

# E-commerce Website's Clickstream Data Analysis

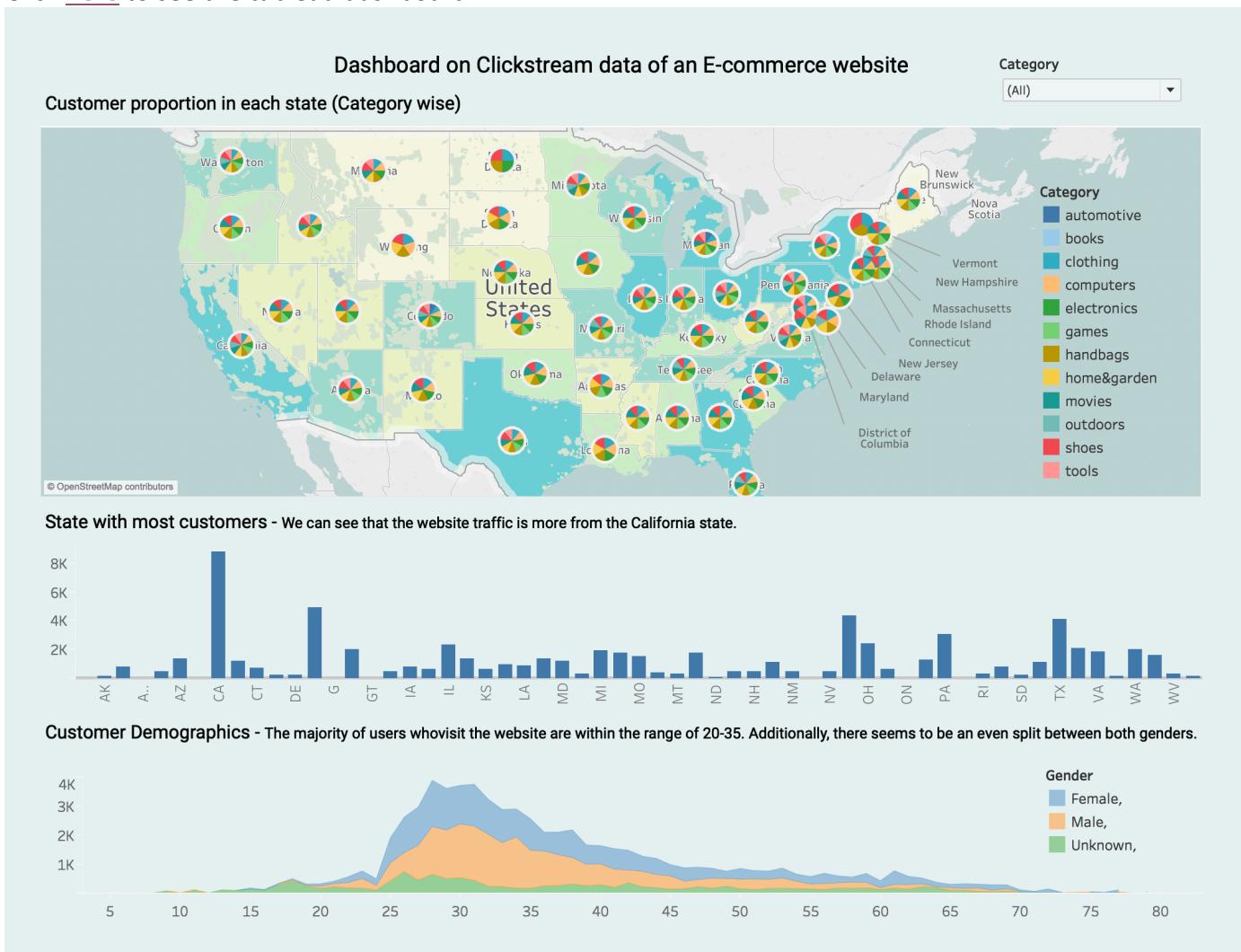
## Hadoop (Data ingestion) and Analysis (Tableau)

### Introduction:

The purpose of this project is to ingest data into HDFS environment and perform queries on those tables with Hive and use Tableau to analyze the clickstream logs files. Our main goal is to perform analysis on the clickstream data for an e-commerce site, so that we can better understand the habits of our customers and understand how to optimize future promotions and advertising.

### E-commerce website clickstream dashboard:

Click [here](#) to see the tableau dashboard



## Contents:

- Overview of Data
- Creating Hive tables in Hadoop environment
- Load Data into a table
- Customizing the table
- Visualizing the clickstream data in Tableau

## Overview of Data:

This E-commerce site has three log files,

1. **Logs.tsv:** Clickstream data is an information trail a user leaves behind while visiting a website. It is typically captured in semi-structured website log files. These website log files contain data elements such as a date and time stamp, the visitor's IP address, the URLs of the pages visited, and a user ID that uniquely identifies the user.

```

1331800486      2012-03-15 01:34:46      28599789619394381      6917530184062522013      FAS-2.8-AS3      N      0      69.76.12.213      1      0      10      http://www.acme.com/SH55126545/
V055177927      {8D0E437E-9249-4DDA-BC4F-C1E5409E3AB}      U      Y      0      0      300      rr.com      15/2/2012 1:7:2 4 420      45      41      Mozilla/5.0 (Windows NT 6.1; WOW64; rv:10.0.2) Gecko/20100101 Firefox/10.0.2      591      0      0
2      11      0      coeur d alene      usa      881      id      0      0      0      0      KXLY      0
120
KXLY

0
1331857433      2012-03-15 17:23:53      2781404195155152050      6917530222178992983      FAS-2.8-AS3      N      0      67.240.15.94      1      0      10      http://www.acme.com/SH55126545/
V055166807      {E3FEB462-CABA-11D4-820E-80A0C9E5BE2D}      Y      Y      2      0      300      rr.com      15/2/2012 19:11:20 4 240      45      2      00011,10020      Mozilla/5.0 (Macintosh; Intel Mac OS X 10_6_8) AppleWebKit/534.52.7 (KHTML, like
Gecko) Version/5.1.2 Safari/534.52.7      71      0      20      59      0      queensbury      usa      532      ny      0      0      0      0      WTN      0
0
120
WTEN

0
1331856380      2012-03-15 17:05:00      2781404195155152050      6917530222178992983      FAS-2.8-AS3      N      0      67.240.15.94      1      0      10      http://www.acme.com/SH55126545/
V055149415      {C3FEB462-CABA-11D4-820E-80A0C9E5BE2D}      Y      Y      2      0      300      rr.com      15/2/2012 19:11:20 4 240      45      2      00011,10020      Mozilla/5.0 (Macintosh; Intel Mac OS X 10_6_8) AppleWebKit/534.52.7 (KHTML, like
Gecko) Version/5.1.2 Safari/534.52.7      71      0      20      45      0      queensbury      usa      532      ny      0      0      1      0      WTN      0
0
120
WTEN

```

2. **Products.tsv:** This tsv file includes the range of products displayed in the webpage.

| url   | category    |
|---|-------------|
| <a href="http://www.acme.com/">http://www.acme.com/</a>   | books       |
| <a href="http://www.acme.com/SH55126545/VD55149415">http://www.acme.com/SH55126545/VD55149415</a> | movies      |
| <a href="http://www.acme.com/SH55126545/VD55163347">http://www.acme.com/SH55126545/VD55163347</a> | games       |
| <a href="http://www.acme.com/SH55126545/VD55165149">http://www.acme.com/SH55126545/VD55165149</a> | electronics |
| <a href="http://www.acme.com/SH55126545/VD55166807">http://www.acme.com/SH55126545/VD55166807</a> | computers   |
| <a href="http://www.acme.com/SH55126545/VD55170364">http://www.acme.com/SH55126545/VD55170364</a> | home&garden |
| <a href="http://www.acme.com/SH55126545/VD55173061">http://www.acme.com/SH55126545/VD55173061</a> | handbags    |
| <a href="http://www.acme.com/SH55126545/VD55177927">http://www.acme.com/SH55126545/VD55177927</a> | clothing    |
| <a href="http://www.acme.com/SH55126545/VD55179433">http://www.acme.com/SH55126545/VD55179433</a> | shoes       |
| <a href="http://www.acme.com/SH55126554/VD55147564">http://www.acme.com/SH55126554/VD55147564</a> | outdoors    |
| <a href="http://www.acme.com/SH5568487/VD55169229">http://www.acme.com/SH5568487/VD55169229</a>   | automotive  |
| <a href="http://www.acme.com/SH5580165/VD55156528">http://www.acme.com/SH5580165/VD55156528</a>   | clothing    |
| <a href="http://www.acme.com/SH5580165/VD55173281">http://www.acme.com/SH5580165/VD55173281</a>   | tools       |
| <a href="http://www.acme.com/SH5582037/VD5582082">http://www.acme.com/SH5582037/VD5582082</a>     | accessories |

3. **Users:** This file has all the user details like gender, and birthdate.

| SWID                                 | BIRTH_DT  | GENDER_CD |
|--------------------------------------|-----------|-----------|
| 0001BDD9-EABF-4D0D-81BD-D9EABFC0D07D | 8-Apr-84  | F         |
| 00071AA7-86D2-4EB9-871A-A786D27EB9BA | 7-Feb-88  | F         |
| 00071B7D-31AF-4D85-871B-7D31AFFD852E | 22-Oct-64 | F         |
| 0007967E-F188-4598-9C7C-E64390482CFB | 1-Jun-66  | M         |
| 000B90B2-92DC-4A7A-8B90-B292DC9A7A71 | 13-Jun-84 | M         |
| 000C1856-994E-476B-8C18-56994E676B29 | 29-Dec-80 | U         |
| 000F36E5-9891-4098-9B69-CEE78483B653 | 24-Mar-85 | F         |
| 00102F3F-061C-4212-9F91-1254F9D6E39F | 1-Nov-91  | F         |
| 0010C6F2-8C04-450E-90C6-F28C04B50E97 | 20-Jun-02 | U         |
| 0011C945-28C4-4D6F-B1E6-6CA7EFC14548 | 13-Nov-87 | F         |
| 001704E0-6CD8-429A-8E0A-89024019CA6A | 10-Jul-91 | M         |
| 001720A4-44E3-43F0-BA8F-2F2E0D7B6275 | 8-Nov-90  | M         |
| 001834AA-7451-49EA-B0E4-ED4A71C97AD1 | 21-Feb-88 | M         |
| 0018B3DA-4763-460A-B67E-1D5168E4DEB6 | 21-Apr-02 | U         |
| 00193DB8-1FE8-440B-8217-BAA2AA5FDD64 | 31-Jul-70 | M         |

## Creating Hive Tables in Hadoop environment:

To store the tables I created a Database “**retailstore\_with\_commos**”

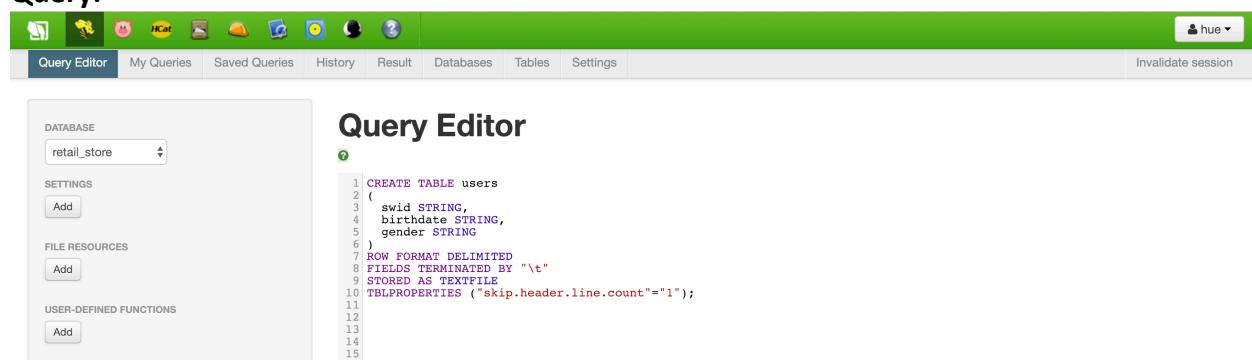
### Query:

```
create database retail_store;
```

### creating Users table:

Since the Users text file is tab separated, we use the below query to create the table under **retailstore** database.

### Query:



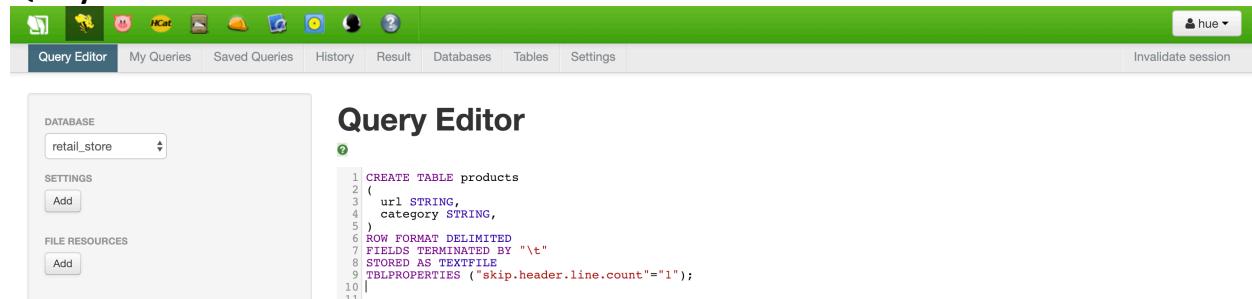
The screenshot shows the Hue Query Editor interface. The top navigation bar includes icons for Home, My Queries, Saved Queries, History, Result, Databases, Tables, Settings, and a user icon. The title bar says "Query Editor". On the left, there's a sidebar with sections for DATABASE (set to "retail\_store"), SETTINGS, FILE RESOURCES, and USER-DEFINED FUNCTIONS, each with an "Add" button. The main area is titled "Query Editor" and contains the following SQL code:

```
CREATE TABLE users
(
    swid STRING,
    birthdate STRING,
    gender STRING
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY "\t"
STORED AS TEXTFILE
TBLPROPERTIES ("skip.header.line.count"="1");
```

### creating Products table:

we use the similar query to create category table,

### Query:

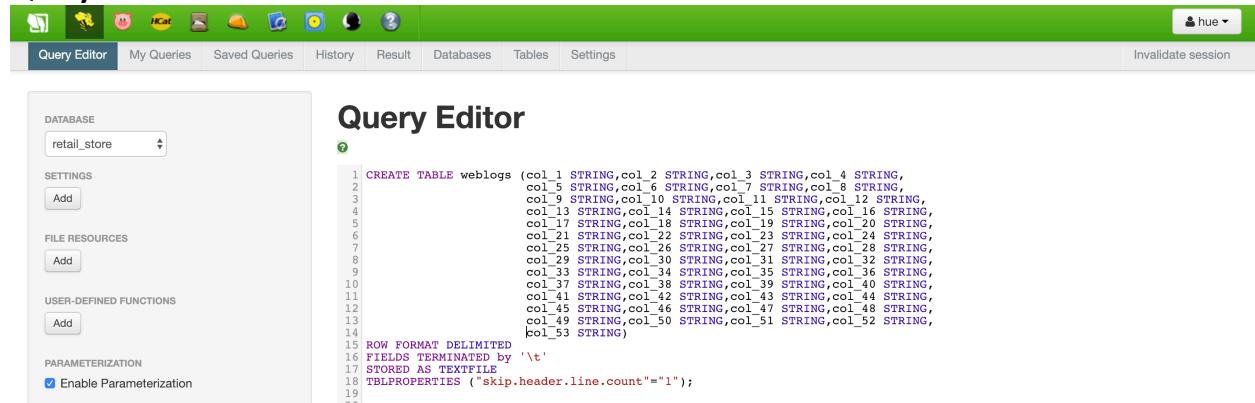


The screenshot shows the Hue Query Editor interface. The top navigation bar includes icons for Home, My Queries, Saved Queries, History, Result, Databases, Tables, Settings, and a user icon. The title bar says "Query Editor". On the left, there's a sidebar with sections for DATABASE (set to "retail\_store"), SETTINGS, FILE RESOURCES, and USER-DEFINED FUNCTIONS, each with an "Add" button. The main area is titled "Query Editor" and contains the following SQL code:

```
CREATE TABLE products
(
    url STRING,
    category STRING,
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY "\t"
STORED AS TEXTFILE
TBLPROPERTIES ("skip.header.line.count"="1");
```

## creating weblogs table:

Query:



The screenshot shows the Apache Hue Query Editor interface. The top navigation bar includes icons for Home, Help, My Queries, Saved Queries, History, Result, Databases, Tables, Settings, and a user dropdown. Below the navigation is a toolbar with Query Editor, My Queries, Saved Queries, History, Result, Databases, Tables, Settings, and Invalidate session. On the left, there's a sidebar with sections for DATABASE (set to 'retail\_store'), SETTINGS (with an 'Add' button), FILE RESOURCES (with an 'Add' button), USER-DEFINED FUNCTIONS (with an 'Add' button), and PARAMETERIZATION (with a checked 'Enable Parameterization' checkbox). The main area is titled 'Query Editor' and contains the following SQL code:

```
1 CREATE TABLE weblogs (col_1 STRING,col_2 STRING,col_3 STRING,col_4 STRING,
2 col_5 STRING,col_6 STRING,col_7 STRING,col_8 STRING,col_9 STRING,
3 col_10 STRING,col_11 STRING,col_12 STRING,
4 col_13 STRING,col_14 STRING,col_15 STRING,col_16 STRING,
5 col_17 STRING,col_18 STRING,col_19 STRING,col_20 STRING,
6 col_21 STRING,col_22 STRING,col_23 STRING,col_24 STRING,
7 col_25 STRING,col_26 STRING,col_27 STRING,col_28 STRING,
8 col_29 STRING,col_30 STRING,col_31 STRING,col_32 STRING,
9 col_33 STRING,col_34 STRING,col_35 STRING,col_36 STRING,
10 col_37 STRING,col_38 STRING,col_39 STRING,col_40 STRING,
11 col_41 STRING,col_42 STRING,col_43 STRING,col_44 STRING,
12 col_45 STRING,col_46 STRING,col_47 STRING,col_48 STRING,
13 col_49 STRING,col_50 STRING,col_51 STRING,col_52 STRING,
14 col_53 STRING)
15 ROW FORMAT DELIMITED
16 FIELDS TERMINATED BY '\t'
17 STORED AS TEXTFILE
18 TBLPROPERTIES ('skip.header.line.count'='1');
19
^~
```

Three tables are created through Hive interface and stored in Hadoop environment.

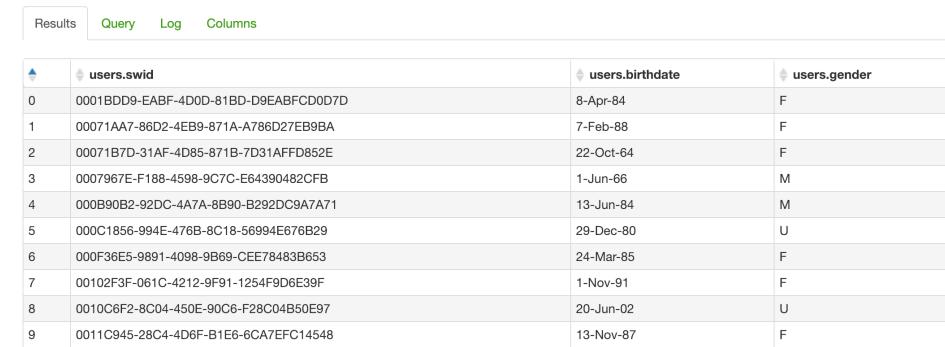
## Loading data into hive tables:

Import the text file into HDFS and use the below query to load in to the users table.

Query:

```
1 LOAD DATA INPATH '/user/hue/rajesh/WeblogAnalytics_1/users.tsv' OVERWRITE INTO TABLE users;
2
```

## Table structure:



The screenshot shows the Apache Hue Results interface. The top navigation bar includes icons for Home, Help, My Queries, Saved Queries, History, Result, Databases, Tables, Settings, and a user dropdown. Below the navigation is a toolbar with Results, Query, Log, and Columns. The main area displays the schema of the 'users' table:

|   | users.swid                           | users.birthdate | users.gender |
|---|--------------------------------------|-----------------|--------------|
| 0 | 0001BDD9-EABF-4D0D-81BD-D9EABFC0D7D  | 8-Apr-84        | F            |
| 1 | 00071AA7-86D2-4EB9-871A-A786D27EB9BA | 7-Feb-88        | F            |
| 2 | 00071B7D-31AF-4D85-871B-7D31AFFD852E | 22-Oct-64       | F            |
| 3 | 0007967E-F188-4598-9C7C-E64390482CFB | 1-Jun-66        | M            |
| 4 | 000B90B2-92DC-4A7A-8B90-B292DC9A7A71 | 13-Jun-84       | M            |
| 5 | 000C1856-994E-476B-8C18-56994E676B29 | 29-Dec-80       | U            |
| 6 | 000F36E5-9891-4098-9B69-CEE78483B653 | 24-Mar-85       | F            |
| 7 | 00102F3F-061C-4212-9F91-1254F9D6E39F | 1-Nov-91        | F            |
| 8 | 0010C6F2-8C04-450E-90C6-F28C04B50E97 | 20-Jun-02       | U            |
| 9 | 0011C945-28C4-406F-B1E6-6CA7EFC14548 | 13-Nov-87       | F            |

As we can see that gender column is not that clean so let's write a query to convert 'F' to Female, 'M' to Male and 'U' to Unknown.

```
1 INSERT OVERWRITE TABLE users
2 SELECT swid,
3       birthdate,
4       CASE WHEN users.gender='F' THEN 'Female'
5             WHEN users.gender='M' THEN 'Male'
6             WHEN users.gender='U' THEN 'Unknown'
7             END AS gender
8 FROM users;
```

Now the table look like,

|   | ◆ retail_store.users.swid            | ◆ retail_store.users.birthdate | ◆ retail_store.users.gender |
|---|--------------------------------------|--------------------------------|-----------------------------|
| 0 | 00071AA7-86D2-4EB9-871A-A786D27EB9BA | 7-Feb-88                       | Female                      |
| 1 | 00071B7D-31AF-4D85-871B-7D31AFFD852E | 22-Oct-64                      | Female                      |
| 2 | 0007967E-F188-4598-9C7C-E64390482CFB | 1-Jun-66                       | Male                        |
| 3 | 000B90B2-92DC-4A7A-8B90-B292DC9A7A71 | 13-Jun-84                      | Male                        |
| 4 | 000C1856-994E-476B-8C18-56994E676B29 | 29-Dec-80                      | Unknown                     |
| 5 | 000F36E5-9891-4098-9B69-CEE78483B653 | 24-Mar-85                      | Female                      |

## Loading Product table:

Query:

```
1 LOAD DATA INPATH '/user/hue/rajesh/WeblogAnalytics_1/product.tsv' OVERWRITE INTO TABLE products;
```

|   | ◆ retailstore_with_commos.products.url    | ◆ retailstore_with_commos.products.category |
|---|---|---|
| 0 | http://www.acme.com/                      | books                                       |
| 1 | http://www.acme.com/SH55126545/VD55149415 | movies                                      |
| 2 | http://www.acme.com/SH55126545/VD55163347 | games                                       |
| 3 | http://www.acme.com/SH55126545/VD55165149 | electronics                                 |
| 4 | http://www.acme.com/SH55126545/VD55166807 | computers                                   |
| 5 | http://www.acme.com/SH55126545/VD55170364 | home&garden                                 |
| 6 | http://www.acme.com/SH55126545/VD55173061 | handbags                                    |
| 7 | http://www.acme.com/SH55126545/VD55177927 | clothing                                    |

## Loading Weblogs table:

Query:

```
LOAD DATA INPATH '/user/hue/rajesh/WeblogAnalytics_1/logs.tsv' OVERWRITE INTO TABLE weblogs;
```

|   | ◆ retail_store.weblogs.col_1 | ◆ retail_store.weblogs.col_2 | ◆ retail_store.weblogs.col_3 | ◆ retail_store.weblogs.col_4 | ◆ retail_store.weblogs.col_5 |
|---|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|
| 0 | 1331800486                   | 2012-03-15 01:34:46          | 2859997896193943381          | 6917530184062522013          | FAS-2.8-AS3                  |
| 1 | 1331857433                   | 2012-03-15 17:23:53          | 2781404195155152050          | 6917530222178992983          | FAS-2.8-AS3                  |
| 2 | 1331856300                   | 2012-03-15 17:05:00          | 2781404195155152050          | 6917530222178992983          | FAS-2.8-AS3                  |
| 3 | 1331800073                   | 2012-03-15 01:27:53          | 2829533710344657576          | 6917530841727896174          | FAS-2.8-AS3                  |
| 4 | 1331802578                   | 2012-03-15 02:09:38          | 5279782221568288854          | 732692051711230006           | FAS-2.8-AS3                  |
| 5 | 1331831882                   | 2012-03-15 10:18:02          | 2754911479663756329          | 6917530150775072471          | FAS-2.8-AS3                  |
| 6 | 1331836722                   | 2012-03-15 11:38:42          | 2838741409408826824          | 6917530129301952401          | FAS-2.8-AS3                  |

## Data cleaning/Customization:

We can see from the weblogs table there are lot of unwanted data lets clean up the table and get it into a clean format.

```
1 CREATE TABLE weblogs_updated AS
2 SELECT
3   CAST(col_2 as STRING) as ts,
4   CAST(col_8 as STRING) as ip,
5   CAST(col_13 as STRING) as url,
6   CAST(regexp_replace(col_14, '[^a-zA-Z0-9\u00E0-\u00FC ]+', '') as STRING) as swid,
7   CAST(UPPER(col_50) as STRING) as city,
8   CAST(UPPER(col_51) as STRING) as country,
9   CAST(UPPER(col_53) as STRING) as state
10 FROM weblogs;
```

| <code>retail_store.weblogs_updated.ts</code> | <code>retail_store.weblogs_updated.ip</code> | <code>retail_store.weblogs_updated.url</code> | <code>retail_store.weblogs_updated.swid</code> | <code>retail_store.weblogs_updated.city</code> | <code>retail</code> |
|--|--|---|--|--|---------------------|
| 2012-03-15 01:34:46                          | 69.76.12.213                                 | http://www.acme.com/SH55126545/VD55177927     | 8D0E437E92494DDABC4FC1E5409E3A3B               | COEUR D ALENE                                  | USA                 |
| 2012-03-15 17:23:53                          | 67.240.15.94                                 | http://www.acme.com/SH55126545/VD55166807     | E3FEBAA62CABA11D4820E00A0C9E58E2D              | QUEENSBURY                                     | USA                 |
| 2012-03-15 17:05:00                          | 67.240.15.94                                 | http://www.acme.com/SH55126545/VD55149415     | E3FEBAA62CABA11D4820E00A0C9E58E2D              | QUEENSBURY                                     | USA                 |
| 2012-03-15 01:27:53                          | 98.234.107.75                                | http://www.acme.com/SH55126545/VD55179433     | 49E0D2EE1D5748C5A27D7660C78CB55C               | SUNNYVALE                                      | USA                 |
| 2012-03-15 02:09:38                          | 75.85.165.38                                 | http://www.acme.com/SH55126545/VD55179433     | F6F8B4604204C26A32CB93826DCB99                 | SAN DIEGO                                      | USA                 |
| 2012-03-15 10:18:02                          | 71.53.206.175                                | http://www.acme.com/SH55126545/VD55166807     | AC4D6E522F0B48E29CA0770B68222711               | CHARLOTTESVILLE                                | USA                 |
| 2012-03-15 11:38:42                          | 97.96.62.161                                 | http://www.acme.com/SH55126545/VD55179433     | 56C1008E60954BE1A58B9523459F295C               | PARRISH  | USA                 |
| 2012-03-15 12:42:59                          | 129.119.158.240                              | http://www.acme.com/SH55126545/VD55170364     | CC5F240A9B9E4288B7AF91521D6F7A1C               | DALLAS   | USA                 |
| 2012-03-15 11:02:49                          | 96.241.99.50                                 | http://www.acme.com/SH55126545/VD55179433     | 303AADFF4CB8480CA757213BEEB051BC               | CAPITOL HEIGHTS                                | USA                 |
| 2012-03-15 11:14:20                          | 96.241.99.50                                 | http://www.acme.com/SH55126545/VD55179433     | 303AADFF4CB8480CA757213BEEB051BC               | CAPITOL HEIGHTS                                | USA                 |
| 2012-03-15 10:38:37                          | 24.187.64.39                                 | http://www.acme.com/SH55126545/VD55179433     | 66D56071EA8B4AA6FD0D2BFA1DD47EC                | NEW BRUNSWICK                                  | USA                 |
| 2012-03-15 09:52:16                          | 98.184.170.44                                | http://www.acme.com/SH55126545/VD55179433     | 04721673E4224478AEB1966F7F9CAB76               | TULSA  | USA                 |
| 2012-03-15 10:57:04                          | 75.135.144.63                                | http://www.acme.com/SH55126545/VD55179433     | FC431962DA5F4CA3A7D2B43DC7E156C6               | ROCKFORD                                       | USA                 |
| 2012-03-15 10:32:20                          | 67.191.202.209                               | http://www.acme.com/SH55126545/VD55177927     | A2DB48C439234D278D37E315AA165C74               | MARIETTA                                       | USA                 |
| 2012-03-15 11:33:13                          | 71.53.206.175                                | http://www.acme.com/SH55126545/VD55170364     | AC4D6E522F0B48E29CA0770B68222711               | CHARLOTTESVILLE                                | USA                 |

## Join Data from Multiple Tables:

Let's play with our data further, taking specific fields from different tables and creating a custom table "webanalytics" from them.

```

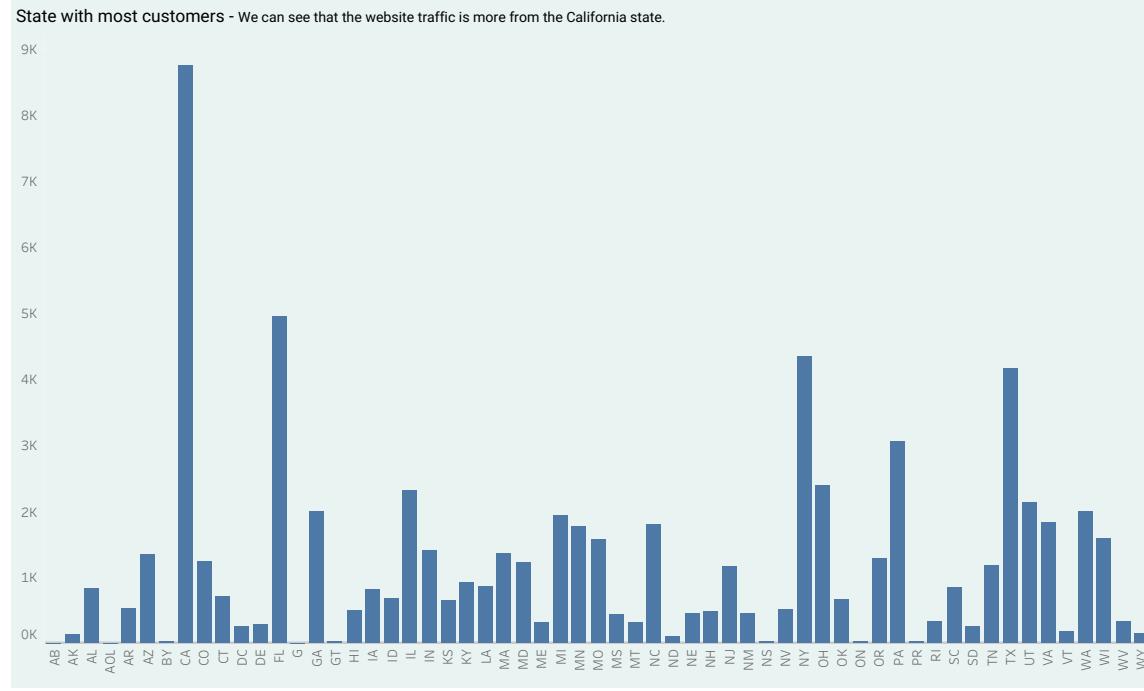
1 CREATE TABLE webanalytics AS
2 SELECT to_date(o.ts) logdate, o.url, o.ip, o.city, o.state,
3 o.country, p.category, CAST(datediff(from_unixtime(unix_timestamp()),
4 from_unixtime(unix_timestamp(u.birthdate, 'dd-MMM-yy')))) / 365 AS INT) age, u
5 FROM weblogs_updated o
6 INNER JOIN retailstore_with_commos.products p
7 ON o.url = p.url
8 LEFT OUTER JOIN users u
9 ON o.swid = u.swid;

```

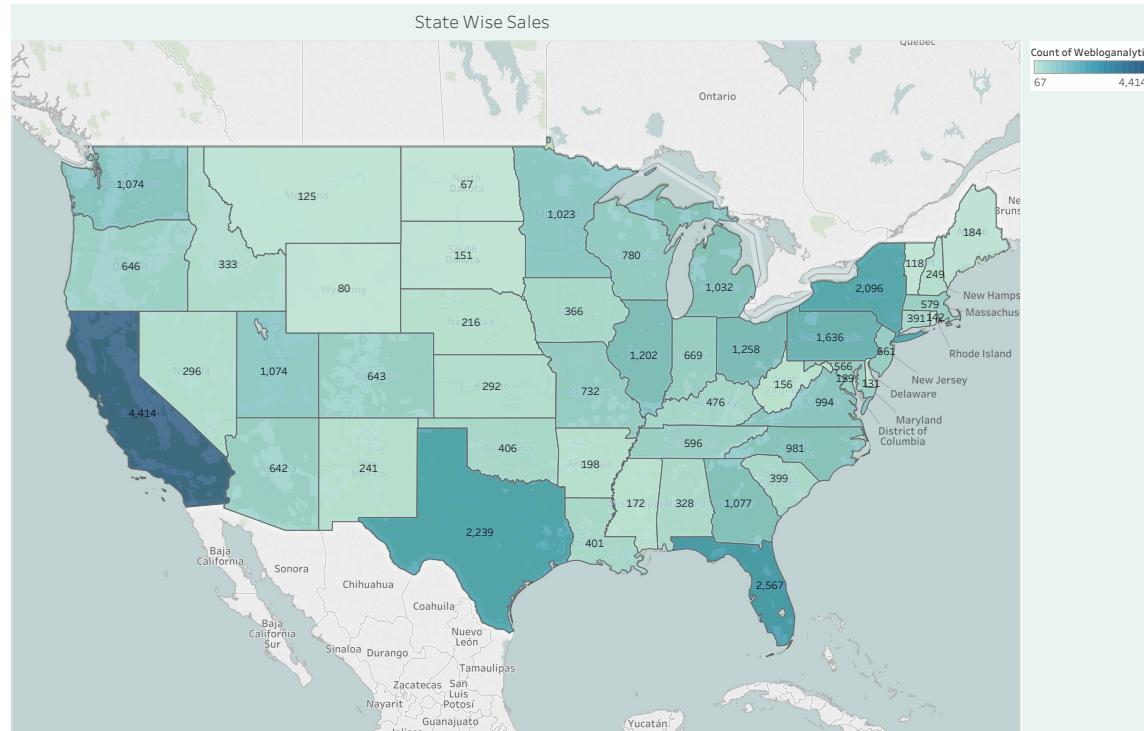
| <code>retail_store.webanalytics.logdate</code> | <code>retail_store.webanalytics.url</code> | <code>retail_store.webanalytics.ip</code> | <code>retail_store.webanalytics.city</code> | <code>retail_store.webanalytics.state</code> | <code>retail_store.webanalytics.state</code> | <code>retail_store.webanalytics.country</code> | <code>retail_store.webanalytics.category</code> | <code>retail_store.webanalytics.age</code> | <code>retail_store.webanalytics.gender</code> |
|--|--|---|---|--|--|--|---|--|---|
| 2012-03-15                                     | http://www.acme.com/SH55126545/VD55177927  | 69.76.12.213                              | COEUR D ALENE                               | ID   | NY   | USA  | clothing  | 37   | Female  |
| 2012-03-15                                     | http://www.acme.com/SH55126545/VD55166807  | 67.240.15.94                              | QUEENSBURY                                  | NY   | NY   | USA  | computers                                       | 36   | Male  |
| 2012-03-15                                     | http://www.acme.com/SH55126545/VD55149415  | 67.240.15.94                              | QUEENSBURY                                  | NY   | CA   | USA  | movies  | 36   | Male  |
| 2012-03-15                                     | http://www.acme.com/SH55126545/VD55179433  | 98.234.107.75                             | SUNNYVALE                                   | CA   | CA   | USA  | shoes   | 22   | Male  |
| 2012-03-15                                     | http://www.acme.com/SH55126545/VD55179433  | 75.85.165.38                              | SAN DIEGO                                   | CA   | CA   | USA  | shoes   | 29   | Female  |
| 2012-03-15                                     | http://www.acme.com/SH55126545/VD55166807  | 71.53.206.175                             | CHARLOTTESVILLE                             | VA   | VA   | USA  | computers                                       | 27   | Female  |
| 2012-03-15                                     | http://www.acme.com/SH55126545/VD55179433  | 97.96.62.161                              | PARRISH                                     | FL   | FL   | USA  | shoes   | 46   | Female  |
| 2012-03-15                                     | http://www.acme.com/SH55126545/VD55170364  | 24.187.64.39                              | DALLAS                                      | TX   | TX   | USA  | home&garden                                     | 28   | Female  |
| 2012-03-15                                     | http://www.acme.com/SH55126545/VD55179433  | 96.241.99.50                              | CAPITOL HEIGHTS                             | MD   | MD   | USA  | shoes   | 44   | Female  |
| 2012-03-15                                     | http://www.acme.com/SH55126545/VD55179433  | 96.241.99.50                              | CAPITOL HEIGHTS                             | MD   | MD   | USA  | shoes   | 44   | Female  |
| 2012-03-15                                     | http://www.acme.com/SH55126545/VD55179433  | 24.187.64.39                              | NEW BRUNSWICK                               | NJ   | NJ   | USA  | shoes   | 44   | Male  |

## Tableau- Clickstream data visualization:

- Based on the clickstream data the below visualization shows us that most of the users are from California state.

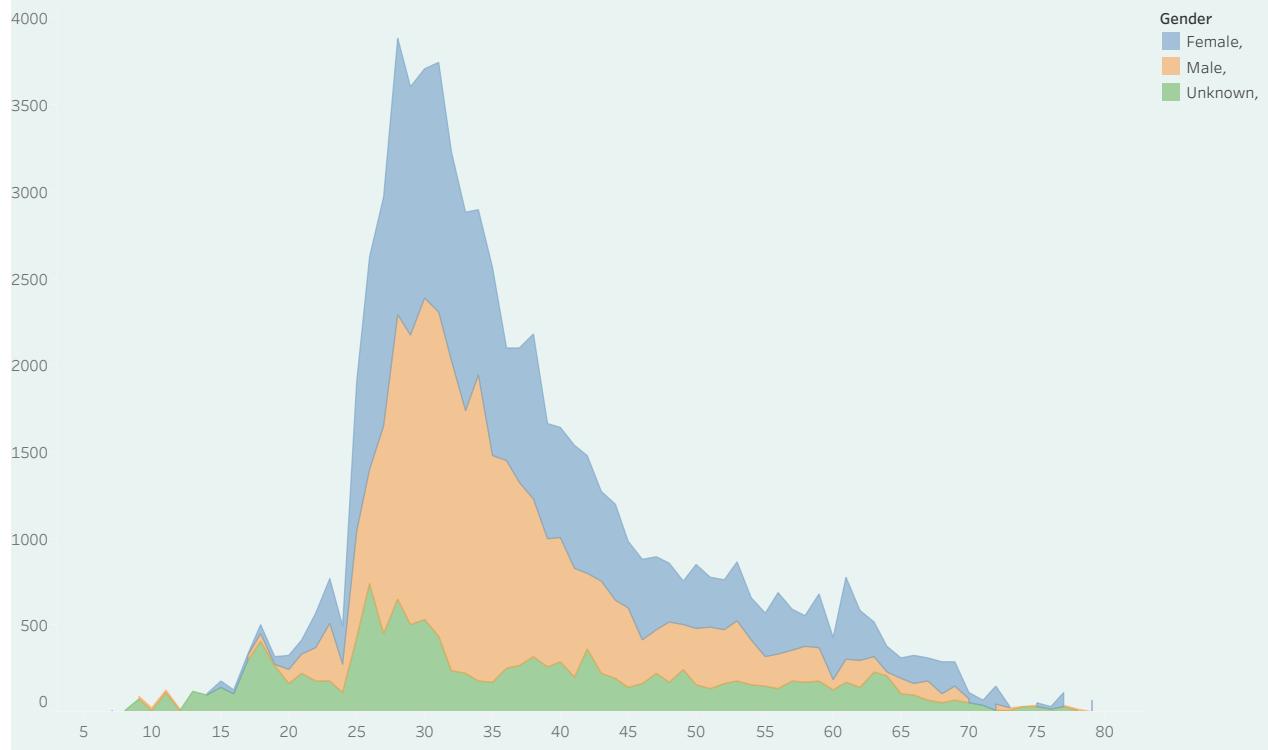


- Count of sales in each state for the month of March 2012



- Majority of users who visit the website are within age range of 20-30.

**Customer Demographics** - The majority of users who visit the website are within the range of 20-35. Additionally, there seems to be an even split between both genders.

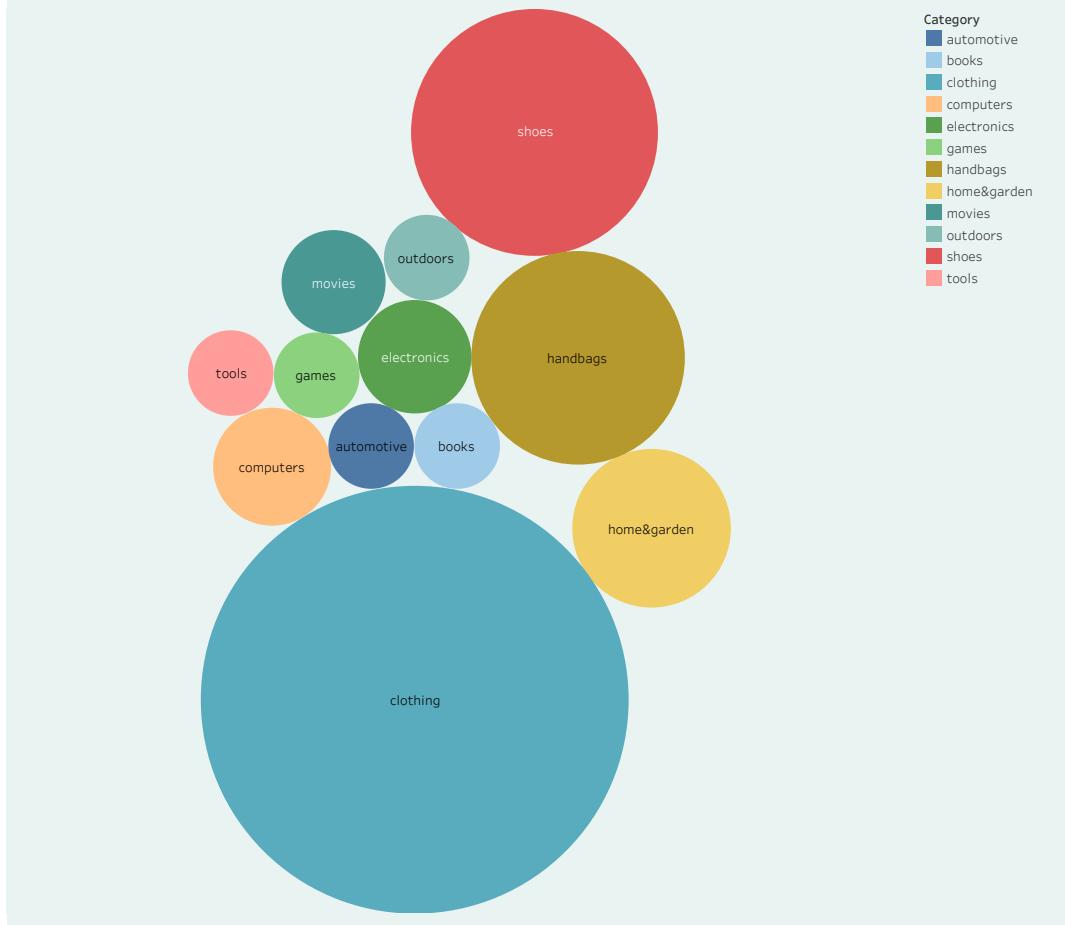


- Based on the below plot we can see that the E-commerce site has strong hold in clothing.
- Many users prefer purchasing clothing over the other categories.

Top 3 categories are,

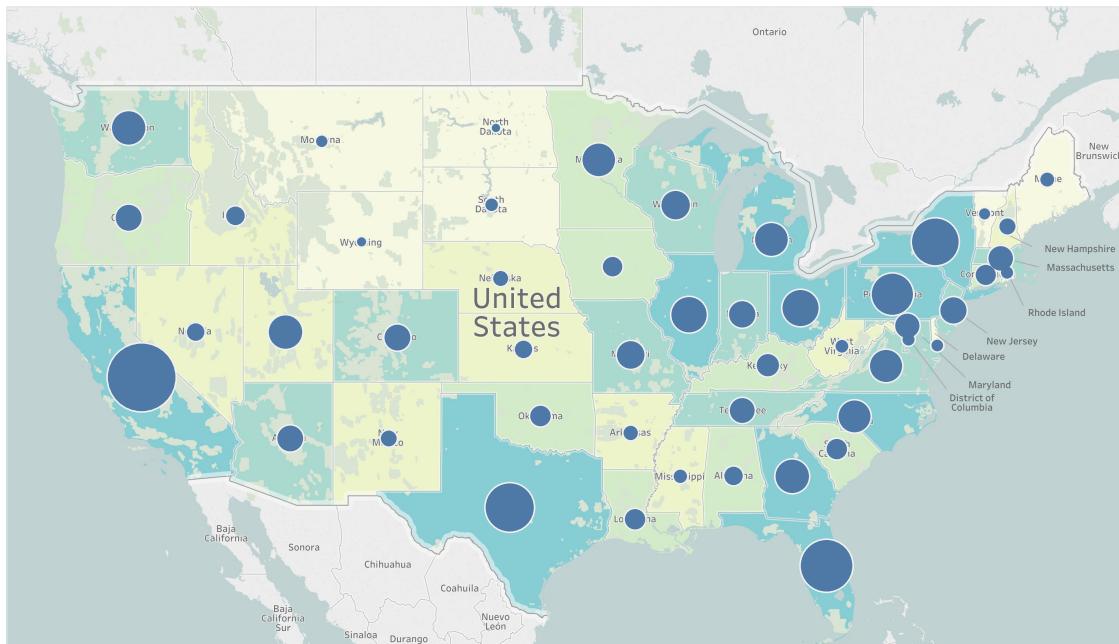
1. Clothing
2. Shoes
3. Handbags

Top categories- Higher the size of a category more the users interested in.



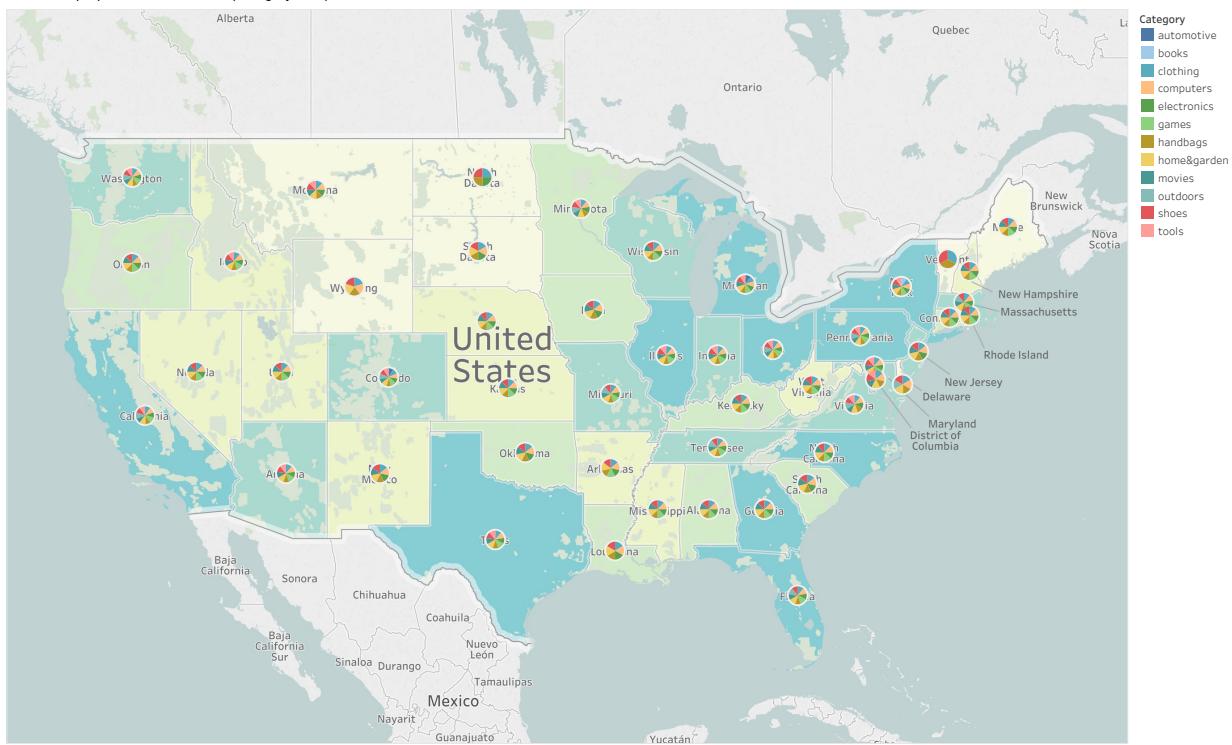
- We can see California comes in top position for clothing business as well.

State wise sale in clothing



## Category wise distribution in each state:

Customer proportion in each state (Category wise)



## Recommendations/Future actions:

- Users are more in California when compared with other states. This could be because of various reasons like there are many IT engineers located around the bay area who have more online purchasing power than the other people.
- Seems like most of the users are in the age range of 25 to 40. It could either be because of the nature of the site that caters only to young people or it could be possible that older people find it difficult to navigate through the site for a meaningful conversion.