



CIS5200 Lab Tutorial Presentation



Authors: Anusha Valasapalli
Lohitha Karmuru
Ruta Antaliya
Surya Kiran Golagani

Instructor: Jongwook Woo

Date: 05/11/2022

Lab Tutorial

Anusha(avalasa@calstatela.edu)
Lohitha(nkarmur@calstatela.edu)
Ruta(rantali@calstatela.edu)
SuryaKiran(sgolaga2@calstatela.edu)
03/05/2022

Pain Pills in the USA

Objectives

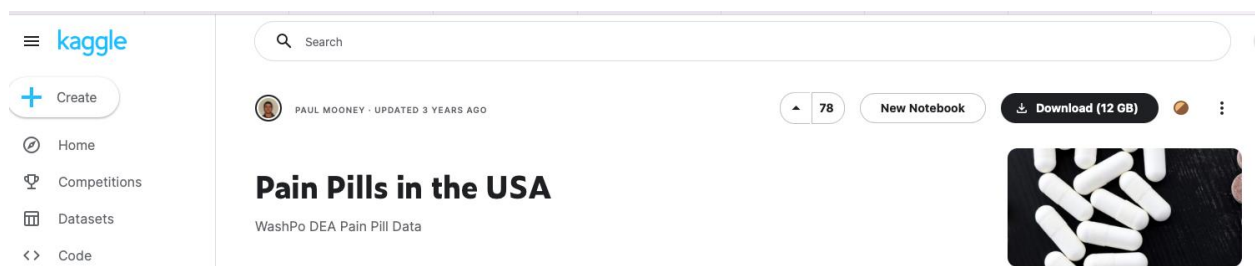
- Download the Pain pills in the USA dataset from the Kaggle website
- Copy the dataset from the local system to the Linux system
- Create HDFS directories
- Copy the data from Linux to HDFS
- Create hive tables
- HiveQL commands to perform the analysis
- Visualization using tableau and excel

Specifications

- **Platform:** AWS EMR
- **CPU Speed:** 2.50GHz
- **# Of CPU Cores:**4
- **# Of nodes:** 7
- **Total Memory Size:** 500GB

STEP 1: Download the Pain pills in the USA dataset from the Kaggle website

Link: <https://www.kaggle.com/datasets/paultimothymooney/pain-pills-in-the-usa>



STEP 2: Copy the dataset from the local system to the Linux system

scp /Users/anushavalasapalli/Desktop/arcos_all_washpost.tsv [avalasa@52.12.203.116:~/](mailto:avalasa@52.12.203.116)

```
Last login: Tue May 3 18:18:10 on ttys001
anushavalasapalli@Kalyanams-MBP ~ % scp /Users/anushavalasapalli/Desktop/arcos_all_washpost.tsv avalasa@52.12.203.116:~/
avalasa@52.12.203.116's password:
/Users/anushavalasapalli/Desktop/arcos_all_washpost.tsv: No such file or directory
anushavalasapalli@Kalyanams-MBP ~ % scp /Users/anushavalasapalli/Desktop/arcos_all_washpost.tsv avalasa@52.12.203.116:~/
avalasa@52.12.203.116's password:
arcos_all_washpost.tsv                                83% 62GB 20.9MB/s 09:59 ETAclien
t_loop: send disconnect: Broken pipe
lost connection
anushavalasapalli@Kalyanams-MBP ~ % scp /Users/anushavalasapalli/Desktop/arcos_all_washpost.tsv avalasa@52.12.203.116:~/
avalasa@52.12.203.116's password:
arcos_all_washpost.tsv                                100% 75GB 41.4MB/s 30:43
anushavalasapalli@Kalyanams-MBP ~ %
```

STEP 3: Create 4 HDFS directories for pain pills in the USA CITY, STATE, and popular pain pills in all the states and particular states like CA, TX, FL

```
hdfs dfs -mkdir /user/avalasa/PainPillsInTheUSA
hdfs dfs -ls /user/avalasa/PainPillsInTheUSA
```

```
hdfs dfs -mkdir /user/avalasa/PainPillsInTheUSA/PainPillsInTheState
hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/PainPillsInTheState
```

```
hdfs dfs -mkdir /user/avalasa/PainPillsInTheUSA/PainPillsInTheCity
hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/PainPillsInTheCity
```

```
hdfs dfs -mkdir /user/avalasa/PainPillsInTheUSA/MostPopularPainPills
hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/MostPopularPainPills
```

```
hdfs dfs -mkdir /user/avalasa/PainPillsInTheUSA/MostPopularPainPillsInParticularStates
hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/MostPopularPainPillsInParticularStates
```

```

-bash-4.2$ ls
arcos_all_washpost.tsv
-bash-4.2$ hdfs dfs -ls
-bash-4.2$ hdfs dfs -mkdir /user/avalasa/PainPillsInTheUSA
-bash-4.2$ hdfs dfs -ls /user/avalasa/
Found 1 items
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:23 /user/avalasa/PainPillsInTheUSA
-bash-4.2$ hdfs dfs -mkdir /user/avalasa/PainPillsInTheUSA/PainPillsInTheState
-bash-4.2$ hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/
Found 1 items
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:25 /user/avalasa/PainPillsInTheUSA/PainPillsInTheState
-bash-4.2$ hdfs dfs -mkdir /user/avalasa/PainPillsInTheUSA/PainPillsInTheCity
-bash-4.2$ hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/
Found 2 items
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:26 /user/avalasa/PainPillsInTheUSA/PainPillsInTheCity
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:25 /user/avalasa/PainPillsInTheUSA/PainPillsInTheState
-bash-4.2$ hdfs dfs -mkdir /user/avalasa/PainPillsInTheUSA/MostPopularPainPills
-bash-4.2$ hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/
Found 3 items
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:28 /user/avalasa/PainPillsInTheUSA/MostPopularPainPills
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:26 /user/avalasa/PainPillsInTheUSA/PainPillsInTheCity
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:25 /user/avalasa/PainPillsInTheUSA/PainPillsInTheState
-bash-4.2$ hdfs dfs -mkdir /user/avalasa/PainPillsInTheUSA/MostPopularPainPillsInParticularStates
-bash-4.2$ hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/MostPopularPainPillsInParticularStates
-bash-4.2$ hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/
Found 4 items
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:28 /user/avalasa/PainPillsInTheUSA/MostPopularPainPills
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:30 /user/avalasa/PainPillsInTheUSA/MostPopularPainPillsInParticul
arStates
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:26 /user/avalasa/PainPillsInTheUSA/PainPillsInTheCity
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:25 /user/avalasa/PainPillsInTheUSA/PainPillsInTheState
-bash-4.2$ hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/
Found 4 items
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:28 /user/avalasa/PainPillsInTheUSA/MostPopularPainPills
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:30 /user/avalasa/PainPillsInTheUSA/MostPopularPainPillsInParticul
arStates
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:26 /user/avalasa/PainPillsInTheUSA/PainPillsInTheCity
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:25 /user/avalasa/PainPillsInTheUSA/PainPillsInTheState
-bash-4.2$ ls

```

STEP 4: Copy the data from Linux to HDFS

`hdfs dfs -put arcos_all_washpost.tsv/user/avalasa/PainPillsInTheUSA/`

```

arcos_all_washpost.tsv
-bash-4.2$ rm arcos_all_washpost.tsv
-bash-4.2$ ls
-bash-4.2$ hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/
Found 4 items
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:28 /user/avalasa/PainPillsInTheUSA/MostPopularPainPills
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:30 /user/avalasa/PainPillsInTheUSA/MostPopularPainPillsInParticul
arStates
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:26 /user/avalasa/PainPillsInTheUSA/PainPillsInTheCity
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:25 /user/avalasa/PainPillsInTheUSA/PainPillsInTheState
-bash-4.2$ ls
arcos_all_washpost.tsv
-bash-4.2$ hdfs dfs -put arcos_all_washpost.tsv /user/avalasa/PainPillsInTheUSA/

```

`hdfs dfs -cat /user/avalasa/PainPillsInTheUSA/arcos_all_washpost.tsv`

`hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/`

```
-bash-4.2$ hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/
Found 5 items
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:28 /user/avalasa/PainPillsInTheUSA/MostPopularPainPills
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:30 /user/avalasa/PainPillsInTheUSA/MostPopularPainPillsInParticularStates
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:26 /user/avalasa/PainPillsInTheUSA/PainPillsInTheCity
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:25 /user/avalasa/PainPillsInTheUSA/PainPillsInTheState
-rw-r--r-- 1 avalasa hdfsadmingroup 79996770669 2022-05-04 03:36 /user/avalasa/PainPillsInTheUSA/arcos_all_washpost.tsv
-bash-4.2$
```

```
hdfs dfs -cat /user/avalasa/PainPillsInTheUSA/arcos_all_washpost.tsv | head -n 5
```

STEP 5: Create hive tables

beeline

USE avalasa;

```
CREATE EXTERNAL TABLE if not exists Pain_Pills_In_The_USA(
REPORTER_DEA_NO string,
REPORTER_BUS_ACT string,
REPORTER_NAME string,
REPORTER_ADDL_CO_INFO string,
REPORTER_ADDRESS1 string,
REPORTER_ADDRESS2 string,
REPORTER_CITY string,
REPORTER_STATE string,
REPORTER_ZIP string,
REPORTER_COUNTY string,
BUYER_DEA_NO string,
BUYER_BUS_ACT string,
BUYER_NAME string,
BUYER_ADDL_CO_INFO string,
BUYER_ADDRESS1 string,
BUYER_ADDRESS2 string,
BUYER_CITY string,
BUYER_STATE string,
BUYER_ZIP string,
BUYER_COUNTY string,
TRANSACTION_CODE string,
DRUG_CODE string,
NDC_NO string,
DRUG_NAME string,
QUANTITY DOUBLE,
UNIT string,
ACTION_INDICATOR string,
ORDER_FORM_NO string,
CORRECTION_NO string,
STRENGTH string,
TRANSACTION_DATE string,
CALC_BASE_WT_IN_GM DOUBLE,
DOSAGE_UNIT DOUBLE,
```

```

TRANSACTION_ID DOUBLE,
Product_Name string,
Ingredient_Name string,
Measure string,
MME_Conversion_Factor INT,
Combined_Labeler_Name string,
Revised_Company_Name string,
Reporter_family string,
dos_str DOUBLE)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'
STORED AS TEXTFILE LOCATION '/user/avalasa/PainPillsInTheUSA/'
TBLPROPERTIES ('skip.header.line.count'='1');

```

```

0: jdbc:hive2://ip-172-30-1-120.us-west-2.com> CREATE EXTERNAL TABLE if not exists Pain_Pills_In_The_USA(
. . . . .
-> REPORTER_DEA_NO string,
-> REPORTER_BUS_ACT string,
-> REPORTER_NAME string,
-> REPORTER_ADDL_CO_INFO string,
-> REPORTER_ADDRESS1 string,
-> REPORTER_ADDRESS2 string,
-> REPORTER_CITY string,
-> REPORTER_STATE string,
-> REPORTER_ZIP string,
-> REPORTER_COUNTY string,
-> BUYER_DEA_NO string,
-> BUYER_BUS_ACT string,
-> BUYER_NAME string,
-> BUYER_ADDL_CO_INFO string,
-> BUYER_ADDRESS1 string,
-> BUYER_ADDRESS2 string,
-> BUYER_CITY string,
-> BUYER_STATE string,
-> BUYER_ZIP string,
-> BUYER_COUNTY string,
-> TRANSACTION_CODE string,
-> DRUG_CODE string,
-> NDC_NO string,
-> DRUG_NAME string,
-> QUANTITY DOUBLE,
-> UNIT string,
-> ACTION_INDICATOR string,
-> ORDER_FORM_NO string,
-> CORRECTION_NO string,
-> STRENGTH string,
-> TRANSACTION_DATE string,
-> CALC_BASE_WT_IN_GM DOUBLE,
-> DOSAGE_UNIT DOUBLE,
-> TRANSACTION_ID DOUBLE,
-> Product_Name string,
-> Ingredient_Name string,
-> Measure string,
-> MME_Conversion_Factor INT,
-> Combined_Labeler_Name string,
-> Revised_Company_Name string,
-> Reporter_family string,
-> dos_str DOUBLE
. . . . .

```

```
select count(*) from Pain_Pills_In_The_USA;
```

```

INFO : Map 1: 0(+1)/1 Reducer 2: 0/1
INFO : Map 1: 0(+1)/1 Reducer 2: 0/1
INFO : Map 1: 0(+1)/1 Reducer 2: 0/1
INFO : Map 1: 0(+1)/1 Reducer 2: 0/1
INFO : Map 1: 0(+1)/1 Reducer 2: 0/1
INFO : Map 1: 0(+1)/1 Reducer 2: 0/1
INFO : Map 1: 0(+1)/1 Reducer 2: 0/1
INFO : Map 1: 0(+1)/1 Reducer 2: 0/1
INFO : Map 1: 0(+1)/1 Reducer 2: 0/1
INFO : Map 1: 1/1 Reducer 2: 0(+1)/1
INFO : Map 1: 1/1 Reducer 2: 1/1
INFO : Completed executing command(queryId=hive_20220504034644_62f07b4d-1903-4f69-8643-d217ee159957); Time taken: 259.113 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
+-----+
| _c0 |
+-----+
| 178598026 |
+-----+
1 row selected (259.962 seconds)
0: jdbc:hive2://ip-172-30-1-120.us-west-2.com> select * from Pain_Pills_In_The_USA limit 10;
INFO : Compiling command(queryId=hive_20220504035345_7212137a-888e-46d4-8826-f261f602ee14); select * from Pain_Pills_In_The_USA limit 10

```

```
select BUYER_STATE,sum(QUANTITY), TRANSACTION_DATE
from Pain_Pills_In_The_USA
group by BUYER_STATE, TRANSACTION_DATE limit 50;
```

It takes 21 minutes to select data using group by statement.

```
AK      56.0    01082012
AK      409.0   01062010
AK      176.0   01022011
AK      280.0   01092006
AK      56.0    01012010
AK      238.0   01052007
AK      438.0   01042007
AK      524.0   01082008
AK      94.0    01032010
AK      442.0   01052010
AK      461.0   01082010
AK      70.0    01012007
AK      596.0   01052011
AK      418.0   01052012
AK      537.0   01062009
AK      463.0   01092007
AK      372.0   01032008
AE       8.0    11272006
AK      509.0   01082009
AK      154.0   01022006
AK      582.0   01062011
AK      637.0   01042011
AK      521.0   01052009
AK      284.0   01072009
AK      495.0   01042008
AK      598.0   01022008
AK      206.0   01062006
AK      84.0    01082006
AK      492.0   01022009
AK      390.0   01032007
AK      394.0   01022012
AK      703.0   01042012
AK      476.0   01082007
AK      57.0    01062007
AK      43.0    01072007
AK      257.0   01052006
AK      505.0   01022007
AK      620.0   01072008
AK      407.0   01042006
AK      68.0    01042009
AK      280.0   01032006
AK      323.0   01032012
AK      596.0   01042010
AK      10.0    01072012
AK      589.0   01072010
-----
50 rows selected (1271.492 seconds)
0: jdbc:hive2://ip-172-30-1-132.us-west-2.com>
```

State Pain Pills Count: -

```
DROP TABLE IF EXISTS Pain_Pills_In_The_USA_Per_State;
CREATE TABLE Pain_Pills_In_The_USA_Per_State
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION '/user/avalasa/PainPillsInTheUSA/PainPillsInTheState'
AS
```

```
select BUYER_STATE, sum(QUANTITY) from Pain_Pills_In_The_USA group by
BUYER_STATE order by 2 desc;
select * from Pain_Pills_In_The_USA_Per_State limit 10;
```

```
0: jdbc:hive2://ip-172-30-1-120.us-west-2.com> select * from Pain_Pills_In_The_USA_Per_State limit 10;
INFO : Compiling command(queryId=hive_20220504040616_58bf4b13-31fc-46fa-9c1b-626589ff066a): select * from Pain_Pills_In_The_USA_Per_State limit 10
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldsSchemas:[FieldSchema(name:pain_pills_in_the_usa_per_state.buyer_state, type:string, comment:null), FieldSchema(name:pain_pills_in_the_usa_per_state.cl, type:double, comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20220504040616_58bf4b13-31fc-46fa-9c1b-626589ff066a); Time taken: 0.851 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20220504040616_58bf4b13-31fc-46fa-9c1b-626589ff066a): select * from Pain_Pills_In_The_USA_Per_State limit 10
INFO : Completed executing command(queryId=hive_20220504040616_58bf4b13-31fc-46fa-9c1b-626589ff066a); Time taken: 0.001 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
-----
| pain_pills_in_the_usa_per_state.buyer_state | pain_pills_in_the_usa_per_state.cl |
-----+-----+-----
| CA | 5.4057115E7 |
| FL | 3.9053364E7 |
| TX | 2.5605244E7 |
| NY | 2.2543212E7 |
| PA | 2.0764697E7 |
| OH | 2.0597201E7 |
| NC | 1.6983664E7 |
| TN | 1.5199778E7 |
| GA | 1.417917E7 |
| MI | 1.398949E7 |
-----
10 rows selected (0.876 seconds)
0: jdbc:hive2://ip-172-30-1-120.us-west-2.com>
```

```
hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/PainPillsInTheState
```

```
hdfs dfs -cat /user/avalasa/PainPillsInTheUSA/PainPillsInTheState/000000_0 | tail -n 2
```

```
hdfs dfs -get /user/avalasa/PainPillsInTheUSA/PainPillsInTheState/000000_0  
PainPills_In_The_USA_States.txt
```

```
readlink -f PainPills_In_The_USA_States.txt // to know linux system path
```

```
-bash-4.2$ hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/  
Found 5 items  
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:28 /user/avalasa/PainPillsInTheUSA/MostPopularPainPills  
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:30 /user/avalasa/PainPillsInTheUSA/MostPopularPainPillsInParticularStates  
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:26 /user/avalasa/PainPillsInTheUSA/PainPillsInTheCity  
drwxr-xr-x - avalasa hdfsadmingroup 0 2022-05-04 01:25 /user/avalasa/PainPillsInTheUSA/PainPillsInTheState  
-rw-r--r-- 1 avalasa hdfsadmingroup 79996770669 2022-05-04 03:36 /user/avalasa/PainPillsInTheUSA/arcos_all_washpost.tsv  
-bash-4.2$ hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/PainPillsInTheState/  
Found 1 items  
-rw-r--r-- 1 avalasa hdfsadmingroup 751 2022-05-04 04:04 /user/avalasa/PainPillsInTheUSA/PainPillsInTheState/000000_0  
-bash-4.2$ hdfs dfs -cat /user/avalasa/PainPillsInTheUSA/PainPillsInTheState/000000_0 | tail -n 2  
PW,137.0  
AE,11.0  
-bash-4.2$ ls  
arcos_all_washpost.tsv  
-bash-4.2$ hdfs dfs -get /user/avalasa/PainPillsInTheUSA/PainPillsInTheState/000000_0 PainPills_In_The_USA_States.txt  
-bash-4.2$ ls  
arcos_all_washpost.tsv PainPills_In_The_USA_States.txt  
-bash-4.2$ readlink -f PainPills_In_The_USA_States.txt  
/home/avalasa/PainPills_In_The_USA_States.txt  
-bash-4.2$ █
```

```
scp avalasa@52.12.203.116:/home/avalasa/PainPills_In_The_USA_States.txt .
```

```
PainPills_In_The_USA_States.txt  
anushavalasapalli@Kalyanams-MBP ~ % ls  
Applications Library PainPills_In_The_USA_States.txt PycharmProjects senti_out.xlsx  
Desktop Movies Pictures env top10country10.csv  
Documents Music Postman refersperday12.csv ~$senti_out.xlsx  
Downloads OneDrive - Villanova University Public senti_out.csv  
anushavalasapalli@Kalyanams-MBP ~ % cd Documents  
anushavalasapalli@Kalyanams-MBP Documents % scp avalasa@52.12.203.116:/home/avalasa/PainPills_In_The_USA_States.txt .  
avalasa@52.12.203.116's password:  
PainPills_In_The_USA_States.txt  
anushavalasapalli@Kalyanams-MBP Documents % █  
100% 751 20.8KB/s 00:00
```

City Pain Pills Count:-

```
DROP TABLE IF EXISTS Pain_Pills_In_The_USA_Per_Cities;  
CREATE TABLE Pain_Pills_In_The_USA_Per_Cities  
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','  
STORED AS TEXTFILE LOCATION '/user/avalasa/PainPillsInTheUSA/PainPillsInTheCity'  
AS  
select BUYER_CITY, sum(QUANTITY)  
from Pain_Pills_In_The_USA  
group by BUYER_CITY order by 2 desc;
```



```
0: jdbc:hive2://ip-172-30-1-120.us-west-2.com> DROP TABLE IF EXISTS Pain_Pills_In_The_USA_Per_Cities;
INFO : Compiling command(queryId=hive_20220504043053_eb66f423-78ee-4449-a42a-9e5f9186b441): DROP TABLE IF EXISTS Pain_Pills_In_The_USA_Per_Cities
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO : Completed compiling command(queryId=hive_20220504043053_eb66f423-78ee-4449-a42a-9e5f9186b441); Time taken: 0.143 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20220504043053_eb66f423-78ee-4449-a42a-9e5f9186b441): DROP TABLE IF EXISTS Pain_Pills_In_The_USA_Per_Cities
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20220504043053_eb66f423-78ee-4449-a42a-9e5f9186b441); Time taken: 0.327 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
No rows affected (0.48 seconds)
0: jdbc:hive2://ip-172-30-1-120.us-west-2.com> CREATE TABLE Pain_Pills_In_The_USA_Per_Cities
. . . . .> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
. . . . .> STORED AS TEXTFILE LOCATION '/user/avalasa/PainPillsInTheUSA/PainPillsInTheCity'
. . . . .> AS
. . . . .> select BUYER_CITY, sum(QUANTITY) from Pain_Pills_In_The_USA group by BUYER_CITY order by 2 desc;
```

select * from Pain_Pills_In_The_USA_Per_Cities limit 10;

```
11 rows selected (0.569 seconds)
0: jdbc:hive2://ip-172-30-1-120.us-west-2.com> select * from Pain_Pills_In_The_USA_Per_Cities limit 10;
INFO : Compiling command(queryId=hive_20220504045512_341a60c2-f901-4e75-b070-3af9143c3456): select * from Pain_Pills_In_The_USA_Per_Cities limit 10
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:pain_pills_in_the_usa_per_cities.buyer_city, type:string, comment:null), FieldSchema(name:pain_pills_in_the_usa_per_cities.c1, type:double, comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20220504045512_341a60c2-f901-4e75-b070-3af9143c3456); Time taken: 0.812 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20220504045512_341a60c2-f901-4e75-b070-3af9143c3456): select * from Pain_Pills_In_The_USA_Per_Cities limit 10
INFO : Completed executing command(queryId=hive_20220504045512_341a60c2-f901-4e75-b070-3af9143c3456); Time taken: 0.001 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
```

pain_pills_in_the_usa_per_cities.buyer_city	pain_pills_in_the_usa_per_cities.c1
LIVERMORE	5266209.0
NORTH CHARLESTON	4750679.0
LAS VEGAS	3588555.0
TAMPA	3150503.0
HOUSTON	2783058.0
PHILADELPHIA	2677515.0
COLUMBUS	2397638.0
PHOENIX	2344390.0
PORTLAND	2311690.0
JACKSONVILLE	2300874.0

```
10 rows selected (0.834 seconds)
```

hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/PainPillsInTheCity

```
-bash-4.2$ hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/PainPillsInTheCity
Found 1 items
-rw-r--r-- 1 avalasa hdfsadmin group 187768 2022-05-04 04:42 /user/avalasa/PainPillsInTheUSA/PainPillsInTheCity/000000_0
```

hdfs dfs -cat /user/avalasa/PainPillsInTheUSA/PainPillsInTheCity/000000_0 | tail -n 5

```
-bash-4.2$ hdfs dfs -cat /user/avalasa/PainPillsInTheUSA/PainPillsInTheCity/000000_0 | tail -n 5
OSINING,1.0
BARNSTEAD,1.0
POLK,1.0
HADDAM,1.0
```

hdfs dfs -get /user/avalasa/PainPillsInTheUSA/PainPillsInTheCity/000000_0
PainPills_In_The_USA_Cities.txt

```
-bash-4.2$ hdfs dfs -get /user/avalasa/PainPillsInTheUSA/PainPillsInTheCity/000000_0 PainPills_In_The_USA_Cities.txt
-bash-4.2$ ls
arcos_all_washpost.tsv PainPills_In_The_USA_Cities.txt PainPills_In_The_USA_States.txt
-bash-4.2$
```

scp avalasa@52.12.203.116:/home/avalasa/PainPills_In_The_USA_Cities.txt .

```
anushavalasapalli@Kalyanams-MBP Documents % scp avalasa@52.12.203.116:/home/avalasa/PainPills_In_The_USA_Cities.txt .
avalasa@52.12.203.116's password:
PainPills_In_The_USA_Cities.txt
anushavalasapalli@Kalyanams-MBP Documents %
```

100% 183KB 971.0KB/s 00:00

MostPopularPainPills:

```
DROP TABLE IF EXISTS Most_Popular_Pain_Pills_In_The_USA;
CREATE TABLE Most_Popular_Pain_Pills_In_The_USA
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION '/user/avalasa/PainPillsInTheUSA/MostPopularPainPills'
AS
select DRUG_NAME, sum(QUANTITY)
from Pain_Pills_In_The_USA
where DRUG_NAME is not null group by DRUG_NAME order by 2 desc;
```

```
0: jdbc:hive2://ip-172-30-1-120.us-west-2.com> CREATE TABLE Most_Popular_Pain_Pills_In_The_USA
. . . . .> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
. . . . .> STORED AS TEXTFILE LOCATION '/user/avalasa/PainPillsInTheUSA/MostPopularPainPills'
. . . . .> AS
. . . . .> select DRUG_NAME, sum(QUANTITY) from Pain_Pills_In_The_USA where DRUG_NAME is not null group by DRUG_NAME order
by 2
. . . . .> ;
```

```
select * from Most_Popular_Pain_Pills_In_The_USA limit 10;
```

```
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20220504062107_f68ae52c-4f03-4d05-9d9e-f9a02ad88bc6): select * from Most_Popular_Pain_Pills_In_The_USA limit 10
INFO : Completed executing command(queryId=hive_20220504062107_f68ae52c-4f03-4d05-9d9e-f9a02ad88bc6); Time taken: 0.001 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
```

most_popular_pain_pills_in_the_usa.drug_name	most_popular_pain_pills_in_the_usa.c1
OXYCODONE	2.35667016E8
HYDROCODONE	2.44599198E8

2 rows selected (0.915 seconds)

```
hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/MostPopularPainPills
hdfs dfs -cat /user/avalasa/PainPillsInTheUSA/MostPopularPainPills/000000_0 | tail -n 2
hdfs dfs -get /user/avalasa/PainPillsInTheUSA/MostPopularPainPills/000000_0
PainPills_In_The_USA_Particular_Cities.txt
```

```
-bash-4.2$ hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/MostPopularPainPills
Found 1 items
-rw-r--r-- 1 avalasa hdfsadmingroup 48 2022-05-04 06:18 /user/avalasa/PainPillsInTheUSA/MostPopularPainPills/000000_0
-bash-4.2$ hdfs dfs -cat /user/avalasa/PainPillsInTheUSA/MostPopularPainPills/000000_0 | tail -n 2
OXYCODONE,2.35667016E8
HYDROCODONE,2.44599198E8
-bash-4.2$ hdfs dfs -get /user/avalasa/PainPillsInTheUSA/MostPopularPainPills/000000_0 PainPills_In_The_USA_Particular_Cities.txt
-bash-4.2$ readlink -f PainPills_In_The_USA_Particular_Cities.txt
/home/avalasa/PainPills_In_The_USA_Particular_Cities.txt
```

```
scp avalasa@52.12.203.116:/home/avalasa/PainPills_In_The_USA_Particular_Cities.txt .
```

```
anushavalasapalli@Kalyanams-MBP Documents % scp avalasa@52.12.203.116:/home/avalasa/PainPills_In_The_USA_Particular_Cities.txt .
avalasa@52.12.203.116's password:
PainPills_In_The_USA_Particular_Cities.txt
```

100% 48 1.4KB/s 00:00

Most Popular Pain Pills in FL, TX and CA

```
DROP TABLE IF EXISTS Popular_Pain_Pills_In_FL_TX_CA;
CREATE TABLE Popular_Pain_Pills_In_FL_TX_CA
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION
'/user/avalasa/PainPillsInTheUSA/MostPopularPainPillsInParticularStates'
AS
select DRUG_NAME, sum(QUANTITY)
from Pain_Pills_In_The_USA
where DRUG_NAME is not null and
(lower(BUYER_STATE) like "ca" or
lower(BUYER_STATE) like "fl"
or lower(BUYER_STATE) like "tx")
group by DRUG_NAME order by 2 desc;
```

```
0: jdbc:hive2://ip-172-30-1-120.us-west-2.com> CREATE TABLE Popular_Pain_Pills_In_FL_TX_CA
. . . . .> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
. . . . .> STORED AS TEXTFILE LOCATION '/user/avalasa/PainPillsInTheUSA/MostPopularPainPillsInParticularStates'
. . . . .> AS
. . . . .> select DRUG_NAME, sum(QUANTITY) from Pain_Pills_In_The_USA where DRUG_NAME is not null and lower(BUYER_S
TATE) like "ca" or lower(BUYER_STATE) like "fl" or lower(BUYER_STATE) like "tx" group by DRUG_NAME order by 2 desc;
INFO : Compiling command(queryId=hive_20220504062311_090c5649-d47c-4e88-b3eb-7e96d05ca088): CREATE TABLE Popular_Pain_Pills_In_FL_TX_CA
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION '/user/avalasa/PainPillsInTheUSA/MostPopularPainPillsInParticularStates'
AS
select DRUG_NAME, sum(QUANTITY) from Pain_Pills_In_The_USA where DRUG_NAME is not null and lower(BUYER_STATE) like "ca" or lower(BUYER_STATE) like "fl"
or lower(BUYER_STATE) like "tx" group by DRUG_NAME order by 2 desc
```

select * from Popular_Pain_Pills_In_FL_TX_CA limit 10;

```
INFO : Compiling command(queryId=hive_20220504063614_c323dec6-d879-4e4b-93f1-8b13f19f75e3): select * from Popular_Pain_Pills_In_FL_TX_CA limit 10
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:popular_pain_pills_in_fl_tx_ca.drug_name, type:string, comment:null), FieldSchema(n
ame:popular_pain_pills_in_fl_tx_ca._c1, type:double, comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20220504063614_c323dec6-d879-4e4b-93f1-8b13f19f75e3); Time taken: 0.893 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20220504063614_c323dec6-d879-4e4b-93f1-8b13f19f75e3): select * from Popular_Pain_Pills_In_FL_TX_CA limit 10
INFO : Completed executing command(queryId=hive_20220504063614_c323dec6-d879-4e4b-93f1-8b13f19f75e3); Time taken: 0.001 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager

+-----+-----+
| popular_pain_pills_in_fl_tx_ca.drug_name | popular_pain_pills_in_fl_tx_ca._c1 |
+-----+-----+
| HYDROCODONE                             | 7.4041178E7                         |
| OXYCODONE                               | 4.4674545E7                         |
+-----+-----+
```

```
hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/MostPopularPainPillsInParticularStates
hdfs dfs -cat /user/avalasa/PainPillsInTheUSA/MostPopularPainPillsInParticularStates/000000_0
| tail -n 2
hdfs dfs -get user/avalasa/PainPillsInTheUSA/MostPopularPainPillsInParticularStates/000000_0
PainPills_In_The_Particular_States_FL_TX_CA.txt
```

```
bash-4.2$ hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/MostPopularPainPillsInParticularStates
Found 1 items
-rw-r--r-- 1 avalasa hdfsadmin group 46 2022-05-04 06:34 /user/avalasa/PainPillsInTheUSA/MostPopularPainPillsInParticularStates/000000_0
bash-4.2$ hdfs dfs -cat /user/avalasa/PainPillsInTheUSA/MostPopularPainPillsInParticularStates/000000_0 | tail -n 2
HYDROCODONE,7.4041178E7
OXYCODONE,4.4674545E7
bash-4.2$ hdfs dfs -get /user/avalasa/PainPillsInTheUSA/MostPopularPainPillsInParticularStates/000000_0 PainPills_In_The_Particular_States_FL_TX_CA.txt
bash-4.2$
bash-4.2$ readlink -f PainPills_In_The_Particular_States_FL_TX_CA.txt
```

scp [avalasa@52.12.203.116:/home/avalasa/PainPills_In_The_Particular_States_FL_TX_CA.txt](#)

```
anushavalasapalli@Kalyanams-MBP Documents % scp avalasa@52.12.203.116:/home/avalasa/PainPills_In_The_Particular_States_FL_TX_CA.txt .
avalasa@52.12.203.116's password:
PainPills_In_The_Particular_States_FL_TX_CA.txt                                100% 46   1.4KB/s   00:00
anushavalasapalli@Kalyanams-MBP Documents % █
```

```
Last login: Tue May 3 20:33:36 on ttys000
anushavalasapalli@Kalyanams-MBP ~ % scp avalasa@52.12.203.116:/home/avalasa/PainPills_In_The_USA_States.txt .
avalasa@52.12.203.116's password:
PainPills_In_The_USA_States.txt                                100% 751   21.8KB/s   00:00
anushavalasapalli@Kalyanams-MBP ~ % ls
Applications      Library           PainPills_In_The_USA_States.txt  PycharmProjects      senti_out.xlsx
Desktop           Movies            Pictures                     env                   top10country10.csv
Documents         Music             Postman                      refersperday12.csv   ~$senti_out.xlsx
Downloads         OneDrive - Villanova University Public      senti_out.csv
anushavalasapalli@Kalyanams-MBP ~ % cd Documents
anushavalasapalli@Kalyanams-MBP Documents % scp avalasa@52.12.203.116:/home/avalasa/PainPills_In_The_USA_States.txt .
avalasa@52.12.203.116's password:
PainPills_In_The_USA_States.txt                                100% 751   20.8KB/s   00:00
anushavalasapalli@Kalyanams-MBP Documents % scp avalasa@52.12.203.116:/home/avalasa/PainPills_In_The_USA_Cities.txt .
avalasa@52.12.203.116's password:
PainPills_In_The_USA_Cities.txt                                100% 183KB 971.0KB/s   00:00
anushavalasapalli@Kalyanams-MBP Documents % scp avalasa@52.12.203.116:/home/avalasa/PainPills_In_The_USA_Particular_Cities.txt .
avalasa@52.12.203.116's password:
PainPills_In_The_USA_Particular_Cities.txt                    100% 48    1.4KB/s    00:00
anushavalasapalli@Kalyanams-MBP Documents % scp avalasa@52.12.203.116:/home/avalasa/Popular_PainPills_In_The_USA.txt .
avalasa@52.12.203.116's password:
scp: /home/avalasa/Popular_PainPills_In_The_USA.txt: No such file or directory
anushavalasapalli@Kalyanams-MBP Documents % scp avalasa@52.12.203.116:/home/avalasa/PainPills_In_The_Particular_States_FL_TX_CA.txt .
avalasa@52.12.203.116's password:
PainPills_In_The_Particular_States_FL_TX_CA.txt                100% 46    1.4KB/s    00:00
anushavalasapalli@Kalyanams-MBP Documents % █
```

Yearly Pain pills consumption in FL and CA

DROP TABLE IF EXISTS PainPills_In_Particular_States_year;

CREATE TABLE PainPills_In_Particular_States_year
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION
'/user/avalasa/PainPillsInTheUSA/PainPillsInParticularStatesbyyear'
AS

select BUYER_STATE, sum(QUANTITY) AS Drug_Count,
SUBSTRING(TRANSACTION_DATE, 5, 4) AS year
from Pain_Pills_In_The_USA
where (BUYER_STATE like "CA" or BUYER_STATE like "FL")
GROUP BY BUYER_STATE, SUBSTRING(TRANSACTION_DATE, 5, 4);

select * from PainPills_In_Particular_States_year ORDER BY BUYER_STATE DESC;

```
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20220513000411_98f4ed08-3942-45a5-b64b-be87a7c35764): select * from PainPills_In_Particular_States_year ORDER BY BUYER_STATE DESC
INFO : Query ID = hive_20220513000411_98f4ed08-3942-45a5-b64b-be87a7c35764
INFO : Total jobs = 1
INFO : Launching Job 1 out of 1
INFO : Starting task [stage-1:MAPRED] in serial mode
INFO : Subscribed to counters: {} for queryId: hive_20220513000411_98f4ed08-3942-45a5-b64b-be87a7c35764
INFO : Session is already open
INFO : Dag name: select * from PainPills_In_Particular_States_year ORDER BY BUYER_STATE DESC (Stage-1)
INFO : Status: Running (Executing on YARN cluster with App id application_1652387077726_0003)

INFO : Map 1: 0/1 Reducer 2: 0/1
INFO : Map 1: 0(-1)/1 Reducer 2: 0/1
INFO : Map 1: 1/1 Reducer 2: 0/1
INFO : Map 1: 1/1 Reducer 2: 1/1
INFO : Completed executing command(queryId=hive_20220513000411_98f4ed08-3942-45a5-b64b-be87a7c35764); Time taken: 5.289 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
```

painpills_in_particular_states_year.buyer_state	painpills_in_particular_states_year.drug_count	painpills_in_particular_states_year.year
FL	3557829.0	2006
FL	3734224.0	2007
FL	6280861.0	2010
FL	5196573.0	2009
FL	4338617.0	2008
FL	4099095.0	2012
FL	5501231.0	2011
CA	5457999.0	2007
CA	6740300.0	2010
CA	7797757.0	2011
CA	8129629.0	2012
CA	5837527.0	2008
CA	6098397.0	2009
CA	5035609.0	2006

```
14 rows selected (6.196 seconds)
0: jdbc:hive2://ip-172-30-1-132.us-west-2.com>
```

hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/PainPillsInParticularStatesbyyear
hdfs dfs -cat /user/avalasa/PainPillsInTheUSA/PainPillsInParticularStatesbyyear/000004_0 | tail
-n 2

```
MINGW64/c/Users/mauli
https://aws.amazon.com/amazon-linux-2/
49 package(s) needed for security, out of 88 available
Run "sudo yum update" to apply all updates.

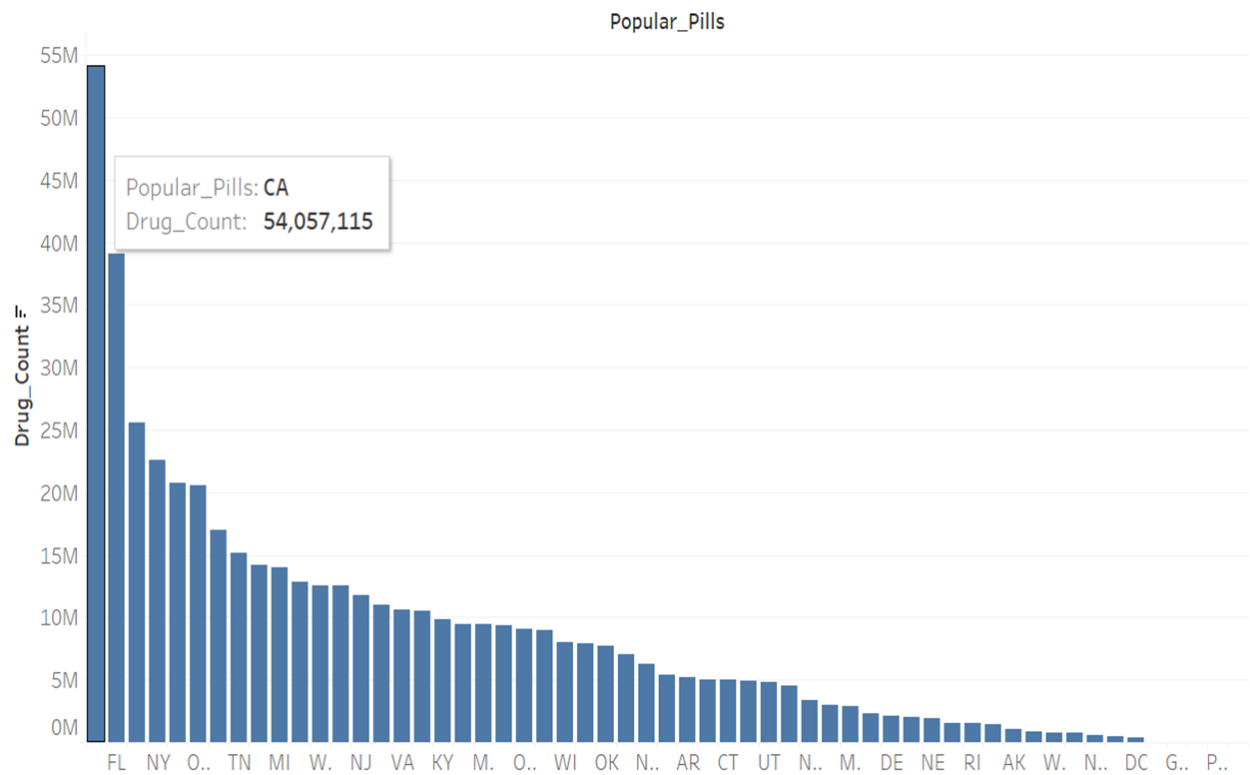
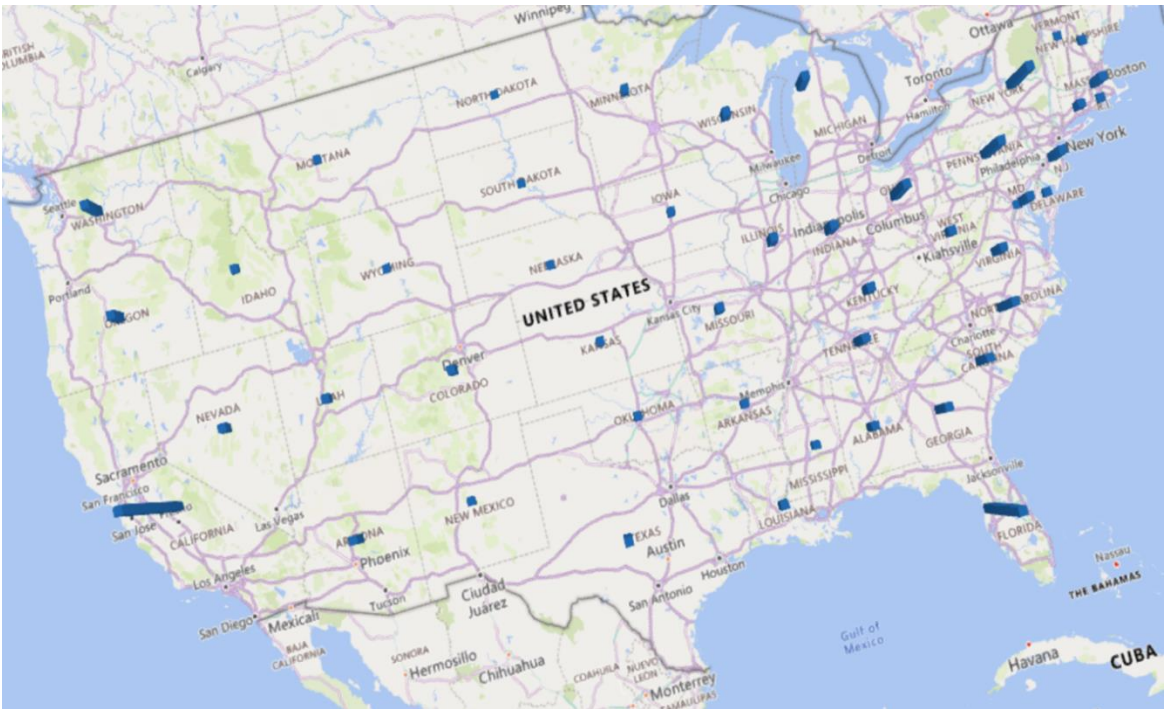
EEEEEEEEEEEEEEEEEEEE MMMMMMM RRRRRRRRRRRRRRR
E:::EEEEEEEEEEEEEEEE M:::MM M:::MM R:::RRRRRRRRRR
EE:::EEEEEEEEEEEEEEEE M:::MM M:::MM R:::RRRRRRRRRR
E:::E EEEEE M:::MM M:::MM RRR::R R:::R
E:::E EEEEE M:::MM M:::MM R:::R R:::R
E:::EEEEEEEEEEEE M:::MM M:::MM M:::MM R:::RRRRRRRRRR
E:::EEEEEEEEEEEE M:::MM M:::MM M:::MM R:::RRRRRRRRRR
E:::EEEEEEEEEEEE M:::MM M:::MM M:::MM R:::RRRRRRRRRR
E:::E EEEEE M:::MM M:::MM M:::MM R:::R R:::R
E:::E EEEEE M:::MM M:::MM M:::MM R:::R R:::R
EE:::EEEEEEEEEEEE M:::MM M:::MM M:::MM R:::R R:::R
E:::EEEEEEEEEEEE M:::MM M:::MM M:::MM R:::R R:::R
EEEEEEEEEEEEEEEEEEEE MMMMMMM RRRRRRRRRRRRRRR

-bash-4.2$ hdfs dfs -ls /user/avalasa/PainPillsInTheUSA/
Found 5 items
drwxr-xr-x - avalasa hdfsadmin 0 2022-05-12 16:49 /user/avalasa
/PainPillsInTheUSA/MostPopularPainPills
drwxr-xr-x - avalasa hdfsadmin 0 2022-05-12 23:59 /user/avalasa
/PainPillsInTheUSA/PainPillsInParticularStatesbyyear
drwxr-xr-x - avalasa hdfsadmin 0 2022-05-12 16:49 /user/avalasa
/PainPillsInTheUSA/PainPillsInTheCity
drwxr-xr-x - avalasa hdfsadmin 0 2022-05-12 16:48 /user/avalasa
/PainPillsInTheUSA/PainPillsInTheState
-rw-r--r-- 1 avalasa hdfsadmin 66343190528 2022-05-12 17:19 /user/avalasa
/PainPillsInTheUSA/arcs_all_washpost.tsv
-bash-4.2$ hdfs dfs -cat /user/avalasa/PainPillsInTheUSA/PainPillsInParticularStatesbyyear/000004_0 | tail -n 2
CA,5457999.0,2007
```

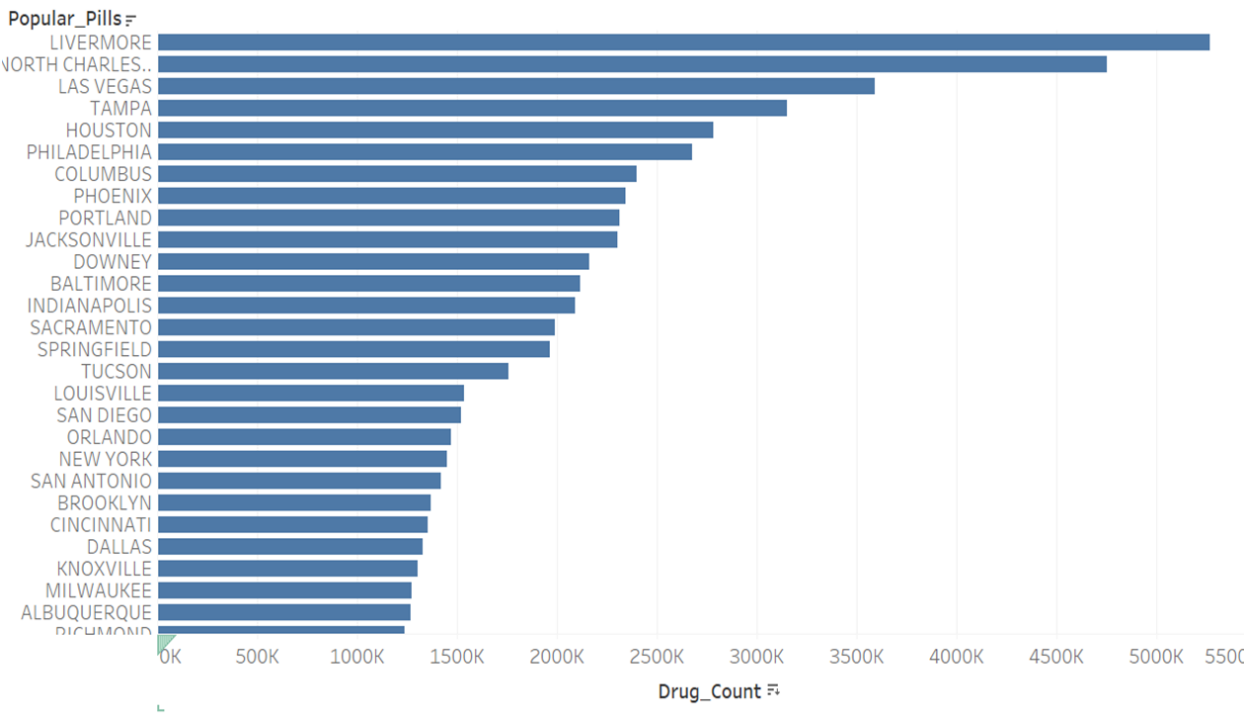
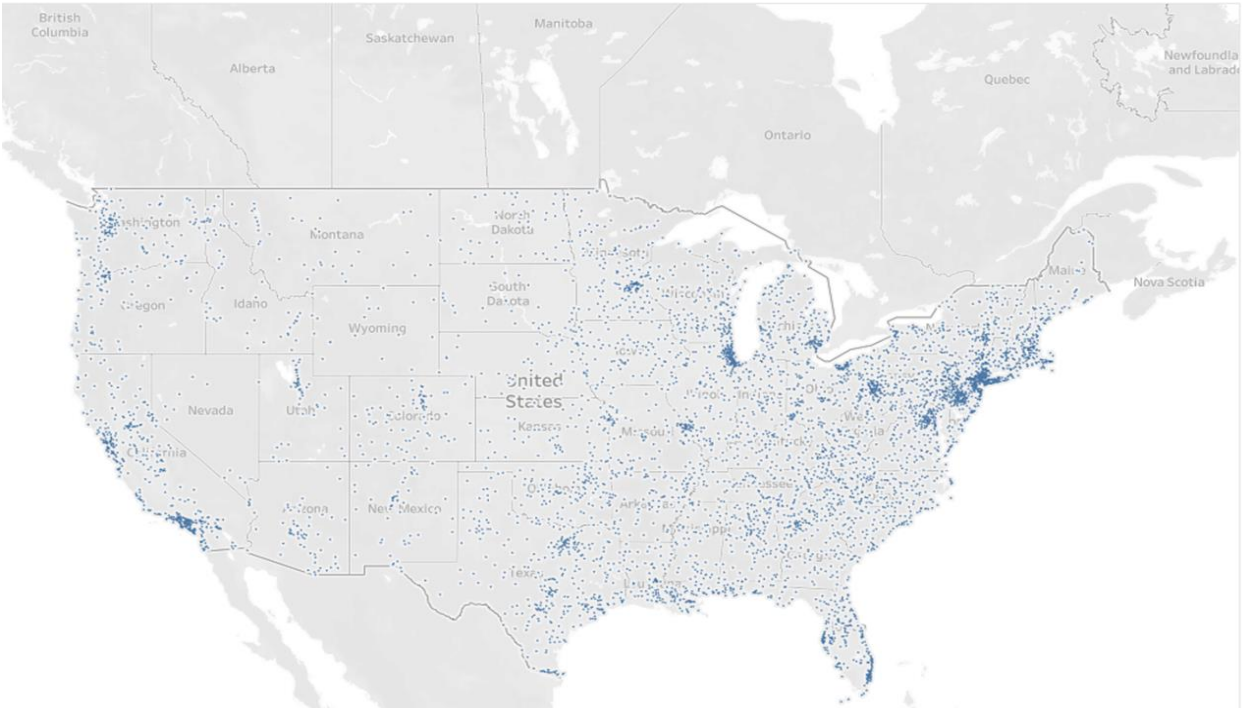
hdfs dfs -get /user/avalasa/PainPillsInTheUSA/PainPillsInParticularStatesbyyear/000004_0
PainPillsInParticularStatesbyyear.txt
readlink -f PainPills_In_Particular_States_year.txt// to know linux system path
scp avalasa@35.87.210.247:/home/avalasa/PainPillsInParticularStatesbyyear.txt .

Visualization

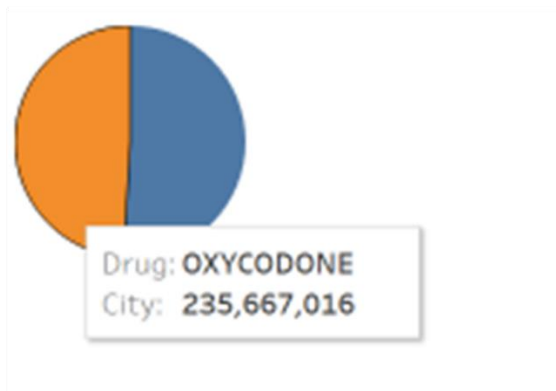
State Pain Pills Count:-



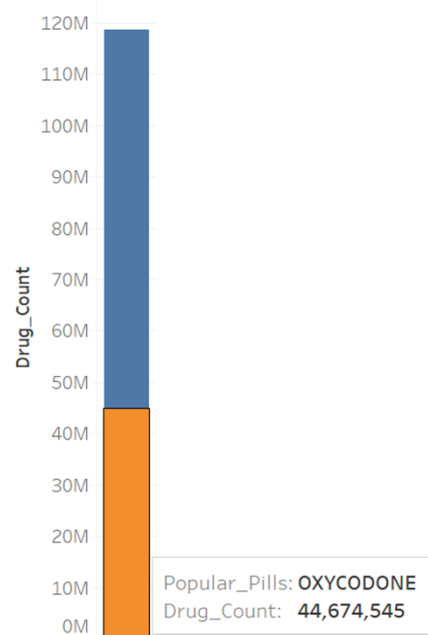
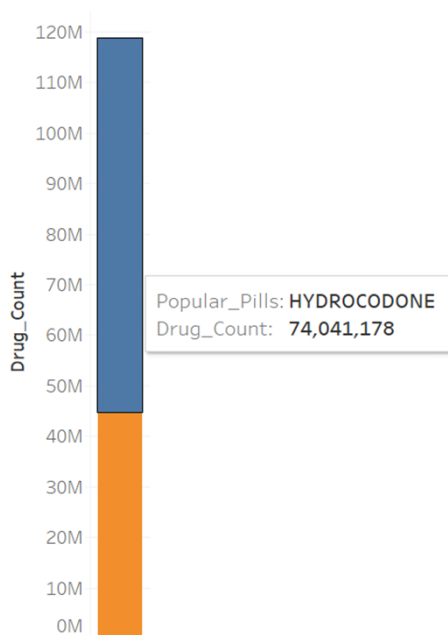
City Pain Pills Count:-



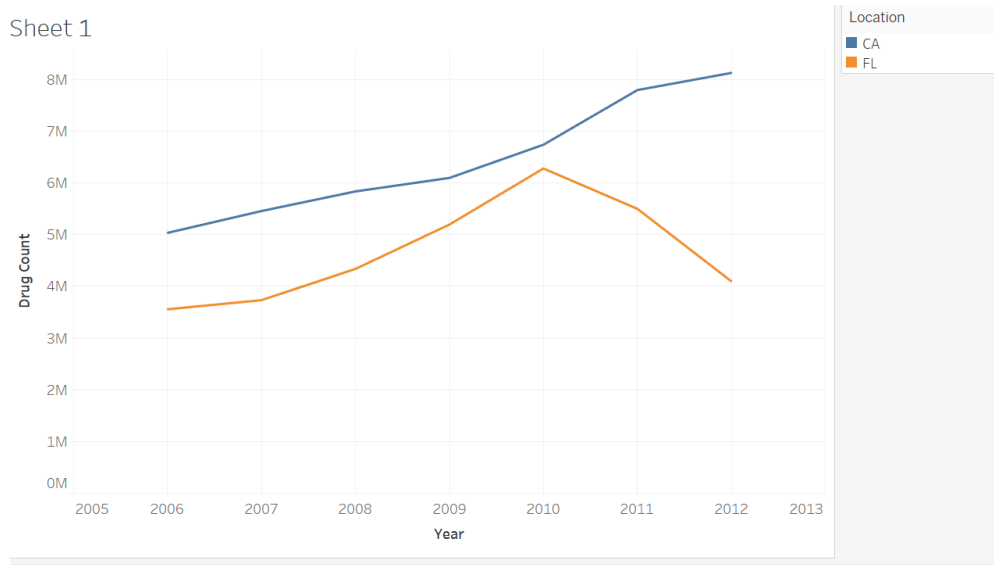
Most Popular Pain Pills:



Most Popular Pain Pills in FL, TX, and CA



Total Pain Pills Sold Per Day Per States (California and Florida)



References:

URL of Data Source

<http://www.calstatela.edu>

URL of Github

<https://github.com/ruta2709/Pain-Pills-in-USA.git>

URL of References

1. <https://www.kaggle.com/datasets/paultimothymooney/pain-pills-in-the-usa>
2. <https://www.kaggle.com/code/paultimothymooney/pain-pills-in-the-usa-dask-plotly>

