115454388
243	242.47700053563278
492	488.28334354894434
260	258.0292152387924
170	171.56264302067007
170	171.56264302067007
214	215.8578611022839