

HEALTH INSURANCE MARKETPLACE DATA ANALYSIS USING HADOOP & HIVE (Oracle Cloud)

A) HADOOP

Step1: Download the source file from CMS website (Center for Medical Services) where it has a dataset of all the insurance plan for throughout years.

For our Analysis, we took 2017 to 2021 dataset to compare and identify which insurance plan would be better for individual/family/employee.

The files to download are as follows:

1. Benefits and Cost Sharing PUF
2. Rate PUF
3. Plan Attributes PUF
4. Business Rules PUF
5. Service Area PUF
6. Network
7. Machine-readable URL PUF
8. Transparency in Coverage PUF

<https://www.cms.gov/ccio/resources/data-resources/marketplace-puf>

Step2: Upload the files in linux system:

`scp "filepath\file.csv" username@serveradd:/tmp`

```
AD+rbhoga1@STU-PF2XYGFC MINGW64 ~  
$ scp "C:\Users\rbhoga1\Desktop\5200 combined files\project_final\machine_readable_url.csv" rbhoga1@144.24.14.145:/tmp  
rbhoga1@144.24.14.145's password:  
C:\Users\rbhoga1\Desktop\5200 combined files\project_final\machine_readable_url.csv 100% 206KB 126.8KB/s
```

Step2: Transfer the file from linux to hdfs

Log in to your hdfs server:

Ssh username@serveradd

username@serveradd password:

```
AD+rbhoga1@STU-PF2XYGFC MINGW64 ~  
$ ssh rbhoga1@144.24.14.145  
rbhoga1@144.24.14.145's password:  
Last login: Wed Dec 7 10:33:52 2022 from 219.sub-174-193-194.myvzw.com  
~-bash-4.2$ |
```

`cd /tmp`

HEALTH INSURANCE MARKETPLACE DATA ANALYSIS USING HADOOP & HIVE (Oracle Cloud)

ls /tmp (to view your uploaded files)

```
bash-4.2$ cd
bash-4.2$ pwd
home/rbhoga1
bash-4.2$ cd /tmp
bash-4.2$ ls /tmp
```

```
machine_readable_url.csv
output
pym-p0a47ex
pym-BIVM8m
pym-cIec9J
pym-NLojTz
pym-qCZG1h
pym-Xzfkeo
report.json
snappy-1.0.5-5fd1aeda-b749-4081-b289-22c117049a00-libsnappyjava.so
systemd-private-3207b8db87a24ac9b7ceb8bf19353586-chronyd.service-w5XMXn
systemd-private-3207b8db87a24ac9b7ceb8bf19353586-cups.service-rmMLWK
systemd-private-3207b8db87a24ac9b7ceb8bf19353586-unified-monitoring-agent.s
```

Make directory in hdfs

hdfs dfs -mkdir Project/ ---(Project is main folder name)

hdfs dfs -mkdir Project/file --(file is subfolder)

hdfs dfs -put file.csv Project/file/

To view folders and files:

Hdfs dfs -ls

```
-bash-4.2$ hdfs dfs -ls
Found 3 items
drwx----- - rbhoga1 hdfs      0 2022-12-07 11:02 .Trash
drwxr-xr-x - rbhoga1 hdfs      0 2022-11-10 02:35 .hiveJars
drwxr-xr-x - rbhoga1 hdfs      0 2022-11-29 21:02 Health
```

Hdfs dfs -ls Health/

```
-bash-4.2$ hdfs dfs -ls Health/
Found 8 items
drwxr-xr-x - rbhoga1 hdfs      0 2022-11-29 22:53 Health/benefits_cost_sharing
drwxr-xr-x - rbhoga1 hdfs      0 2022-11-29 22:56 Health/business_rules
drwxr-xr-x - rbhoga1 hdfs      0 2022-11-29 22:52 Health/machine_readable_url
drwxr-xr-x - rbhoga1 hdfs      0 2022-11-29 22:11 Health/network
drwxr-xr-x - rbhoga1 hdfs      0 2022-11-29 22:18 Health/plan_attributes
drwxr-xr-x - rbhoga1 hdfs      0 2022-11-29 22:32 Health/rate
drwxr-xr-x - rbhoga1 hdfs      0 2022-11-29 22:19 Health/service_area
drwxr-xr-x - rbhoga1 hdfs      0 2022-11-30 00:18 Health/transparency_in_coverage
```

Step3: Once the files are copied from linux to hdfs, remove the files in linux :

rm -file.csv

HEALTH INSURANCE MARKETPLACE DATA ANALYSIS USING HADOOP & HIVE (Oracle Cloud)

```
-bash-4.2$ rm machine_readable_url.csv
-bash-4.2$ ls /tmp
03d7e569-d376-4a64-b9c4-bbb8cca6cd35_resources
2417d95f-6634-4b31-87ca-e0ccfcdcd158_resources
2ea2bbcf-333c-45a5-a22a-04c6b9316ebf_resources
4bf07050-93e4-4012-a977-50b233a28eb6_resources
4bf2b422-c146-4140-92dd-377cb5e7a860_resources
76c1f38e-0efe-4571-99a4-f7d1d3460267_resources
```

Step4: Connect to Hive using command:

beeline;

```
-bash-4.2$ beeline;
SLF4J: class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/odh/1.1.2/hive/lib/log4j-slf4j-impl-2.17.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/odh/1.1.2/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Connecting to jdbc:hive2://bigdaiwn0.sub02180640120.trainingvcn.oraclevcn.com:2181,bigdaiwn0.sub02180640120.trainingvcn.oraclevcn.com:2181,bigdaiwn0.sub02180640120.trainingvcn.oraclevcn.com:2181/default;password=rbhogal;serviceDiscoveryMode=zooKeeper;user=rbhogal;zooKeeperNamespace=hiveserver2
22/12/07 21:15:41 [main-EventThread]: ERROR impls.EnsembleTracker: Invalid config event received: {server.1=bigdaiwn0.sub02180640120.trainingvcn.oraclevcn.com:2838:3888:participant, version=0, server.3=bigdaiwn0.sub02180640120.trainingvcn.oraclevcn.com:2888:3888:participant, server.2=bigdaiwn0.sub02180640120.trainingvcn
```

B) HIVE

Step5: use your database

use username;

```
0: jdbc:hive2://bigdaiwn0.sub02180640120.traib> use rbhogal;
INFO : Compiling command(queryId=hive_20221207211901_1c895217-ebe8-4049-86d3-134e6dcf608b): use rbhogal
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_20221207211901_1c895217-ebe8-4049-86d3-134e6dcf608b); Time taken: 0.025 s
econds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20221207211901_1c895217-ebe8-4049-86d3-134e6dcf608b): use rbhogal
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20221207211901_1c895217-ebe8-4049-86d3-134e6dcf608b); Time taken: 0.214 s
econds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
No rows affected (0.291 seconds)
0: jdbc:hive2://bigdaiwn0.sub02180640120.traib> |
```

Step6: See the existing tables

show tables;

```
0: jdbc:hive2://bigdaiwn0.sub02180640120.traib> show tables;
INFO : Compiling command(queryId=hive_20221207211935_a5bf16a0-da8e-4d66-8a86-75eee62e2641): show tables
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:tab_name, type:string, comment:from deserial),], properties:null)
INFO : Completed compiling command(queryId=hive_20221207211935_a5bf16a0-da8e-4d66-8a86-75eee62e2641); Time taken: 0.025 s
econds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20221207211935_a5bf16a0-da8e-4d66-8a86-75eee62e2641): show tables
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20221207211935_a5bf16a0-da8e-4d66-8a86-75eee62e2641); Time taken: 0.291 s
econds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager

+-----+
| tab_name |
+-----+
| a        |
| b1       |
| bb       |
| benefits |
| benefits_cost_sharing |
| br       |
+-----+
```

HEALTH INSURANCE MARKETPLACE DATA ANALYSIS USING HADOOP & HIVE (Oracle Cloud)

Step7: Create new tables for all the csv files uploaded in hdfs.

Note: Since the file path is hdfs, we need to retain the Project folder as data gets populated from the hdfs to hive tables.

CREATE EXTERNAL TABLE if not exists table_name

(col1 datatype,

col2 datatype,

col3 datatype,

.

.

.

colN datatype

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE LOCATION '/user/username/Project/file/' --file path of hdfs folder

TBLPROPERTIES ('skip.header.line.count'='1');

```
0: jdbc:hive2://bigdaiwn0.sub02180640120.tra1> CREATE EXTERNAL TABLE if not exists machine_readable_url12
...> (year int,
...> State string,
...> Issuer_ID int,
...> URL_Submitted string,
...> Tech_POC_Email string
...> )
...> ROW FORMAT DELIMITED
...> FIELDS TERMINATED BY ','
...> STORED AS TEXTFILE LOCATION '/user/rbhogal/Health/machine_readable_url1/'
...> TBLPROPERTIES ('skip.header.line.count'='1');
INFO : Compiling command(queryId=hive_20221207212119_ccee1e8a-16d1-4a8c-87f2-e1efd94f5b4b): CREATE EXTERNAL TABLE if not exists machine_readable_url12
(year int,
State string,
Issuer_ID int,
URL_Submitted string,
Tech_POC_Email string
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION '/user/rbhogal/Health/machine_readable_url1/'
TBLPROPERTIES ('skip.header.line.count'='1')
INFO : Concurency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldsSchemas:null, properties:null)
INFO : Completed compiling command(queryId=hive_20221207212119_ccee1e8a-16d1-4a8c-87f2-e1efd94f5b4b); Time taken: 0.03 seconds
INFO : Concurency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20221207212119_ccee1e8a-16d1-4a8c-87f2-e1efd94f5b4b): CREATE EXTERNAL TABLE if not exists machine_readable_url12
(year int,
State string,
Issuer_ID int,
URL_Submitted string,
Tech_POC_Email string
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION '/user/rbhogal/Health/machine_readable_url1/'
TBLPROPERTIES ('skip.header.line.count'='1')
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20221207212119_ccee1e8a-16d1-4a8c-87f2-e1efd94f5b4b); Time taken: 0.145 seconds
INFO : OK
INFO : Concurency mode is disabled, not creating a lock manager
No rows affected (0.187 seconds)
```

See the table contents

HEALTH INSURANCE MARKETPLACE DATA ANALYSIS USING HADOOP & HIVE (Oracle Cloud)

Select * from table_name limit 10; Select * from machine_readable_url limit 10;

```
INFO : Compiling command(queryId=hive_20221207212311_sceb272-4541-41f6-9c90-2ce5772ec26b): select * from machine_readable_url12 limit 10
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retry = false)
INFO : Returning Hive schema: Schema(fieldsSchemas:[FieldsSchema(name:machine_readable_url12.year, type:int, comment:null), FieldsSchema(name:machine_readable_url12.state, type:string, comment:null), FieldsSchema(name:machine_readable_url12.tech_poc_email, type:string, comment:null), FieldsSchema(name:machine_readable_url12.issuer_id, type:int, comment:null), FieldsSchema(name:machine_readable_url12.url_submitted, type:string, comment:null), FieldsSchema(name:machine_readable_url12.url_submitted, type:string, comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20221207212311_sceb272-4541-41f6-9c90-2ce5772ec26b); Time taken: 0.316 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20221207212311_sceb272-4541-41f6-9c90-2ce5772ec26b): select * from machine_readable_url12 limit 10
INFO : Completed executing command(queryId=hive_20221207212311_sceb272-4541-41f6-9c90-2ce5772ec26b); Time taken: 0.0 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
```

machine_readable_url12.tech_poc_email	machine_readable_url12.year	machine_readable_url12.state	machine_readable_url12.issuer_id	machine_readable_url12.url_submitted	machine_readable_url12.url_submitted
al.th.com	2017	AK	21989	https://www.modahealth.com/cms-data-index.json	jessica.wagner@modahealth.com
son jim.tedford@premera.com	2017	AK	38344	https://fm.formularynavigator.com/jsonFiles/publish/11/47/cms-data-index.json	
al.th.com	2017	AK	73836	https://www.modahealth.com/cms-data-index.json	jessica.wagner@modahealth.com
al.th.com	2017	AK	74819	http://www.bestlife.com/exchange/cms-data-index.json	dpitman@bestlife.com
m	2017	AL	12538	http://www.bestlife.com/exchange/cms-data-index.json	dpitman@bestlife.com
m	2017	AL	44580	https://api.humana.com/v1/cms/index.json	pboulet@humana.com
stions@bcbsal.org	2017	AL	46944	https://www.bcbsal.org/cms/cms-data-index.json	HIMMachineReadableQueue
al.com	2017	AL	82285	https://www.deltadental.com/CMSDirectory/index.json	bhenderson@deltadental.com
m	2017	AR	26904	http://www.bestlife.com/exchange/cms-data-index.json	dpitman@bestlife.com
al.com	2017	AR	28348	https://www.deltadental.com/CMSDirectory/index.json	bhenderson@deltadental.com

Step8: Once all the tables are created, we need to create summary tables

- 1) using joins
- 2) using group by clause
- 3) conditions (where clause, having clause ,case when or nested if-else statements, regex_expression)

ex:

create table ques123 as

select * from b1 left join f12

on (b1.BusinessYear = f12.year and b1.StateCode =f12.state and b1.issuerid =f12.id);

```
D:\jdbc:hive2://bigdata01.sub02180640120.tra> create table ques123 as
select *
from b1 left join f12
on (b1.BusinessYear = f12.year and b1.StateCode =f12.state and b1.issuerid =f12.id);
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retry = false)
INFO : Returning Hive schema: Schema(fieldsSchemas:[FieldsSchema(name:b1.businessyear, type:int, comment:null), FieldsSchema(name:b1.statecode, type:string, comment:null), FieldsSchema(name:b1.issuerid, type:int, comment:null), FieldsSchema(name:b1.benefitname, type:string, comment:null), FieldsSchema(name:b1.isehb, type:string, comment:null), FieldsSchema(name:b1.exclusions, type:string, comment:null), FieldsSchema(name:f12.year, type:int, comment:null), FieldsSchema(name:f12.state, type:string, comment:null), FieldsSchema(name:f12.tobacco, type:string, comment:null), FieldsSchema(name:f12.id, type:int, comment:null), FieldsSchema(name:f12.age_new, type:string, comment:null), FieldsSchema(name:f12.individualrate, type:decimal(38,18), comment:null), FieldsSchema(name:f12.individual_and_dependents, type:decimal(38,18), comment:null), FieldsSchema(name:f12.couple, type:decimal(38,18), comment:null), FieldsSchema(name:f12.individual_and_dependents, type:decimal(38,18), comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20221207212519_efdbf442-741c-4bd1-b41f-b656c519a371); Time taken: 1.257 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20221207212519_efdbf442-741c-4bd1-b41f-b656c519a371): create table ques123 as
select *
from b1 left join f12
on (b1.BusinessYear = f12.year and b1.StateCode =f12.state and b1.issuerid =f12.id)
INFO : Query ID = hive_20221207212519_efdbf442-741c-4bd1-b41f-b656c519a371
INFO : Total jobs = 1
INFO : Launching Job 1 out of 1
```

Step9: View the tables to check whether we are getting proper values or not.

select * from ques123 limit 10;

HEALTH INSURANCE MARKETPLACE DATA ANALYSIS USING HADOOP & HIVE (Oracle Cloud)

```
0: jdbc:hive2://bigdaiwn0.sub02180640120.tra1> select * from ques123 limit 10;
INFO : Compiling command(queryId=hive_20221207212753_667f97ba-084f-4ce6-aac3-2d635579cf57): select * from ques123 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:ques123.businessyear, type:int, comment:null), FieldSchema(name:ques123.statecode, type:string, comment:null), FieldSchema(name:ques123.issuerid, type:int, comment:null), FieldSchema(name:ques123.benefitname, type:string, comment:null), FieldSchema(name:ques123.iseshb, type:string, comment:null), FieldSchema(name:ques123.iscovered, type:string, comment:null), FieldSchema(name:ques123.exclusions, type:string, comment:null), FieldSchema(name:ques123.year, type:int, comment:null), FieldSchema(name:ques123.state, type:string, comment:null), FieldSchema(name:ques123.tobacco, type:string, comment:null), FieldSchema(name:ques123.id, type:int, comment:null), FieldSchema(name:ques123.age_new, type:string, comment:null), FieldSchema(name:ques123.individualrate, type:decimal(38,18), comment:null), FieldSchema(name:ques123.individualtobaccorate, type:decimal(38,18), comment:null), FieldSchema(name:ques123.couple, type:decimal(38,18), comment:null), FieldSchema(name:ques123.individual_and_dependents, type:decimal(38,18), comment:null)], properties:[]))
INFO : Completed compiling command(queryId=hive_20221207212753_667f97ba-084f-4ce6-aac3-2d635579cf57); Time taken: 0.423 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20221207212753_667f97ba-084f-4ce6-aac3-2d635579cf57): select * from ques123 limit 10
INFO : Completed executing command(queryId=hive_20221207212753_667f97ba-084f-4ce6-aac3-2d635579cf57); Time taken: 0.001 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
```

ques123.businessyear	ques123.statecode	ques123.issuerid	ques123.benefitname	ques123.iseshb	ques123.iscovered	ques123.exclusions	ques123.year
2017	AK	38344	Dental Check-Up for Children 65 and over	NULL	Covered	-1	2017
2017	AK	38344	Dental Check-Up for Children 15-44	683.7883333333333333	Covered	-1	2017
2017	AK	38344	Dental Check-Up for Children 15-44	744.837837837837838	Covered	-1	2017
2017	AK	38344	Dental Check-Up for Children 65 and over	NULL	Covered	-1	2017
2017	AK	38344	Dental Check-Up for Children 45-64	895.9696969696969697	Covered	-1	2017
2017	AK	38344	Dental Check-Up for Children 45-64	900.224489795918367347	Covered	-1	2017
2017	AK	38344	Eye Glasses for Children	Yes	Covered	-1	2017

Step10: To download the summary files from hive into the hdfs system,

```
INSERT OVERWRITE DIRECTORY '/user/username/output/' ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' SELECT * from ques123;
```

(Note: in your hdfs, folder of output would be created with summary table ques123 data)

```
0: jdbc:hive2://bigdaiwn0.sub02180640120.tra1> INSERT OVERWRITE DIRECTORY '/user/rbhogal/output/' ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' SELECT * from ques123
INFO : Compiling command(queryId=hive_20221207212944_fb5175dd-1cf5-490d-837b-14c27db51762): INSERT OVERWRITE DIRECTORY '/user/rbhogal/output/' ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' SELECT * from ques123
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldsSchemas:[FieldSchema(name:ques123.businessyear, type:int, comment:null), FieldSchema(name:ques123.statecode, type:string, comment:null), FieldSchema(name:ques123.issuerid, type:int, comment:null), FieldSchema(name:ques123.benefitname, type:string, comment:null), FieldSchema(name:ques123.iseshb, type:string, comment:null), FieldSchema(name:ques123.iscovered, type:string, comment:null), FieldSchema(name:ques123.exclusions, type:string, comment:null), FieldSchema(name:ques123.year, type:int, comment:null), FieldSchema(name:ques123.state, type:string, comment:null), FieldSchema(name:ques123.tobacco, type:string, comment:null), FieldSchema(name:ques123.id, type:int, comment:null), FieldSchema(name:ques123.age_new, type:string, comment:null), FieldSchema(name:ques123.individualrate, type:decimal(38,18), comment:null), FieldSchema(name:ques123.individualtobaccorate, type:decimal(38,18), comment:null), FieldSchema(name:ques123.couple, type:decimal(38,18), comment:null), FieldSchema(name:ques123.individual_and_dependents, type:decimal(38,18), comment:null)], properties:[]))
```

Step11: Once the file is stored in hdfs, download the file from hdfs to linux system.

```
hdfs dfs -get output/file
```

```
ex hdfs dfs -get output/000000_0
```

```
Found 4 items
drwx----- - rbhogal hdfs 0 2022-12-07 11:02 .Trash
drwxr-xr-x - rbhogal hdfs 0 2022-11-10 02:35 .hiveJars
drwxr-xr-x - rbhogal hdfs 0 2022-11-29 21:02 Health
drwxr-xr-x - rbhogal hdfs 0 2022-12-07 21:30 output
```

```
-bash-4.2$ hdfs dfs -get output/000000_0
-bash-4.2$ ls
000000_0 br_rate.csv driver_data.zip _MACOSX ques1.csv ques41.csv ratings_2013.txt top10country.csv
br_rate1.csv driver_data driver_event.txt products.tsv ques2.csv ratings_2012.txt tmp tweets_alphago.csv
```

Step12: Download the file from linux to our local pc

```
scp username@serveradd:/home/username/000000_0 filename.csv
```

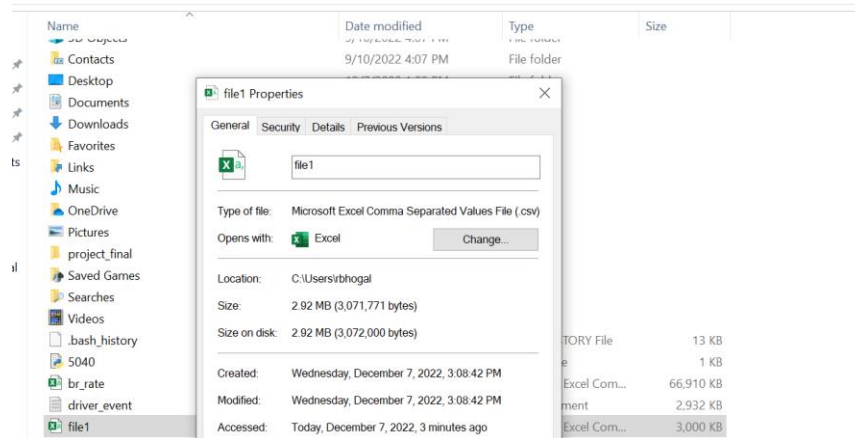
000000_0 will be saved as filename.csv or file1.csv as shown below in our C: Drive > Users > file1.csv

HEALTH INSURANCE MARKETPLACE DATA ANALYSIS USING HADOOP & HIVE (Oracle Cloud)

```
AD+rbhoga1@STU-PF2XYGFC MINGW64 ~
$ scp rbhoga1@144.24.14.145:/home/rbhoga1/000000_0 file1.csv
rbhoga1@144.24.14.145's password:
000000_0
```

100% 3000KB

This PC > Windows (C:) > Users > rbhoga1



Step13: Once the file is downloaded, upload the file in any visualization tool(Excel/Powerbi/Tableau)

(Note: Add the column name in the top row.)

Analyze the data, through different filters, sorting and various visualization charts to solve your defined problem statement

AutoSave file1 Bhogale, Rahul Sarvottam

File Home Insert Page Layout Formulas Data Review View Help Power Pivot

Clipboard Font Alignment Number Conditional Formatting Styles Cells Editing Analysis Sensitivity

POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.

H13 2017

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1																			
2		2017 AK		38344 Inpatient	60.00% CcYes		-1	2017 AK	No Prefer	38344 15-44	683.7883		0	0	0	0			
3		2017 AK		38344 Inpatient	60.00% CcYes		-1	2017 AK	No Prefer	38344 65 and ovi	\N		0	0	0	0			
4		2017 AK		38344 Inpatient	60.00% CcYes		-1	2017 AK	No Prefer	38344 45-64	900.2245		0	0	0	0			
5		2017 AK		38344 ChemotheYes	Covered		-1	2017 AK	Tobacco U	38344 65 and ovi	\N		0	0	0	0			
6		2017 AK		38344 ChemotheYes	Covered		-1	2017 AK	No Prefer	38344 15-44	683.7883		0	0	0	0			
7		2017 AK		38344 ChemotheYes	Covered		-1	2017 AK	Tobacco U	38344 15-44	744.8378	777.8295		0	0	0	0		
8		2017 AK		38344 ChemotheYes	Covered		-1	2017 AK	No Prefer	38344 65 and ovi	\N		0	0	0	0			
9		2017 AK		38344 ChemotheYes	Covered		-1	2017 AK	Tobacco U	38344 45-64	895.9697	924.8182		0	0	0	0		
10		2017 AK		38344 ChemotheYes	Covered		-1	2017 AK	No Prefer	38344 45-64	900.2245		0	0	0	0			
11		2017 AK		38344 Infertility	-1 Not Cover		-1	2017 AK	Tobacco U	38344 65 and ovi	\N		0	0	0	0			

AutoSave file1 - Saved Bhogale, Rahul Sarvottam

File Home Insert Page Layout Formulas Data Review View Help Power Pivot

Get Data From Text/CSV Recent Sources From Web Existing Connections From Table/Range

Queries & Connections Refresh All Properties Edit Links

Stocks Currencies

Sort Filter

Text to Columns Advanced

Data Tools What-If Analysis Forecast Sheet Outline

POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.

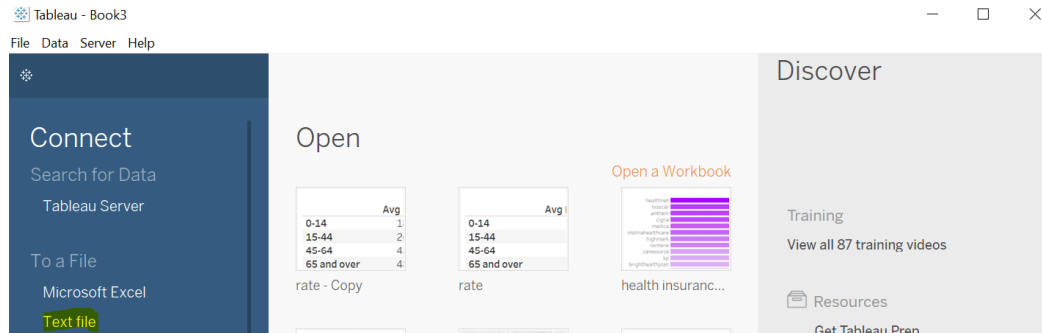
O10 0

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	year	state	id	benefit	isheh	iscoveri	exclusiv	year	state	Tobacco	id	age gro	ind	ind n to	couple	ind n d	couple	lep	
2		2017 AK		38344 Inpatient	60.00% CcYes		-1	2017 AK	No Prefer	38344 15-44	683.7883		0	0	0	0			
3		2017 AK		38344 Inpatient	60.00% CcYes		-1	2017 AK	No Prefer	38344 65 and ovi	\N		0	0	0	0			
4		2017 AK		38344 Inpatient	60.00% CcYes		-1	2017 AK	No Prefer	38344 45-64	900.2245		0	0	0	0			
5		2017 AK		38344 ChemotheYes	Covered		-1	2017 AK	Tobacco U	38344 65 and ovi	\N		0	0	0	0			
6		2017 AK		38344 ChemotheYes	Covered		-1	2017 AK	No Prefer	38344 15-44	683.7883		0	0	0	0			
7		2017 AK		38344 ChemotheYes	Covered		-1	2017 AK	Tobacco U	38344 15-44	744.8378	777.8295		0	0	0	0		
8		2017 AK		38344 ChemotheYes	Covered		-1	2017 AK	No Prefer	38344 65 and ovi	\N		0	0	0	0			
9		2017 AK		38344 ChemotheYes	Covered		-1	2017 AK	Tobacco U	38344 45-64	895.9697	924.8182		0	0	0	0		
10		2017 AK		38344 ChemotheYes	Covered		-1	2017 AK	No Prefer	38344 45-64	900.2245		0	0	0	0			
11		2017 AK		38344 Infertility	-1 Not Cover		-1	2017 AK	Tobacco U	38344 65 and ovi	\N		0	0	0	0			
12		2017 AK		38344 Infertility	-1 Not Cover		-1	2017 AK	No Prefer	38344 15-44	683.7883		0	0	0	0			

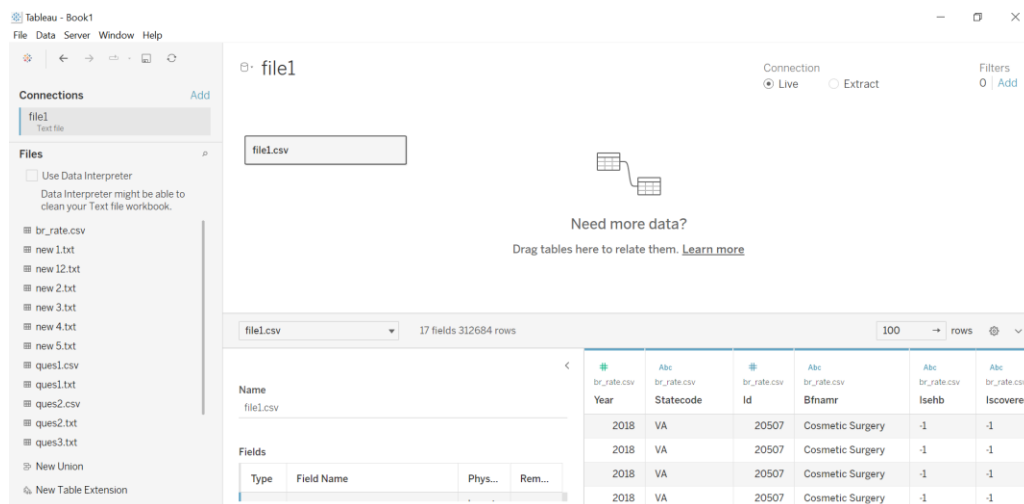
HEALTH INSURANCE MARKETPLACE DATA ANALYSIS USING HADOOP & HIVE (Oracle Cloud)

Prepare the csv file for Tableau visualization.

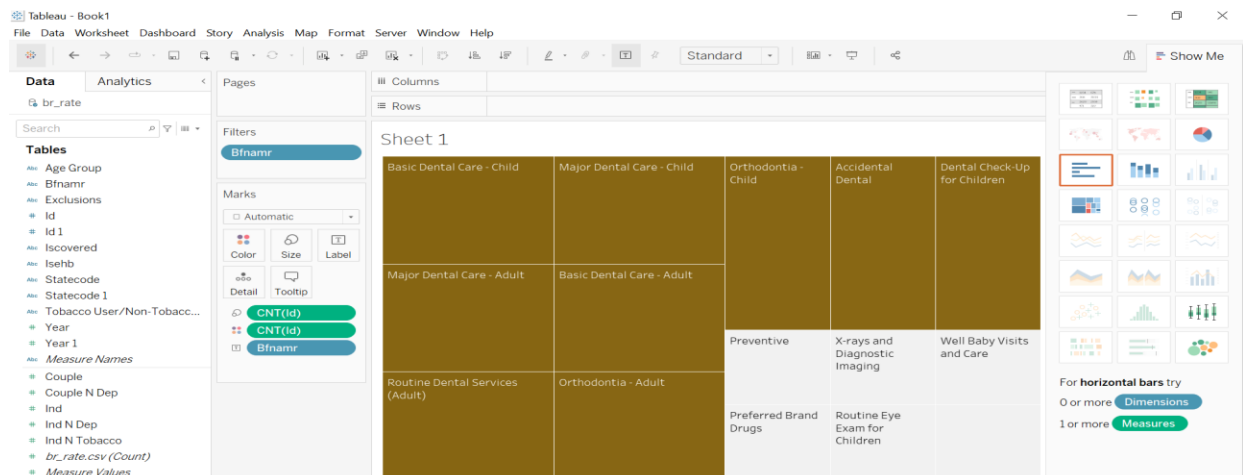
Upload the file through the data source connection



Once the connection is established, the file uploaded would look like.



Click on Sheet1, and build the dashboard similarly to this one where, bfnamr filtered as top 15 count using HeatMap/TreeMap Chart.



HEALTH INSURANCE MARKETPLACE DATA ANALYSIS USING HADOOP & HIVE (Oracle Cloud)

Similarly, we can build a geospatial visualization as well.

