

Local Business and Reviews Data Analysis

Project Tutorial

Hemamalini Madhanguru (hmadhan@calstatela.edu), Mahsa Tayer Farahani (mtayerf@calstatela.edu),
Ruchi Singh (rsingh26@calstatela.edu), Yashaswi Ananth (yananth@calstatela.edu)

Introduction

All local business in today's time, engage with the consumers online for marketing, feedback, repeat business etc. One of the most popular and widely used business and services review platforms are Yelp and Google Local. Data collected from these platforms is the lifeline of the registered businesses. Analyzing this data is the key to critical marketing decisions, fine-tune product/services strategy or improve customer relationship. The dataset spans a variety of businesses such as restaurants, shopping, nightlife, medical, education entertainment, common services, etc. The user reviews collected from Google Local contains valuable information on consumer and business trends. Analysis of this data can derive a lot of inferences about the local business and lists the factors responsible for making a business popular based on the reviews and ratings. In depth analysis of the business reviews can help us understand the sentiments of the customers.

Objective

In this tutorial, you will try to analyze the local business data and reviews to get insights on the popularity of a business and factors responsible for it. To get better insights from the reviews, you will also analyze the sentiments of the customer review. The total size of the dataset is 180 MB. You will use IBM BigInsight and Azure HDInsight platforms and HiveQL and Pig for query processing.

In this hands-on lab, you will learn how to:

- Download data from amazon.aws and upload the data to the Hadoop cluster
 - Analyze the Yelp and Google API dataset and derive insights
 - Use HiveQL and PIG commands to perform the analysis
 - Visualize data using Excel Power View and Tableau
-

Exercise 1: Download data and upload it to the Hadoop cluster

- 1) Download the dataset file (yelp_raw_fall_2016.csv) from Amazon S3 (https://s3.amazonaws.com/hipicdatasets/yelp_raw_fall_2016.csv)
- 2) Import reviews.csv file from the link (<https://docs.google.com/uc?id=0B9kspRX6SWaaMIRvREQ3NmUxOE0&export=download>)
- 3) Sign into your Bluemix account and select 'BigInsights for Apache Hadoop' under services.
- 4) Launch IBM BigInsights and click on Create Cluster.
- 5) Provide the cluster name and required credentials to create a new cluster

The screenshot shows the 'New Cluster' creation form in the IBM Bluemix Data & Analytics interface. The top navigation bar includes 'Docs', a menu icon, and the 'IBM Bluemix Data & Analytics' logo. Below the header, the sub-section 'BigInsights > Instance Name' is visible. The main form has two sections: 'Total Number of Nodes' (set to 2) and 'Estimated Cost per Hour' (Free for Beta). The 'Cluster Name' field contains 'Ruchi7th', and the 'User Name' field contains 'rsingh26'. Both fields have validation messages below them. The 'Password' and 'Confirm Password' fields both contain masked text (*****). A note below the password field specifies the password requirements: at least 12 characters long, containing only letters or numbers, and must contain at least one letter and one number.

- 6) Make sure you select the Spark, Pig, Sqoop and Flume components before creating the cluster.

The created cluster will be shown in active status as follows:

The screenshot shows the IBM Bluemix Data & Analytics Cluster List page. At the top, there are navigation links for 'Docs', 'IBM Bluemix Data & Analytics', 'Catalog', 'Support', and 'Account'. The main area is titled 'Cluster List' under 'BigInsights'. Below this, there are tabs for 'Clusters' (which is selected) and 'Data Store'. A button for 'New Cluster' is visible. The central part of the screen is a table titled 'Clusters (7)' with columns: 'Cluster Name', 'Management Nodes', 'Data nodes', 'High Availability', 'Status', and 'Actions'. The table lists seven clusters:

Cluster Name	Management Nodes	Data nodes	High Availability	Status	Actions
classHadoopCluster	0	0	No	● Expired	
classHadoop	0	0	No	● Expired	
class_Hadoop	0	0	No	● Expired	
classHadoop2	0	0	No	● Expired	
classHadoop4	1	1	No	● Active	
classHadoop1	0	0	No	● Deleted	
classHadoop3	0	0	No	● Deleted	

7) The file can be uploaded into Hadoop using the following commands:

```
$ hdfs dfs -mkdir Project
$ hdfs dfs -mkdir Project/Business
$ hdfs dfs -mkdir Project/Reviews
$ hdfs dfs -put Local_Bussiness2016.csv Project/Business
$ hdfs dfs -put reviews.csv Project/Reviews
```

```
-bash-4.1$ hdfs dfs -mkdir Project
-bash-4.1$ hdfs dfs -mkdir Project/Business
-bash-4.1$ hdfs dfs -mkdir Project/Reviews
-bash-4.1$ ls
Local_Business2016.csv  midPart1      pig_1479933415551.log  reviews.csv
lost+found                midPart1.tgz  products.java
-bash-4.1$ hdfs dfs -put Local_Bussiness2016.csv Project/Business
-bash-4.1$ hdfs dfs -put reviews.csv Project/Reviews
-bash-4.1$ hdfs dfs -ls
Found 4 items
drwx-----  - hmadhan hdfs          0 2016-11-24 06:00 .Trash
drwx-----  - hmadhan hdfs          0 2016-11-24 03:38 .staging
drwxr-xr-x  - hmadhan hdfs          0 2016-11-29 07:27 Project
drwxr-xr-x  - hmadhan hdfs          0 2016-11-24 02:11 dualcore
-bash-4.1$ hdfs dfs -ls /user/hmadhan/Project
Found 2 items
drwxr-xr-x  - hmadhan hdfs          0 2016-11-29 07:45 /user/hmadhan/Project/Business
drwxr-xr-x  - hmadhan hdfs          0 2016-11-29 07:46 /user/hmadhan/Project/Reviews
```

NOTE: Make a note of the path for the files after uploading them for future references.

Exercise 2: Create tables and import data into the tables from the uploaded files

The following Hive statement creates an external table that allows Hive to query data stored in HDFS. External tables preserve the data in the original file format, while allowing Hive to perform queries against the data within the file. The Hive statement below creates a new table named biz by describing the fields within the files, the delimiter (comma) between fields, and the location of the file in Azure Blob Storage. This will allow you to create Hive queries over your data.

Run the following commands in the Hive environment to create the master table and populate data into it:

1) Creating table for Local Business Dataset

```
$ hive  
hive> DROP TABLE IF EXISTS biz;  
  
hive> CREATE EXTERNAL TABLE IF NOT EXISTS biz  
(business_id string, full_address string, Friday_close String, Friday_open String, Tuesday_close  
String, Tuesday_open String, Thursday_close String, Thursday_open String, Sunday_close String,  
Sunday_open String, Saturday_close String, Saturday_open String, Wednesday_close String,  
Wednesday_open String, Monday_close String, Monday_open String, categories_001 String,  
categories_002 String, categories_003 String, categories_004 String, categories_005 String,  
categories_006 String, categories_007 String, categories_008 String, categories_009 String,  
categories_010 String, city String, review_count Int, name String, longitude Double, latitude Double,  
neighborhoods_001 String, neighborhoods_002 String, neighborhoods_003 String, state String, stars  
Double, Accepts_Insurance Boolean, coloring Boolean, africanamerican Boolean, curly Boolean, perms  
Boolean, kids Boolean, extensions Boolean, asian Boolean, straightperms Boolean,  
By_Appointment_Only Boolean, Happy_Hour Boolean, Take_out Boolean, Drive_Thru Boolean,  
Good_For_dessert Boolean, Good_For_latenight Boolean, Good_For_lunch Boolean, Good_For_dinner  
Boolean, Good_For_brunch Boolean, Good_For_breakfast Boolean, Caters Boolean, BYOB Boolean,  
Corkage Boolean, Noise_Level String, Takes_Reservations Boolean, Dietary_Restrictions_dairy_free  
Boolean, Dietary_Restrictions_gluten_free Boolean, Dietary_Restrictions_vegan Boolean,  
Dietary_Restrictions_kosher Boolean, Dietary_Restrictions_halal Boolean, Dietary_Restrictions_soy_free  
Boolean, Dietary_Restrictions_vegetarian Boolean, Delivery Boolean, Dogs_Allowed Boolean,
```

```

Coat_Check Boolean, Smoking String, Ambience_romantic Boolean, Ambience_intimate Boolean,
Ambience_classy Boolean, Ambience_hipster Boolean, Ambience_divey Boolean, Ambience_touristy
Boolean, Ambience_trendy Boolean, Ambience_upscale Boolean, Ambience_casual Boolean,
Parking_garage Boolean, Parking_street Boolean, Parking_validated Boolean, Parking_lot Boolean,
Parking_valet Boolean, Ages_Allowed Boolean, Wheelchair_Accessible Boolean, Open24Hours Boolean,
Music_dj Boolean, Music_background_music Boolean, Music_jukebox Boolean, Music_live Boolean,
Music_video Boolean, Music_karaoke Boolean, Has_TV Boolean, Outdoor_Seating Boolean, Attire
String, Alcohol Boolean, Waiter_Service Boolean, Wi_Fi String, Order_at_Counter Boolean,
Accepts_Credit_Cards Boolean, BYOB_Corkage String, Good_for_Kids Boolean, Good_For_Groups
Boolean, Price_Range Int, Good_For_Dancing Boolean, type String)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION
'/user/hmadhan/Project/Business/'

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

```

Run the below command to view the first 10 rows from the table biz

```
hive> SELECT * FROM biz LIMIT 10;
```

```

hive> select * from biz limit 10;
OK
ad7FFTzCmqQBm2H9_VVtkQ 6645 W Badura Ave Ste 140 Spring Valley Las Vegas, NV 89118 17:00:00 9:00:00 17:00:00
7:00:00 9:00:00 17:00:00 9:00:00 Accessories Fashion Shopping Shoe Stores Swimwear
Saleyla -115.23699 36.060723 Spring Valley NV 5.0 NULL NULL NULL NULL NULL NULL false false true
ULL NULL NU
ULL NULL FA
ULL NULL NU
F-HBLqU3Kvs891kQ5CjLBQ 420 S Mill Ave Suite 105 Tempe, AZ 85281 23:00:00 10:00:00 21:00:00 10
9:00:00 11:00:00 23:00:00 10:00:00 21:00:00 10:00:00 21:00:00 10:00:00 AC
s & Beauty Supply Beauty & Spas Tempe 15 Lotions & Potions -1
Z 3.5 NULL NULL NULL NULL NULL NULL NULL NULL false NULL NULL NULL NULL NULL NULL NU
ULL NULL NU
ULL NULL NULL false false false false NULL NULL NULL NULL NULL NULL NULL NULL NULL NU
ULL NULL true NULL NULL NULL 2 NULL business
Qd43edttalwCusAF3jULRW 1426 S Tryon St South End Charlotte, NC 28203 18:00:00 10:00:00 18:00:00 10
7:00:00 13:00:00 17:00:00 10:00:00 18:00:00 10:00:00 18:00:00 10:00:00 AC
res Jewelry Charlotte 4 Shu Salon -80.856083 35.216533
C 4.0 NULL NULL NULL NULL NULL NULL NULL false NULL NULL NULL NULL NULL NULL NU
ULL NULL NU
ULL NULL NULL false false false true false NULL NULL NULL NULL NULL NULL NULL NULL NU
ULL NULL true NULL NULL NULL 2 NULL business
goorin-bros-pasadena
ies Hats hats Pasadena NULL Goorin Bros. -118.15143
.5 NULL NU
ULL NULL NU
ULL NULL NU
just-box-alhambra-3
ies Hobby Shops hobbyshops alhambra NULL Just Box
.0 NULL NU
ULL NULL NU
ULL NULL NU

```

2) Creating table for the Reviews Dataset

```

hive> DROP TABLE IF EXISTS reviews;
hive> CREATE EXTERNAL TABLE IF NOT EXISTS reviews (funny int, useful int, cool int, userid string,
reviewid string, stars int, reviewday date, type string, businessid string, comment string)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION
'/user/hmadhan/Project/Reviews' TBLPROPERTIES ('skip.header.line.count'='1');

```

```

hive> DROP TABLE IF EXISTS reviews;
OK
Time taken: 8.893 seconds
hive> CREATE EXTERNAL TABLE IF NOT EXISTS reviews(funny int, useful int, cool int, userid string, rev
iewid string, stars int, reviewday date, type string, businessid string, comment string)
> ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION '/user/hmadhan/Project/Reviews' TBLPROPERT
IES ('skip.header.line.count'='1');
OK
Time taken: 3.489 seconds

```

Run the below command to view the first 10 rows from the table reviews

```

hive> select * from reviews LIMIT 10;

```

```

hive> select * from reviews limit 10;
OK
0      1      0      J0OekBTvoMoiUMsEqzs_Ww  gg1clLMQfz37EBAfFeQNEA  1      2014-05-03      review  wyqw04ZqZyRR6pGk1p
er. Not only is the Las Vegas airport location crooked but so is the corporate office. I had a prepaid reservation booked
the car I informed them that the credit card that I used to pay for the reservation had been cancelled due to fraudulent a
presented to them. I was informed that because I did not have the original card that the reservation is not valid. Even th
months earlier and the replacement card issued by the credit card company for security they would not honor my reservation
a ""walk up"" at double the rate I had already Paid! I told them I was not willing to pay double and they informed me the
as very rude and said he was going to call security if I did not leave the counter while I was trying to reach the corpora
porate office and they are as crooked as this location. They are not willing to refund my money that the charged to my c
ompany can get away with not honoring your pre-paid reservation and then no give you your money back when you refuse to
nt of credit card that goes on this is not an isolated incident.--Instead of having a little compassion they try to bend y
1      4      1      ezzv9rFvT9ETUa_bXdf5wg  SVVZwWC2xCFqRogKYfaGNg  1      2014-05-15      review  wyqw04ZqZyRR6pGk1p
attempting to rent a car from them. ---I made a pre-paid reservation on a debit card. Long story short
0      2      0      bq5ilsRhF7kHqbs0Jce4yg  ODJoaP02eaEjIonryvhnpQ  1      2014-05-15      review  wyqw04ZqZyRR6pGk1p
""pay"" for your car rental but then when you get there you find out that there is an additional 25 dollar per day fee. So
r ""canceling"" your car rental or paying the $25/day fee. It is a complete joke. Do not rent from this car rental place.
0      2      0      d3jZgRWSYZA6u6_Q0CJAA  tVn6ER-048qK9104FgVa9g  1      2014-05-15      review  wyqw04ZqZyRR6pGk1p
ented tons of cars and never had them do this. Don't rent from them unless you want to pay double what you agreed to. Doub
e. No. Go to a legitimate rental place where the prices are the prices."
0      0      0      hc-8ax0bpNPEWAVx8toPFw  dz1amcj8crv1czfKSk_EBA  1      2014-05-19      review  wyqw04ZqZyRR6pGk1p
rent a car without a proof of residency .Guy at the counter asked me to show him an electricity bill that would match the
th me
0      0      0      qMOqG6m9strqoFqeNF91Pg  iSb_ZNs0WPo28RbkUC6Wrg  1      2014-05-22      review  wyqw04ZqZyRR6pGk1p
0      0      0      KVibIbPsixtxYSRIGXR4Ig  NZ4BaTq3wMdldPmVJzZ84Q  1      2014-05-23      review  wyqw04ZqZyRR6pGk1p
I needed to go on a last minute trip to San Diego today due to his mother being sick. We reserved our Van online through
When we got there to pick up our reservation they told us our car type wasnt guaranteed... ok fine. They told us if we
sedona then it would be an extra 10 bucks a day. I said hell no
0      1      1      VELpw7xa0OG1luo5NQdiA  UBOOddKjLaFmMI03EzW3Hg  1      2014-05-26      review  wyqw04ZqZyRR6pGk1p
e hand !!!! bait and switch !!! I was supposed to be charge $49 and when it came down to it I was charged $375!!! Here is
any car company I have ever rented from. As for Marcel
0      2      0      d7_LFbdCABDTwGjYcmwMMQ  H66W3JygOyAVpC0XxwsURA  1      2014-05-31      review  wyqw04ZqZyRR6pGk1p
1hr just to be told I needed a utility bill (I'm local) to be able to rent a car! Bad customer service! If I could put Zer

```

3) To generate the table reviews_analysis that will contain the sentiment analysis data

For sentiment analysis, you need to create a table dictionary, which has polarity to show each word's meaning implied as positive or negative. To do that, run the below command:

```

hive> CREATE EXTERNAL TABLE if not exists dictionary (

```

```

type string, length int, word string, pos string, stemmed string, polarity string)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'
STORED AS TEXTFILE LOCATION '/user/hmadhan/dictionary';

hive> CREATE VIEW IF NOT EXISTS I1 AS
    SELECT reviewid, businessid, words FROM reviews
    LATERAL VIEW EXPLODE(sentences(lower(comment))) dummy AS words;

hive> CREATE VIEW IF NOT EXISTS I2 AS SELECT reviewid, businessid, I2.word,
CASE d.polarity
WHEN 'negative' THEN -1
WHEN 'positive' THEN 1
ELSE 0 END AS polarity
FROM I2 LEFT OUTER JOIN dictionary d ON I2.word = d.word;

hive> CREATE TABLE IF NOT EXISTS review_sentiment_biz
    STORED AS orc AS SELECT reviewid, businessid,
CASE
WHEN sum (polarity) > 0 THEN 'positive'
WHEN sum (polarity) < 0 THEN 'negative'
ELSE 'neutral' end AS sentiment
FROM I3 WHERE reviewid <> '#NAME?' AND reviewid <> '#VALUE!' AND businessid <> '#NAME?'
AND businessid <> '#REF!'
GROUP BY businessid, reviewid;

hive> CREATE TABLE IF NOT EXISTS review_sentiment_rev
    STORED AS orc AS SELECT
reviewid, businessid,
CASE
WHEN sum (polarity) > 0 THEN 'positive'
WHEN sum (polarity) < 0 THEN 'negative'
ELSE 'neutral' end AS sentiment
FROM I3 WHERE reviewid <> '#NAME?' AND reviewid <> '#VALUE!' AND businessid <> '#NAME?'
AND businessid <> '#REF!'
GROUP BY reviewid, businessid;

```

To view the contents of the table review_analysis, use the following command:

```
hive> SELECT * FROM review_analysis LIMIT 10;
```

```

hive> select * from review_analysis limit 10;
OK
0    1    0    J0OekBTvoMoiUMsEqzs_Ww  gq1clLMQfz37EBAfFeQNEA  1    wyqw04ZqZyRR6pGk1pTBrA  positive
1    4    1    eZZv9rFVT9ETUa_bXdf5wg  SVVZwNC2xCfRgOgKYfaGNg  1    wyqw04ZqZyRR6pGk1pTBrA  positive
0    2    0    bq5ilsRhF7kHqbs0Jce4yg  ODJoaP02eaEjIonryvhnpQ  1    wyqw04ZqZyRR6pGk1pTBrA  positive
0    2    0    d3jZgRWSYZA6u6-_QCJAA  tVn6ER-048gk9l04FgVa9g  1    wyqw04ZqZyRR6pGk1pTBrA  neutral
0    0    0    hc-8ax0bpNPEWAVx8toPFw  dz1amcj8crvlozfKSk_EBA  1    wyqw04ZqZyRR6pGk1pTBrA  positive
0    0    0    qMoqG6m9strqoFqeNB91Pg  iSb_ZNs0WPo28RbkUC6Wrg  1    wyqw04ZqZyRR6pGk1pTBrA  neutral
0    0    0    KVibIbPsixtYsRIGXR4Ig  NZ4BaTq3wMdldPmVJZz84Q  1    wyqw04ZqZyRR6pGk1pTBrA  negative
0    1    1    VELpWw7xaOOGLluo5NQdiA  UBOOddKjLaFmMI03EzW3Hg  1    wyqw04ZqZyRR6pGk1pTBrA  negative
0    2    0    d7_LFbdCABDTwGjYcmwMMQ  H66W3JygoAvpc0XxwsURA  1    wyqw04ZqZyRR6pGk1pTBrA  positive
0    2    0    b60qyz1vpleqtrZ-y7h0ug  e_ffa8LdbcSz1gvCybDzRQ  1    wyqw04ZqZyRR6pGk1pTBrA  neutral
Time taken: 0.081 seconds, Fetched: 10 row(s)

```

Exercise 3: Break down the data set into different categories of Local Business

The Local Business dataset is extremely diverse and vast. To ease the task of data analysis, we broadly break down the local businesses into 6 major categories namely Education, Entertainment, Food, Medical, Services and Shopping. These tables are achieved by separating the distinct values in the categories_001 attribute column into suitable and relevant categories.

Create table and populate data for different categories using the below commands:

- **Education:** All business under schools, colleges, classes, learning institutions, library have been put under this category.

```

hive> DROP TABLE IF EXISTS Education;
hive> CREATE EXTERNAL TABLE IF NOT EXISTS Education(
  business_id string, categories_001 String, city String, name String, stars Double, review_count int)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION
'/user/hmadhan/Business/Education';

hive> INSERT OVERWRITE TABLE Education
SELECT business_id, categories_001,city,name ,stars, review_count from biz where
categories_001=='Adult Education' or categories_001=='Art Schools' or categories_001=='Barre Classes'
or categories_001=='Cardio Classes' or categories_001=='Career Counseling' or
categories_001=='Childbirth Education' or categories_001=='Colleges & Universities' or
categories_001=='Cooking Schools' or categories_001=='Cosmetology Schools' or categories_001=='CPR
Classes' or categories_001=='Cycling Classes' or categories_001=='Dance Schools' or
categories_001=='Driving Schools' or categories_001=='Education' or categories_001=='Educational
ServicesElementary Schools' or categories_001=='Elementary Schools' or categories_001=='First Aid

```

```
Classes' or categories_001=='Golf Lessons' or categories_001=='Libraries' or categories_001=='library' or
categories_001=='Massage Schools' or categories_001=='Middle Schools & High Schools' or
categories_001=='Newspapers & Magazines' or categories_001=='Parenting Classes' or
categories_001=='Pet Training' or categories_001=='Pole Dancing Classes' or
categories_001=='Preschools' or categories_001=='Private Tutors' or categories_001=='Swimming
Lessons/Schools' or categories_001=='Tutoring Centers' or categories_001=='university' or
categories_001=='University Housing' or categories_001=='Used Bookstore';
```

To view the contents of the table Education, use the following command:

```
hive> SELECT * FROM Education LIMIT 10;
```

```
hive> select * from Education limit 10;
OK
mMSIJUyvnh0mXQBcj8T-yQ  Adult Education Las Vegas      Nevada School of Massage Therapy    2.5   10
el-monte-rosemead-adult-school-el-monte Adult Education El monte          El Monte-Rosemead Adult School  3.0   NULL
new-horizons-computer-learning-centers-burbank-4  Adult Education burbank New Horizons Computer Learning Centers 5.
0      NULL
baldwin-park-adult-education-computer-and-network-training-baldwin-park Adult Education baldwin park Baldwin Park Adult
Education - Computer and Network Training      3.5   NULL
downey-adult-school-downey-3  Adult Education downey Downey Adult School    3.5   NULL
toastmasters-club-587-downey-2  Adult Education downey Toastmasters Club 587  5.0   NULL
4T6CsZE4BSTM2fCEOJkWmg Art Schools  Montr_al  Au Papier Japonais  4.5   8
6-C75pMHTs1ni6MimMax-g Art Schools  Las Vegas  Social Paintbrush  4.5   11
6q4UCDeSOami_nWKge8z3w Art Schools  Scottsdale Ursula's Piano Lesson Studio 5.0   5
6SomDsOz9QY5TiPwVnu4lw Art Schools  Las Vegas  Berns Music School   3.5   3
Time taken: 0.084 seconds, Fetched: 10 row(s)
```

- **Entertainment:** All business under amusement parks, music places, clubs, zoo etc are put under this category.

```
hive> DROP TABLE IF EXISTS Entertainment;
```

```
hive> CREATE EXTERNAL TABLE IF NOT EXISTS Entertainment( business_id string, categories_001 String,
city String, name String, stars Double, review_count int)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION
'/user/hmadhan/Business/Entertainment';
```

```
hive> INSERT OVERWRITE TABLE Entertainment
SELECT business_id,categories_001,city,name,stars, review_count
FROM biz WHERE categories_001=='Active Life' or categories_001=='Adult' or categories_001=='Adult
Entertainment' or categories_001=='Amusement Parks' or categories_001=='amusement_park' or
categories_001=='Archery' or categories_001=='Art Galleries' or categories_001=='art_gallery' or
```

```
categories_001=='Arts & Entertainment' or categories_001=='Beaches' or categories_001=='Bingo Halls' or categories_001=='Bowling' or categories_001=='bowling_alley' or categories_001=='casino' or categories_001=='Casinos' or categories_001=='Cinema' or categories_001=='Comedy Clubs' or categories_001=='Comic Books' or categories_001=='Country Clubs' or categories_001=='Country Dance Halls' or categories_001=='Dance Clubs' or categories_001=='Day Camps' or categories_001=='Go Karts' or categories_001=='Golf' or categories_001=='Hiking' or categories_001=='Karaoke' or categories_001=='Kids Activities' or categories_001=='Leisure Centers' or categories_001=='Magicians' or categories_001=='movie_theater' or categories_001=='museum' or categories_001=='Museums' or categories_001=='night_club' or categories_001=='Nightlife' or categories_001=='Opera & Ballet' or categories_001=='Paintball' or categories_001=='Recreation Centers' or categories_001=='Resorts' or categories_001=='Scuba Diving' or categories_001=='Social Clubs' or categories_001=='Tennis' or categories_001=='zoo' or categories_001=='Zoos';
```

To view the contents of the table Entertainment, use the following command:

```
hive> SELECT * FROM Entertainment LIMIT 10;
```

```
hive> select * from Entertainment limit 10;
OK
szJVgpRqcISuzdLfUfEw Active Life Scottsdale Fiddlesticks Family Fun Park 3.5 3
0bPPbUg001AWEZ7DR9fHg Active Life Phoenix Lee's ATA Martial Arts 5.0 4
1QQZuf4zZoyFCvXc0o6Vg Active Life Scottsdale Chaparral Off Leash Area 4.5 144
1Z2Zj5eotjc2peYxrGZLw Active Life Peoria Go Sail Arizona 5.0 3
2m5TaARSkFcq3Q1o-YBmw Active Life Goodyear Monroy Training 5.0 3
2uA9B6jgrtrQFaY0FIEwhQ Active Life Huntersville Huntersville Family Fitness & Aquatics 3.5 10
4PneR-GyiymyNlwQ37V-w Active Life Phoenix LA Fitness 2.5 16
8pNViFb5KglecXM4j23iA Active Life Tempe Sumits Yoga Tempe 4.5 5
8tUKyMRMMt1YSlac04eqg Active Life Karlsruhe Schiffsrestaurant Vater Rhein 4.5 4
APW9L14uzlxANoSontL5Q Active Life Henderson Soccer Zone 4.0 6
Time taken: 0.06 seconds, Fetched: 10 row(s)
```

- **Food:** All business under restaurants, bakeries, cafes, pizza places etc are put under this category.

```
hive> DROP TABLE IF EXISTS Food;
```

```
hive> CREATE EXTERNAL TABLE IF NOT EXISTS Food( business_id string, categories_001 String, city String, name String, stars Double, review_count int)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION
'/user/hmadhan/Business/Food';
```

```
hive> INSERT OVERWRITE TABLE Food
```

```
SELECT business_id, categories_001, city, name , stars, review_count
FROM biz WHERE categories_001 == 'Wine Bars' or categories_001 == 'Vietnamese' or categories_001
== 'vegetarian' or categories_001 == 'vegan' or categories_001 == 'Turkish' or categories_001 == 'Thai'
or categories_001 == 'Tex-Mex' or categories_001 == 'Tea Rooms' or categories_001 == 'Tapas/Small
Plates' or categories_001=='Tapas Bars' or categories_001=='Taiwanese' or
categories_001=='Szechuan' or categories_001=='Sushi Bars' or categories_001=='Steakhouses' or
categories_001=='Soup' or categories_001=='Soul Food' or categories_001=='Seafood' or
categories_001=='Sandwiches' or categories_001=='Salad' or categories_001=='Russian' or
categories_001=='Restaurants' or categories_001=='restaurant' or categories_001=='Pubs' or
categories_001=='Pretzels' or categories_001=='Portuguese' or categories_001=='Pizza' or
categories_001=='Pakistani' or categories_001=='Nightlife' or categories_001=='night_club' or
categories_001=='Moroccan' or categories_001=='Mongolian' or categories_001=='Modern European'
or categories_001=='Middle Eastern' or categories_001=='Mexican'
or categories_001=='Mediterranean' or categories_001=='meal_takeaway' or
categories_001=='Live/Raw Food' or categories_001=='liquor_store' or categories_001=='Lebanese' or
categories_001=='Kosher' or categories_001=='Korean' or categories_001=='Kebab' or
categories_001=='Juice Bars & Smoothies' or categories_001=='Japanese' or categories_001=='Italian'
or categories_001=='Irish' or categories_001=='Indian' or categories_001=='Ice Cream & Frozen
Yogurt' or categories_001=='Hot Dogs' or categories_001=='Hawaiian' or categories_001=='Greek' or
categories_001=='German' or categories_001=='Gay Bars' or categories_001=='Gastropubs' or
categories_001=='French' or categories_001=='Food Trucks' or categories_001=='Food Stands' or
categories_001=='Food Court' or categories_001=='Food' or categories_001=='Fast Food' or
categories_001=='Ethiopian' or categories_001=='Donuts' or categories_001=='Diners' or
categories_001=='Dim Sum' or categories_001=='Desserts' or categories_001=='Cuban' or
categories_001=='Creperies' or categories_001=='Comfort Food' or categories_001=='Coffee & Tea' or
categories_001=='Chinese' or categories_001=='Chicken Wings' or categories_001=='Cheesesteaks' or
categories_001=='Cheese Shops' or categories_001=='Champagne Bars' or categories_001=='Caribbean'
or categories_001=='Canadian (New)' or categories_001=='Cajun/Creole' or categories_001=='Cafes' or
categories_001=='cafe' or categories_001=='Burgers' or categories_001=='Buffets'
or categories_001=='British' or categories_001=='Breakfast & Brunch' or categories_001=='Brasseries'
or categories_001=='Bistros' or categories_001=='Belgian' or categories_001=='Beer Garden' or
categories_001=='Bars' or categories_001=='Bangladeshi' or categories_001=='Bakeries' or
```

```
categories_001=='Asian Fusion' or categories_001=='American (Traditional)' or  
categories_001=='American (New)' or categories_001=='African';
```

To view the contents of the table Food, use the following command:

```
hive> SELECT * FROM Food LIMIT 10;
```

```
hive> select * from Food limit 10;  
OK  
67I_MP5CQFuEbX4T44fQw  African Glendale      Zanzibar African Restaurant    2.5    7  
FujnmNjFpU5WQC1Nqd7ahg  African Charlotte    Floyd's Restaurant II       2.5    11  
nWTfbL-ilQiQ9iQwPnViQw  African Phoenix Love Cuisine 4.5     87  
VhhEhn4XlaxmRPkuL3Wq0w  African Edinburgh     Morocco Walima        4.0     4  
vNL25b0LJj38Z2_H1aT3bA  African Montr_al     La Goulette        4.0     6  
revolutionario-north-african-tacos-los-angeles  African Los angeles   Revolutionario: North African Tacos  4.5    NU  
LL  
the-briks-los-angeles-2 African Los angeles   The Briks          4.0     NULL  
4_Qn9WZVwofeaajgjQG2Zw  American (New)    Charlotte    Fahrenheit      3.0     264  
EdsQOKaTuCtZ5y5A2m0rw  American (New)    Phoenix Arby's    3.0     5  
fcEpudbiTB4o9WLJ-egTA  American (New)    Henderson   Grand Caf_      2.5     89  
Time taken: 0.082 seconds, Fetched: 10 row(s)
```

- **Medical:** All business under health related business, dentist, doctors, pharmacy etc are put under this category.

```
hive> DROP TABLE IF EXISTS Medical;
```

```
hive> CREATE EXTERNAL TABLE IF NOT EXISTS Medical ( business_id string, categories_001 String, city String, name String, stars Double, review_count int) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION '/user/hmadhan/Business/Medical';
```

```
hive> INSERT OVERWRITE TABLE Medical
```

```
SELECT business_id, categories_001, city, name, stars, review_count  
FROM biz WHERE categories_001=='Acupuncture' or categories_001=='Allergists' or  
categories_001=='Anesthesiologists' or categories_001=='Cardio Classes' or  
categories_001=='Cardiologists' or categories_001=='Childbirth Education' or  
categories_001=='Cosmetic Dentists' or categories_001=='Cosmetic Surgeons' or  
categories_001=='Counseling & Mental Health' or categories_001=='CPR Classes' or  
categories_001=='dentist' or categories_001=='Dentists' or categories_001=='Dermatologists' or  
categories_001=='Diagnostic Imaging' or categories_001=='Diagnostic Services' or  
categories_001=='Dialysis Clinics' or categories_001=='doctor' or categories_001=='Doctors' or  
categories_001=='Ear Nose & Throat' or categories_001=='Endocrinologists' or  
categories_001=='Endodontists' or categories_001=='Fertility' or categories_001=='First Aid Classes' or
```

```

categories_001=='Gastroenterologist' or categories_001=='General Dentistry' or
categories_001=='health' or categories_001=='Health & Medical' or categories_001=='Health Markets'
or categories_001=='hospital' or categories_001=='Hospitals' or categories_001=='Medical Centers' or
categories_001=='Medical Spas' or categories_001=='Medical Supplies' or categories_001=='Medical
Transportation' or categories_001=='Neurologist' or categories_001=='Nurse Practitioner' or
categories_001=='Nutritionists' or categories_001=='Obstetricians & Gynecologists' or
categories_001=='Oncologist' or categories_001=='Ophthalmologists' or
categories_001=='Optometrists' or categories_001=='Oral Surgeons' or
categories_001=='Orthodontists' or categories_001=='Orthopedists' or categories_001=='Orthotics' or
categories_001=='Osteopathic Physicians' or categories_001=='Pediatric Dentists' or
categories_001=='Pediatricians' or categories_001=='Periodontists' or categories_001=='pharmacy' or
categories_001=='Physical Therapy' or categories_001=='physiotherapist' or
categories_001=='PodiatristsProctologists' or categories_001=='Prosthodontists' or
categories_001=='Psychologists' or categories_001=='Radiologists' or
categories_001=='Rheumatologists' or categories_001=='Speech Therapists' or categories_001=='Spine
Surgeons' or categories_001=='Surgeons' or categories_001=='Urologists' or
categories_001=='Vitamins & Supplements';

```

To view the contents of the table Medical, use the following command:

```
hive> SELECT * FROM Medical LIMIT 10;
```

```

hive> select * from Medical limit 10;
OK
_pqG5sNH7C6EUZs6MqQOvg  Acupuncture    Gilbert Prosperity Health      5.0      4
_VY1Qs19eR0K0FTbACrxBA  Acupuncture    Charlotte      Omni Wellness Center   3.5      3
_x4hVsoORIjmqjMMX_MZRA  Acupuncture    Las Vegas     Setsuko H Sage; OMD       5.0      5
01a4YOjiBbPUBxvgX6F7VQ  Acupuncture    Pittsburgh    Chinese Acupuncture & Herbs Center LLC 5.0      3
08DRbWJGN_s29gXi3qNroA  Acupuncture    Scottsdale   The Anti-Aging Spot      5.0      6
0JXEak05-7FJ7de2_27FvQ  Acupuncture    Charlotte    Charlotte Community Acupuncture 4.0      5
1fE_QMLi4okKcpJRFaRfkq  Acupuncture    Scottsdale   Innova Acupuncture      5.0      3
1MJV-Jq2xzhfkAaiosixbw  Acupuncture    Las Vegas     Wuxin Healing Arts       3.0      3
2B2u6nxHrED_C0s_g-cnsg  Acupuncture    Phoenix Smith Chiropractic Center At Arrowhead 4.5      3
2vmDSfIgDSnVuopjzgE0CQ  Acupuncture    Scottsdale   Victoria Mogilner Total Rejuvenation Center 2.0      4
Time taken: 0.069 seconds, Fetched: 10 row(s)

```

- **Services:** All business under home services, delivery services, repair shops etc are put under this category.

```
hive> DROP TABLE IF EXISTS Services;
```

```
hive> CREATE EXTERNAL TABLE IF NOT EXISTS Services( business_id string, categories_001 String, city String, name String, stars Double, review_count int) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION '/user/hmadhan/Business/Services';
```

```
hive> INSERT OVERWRITE TABLE Services
SELECT business_id, categories_001, city, name, stars, review_count
FROM biz WHERE categories_001=='Air Duct Cleaning' or categories_001=='Aircraft Repairs' or
categories_001=='Appliances & Repair' or categories_001=='Appraisal Services' or
categories_001=='Aquarium Services' or categories_001=='Auto Customization' or
categories_001=='Auto Detailing' or categories_001=='Auto Glass Services' or categories_001=='Bike
Rentals' or categories_001=='Bike Repair/Maintenance' or categories_001=='Blow Dry/Out Services' or
categories_001=='Car Brokers' or categories_001=='Car Rental' or categories_001=='Car Stereo
Installation' or categories_001=='Car Wash' or categories_001=='car_rental' or
categories_001=='car_repair' or categories_001=='car_wash' or categories_001=='Carpet Cleaning' or
categories_001=='Carpet Installation' or categories_001=='Chimney Sweeps' or
categories_001=='Commercial Truck Repair' or categories_001=='Community Service/Non-Profit' or
categories_001=='Countertop Installation' or categories_001=='Couriers & Delivery Services' or
categories_001=='Demolition Services' or categories_001=='Door Sales/Installation' or
categories_001=='Drywall Installation & Repair' or categories_001=='Event Photography' or
categories_001=='Event Planning & Services' or categories_001=='Eyelash Service' or
categories_001=='Farm Equipment Repair' or categories_001=='Financial Advising' or
categories_001=='Financial Services' or categories_001=='Fire Departments' or categories_001=='Fire
Protection Services' or categories_001=='Fireplace Services' or categories_001=='Food Delivery
Services' or categories_001=='Funeral Services & Cemeteries' or categories_001=='Garage Door
Services' or categories_001=='Gas & Service Stations' or categories_001=='Gutter Services' or
categories_001=='Home Cleaning' or categories_001=='Home Services' or categories_001=='Home
Theatre Installation' or categories_001=='Home Window Tinting' or categories_001=='House Sitters' or
categories_001=='Insulation Installation' or categories_001=='IT Services & Computer Repair' or
categories_001=='Jewelry Repair' or categories_001=='Junk Removal & Hauling' or
categories_001=='Keys & Locksmiths' or categories_001=='Legal Services' or categories_001=='Lice
Services' or categories_001=='Local Services' or categories_001=='meal_delivery' or
categories_001=='Mobile Phone Repair' or categories_001=='Mortgage Brokers' or
```

```

categories_001=='Motorcycle Repair' or categories_001=='Musical Instrument Services' or
categories_001=='Party & Event Planning' or categories_001=='Payroll Services' or
categories_001=='Pest Control' or categories_001=='Pet Boarding/Pet Sitting' or categories_001=='Pet
Groomers' or categories_001=='Photography Stores & Services' or categories_001=='Piano Services' or
categories_001=='plumber' or categories_001=='Plumbing' or categories_001=='Pool & Hot Tub
Service' or categories_001=='Printing Services' or categories_001=='Private Investigation' or
categories_001=='Private Tutors' or categories_001=='Professional Services' or categories_001=='Public
Services & Government' or categories_001=='Public Transportation' or categories_001=='Real Estate
Services' or categories_001=='real_estate_agency' or categories_001=='Refinishing Services' or
categories_001=='Security Services' or categories_001=='Shoe Repair' or categories_001=='Shredding
Services' or categories_001=='Solar Installation' or categories_001=='Tax Services' or
categories_001=='Television Service Providers' or categories_001=='Threading Services' or
categories_001=='Translation Services' or categories_001=='Transmission Repair' or
categories_001=='Travel Services' or categories_001=='Tree Services' or categories_001=='Vacation
Rental Agents' or categories_001=='Valet Services' or categories_001=='Watch Repair' or
categories_001=='Water Delivery' or categories_001=='Water Heater Installation/Repair' or
categories_001=='Water Purification Services' or categories_001=='Wedding Planning' or
categories_001=='Window Washing' or categories_001=='Windows Installation' or
categories_001=='Windshield Installation & Repair';

```

To view the contents of the table Services, use the following command:

```
hive> SELECT * FROM Services LIMIT 10;
```

```

hive> select * from Services limit 10;
OK
alhambra-comfort-ac-alhambra-2 Air Duct Cleaning      alhambra      Alhambra Comfort AC    5.0    NULL
rodeo-air-duct-cleaning-beverly-hills-2 Air Duct Cleaning      Beverly hills  Rodeo Air Duct Cleaning 5.0    NULL
aircraft-ground-support-el-monte     Aircraft Repairs      El monte     Aircraft Ground Support 5.0   NULL
repair-twist-of-pasadena-pasadena-4 Appliances & Repair  Pasadena     Repair Twist of Pasadena 5.0    NU
LL
pasadena-appliance-guy-pasadena-2   Appliances & Repair  Pasadena     Pasadena Appliance Guy  5.0    NULL
repairbest-appliance-repair-pasadena-2 Appliances & Repair  Pasadena     Repairbest Appliance Repair 4.5    NU
LL
alexs-appliance-repair-alhambra   Appliances & Repair  alhambra      Alex's Appliance Repair 5.0    NULL
america-technicians-alhambra       Appliances & Repair  alhambra      America Technicians  4.5    NULL
garvey-appliances-el-monte        Appliances & Repair  El monte     Garvey Appliances   4.0    NULL
moms-appliance-repairs-el-monte   Appliances & Repair  El monte     Mom's Appliance Repairs 5.0   NULL
Time taken: 0.031 seconds, Fetched: 10 row(s)

```

➤ **Shopping:** All business shops, stores, mall, shopping centers etc are put under this category.

```
hive> DROP TABLE IF EXISTS Shopping;
```

```
hive> CREATE EXTERNAL TABLE IF NOT EXISTS Shopping( business_id string, categories_001 String, city String, name String, stars Double, review_count int) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION '/user/hmadhan/Business/Shopping';
```

```
hive> INSERT OVERWRITE TABLE Shopping
SELECT business_id, categories_001,city,name ,stars, review_count
FROM biz WHERE categories_001=='Appliances' or categories_001=='Bespoke Clothing' or
categories_001=='bicycle_store' or categories_001=='Body Shops' or categories_001=='book_store' or
categories_001=='Bookstores' or categories_001=='clothing_store' or categories_001=='Costumes' or
categories_001=='Department Stores' or categories_001=='department_store' or
categories_001=='Dive Shops' or categories_001=='electronics_store' or categories_001=='Fabric
Stores' or categories_001=='Flowers & Gifts' or categories_001=='furniture_store' or
categories_001=='Gift Shops' or categories_001=='Guitar Stores' or categories_001=='Head Shops' or
categories_001=='home_goods_store' or categories_001=='Outlet Stores' or categories_001=='Paint
Stores' or categories_001=='Pawn Shops' or categories_001=='Pet Stores' or
categories_001=='pet_store' or categories_001=='Piano Stores' or categories_001=='Pop-up Shops' or
categories_001=='Shoe Stores' or categories_001=='shoe_store' or categories_001=='Shopping' or
categories_001=='Shopping Centers' or categories_001=='shopping_mall' or categories_001=='Souvenir
Stores' or categories_001=='Surf Shop' or categories_001=='Thrift Stores' or categories_001=='Tobacco
Stores' or categories_001=='Toy Stores' or categories_001=='Trophy Shops' or categories_001=='Vape
Stores';
```

To view the contents of the table Shopping, use the following command:

```
hive> SELECT * FROM Shopping LIMIT 10;
```

```

hive> select * from Shopping limit 10;
OK
TgITBcOoarnIcMA82iKzMQ    Appliances      Chandler      The Great Indoors      3.5      10
the-home-depot-alhambra-2    Appliances      alhambra      The Home Depot      3.0      NULL
sears-home-appliance-showroom-alhambra-2    Appliances      alhambra      Sears Home Appliance Showroom  2.5      NU
LL
1234buy-com-el-monte-2    Appliances      El monte      1234buy.com      5.0      NULL
san-yang-pai-co-san-gabriel    Appliances      San Gabriel      San Yang Pai Co 3.0      NULL
bay-cities-kitchens-and-appliances-beverly-hills    Appliances      Beverly hills      Bay Cities Kitchens & Appliances 4
.5      NULL
action-sales-monterey-park-2    Appliances      Monterey park      Action Sales      4.5      NULL
savon-appliances-burbank    Appliances      burbank Savon Appliances      4.5      NULL
warehouse-discount-center-burbank    Appliances      burbank Warehouse Discount Center 4.0      NULL
johnstone-supply-of-baldwin-park-baldwin-park    Appliances      baldwin park      Johnstone Supply of Baldwin Park 3.
0      NULL
Time taken: 0.072 seconds, Fetched: 10 row(s)

```

Exercise 3: Hive QL commands to perform the analysis

- 1) To show the most popular category of business:

The total number of review count for each category on business to compare the popularity of the business categories:

```

select sum(review_count) from Education;
select sum(review_count) from Entertainment;
select sum(review_count) from Food;
select sum(review_count) from Medical;
select sum(review_count) from Services;
select sum(review_count) from Shopping;

```

Put the results in excel for visualization

Category	Review_count
Food	31572
Education	527
Entertainment	5232
Medical	4224
Services	8046
Shopping	4682

- 2) To show the cities which have been rated five stars in the different categories:

```

EDUCATION select business_id, city, name, stars, review_count from
Education where stars=5 group by business_id, city, name, stars,
review_count order by review_count desc limit 10;

```

```

mMSIJuYvnh0mXQBCj8T-yQ Adult Education Las Vegas Nevada School of Massage Therapy 10 2.5
el-monte-rosemead-adult-school-el-monte Adult Education El monte El Monte-Rosemead Adult School NULL 3.0
new-horizons-computer-learning-centers-burbank-4 Adult Education burbank New Horizons Computer Learning Centers NU
LL 5.0
baldwin-park-adult-education-computer-and-network-training-baldwin-park Adult Education baldwin park Baldwin Park Adult
Education - Computer and Network Training NULL 3.5
downey-adult-school-downey-3 Adult Education downey Downy Adult School NULL 3.5
toastmasters-club-587-downey-2 Adult Education downey Toastmasters Club 587 NULL 5.0
4T6CsZE4BSTM2FCEOJkWMg Art Schools Montr_al Au Papier Japonais 8 4.5
6-C75pMHTs1ni6MinMax-g Art Schools Las Vegas Social Paintbrush 11 4.5
6q4UCDeSOami_nWKge8z3w Art Schools Scottsdale Ursula's Piano Lesson Studio 5 5.0
6SomDs0z9QYsTiPwVhu4lw Art Schools Las Vegas Berns Music School 3 3.5
Time taken: 1.678 seconds, Fetched: 10 row(s)

```

ENTERTAINMENT select business_id, city, name, stars, review_count from Entertainment where stars=5 group by business_id, city, name, stars, review_count order by review_count desc limit 10;

```

_szJVgpRgcISuzdLfUfEW Active Life Scottsdale Fiddlesticks Family Fun Park 3 3.5
_0bPPBuG001AwEZ7DR9fHg Active Life Phoenix Lee's ATA Martial Arts 4 5.0
_1QQzuf4zZoyFCvXc0o6Vg Active Life Scottsdale Chaparral Off Leash Area 144 4.5
_122zj5eoJc2peYxrGzLw Active Life Peoria Go Sail Arizona 3 5.0
_2m5TaARSkFcq3Q1O-YBmw Active Life Goodyear Monroy Training 3 5.0
_2uA9B6jgtrQFaY0FIEwhQ Active Life Huntersville Huntersville Family Fitness & Aquatics 10 3.5
_4PneR-GyiymyNlwQ37V-w Active Life Phoenix LA Fitness 16 2.5
_8pNViFB5KglecXM4j23iA Active Life Tempe Sumits Yoga Tempe 5 4.5
_8tUKyMRMMT1yslac04egg Active Life Karlsruhe Schiffsrestaurant Vater Rhein 4 4.5
_APW9L14uzlxANosOntl5Q Active Life Henderson Soccer Zone 6 4.0
Time taken: 3.585 seconds, Fetched: 10 row(s)

```

FOOD AND DRINKS select business_id, city, name, stars, review_count from Food where stars=5 group by business_id, city, name, stars, review_count order by review_count desc limit 10;

```

671 MPG5CQFuEbX4T44fQw African Glendale Zanzibar African Restaurant 2.5 7
FujnmNjFpU5WQC1Nqd7ahg African Charlotte Floyd's Restaurant II 2.5 11
mWTFbL-i1QiQ9iQwPnViQw African Phoenix Love Cuisine 4.5 87
VnhEhn4XlaxmRPkuL3Wq0W African Edinburgh Morocco Walima 4.0 4
vNL25b0Lj38Z2_H1aT3bA African Montr_al La Goulette 4.0 6
revolutionario-north-african-tacos-los-angeles African Los angeles Revolutionario: North African Tacos 4.5 NU
LL
the-briks-los-angeles-2 African Los angeles The Briks 4.0 NULL
_4_Qn9WZvwofeaQjgQG2Zw American (New) Charlotte Fahrenheit 3.0 264
_EdSqokaTuCTz5y5A2mOrw American (New) Phoenix Arby's 3.0 5
_fcEpudbiTB4o9wLJ-egTA American (New) Henderson Grand Caf_ 2.5 89
Time taken: 26.683 seconds, Fetched: 10 row(s)

```

MEDICAL select business_id, city, name, stars, review_count from Medical where stars=5 group by business_id, city, name, stars, review_count order by review_count desc limit 10;

```

_pgg5sNH7C6EUzs6MqQ0vg Acupuncture Gilbert Prosperity Health 4 5.0
_VYlqs19eR0KOFTbACrxBA Acupuncture Charlotte Omni Wellness Center 3 3.5
_x4hvsoORIJmcjmMx_MZRA Acupuncture Las Vegas Setsuko H Sage: OMD 5 5.0
01a4YojiBbPUbxvgX6F7VQ Acupuncture Pittsburgh Chinese Acupuncture & Herbs Center LLC 3 5.0
08DRbWJGN_s29gXi3qNroA Acupuncture Scottsdale The Anti-Aging Spot 6 5.0
0JXEak05-7FJ7de2_27FvQ Acupuncture Charlotte Charlotte Community Acupuncture 5 4.0
1fe_QMLI4okKcpJRFFkg Acupuncture Scottsdale Innova Acupuncture 3 5.0
1MVJ-Jq2xzhfkAaiOSixbw Acupuncture Las Vegas Wuxin Healing Arts 3 3.0
2B2u6nxHrED_C0s_g-cnsg Acupuncture Phoenix Smith Chiropractic Center At Arrowhead 3 4.5
2vmDsfIgDSnVuOpjzgE0CQ Acupuncture Scottsdale Victoria Mogilner Total Rejuvenation Center 4 2.0
Time taken: 4.811 seconds, Fetched: 10 row(s)

```

SERVICES select business_id, city, name, stars, review_count from Services where stars=5 group by business_id, city, name, stars,

```
review_count order by review_count desc limit 12;
```

alhambra-comfort-ac-alhambra-2	Air Duct Cleaning	alhambra	Alhambra Comfort AC	NULL	5.0
rodeo-air-duct-cleaning-beverly-hills-2	Air Duct Cleaning	Beverly hills	Rodeo Air Duct Cleaning	NULL	5.0
aircraft-ground-support-el-monte	Aircraft Repairs	El monte	Aircraft Ground Support	NULL	5.0
repair-twist-of-pasadena-pasadena-4	Appliances & Repair	Pasadena	Repair Twist of Pasadena	NULL	5
0					
pasadena-appliance-guy-pasadena-2	Appliances & Repair	Pasadena	Pasadena Appliance Guy	NULL	5.0
repairbest-appliance-repair-pasadena-2	Appliances & Repair	Pasadena	Repairbest Appliance Repair	NULL	4
5					
alexs-appliance-repair-alhambra	Appliances & Repair	alhambra	Alex's Appliance Repair	NULL	5.0
america-technicians-alhambra	Appliances & Repair	alhambra	America Technicians	NULL	4.5
garvey-appliances-el-monte	Appliances & Repair	El monte	Garvey Appliances	NULL	4.0
moms-appliance-repairs-el-monte	Appliances & Repair	El monte	Mom's Appliance Repairs	NULL	5.0
Time taken: 20.324 seconds, Fetched: 10 row(s)					

SHOPPING select business_id, city, name, stars, review_count from Shopping where stars=5 group by business_id, city, name, stars, review_count order by review_count desc limit 10;

TgITBc0oarnIcMA82iKzMQ	Appliances	Chandler	The Great Indoors	10	3.5
the-home-depot-alhambra-2	Appliances	alhambra	The Home Depot	NULL	3.0
sears-home-appliance-showroom-alhambra-2	Appliances	alhambra	Sears Home Appliance Showroom	NULL	2.
5					
1234buy-com-el-monte-2	Appliances	El monte	1234buy.com	NULL	5.0
san-yang-pai-co-san-gabriel	Appliances	San gabriel	San Yang Pai Co	NULL	3.0
bay-cities-kitchens-and-appliances-beverly-hills	Appliances	Beverly hills	Bay Cities Kitchens & Appliances	NULL	4.5
ULL					
action-sales-monterey-park-2	Appliances	Monterey park	Action Sales	NULL	4.5
savon-appliances-burbank	Appliances	burbank	Savon Appliances	NULL	4.5
warehouse-discount-center-burbank	Appliances	burbank	Warehouse Discount Center	NULL	4.0
johnstone-supply-of-baldwin-park-baldwin-park	Appliances	baldwin park	Johnstone Supply of Baldwin Park	NULL	3.0
LL					
Time taken: 3.88 seconds, Fetched: 10 row(s)					

3) Analysis of the sub categories of shopping

Execute the query to get the list of all the sub categories of shopping and export the result from Ambari.

```
select business_id, categories_001, city, name, review_count,
stars from Shopping Join biz on Shopping.business_id =
biz.business_id;
```

Queried data in the excel.

BusinessId	Category	City	Name	ReviewCount	stars
vJtLcSMINgvH11_rL1K2Q	Shopping	Anjou	Costco Anjou	4	3.5
HMDymN9TbFUln9AkqHMpVw	Department Stores	Anthem	Polo Ralph Lauren Factory Store	6	3
_JzjW218jlvL6C14yMDIig	Shopping	Anthem	Cvs	3	4
IrZZOcsO8AvoRzwhrcdazA	Shopping	Anthem	Guess	4	4
L1DLgTwAJUNulUJYd9ubqqQ	Shopping	Anthem	Walgreens	5	3
LHwSUFdk1LZzbhSNYwoslw	Shopping	Anthem	Cvs	3	3.5
X-KZvzTme2k5dTQuMd-jxA	Shopping	Anthem	Arizona Eye Care	8	1
XJv7dUb5fMRaZAWVZjhtw	Shopping	Anthem	Outlets At Anthem	78	3.5
bFmzWmaU6hBn4J0gzkjQGw	Tobacco Shops	Anthem	Havana Cigars	11	4
qIPsYQTYaraa48icOUaYGQ	Flowers & Gifts	Apache Junction	Apache Junction Flowers	5	4
9MDaY9VHY1qsCjteoXr9w	Shopping	Apache Junction	Nationwide Vision Centers	3	1
dEeAuK1lhY16RrNt7wjOA	Shopping	Apache Junction	Cvs Pharmacy	5	2
fV7EOjVIFORUoncP1ltVgw	Shopping	Apache Junction	Vapor Nation Az	4	4.5
SZj9k2jVjHUtOSInSsYw3Q	Shopping	Apache Junction	Vapor Cigs Solutions	7	3.5
uc5Slepmspu_1fL-9v1PNA	Shopping	Apache Junction	Walgreens	5	4
v-Us2HMh3w4nBg8aODKY5Q	Shopping	Apache Junction	Blue Nugget	3	3.5
ZuTdf1h4HwgCx0ermgDxtw	Shopping	Apache Junction	Superstition Harley-Davidson	6	3.5
nef2ci4GArhckz7NSU_dRQ	Thrift Stores	Apache Junction	St Vincent De Paul	5	2.5
BWYQQPXmy90Vr1D_wahfmQ	Department Stores	Avondale	Dollar Tree	5	2
FIQz9u8Cy7aTXfMQugUYvA	Department Stores	Avondale	Costco	49	3.5
UPIfEvCACAhZjhKs43AG9g	Flowers & Gifts	Avondale	Petal Pusher	3	5
_qVmvmMt7m9s3vpusYP2y6Q	Shopping	Avondale	Edible Arrangements	9	2
3MyLbgCt6AwF4z-KwoNXVw	Shopping	Avondale	T-Mobile	3	1
7o39-oQa08c1ojMjZLbQ	Shopping	Avondale	Walgreens	5	3
9edfbGejEl0OctuxArLyA	Shopping	Avondale	Fry's & Fry's Marketplace	11	2.5
B6Q25MB9iiPvhGjBztPHLg	Shopping	Avondale	Xpress Care Pharmacy	7	4.5
fa50r2YZnK6UG2Qbp508HA	Shopping	Avondale	Spirit Halloween Store	3	2
facDeMpy1JbKXUzoP7V2cQ	Shopping	Avondale	Verizon	15	3
FTV78oT5bgaH1YSEZVEzA	Shopping	Avondale	Hobby Lobby	28	4
na1p1WO2Fdv83zA1pM6JTg	Shopping	Avondale	T-Mobile	7	4

4) To find the maximum number of business reviewed by a reviewer:

We count the number of business reviewed by every reviewer and sort it in descending order.

```
select userid, count(businessid) cb from review_analysis group by
userid order by cb desc limit 10;
```

OKsj1ATHNWua2Pr4HStQ 225

4TkB_iDShmg41Y_QW9nw 200

4fx3LBeXoE88gDTK6TKQ 160

65q1FpAL_UQtVZ2PTGew 150

9jRaeY1xK-219r9fVQWA 135

03UjydLhAeSH1TJn_6nYA 130

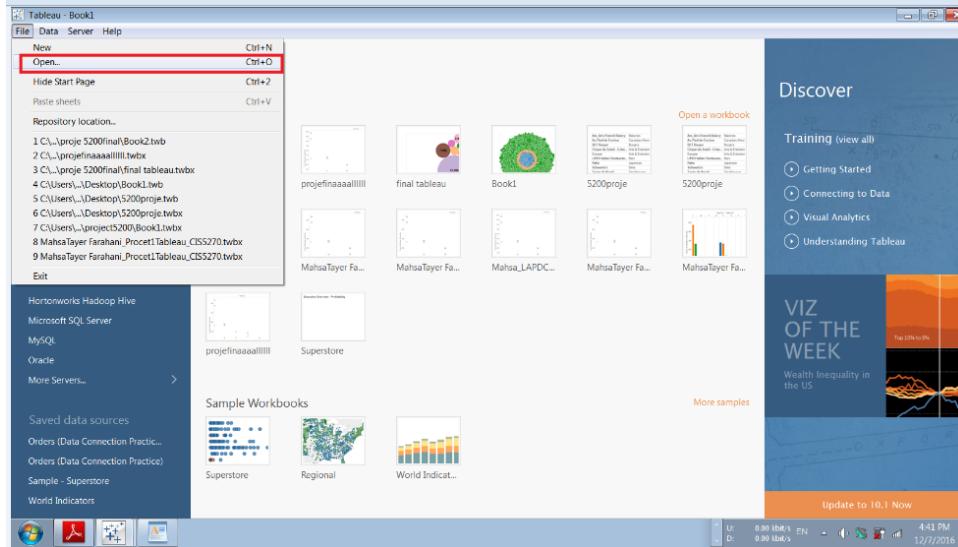
05nNg3337uQFUwFFPH36Q 120

0DRa69RLdSSNrc57P_5ew 100

Exercise 5: Visualization

REVIEW COUNT FOR BASED ON EACH CATEGORY

- 1) Open Tableau. Open excel sheet (File > Open, select the excel sheet).



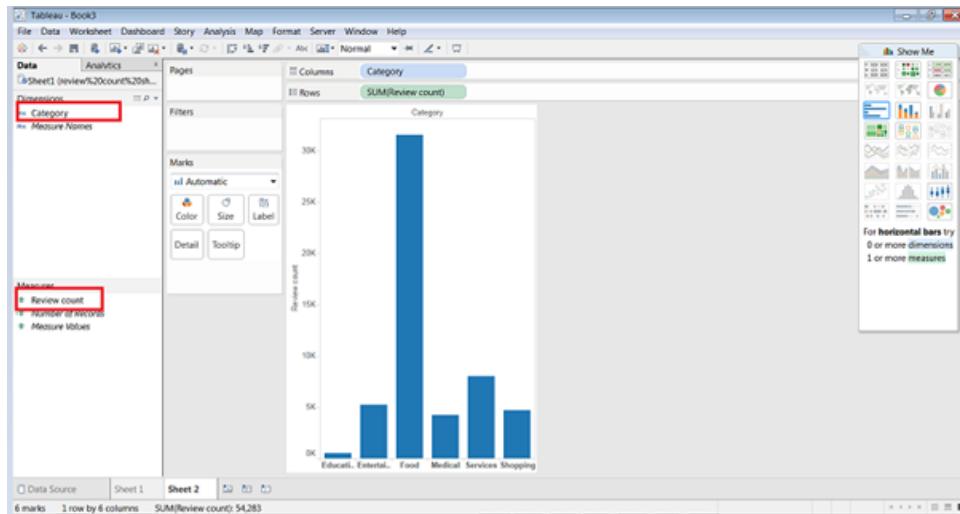
- 2) Select the sheet in the workbook with the results of Analysis 1. Details of the sheet can be seen at the bottom of the window. Select "Go to Worksheet".

A screenshot of the Tableau software showing a worksheet titled 'Sheet1 (review%20count%20sheet)'. The worksheet displays a data table with columns 'Category' and 'Review count'. The data is as follows:

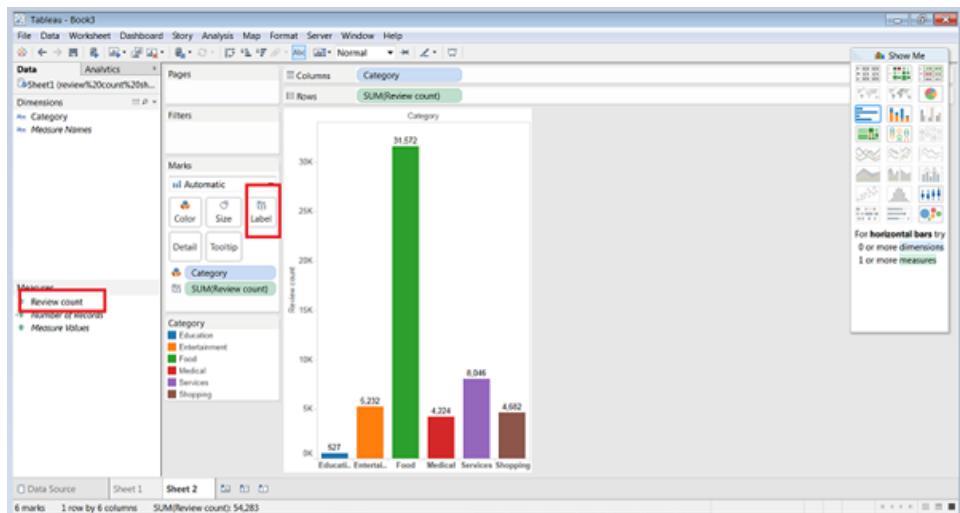
Category	Review count
Food	31,572
Education	527
Entertainment	5,232
Medical	4,224
Services	8,946
Shopping	4,682

At the bottom of the worksheet, there is a toolbar with icons for 'Go to Worksheet', 'Sheet 1', and other navigation options. The 'Go to Worksheet' icon is highlighted with a red box.

- 3) Drag Category from Dimension into the Column and Review Count from Measures into the Row.

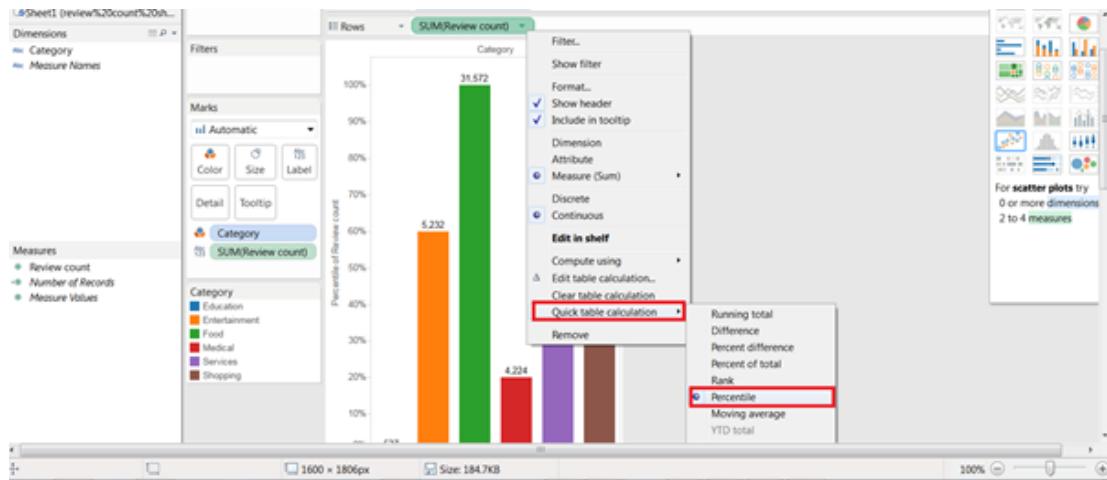


- 4) Drag category into Color of Marks box to assign different colors to different categories. Drag Review Count onto Label of mark box in order to show the number of reviews for each category.
- As we can see that food has the highest number of reviews and education has the lowest number of reviews.

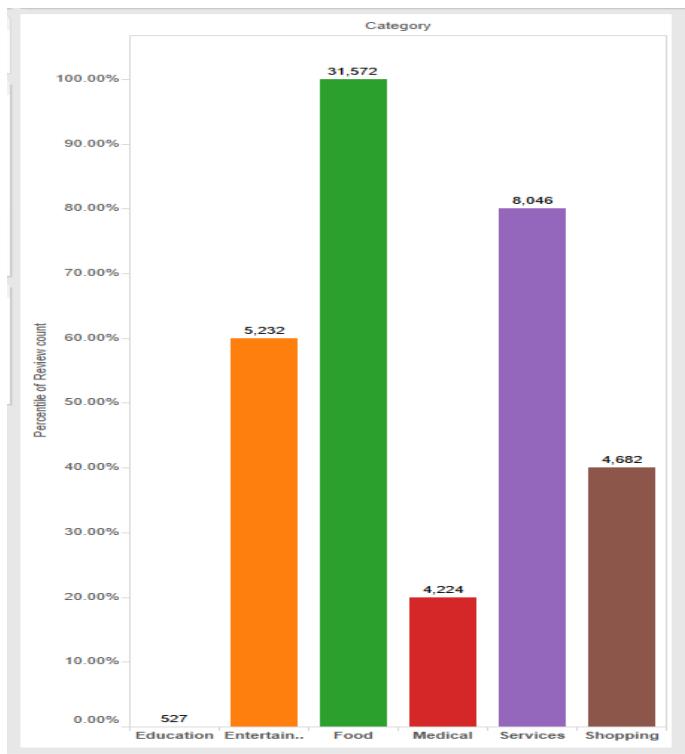


- 5) In order to show the number of reviews for each category in percentage. Right Click the review count in row and from the quick table calculation select percentage. As shown in this figure,

percentage appears on the Y axis.

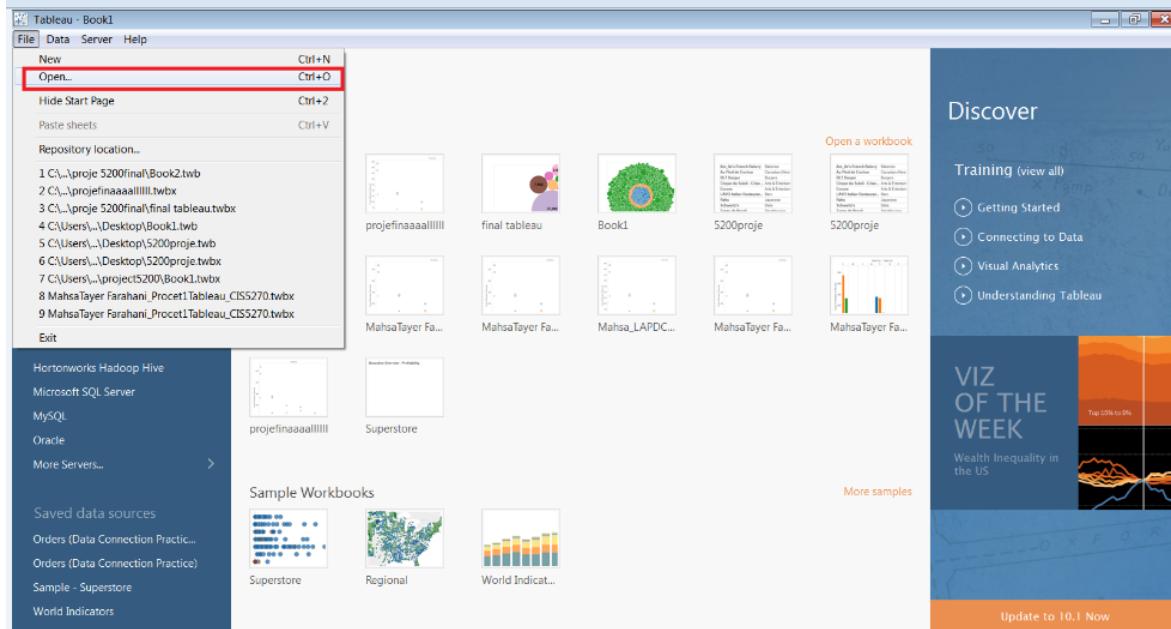


Final Visualization



TOP BUSINESSES IN THE SIX CATEGORIES

- 1) In order to visualize the dataset in Tableau open the excel sheet in tableau. From file tab - choose open and browse the excel sheet.

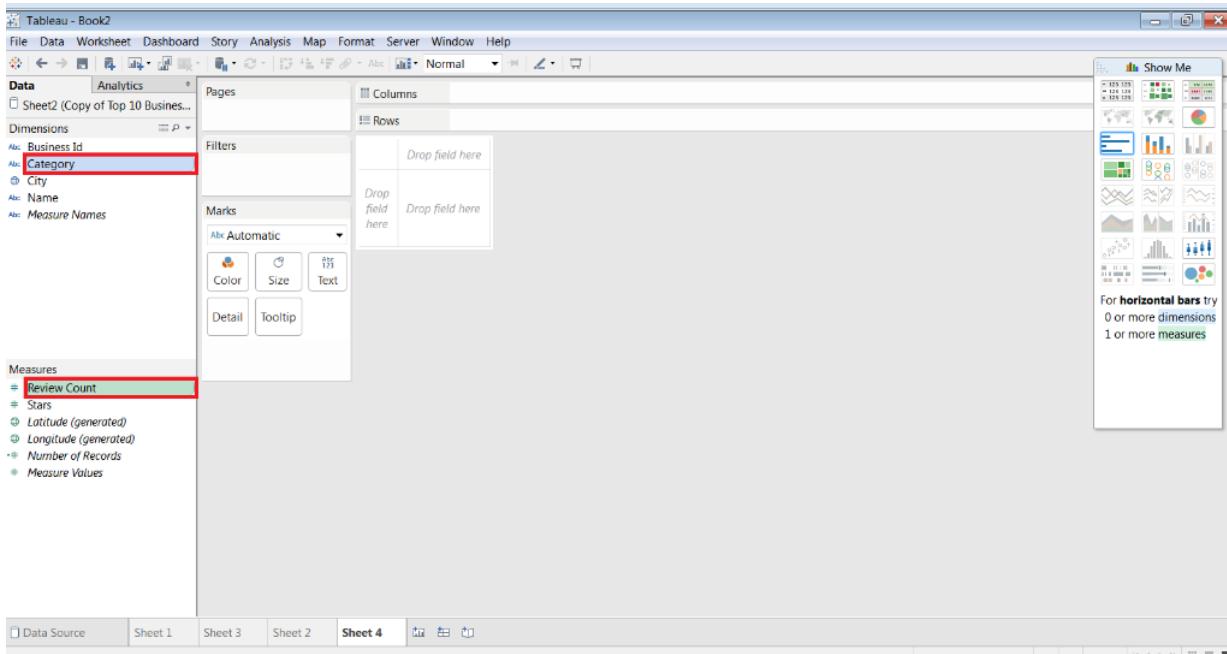


- 2) Select the sheet in the workbook with the results of Analysis 2. Details of the sheet can be seen at the bottom of the window. Select “Go to Worksheet”.

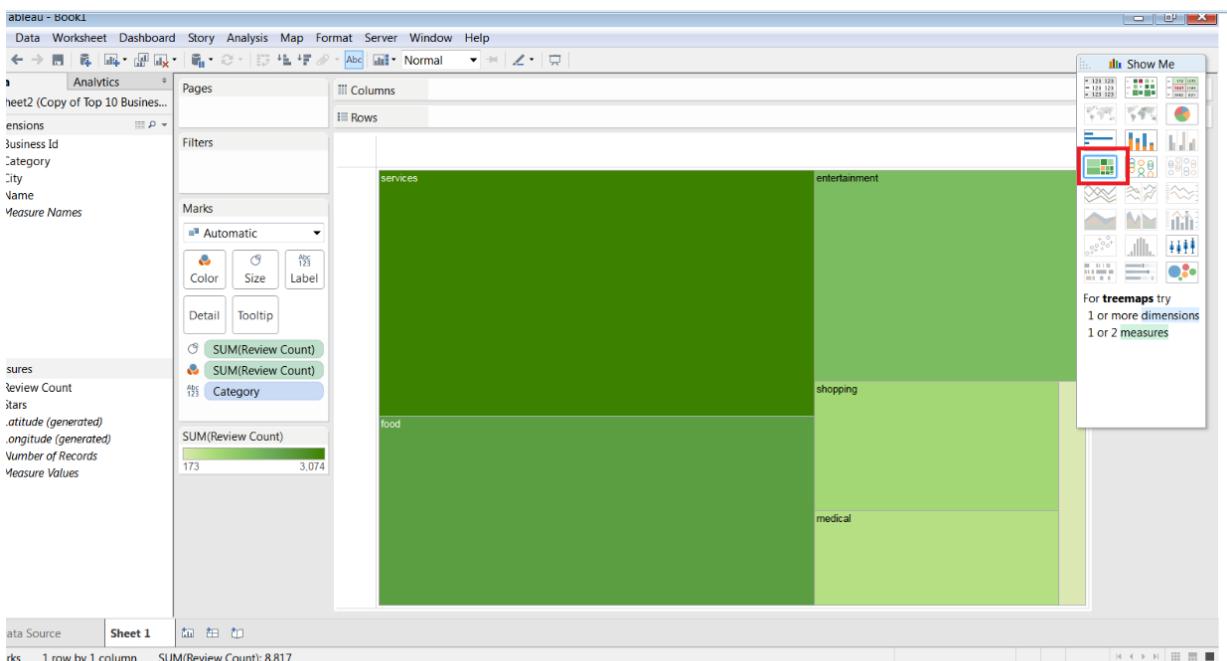
The screenshot shows the Tableau data source editor. On the left, there is a 'Sheets' pane with a search bar and a list of sheets: 'Sheet2', 'New Union', and 'Sheet1'. The 'New Union' sheet is currently selected. The main area displays a table with the following columns: Business Id, City, Name, Review Count, Stars, and Category. At the bottom of the table view, there is a 'Go to Worksheet' button, which is highlighted with a red box. The bottom navigation bar includes buttons for 'Data Source', 'Sheet 1', and other worksheet navigation.

Business Id	City	Name	Review Count	Stars	Category
NuJcBRt63p2OH...	Las Vegas	Art of Flavors	359	5	food
JXUX_oiCrHm6b1...	North Las Veg...	Poke Express	315	5	food
VqU4PiURJjgFS1...	Las Vegas	Brew Tea Bar	306	5	food
8HQ8clouGgee9...	Gilbert	Frost Gelato	260	5	food
PXFEDPxwSfkKoh...	Las Vegas	Dutch Bros. Coffee	241	5	food
DKBtEVkLjmT25...	Phoenix	Ten Handcrafted ...	232	5	food
kdPxx4mVjaDNj...	Las Vegas	J Karaoke Bar	192	5	food
zHFQpbngzb3M0...	Montral	Kem CoBa	156	5	food
84NQRcMC0ldIC8...	Mesa	Gelato Dolce Vita	149	5	food

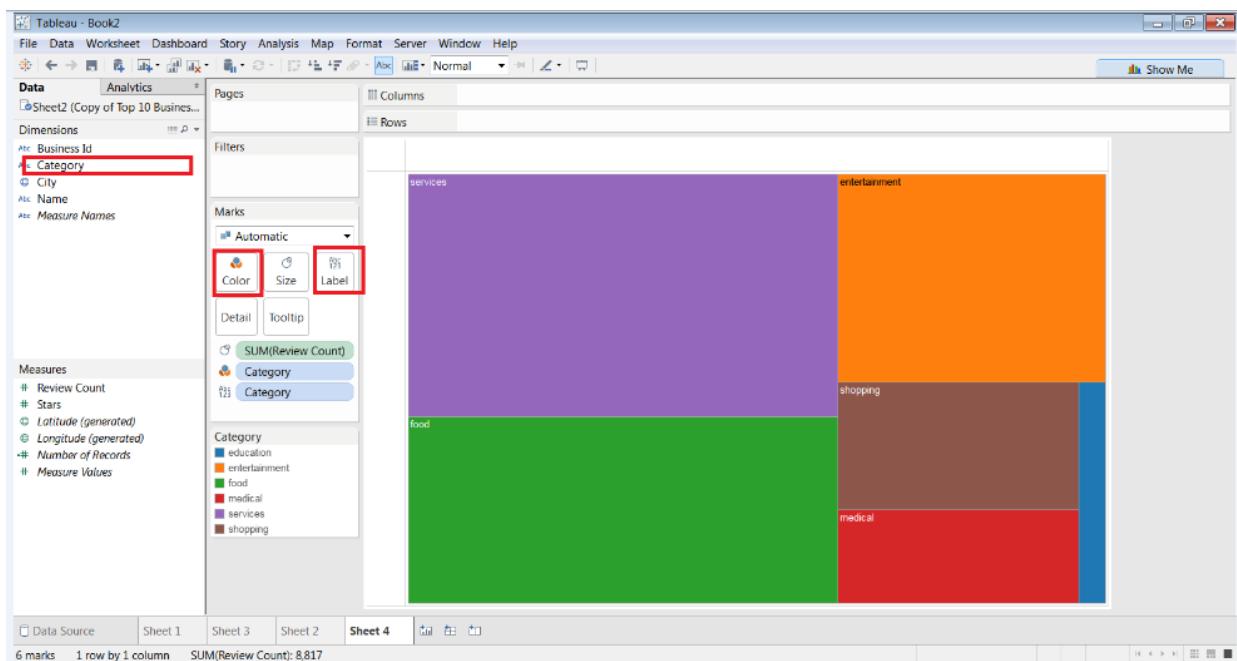
- 3) Select Category from Dimensions and drag to columns and Review Count from measure and drag it to rows.



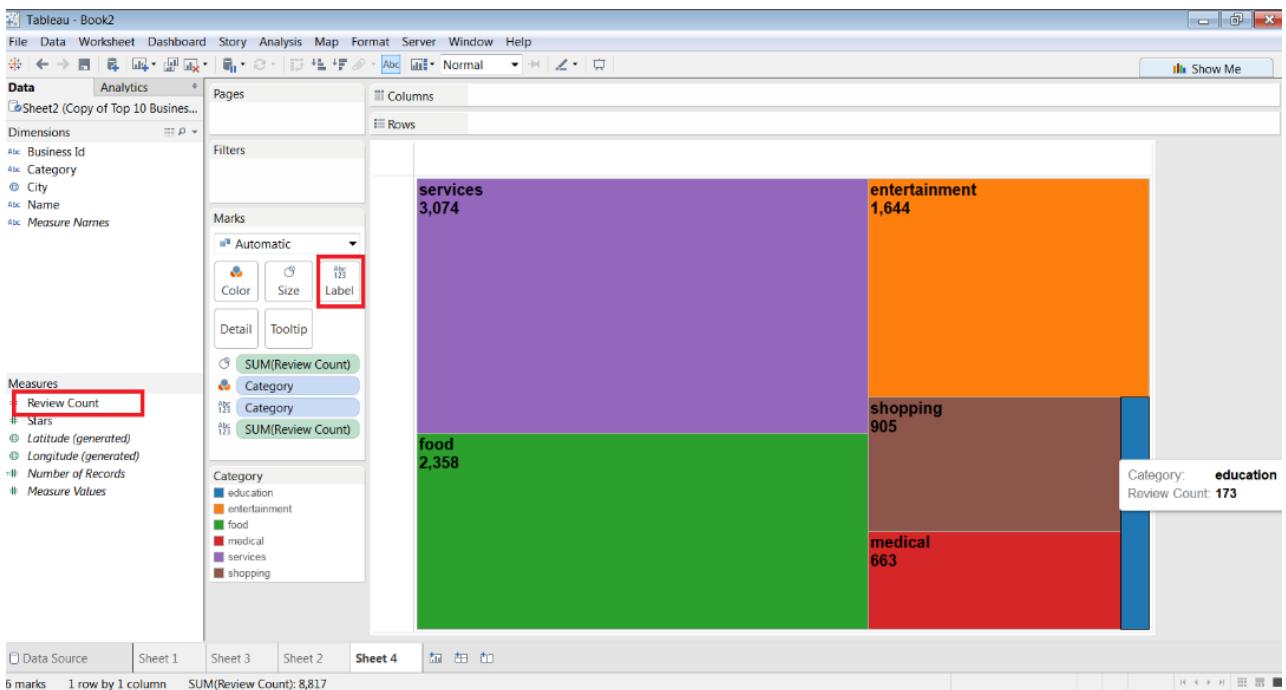
4) Select the tree map from the Show Me on the top right of the window.



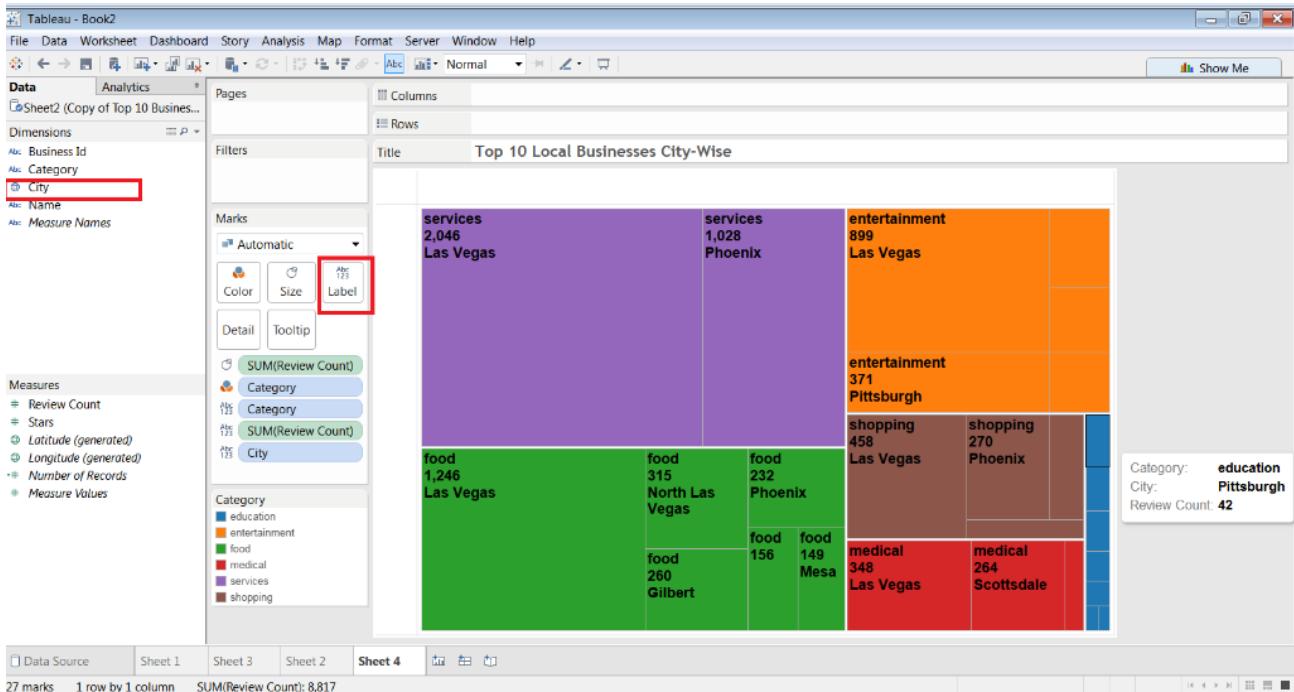
5) Drag Category to color and to the labels. Thus we can see there are different categories with different colors for education, entertainment, food, medical, services and shopping.



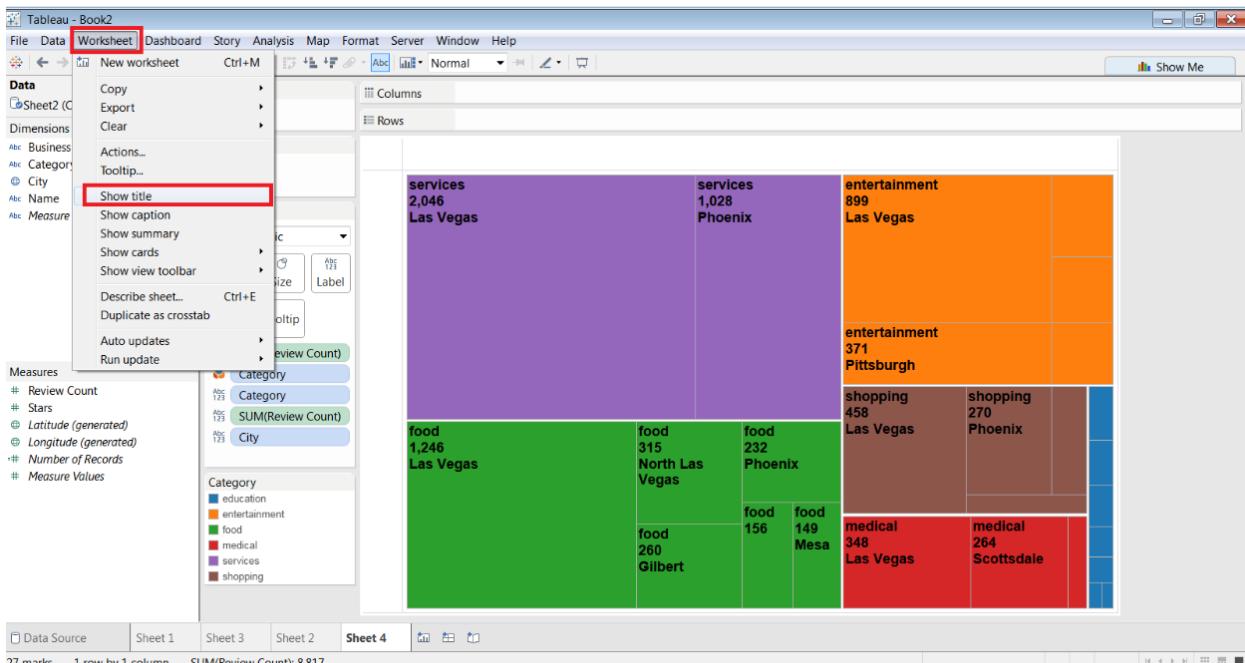
6) Drag reviews count to labels in order to analyze the number of reviews for each category



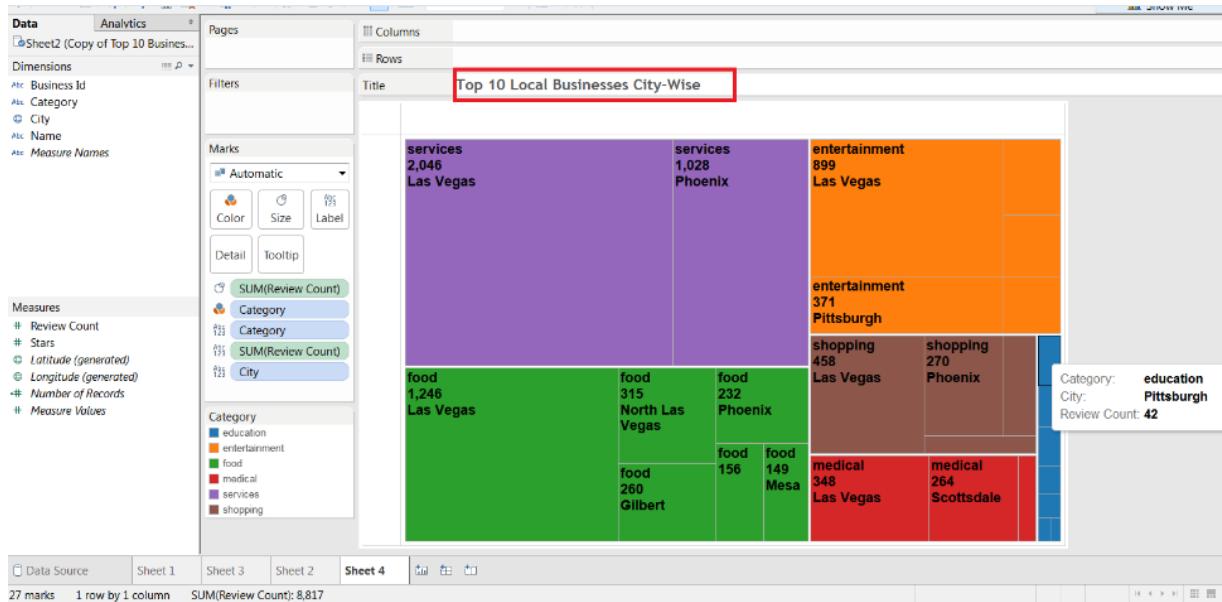
7) Drag city to label as we want to know the number of reviews for each business category in various city. The Tree Map depicts Las Vegas has the highest number of reviews in almost every category .



- 8) In order to give a title to our visualization go to worksheet tab and select show title to write the title for the insight.

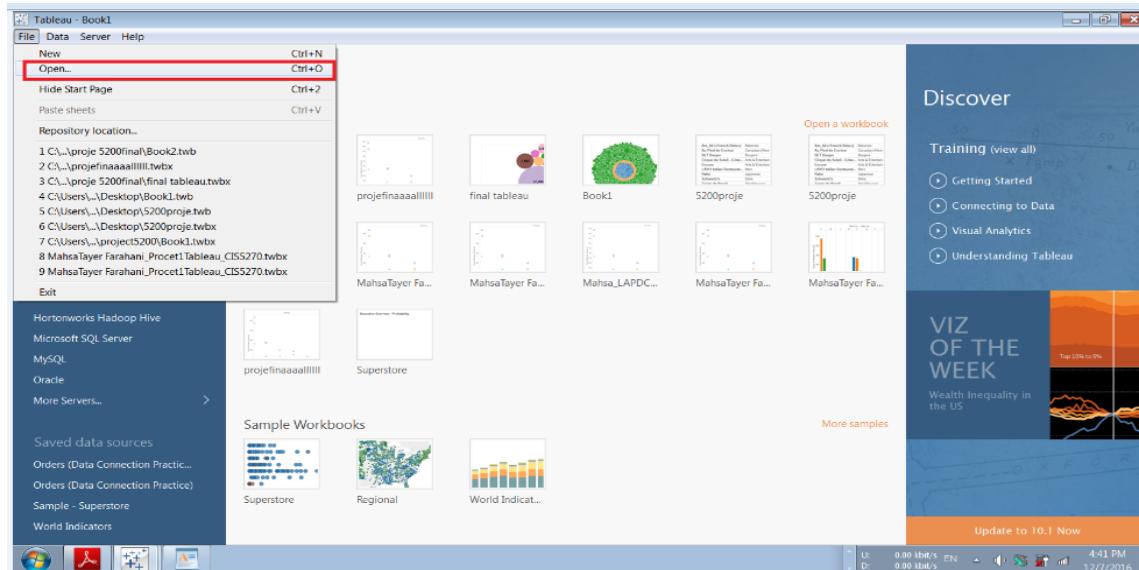


We can now see the title being saved in the title tab. Save the file after completing the visualization.



REVIEW COUNT FOR SUBCATEGORY FOR SHOPPING

1) Open Tableau. Open excel sheet (File > Open, select the excel sheet).



2) Select the Sheet 1 in the Shopping excel with the results of Analysis 3. Details of the sheet can be seen at the bottom of the window. Select “Go to Worksheet”.

The screenshot shows the Tableau interface with the following details:

- Top Bar:** Tableau - Book3, File, Data, Server, Window, Help.
- Connections:** Connected to Excel.
- Panels:**
 - Data Sources:** Shows 'Sheet1.xlsx' with a red box around it.
 - Sheets:** Shows 'Sheet1' with a red box around the 'Go to Worksheet' button.
 - Worksheet Area:** Displays a table with columns: Sheet1, Business Id, Category, City, Name, Review Count, and Stars. The table contains 10 rows of data.
 - Bottom:** Data Source, Go to Worksheet, Sheet 1, and other buttons.

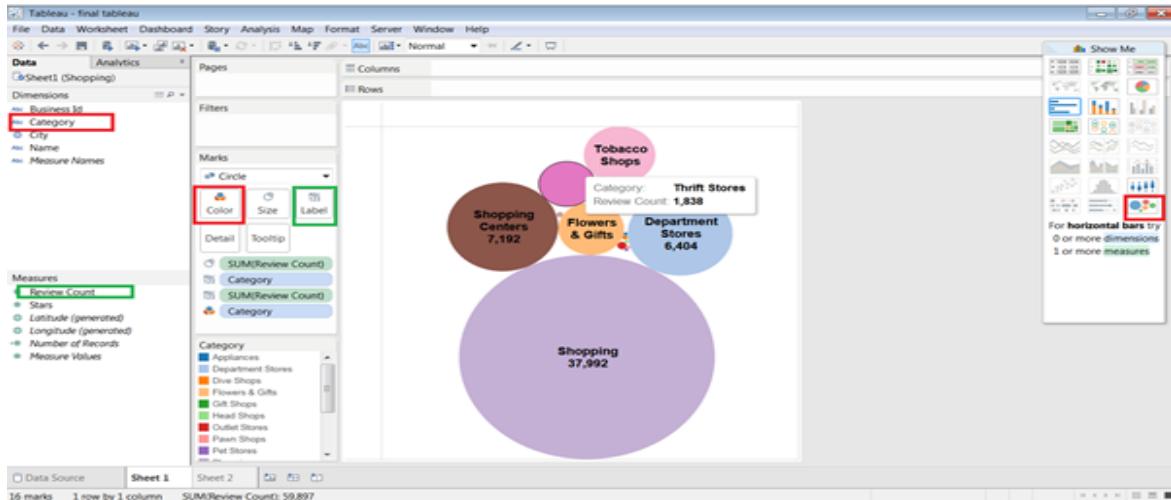
3) Drag Category to color in the marks box and Review Count to the label.

The screenshot shows the Tableau interface with the following details:

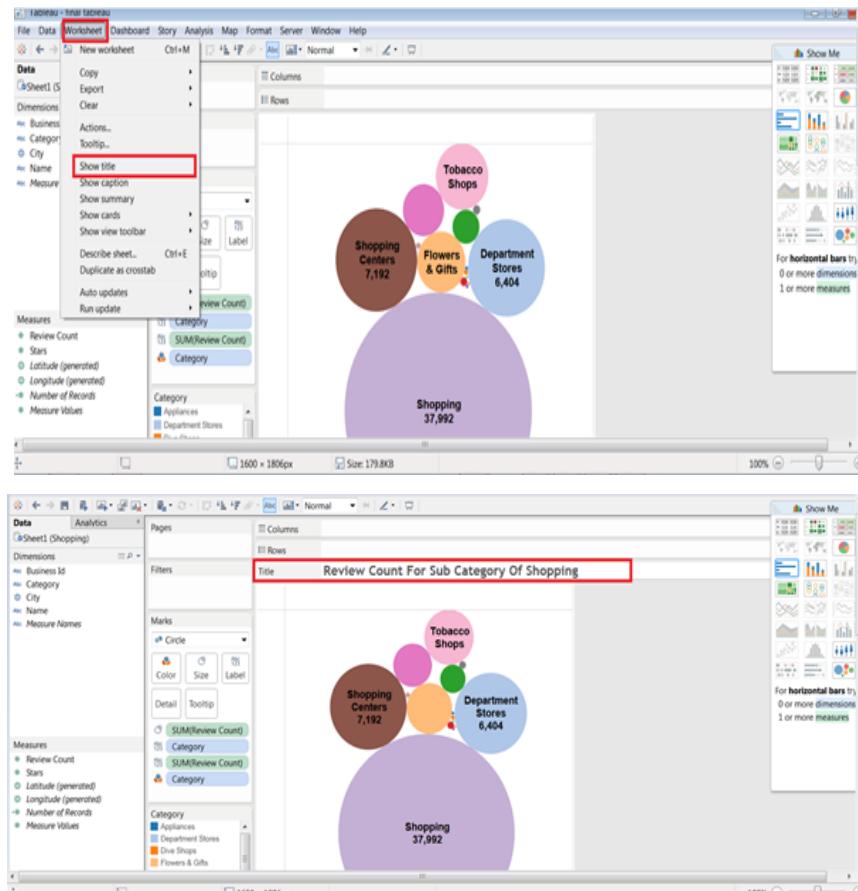
- Dimensions:** Business Id, Category (highlighted with a red box), City, Name, Measure Names.
- Measures:** Review Count (highlighted with a green box), Stars, Latitude (generated), Longitude (generated), Number of Records, Measure Values.
- Marks:** Circle, Color (selected for Category), Size, Label (selected for Review Count).
- Labels:** Category (list of categories: Appliances, Department Stores, Drive Shops, Flowers & Gifts, Gift Shops, Head Shops, Outlet Stores, Pawn Shops, Pet Stores, Shopping, Shopping Centers, Surf Shop).
- Right Panel:** Shows a preview of horizontal bars with the text: "For horizontal bars to 0 or more dimension 1 or more measures".

4) Select the bubble chart in show me tab. The bubbles represent various sub categories of shopping

with their number of reviews.

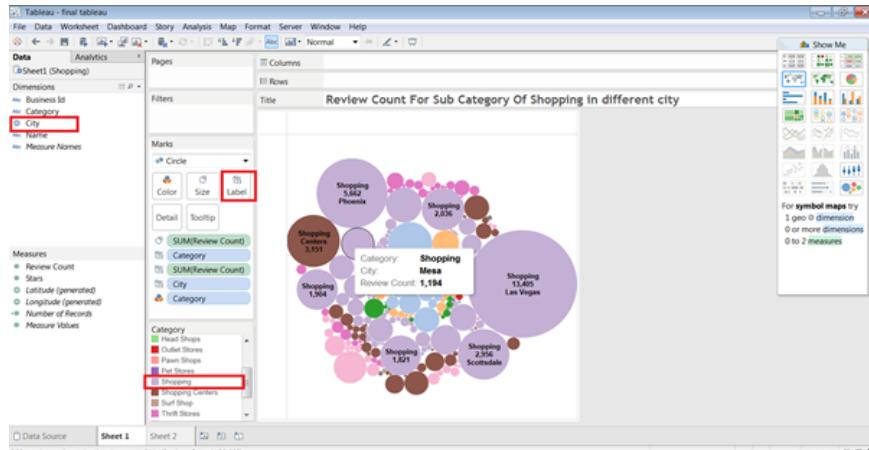


5) Add title for this visualization. Go to worksheet tab in the top left side and select show title.

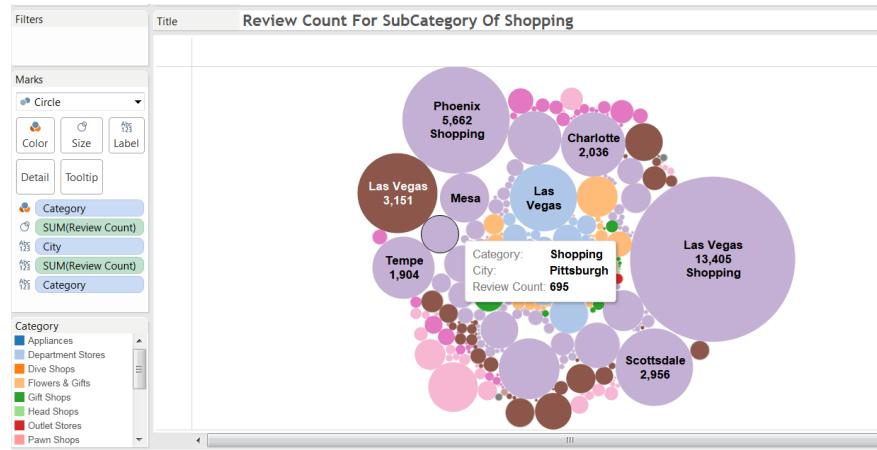


6) To analyze the review count for sub category of shopping in different cities, we have to drag City to the labels as it shows in the below screenshot. As we can see in the in the category list, shopping in

purple color has the highest number of reviews and Las Vegas has the highest number of reviews for shopping, compared to other cities.

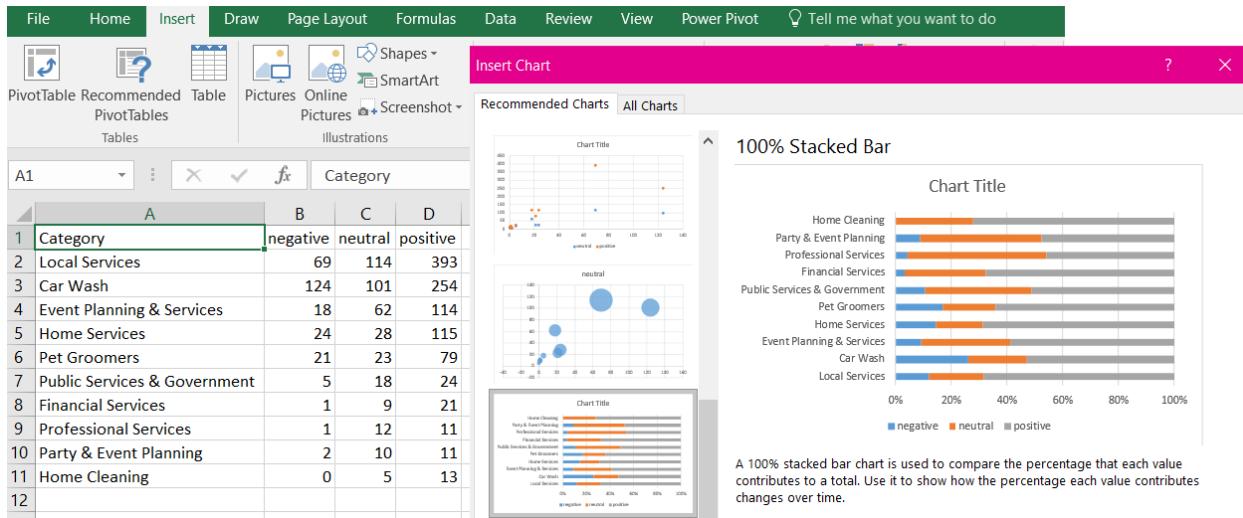


Final visualization

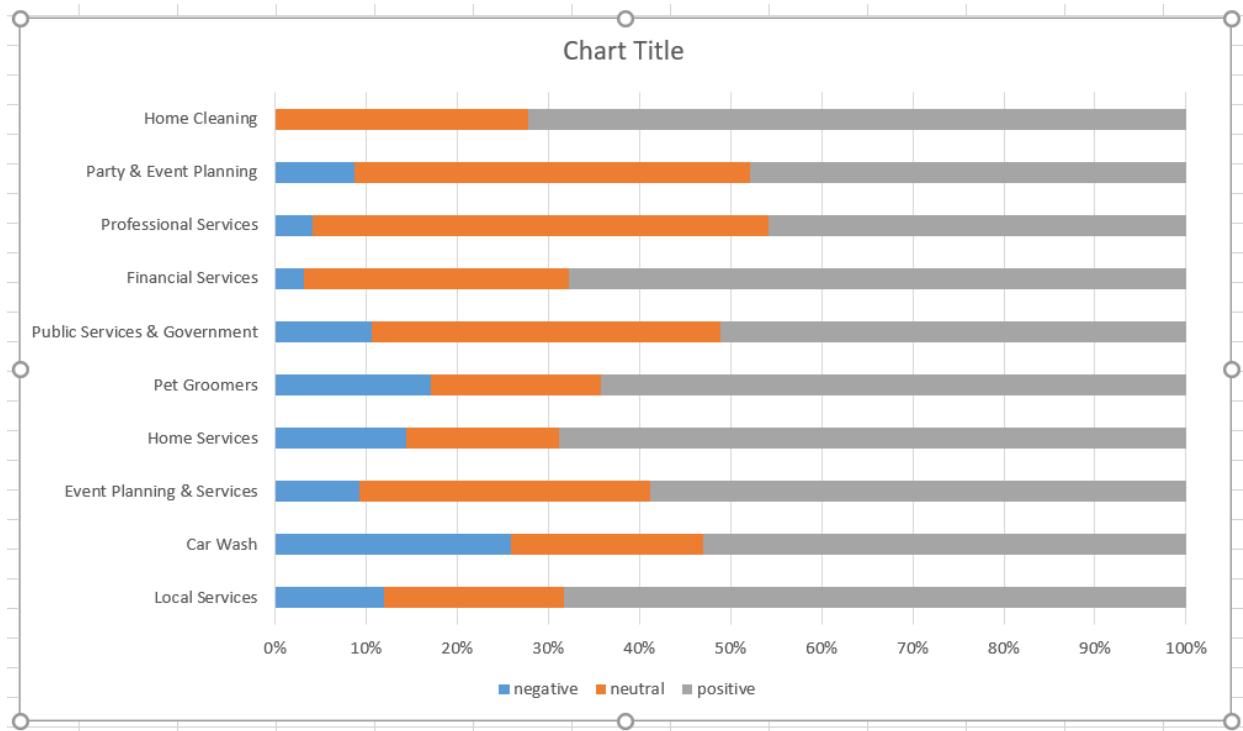


SENTIMENT ANALYSIS FOR ALL SERVICES IN EXCEL

- 1) Open the .CSV file with the output data and click on Insert->Recommended Charts and select the 100% Stacked Bar Chart. In the preview box, click on OK to select chart.



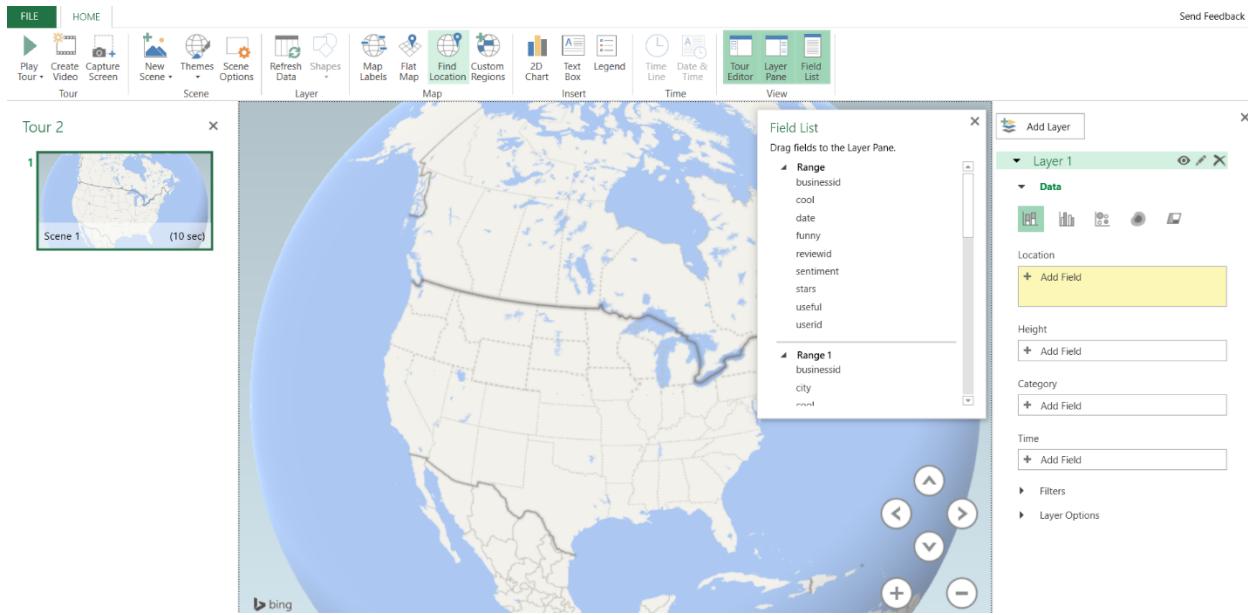
2) After you click on OK, the data will be represented in the form of Stacked Bar chart as follows:



GEO-SPATIAL VISUALIZATION OF THE USER SENTIMENTS

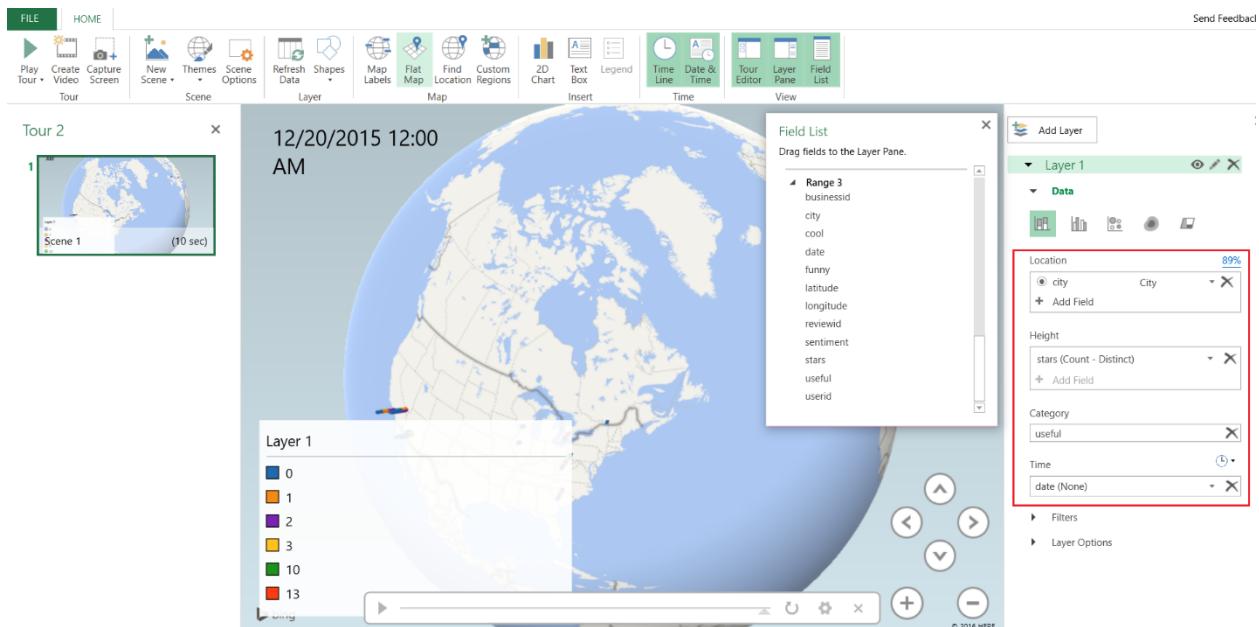
To create a Geospatial visualization video, follow the below given steps.

- 1) Open the .CSV file with the query output and save it as an excel workbook (.XLSX) on your local machine.
- 2) To create a new map tour, click on Insert->3D Map and select a new tour. The map tour window looks something like this.



- 3) Drag and drop the following attributes from the field list and associate them with the attributes in the Layer 1 on the right hand side.
- Location – city
 - Height – stars
 - Category – useful
 - Time – date

Click on 'Play' to observe the changes.





REFERENCES

1. Create HDInsight at Azure: <https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-tutorial-get-started-windows>
2. <http://hadooptutorial.info/tableau-integration-with-hadoop/>
3. <http://hortonworks.com/hadoop-tutorial/making-things-tick-with-tableau/>
4. <http://hortonworks.com/blog/how-to-integrate-tableau-and-hadoop-with-hortonworks-data-platform/>
5. <http://hortonworks.com/hadoop-tutorial/business-discovery-and-visualizing-your-data-in-hdp-using-qlikview/>