CS 185C – Introduction to Big Data Analytics

TripAdvisor Hotel Reviews

Submitted to : Prof. Peter Zadrozny

Submitted by : Team#2

Sailee Choudhary

Tin Hon Ng

Anumeha Umang Shah

Date: 11/19/2014

Server : http://216.121.58.230:8000/

username : admin

Password for root : Noeasywayout203

Contents

Introduction

Description

In today's fast paced digital world, customer reviews are so important. After all, when you look for dinner, place to stay, or entertainment, most likely you look for reviews to help you make your decision. Speaking of reviews, TripAdvisor is one of the most popular travel websites in the US providing people with reviews on travel-related content. Since the websites services are free to the users, users can provide as many reviews as they want. Many people have experience staying in hotel while traveling, no surprise the number reviews on hotels are very high with rich contents.

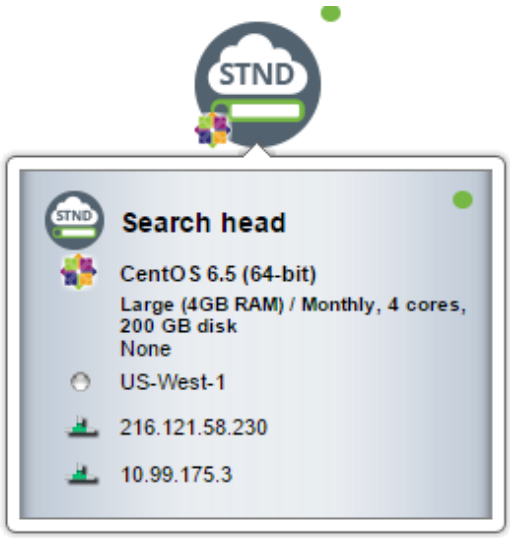
Therefore, our team decided to do an in-depth analysis on the hotel review dataset provided by TripAdvisor. We will apply Splunk technology to perform this analysis. The major objective is to load data into Splunk and come up with appropriate queries to achieve our user stories and suitable conclusions that are useful to others.

Objective

Using the hotel review dataset provided by TripAdvisor, our purpose is to speculate the correlation between different kinds of rating, author, devices they used, etc. Our aim is to discover the interesting facts about our dataset. We apply Splunk technology to perform this analysis. Again the main objective is to load data into Splunk and use it to trigger appropriate queries in order to arrive at suitable conclusions that are relevant and useful to others.

Cluster Description

Server size : Large (4 GB, 4 cores)
OS : CentOS
Architecture : 64-bit



Credentials :

To login to Splunk : http://216.121.58.230:8000/

username : admin

Password for root : Noeasywayout203

Phase 1 - Loading and Verifying Data

Dataset Description

This dataset consists of 878561 reviews from 4333 hotels crawled from TripAdvisor. It consists of two files one with all the reviews for different hotels, and other that contains the information about the hotels. The data is spread over several span of years. According to the survey, 81% of travelers find user reviews important when determining which hotel to stay at during their trip. So, positive feedback from satisfied guests sharing their experiences with your Hotel is very precious. Negative reviews – in contrast – should be diminished and used to improve the quality of the Hotel. So, it is important to have reviews both for customers as well as hotel management. So, we decided to analyze this data.

Where we obtained the data

We extracted our data from Carnegie Mellon University website.

URL : http://www.cs.cmu.edu/~jiweil/html/hotel-review.html

How we obtained the data

Everyone likes to travel! And we always don't have someone who resides where we are travelling. We need to sort shelter in some or the other hotel. It is always better to live in good hotel where you can find all the necessary amenities. It is always irritating to pay loads of money and not get customer satisfaction. Being fond of travelling and exploring new places, we decided to dig deeper into reviews of different hotels. So, with the help of TripAdvisor hotel reviews that we obtained from http://www.cs.cmu.edu/ we could do an in-depth analysis.

Issues with obtaining the data

We used all the publicly available datasets. It required no registration or identification to access, and had no legal restrictions on their use.

Structure and Key Fields of the data set(s)

File	Original Structure	New Structure
review.txt	json	json
offering.txt	json	csv

1. review.txt

{

author: {
id: unique id of each author
location: the home location of the author
num_reviews: number of reviews posted by the author on TripAdvisor
username: username specified by author on TripAdvisor
}

date: date when the review was posted (format : Month date, Year)
date_stayed: date when author stayed at this hotel (format : Month Year)
id: unique review id for each review
num_helpful_votes: count of useful reviews provided by author
offering_id: hotel id where the author stayed

ratings: { ratings provided by the author based on different criteria.
cleanliness: (0-5)
location: (0-5)
overall: (0-5)
rooms: (0-5)
service: (0-5)
sleep_quality: (0-5)
value: (0-5)
}

text: any comments or suggestions that author wants to provide

title: title for the review provided by author
via_mobile: tells if the review was posted via mobile or not

}

2. offering.csv

Field Name	Description
hotel_class	Hotel star rating information
region_id	Region id for the hotel depending on its location
url	The website link for the hotel
address_region	State where hotel is located
address_street-address	Street address for the hotel
address_postal-code	Address postal code
address_locality	City in which the hotel is located
type	Type for the hotel
hotel_id	Unique hotel id for each hotel
name	Name of the hotel

Issues with the data set(s)

1. Both the files, review.txt and offering.txt had a field 'id'.

For, review.txt, 'id' is for 'review id' and it is unique for each event across the whole file.

For, offering.txt, 'id' is for 'hotel id' and this field needs mapping with 'offering_id' in the review.txt because for review.txt, the offering_id specifies the 'hotel id'

So, while pre-processing the data, we changed the field 'id' in offering.txt to 'hotel_id'

2. We wanted to use the file offering.txt as lookup table. Splunk accepts only csv files as lookup tables. Therefore, we had to write a php script to change the json file to csv file.

Pre-processing (ETL) of the data

1. review.txt - This file is in json format which is identified by Splunk. So, no pre-processing was required for this file.

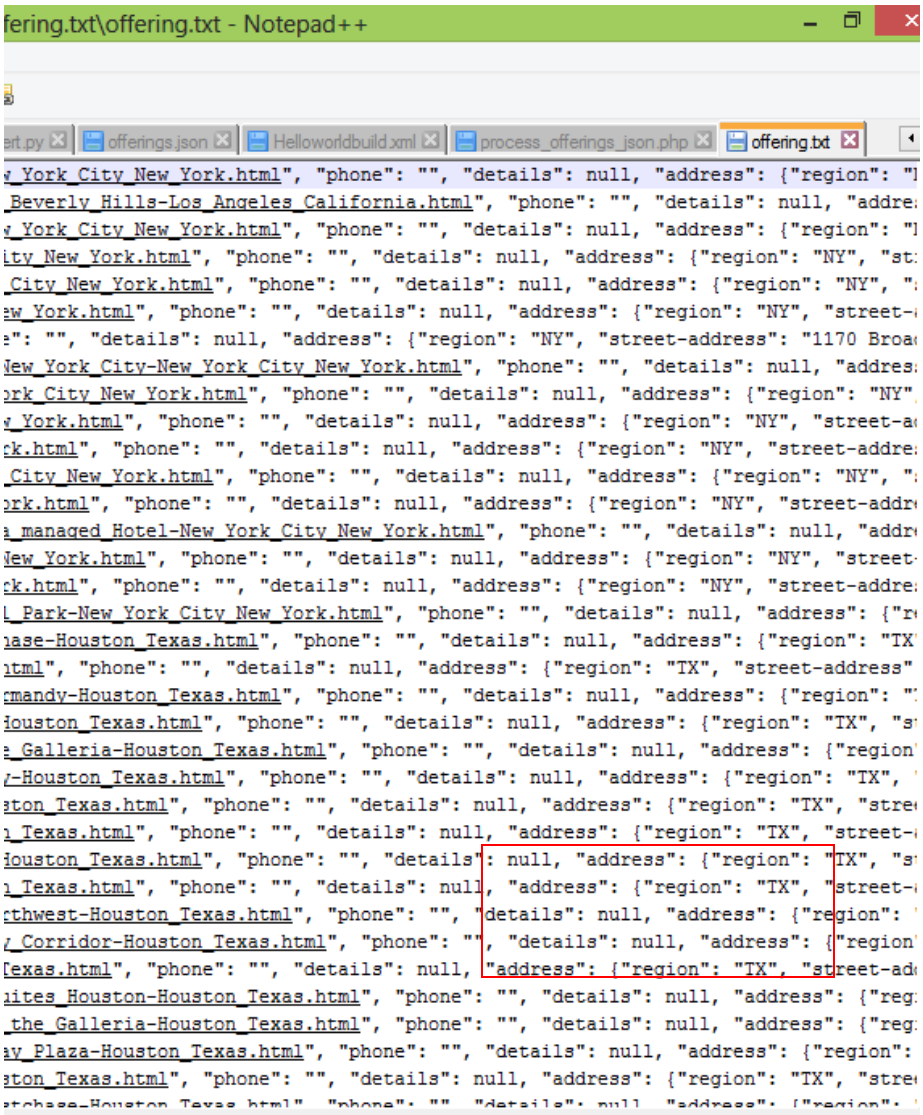
2. offering.txt - This file contains information about hotels and so, we wanted to use this file as lookup table. The file was in json format and to be used as lookup table we had to convert it to csv. We wrote the following script for converting it to csv.

```
<?php

//config vars
$input_json = "offering.txt";
$output_csv = "offering.csv";
$csv_json_map = array(
   'hotel_class'   =>  "hotel_class",
   'region_id'     =>  "region_id",
   'url'           =>  "url",
   'phone'         =>  "phone",
   'details'       =>  "details",
   'address_region'        =>  "address->region",
   'address_street-address'=>  "address->{'street-address'}",
   'address_postal-code'   =>  "address->{'postal-code'}",
   'address_locality'      =>  "address->locality",
   'type'          =>  "type",
   'id'            =>  "hotel_id",
   'name'          =>  "name"
   );

//processing starts
#open input json file
$handle = fopen($input_json, "r") or die("Couldn't get handle");

#open output csv file
$file = fopen($output_csv, 'w');

#write csv header row
$arr_header = array_keys($csv_json_map);
$header = implode(",",$arr_header);
fwrite($file, $header."\n");

if ($handle) {
   while (!feof($handle)) {
      $buffer = fgets($handle, 4096);

      #json_decode input row
      $json = json_decode($buffer);

      #prepare row data
      $row = '';
      foreach($csv_json_map as $csv_field => $json_field)
      {
         $value = '';
         eval("\$value = isset(\$json->$json_field)? \$json->$json_field : '';");
         $row .= '"'.str_replace('"', '""', $value).'",';
```

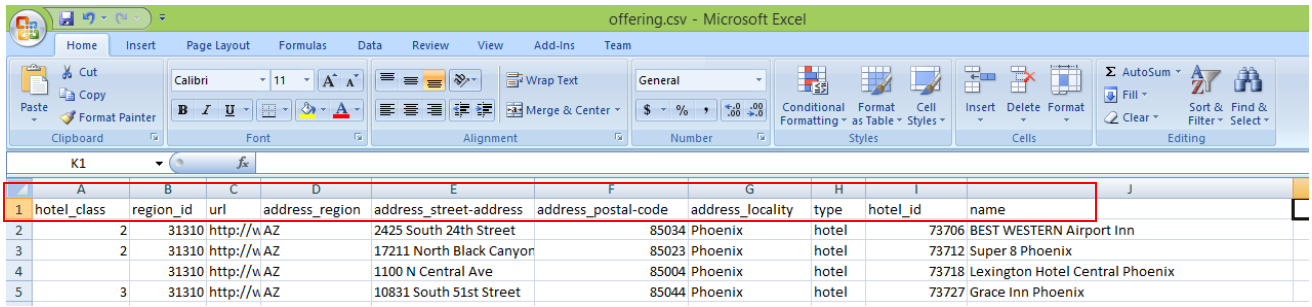
```
        }
        #write row
        fwrite($file, $row."\n");
    }
    fclose($handle);
}
?>
```

Also, while doing this transformation, we changed the field 'id' to 'hotel_id' as in the review.txt, there was field 'id' already used for review id. So, to avoid the confusion, we renamed this field.

In the original json file, the fields, 'details' and 'phone' were null. So, we removed these from our offerings.csv



Columns in final csv : offering.csv

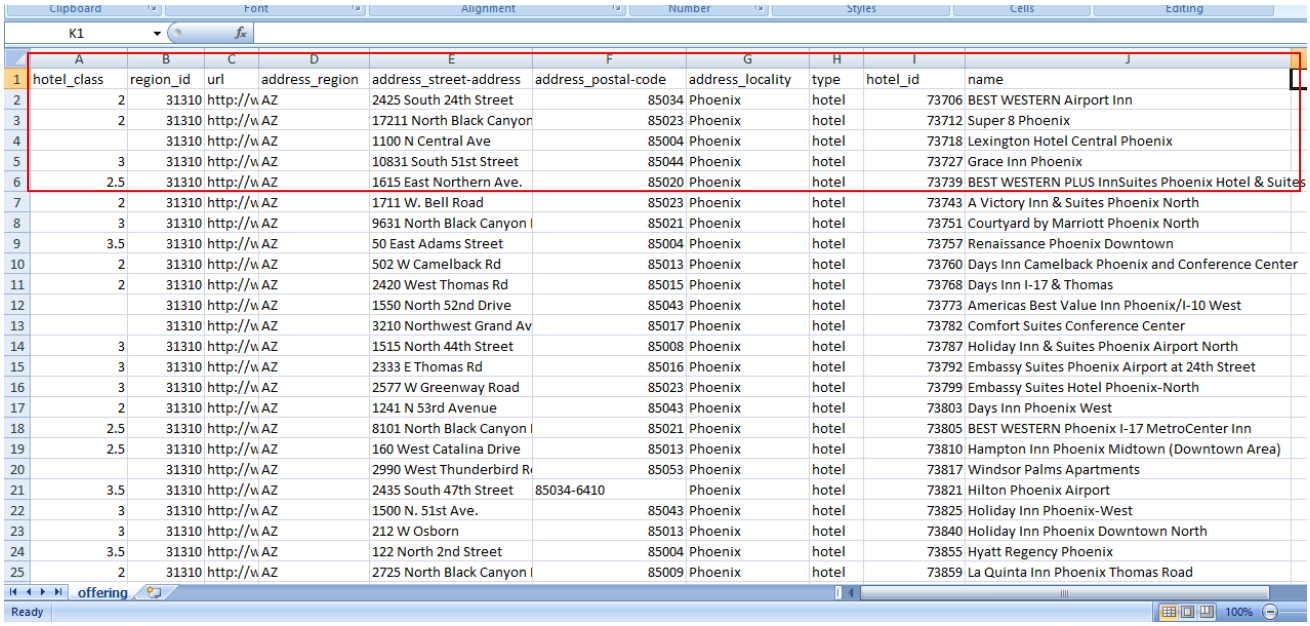


offering.csv - Microsoft Excel

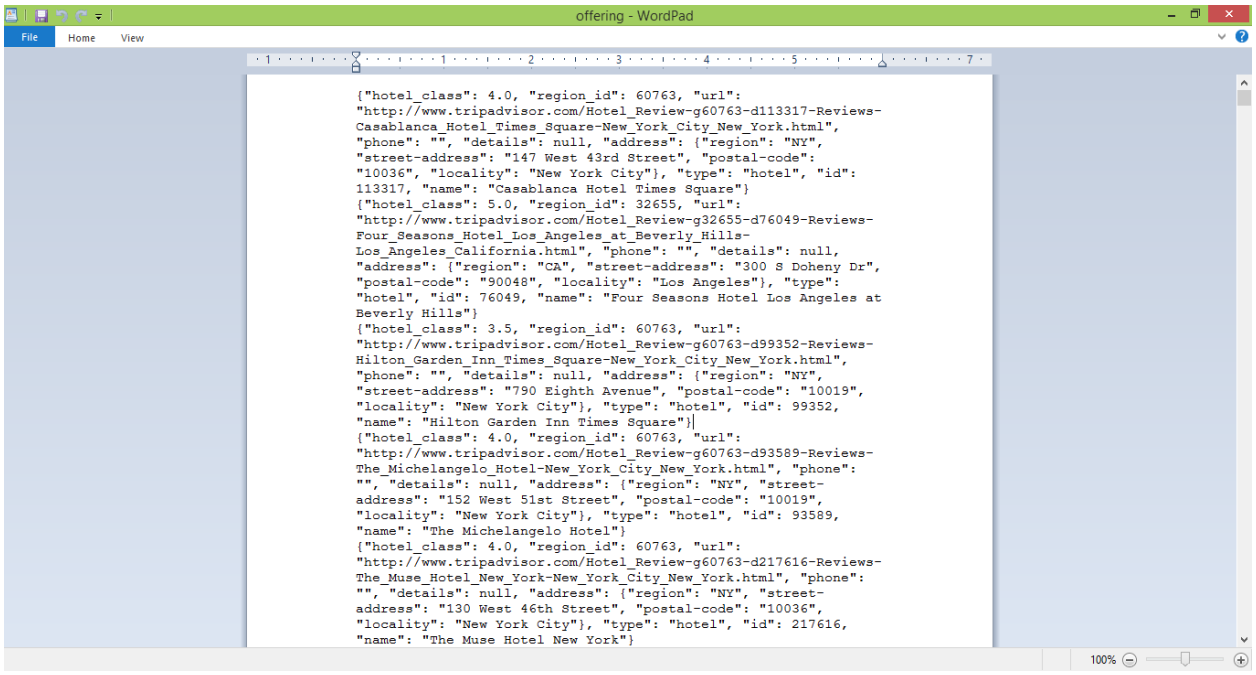
	A	B	C	D	E	F	G	H	I	J
1	hotel_class	region_id	url	address_region	address_street-address	address_postal-code	address_locality	type	hotel_id	name
2	2	31310	http://w	AZ	2425 South 24th Street	85034	Phoenix	hotel	73706	BEST WESTERN Airport Inn
3	2	31310	http://w	AZ	17211 North Black Canyon	85023	Phoenix	hotel	73712	Super 8 Phoenix
4		31310	http://w	AZ	1100 N Central Ave	85004	Phoenix	hotel	73718	Lexington Hotel Central Phoenix
5	3	31310	http://w	AZ	10831 South 51st Street	85044	Phoenix	hotel	73727	Grace Inn Phoenix

Verification if the data was pre-processed correctly :

Check the first 5 records of original and csv files:



	A	B	C	D	E	F	G	H	I	J
1	hotel_class	region_id	url	address_region	address_street-address	address_postal-code	address_locality	type	hotel_id	name
2	2	31310	http://w	AZ	2425 South 24th Street	85034	Phoenix	hotel	73706	BEST WESTERN Airport Inn
3	2	31310	http://w	AZ	17211 North Black Canyon	85023	Phoenix	hotel	73712	Super 8 Phoenix
4		31310	http://w	AZ	1100 N Central Ave	85004	Phoenix	hotel	73718	Lexington Hotel Central Phoenix
5	3	31310	http://w	AZ	10831 South 51st Street	85044	Phoenix	hotel	73727	Grace Inn Phoenix
6	2.5	31310	http://w	AZ	1615 East Northern Ave.	85020	Phoenix	hotel	73739	BEST WESTERN PLUS InnSuites Phoenix Hotel & Suites
7	2	31310	http://w	AZ	1711 W. Bell Road	85023	Phoenix	hotel	73743	A Victory Inn & Suites Phoenix North
8	3	31310	http://w	AZ	9631 North Black Canyon	85021	Phoenix	hotel	73751	Courtyard by Marriott Phoenix North
9	3.5	31310	http://w	AZ	50 East Adams Street	85004	Phoenix	hotel	73757	Renaissance Phoenix Downtown
10	2	31310	http://w	AZ	502 W Camelback Rd	85013	Phoenix	hotel	73760	Days Inn Camelback Phoenix and Conference Center
11	2	31310	http://w	AZ	2420 West Thomas Rd	85015	Phoenix	hotel	73768	Days Inn I-17 & Thomas
12		31310	http://w	AZ	1550 North 52nd Drive	85043	Phoenix	hotel	73773	Americas Best Value Inn Phoenix/I-10 West
13		31310	http://w	AZ	3210 Northwest Grand Av	85017	Phoenix	hotel	73782	Comfort Suites Conference Center
14	3	31310	http://w	AZ	1515 North 44th Street	85008	Phoenix	hotel	73787	Holiday Inn & Suites Phoenix Airport North
15	3	31310	http://w	AZ	2333 E Thomas Rd	85016	Phoenix	hotel	73792	Embassy Suites Phoenix Airport at 24th Street
16	3	31310	http://w	AZ	2577 W Greenway Road	85023	Phoenix	hotel	73799	Embassy Suites Hotel Phoenix-North
17	2	31310	http://w	AZ	1241 N 53rd Avenue	85043	Phoenix	hotel	73803	Days Inn Phoenix West
18	2.5	31310	http://w	AZ	8101 North Black Canyon	85021	Phoenix	hotel	73805	BEST WESTERN Phoenix I-17 MetroCenter Inn
19	2.5	31310	http://w	AZ	160 West Catalina Drive	85013	Phoenix	hotel	73810	Hampton Inn Phoenix Midtown (Downtown Area)
20		31310	http://w	AZ	2990 West Thunderbird R	85053	Phoenix	hotel	73817	Windsor Palms Apartments
21	3.5	31310	http://w	AZ	2435 South 47th Street	85034-6410	Phoenix	hotel	73821	Hilton Phoenix Airport
22	3	31310	http://w	AZ	1500 N. 51st Ave.	85043	Phoenix	hotel	73825	Holiday Inn Phoenix-West
23	3	31310	http://w	AZ	212 W Osborn	85013	Phoenix	hotel	73840	Holiday Inn Phoenix Downtown North
24	3.5	31310	http://w	AZ	122 North 2nd Street	85004	Phoenix	hotel	73855	Hyatt Regency Phoenix
25	2	31310	http://w	AZ	2725 North Black Canyon	85009	Phoenix	hotel	73859	La Quinta Inn Phoenix Thomas Road



They are exactly same.

Also, checking the number of records

4330	3	60878	http://w	WA	300 Roy Street	98109	Seattle	hotel	1723585	The Maxwell Hotel Seattle
4331	3	60878	http://w	WA	612 2nd Avenue	98104	Seattle	hotel	1724250	Courtyard Seattle Downtown
4332	3	60878	http://w	WA	110 6th Avenue North	98109	Seattle	hotel	1755673	Hyatt Place Seattle/Downtov
4333		60878	http://w	WA	9613 Aurora Ave N	98103	Seattle	hotel	2095177	Columbus Motor Inn
4334		60878	http://w	WA	9501 Aurora Avenue Nort	WA 98103	Seattle	hotel	3291041	Crown Inn
4335										
4336										
4337										

```
[root@search-head ~]# wc -l offering.txt
4333 offering.txt
[root@search-head ~]#
```

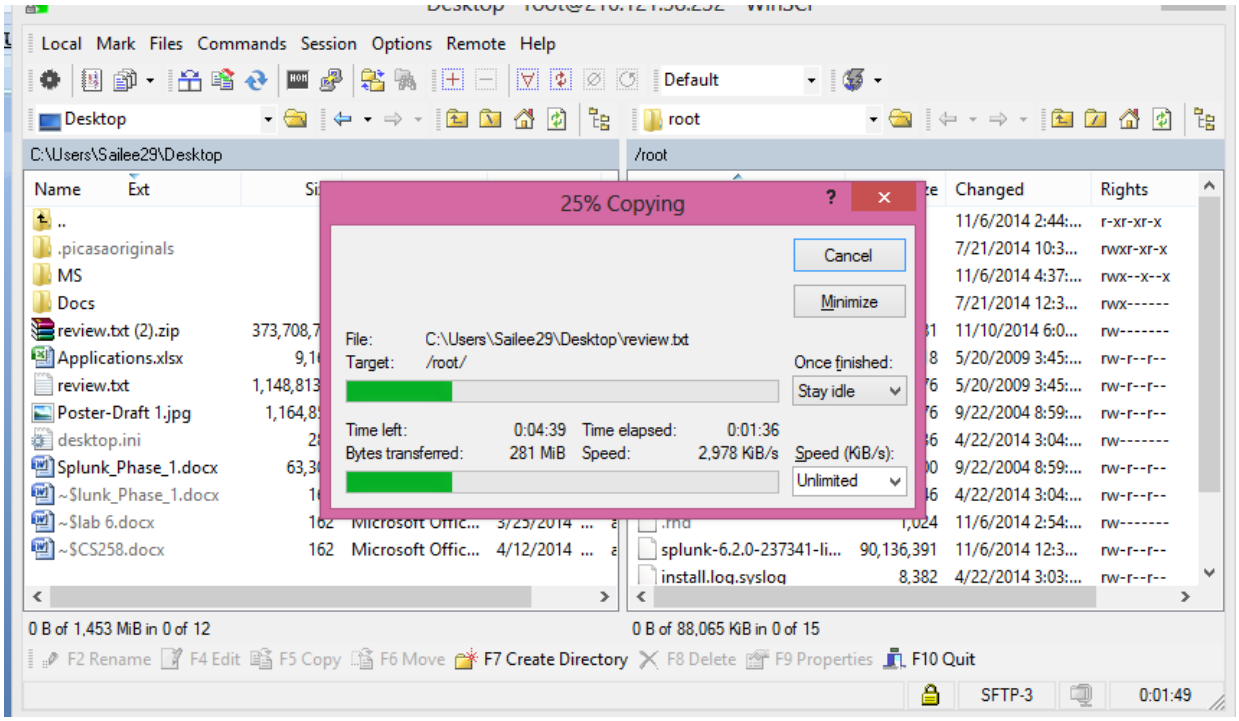
As in excel, the first line is for column name, so we have one extra record in csv file.

Loading the data

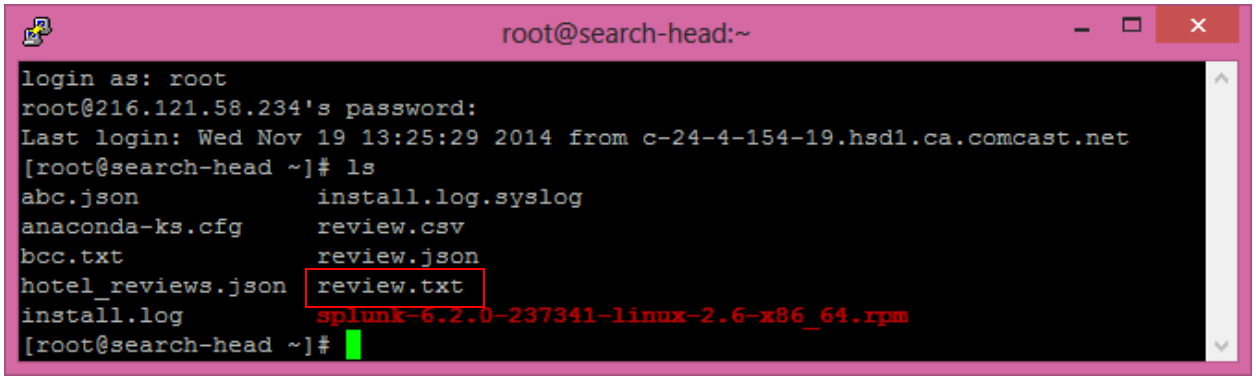
1. Loading the reviews table

Step 1 : Transfer the data from local machine to server root directory.

Use WinSCP to transfer data from local machine to search head.



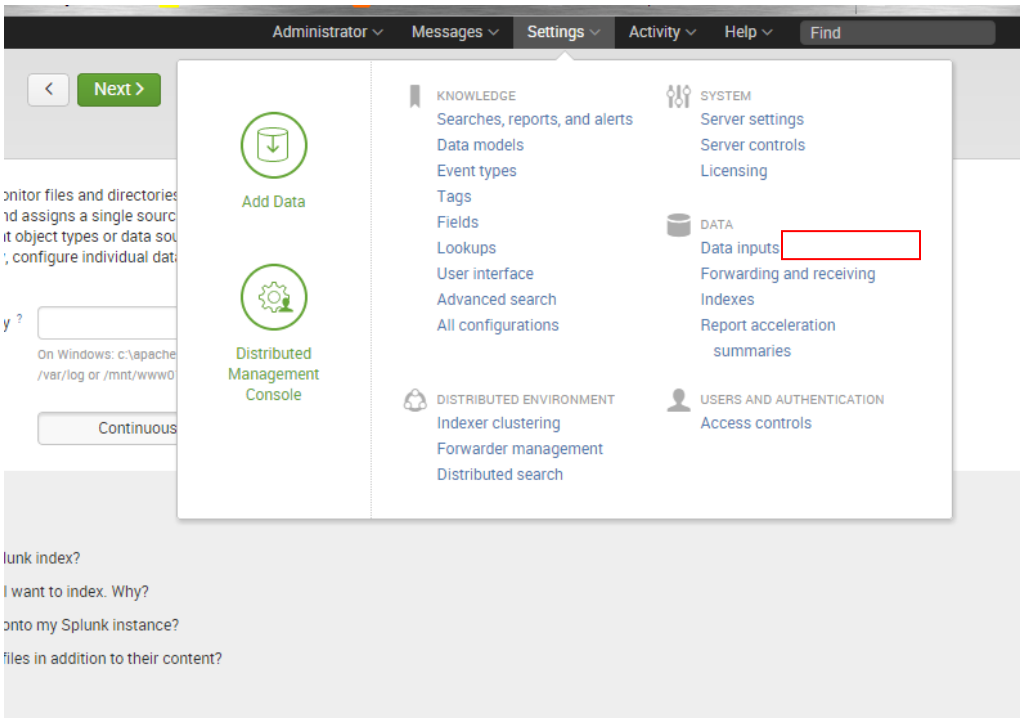
Step 2: Login to search head cluster and check if the data file is transferred.



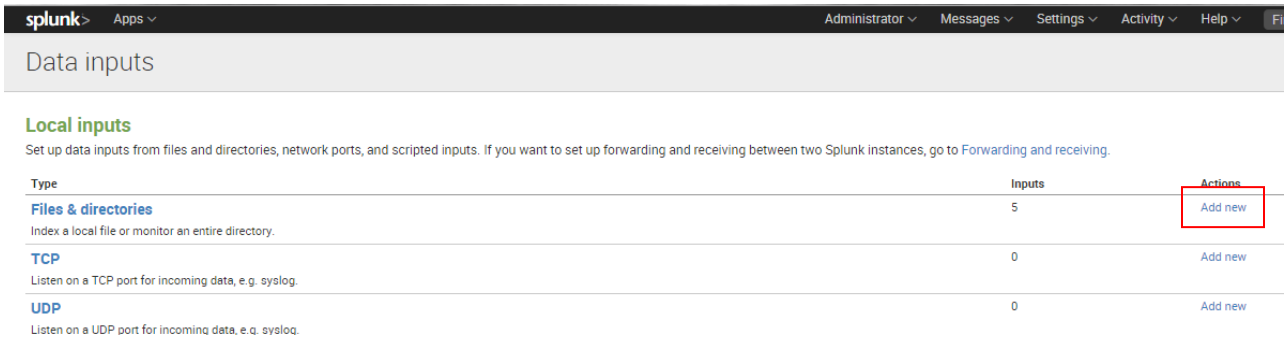
Step 3 : Login to the web interface of search head.

URL : http://216.121.58.230:8000/
username : admin
password : Noeasywayout203

Step 4 : Go to 'Settings' and click on 'Data inputs'.



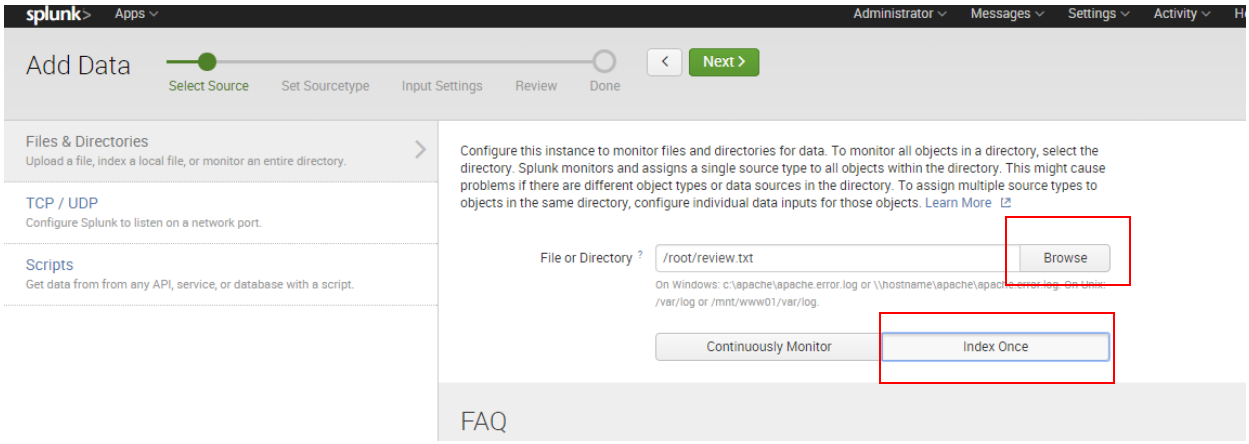
Step 5 : Click on 'Add new' that resides besides 'Files & directories'.

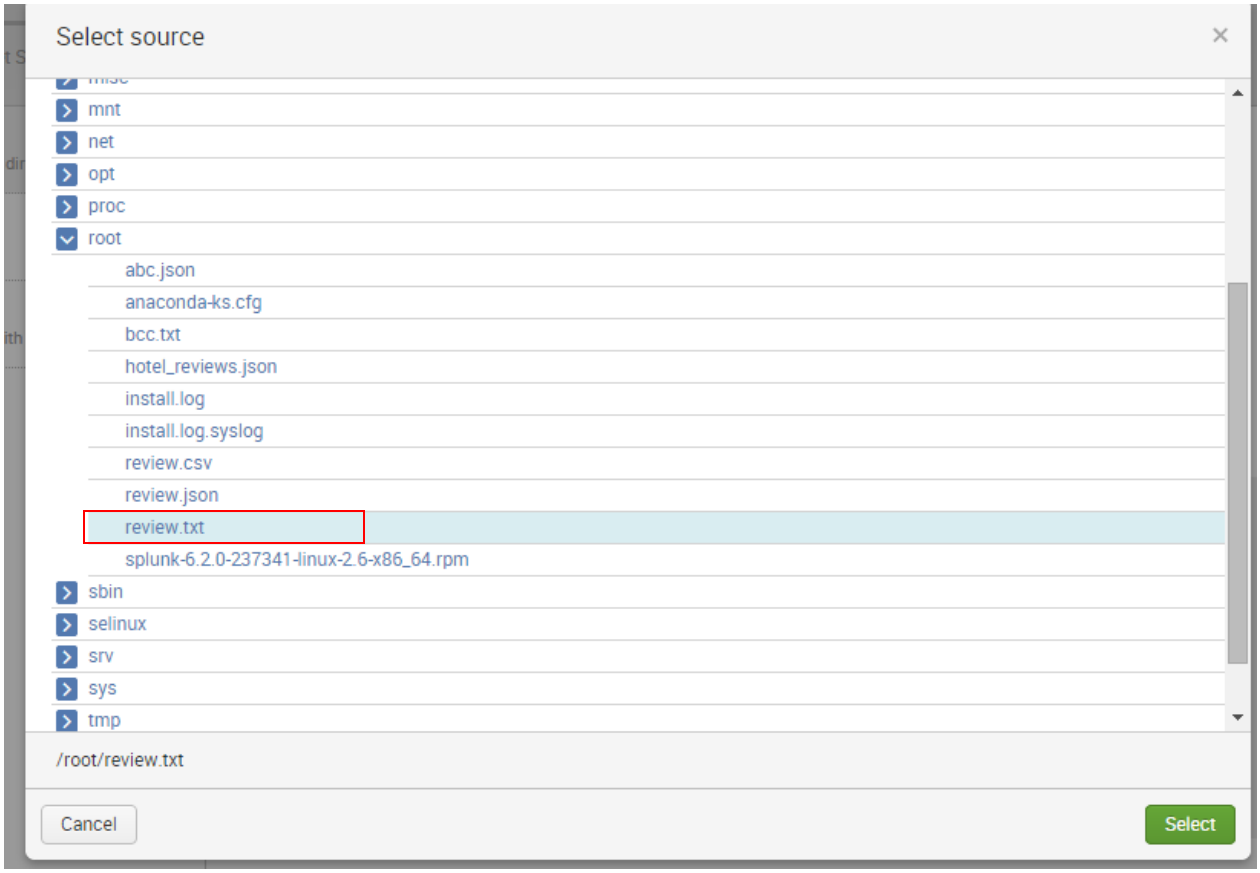


Step 6 : Click on 'Browse' and on the next screen, go to the 'root' directory and select the file to be loaded.

File loaded : "review.txt"

Click on 'Index once' and click 'Next'





Step 7 : Set Sourcetype.

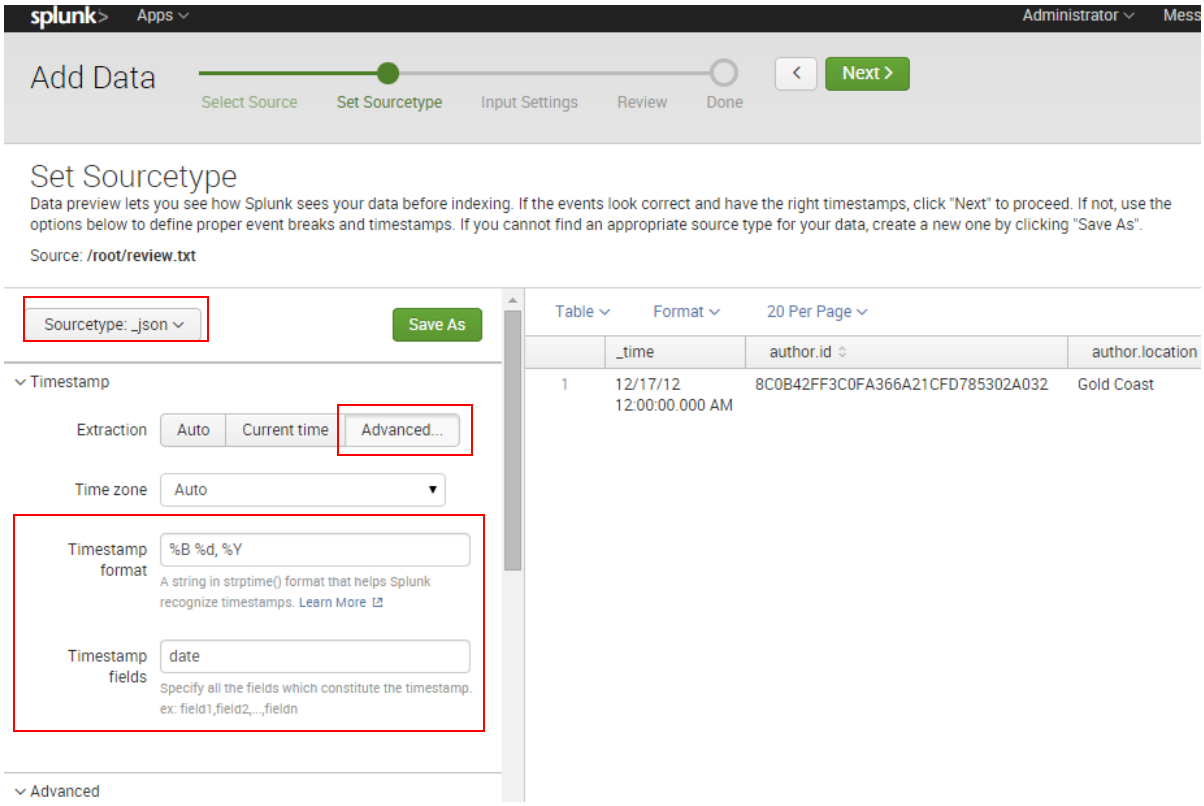
(1) Select sourcetype by first going to 'Structured' and then '**_json**'

(2) Timestamp :

Select 'Advanced' and specify the timestamp format and fields.

Timestamp format : **%B %d, %Y**

Timestamp fields : date

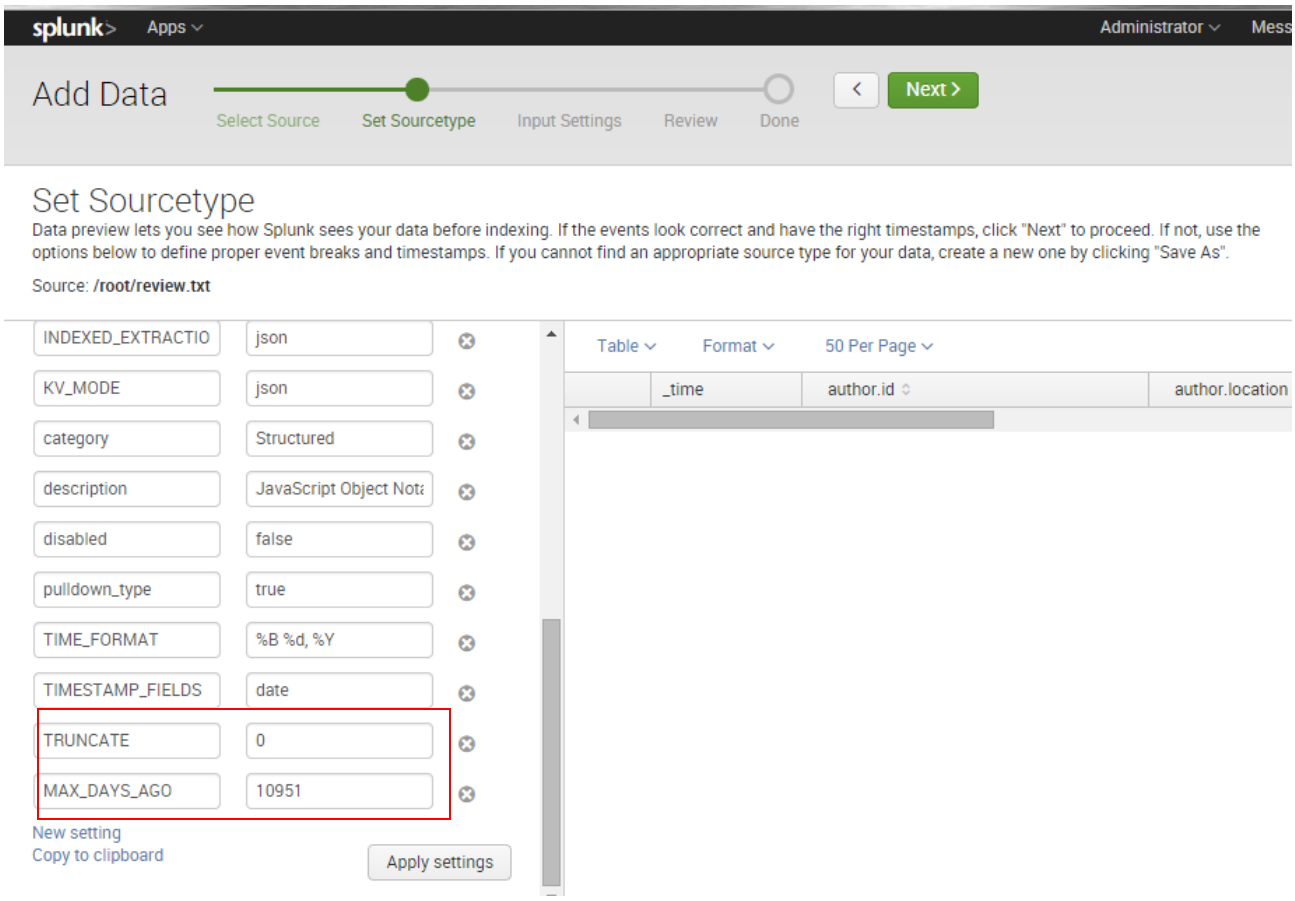


(3) Select 'Advanced' and click on 'New setting'

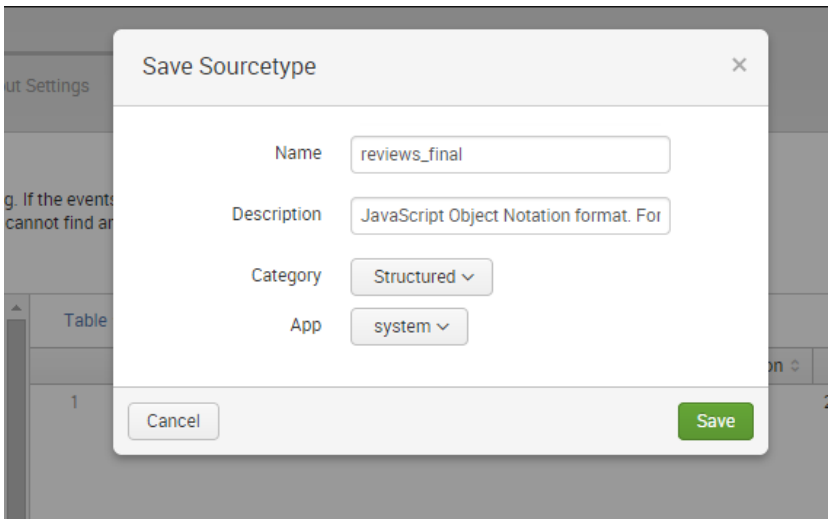
Add an extra field:

- Name : TRUNCATE
  Value : 0
  (Some of the events in our data were too long because of long comments in the 'text' field. So, few events were not loaded properly. In order to avoid this, we specified a setting to never truncate any event in props.conf)

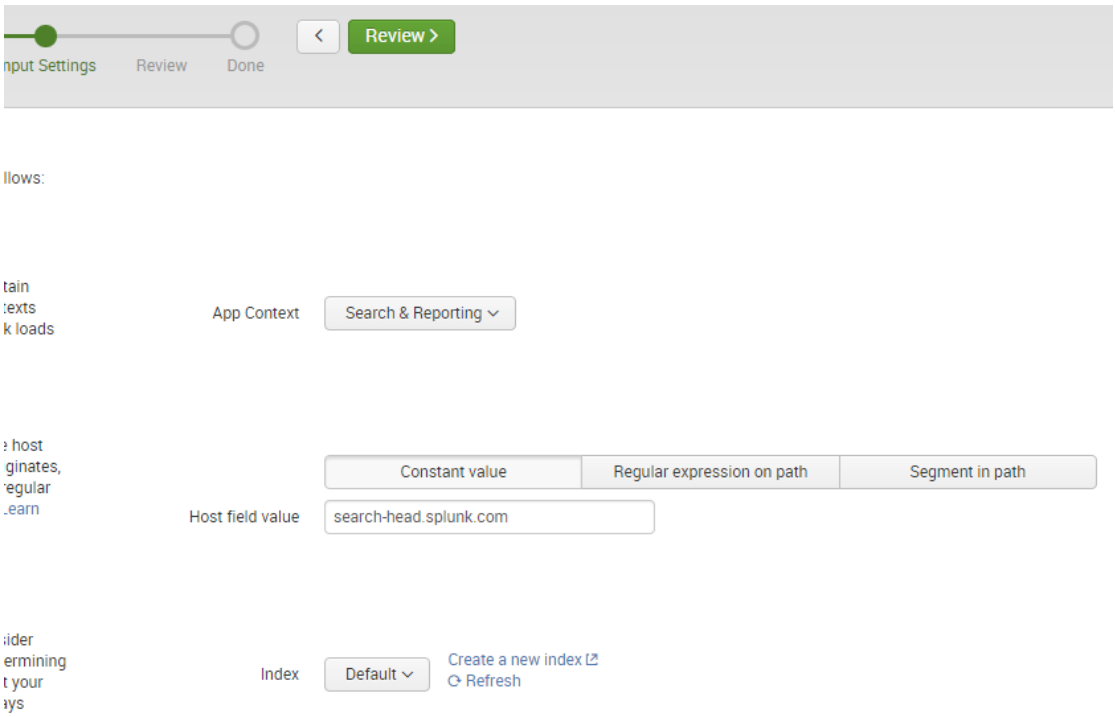
- Name : MAX_DAYS_AGO
  Value : 10951
  (This field is used for specifying for how many days before the current date should Splunk index the data. Since, our data was starting from year 2001, we set this value to the maximum value(10951). Default value is 2000)



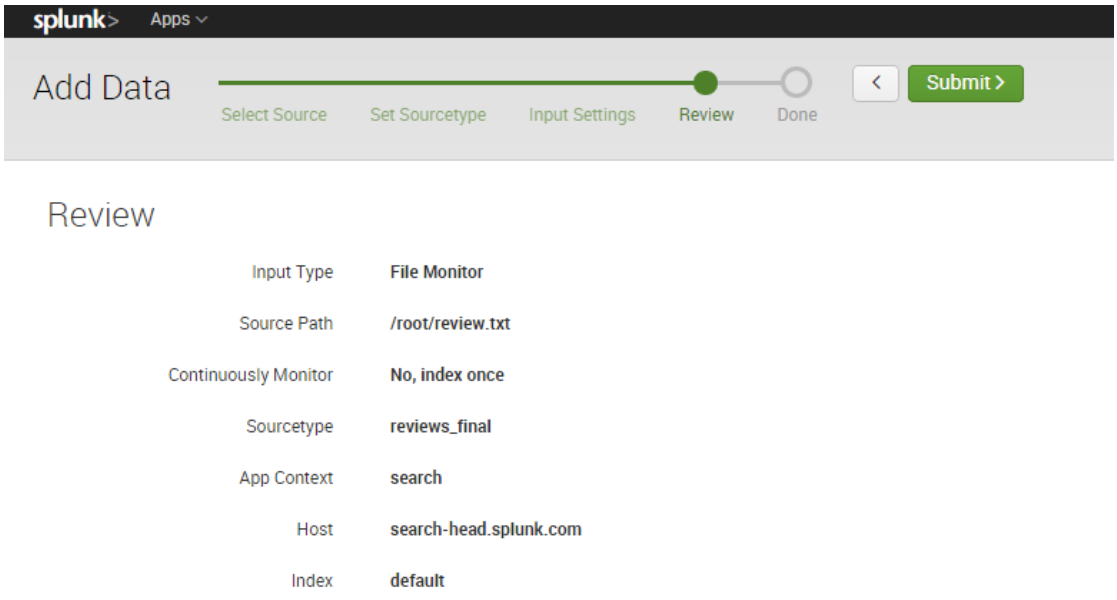
Step 8 : Click on 'Save as' and save the sourcetype.



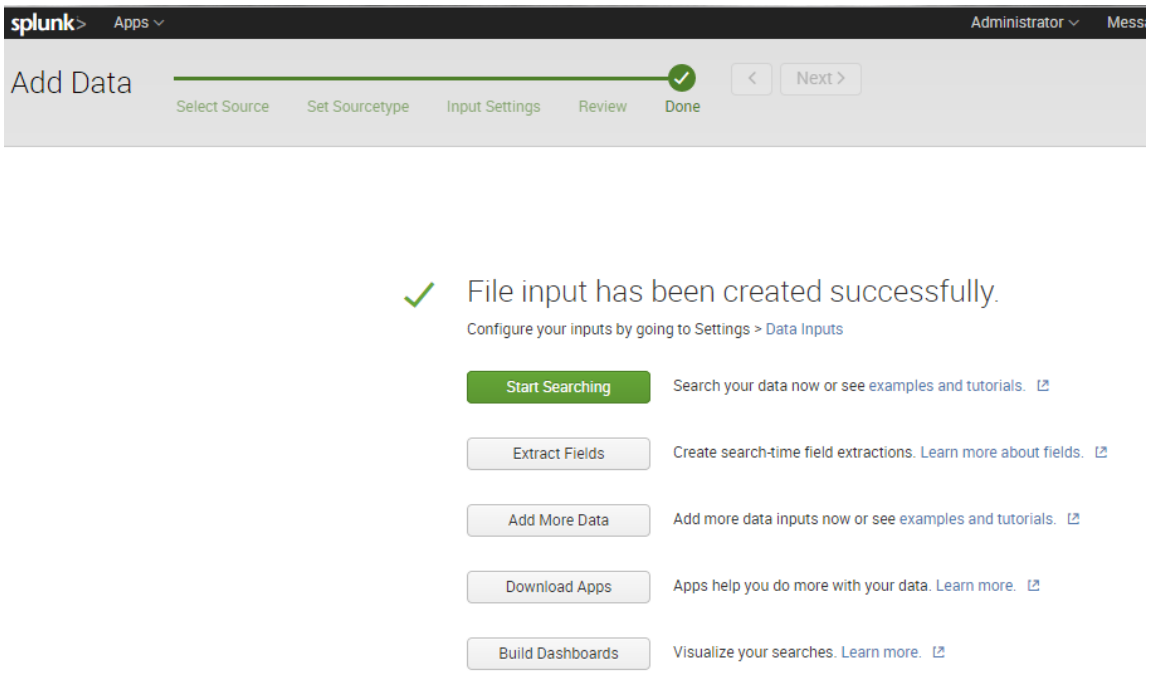
Step 9 : Specify the input settings and click on 'Review'



Step 10: If everything is correct, click on 'Submit' else go back and do the necessary changes.



Step 11: Start Searching after you see the following screen.



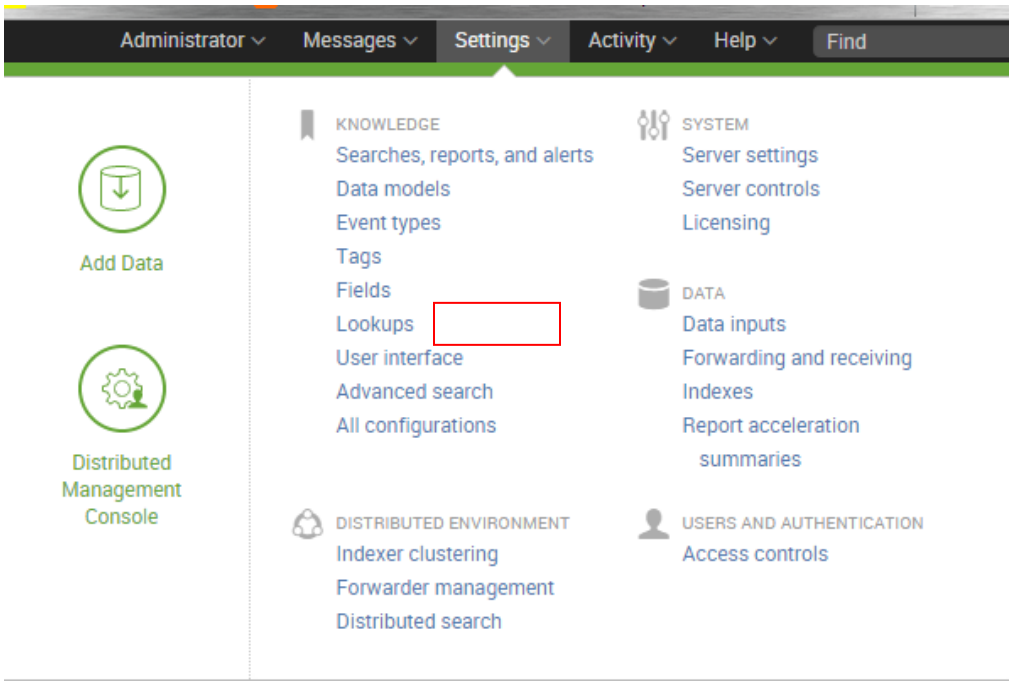
2. Loading the hotel information table

The hotel information table provides information about the hotels such as address, url, name, etc. The hotel_id field of this table connects with the

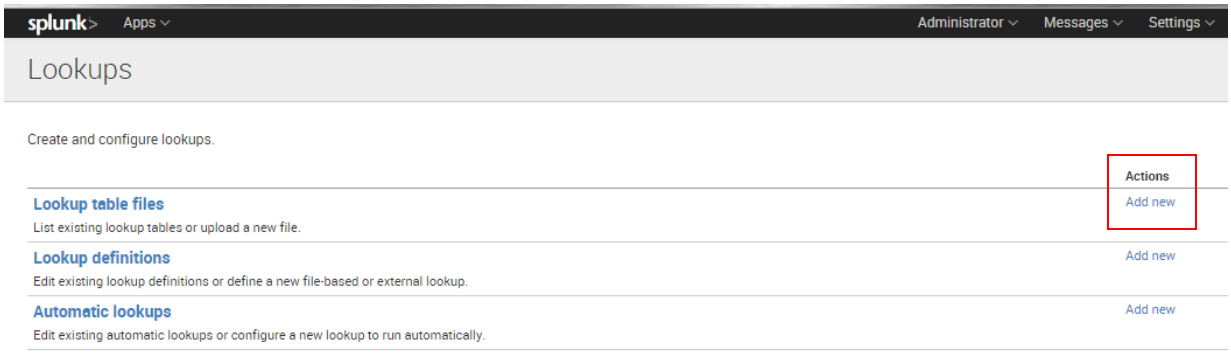
Step 1 : Login to the web interface of search head.

URL : http://216.121.58.230:8000/
username : admin
password : Noeasywayout203

Step 2 : Go to 'Settings' and click on 'Lookups'.

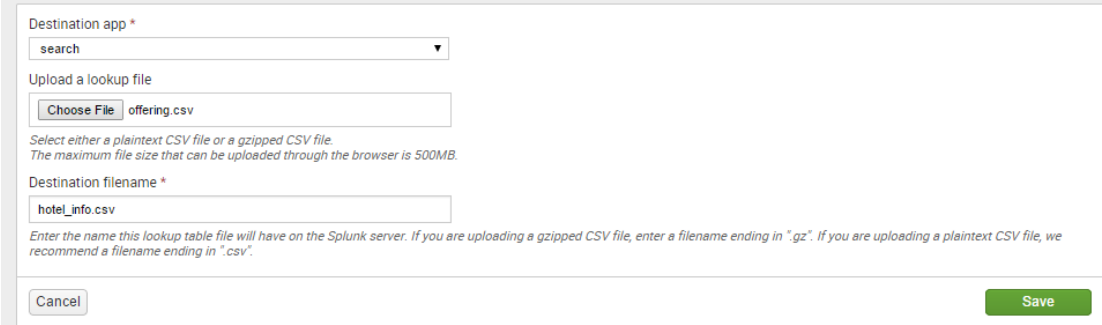


Step 3 : Click on 'Add new' adjacent to Lookup table files.

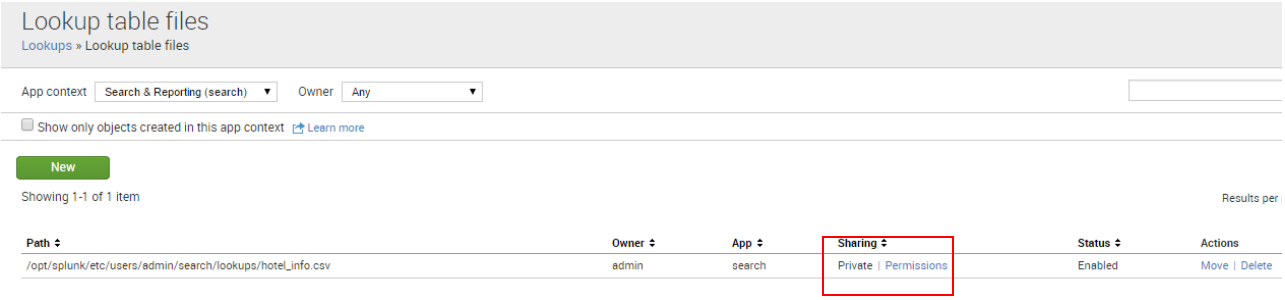


Step 4 : Fill out the details for lookup table files. Browse the file that you want to use as lookup table.

Destination app : search

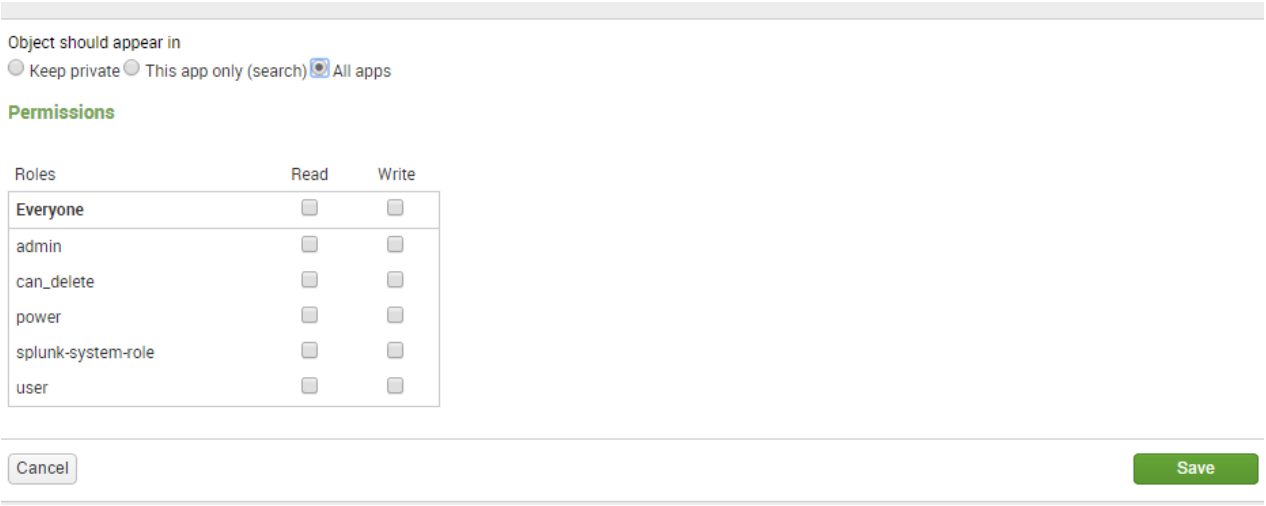


Step 5 : The new lookup table uploaded should be reflected in the lookup table files. Click on 'Permissions' for this file.

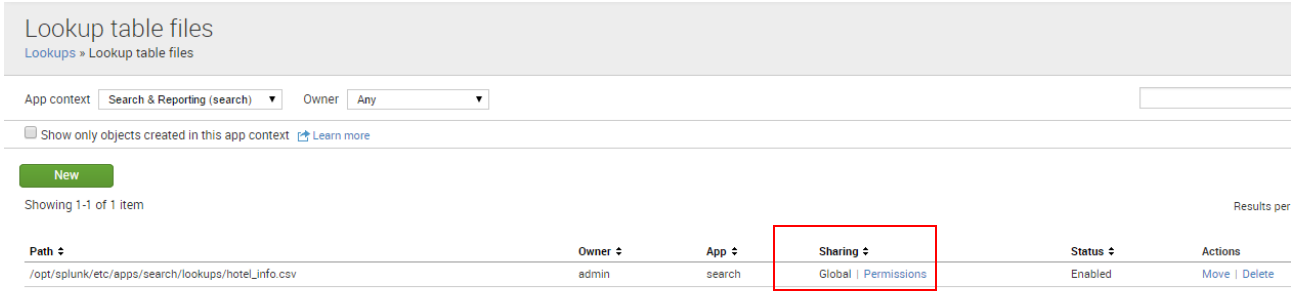


Step 6 : Give the appropriate permissions for the file and click on 'Save'

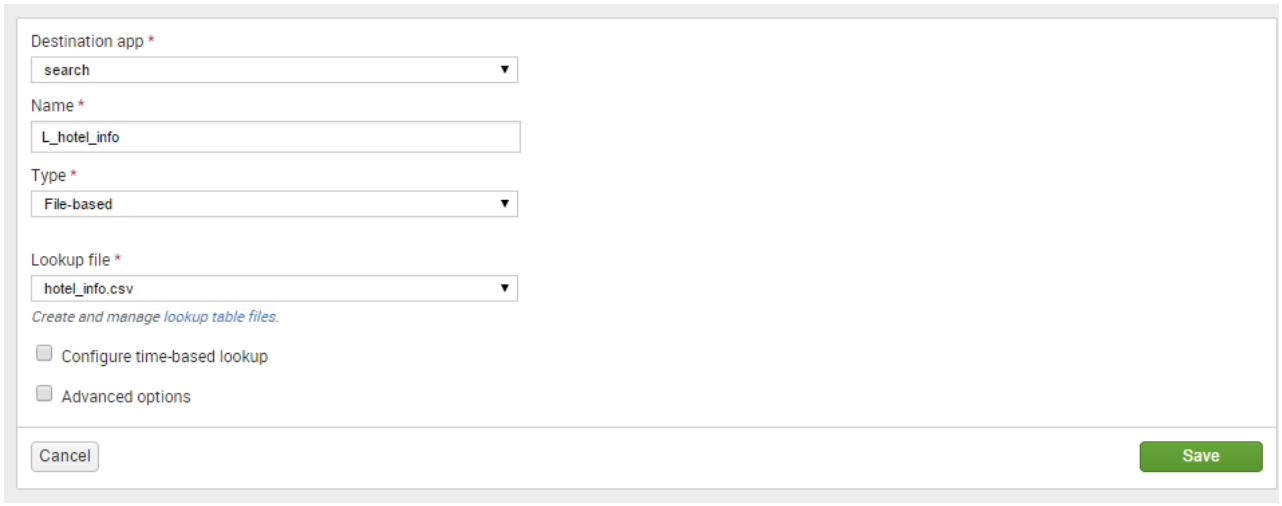
Here, we made the lookup table accessible to all the apps.



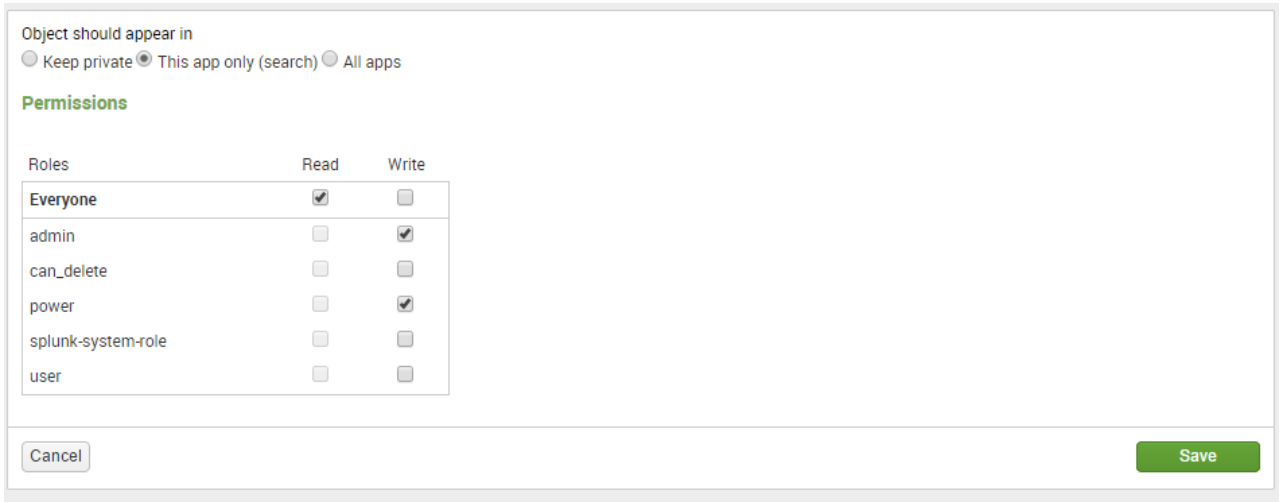
Step 7 : On clicking on save, the file permissions should be reflected in the lookup table files



Step 8 : Next, Go to 'Settings', then 'Lookups' and then click on 'Add New' besides 'Lookup definitions'

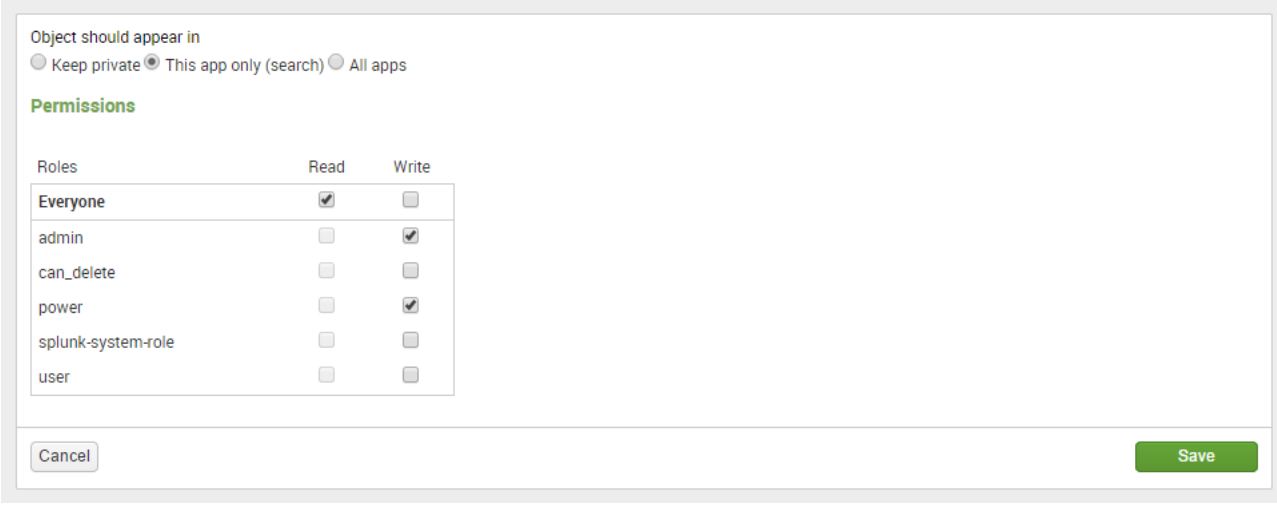


Save this Lookup definition. It should appear in 'Lookup definitions'. Click on 'Permissions' and select 'This app only (search)' and save.



Step 9 : Go to 'Settings', then 'Lookups' and then click on 'Add New' besides ' Automatic Lookups' and configure the fields and 'Save'.

Step 10 : Edit the Permissions and make this automatic lookup table available to Search App.



Step 11 : The automatic lookup table should be reflected in 'Automatic lookups'

Automatic lookup table :

Field Name	Field in Automatic lookup table
hotel_class	hotel_class
region_id	hotel_region_id
url	hotel_url
address_region	hotel_state
address_street-address	hotel_street_address
address_postal-code	hotel_postal
address_locality	hotel_city
type	hotel_type
hotel_id	hotel_id
name	hotel_name

Data Verification

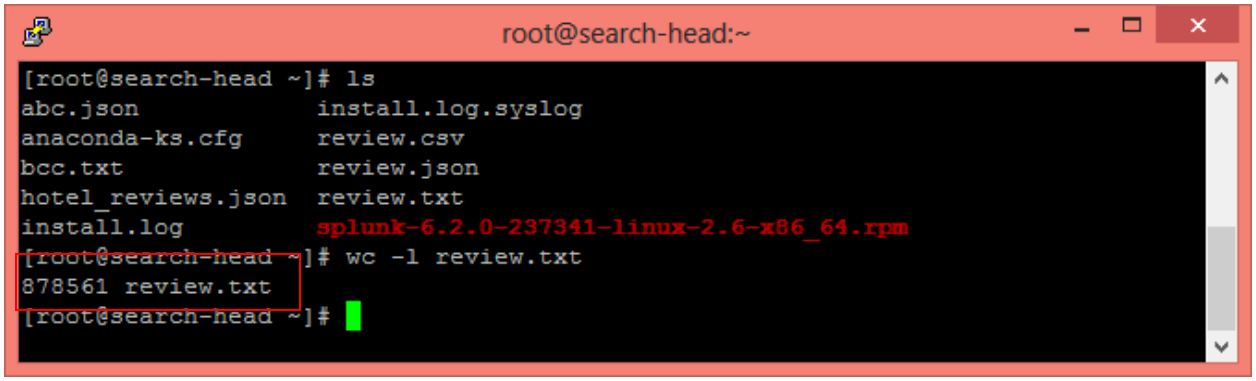
1. Verification of count of events

1. Verification for count of table review.txt

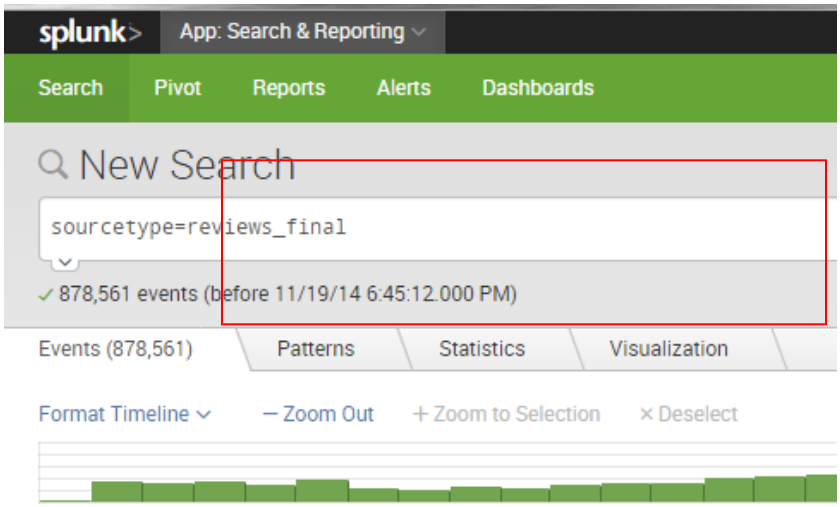
review.txt contains all the hotel reviews from TripAdvisor. There are 878561 reviews in this file. We first verify if all the reviews have been uploaded correctly.

This is the snapshot of the count of records in the original file. (review.txt)

```
# wc -l review.txt
```



We now check if there are same number of events in Splunk :

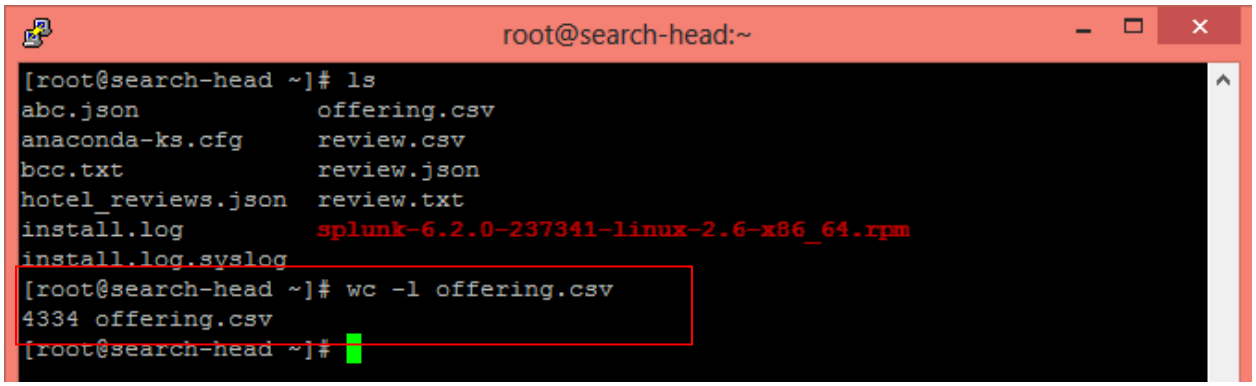


The total number of events in both the cases are same. So, we can conclude that all the events were loaded correctly

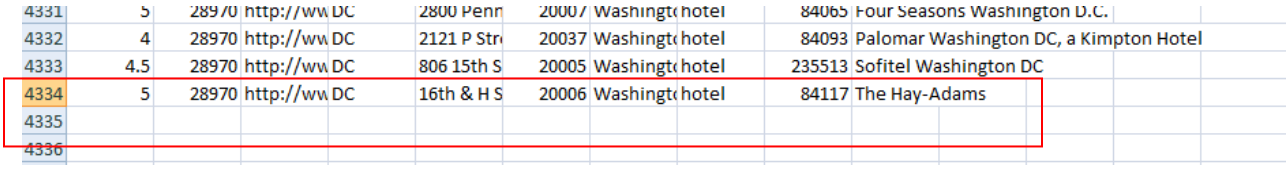
2. Verification for lookup table hotel_info.csv

This is the lookup table and so it was converted to csv while pre-processing. We count the number of records in this file using the command:

wc -l offering.csv



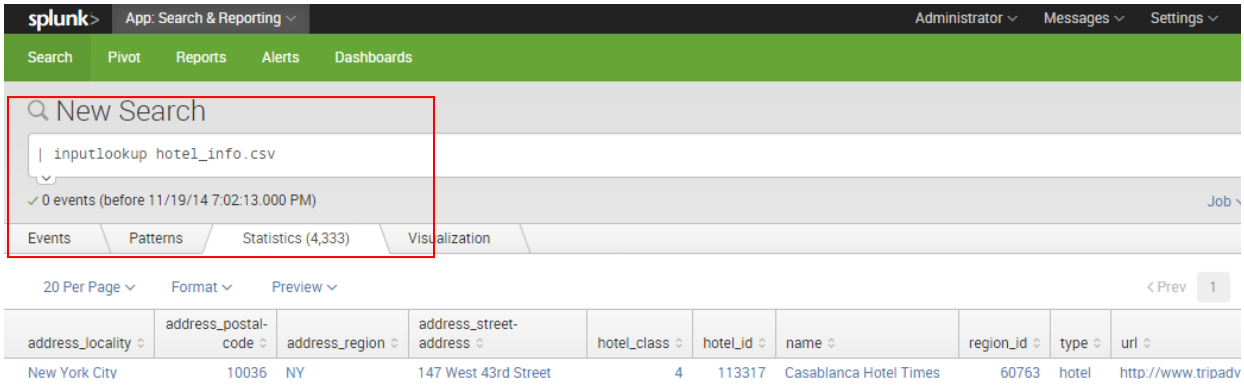
Also, we can have screenshot from excel file.



4331	5	28970	http://ww	DC	2800 Penn	20007	Washingt(	hotel	84065	Four Seasons Washington D.C.
4332	4	28970	http://ww	DC	2121 P Str	20037	Washingt(	hotel	84093	Palomar Washington DC, a Kimpton Hotel
4333	4.5	28970	http://ww	DC	806 15th S	20005	Washingt(	hotel	235513	Sofitel Washington DC
4334	5	28970	http://ww	DC	16th & H S	20006	Washingt(	hotel	84117	The Hay-Adams
4335										
4336										

Both of the above snapshots show one extra record as row for column names is also included in the file.

Let us verify if the lookup table was correctly uploaded to Splunk.



The number of records match for both the files!

2. Verification of fields

1. review.txt

All the interesting fields collected from our data were properly loaded into splunk. The field used for timestamp is date. All the other fields are properly reflected.

```
{ [-]
   author: { [-]
     id: 631AAF84885D8AA48F4876100F92CEEB
     location: New York, New York, United States
     num_reviews: 1
     username: Danny K
   }
   date: December 20, 2012
   date_stayed: December 2012
   id: 147785077
   num_helpful_votes: 0
   offering_id: 120556
   ratings: { [-]
     cleanliness: 5
     location: 5
     overall: 5
     rooms: 5
     service: 5
     sleep_quality: 5
     value: 5
   }
   text: We arrived late at night and immediately were treated with everything we needed by NIkki and Adam. We were here for three nights and
there wasn't anything we needed that the hotel wasn't able to help us with. They even provide a whole bunch of complimentary travel items you
might forget (phone chargers, tooth brushes, contact solution etc.). They also have a very nice restaurant and bar in lobby which was very
enjoyable to visit. I highly recommend the George hotel.
   title: "Unbelievable customer service"
   via_mobile: false
}
```


Show as raw text

Interesting Fields

- a author.id 100+
- a author.location 100+
- a author.num_cities 100+
- a author.num_helpful_votes 100+
- a author.num_reviews 100+
- a author.num_type_reviews 100+
- a author.username 100+
- a date 100+
- # date_mday 31
- a date_month 5
- a date_stayed 100+
- a date_wday 7
- # date_year 1
- a date_zone 1
- a id 100+
- a index 1
- # linecount 1
- a num_helpful_votes 56
- a offering_id 100+
- a punct 100+
- a ratings.cleanliness 5
- a ratings.location 5
- a ratings.overall 5
- a ratings.rooms 5
- a ratings.service 5
- a ratings.sleep_quality 5
- a ratings.value 5
- a splunk_server 1
- a text 100+

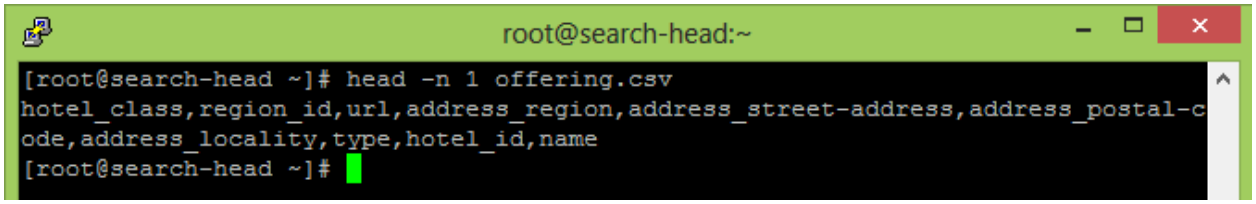
2. hotel_info.csv

To check if all the fields that are present in offering.csv are correctly loaded in Splunk, we first look at the fields in original file.

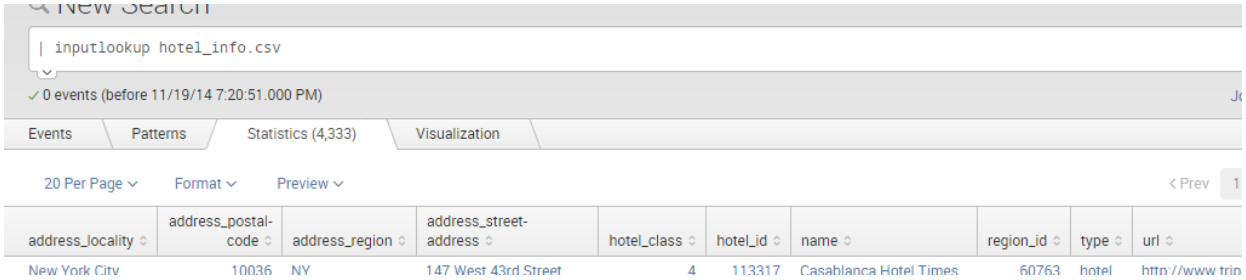
	A	B	C	D	E	F	G	H	I	J	K
1	hotel_class	region_id	url	address_region	address_street-address	address_postal-code	address_locality	type	hotel_id	name	
2	4	60763	http://w	NY	147 West 43rd Street	10036	New York City	hotel	113317	Casablanca Hotel T	

We also can have a look at the first line of this csv using command :

head -n 1 offering.csv



Let us check if all these fields are present in the uploaded lookup table in Splunk.



All the fields are properly reflected in hotel_info.csv.

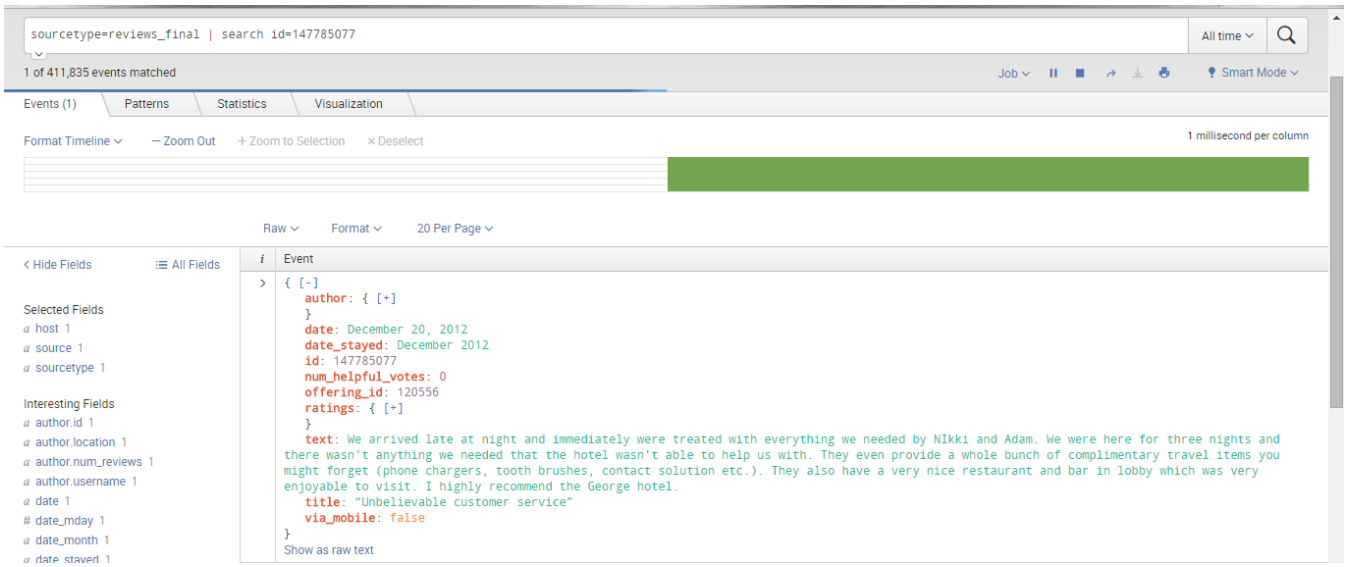
3. Verification of data

We queried the data on some unique fields like 'id' and cross verified if the same data exists for the same file within the original file.

1. review.txt

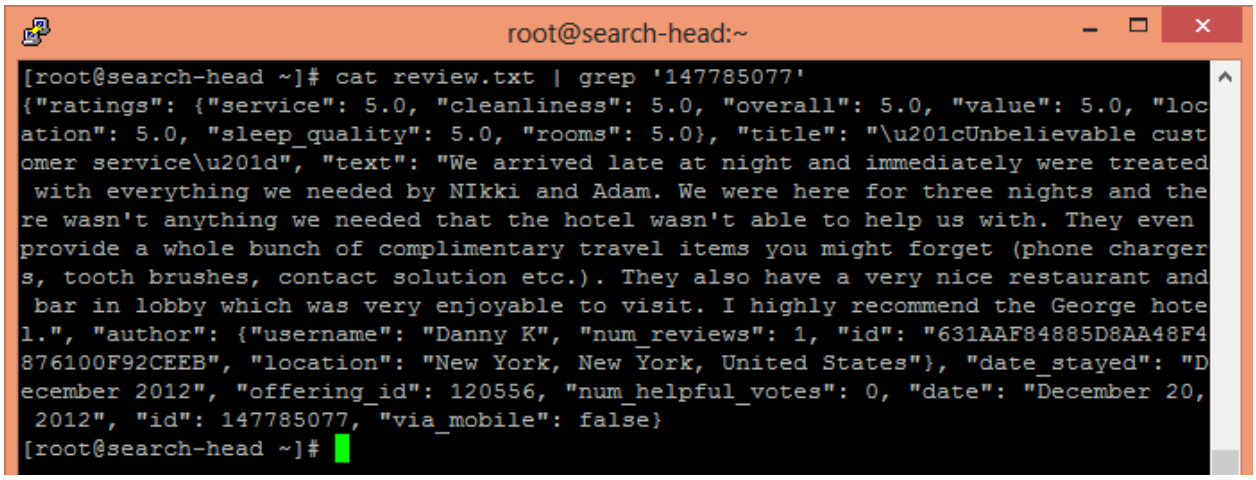
We checked the data for record with id = '147785077'

sourcetype=reviews_final | search id=147785077



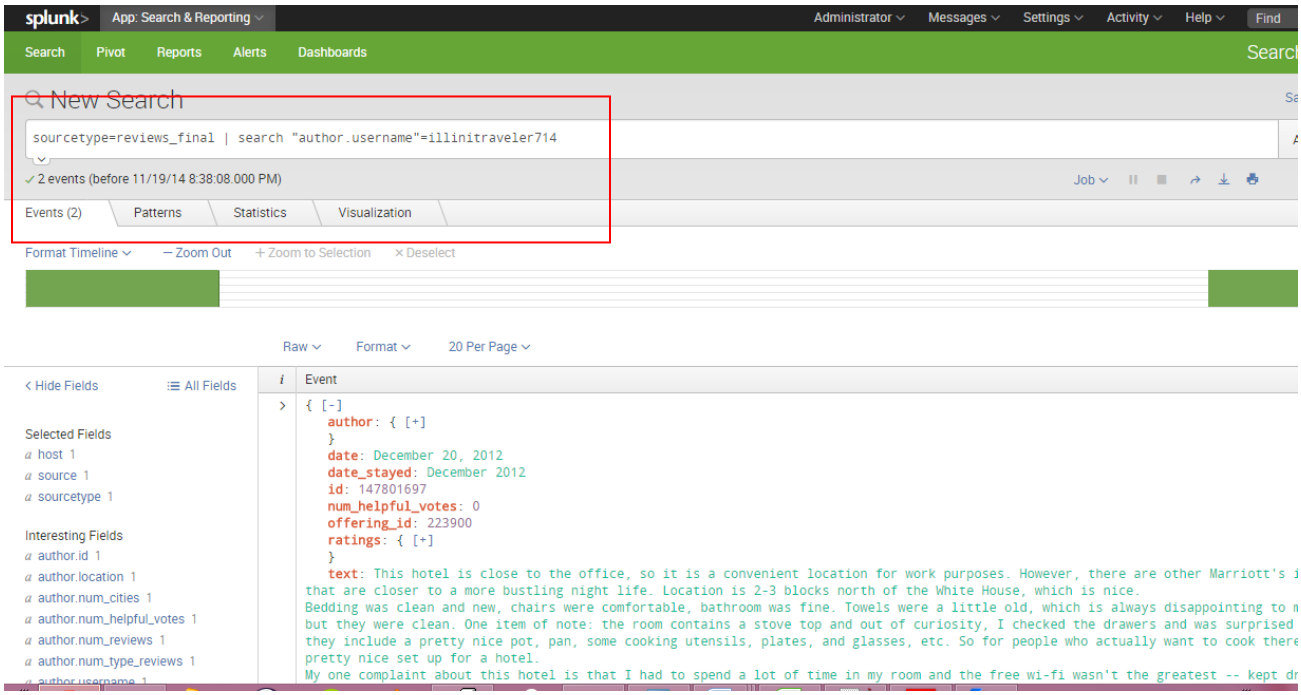
We compared this record with the record in original file.

cat review.txt | grep '147785077'



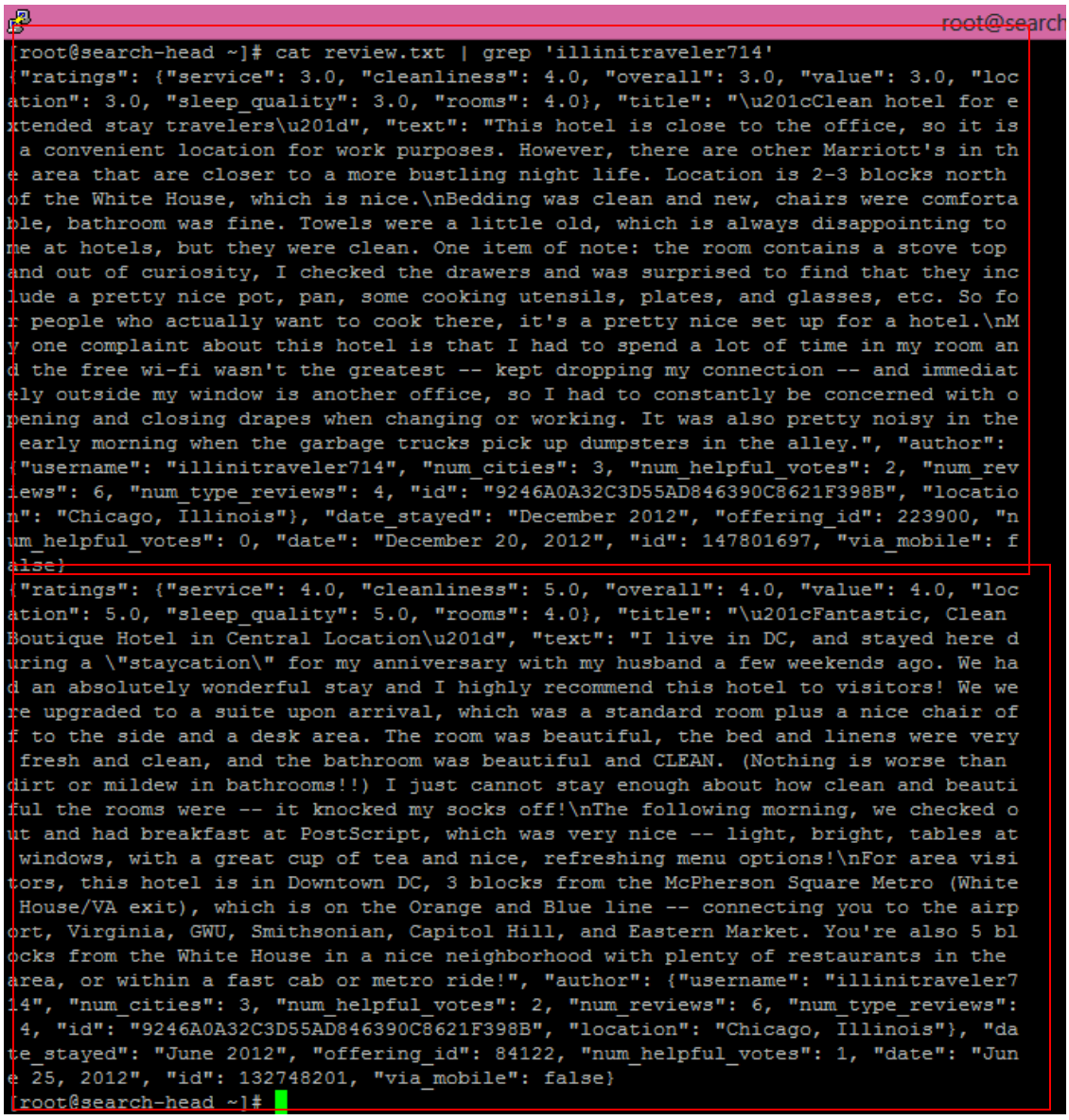
Next, we verified the reviews posted by the author with username = 'illinitraveler714'. We found two matching events in Splunk.

sourcetype=reviews_final | search "author.username"=illinitraveler714



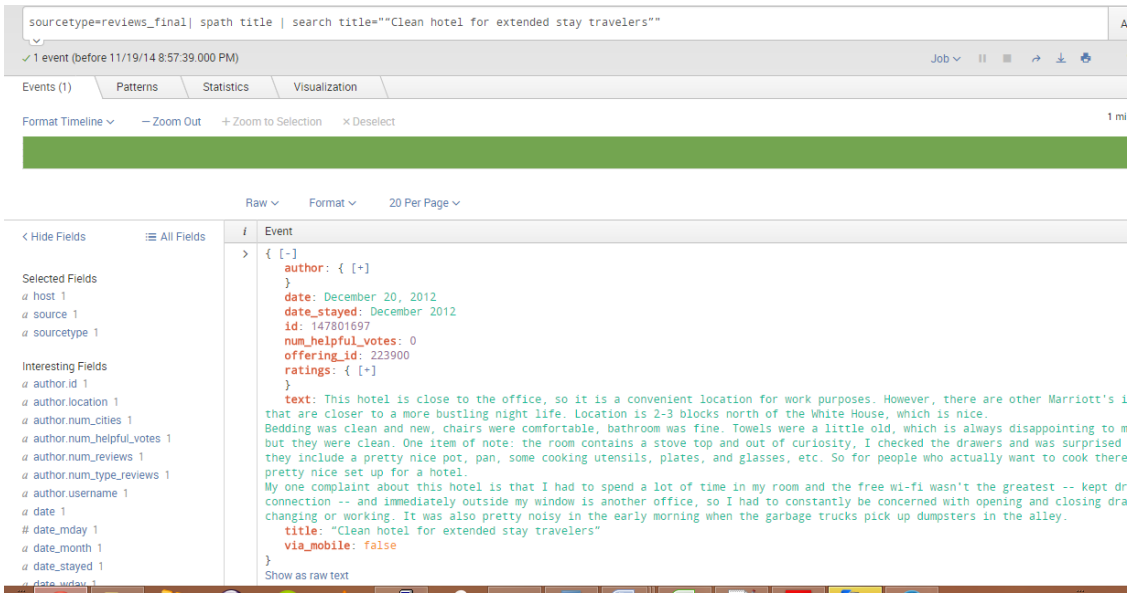
To cross-verify, we checked if these two records were the same in original file.

cat review.txt | grep 'illinitraveler714'

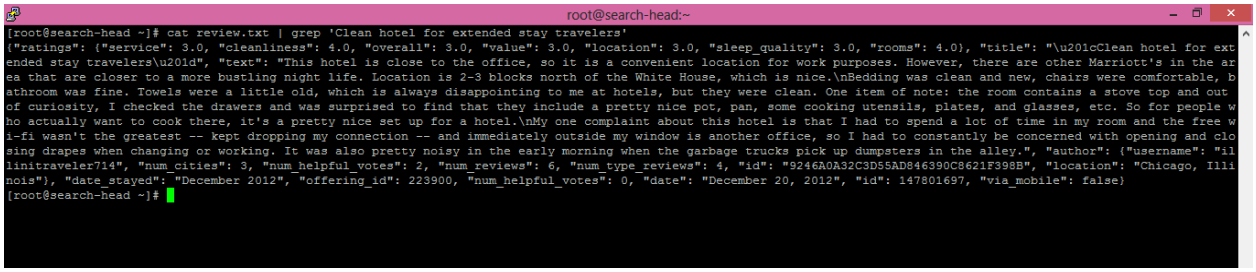


Next, we verified the reviews posted by the title = ' Clean hotel for extended stay travelers'. We found one matching event in Splunk.

sourcetype=reviews_final | spath title | search title=" Clean hotel for extended stay travelers "



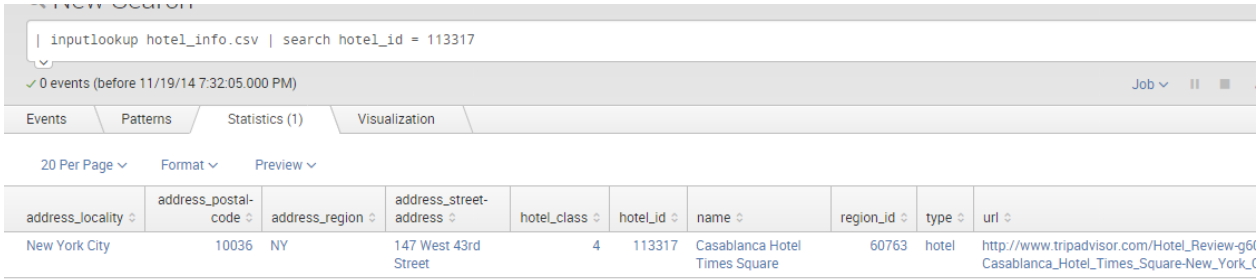
To cross-verify, we checked if these two records were the same in original file.



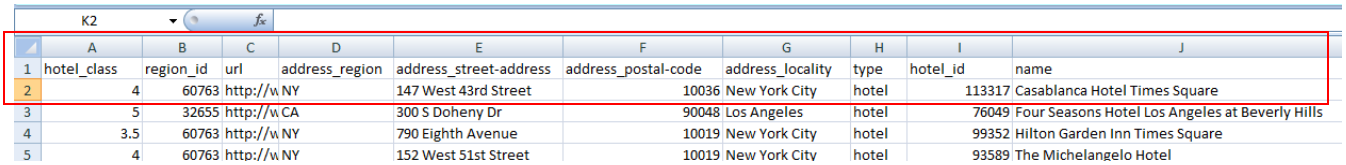
2. hotel_info.csv

We checked the record with hotel id = '113317' in Splunk :

| inputlookup hotel_info.csv | search hotel_id = 113317



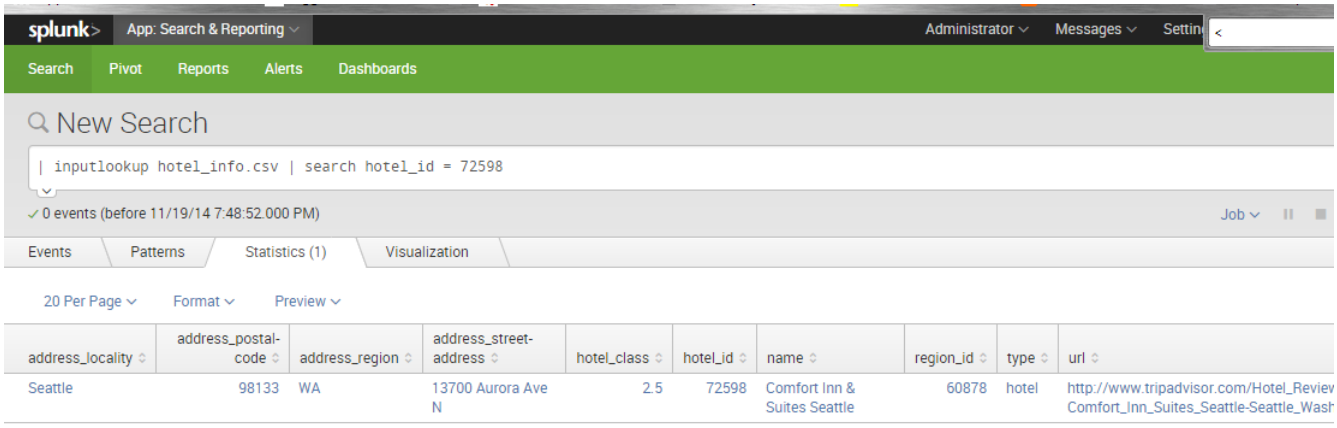
We verified if this was correct by checking it against the csv file.

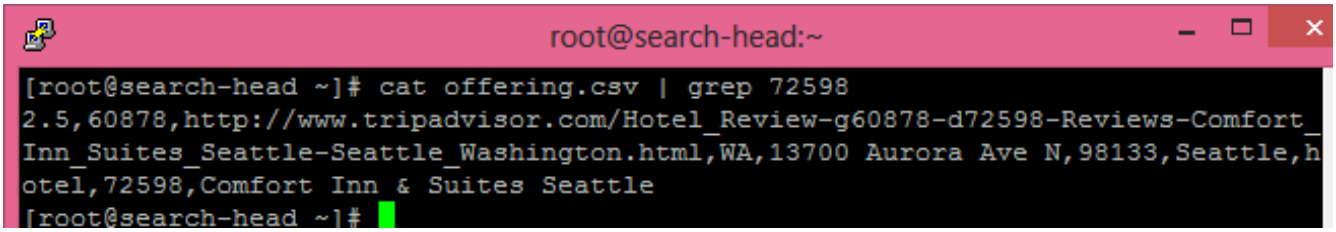


	A	B	C	D	E	F	G	H	I	J
1	hotel_class	region_id	url	address_region	address_street-address	address_postal-code	address_locality	type	hotel_id	name
2	4	60763	http://w	NY	147 West 43rd Street	10036	New York City	hotel	113317	Casablanca Hotel Times Square
3	5	32655	http://w	CA	300 S Doheny Dr	90048	Los Angeles	hotel	76049	Four Seasons Hotel Los Angeles at Beverly Hills
4	3.5	60763	http://w	NY	790 Eighth Avenue	10019	New York City	hotel	99352	Hilton Garden Inn Times Square
5	4	60763	http://w	NY	152 West 51st Street	10019	New York City	hotel	93589	The Michelangelo Hotel

For record with hotel_id = '72598'

| inputlookup hotel_info.csv | search hotel_id = 72598

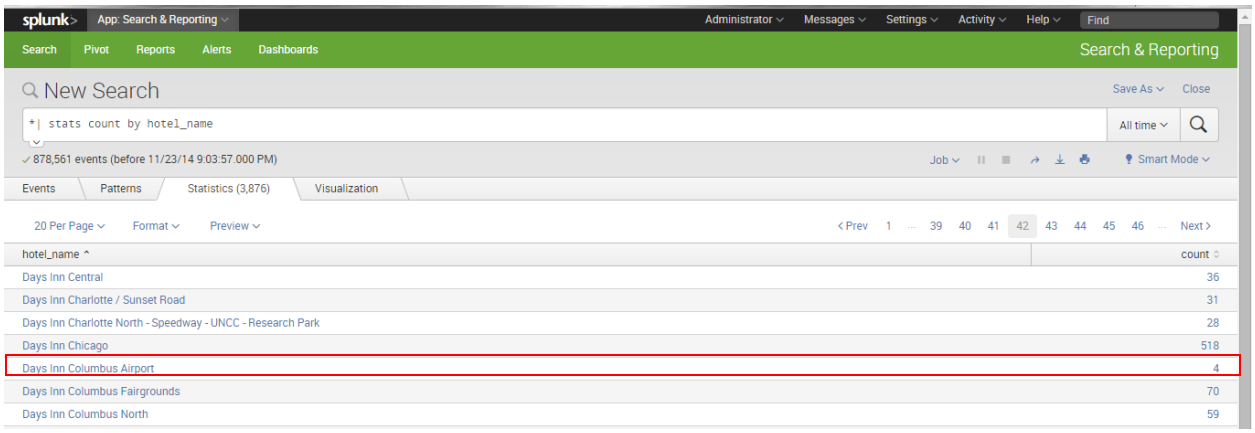




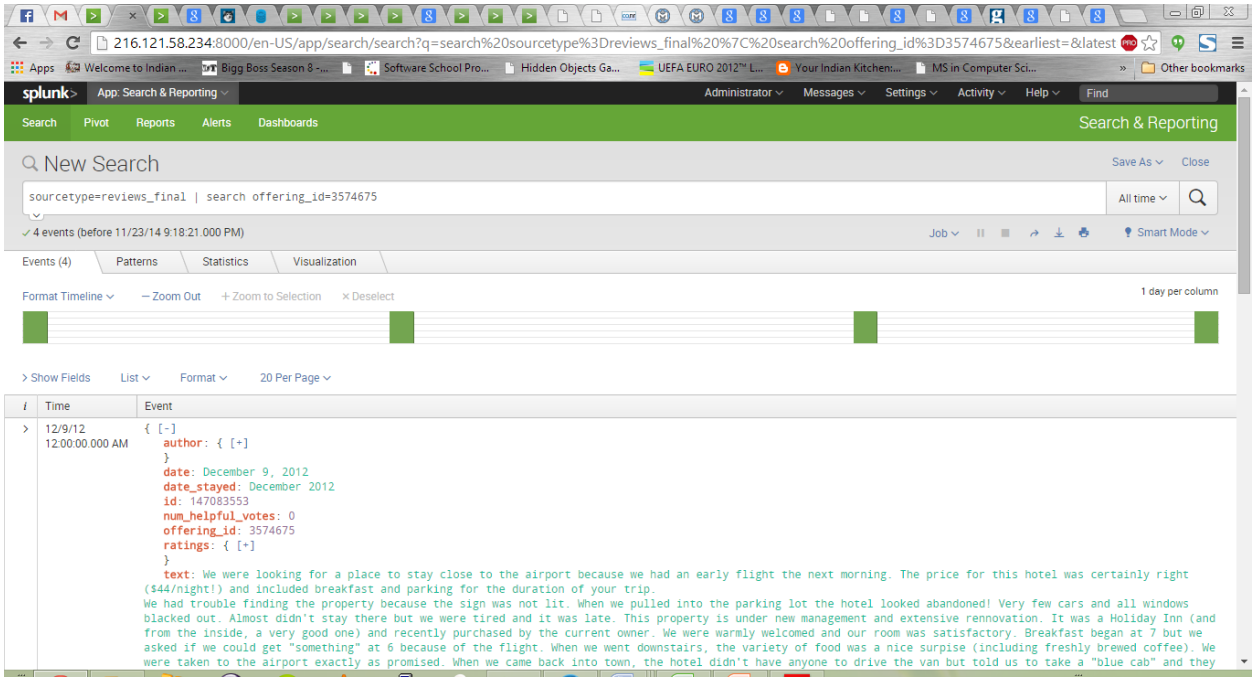
3. Check if lookup table works correctly with the table review.txt.

We executed a query to see if the number of the events with the hotel name are same as the count corresponding to hotel name.

*| stats count by hotel_name



Find the offering



On comparing it with the original file, the corresponding name matched with the hotel_id

Phase 2 - Building and Verifying Queries

1. Number of reviews for each year

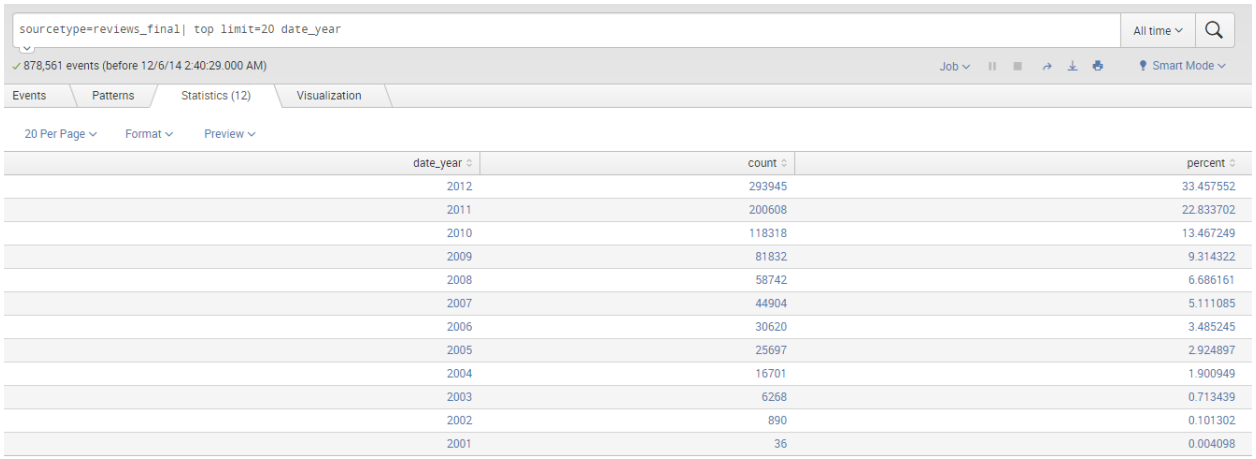
Objective of the query: To count the total number of reviews that were made in each year.

Description of the query :

Our dataset consists of reviews from years 2001 to 2012. We wanted to know the count of reviews in each year.

Query :

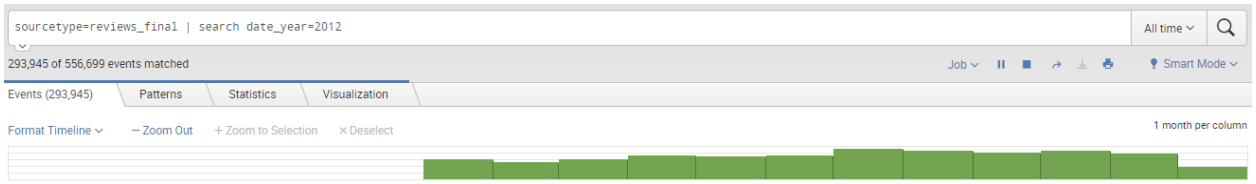
sourcetype=reviews_final| top limit=20 date_year



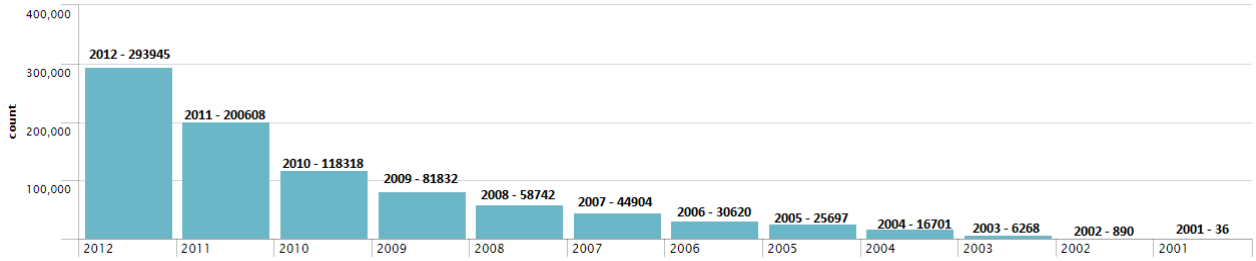
date_year	count	percent
2012	293945	33.457552
2011	200608	22.833702
2010	118318	13.467249
2009	81832	9.314322
2008	58742	6.686161
2007	44904	5.111085
2006	30620	3.485245
2005	25697	2.924897
2004	16701	1.900949
2003	6268	0.713439
2002	890	0.101302
2001	36	0.004098

Verification: We verified this data corresponding to year 2012.

sourcetype=reviews_final | search date_year = 2012



Visualization



2. State wise count for reviews

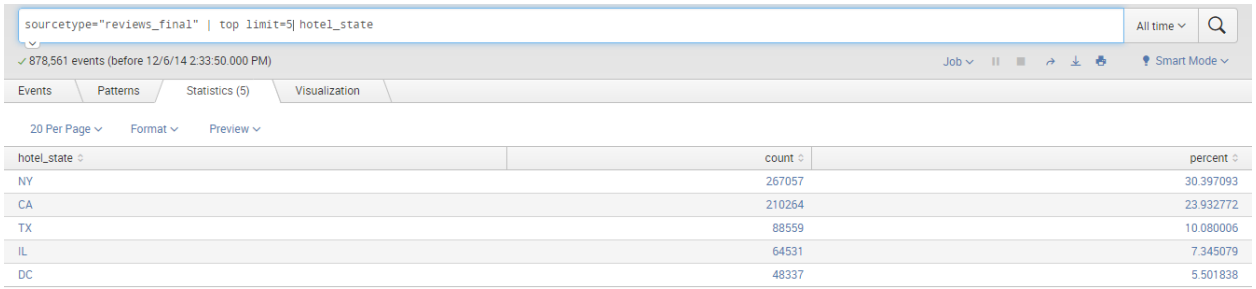
Objective of the query : To find top 5 states for with maximum number of reviews.

Description of the query :

The hotels in our dataset our spread over 17 states. We wanted to find the top five states with highest number of reviews that are recorded.

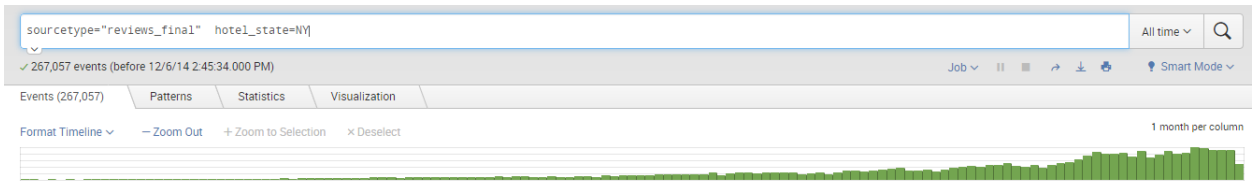
Query :

```
# sourcetype="reviews_final" | top limit=5 hotel_state
```



Verification: We verified the data for the state of NY.

```
# sourcetype="reviews_final"  hotel_state=NY
```



Visualization

3. City wise count for reviews

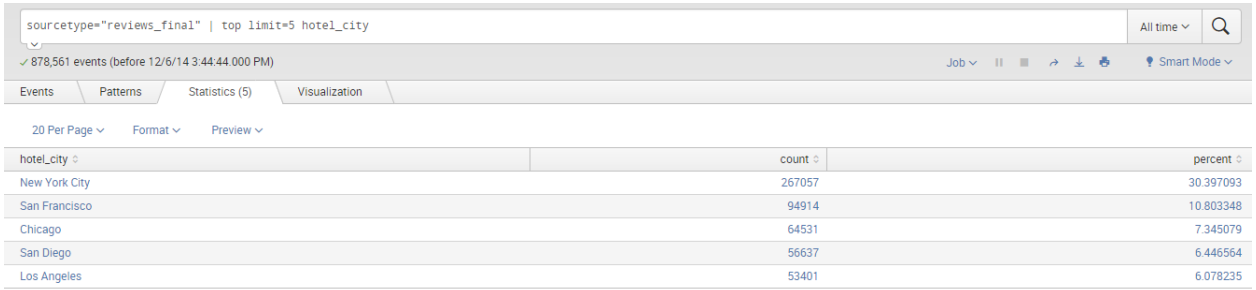
Objective of the query : To find top 5 cities for with maximum number of reviews.

Description of the query :

The hotels in our dataset our spread over multiple cities. We wanted to find the top five cities with highest number of reviews that are recorded.

Query :

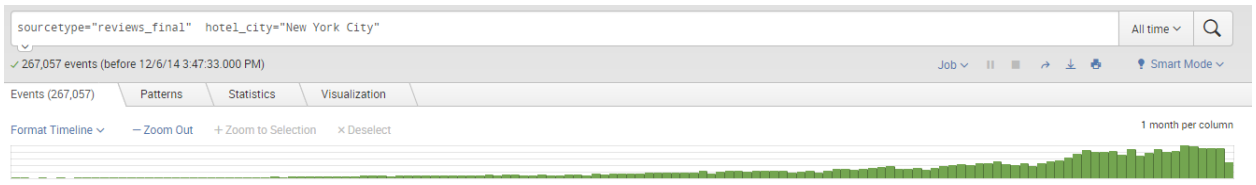
```
# sourcetype="reviews_final" | top limit=5 hotel_city
```



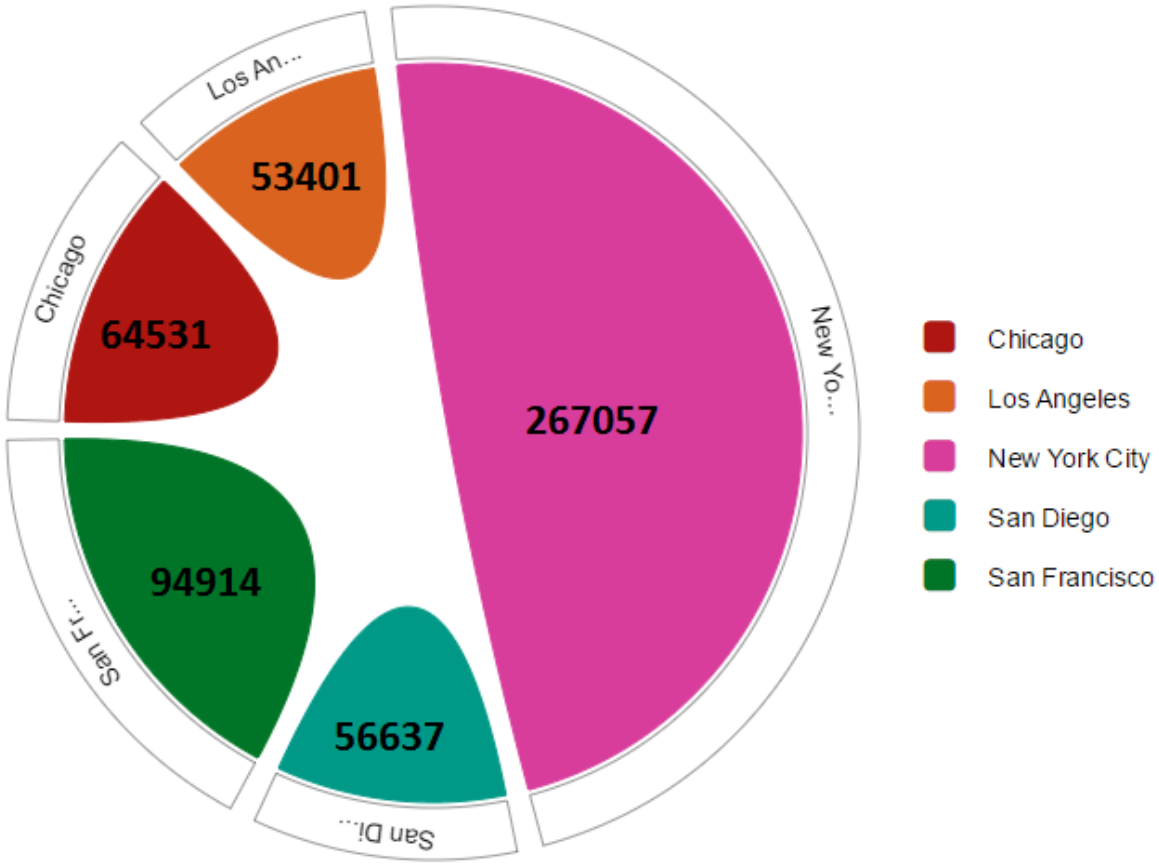
hotel_city	count	percent
New York City	267057	30.397093
San Francisco	94914	10.803348
Chicago	64531	7.345079
San Diego	56637	6.446564
Los Angeles	53401	6.078235

Verification: We verified the data for the New York City.

```
# sourcetype="reviews_final"  hotel_city="New York City"
```



Visualization

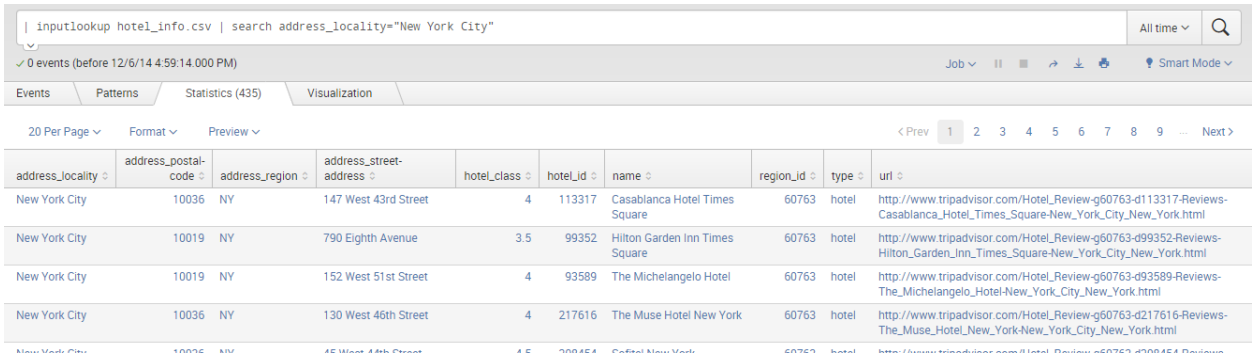


From the above two queries, for top five states and top five cities, we can see that the statistics for NY state and New York City are same. The number of reviews for both of them are 267057. So, to be sure that this statistics are correct, we verified it using some queries which are as follows :

Query 1 :

| inputlookup hotel_info.csv | search address_locality="New York City"

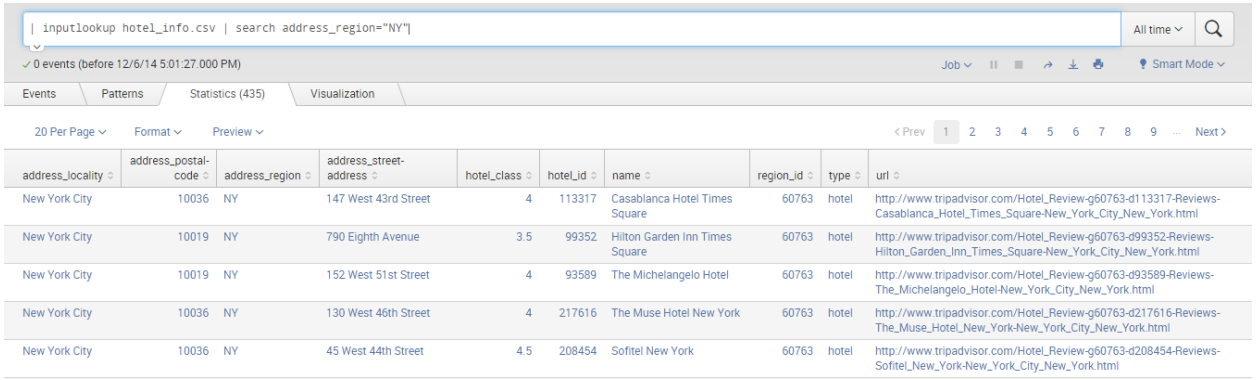
Result :



Query 2 :

| inputlookup hotel_info.csv | search address_region="NY"

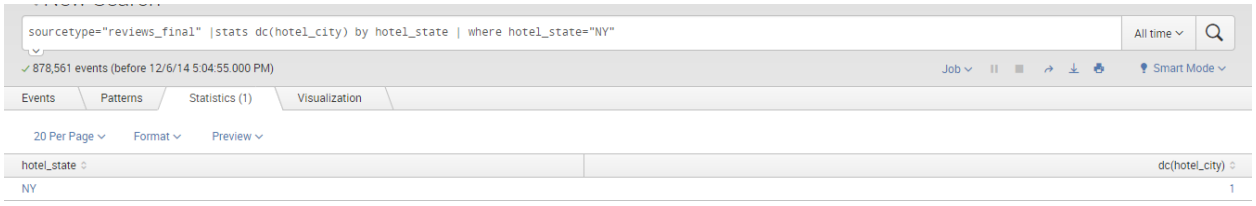
Result :



Query 3 :

sourcetype="reviews_final" | stats dc(hotel_city) by hotel_state | where hotel_state="NY"

Result :



4. To find the number of cities for which reviews are recorded in each state

Objective of the query : To find the number of cities in each state whose hotels are reviewed.

Description of the query :

From the statistic of New York City and NY state, we were curious to find if there are any other states for whom hotel from only one city are reviewed.

Query:

sourcetype="reviews_final" |stats dc(hotel_city) by hotel_state

```
sourcetype="reviews_final" |stats dc(hotel_city) by hotel_state
```

All time

✓ 878,561 events (before 12/6/14 5:21:47.000 PM)

Job | Smart Mode

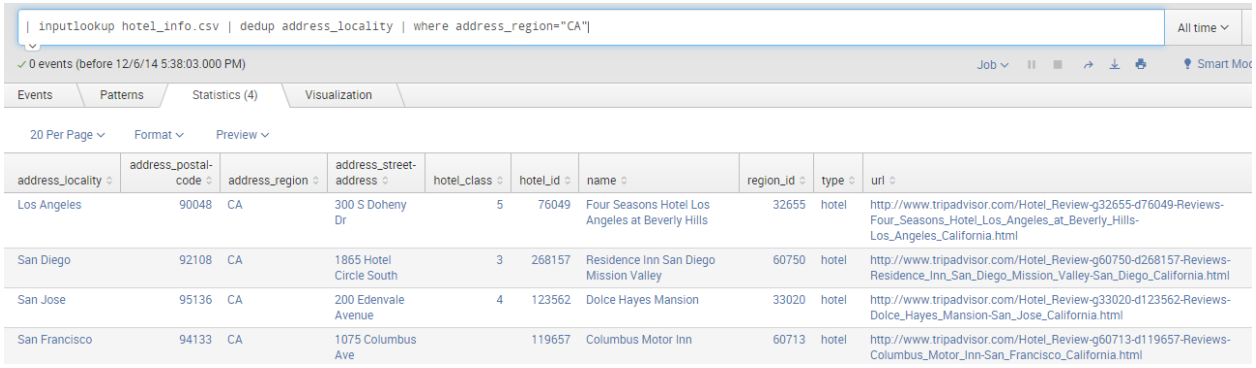
Events | Patterns | Statistics (17) | Visualization

20 Per Page | Format | Preview

hotel_state	dc(hotel_city)
TX	6
CA	4
AZ	1
CO	1
DC	1
FL	1
IL	1
IN	1
MA	1
MD	1
MI	1
NC	1
NY	1
OH	1
PA	1

Verification Query 1:

| inputlookup hotel_info.csv | dedup address_locality | where address_region="CA"



```
| inputlookup hotel_info.csv | dedup address_locality | where address_region="CA"
```

All time

✓ 0 events (before 12/6/14 5:38:03.000 PM)

Job | Smart Mod

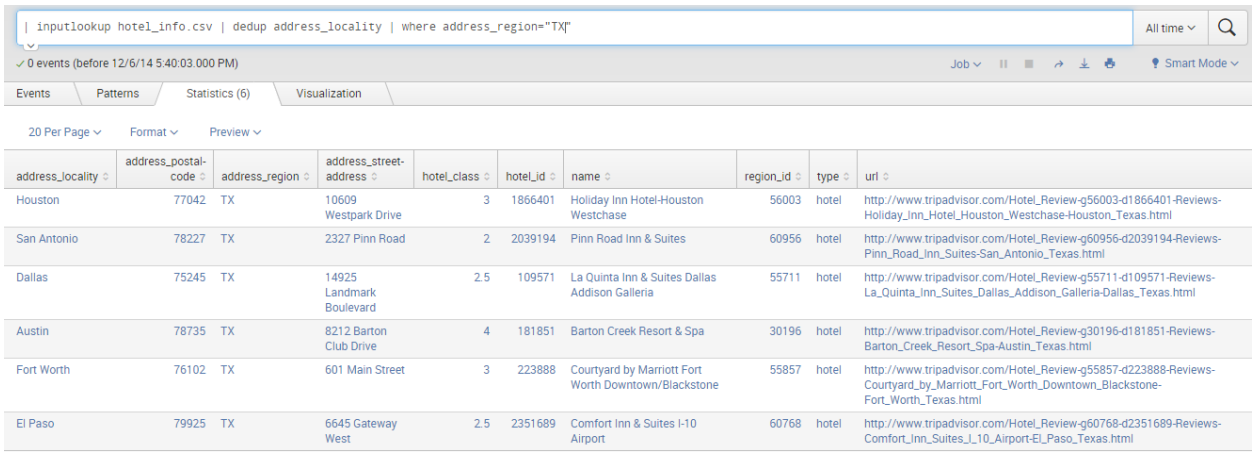
Events | Patterns | Statistics (4) | Visualization

20 Per Page | Format | Preview

address_locality	address_postal-code	address_region	address_street-address	hotel_class	hotel_id	name	region_id	type	url
Los Angeles	90048	CA	300 S Doheny Dr	5	76049	Four Seasons Hotel Los Angeles at Beverly Hills	32655	hotel	http://www.tripadvisor.com/Hotel_Review-g32655-d76049-Reviews-Four_Seasons_Hotel_Los_Angeles_at_Beverly_Hills-Los_Angeles_California.html
San Diego	92108	CA	1865 Hotel Circle South	3	268157	Residence Inn San Diego Mission Valley	60750	hotel	http://www.tripadvisor.com/Hotel_Review-g60750-d268157-Reviews-Residence_Inn_San_Diego_Mission_Valley-San_Diego_California.html
San Jose	95136	CA	200 Edenvale Avenue	4	123562	Dolce Hayes Mansion	33020	hotel	http://www.tripadvisor.com/Hotel_Review-g33020-d123562-Reviews-Dolce_Hayes_Mansion-San_Jose_California.html
San Francisco	94133	CA	1075 Columbus Ave		119657	Columbus Motor Inn	60713	hotel	http://www.tripadvisor.com/Hotel_Review-g60713-d119657-Reviews-Columbus_Motor_Inn-San_Francisco_California.html

Verification Query 2:

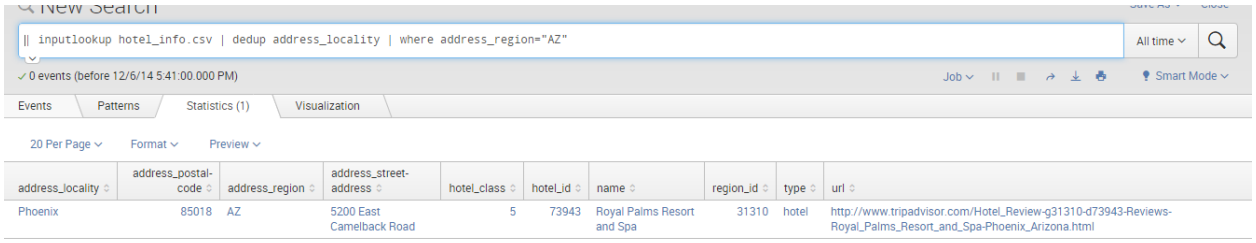
| inputlookup hotel_info.csv | dedup address_locality | where address_region="TX"



address_locality	address_postal-code	address_region	address_street-address	hotel_class	hotel_id	name	region_id	type	url
Houston	77042	TX	10609 Westpark Drive	3	1866401	Holiday Inn Hotel-Houston Westchase	56003	hotel	http://www.tripadvisor.com/Hotel_Review-g56003-d1866401-Reviews-Holiday_Inn_Hotel_Houston_Westchase-Houston_Texas.html
San Antonio	78227	TX	2327 Pinn Road	2	2039194	Pinn Road Inn & Suites	60956	hotel	http://www.tripadvisor.com/Hotel_Review-g60956-d2039194-Reviews-Pinn_Road_Inn_Suites-San_Antonio_Texas.html
Dallas	75245	TX	14925 Landmark Boulevard	2.5	109571	La Quinta Inn & Suites Dallas Addison Galleria	55711	hotel	http://www.tripadvisor.com/Hotel_Review-g55711-d109571-Reviews-La_Quinta_Inn_Suites_Dallas_Addison_Galleria-Dallas_Texas.html
Austin	78735	TX	8212 Barton Club Drive	4	181851	Barton Creek Resort & Spa	30196	hotel	http://www.tripadvisor.com/Hotel_Review-g30196-d181851-Reviews-Barton_Creek_Resort_Spa-Austin_Texas.html
Fort Worth	76102	TX	601 Main Street	3	223888	Courtyard by Marriott Fort Worth Downtown/Blackstone	55857	hotel	http://www.tripadvisor.com/Hotel_Review-g55857-d223888-Reviews-Courtyard_by_Marriott_Fort_Worth_Downtown_Blackstone-Fort_Worth_Texas.html
El Paso	79925	TX	6645 Gateway West	2.5	2351689	Comfort Inn & Suites I-10 Airport	60768	hotel	http://www.tripadvisor.com/Hotel_Review-g60768-d2351689-Reviews-Comfort_Inn_Suites_I_10_Airport-El_Paso_Texas.html

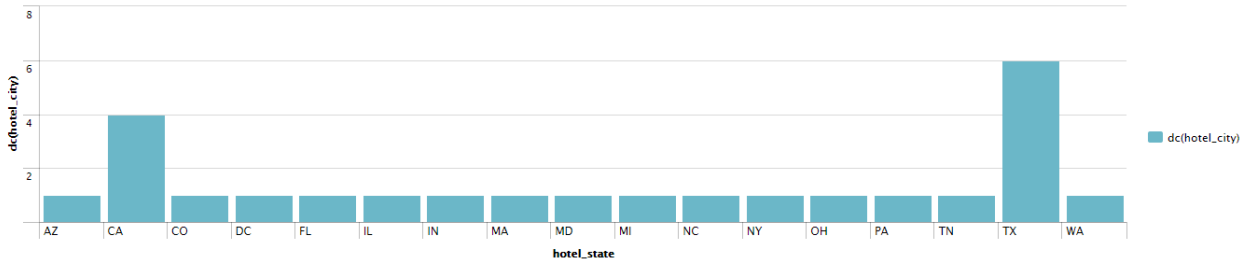
Verification Query 3:

| inputlookup hotel_info.csv | dedup address_locality | where address_region="AZ"



address_locality	address_postal-code	address_region	address_street-address	hotel_class	hotel_id	name	region_id	type	url
Phoenix	85018	AZ	5200 East Camelback Road	5	73943	Royal Palms Resort and Spa	31310	hotel	http://www.tripadvisor.com/Hotel_Review-g31310-d73943-Reviews-Royal_Palms_Resort_and_Spa-Phoenix_Arizona.html

Visualization



5. The region that has the most number of hotel reviews according to postal code

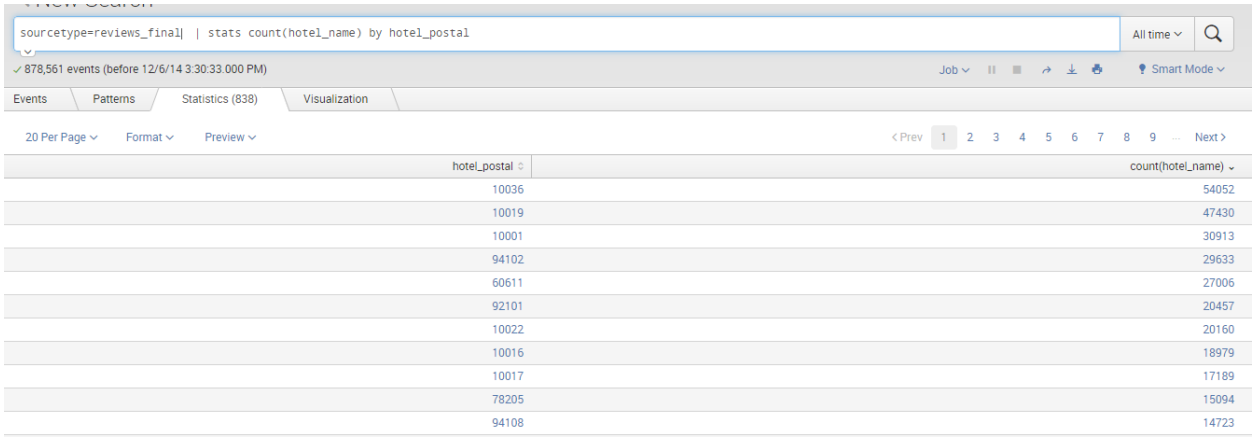
Objective of the query : To find the postal regions for which the highest number of reviews are recorded.

Description of the query :

Every state is divided into multiple postal codes. The objective of this query is to find the postal region for which highest number of reviews are recorded.

Query :

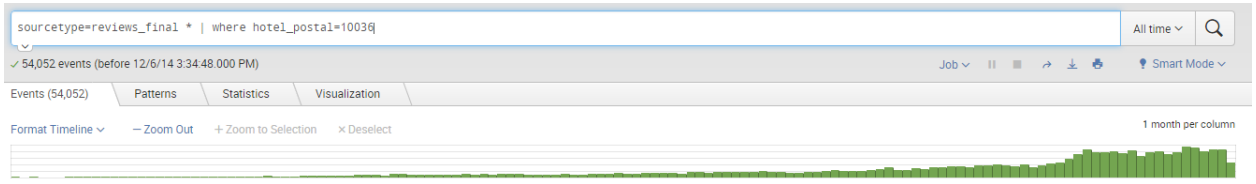
```
# sourcetype=reviews_final  | stats count(hotel_name) by hotel_postal
```



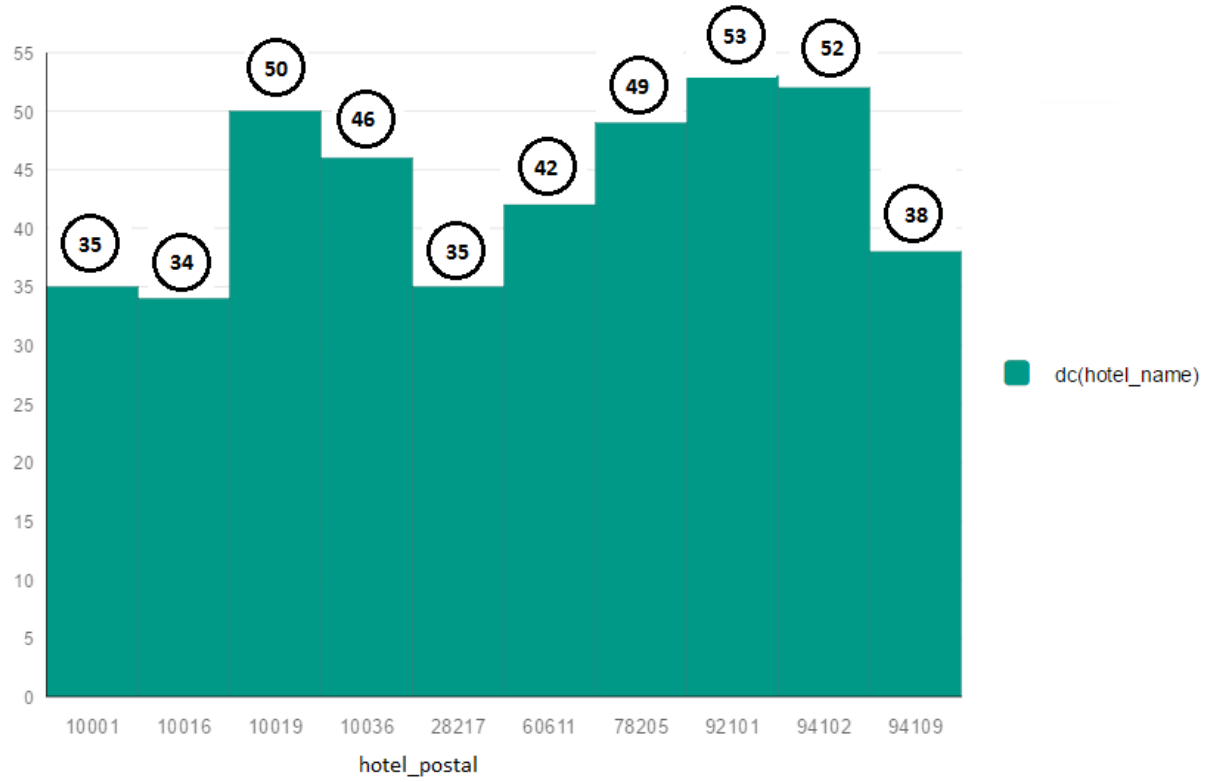
hotel_postal	count(hotel_name)
10036	54052
10019	47430
10001	30913
94102	29633
60611	27006
92101	20457
10022	20160
10016	18979
10017	17189
78205	15094
94108	14723

Verification : The total number of reviews for hotels having postal code as 10036 were verified.

```
# sourcetype=reviews_final * | where hotel_postal=10036
```

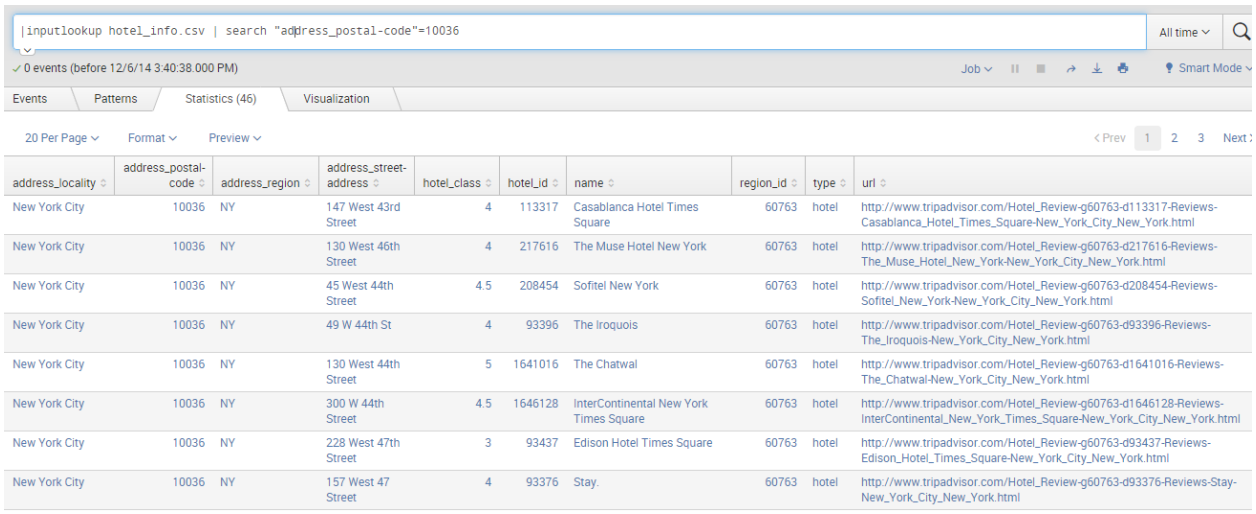


Visualization



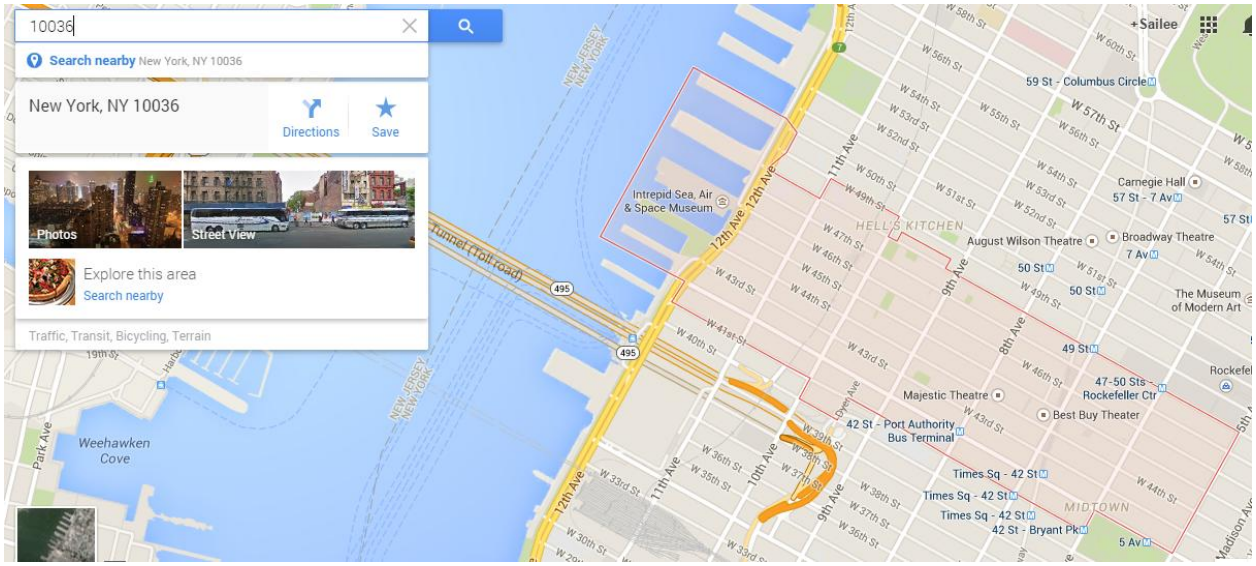
There are 46 hotels that have this postal code :

|inputlookup hotel_info.cv | search "adress_postal-code"=10036



address_locality	address_postal-code	address_region	address_street-address	hotel_class	hotel_id	name	region_id	type	url
New York City	10036	NY	147 West 43rd Street	4	113317	Casablanca Hotel Times Square	60763	hotel	http://www.tripadvisor.com/Hotel_Review-g60763-d113317-Reviews-Casablanca_Hotel_Times_Square-New_York_City_New_York.html
New York City	10036	NY	130 West 46th Street	4	217616	The Muse Hotel New York	60763	hotel	http://www.tripadvisor.com/Hotel_Review-g60763-d217616-Reviews-The_Muse_Hotel_New_York-New_York_City_New_York.html
New York City	10036	NY	45 West 44th Street	4.5	208454	Sofitel New York	60763	hotel	http://www.tripadvisor.com/Hotel_Review-g60763-d208454-Reviews-Sofitel_New_York-New_York_City_New_York.html
New York City	10036	NY	49 W 44th St	4	93396	The Iroquois	60763	hotel	http://www.tripadvisor.com/Hotel_Review-g60763-d93396-Reviews-The_Iroquois-New_York_City_New_York.html
New York City	10036	NY	130 West 44th Street	5	1641016	The Chatwal	60763	hotel	http://www.tripadvisor.com/Hotel_Review-g60763-d1641016-Reviews-The_Chatwal-New_York_City_New_York.html
New York City	10036	NY	300 W 44th Street	4.5	1646128	InterContinental New York Times Square	60763	hotel	http://www.tripadvisor.com/Hotel_Review-g60763-d1646128-Reviews-InterContinental_New_York_Times_Square-New_York_City_New_York.html
New York City	10036	NY	228 West 47th Street	3	93437	Edison Hotel Times Square	60763	hotel	http://www.tripadvisor.com/Hotel_Review-g60763-d93437-Reviews-Edison_Hotel_Times_Square-New_York_City_New_York.html
New York City	10036	NY	157 West 47 Street	4	93376	Stay.	60763	hotel	http://www.tripadvisor.com/Hotel_Review-g60763-d93376-Reviews-Stay-New_York_City_New_York.html

This is the postal code for city of New York.



6. Month that has the most number of reviews in all the years

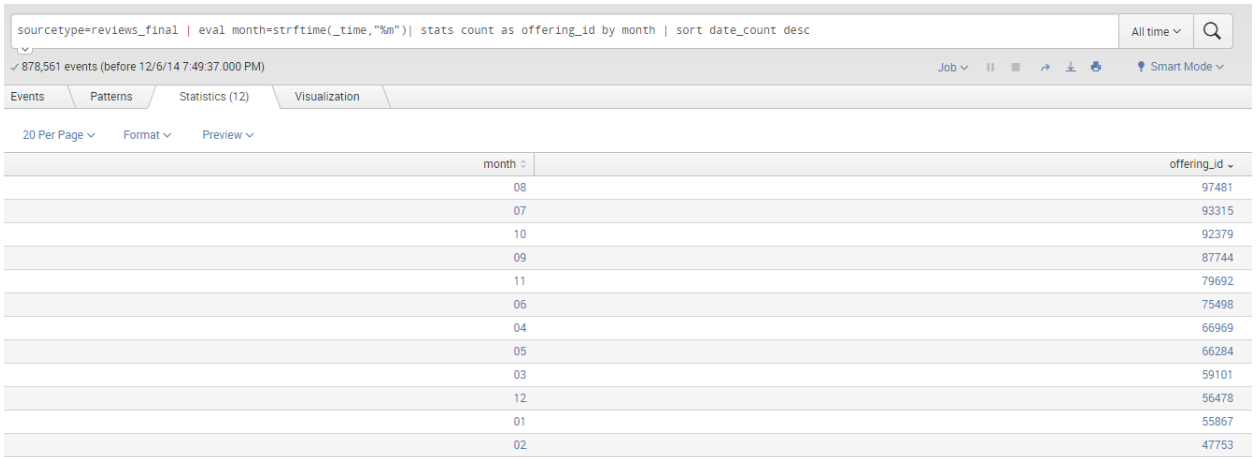
Objective of the query : To find the month for which the total number of reviews recorded are highest in all the years.

Description of the query:

The intent of this query is to find which month has the highest total count for number of reviews recorded.

Query :

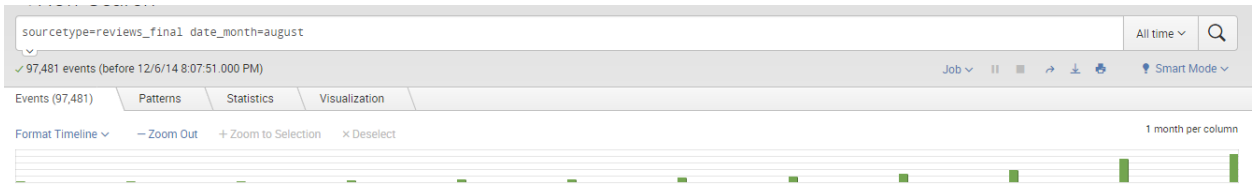
```
# sourcetype=reviews_final | eval month=strftime(_time,"%m")| stats count as offering_id by month | sort date_count desc
```



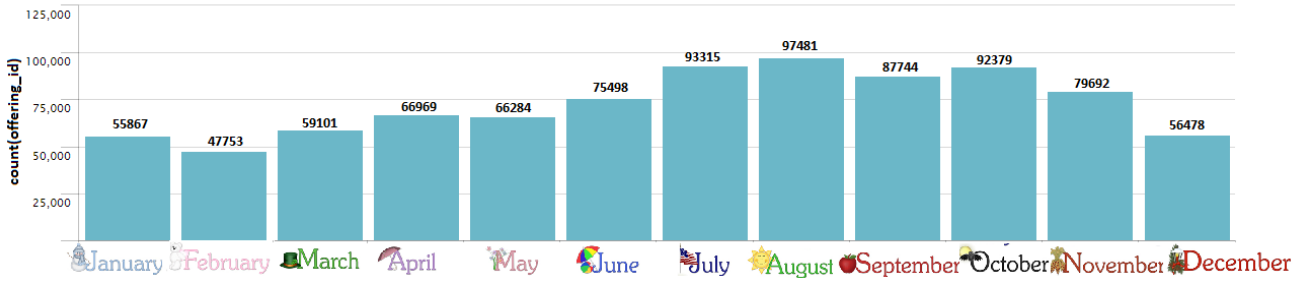
month	offering_id
08	97481
07	93315
10	92379
09	87744
11	79692
06	75498
04	66969
05	66284
03	59101
12	56478
01	55867
02	47753

Verification: We verified this for the month of August.

sourcetype=reviews_final date_month=august



Visualization



7. Month wise count of reviews for years 2010 to 2012

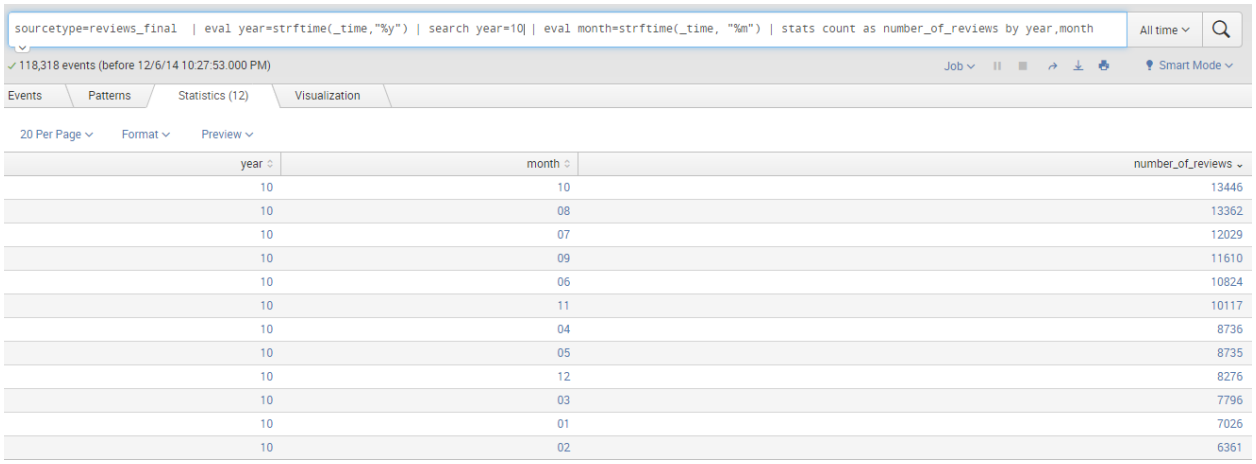
Objective of the query : To find the number of reviews recorded for each month for the years 2010, 2011 and 2012.

Description of the query :

To know if there was any specific month in each year where most reviews were recorded.

For year 2010:

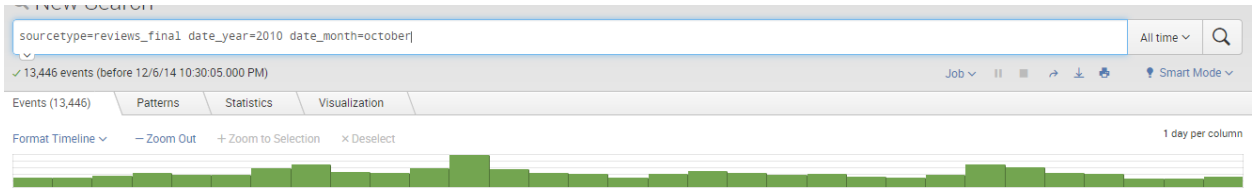
sourcetype=reviews_final | eval year=strftime(_time,"%y") | search year=10 | eval month=strftime(_time, "%m") | stats count as number_of_reviews by year,month



year	month	number_of_reviews
10	10	13446
10	08	13362
10	07	12029
10	09	11610
10	06	10824
10	11	10117
10	04	8736
10	05	8735
10	12	8276
10	03	7796
10	01	7026
10	02	6361

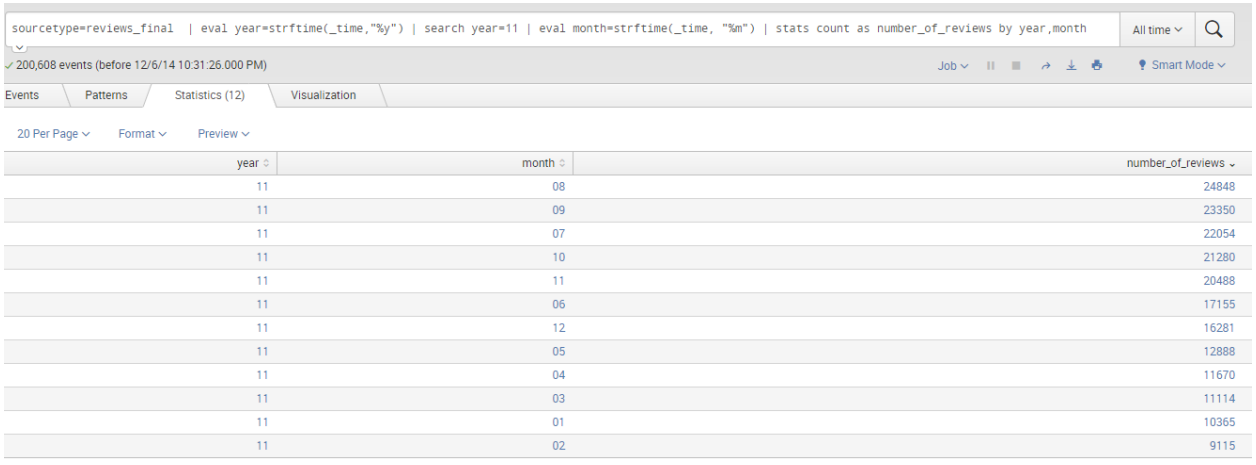
Verification :

sourcetype=reviews_final date_year=2010 date_month=october



For year 2011:

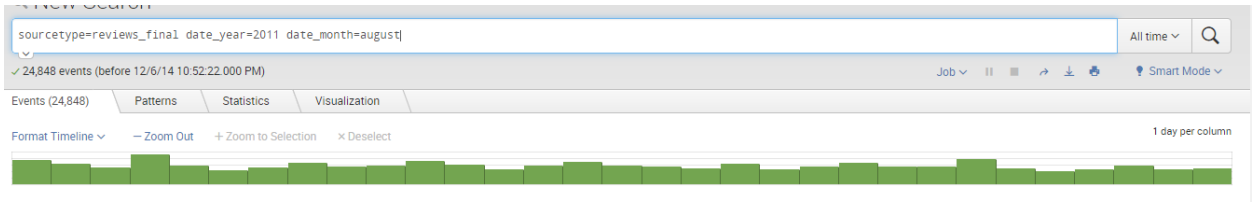
sourcetype=reviews_final | eval year=strftime(_time,"%y") | search year=11 | eval month=strftime(_time, "%m") | stats count as number_of_reviews by year,month



year	month	number_of_reviews
11	08	24848
11	09	23350
11	07	22054
11	10	21280
11	11	20488
11	06	17155
11	12	16281
11	05	12888
11	04	11670
11	03	11114
11	01	10365
11	02	9115

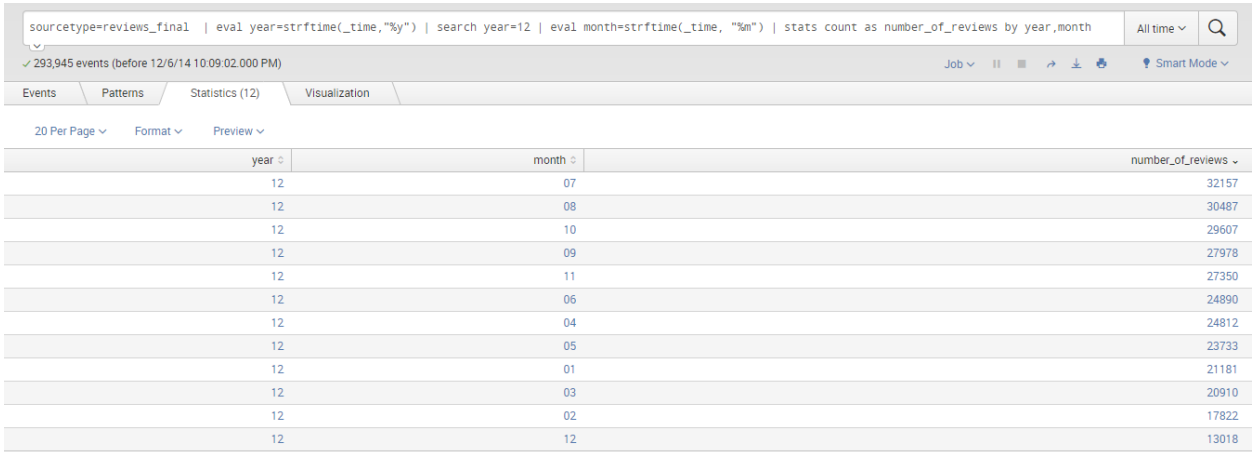
Verification:

sourcetype=reviews_final date_year=2011 date_month=august



For year 2012:

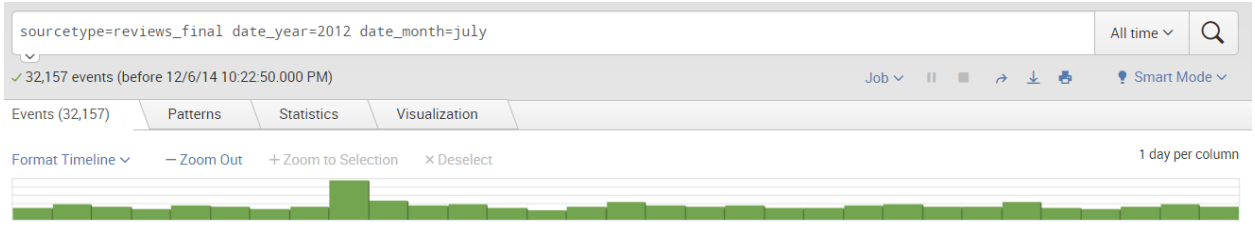
sourcetype=reviews_final | eval year=strftime(_time,"%y") | search year=12 | eval month=strftime(_time, "%m") | stats count as number_of_reviews by year,month



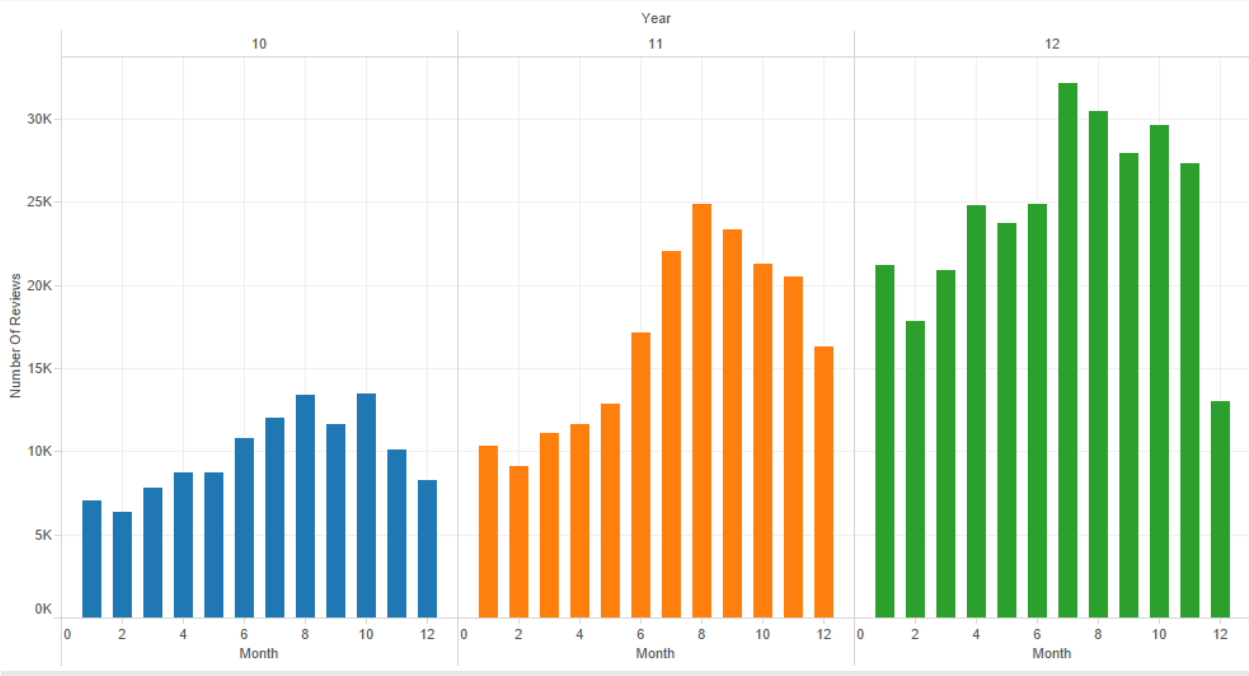
year	month	number_of_reviews
12	07	32157
12	08	30487
12	10	29607
12	09	27978
12	11	27350
12	06	24890
12	04	24812
12	05	23733
12	01	21181
12	03	20910
12	02	17822
12	12	13018

Verification:

sourcetype=reviews_final date_year=2012 date_month=july



Visualization



Result :

The most number of reviews in each year are recorded for the months of July, August, September and October.

8. Top 10 hotels that have the most reviews

Objective of the query: To find the hotels that have the most number of reviews.

Description of the query:

The intent of this query is to find the hotels that have the highest number of reviews from the customers.

Query:

```
# sourcetype=reviews_final* | stats count(id) by offering_id  | lookup hotel_info.csv hotel_id as offering_id output name | sort limit=5 -count(id) | fields name, count(id) | rename count(id) as number_reviews
```

```
sourcetype=reviews_final* | stats count(id) by offering_id  | lookup hotel_info.csv hotel_id as offering_id output name | sort limit=10 -count(id) | fields name, count(id) | rename count(id) as number_reviews
```

All time

878,561 events (before 12/7/14 12:19:36.000 AM)

Job | Smart Mode

Events | Patterns | Statistics (10) | Visualization

20 Per Page | Format | Preview

name	number_reviews
Hotel Pennsylvania New York	21824
Park Central	16036
The New Yorker Hotel	14904
Waldorf Astoria New York	14136
Hudson New York	13540
The Roosevelt Hotel	12872
Affinia Manhattan	12680
Edison Hotel Times Square	12136
Hilton New York	12016
Sofitel New York	11592

Verification:

sourcetype=reviews_final* | stats count(id) by offering_id | lookup hotel_info.csv hotel_id as offering_id output name | search name = "Hotel Pennsylvania New York"

```
sourcetype=reviews_final* | stats count(id) by offering_id  | lookup hotel_info.csv hotel_id as offering_id output name | search name = "Hotel Pennsylvania New York"
```

878,561 events (before 12/7/14 12:22:23.000 AM)

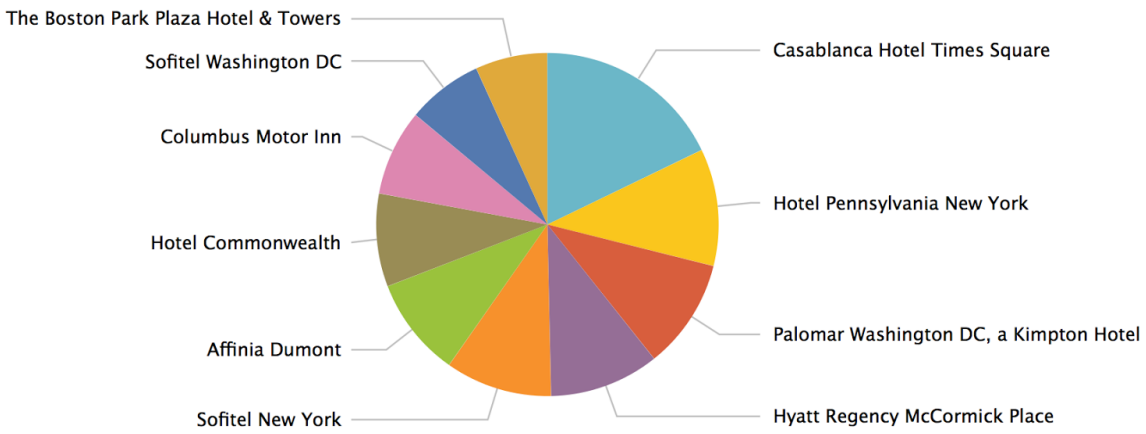
Job

Events | Patterns | Statistics (1) | Visualization

20 Per Page | Format | Preview

offering_id	count(id)	name
214197	21824	Hotel Pennsylvania New York

Visualization



9. Top 10 hotels that have the most positive reviews(2.5 + on overall rating)

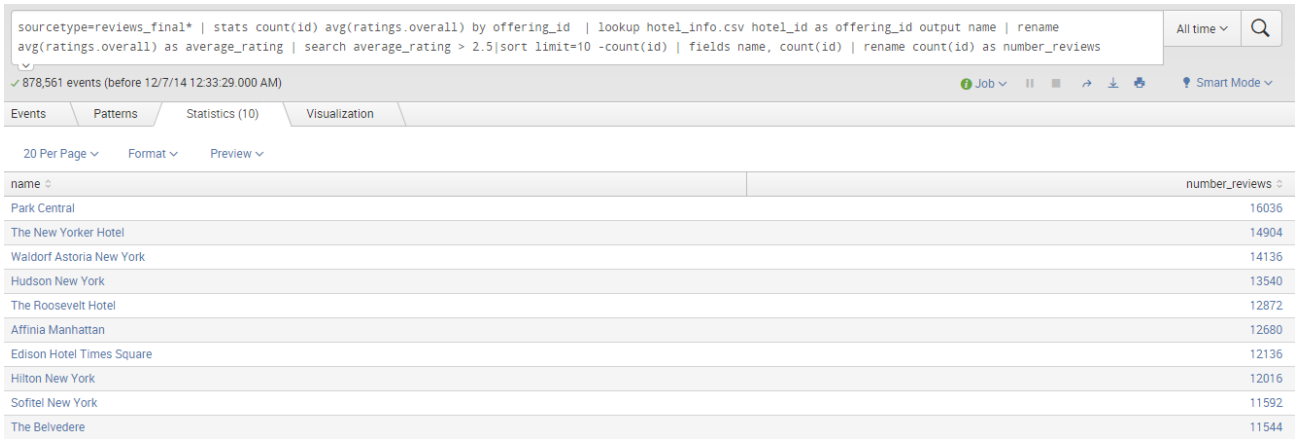
Objective of the query: To find the hotels that have the most number of positive reviews.

Description of the query:

The intent of this query is to find the hotels that have the highest number of positive reviews from the customers.

Query:

sourcetype=reviews_final* | stats count(id) avg(ratings.overall) by offering_id | lookup hotel_info.csv hotel_id as offering_id output name | rename avg(ratings.overall) as average_rating | search average_rating > 2.5|sort limit=10 -count(id) | fields name, count(id) | rename count(id) as number_reviews



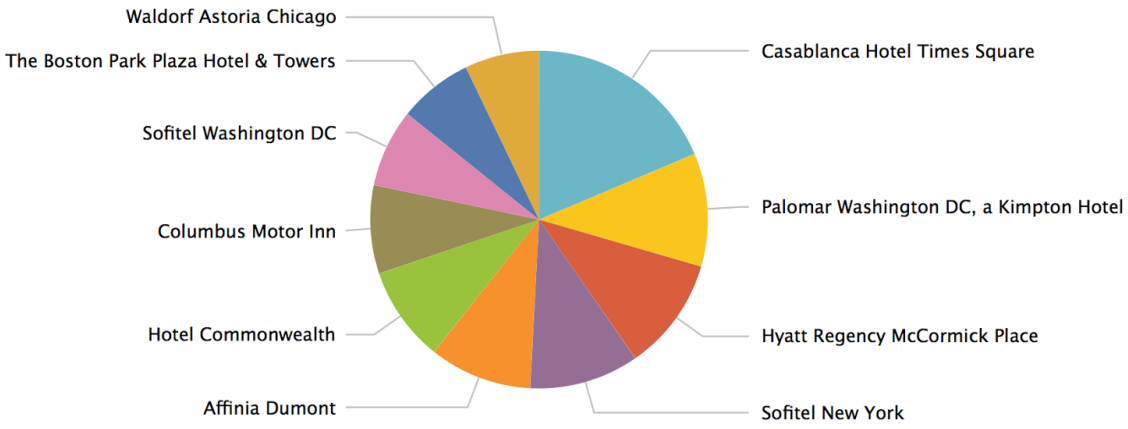
name	number_reviews
Park Central	16036
The New Yorker Hotel	14904
Waldorf Astoria New York	14136
Hudson New York	13540
The Roosevelt Hotel	12872
Affinia Manhattan	12680
Edison Hotel Times Square	12136
Hilton New York	12016
Sofitel New York	11592
The Belvedere	11544

Verification:

sourcetype=reviews_final* | stats count(id) avg(ratings.overall) by offering_id | lookup hotel_info.csv hotel_id as offering_id output name | search name="Park Central”

offering_id	count(id)	avg(ratings.overall)	name
93520	16036	3.496383	Park Central

Visualization



10. Hotels that have the most negative reviews(2.5 - on overall rating)

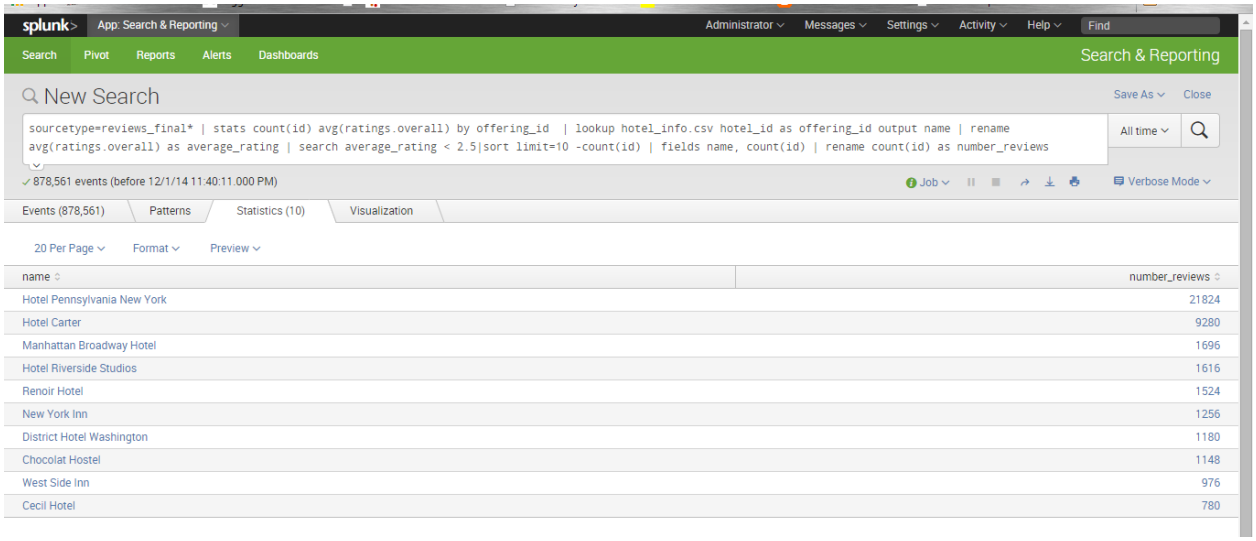
Objective of the query: To find the hotels that have the most number of reviews.

Description of the query:

The intent of this query is to find the hotels that have the highest number of reviews from the customers.

Query:

```
# sourcetype=reviews_final* | stats count(id) avg(ratings.overall) by offering_id  | lookup hotel_info.csv hotel_id as offering_id output name | rename avg(ratings.overall) as average_rating | search average_rating < 2.5|sort limit=10 -count(id) | fields name, count(id) | rename count(id) as number_reviews
```



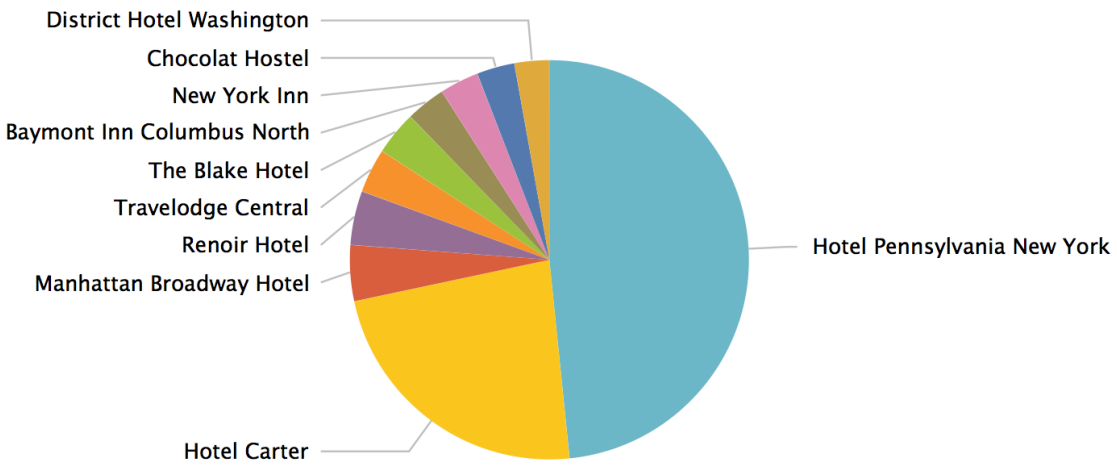
name	number_reviews
Hotel Pennsylvania New York	21824
Hotel Carter	9280
Manhattan Broadway Hotel	1696
Hotel Riverside Studios	1616
Renoir Hotel	1524
New York Inn	1256
District Hotel Washington	1180
Chocolat Hostel	1148
West Side Inn	976
Cecil Hotel	780

Verification :

sourcetype=reviews_final* | stats count(id) avg(ratings.overall) by offering_id | lookup hotel_info.csv hotel_id as offering_id output name | search name="Hotel Pennsylvania New York"

offering_id	count(id)	avg(ratings.overall)	name
214197	21824	2.469391	Hotel Pennsylvania New York

Visualization



11. Top 10 useful reviewers

Objective of the query: To find the hotels that have the most number of reviews.

Description of the query:

The intent of this query is to find the hotels that have the highest number of reviews from the customers.

Query:

```
# sourcetype=reviews_final author.username != "" | stats sum(num_helpful_votes) by author.username | sort limit=10 -sum(num_helpful_votes)
```

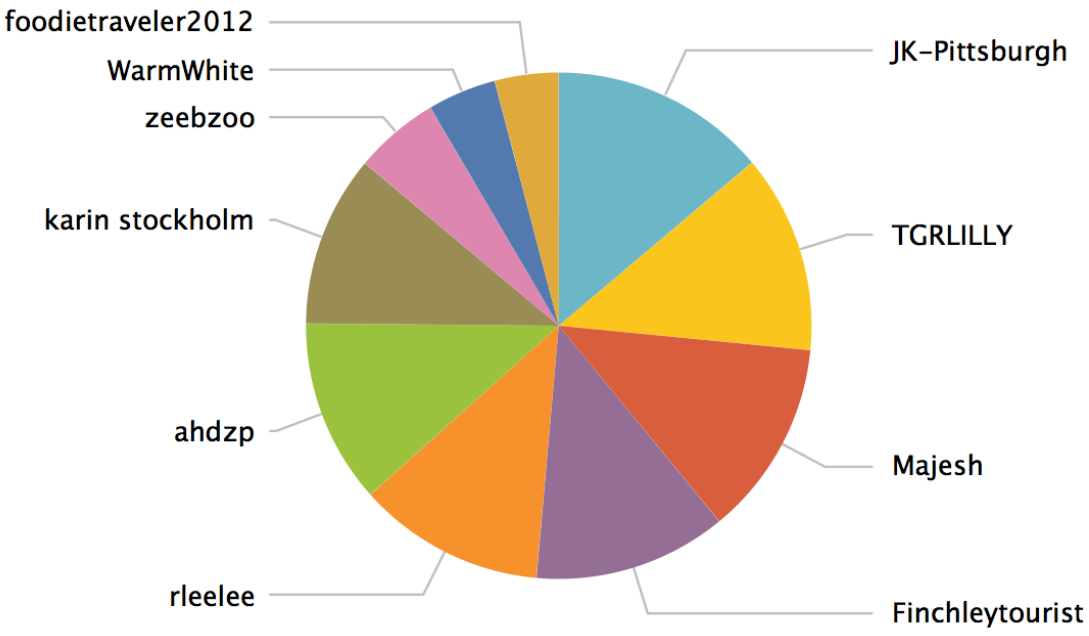
author.username	sum(num_helpful_votes)
JK-Pittsburgh	2060
Majesh	1860
rleelee	1764
foodietraveler2012	464
Michael S	432
zeebzoo	412
WillyJ	376
vudubabz	340
AIB1	324
BosPretzel123	316

Verification :

```
# sourcetype=reviews_final author.username = "JK-Pittsburgh" | stats sum(num_helpful_votes) by author.username
```

author.username	sum(num_helpful_votes)
JK-Pittsburgh	2060

Visualization



12. Hotel with most number of overall ratings as 5.0

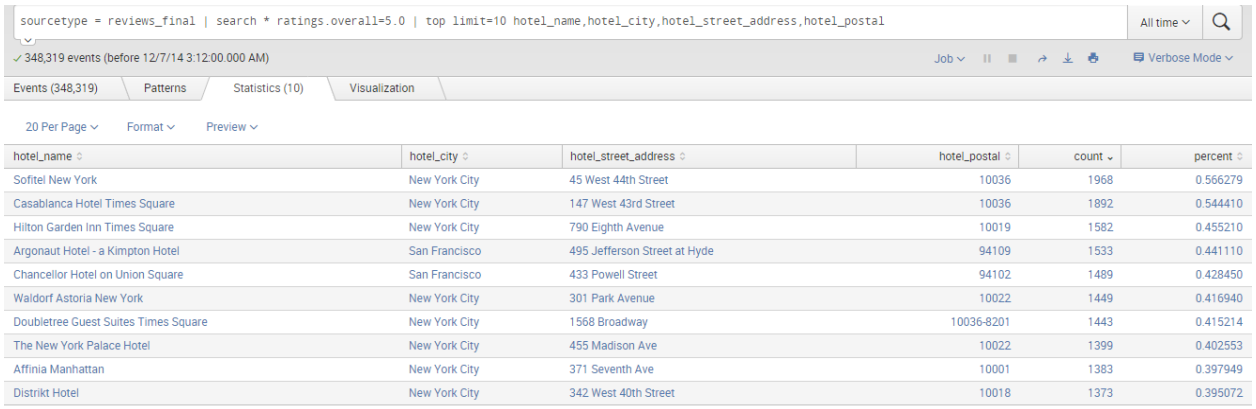
Objective of the query : To find the hotel, and its location who has the most number of overall ratings as 5

Description of the query :

We wanted to find the hotel which has been rated the best in overall ratings by the maximum number of customers.

Query:

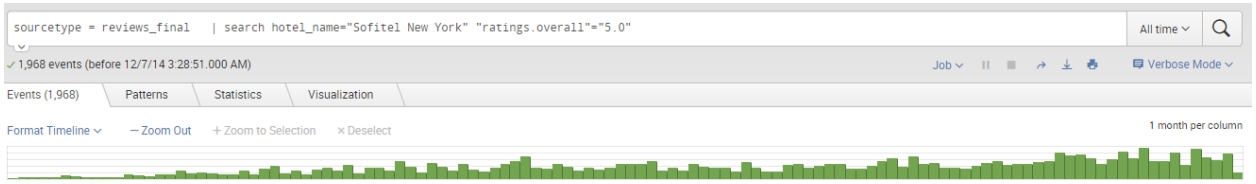
```
# sourcetype = reviews_final | search * ratings.overall=5.0 | top limit=10 hotel_name,hotel_city,hotel_street_address,hotel_postal
```



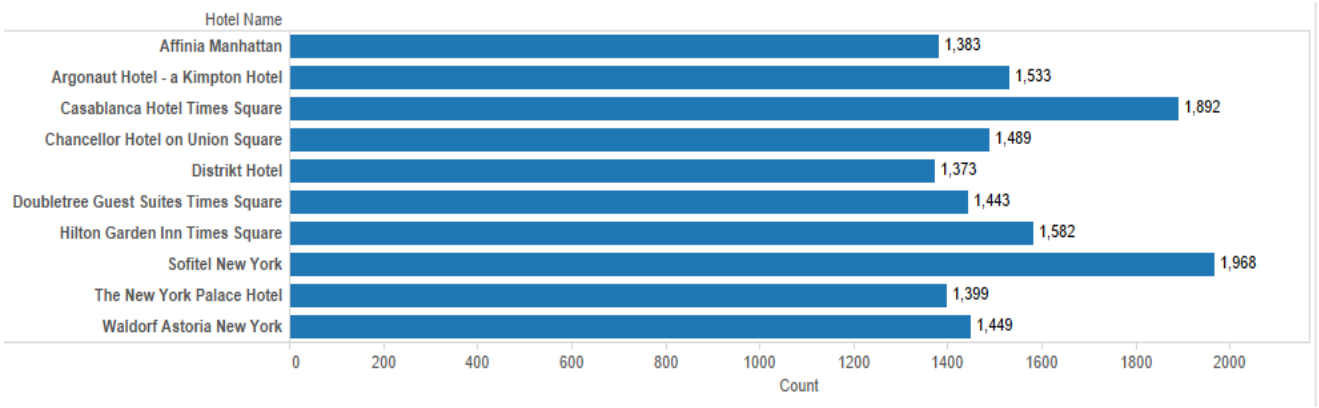
hotel_name	hotel_city	hotel_street_address	hotel_postal	count	percent
Sofitel New York	New York City	45 West 44th Street	10036	1968	0.566279
Casablanca Hotel Times Square	New York City	147 West 43rd Street	10036	1892	0.544410
Hilton Garden Inn Times Square	New York City	790 Eighth Avenue	10019	1582	0.455210
Argonaut Hotel - a Kimpton Hotel	San Francisco	495 Jefferson Street at Hyde	94109	1533	0.441110
Chancellor Hotel on Union Square	San Francisco	433 Powell Street	94102	1489	0.428450
Waldorf Astoria New York	New York City	301 Park Avenue	10022	1449	0.416940
Doubletree Guest Suites Times Square	New York City	1568 Broadway	10036-8201	1443	0.415214
The New York Palace Hotel	New York City	455 Madison Ave	10022	1399	0.402553
Affinia Manhattan	New York City	371 Seventh Ave	10001	1383	0.397949
Distrikt Hotel	New York City	342 West 40th Street	10018	1373	0.395072

Verification:

sourcetype = reviews_final | search * ratings.overall=5.0 | where hotel_name=" Sofitel New York "



Visualization



13. Hotels with most number of overall ratings as 1.0

Objective of the query : To find the hotel, and its location who has the most number of overall ratings as 1.0

Description of the query :

We wanted to find the hotel which has been rated the worst in overall ratings by the maximum number of customers.

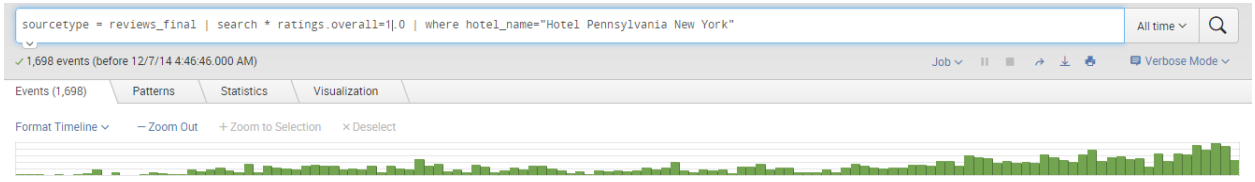
Query:

sourcetype = reviews_final | search ratings.overall=1.0 | top limit=10 hotel_name,hotel_city,hotel_street_address,hotel_postal

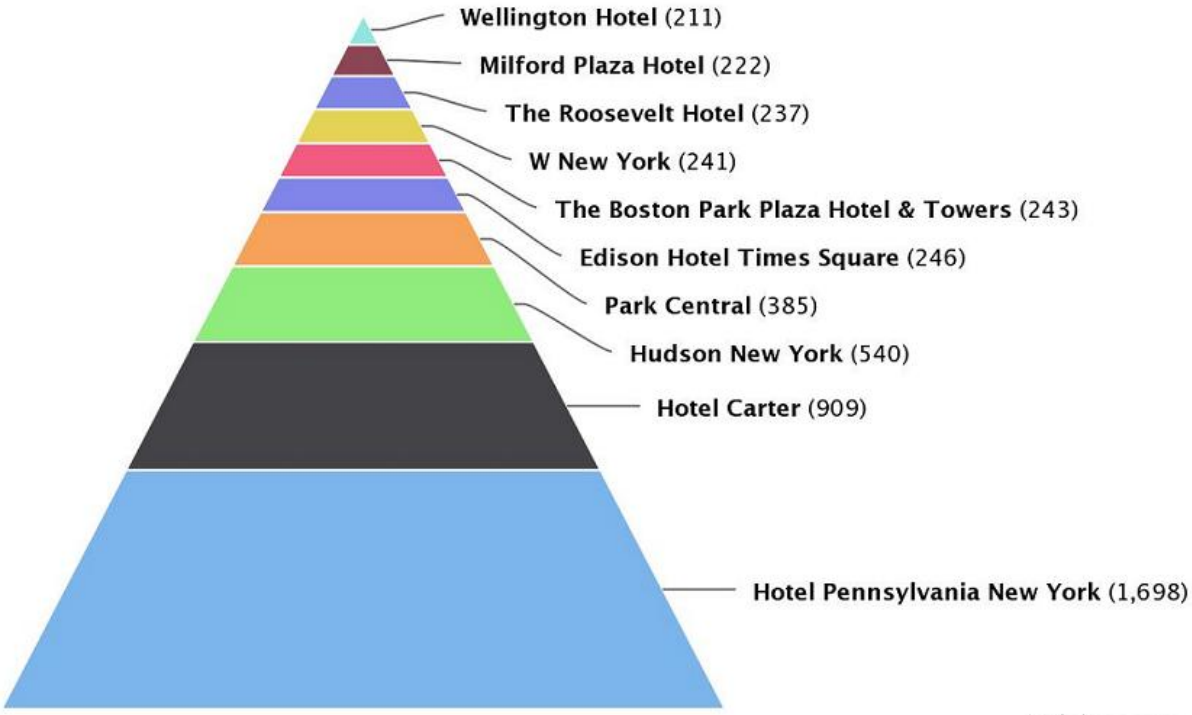
hotel_name	hotel_city	hotel_street_address	hotel_postal	count	percent
Hotel Pennsylvania New York	New York City	401 Seventh Avenue at 33rd Street	10001	1698	3.161364
Hotel Carter	New York City	250 West 43rd Street	10036	909	1.692391
Hudson New York	New York City	356 West 58th Street	10019	540	1.005381
Park Central	New York City	870 Seventh Avenue at 56th Street	10019	385	0.716799
Edison Hotel Times Square	New York City	228 West 47th Street	10036	246	0.458007
The Boston Park Plaza Hotel & Towers	Boston	50 Park Plaza at Arlington Street	2116	243	0.452421
W New York	New York City	541 Lexington Avenue	10022	241	0.448698
The Roosevelt Hotel	New York City	45 East 45th Street at Madison Avenue	10017	237	0.441250
Milford Plaza Hotel	New York City	700 8th Avenue	10036	222	0.413323
Wellington Hotel	New York City	871 7th Ave at W 55th St	10019-3830	211	0.392843

Verification :

sourcetype = reviews_final | search * ratings.overall=1.0 | where hotel_name="Hotel Pennsylvania New York"



Visualization



14. Average of all types of ratings for hotels

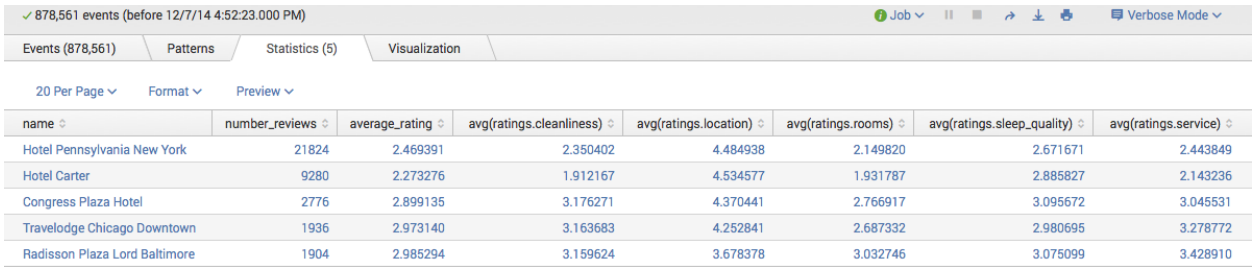
Objective of the query : To find the hotel, and its location who has the most number of overall ratings as 5

Description of the query :

We wanted to find the hotel which has been rated the best in overall ratings by the maximum number of customers.

Query:

```
# sourcetype=reviews_final* | stats count(id) avg(ratings.overall) , avg(ratings.cleanliness), avg(ratings.location), avg(ratings.rooms), avg(ratings.sleep_quality), avg(ratings.service) by offering_id | lookup hotel_info.csv hotel_id as offering_id output name | rename avg(ratings.overall) as average_rating | search average_rating < 3.0|sort limit=5 -count(id) | fields name, hotel_city, count(id), average_rating, avg(ratings.cleanliness), avg(ratings.location), avg(ratings.rooms), avg(ratings.sleep_quality), avg(ratings.service) | rename count(id) as number_reviews
```



✓ 878,561 events (before 12/7/14 4:52:23.000 PM) Job Verbose Mode

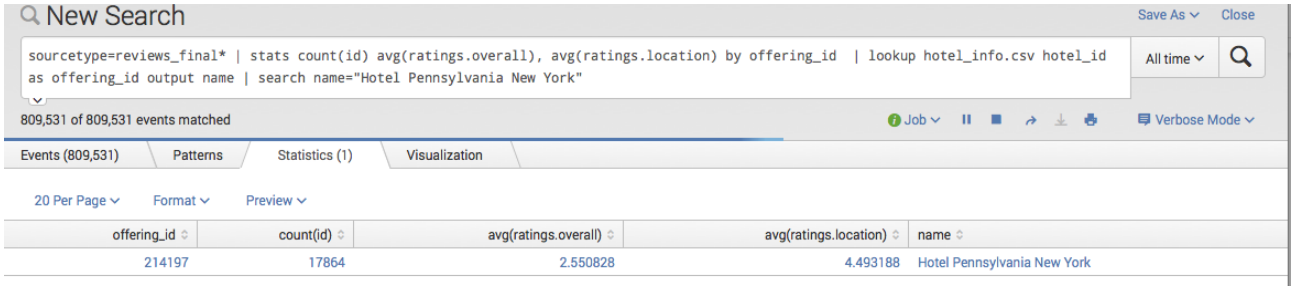
Events (878,561) Patterns Statistics (5) Visualization

20 Per Page Format Preview

name	number_reviews	average_rating	avg(ratings.cleanliness)	avg(ratings.location)	avg(ratings.rooms)	avg(ratings.sleep_quality)	avg(ratings.service)
Hotel Pennsylvania New York	21824	2.469391	2.350402	4.484938	2.149820	2.671671	2.443849
Hotel Carter	9280	2.273276	1.912167	4.534577	1.931787	2.885827	2.143236
Congress Plaza Hotel	2776	2.899135	3.176271	4.370441	2.766917	3.095672	3.045531
Travelodge Chicago Downtown	1936	2.973140	3.163683	4.252841	2.687332	2.980695	3.278772
Radisson Plaza Lord Baltimore	1904	2.985294	3.159624	3.678378	3.032746	3.075099	3.428910

Verification :

sourcetype=reviews_final* | stats count(id) avg(ratings.overall), avg(ratings.location) by offering_id | lookup hotel_info.csv hotel_id as offering_id output name | search name="Hotel Pennsylvania New York"



New Search Save As Close

```
sourcetype=reviews_final* | stats count(id) avg(ratings.overall), avg(ratings.location) by offering_id  | lookup hotel_info.csv hotel_id
as offering_id output name | search name="Hotel Pennsylvania New York"
```

All time

809,531 of 809,531 events matched Job Verbose Mode

Events (809,531) Patterns Statistics (1) Visualization

20 Per Page Format Preview

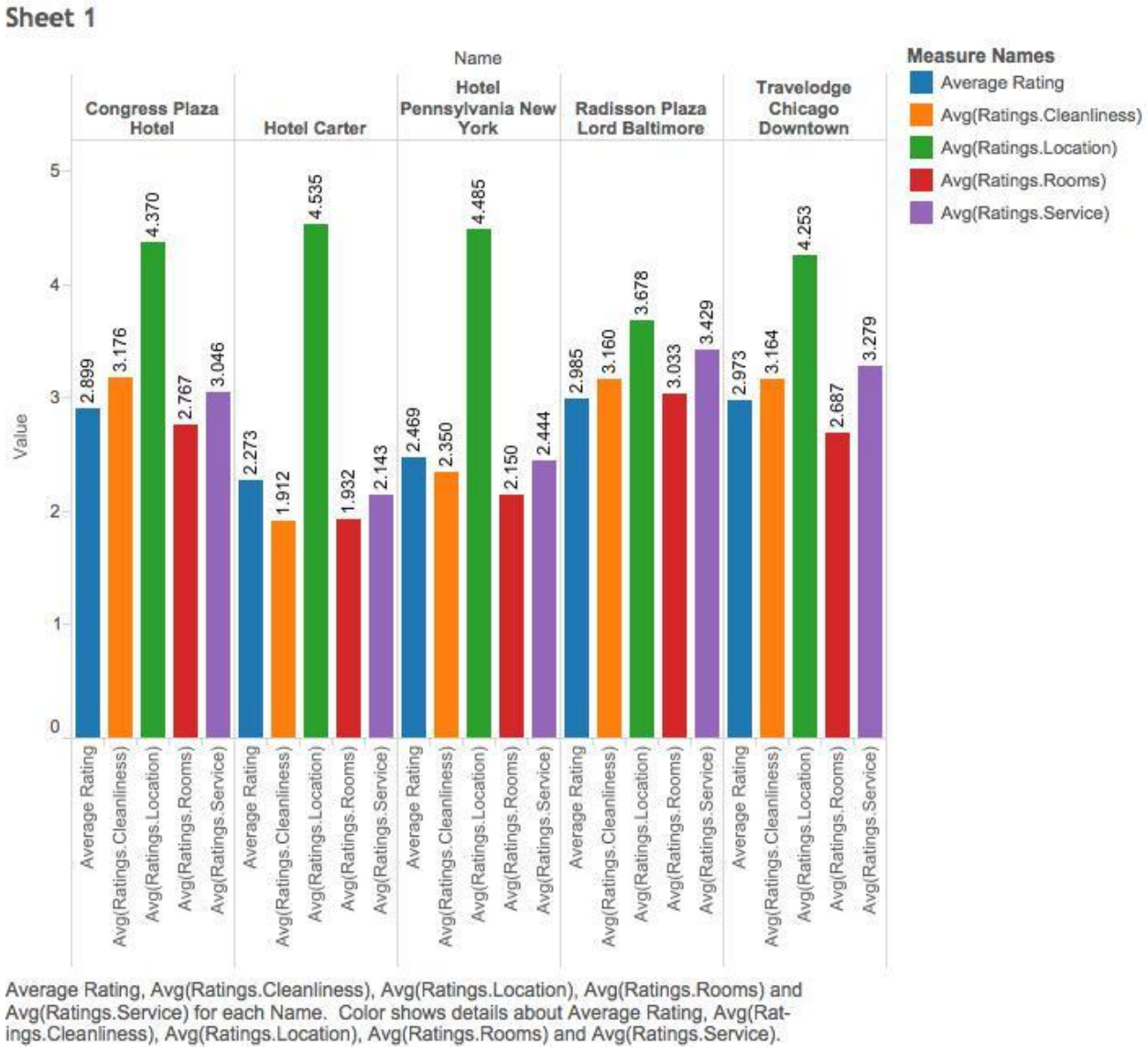
offering_id	count(id)	avg(ratings.overall)	avg(ratings.location)	name
214197	17864	2.550828	4.493188	Hotel Pennsylvania New York

- Here we see that the average ratings for location for the Hotel Pennsylvania New York is very high that is 4.48. This could be one possible reasons that people book this hotel the most even though the hotel has most negative reviews.

- We also observe that the negative rated hotels also have high review counts. What makes them a prefered choice even if the ratings are low. Is it because of location?

- To confirm this we found the average ratings for all the features such as overall, location, cleanliness, sleep_qulaity etc for hotel which are reviewed the most as overall_ratings < 2.5

Visualization



We observe that location ratings for all these hotels are very high. So location of the hotel is one of the compelling reason while making hotel choice besides the price and class.

15. Top 5 hotels having negative reviews but the most number of review count in NYC

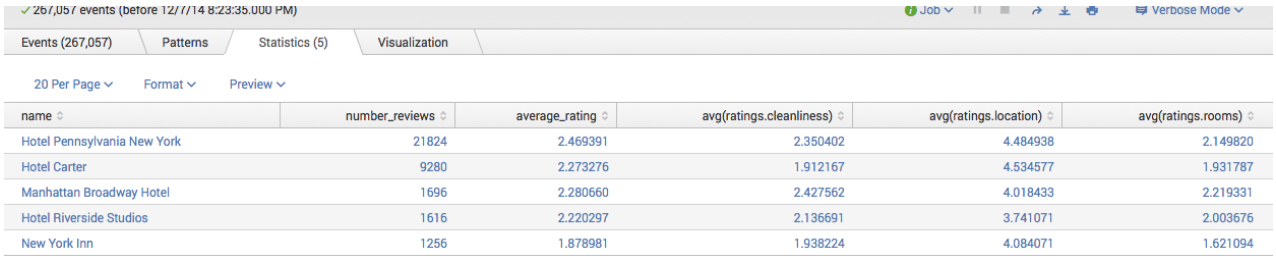
Objective of the query : To find the hotels with most number of reviews negative in New York.

Description of the query :

We wanted to find the hotel which has been rated the best in overall ratings by the maximum number of customers.

Query:

sourcetype=reviews_final |search * "hotel_city" = "New York City" | stats count(id) avg(ratings.overall), avg(ratings.cleanliness), avg(ratings.location), avg(ratings.rooms) by offering_id | lookup hotel_info.csv hotel_id as offering_id output name | rename avg(ratings.overall) as average_rating | search average_rating <= 2.5 | sort limit=5 -count(id) | fields name, count(id), average_rating, avg(ratings.cleanliness), avg(ratings.location), avg(ratings.rooms) | rename count(id) as number_reviews



✓ 267,057 events (before 12/7/14 8:23:35.000 PM) Job Verbose Mode

Events (267,057) Patterns Statistics (5) Visualization

20 Per Page Format Preview

name	number_reviews	average_rating	avg(ratings.cleanliness)	avg(ratings.location)	avg(ratings.rooms)
Hotel Pennsylvania New York	21824	2.469391	2.350402	4.484938	2.149820
Hotel Carter	9280	2.273276	1.912167	4.534577	1.931787
Manhattan Broadway Hotel	1696	2.280660	2.427562	4.018433	2.219331
Hotel Riverside Studios	1616	2.220297	2.136691	3.741071	2.003676
New York Inn	1256	1.878981	1.938224	4.084071	1.621094

Verification :

sourcetype=reviews_final* | stats count(id) avg(ratings.overall) by offering_id | lookup hotel_info.csv hotel_id as offering_id output name | search name="Hotel Pennsylvania New York"

Visualization

offering_id	count(id)	avg(ratings.overall)	name
214197	21824	2.469391	Hotel Pennsylvania New York

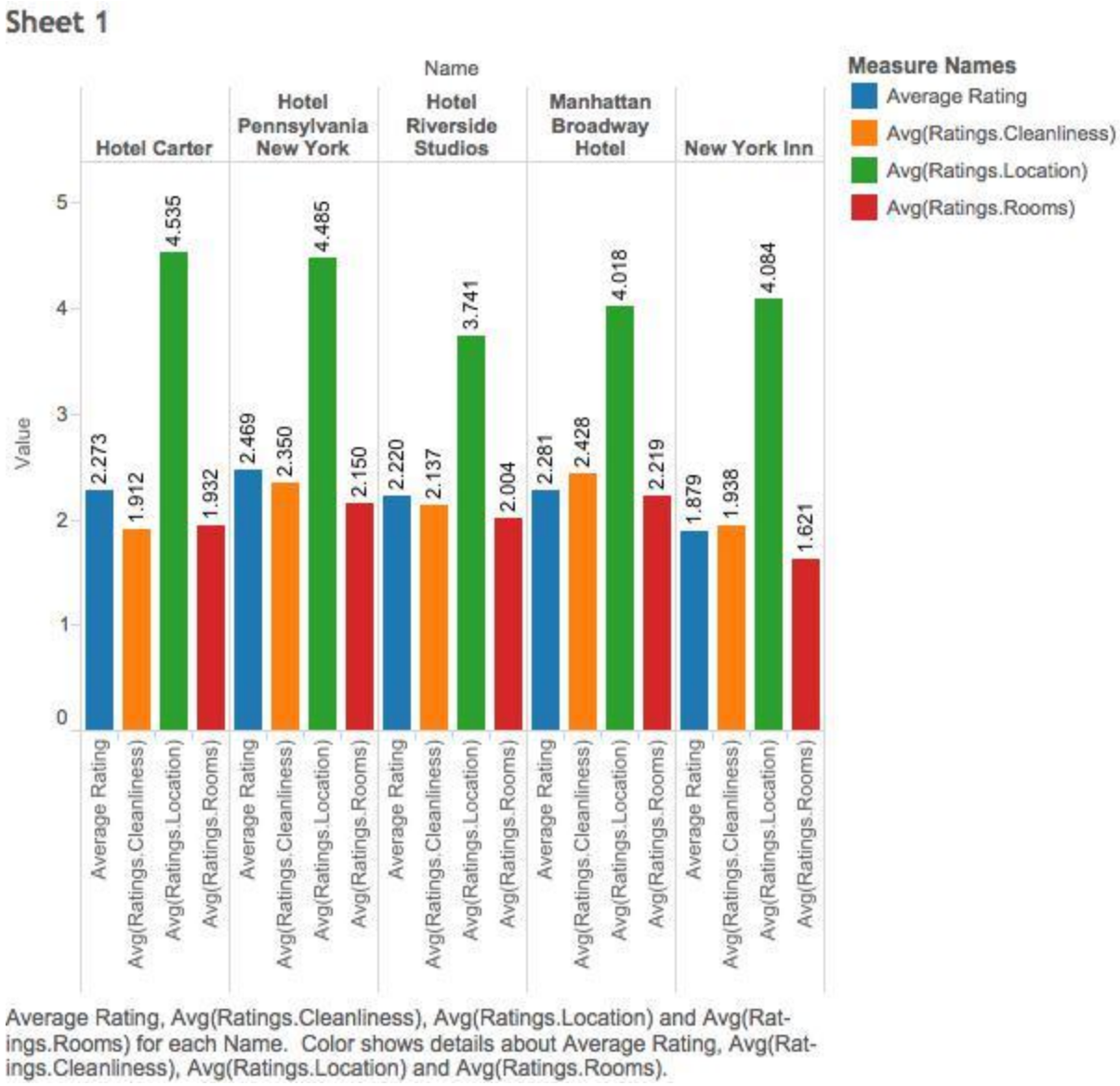
Sheet 1

Name	
Hotel Carter	9,280
Hotel Pennsylvania New York	21,824
Hotel Riverside Studios	1,616
Manhattan Broadway Hotel	1,696
New York Inn	1,256

Number Reviews

1,256 21,824

Sum of Number Reviews broken down by Name. Color shows sum of Number Reviews. The marks are labeled by sum of Number Reviews.



16. Top 10 haunted hotels based on user's reviews in USA

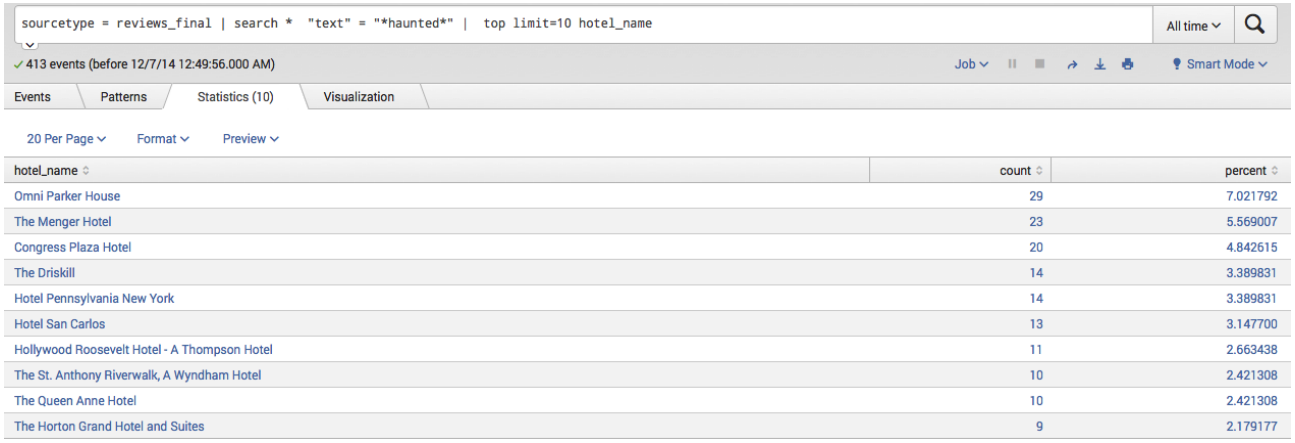
Objective of the query : To find the haunted hotels based on user reviews.

Description of the query :

We wanted to find the hotels which are reviewed as haunted by the maximum number of customers.

Query:

#sourcetype = reviews_final | search * "text" = "*haunted*" AND *ghost* | top limit=10 hotel_name

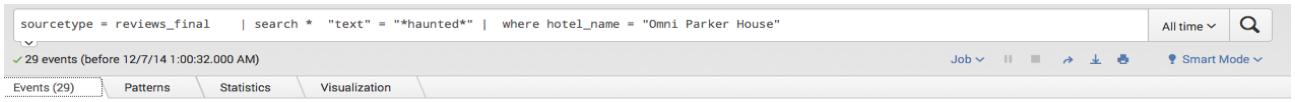


hotel_name	count	percent
Omni Parker House	29	7.021792
The Menger Hotel	23	5.569007
Congress Plaza Hotel	20	4.842615
The Driskill	14	3.389831
Hotel Pennsylvania New York	14	3.389831
Hotel San Carlos	13	3.147700
Hollywood Roosevelt Hotel - A Thompson Hotel	11	2.663438
The St. Anthony Riverwalk, A Wyndham Hotel	10	2.421308
The Queen Anne Hotel	10	2.421308
The Horton Grand Hotel and Suites	9	2.179177

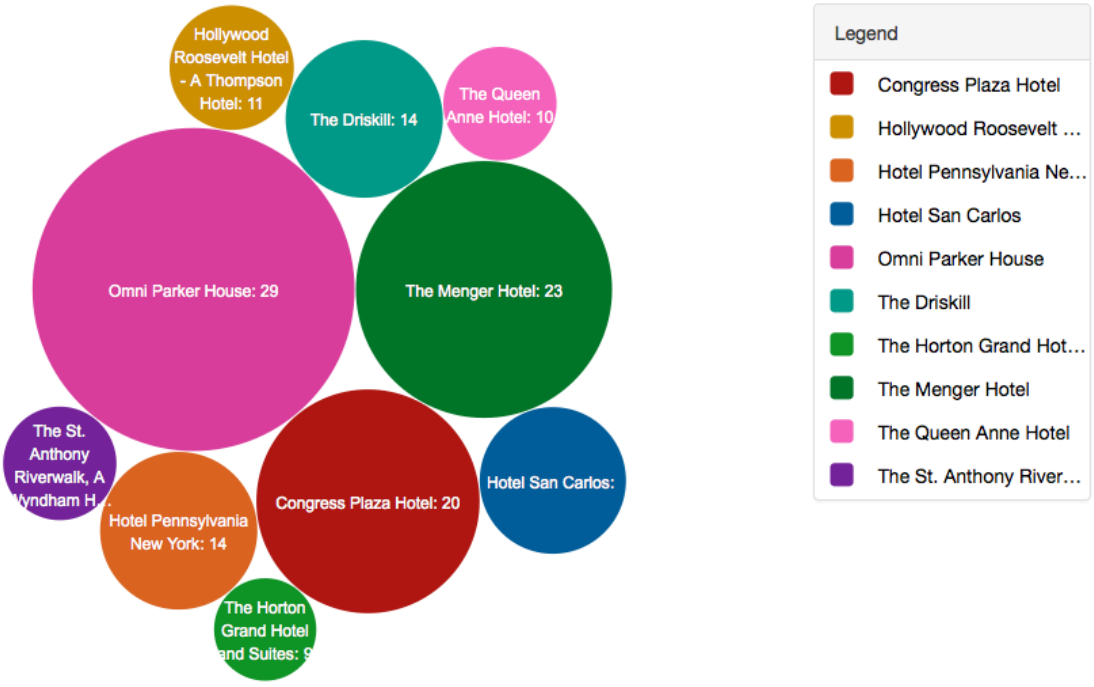
Verification:

We verified it for the hotel Omni Parker House.

sourcetype=reviews_final | search* "text" = "*haunted*" | where hotel_name = "Omni Parker House"



Visualization



17. Top 10 haunted hotels based on user's reviews in NY

Objective of the query : To find the haunted hotels in NY based on user reviews.

Description of the query :

We wanted to find the hotels which are reviewed as haunted by the maximum number of customers.

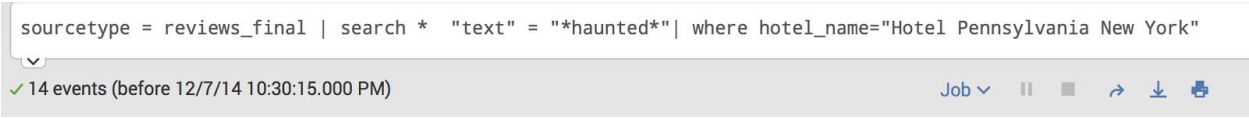
Query:

```
# sourcetype = reviews_final | search *  "text" = "*haunted*"| where hotel_state = "NY" |  top limit=10 hotel_name hotel_street_address hotel_state hotel_postal hotel_city
```

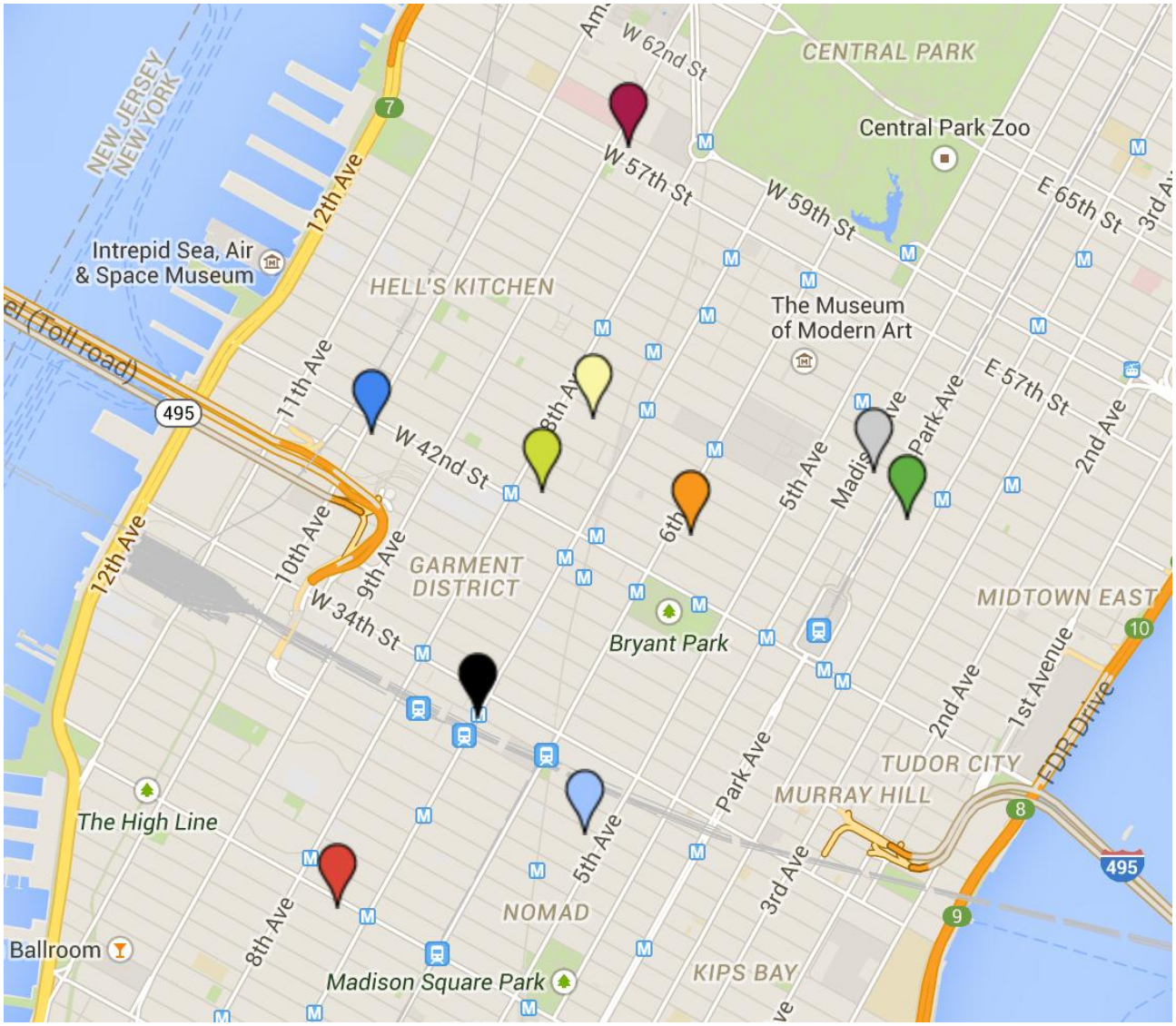
hotel_name	hotel_street_address	hotel_state	hotel_postal	hotel_city	count	percent
Hotel Pennsylvania New York	401 Seventh Avenue at 33rd Street	NY	10001	New York City	14	26.923077
Hotel Carter	250 West 43rd Street	NY	10036	New York City	4	7.692308
Chelsea Hotel	222 W 23rd St	NY	10011	New York City	4	7.692308
Wolcott Hotel	4 West 31st Street	NY	10001	New York City	2	3.846154
Hudson New York	356 West 58th Street	NY	10019	New York City	2	3.846154
Edison Hotel Times Square	228 West 47th Street	NY	10036	New York City	2	3.846154
Algonquin Hotel Times Square, Autograph Collection	59 W 44th St	NY	10036	New York City	2	3.846154
YOTEL New York at Times Square West	570 10th Ave	NY	10036	New York City	1	1.923077
Waldorf Astoria New York	301 Park Avenue	NY	10022	New York City	1	1.923077
The New York Palace Hotel	455 Madison Ave	NY	10022	New York City	1	1.923077

Verification:

sourcetype= reviews_final | search * "text" = "*haunted*" | where hotel_name = "Hotel Pennsylvania New York"



Visualization



18. The top 20 hotels with bed bugs based on user's reviews

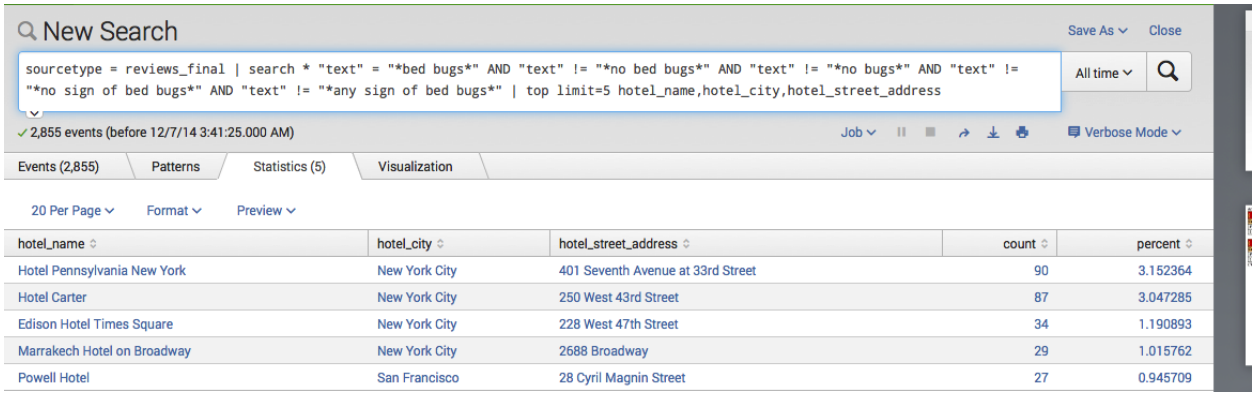
Objective of the query : To find the hotels with bedbugs based on user reviews.

Description of the query :

We wanted to find the hotels which were having maximum bedbugs as reviewed by the customers.

Query:

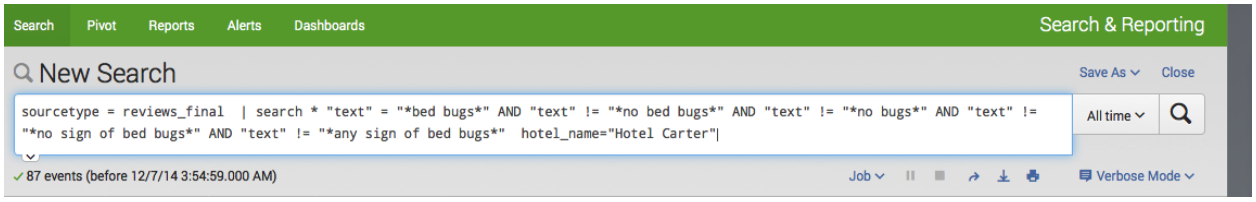
sourcetype = reviews_final | search * "text" = "*bed bugs*" AND "text" != "*no bed bugs*" AND "text" != "*no bugs*" AND "text" != "*no sign of bed bugs*" AND "text" != "*any sign of bed bugs*" | top limit=5 hotel_name,hotel_city,hotel_street_address



hotel_name	hotel_city	hotel_street_address	count	percent
Hotel Pennsylvania New York	New York City	401 Seventh Avenue at 33rd Street	90	3.152364
Hotel Carter	New York City	250 West 43rd Street	87	3.047285
Edison Hotel Times Square	New York City	228 West 47th Street	34	1.190893
Marrakech Hotel on Broadway	New York City	2688 Broadway	29	1.015762
Powell Hotel	San Francisco	28 Cyril Magnin Street	27	0.945709

Verification :

sourcetype = reviews_final | search * "text" = "*bed bugs*" AND "text" != "*no bed bugs*" AND "text" != "*no bugs*" AND "text" != "*no sign of bed bugs*" AND "text" != "*any sign of bed bugs*" hotel_name="Hotel Carter"



Visualization