

Project Set up:

Install python 3.7.6 (pyspark 2.4.5 have some issues with 3.8, so recommended to use 3.7.x)

<https://www.python.org/downloads/>

Install anaconda latest version (uses python 3.7.6)

<https://www.anaconda.com/distribution/>

Download code from git:

<https://github.com/rajasekhar-learn/ExplorePython>

## Start HDFS :

rajasekharv@MyPC-RAJ ~/hadoop

**\$ ./sbin/start-dfs.cmd**

Start mysql server:

rajasekharv@MyPC-RAJ /cygdrive/c/Program Files/mysql/bin

**\$ ./mysqld.exe --console**

mysqld: Could not create or access the registry key needed for the MySQL application to log to the Windows EventLog. Run the application with sufficient privileges once to create the key, add the key manually, or turn off logging for that application.

2020-04-10T17:46:17.058092Z 0 [Warning] TIMESTAMP with implicit DEFAULT value is deprecated. Please use --explicit\_defaults\_for\_timestamp server option (see documentation for more details).

2020-04-10T17:46:17.175487Z 0 [Note] --secure-file-priv is set to NULL. Operations related to importing and exporting data are disabled

2020-04-10T17:46:17.276109Z 0 [ERROR] Cannot open Windows EventLog; check privileges, or start server with --log\_syslog=0

2020-04-10T17:46:17.276241Z 0 [Note] C:\Program Files\mysql\bin\mysqld.exe (mysqld 5.7.29) starting as process 15472 ...

2020-04-10T17:46:18.082131Z 0 [Note] InnoDB: Mutexes and rw\_locks use Windows interlocked functions

2020-04-10T17:46:18.082695Z 0 [Note] InnoDB: Uses event mutexes

2020-04-10T17:46:18.083357Z 0 [Note] InnoDB: Memory barrier is not used

2020-04-10T17:46:18.083922Z 0 [Note] InnoDB: Compressed tables use zlib 1.2.11

2020-04-10T17:46:18.107671Z 0 [Note] InnoDB: Number of pools: 1

2020-04-10T17:46:18.146091Z 0 [Note] InnoDB: Not using CPU crc32 instructions

2020-04-10T17:46:18.160187Z 0 [Note] InnoDB: Initializing buffer pool, total size = 128M, instances = 1, chunk size = 128M

2020-04-10T17:46:18.182057Z 0 [Note] InnoDB: Completed initialization of buffer pool

2020-04-10T17:46:18.788020Z 0 [Note] InnoDB: Highest supported file format is Barracuda.

2020-04-10T17:46:24.856924Z 0 [Note] InnoDB: Creating shared tablespace for temporary tables  
2020-04-10T17:46:24.858005Z 0 [Note] InnoDB: Setting file '.\ibtmp1' size to 12 MB. Physically writing the file full; Please wait ...  
2020-04-10T17:46:25.239827Z 0 [Note] InnoDB: File '.\ibtmp1' size is now 12 MB.  
2020-04-10T17:46:25.350165Z 0 [Note] InnoDB: 96 redo rollback segment(s) found. 96 redo rollback segment(s) are active.  
2020-04-10T17:46:25.350568Z 0 [Note] InnoDB: 32 non-redo rollback segment(s) are active.  
2020-04-10T17:46:25.352827Z 0 [Note] InnoDB: Waiting for purge to start  
2020-04-10T17:46:25.403439Z 0 [Note] InnoDB: page\_cleaner: 1000ms intended loop took 7217ms. The settings might not be optimal. (flushed=0 and evicted=0, during the time.)  
2020-04-10T17:46:25.413285Z 0 [Note] InnoDB: 5.7.29 started; log sequence number 6224550  
2020-04-10T17:46:25.420845Z 0 [Note] InnoDB: Loading buffer pool(s) from C:\Program Files\mysql\data\ib\_buffer\_pool  
2020-04-10T17:46:25.490095Z 0 [Note] Plugin 'FEDERATED' is disabled.  
2020-04-10T17:46:26.645786Z 0 [Note] InnoDB: Buffer pool(s) load completed at 200410 23:16:26  
2020-04-10T17:46:27.043685Z 0 [Note] Found ca.pem, server-cert.pem and server-key.pem in data directory. Trying to enable SSL support using them.  
2020-04-10T17:46:27.044516Z 0 [Note] Skipping generation of SSL certificates as certificate files are present in data directory.  
2020-04-10T17:46:27.214434Z 0 [Warning] CA certificate ca.pem is self signed.  
2020-04-10T17:46:27.217049Z 0 [Note] Skipping generation of RSA key pair as key files are present in data directory.  
2020-04-10T17:46:27.448762Z 0 [Note] Server hostname (bind-address): '\*'; port: 3306  
2020-04-10T17:46:27.453526Z 0 [Note] IPv6 is available.  
2020-04-10T17:46:27.506573Z 0 [Note] - '::' resolves to '::';  
2020-04-10T17:46:27.507335Z 0 [Note] Server socket created on IP: '::'.  
2020-04-10T17:46:31.121062Z 0 [Note] Event Scheduler: Loaded 0 events  
2020-04-10T17:46:31.121781Z 0 [Note] C:\Program Files\mysql\bin\mysqld.exe: ready for connections.  
Version: '5.7.29' socket: " port: 3306 MySQL Community Server (GPL)  
2020-04-10T17:47:31.874853Z 2 [Note] Access denied for user 'root'@'localhost' (using password: NO)

## Connect mysql :

rajasekharv@MyPC-RAJ /cygdrive/c/Program Files/mysql/bin

**\$ ./mysql -u root -p**

Enter password: \*\*\*\*

Welcome to the MySQL monitor. Commands end with ; or \g.

Your MySQL connection id is 4

Server version: 5.7.29 MySQL Community Server (GPL)

Copyright (c) 2000, 2020, Oracle and/or its affiliates. All rights reserved.

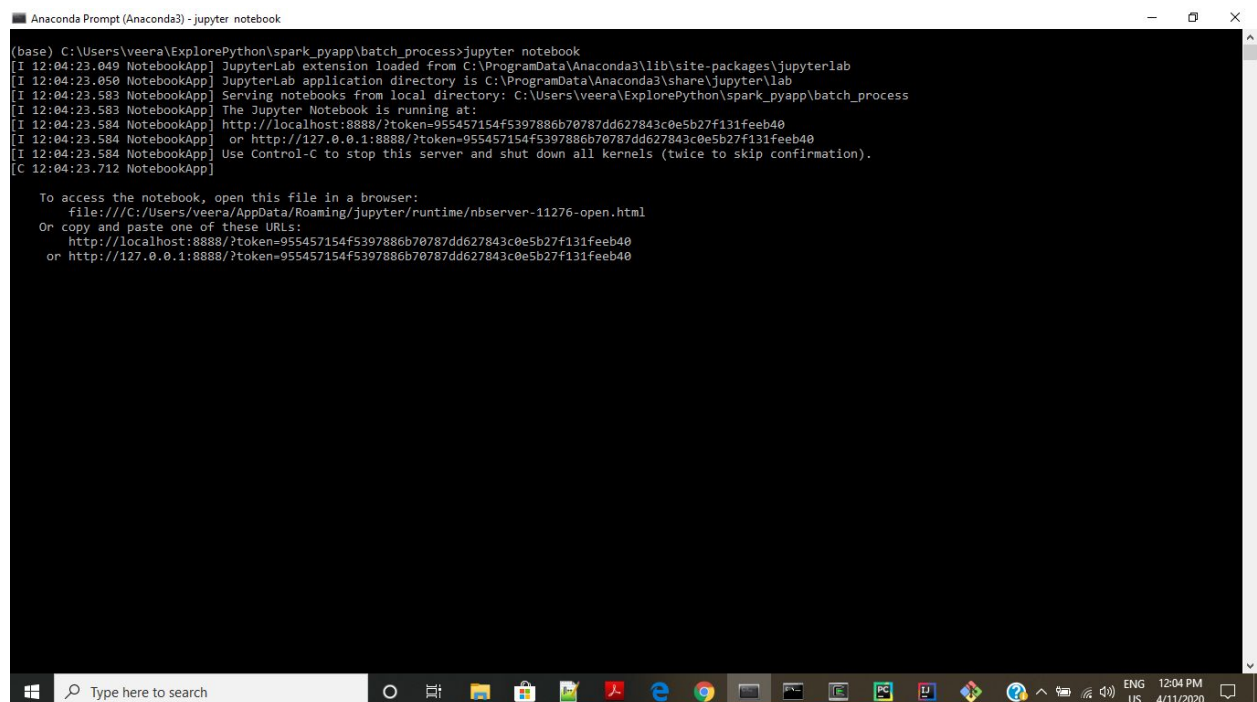
Oracle is a registered trademark of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

```
mysql> use hivedb;  
Database changed
```

Open jupyter notebook using anaconda prompt:

### Jupyter notebook

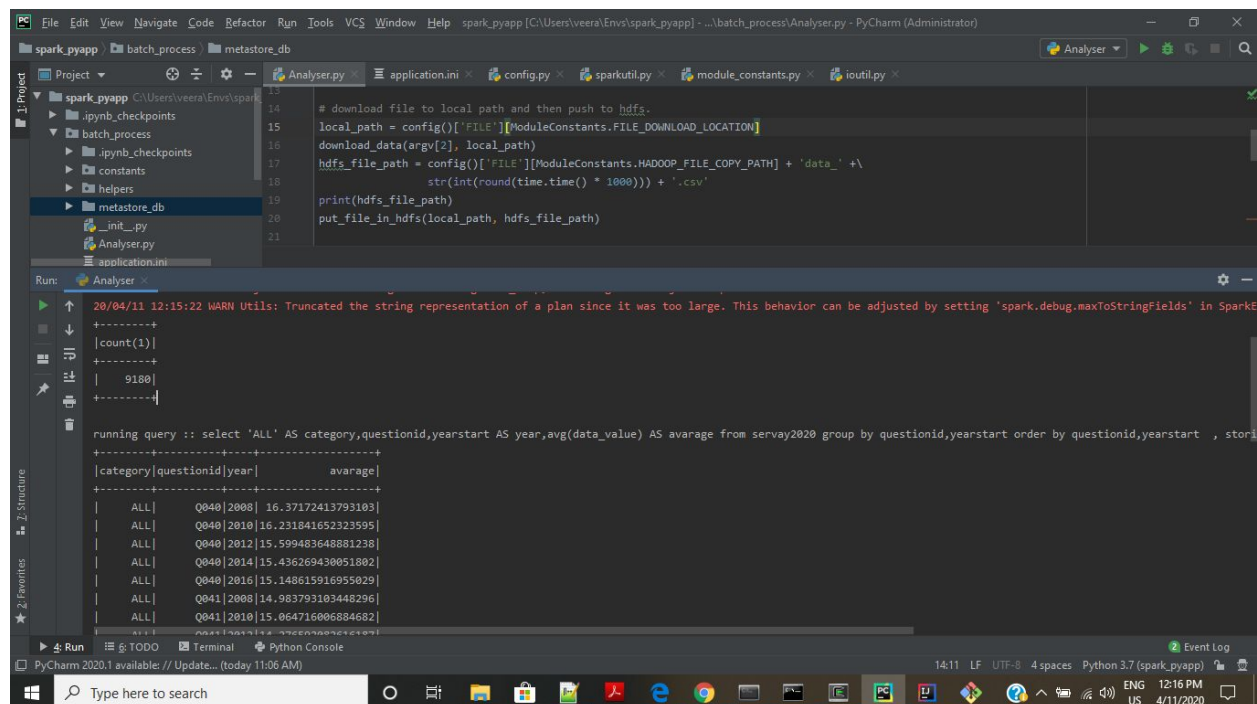
A screenshot of an Anaconda Prompt window titled "Anaconda Prompt (Anaconda3) - jupyter notebook". The terminal shows the command "(base) C:\Users\veera\ExplorePython\spark\_pyapp\batch\_process>jupyter notebook" and its output. The output includes log messages from JupyterLab and the Jupyter Notebook application, indicating that the server is running at http://localhost:8888. It also provides a file path to open in a browser and two URLs to access the notebook. The Windows taskbar is visible at the bottom, showing the search bar and various application icons.

```

Anaconda Prompt (Anaconda3) - jupyter notebook
(base) C:\Users\veera\ExplorePython\spark_pyapp\batch_process>jupyter notebook
[I 12:04:23.049 NotebookApp] JupyterLab extension loaded from C:\ProgramData\Anaconda3\lib\site-packages\jupyterlab
[I 12:04:23.050 NotebookApp] JupyterLab application directory is C:\ProgramData\Anaconda3\share\jupyter\lab
[I 12:04:23.583 NotebookApp] Serving notebooks from local directory: C:\Users\veera\ExplorePython\spark_pyapp\batch_process
[I 12:04:23.583 NotebookApp] The Jupyter Notebook is running at:
[I 12:04:23.584 NotebookApp] http://localhost:8888/?token=955457154f5397886b70787dd627843c0e5b27f131feeb40
[I 12:04:23.584 NotebookApp] or http://127.0.0.1:8888/?token=955457154f5397886b70787dd627843c0e5b27f131feeb40
[I 12:04:23.584 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 12:04:23.712 NotebookApp]

To access the notebook, open this file in a browser:
file:///C:/Users/veera/AppData/Roaming/jupyter/runtime/nbserver-11276-open.html
Or copy and paste one of these URLs:
http://localhost:8888/?token=955457154f5397886b70787dd627843c0e5b27f131feeb40
or http://127.0.0.1:8888/?token=955457154f5397886b70787dd627843c0e5b27f131feeb40
```

## IntelliJ execution:



## IntelliJ execution logs:

C:\Users\veera\Env\spark\_pyapp\Scripts\python.exe

C:/Users/veera/Env/spark\_pyapp/batch\_process/Analysers.py 1

<https://chronicdata.cdc.gov/views/735e-byxc/rows.csv?accessType=DOWNLOAD>

hdfs://localhost:9000/

hdfs://localhost:9000/data\_1586587303346.csv

Running shell command: hdfs dfs -put C:\\Users\\veera\\Desktop\\csv\\data.csv

hdfs://localhost:9000/data\_1586587303346.csv

0 b" b"

Setting default log level to "WARN".

To adjust logging level use `sc.setLogLevel(newLevel)`. For SparkR, use `setLogLevel(newLevel)`.

20/04/11 12:12:30 WARN ObjectStore: Version information not found in metastore.

hive.metastore.schema.verification is not enabled so recording the schema version 1.2.0

20/04/11 12:12:30 WARN ObjectStore: Failed to get database default, returning

NoSuchObjectException

20/04/11 12:12:33 WARN Utils: Truncated the string representation of a plan since it was too

large. This behavior can be adjusted by setting 'spark.debug.maxToStringFields' in

SparkEnv.conf.

20/04/11 12:12:36 WARN ObjectStore: Failed to get database global\_temp, returning

NoSuchObjectException

+-----+

```
|count(1)|
+-----+
|  9180|
+-----+
```

running query :: select 'ALL' AS category,questionid,yearstart AS year,avg(data\_value) AS  
avarage from servay2020 group by questionid,yearstart order by questionid,yearstart , storing  
results in :: servay\_analysis\_all\_results

```
+-----+-----+----+-----+
|category|questionid|year|      avarage|
+-----+-----+----+-----+
|  ALL|    Q040|2008| 16.37172413793103|
|  ALL|    Q040|2010|16.231841652323595|
|  ALL|    Q040|2012|15.599483648881238|
|  ALL|    Q040|2014|15.436269430051802|
|  ALL|    Q040|2016|15.148615916955029|
|  ALL|    Q041|2008|14.983793103448296|
|  ALL|    Q041|2010|15.064716006884682|
|  ALL|    Q041|2012|14.276592082616187|
|  ALL|    Q041|2014|13.943350604490504|
|  ALL|    Q041|2016| 13.53235294117647|
|  ALL|    Q060|2008|13.218181818181813|
|  ALL|    Q060|2010|13.046984126984121|
|  ALL|    Q060|2012|11.963057324840753|
|  ALL|    Q060|2014|11.450638977635784|
|  ALL|    Q060|2016|11.577635782747599|
+-----+-----+----+-----+
```

running query :: select 'FEMALE' AS category,questionid,yearstart AS year,avg(data\_value) AS  
avarage from servay2020 where gender ='Female' group by questionid,yearstart order by  
questionid,yearstart , storing results in :: servay\_analysis\_female\_results

```
+-----+-----+----+-----+
|category|questionid|year|      avarage|
+-----+-----+----+-----+
| FEMALE|    Q040|2008| 16.2962962962963|
| FEMALE|    Q040|2010| 16.23518518518518|
| FEMALE|    Q040|2012|15.648148148148145|
| FEMALE|    Q040|2014|15.411111111111113|
| FEMALE|    Q040|2016| 15.39259259259259|
| FEMALE|    Q041|2008| 13.91666666666667|
| FEMALE|    Q041|2010|14.161111111111111|
| FEMALE|    Q041|2012|13.594444444444445|
```

Process finished with exit code 0

```
jupyter app-notebook - Jupyter | + v
```

localhost:8888/notebooks/app-notebook.ipynb

```
In [1]: Import Findspark
import spark
findspark.find()

In [2]: Import libraries
from pyspark.conf import SparkConf
from pyspark.context import SparkContext
from pyspark.sql import SQLContext
from constants.mobile_contacts import MobileContacts

In [3]: local_path = "data_notebook.csv"
download_data("https://covid19data.cdc.gov/covid19-hypoc/usa-us-forecast?type=COVID-19", local_path)
HDFS_file_path = config["file"][MobileConstants.HADOOP_FILE_COPY_PATH] + local_path
df[isint(round(time.time()) * 1000)] = '-csv'
print(df_file_path)
Hdfs://localhost:8000/data_5160580931046.csv

In [4]: put_file_in_hdfs(local_path, HDFS_file_path)

Running shell command: hdfs dfs -put data_notebook.csv Hdfs://localhost:8000/data_5160580931046.csv
0 0" 0"

In [5]: batch_data_frame = SparkRDD.load_data("com.databricks.spark.csv", HDFS_file_path)

In [6]: [batch_data_frame = batch_data_frame \
    .selectColumnNames("Sign_Confidence_Limit", "High_Confidence_Limit") \
    .alias(batch_data_frame.alias())]

Out[6]: DataFrame[YearofStart: string, YearofEnd: string, LocationID: string, LocationName: string, DataSource: string,
Class: string, Topic: string, Question: string, Data_Value_Int: string, Data_Value_Type: string, Data_Value: double, Data_Valu
e_Alt: string, Data_Value_Potential_Symbol: string, Data_Value_Potential: string, Low_Confidence_Limit: string, High_Confidence_L
imit: string, Sample_Size: string, Total: string, Age: string, Gender: string, Race/Ethnicity: string, Occupation: string, Cla
ss_ID: string, Traveler: string, QuestionID: string, SubsetValuePath: string, Location: string, StratificationCategory: strin
g, Stratification: string, StratificationCategoryID: string, StratificationID: string]

In [7]: [batch_data_frame.write.mode("overwrite").saveAsTable(config["hive"]["MobileConstants.MDVE_APP_TABLE"])]

In [8]: query = "select count(*) from " + config["hive"][MobileConstants.MDVE_APP_TABLE]
SparkRDD.execute_query(query)

*****
|count|
|-----|
|  9848|
*****

In [9]: queries = config["queries"][MobileConstants.QDRRDS].split("\n")
tables = config["queries"][MobileConstants.RDRRD_TABLES].split(',')
SparkRDD.execute_query_store_results(queries, tables)

running query is select All AS category,monthid,yearstart AS year,avg(data_value) AS average from survey2020 group by mon
thid,yearstart order by monthid,yearstart -- storing results in survey_analysis_results

*****
|category|monthid|year|average|
|-----|-----|-----|-----|
|All|2020|12|3727941379288|
```

Home Page - Select or create app-notebook - Jupyter

localhost:8888/notebooks/app-notebook.ipynb

jupyter app-notebook Last Checkpoint: 2 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
In [7]: batch_data_frame.write.mode("overwrite").saveAsTable(config()['HIVE']["ModuleConstants.HIVE_APP_TABLE"])
```

```
In [8]: query = 'select count(*) from ' + config()['HIVE']["ModuleConstants.HIVE_APP_TABLE"]
SparkUtil.execute_show_query(query)
```

```
+-----+
|count(1)|
+-----+
|      9180|
+-----+
```

```
In [9]: queries = config()['QUERIES']["ModuleConstants.QUERIES"].split('^')
tabales = config()['QUERIES']["ModuleConstants.RESULT_TABLES"].split(',')
SparkUtil.execute_query_store_results(queries, tabales)
```

running query :: select 'ALL' AS category,questionid,yearstart AS year,avg(data\_value) AS avarage from servay2020 group by ques  
tionid,yearstart order by questionid,yearstart , storing results in :: servay\_analysis\_all\_results

```
+-----+
|category|questionid|year|          avarage|
+-----+
|ALL|Q040|2008| 16.37172413793103|
|ALL|Q040|2010| 16.231841652323595|
|ALL|Q040|2012| 15.599483648881238|
|ALL|Q040|2014| 15.436269430051802|
|ALL|Q040|2016| 15.148615916955029|
|ALL|Q041|2008| 14.983793103448296|
|ALL|Q041|2010| 15.064716006884682|
|ALL|Q041|2012| 14.276592082616187|
```

Home Page - Select or create app-notebook - Jupyter

localhost:8888/notebooks/app-notebook.ipynb

jupyter app-notebook Last Checkpoint: 2 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
In [9]: queries = config()['QUERIES']["ModuleConstants.QUERIES"].split('^')
tabales = config()['QUERIES']["ModuleConstants.RESULT_TABLES"].split(',')
SparkUtil.execute_query_store_results(queries, tabales)
```

running query :: select 'ALL' AS category,questionid,yearstart AS year,avg(data\_value) AS avarage from servay2020 group by ques  
tionid,yearstart order by questionid,yearstart , storing results in :: servay\_analysis\_all\_results

```
+-----+
|category|questionid|year|          avarage|
+-----+
|ALL|Q040|2008| 16.37172413793103|
|ALL|Q040|2010| 16.231841652323595|
|ALL|Q040|2012| 15.599483648881238|
|ALL|Q040|2014| 15.436269430051802|
|ALL|Q040|2016| 15.148615916955029|
|ALL|Q041|2008| 14.983793103448296|
|ALL|Q041|2010| 15.064716006884682|
|ALL|Q041|2012| 14.276592082616187|
|ALL|Q041|2014| 13.943350604490504|
|ALL|Q041|2016| 13.53235294117647|
|ALL|Q060|2008| 13.218181818181813|
|ALL|Q060|2010| 13.046984126984121|
|ALL|Q060|2012| 11.963057324840753|
|ALL|Q060|2014| 11.450638977635784|
|ALL|Q060|2016| 11.577635782747599|
```

running query :: select 'FEMALE' AS category,questionid,yearstart AS year,avg(data\_value) AS avarage from servay2020 where gend  
er = 'Female' group by questionid,yearstart order by questionid,yearstart , storing results in :: servay\_analysis\_female\_results

```
+-----+
|category|questionid|year|          avarage|
+-----+
```



Home Page - Select or creat app-notebook - Jupyter X +

localhost:8888/notebooks/app-notebook.ipynb

jupyter app-notebook Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```

ALL|Q000|2008|15.218181818181818|
ALL|Q000|2010|13.046984126984121|
ALL|Q000|2012|11.963057324840753|
ALL|Q000|2014|11.450638977635784|
ALL|Q000|2016|11.577635782747599|
+-----+
running query :: select 'FEMALE' AS category,questionid,yearstart AS year,avg(data_value) AS avarage from servay2020 where gend
er ='Female' group by questionid,yearstart order by questionid,yearstart , storing results in :: servay_analysis_female_results
+-----+
|category|questionid|year|          avarage|
+-----+
| FEMALE|Q040|2008| 16.2962962962963|
| FEMALE|Q040|2010| 16.23518518518518|
| FEMALE|Q040|2012| 15.648148148148145|
| FEMALE|Q040|2014| 15.411111111111113|
| FEMALE|Q040|2016| 15.39259259259259|
| FEMALE|Q041|2008| 13.916666666666667|
| FEMALE|Q041|2010| 14.161111111111111|
| FEMALE|Q041|2012| 13.594444444444445|
| FEMALE|Q041|2014| 13.198148148148144|
| FEMALE|Q041|2016| 12.957407407407413|
| FEMALE|Q060|2008| 12.72962962962963|
| FEMALE|Q060|2010| 12.38148148148148|
| FEMALE|Q060|2012| 11.264814814814816|
| FEMALE|Q060|2014| 10.725925925925926|
| FEMALE|Q060|2016| 10.996296296296292|
+-----+

In [12]: SparkUtil.spark_session().stop()

```

Type here to search

11:51 AM 4/11/2020

Home Page - Select or creat app-notebook - Jupyter X +

localhost:8888/notebooks/app-notebook.ipynb

jupyter app-notebook Last Checkpoint: 3 hours ago (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```

In [7]: batch_data_frame.write.mode("overwrite").saveAsTable(config["HIVE"]['MetdataConstants.MDVE_APP_TABLE'])

In [8]: query = "select count(*) from " + config["HIVE"]['MetdataConstants.MDVE_APP_TABLE']
SparkUtil.execute_show_query(query)

+-----+
|count(*)|
+-----+
|      9388|
+-----+

In [9]: queries = config["QUESTIONS"]['MetdataConstants.QUESTIONS'].split(',')
         tables = config["QUESTIONS"]['MetdataConstants.QUESTIONS'].split(',')
SparkUtil.execute_show_results(queries, tables)

running query :: select 'ALL' AS category,questionid,yearstart AS year,avg(data_value) AS avarage from servay2020 group by ques
tionid,yearstart order by questionid,yearstart , storing results in :: servay_analysis_all_results
+-----+
|category|questionid|year|          avarage|
+-----+
| ALL|Q000|2008| 16.217341731582|
| ALL|Q000|2010| 16.213463222395|
| ALL|Q000|2012| 15.109614881220|
| ALL|Q000|2014| 15.4028343081582|
| ALL|Q000|2016| 15.1481510915629|
| ALL|Q041|2008| 14.36379338448295|
| ALL|Q041|2010| 14.36473889889888|
| ALL|Q041|2012| 14.27633082616187|
| ALL|Q041|2014| 14.36133883889888|
| ALL|Q041|2016| 13.5321329413767|
| ALL|Q060|2008| 12.23383883889888|
| ALL|Q060|2010| 13.48408423884121|
| ALL|Q060|2012| 12.363057324840753|
| ALL|Q060|2014| 12.40818077635784|
| ALL|Q060|2016| 12.577635782747599|
+-----+

running query :: select 'FEMALE' AS category,questionid,yearstart AS year,avg(data_value) AS avarage from servay2020 where gend
er ='Female' group by questionid,yearstart order by questionid,yearstart , storing results in :: servay_analysis_female_results
+-----+
|category|questionid|year|          avarage|
+-----+
| FEMALE|Q040|2008| 16.2962962962963|
| FEMALE|Q040|2010| 16.23518518518518|
| FEMALE|Q040|2012| 15.648148148148145|
| FEMALE|Q040|2014| 15.411111111111113|
| FEMALE|Q040|2016| 15.39259259259259|
| FEMALE|Q041|2008| 13.916666666666667|
| FEMALE|Q041|2010| 14.161111111111111|
| FEMALE|Q041|2012| 13.594444444444445|
| FEMALE|Q041|2014| 13.198148148148144|
| FEMALE|Q041|2016| 12.957407407407413|
| FEMALE|Q060|2008| 12.72962962962963|
| FEMALE|Q060|2010| 12.38148148148148|
| FEMALE|Q060|2012| 11.264814814814816|
| FEMALE|Q060|2014| 10.725925925925926|
| FEMALE|Q060|2016| 10.996296296296292|
+-----+

In [10]: SparkUtil.spark_session().stop()

```

Type here to search

12:07 PM 4/11/2020

Hive metastore db tables:



```
/cydrive/c/Program Files/mysql/bin
skewed_col_names
skewed_col_value_loc_map
skewed_string_list
skewed_string_list_values
skewed_values
sort_cols
tbl_col_stats
table_params
tbl_col_privs
tbl_privs
tbls
txn_components
txns
type_fields
types
version
write_set
-----
57 rows in set (0.11 sec)

mysql> select * from tbls;
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| TBL_ID | CREATE_TIME | DB_ID | LAST_ACCESS_TIME | OWNER | RETENTION | SD_ID | TBL_NAME | TBL_TYPE | VIEW_EXPANDED_TEXT | VIEW_ORIGINAL_TEXT | IS_REWRITE_ENAB |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 17 | 1584638317 | 1 | 0 | rajasekharv | 0 | 17 | servay_analysis_results | MANAGED_TABLE | NULL | NULL | |
| 56 | 1584923534 | 1 | 0 | rajasekharv | 0 | 56 | servay_analysis_all_results | MANAGED_TABLE | NULL | NULL | |
| 57 | 1584923554 | 1 | 0 | rajasekharv | 0 | 57 | servay_analysis_female_results | MANAGED_TABLE | NULL | NULL | |
| 61 | 1585658618 | 1 | 0 | rajasekharv | 0 | 61 | survey_data_dump | MANAGED_TABLE | NULL | NULL | |
| 96 | 1586049227 | 1 | 0 | rajasekharv | 0 | 96 | survey_analysis_all_results | MANAGED_TABLE | NULL | NULL | |
| 97 | 1586049251 | 1 | 0 | rajasekharv | 0 | 97 | survey_analysis_female_results | MANAGED_TABLE | NULL | NULL | |
| 111 | 1586543261 | 1 | 0 | rajasekharv | 0 | 111 | survey001 | MANAGED_TABLE | NULL | NULL | |
| 116 | 1586572421 | 1 | 0 | rajasekharv | 0 | 116 | servay_data_dump | MANAGED_TABLE | NULL | NULL | |
| 126 | 1586574323 | 1 | 0 | rajasekharv | 0 | 126 | servay2020 | MANAGED_TABLE | NULL | NULL | |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
9 rows in set (0.13 sec)

mysql>
```

## Mysql results tables:

```
/cydrive/c/Program Files/mysql/bin
mysql> select * from survey_analysis_all_results;
+-----+-----+-----+-----+
| category | questionid | year | avarage |
+-----+-----+-----+-----+
| ALL | Q040 | 2012 | 15.599483648881241 |
| ALL | Q040 | 2014 | 15.436269430051802 |
| ALL | Q040 | 2008 | 16.37172413793103 |
| ALL | Q040 | 2010 | 16.2218416523236 |
| ALL | Q040 | 2016 | 15.148615916955029 |
| ALL | Q041 | 2008 | 14.983793103448294 |
| ALL | Q041 | 2010 | 15.064716006884684 |
| ALL | Q041 | 2012 | 14.276592082616189 |
| ALL | Q041 | 2014 | 13.943550604490506 |
| ALL | Q060 | 2008 | 13.218181818181815 |
| ALL | Q041 | 2016 | 13.532352941176468 |
| ALL | Q060 | 2010 | 13.046984126984121 |
| ALL | Q060 | 2012 | 11.961057324840754 |
| ALL | Q060 | 2014 | 11.450638977635782 |
| ALL | Q060 | 2016 | 11.5776357827476 |
+-----+-----+-----+-----+
15 rows in set (0.00 sec)

mysql> select * from survey_analysis_female_results;
+-----+-----+-----+-----+
| category | questionid | year | avarage |
+-----+-----+-----+-----+
| FEMALE | Q040 | 2012 | 15.648148148148147 |
| FEMALE | Q040 | 2010 | 16.23518518518518 |
| FEMALE | Q040 | 2008 | 16.296296296296298 |
| FEMALE | Q040 | 2016 | 15.39259259259259 |
| FEMALE | Q040 | 2014 | 15.411111111111111 |
| FEMALE | Q041 | 2008 | 13.916666666666668 |
| FEMALE | Q041 | 2010 | 14.161111111111111 |
| FEMALE | Q041 | 2012 | 13.594444444444447 |
| FEMALE | Q041 | 2014 | 13.198148148148144 |
| FEMALE | Q041 | 2016 | 12.957407407407413 |
| FEMALE | Q060 | 2008 | 12.729629629629626 |
| FEMALE | Q060 | 2010 | 12.581481481481481 |
| FEMALE | Q060 | 2012 | 11.264814814814816 |
| FEMALE | Q060 | 2014 | 10.725925925925925 |
| FEMALE | Q060 | 2016 | 10.99629629629629 |
+-----+-----+-----+-----+
15 rows in set (0.00 sec)

mysql> show tables;
+-----+
| Tables_in_metricsdb |
+-----+
| survey_analysis_all_results |
| survey_analysis_female_results |
+-----+
```

After code refactoring final note book: analyser-notebook

[illegible]