2.2.5 Data Visualization

2.2.5 Visualization of Dataset

Credit Card System

Req-2.2.5

Data exploration and visualization

Functional Requirements Use Hive Query and Hive Visualization tool. 1) Find the top 20 zip codes(hint: branch\_zip) by total transaction value 2) Find total transaction value for each transaction type by Quarter in 2018 Hint: Find quarter from 'creditcard' table using month or use 'time' table if you already added transaction\_id column there.

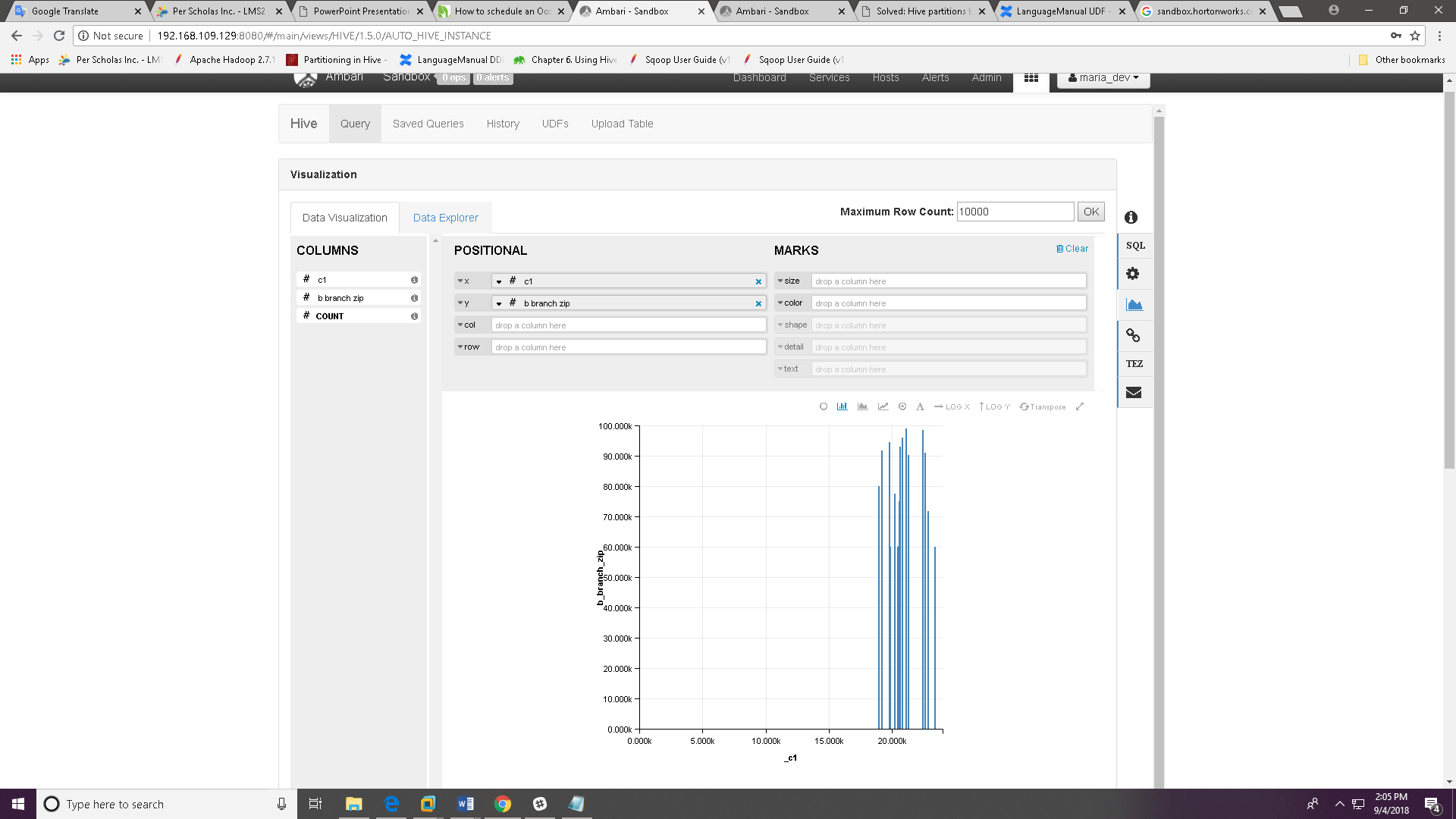
1) select b.branch\_zip, sum(transaction\_value) from cdw\_sapp\_f\_credit\_card c

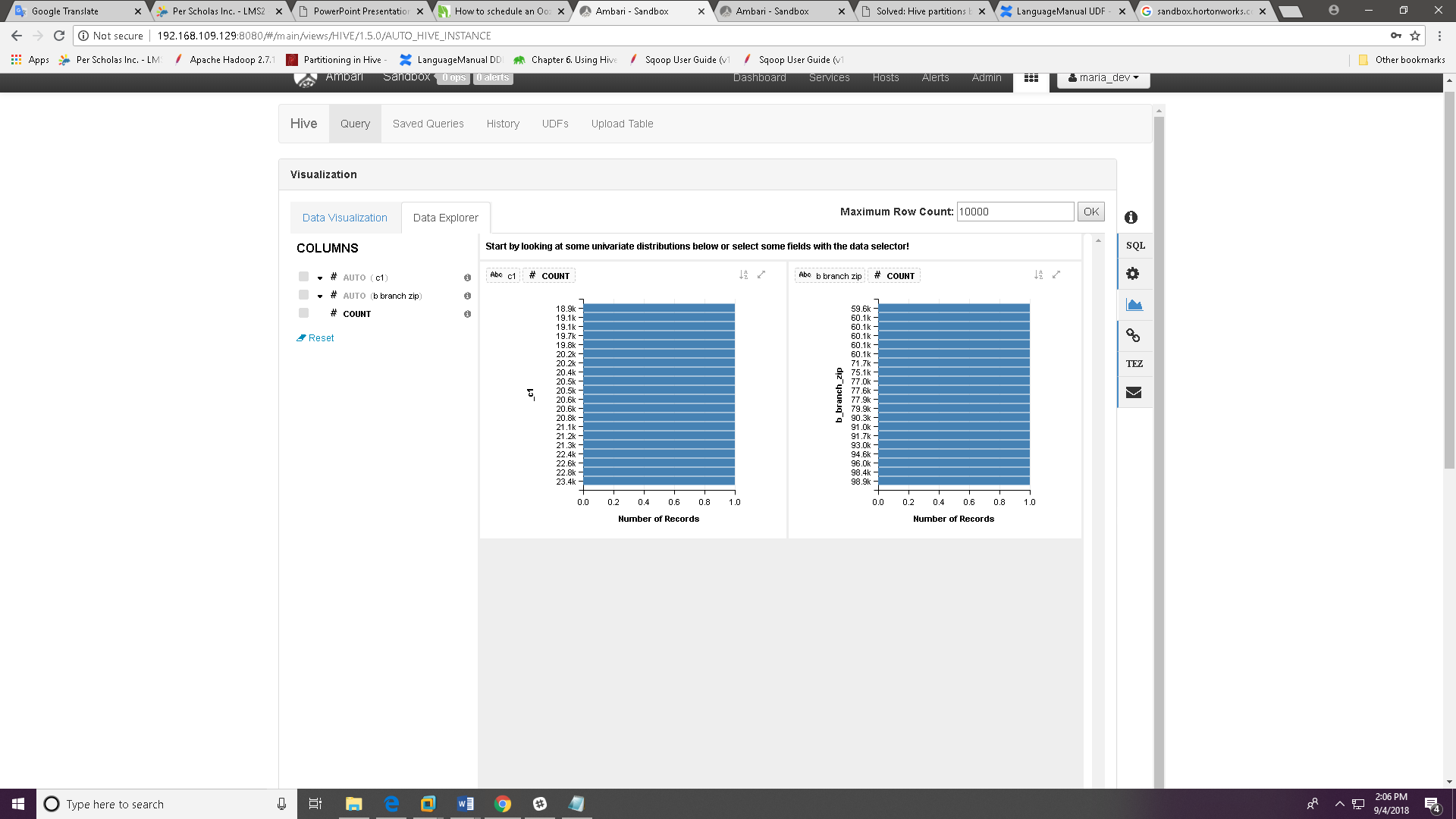
join cdw\_sapp\_d\_branch b on (c.branch\_code=b.branch\_code)

group by b.branch\_zip

order by b.branch\_zip desc limit 20;

|  |  |
| --- | --- |
| b.branch\_zip | \_c1 |
| 98908 | 21082.58 | 71730 | 22772.8 | 90278 | 21261.37 |
| 98444 | 22356.68 | 60148 | 20453.29 | 79930 | 18923.41 |
| 95993 | 20757.48 | 60103 | 20414.61 | 77904 | 21234.44 |
| 94580 | 19735.83 | 60091 | 23350.68 | 77566 | 20156.75 |
| 93035 | 20571.75 | 60089 | 19142.95 | 77016 | 20605.86 |
| 91740 | 19112.13 | 60060 | 19774.43 | 75088 | 20544.27 |
| 91010 | 22583.27 | 59601 | 20189.63 |





2)

sqoop import --connect jdbc:mysql://localhost/CDW\_SAPP --driver com.mysql.jdbc.Driver --query 'select concat(YEAR,lpad(MONTH,2,'0'),lpad(DAY,2,'0')) AS ID,DAY,MONTH,quarter(concat(YEAR,lpa

d(MONTH,2,0),lpad(DAY,2,0))) as QUARTER,YEAR,transaction\_id from CDW\_SAPP\_CREDITCARD WHERE $CONDITIONS' --target-dir /user/maria\_dev/Credit\_card\_System/time --fields-terminated-by '\t' --split-by ID -m1

drop table TIME;

create table TIME

( TIMEID VARCHAR(8), DAY INT, MONTH INT,

QUARTER VARCHAR(8), YEAR INT, TRANSACTION\_ID INT)

row format delimited fields terminated by '\t'

STORED AS TEXTFILE

Tblproperties("skip.header.line.count"="3");

LOAD DATA INPATH '/user/maria\_dev/Credit\_card\_System/time'

OVERWRITE INTO TABLE TIME;drop table CDW\_SAPP\_D\_TIME;

create external table CDW\_SAPP\_D\_TIME

( TIMEID VARCHAR(8), DAY INT, MONTH INT,

YEAR INT,TRANSACTION\_ID VARCHAR(4))

PARTITIONED BY (QUARTER VARCHAR(8))

row format delimited fields terminated by '\t'

LINES TERMINATED BY '\n'

location '/user/maria\_dev/Credit\_card\_System/CDW\_SAPP\_D\_TIME/';

SET hive.exec.dynamic.partition=true;

SET hive.exec.dynamic.partition.mode=nonstrict;

INSERT OVERWRITE TABLE CDW\_SAPP\_D\_TIME

PARTITION (QUARTER)

SELECT TIMEID, DAY, MONTH,

YEAR, TRANSACTION\_ID,QUARTER

FROM TIME;

select t.quarter, t.year, transaction\_type, sum(transaction\_value)

from cdw\_sapp\_f\_credit\_card c

join CDW\_SAPP\_D\_TIME t on (c.transaction\_id=t.transaction\_id)

group by t.quarter,t.year, c.transaction\_type

having t.year=2018

order by t.quarter,t.year, c.transaction\_type desc;

|  |  |  |  |
| --- | --- | --- | --- |
| t.quarter | t.year | transaction\_type | \_c3 |
| 1 | 2018 | Test | 85990 | 3 | 2018 | Test | 87521.19 |
| 1 | 2018 | Healthcare | 80736.7 | 3 | 2018 | Healthcare | 86299.67 |
| 1 | 2018 | Grocery | 76323.14 | 3 | 2018 | Grocery | 80955.42 |
| 1 | 2018 | Gas | 89927.84 | 3 | 2018 | Gas | 83453.32 |
| 1 | 2018 | Entertainment | 89341.24 | 3 | 2018 | Entertainment | 88121.49 |
| 1 | 2018 | Education | 80436.57 | 3 | 2018 | Education | 85844.52 |
| 1 | 2018 | Bills | 85816.21 | 3 | 2018 | Bills | 89673.01 |
| 2 | 2018 | Test | 86375.73 | 4 | 2018 | Test | 92064.83 |
| 2 | 2018 | Healthcare | 81910.16 | 4 | 2018 | Healthcare | 80670.91 |
| 2 | 2018 | Grocery | 80313.77 | 4 | 2018 | Grocery | 80987.24 |
| 2 | 2018 | Gas | 86317.78 | 4 | 2018 | Gas | 87481.44 |
| 2 | 2018 | Entertainment | 85485.3 | 4 | 2018 | Entertainment | 88950.8 |
| 2 | 2018 | Education | 84131.83 | 4 | 2018 | Education | 82839.13 |
| 2 | 2018 | Bills | 89326.39 | 4 | 2018 | Bills | 88255.17 |

