Online-Coffee shop-Data Analytics

DATA ANALYSIS PROJECT PROPOSAL

FOR COFFEEKING

04/2020

CHENG GONG

Section 1.

Which client/dataset did you select and why?

CoffeeKing is a new startup coffee company providing a unique and novel experience to their customers. They want to appeal to a wide variety of clientele.

This project is an online data analytics for coffeeshops, including business address, business hours, user reviews, star rating , etc.

The analysing result could provide client a **Targeted and personalized marketing Plan**.

- ONLINE-COFFEE SHOP-DATA ANALYTICS

Describe the steps you took to import and clean the data.

1. Download dataset from Yelp Dataset

2. Import data to Jupiter notebook. Because it is a large JSON file, I used spark SQL to read and transfer data.

3. Depending on ERD, I loaded in user, review and business tables, joined 3 tables and kept the columns that I need and filtered names with key-word 'coffee' . Then dropped all the null values to get a clean dataset.

Perform initial exploration of data and provide some screenshots or display some stats of the data you are looking at.

```python
In [2]: import json
        import pandas as pd

        import pyspark
        import random

        sc = pyspark.SparkContext(appName="Pi")
```

```python
In [28]: from IPython.core.display import display, HTML
         display(HTML("<style>.container { width:100% !important; }</style>"))
```

```python
In [18]: from pyspark.sql import SparkSession

         spark = SparkSession \
             .builder \
             .appName("Python Spark SQL basic example") \
             .getOrCreate()

         # df = spark.read.csv("/home/stp/test1.csv", header=True, sep="|");
         spark
```

Out[18]: **SparkSession - in-memory**

SparkContext

Spark UI

Version
v2.4.5

Master
local[*]

AppName
Pi

```python
In [30]: review_df=spark.read.json('F:/Download/yelp_academic_dataset/yelp_academic_dataset_review.json')
```

```python
In [31]: review_df.show()
```

```
+--------------------+----+-------------------+-----+--------------------+-----+--------------------+------+--------------------+
|         business_id|cool|               date|funny|           review_id|stars|                text|useful|             user_id|
+--------------------+----+-------------------+-----+--------------------+-----+--------------------+------+--------------------+
|-MhfebM0QIsKt87iD...|   0|2015-04-15 05:21:16|    0|xQY8N_XvtGbearJ5X...|  2.0|As someone who ha...|     5|OwjRMXRC0KyPrI1cj...|
|1brU8StCq3yDfr-QM...|   0|2013-12-07 03:16:52|    1|UmFMZ8PyXZTY2Qcwz...|  1.0|I am actually hor...|     1|nIJD_7ZXHq-FX8byP...|
|HQl28KMwrEKHqhFrr...|   0|2015-12-05 03:18:11|    0|LG2ZaYiOgpr2DK_90...|  5.0|I love Deagan's. ...|     1|V34qejxNsCbcgD8C0...|
|5JxlZaqCnk1MnbgRi...|   0|2011-05-27 05:30:52|    0|i6g_oA9Yf9Y31qt0w...|  1.0|Dismal, lukewarm,...|     0|ofKDkJKXSKZXu5xJN...|
|IS4cv902ykd8wj1TR...|   0|2017-01-14 21:56:57|    0|6TdNDKywdbjoTkize...|  4.0|Oh happy day, fin...|     0|UgMW8bLE0QMJDCkQ1...|
|nlxHRv1zXGT0cOK51...|   0|2013-05-07 07:25:25|    0|L20_INwlrRuoXO5KS...|  5.0|This is definitel...|     2|5vD2kmE25YBrbayKh...|
|Pthe4qk5xh4n-ef-9...|   0|2015-11-05 23:11:05|    0|ZayJ1zWyWgY9S_TRL...|  5.0|Really good place...|     1|aq_ZxGHiri48TUXJ1...|
|FNCJpSn0tL9iqoY3J...|   0|2017-07-18 18:31:54|    0|lpFIJYpsvDxyph-kP...|  5.0|Awesome office an...|     0|dsd-KNYKMpx6ma_sR...|
|e_BiI4ej1CW1F0EyV...|   0|2015-02-16 06:48:47|    0|JA-xnyHytKiOIHl_z...|  5.0|Most delicious au...|     0|P6apihD4ASf1vpPxH...|
|Ws8V970-mQt2X9CwC...|   1|2009-10-13 04:16:41|    0|z4BCgTkfNtCu4XY5L...|  4.0|I have been here ...|     3|jOERvhmK6_1o_XGUB...|
|PA61Rwk3AMwOEXHev...|   0|2013-12-28 21:02:55|    0|TfVth7UNfgi1v4J3e...|  5.0|Maria is VERY goo...|     0|s5j_CRBWDCCMDJ6r7...|
|1-nL4BmhzpZjcavoo...|   1|2015-10-17 01:38:13|    1|Tyx7AxYQfSRnBFUIX...|  4.0|ORDER In (Deliver...|     2|HJECayULRM-6xh2GC...|
|Naa6E0YU0Wr7jCuCE...|   0|2015-07-03 21:48:51|    0|wJMjt5C2y1RKgY0Xb...|  5.0|We purchased new ...|     3|1YIQGP-a534nyksaw...|
|Ns4tjgLfqR1qawG1N...|   0|2016-06-11 22:00:11|    0|QCxPzh7cuxJrLd6A_...|  5.0|Everything that m...|     0|qftVgPj_kRTi1dMDj...|
|Z1CSsWS07Ju1SBIQ1...|   0|2015-05-26 10:36:47|    0|qWHp2121ysENZ0bh6...|  5.0|Called for a 5:15...|     1|51b0P0g2t-AkMFx66...|
|7Ka9Pd8X9SRHs1D5E...|   0|2017-08-07 21:36:36|    0|mjbs5CL4eMu4o6_Vt...|  1.0|If I could give 1...|     0|TF4C-F5iqavACQgKT...|
|d4qwVw4PcN-_2mK2o...|   0|2015-02-02 06:28:00|    0|bVTjZgRNq8Toxzvti...|  1.0|10pm on a super b...|     0|2hRe26HSCAWbFRn5W...|
|oVuZt1CFg_zF090Nh...|   1|2018-02-01 19:15:00|    0|Ne_2CSfcKIqXHmv_K...|  4.0|A close friend wa...|     3|6sJN_H1M_uwpfLJ1p...|
|_iGvLfEsqDwPUxRUA...|   0|2017-06-28 00:39:18|    0|Hy-gUXQh3RVhE8FLH...|  1.0|Tried to have my ...|     0|kMkWON21mwOs-M-fw...|
|poSV39UqEg-gpESXa...|   0|2018-03-04 01:03:53|    0|UGErdm6bt48SXTVwJ...|  3.0|My husband and I ...|     0|QodunSzok4nIYFNrT...|
+--------------------+----+-------------------+-----+--------------------+-----+--------------------+------+--------------------+
only showing top 20 rows
```

Get clean dataset

In [279]: `analysis_df=spark.sql("SELECT a.business_id,categories, name, postal_code,city, review_count, stars, text, review_id, useful,Friday, Monday, Saturday, Sunday, Thursday, Tuesday, Wednesday FROM analysis_df b JOIN analysis_hours_spl`

In [280]: `analysis_df.show()`

```
+----------------+-----------------+-----------+-----------+-------+------------+-----+----------------+----------------+------+---------+---------+--------+--------+---------+---------+---------+
|     business_id|       categories|       name|postal_code|   city|review_count|stars|            text|       review_id|useful|   Friday|   Monday|Saturday|  Sunday| Thursday|  Tuesday|Wednesday|
+----------------+-----------------+-----------+-----------+-------+------------+-----+----------------+----------------+------+---------+---------+--------+--------+---------+---------+---------+
|JTcsBnLgSh5vNVLD7...|Food, Restaurants...|Supercoffee|    M6N 3S3|Toronto|          45|  5.0|Love this place! ...|oNO_NaJtUse4Sn6hE...|     2|6:30-19:0|6:30-18:0|8:0-19:0|8:0-19:0|6:30-19:0|8:0-18:0| 8:0-18:0|
|JTcsBnLgSh5vNVLD7...|Food, Restaurants...|Supercoffee|    M6N 3S3|Toronto|          45|  5.0|I work nearby and...|AHHkBBBjRGJBbfeGO...|     0|6:30-19:0|6:30-18:0|8:0-19:0|8:0-19:0|6:30-19:0|8:0-18:0| 8:0-18:0|
|JTcsBnLgSh5vNVLD7...|Food, Restaurants...|Supercoffee|    M6N 3S3|Toronto|          45|  4.0|Very new place, b...|a81xjizNu1OtYWQR7...|     1|6:30-19:0|6:30-18:0|8:0-19:0|8:0-19:0|6:30-19:0|8:0-18:0| 8:0-18:0|
|JTcsBnLgSh5vNVLD7...|Food, Restaurants...|Supercoffee|    M6N 3S3|Toronto|          45|  4.0|A much needed sma...|Y-stOmqwHlV5fZm5q...|     1|6:30-19:0|6:30-18:0|8:0-19:0|8:0-19:0|6:30-19:0|8:0-18:0| 8:0-18:0|
|JTcsBnLgSh5vNVLD7...|Food, Restaurants...|Supercoffee|    M6N 3S3|Toronto|          45|  4.0|Great promise to ...|pXIbtZesoON4ALPa_...|     2|6:30-19:0|6:30-18:0|8:0-19:0|8:0-19:0|6:30-19:0|8:0-18:0| 8:0-18:0|
|JTcsBnLgSh5vNVLD7...|Food, Restaurants...|Supercoffee|    M6N 3S3|Toronto|          45|  4.0|Neighborhood coff...|EK3gqOPhtKFn9T2Rs...|     1|6:30-19:0|6:30-18:0|8:0-19:0|8:0-19:0|6:30-19:0|8:0-18:0| 8:0-18:0|
|JTcsBnLgSh5vNVLD7...|Food, Restaurants...|Supercoffee|    M6N 3S3|Toronto|          45|  5.0|Absolutely wonder...|MqYebCiZDO1_xWNgE...|     1|6:30-19:0|6:30-18:0|8:0-19:0|8:0-19:0|6:30-19:0|8:0-18:0| 8:0-18:0|
|JTcsBnLgSh5vNVLD7...|Food, Restaurants...|Supercoffee|    M6N 3S3|Toronto|          45|  5.0|The coffee here m...|HLQ3LkwkRAAOqpwTB...|     1|6:30-19:0|6:30-18:0|8:0-19:0|8:0-19:0|6:30-19:0|8:0-18:0| 8:0-18:0|
|JTcsBnLgSh5vNVLD7...|Food, Restaurants...|Supercoffee|    M6N 3S3|Toronto|          45|  5.0|Amazing varieties...|BP_TO5Y_8QmzBFuj9...|     0|6:30-19:0|6:30-18:0|8:0-19:0|8:0-19:0|6:30-19:0|8:0-18:0| 8:0-18:0|
|JTcsBnLgSh5vNVLD7...|Food, Restaurants...|Supercoffee|    M6N 3S3|Toronto|          45|  5.0|I passed this pla...|3ykpXhOb1AakhxKmx...|     2|6:30-19:0|6:30-18:0|8:0-19:0|8:0-19:0|6:30-19:0|8:0-18:0| 8:0-18:0|
|JTcsBnLgSh5vNVLD7...|Food, Restaurants...|Supercoffee|    M6N 3S3|Toronto|          45|  4.0|A rustic coffee s...|EVUwOxvqGXb-L3P1M...|     3|6:30-19:0|6:30-18:0|8:0-19:0|8:0-19:0|6:30-19:0|8:0-18:0| 8:0-18:0|
|JTcsBnLgSh5vNVLD7...|Food, Restaurants...|Supercoffee|    M6N 3S3|Toronto|          45|  5.0|The only place it...|-3vt_r18Uahx1Qsne...|     0|6:30-19:0|6:30-18:0|8:0-19:0|8:0-19:0|6:30-19:0|8:0-18:0| 8:0-18:0|
|JTcsBnLgSh5vNVLD7...|Food, Restaurants...|Supercoffee|    M6N 3S3|Toronto|          45|  5.0|I discovered this...|9Zm4FSHnJ_SLaKywY...|     1|6:30-19:0|6:30-18:0|8:0-19:0|8:0-19:0|6:30-19:0|8:0-18:0| 8:0-18:0|
|JTcsBnLgSh5vNVLD7...|Food, Restaurants...|Supercoffee|    M6N 3S3|Toronto|          45|  4.0|Always order the ...|C9S2kuvGJ8f9W2HSM...|     0|6:30-19:0|6:30-18:0|8:0-19:0|8:0-19:0|6:30-19:0|8:0-18:0| 8:0-18:0|
|JTcsBnLgSh5vNVLD7...|Food, Restaurants...|Supercoffee|    M6N 3S3|Toronto|          45|  5.0|Supercoffee has b...|8J3wSZQZ_WGFyPOG7...|     1|6:30-19:0|6:30-18:0|8:0-19:0|8:0-19:0|6:30-19:0|8:0-18:0| 8:0-18:0|
|JTcsBnLgSh5vNVLD7...|Food, Restaurants...|Supercoffee|    M6N 3S3|Toronto|          45|  5.0|Really enjoyed my...|BXzURHLO6Y1ZM1_1C...|     3|6:30-19:0|6:30-18:0|8:0-19:0|8:0-19:0|6:30-19:0|8:0-18:0| 8:0-18:0|
|JTcsBnLgSh5vNVLD7...|Food, Restaurants...|Supercoffee|    M6N 3S3|Toronto|          45|  5.0|A really chill ca...|s6peL2dziSL5LgCg7...|     6|6:30-19:0|6:30-18:0|8:0-19:0|8:0-19:0|6:30-19:0|8:0-18:0| 8:0-18:0|
|JTcsBnLgSh5vNVLD7...|Food, Restaurants...|Supercoffee|    M6N 3S3|Toronto|          45|  4.0|Very nice place t...|GvSBTk_4n-IwUf_Me...|     0|6:30-19:0|6:30-18:0|8:0-19:0|8:0-19:0|6:30-19:0|8:0-18:0| 8:0-18:0|
|JTcsBnLgSh5vNVLD7...|Food, Restaurants...|Supercoffee|    M6N 3S3|Toronto|          45|  5.0|Good vibes, good ...|8obcqj6Cp23X6eENI...|     0|6:30-19:0|6:30-18:0|8:0-19:0|8:0-19:0|6:30-19:0|8:0-18:0| 8:0-18:0|
|JTcsBnLgSh5vNVLD7...|Food, Restaurants...|Supercoffee|    M6N 3S3|Toronto|          45|  5.0|Very happy with t...|AnPZfxlnzC_QSNJLt...|     2|6:30-19:0|6:30-18:0|8:0-19:0|8:0-19:0|6:30-19:0|8:0-18:0| 8:0-18:0|
+----------------+-----------------+-----------+-----------+-------+------------+-----+----------------+----------------+------+---------+---------+--------+--------+---------+---------+---------+
only showing top 20 rows
```

Create an ERD or proposed ERD to show the relationships of the data you are exploring.
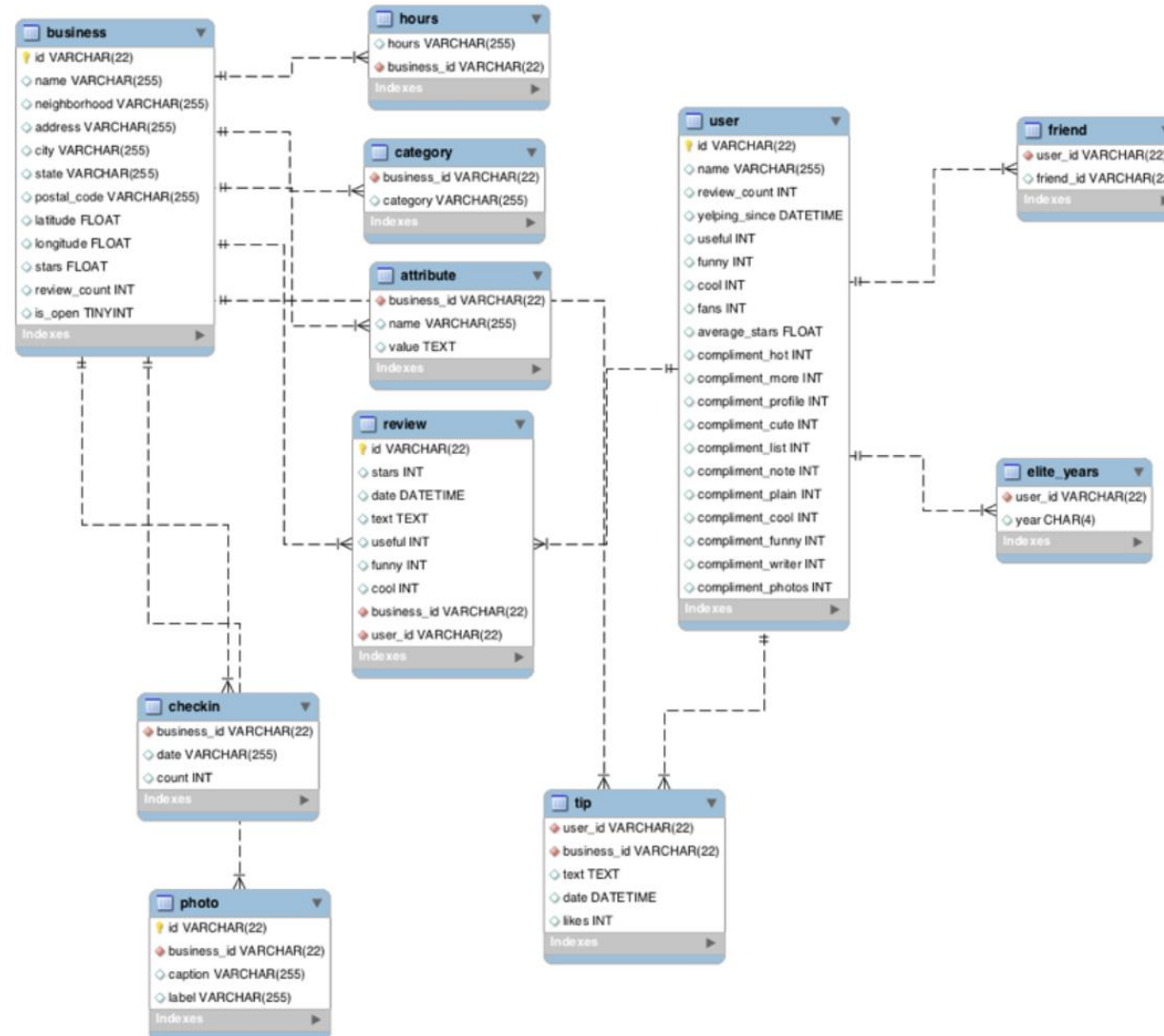
Section 2.

Description

This analysis is facing to coffee shops and restaurants business. The analysis is based on users' reviews for different areas, different styles of the shops. The analysing result could provide startup companies and investors a targeted marketing plan.

The information from users' reviews, we could try to get connections with the business conditions.

Can we estimate which area is the best shop address from the most review counts shops and the highest star rating shops?

Can we get the information from review text to know which field is customers most focusing on?

Hypothesis and Approach

From the dataset, I would like to explore the connections between city and review counts and average stars. Could it show us which area people like coffee the most, and which area has the best coffee shops?

From the review text, could we find the most popular key words to know what is customers focusing on? E.g. coffee taste, environment or the service attitude.

For this analysis, I would like to look at the area first to provide the best shop place for clients.

And I would like to check if the business hours impact on the paucity of customers.

Then we could analyse people from different areas might focus on different points of the shops.

I think a lot more answers will be found from this relational data model.

From the dataset, I find the cities with the shops have most reviews. It means the coffee business is popular in those areas. But the postal codes are all different from each other, so we could not get the information from it which is different with my hypothesis.

```
+-------------------+-------------+---------+-----------+
|               name|review_counts|     city|postal_code|
+-------------------+-------------+---------+-----------+
|Snooze, An A.M. E...|     12682120|  Phoenix|      85016|
|        Lux Central|      5047640|  Phoenix|      85012|
|Amélie's French B...|      3348416| Charlotte|     28205|
|Butters Pancakes ...|      2983750|Scottsdale|     85258|
|          The Henry|      2295748|  Phoenix|      85018|
|     Liberty Market|      2167793|  Gilbert|      85234|
|        Kung Fu Tea|      2015244| Las Vegas|     89146|
|       Omelet House|      1938784| Las Vegas|     89102|
|     BabyStacks Cafe|     1864128| Las Vegas|     89128|
|Breakfast Club- S...|      1767108|Scottsdale|     85251|
|      Baguette Cafe|      1748808| Las Vegas|     89113|
|     Sunrise Coffee|      1641430| Las Vegas|     89120|
|     The Bagel Cafe|      1527526| Las Vegas|     89145|
|Jean Philippe Pat...|      1522635| Las Vegas|     89109|
|   Makers & Finders|      1513824| Las Vegas|     89104|
|Grouchy John's Co...|      1479408| Las Vegas|     89123|
|Gäbi Coffee & Bakery|     1394617| Las Vegas|     89146|
|Snooze, An A.M. E...|      1385248|    Tempe|      85281|
|TIABI Coffee & Wa...|      1354700| Las Vegas|     89119|
|           PublicUs|      1304844| Las Vegas|     89101|
+-------------------+-------------+---------+-----------+
only showing top 20 rows
```

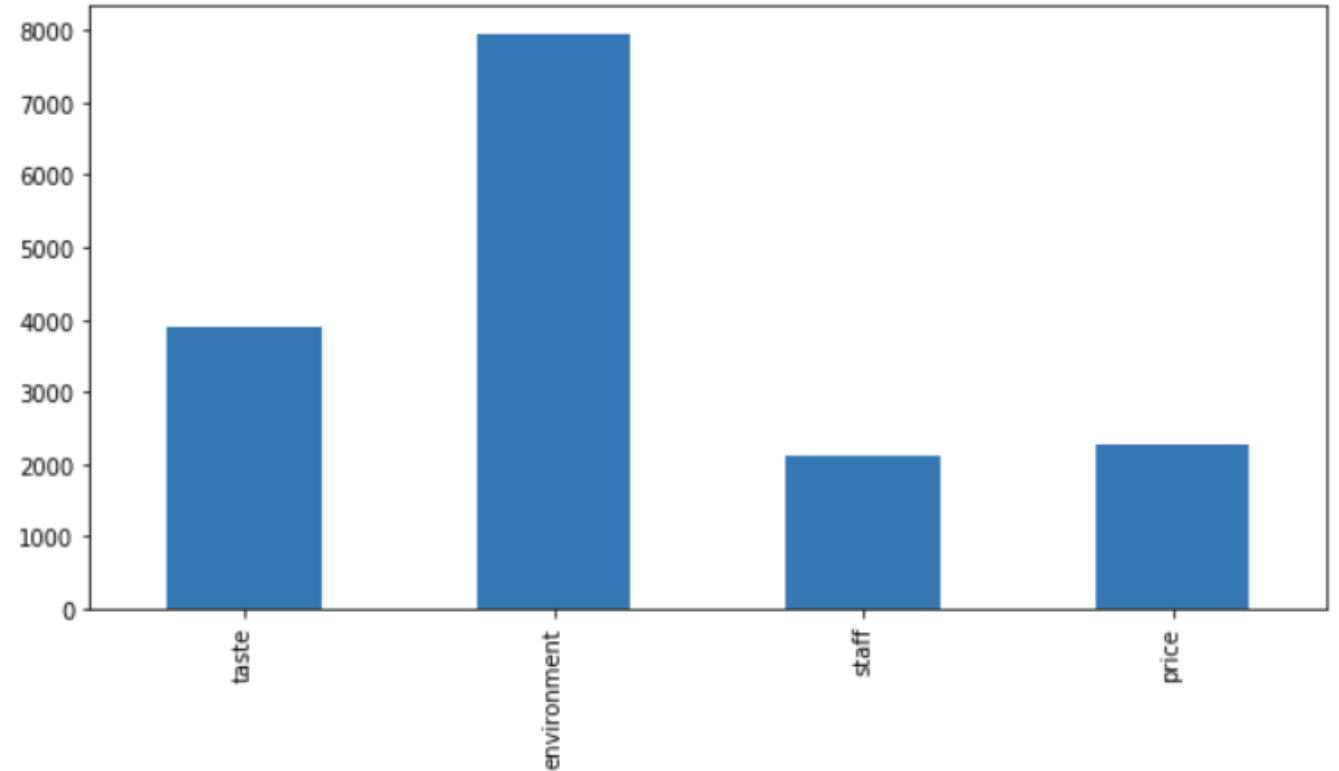|	city	review_counts
0	Phoenix	12682120
1	Phoenix	5047640
2	Charlotte	3348416
3	Scottsdale	2983750
4	Phoenix	2295748
5	Gilbert	2167793
6	Las Vegas	2015244
7	Las Vegas	1938784
8	Las Vegas	1864128
9	Scottsdale	1767108
10	Las Vegas	1748808
11	Las Vegas	1641430

The second information we could get is from dataset is from costumers' review text for the top star rating shops. I picked some key words: coffee taste, environment, staff attitude and price, to analyze what is the most important impact of customer experience about why people love these shops.

The result is different from my hypothesis, which price is the least important impact.

	name	city	avg_star	taste	environment	staff	price
0	Ro & Joe's Hot Dog Hut	Phoenix	5.000000	0	1	0	0
1	Ming Wah Hong	Montréal	5.000000	0	2	0	3
2	Timothys World Coffee	Etobicoke	5.000000	3	2	0	0
3	Icki Sticki	Mount Horeb	5.000000	3	1	0	0
4	Dicey Business	Mississauga	5.000000	2	2	2	1
5	The Knitting Loft	Toronto	5.000000	2	2	1	1
6	Classic Juice	Toronto	5.000000	0	5	1	1
7	High Society Boutique	Cleveland	5.000000	0	1	1	0
8	Bar à Gaufre	Montréal	5.000000	1	2	0	0
9	Cardinal Press	Newmarket	5.000000	3	2	2	0
10	Deseo Coffee Shack	Phoenix	5.000000	11	8	0	1
11	Épices de Cru	Montréal	5.000000	0	4	5	3
12	CrepeStar Dessert Cafe & Bistro - Aurora	Aurora	5.000000	0	4	2	1
13	Teashop	Montréal	5.000000	1	3	0	0
14	William's Coffee Co	Phoenix	5.000000	7	3	4	0
15	Candy Cravers Coffee	Litchfield Park	5.000000	2	0	0	0
16	Equal Exchange Espresso Bar	Cleveland	5.000000	3	2	0	0
17	El Güero Birria de Chivo	Phoenix	5.000000	1	4	1	0
18	The Maker Bean Cafe	Toronto	5.000000	3	2	1	1
19	Sugar Miracles	Toronto	5.000000	4	6	0	1
20	Mama Baer's	Gilbert	5.000000	7	2	0	0
21	Pret A Manger	Las Vegas	5.000000	2	3	0	0
22	Jojos Smokeless World	North Olmsted	5.000000	0	4	1	3
23	Coffee Cove	Scottsdale	5.000000	2	1	2	1
24	Restore Health Bar & Juicery	Shaker Heights	5.000000	0	1	1	0
25	Sahara Sweet	Phoenix	5.000000	0	2	0	0
26	Buckeye Chocolate Bainbridge Café & Coffee Shop	Chagrin Falls	5.000000	9	7	1	2
27	Reginald's Coffee	Bethel Park	5.000000	24	13	2	2

Then I used the column is_open to get the relationship between how many shops are shutdown now with the cities and customers' reviews. It is important to find out the negative impacts for the coffee business.

	name	city	shutdown	taste	environment	staff	price
0	Jean Philippe Patisserie	Las Vegas	1245	199	502	55	228
1	Max Brenner	Las Vegas	983	41	480	52	161
2	Payard Pâtisserie & Bistro	Las Vegas	640	154	234	35	143
3	Serenade Coffee Bar & Dessert	Las Vegas	590	296	343	76	74
4	Duza's Kitchen	Phoenix	523	91	241	102	56
5	Buddha Belly Deli	Henderson	494	30	238	62	52
6	Orange Table	Scottsdale	472	124	248	65	34
7	Crepe Shack & Waffles	Henderson	453	11	262	94	41
8	JoBot Coffee & Diner	Phoenix	449	272	226	48	34
9	Sambalatte	Las Vegas	401	285	158	47	67
10	Cafe Summer	Las Vegas	378	30	211	36	28
11	Salted Malted Bakery & Creamery	North Las Vegas	366	22	168	74	48
12	Owen's Bagel & Deli	Charlotte	362	62	163	61	30
13	Holley's Cuppa	Las Vegas	351	263	164	63	31
14	Jozo Drink Cafe	Henderson	346	147	203	46	38
15	The Beat Coffeehouse & Records	Las Vegas	342	247	189	48	40
16	French Grocery	Phoenix	342	91	179	48	37
17	Tea Time Café	Las Vegas	332	25	187	64	29
18	Pot Tea Social House	Las Vegas	322	1	151	44	44
19	Snow White Cafe	Las Vegas	291	64	155	22	30
20	Lox, Stock & Bagel	Scottsdale	285	47	158	36	46
21	Chatime	Las Vegas	281	21	154	51	11
22	Portland Variety	Toronto	278	52	137	36	58
23	Joe Maxx Coffee	Las Vegas	277	197	123	48	10
24	D'Lite Healthy On The Go	Tempe	269	37	110	36	49
25	Arepa Café	Toronto	260	13	116	34	52
26	O Face Doughnuts	Las Vegas	259	76	112	42	22
27	Rush Espresso Cafe and Wine Bar	Charlotte	256	178	124	47	28
28	Dutch Bros Coffee	Scottsdale	253	160	72	55	16
29	The Perfect Scoop & Boba Tea	Las Vegas	250	4	129	12	28
30	Panacea	Las Vegas	249	11	120	36	54

In additional, I will try to find out the relationships between the popular coffee shops and the open hours. If I could get the result shows different open hours will affect on the review counts and shop star ratings, it is a good way to provide suggestions to clients.

---

business_id	categories	name	postal_code	city	review_count	stars	text	is_open	review_id	useful	Friday	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday
4x8W4quFhhuTZlsoU...	Fast Food, Restau...	McDonald's	85258	Scottsdale	49	2.0	Have to wait abou...	1	xMG1RtUNMqQNWwGlc...	0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0
4x8W4quFhhuTZlsoU...	Fast Food, Restau...	McDonald's	85258	Scottsdale	49	1.0	I ordered a meal ...	1	1m56Im_oefWW4UY93...	4	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0
4x8W4quFhhuTZlsoU...	Fast Food, Restau...	McDonald's	85258	Scottsdale	49	1.0	This place is a j...	1	tkM7UkbfnzyN1HxAH...	0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0
4x8W4quFhhuTZlsoU...	Fast Food, Restau...	McDonald's	85258	Scottsdale	49	4.0	Have stopped here...	1	wafbiiYUA7AduQFrc...	0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0
4x8W4quFhhuTZlsoU...	Fast Food, Restau...	McDonald's	85258	Scottsdale	49	1.0	I come to this lo...	1	GJYz-KM-VD6pCwKLP...	0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0
4x8W4quFhhuTZlsoU...	Fast Food, Restau...	McDonald's	85258	Scottsdale	49	4.0	On my way to Ikea...	1	B1JBWNia7Bj-MsrpQ...	0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0
4x8W4quFhhuTZlsoU...	Fast Food, Restau...	McDonald's	85258	Scottsdale	49	1.0	Yes, they are ope...	1	L8dRCMHMoByjE41W9...	0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0
4x8W4quFhhuTZlsoU...	Fast Food, Restau...	McDonald's	85258	Scottsdale	49	5.0	Best Mcdonalds EV...	1	vmxsCd9kEVP_uF_HU...	0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0
4x8W4quFhhuTZlsoU...	Fast Food, Restau...	McDonald's	85258	Scottsdale	49	1.0	Defenitely The wo...	1	UvFmIHdFqWZJQjkfh...	1	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0
4x8W4quFhhuTZlsoU...	Fast Food, Restau...	McDonald's	85258	Scottsdale	49	1.0	This is the worst...	1	_CcgGSRBLvGVXgk3J...	0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0
4x8W4quFhhuTZlsoU...	Fast Food, Restau...	McDonald's	85258	Scottsdale	49	5.0	I come to this lo...	1	-kRDuLs3ECpUpCZGW...	0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0
4x8W4quFhhuTZlsoU...	Fast Food, Restau...	McDonald's	85258	Scottsdale	49	3.0	I am filled with ...	1	CkkEKsVQHrzO1DD_i...	2	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0
4x8W4quFhhuTZlsoU...	Fast Food, Restau...	McDonald's	85258	Scottsdale	49	4.0	This is one or th...	1	-fvXOPyYMyJ8uyBZt...	0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0
4x8W4quFhhuTZlsoU...	Fast Food, Restau...	McDonald's	85258	Scottsdale	49	5.0	Awesome service e...	1	FQwQ8T2DmE16QLL9N...	0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0
4x8W4quFhhuTZlsoU...	Fast Food, Restau...	McDonald's	85258	Scottsdale	49	2.0	Fries are saltier...	1	MwR5HQt4TqjzsMk0B...	0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0
4x8W4quFhhuTZlsoU...	Fast Food, Restau...	McDonald's	85258	Scottsdale	49	1.0	Worst McDonald's ...	1	MOHvOmJfnn-Ftx07d...	0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0
4x8W4quFhhuTZlsoU...	Fast Food, Restau_...	McDonald's	85258	Scottsdale	49	1.0	Went for coffee a...	1	G1OdMGxxXLfYXcr5_...	4	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0
4x8W4quFhhuTZlsoU...	Fast Food, Restau...	McDonald's	85258	Scottsdale	49	3.0	McDonalds. Hummm...	1	d5hwEAYO1pWfm1eWT...	1	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0
4x8W4quFhhuTZlsoU...	Fast Food, Restau...	McDonald's	85258	Scottsdale	49	1.0	Eff you McDonalds...	1	X2-QXi1rjnQthAhZY...	22	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0
4x8W4quFhhuTZlsoU...	Fast Food, Restau...	McDonald's	85258	Scottsdale	49	1.0	Terrible. For som...	1	_n4qvBcpEmfPgLTsL...	0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0	5:0-1:0

DEEPER ANALYSIS

Top 20 Coffee Shops

Coffee shops with most reviews distributed in these cities.

As we could see in this table, **Phoenix, Las Vegas and Scottsdale** have the most popular coffee shops. These areas might be the good places for a coffee shop business.
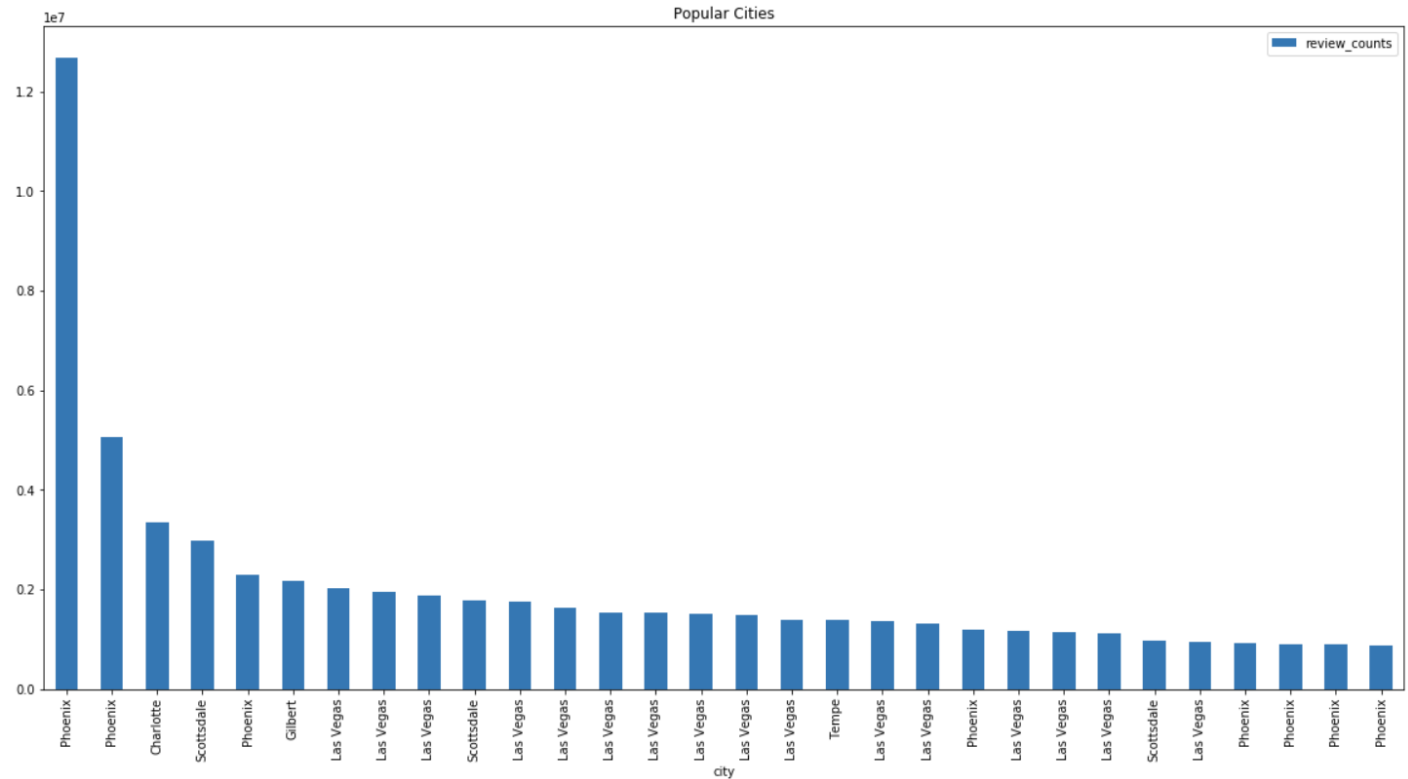
Shutdown Shops Distribution

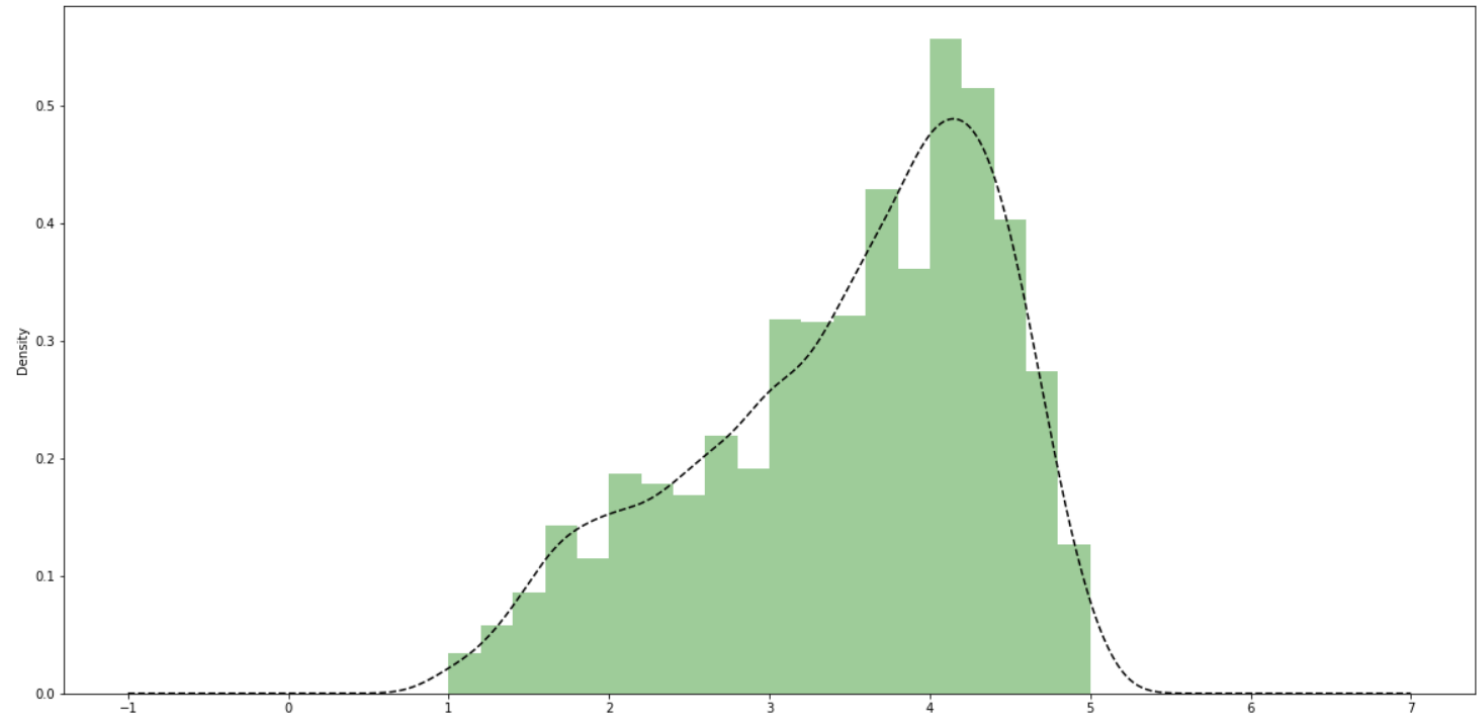This table shows the result of cities have biggest numbers of shutdown coffee shops.

As we could see, Toronto has the biggest number. Also Las Vegas and Phoenix are in the table, because they have the large base.
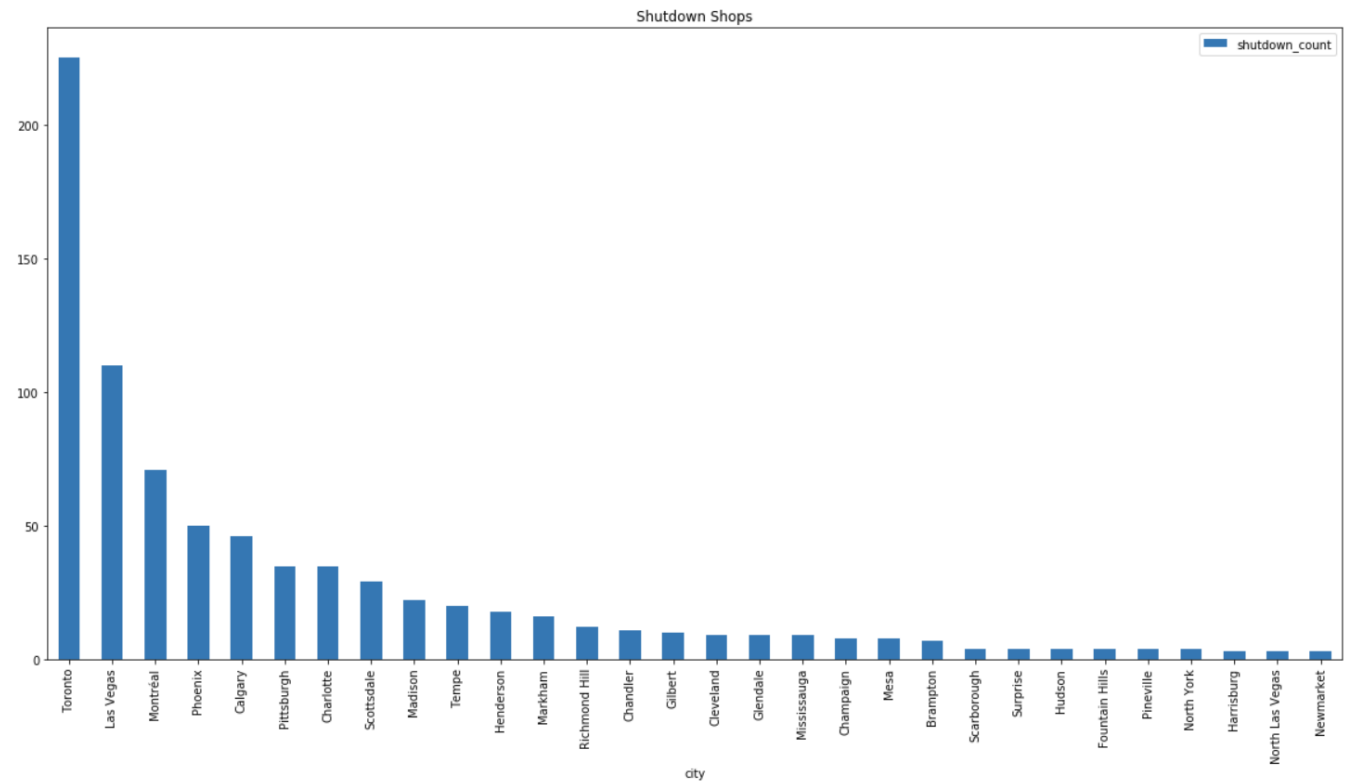
Star Rating
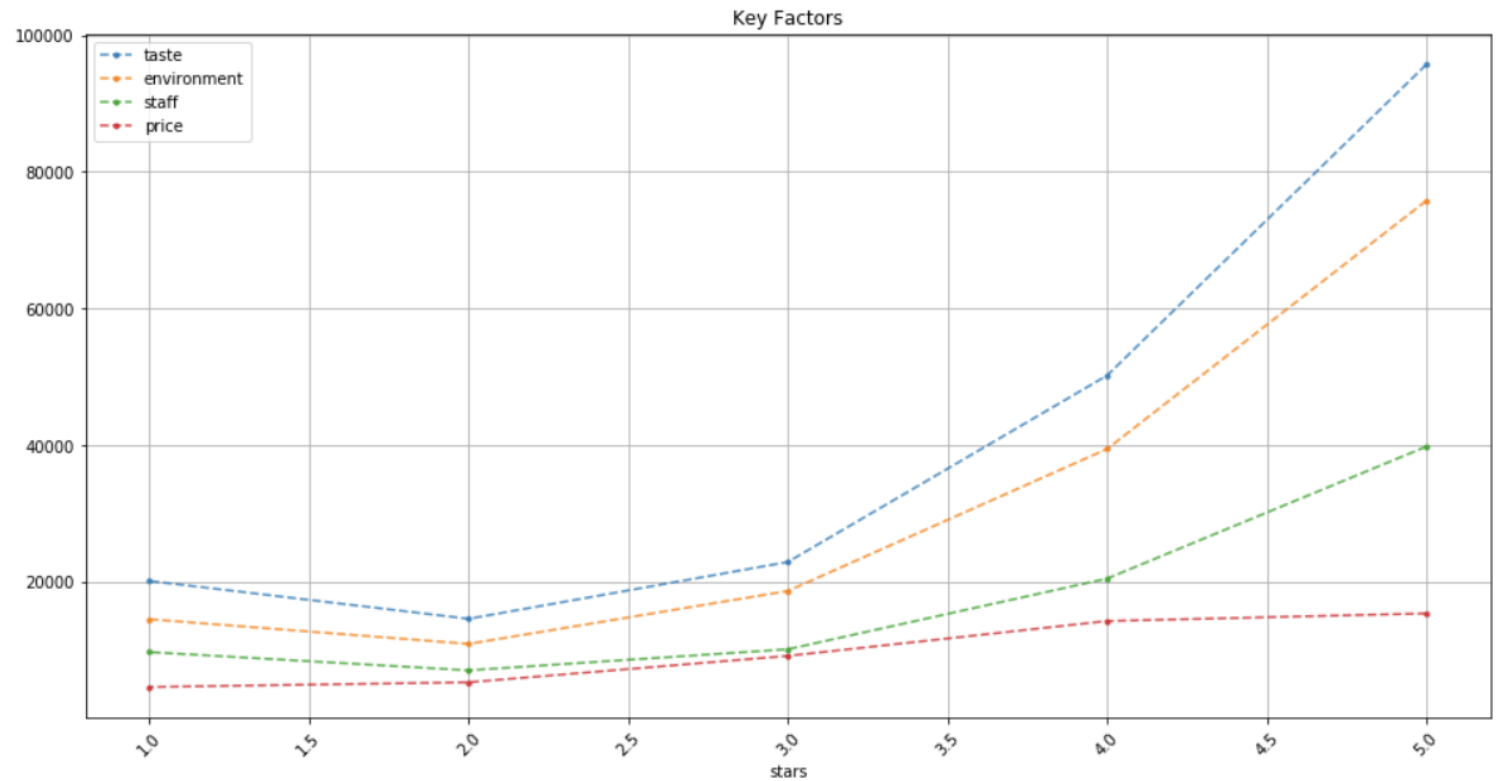
This table shows the density of star ratings.

As we could tell from the table, shops rating around 4.0-4.5 are the most common ones.

Key Factors

Customers' experience is very important to a food business.

This table shows four key factors: taste, environment, service, and price from customer reviews to analyse which part is the customers most focusing on.

Snooze Analysis

	name	city	postal_code	star	Friday	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday
0	Snooze, An A.M. Eatery	Tempe	85281	4.356661	6:30-14:30	6:30-14:30	6:30-14:30	6:30-14:30	6:30-14:30	6:30-14:30	6:30-14:30
1	Snooze, An A.M. Eatery	Phoenix	85016	4.151330	6:30-14:30	6:30-14:30	6:30-14:30	6:30-14:30	6:30-14:30	6:30-14:30	6:30-14:30
2	Snooze, an A.M. Eatery	Scottsdale	85254	4.317317	6:30-15:0	0:0-0:0	6:30-15:0	6:30-15:0	6:30-15:0	6:30-15:0	6:30-15:0
3	Snooze, An A.M. Eatery	Phoenix	85044	4.078431	6:30-14:30	6:30-14:30	6:30-14:30	6:30-14:30	6:30-14:30	6:30-14:30	6:30-14:30
4	Snooze an A.M. Eatery	Gilbert	85234	3.887339	6:30-14:30	6:30-14:30	6:30-14:30	6:30-14:30	6:30-14:30	6:30-14:30	6:30-14:30

Snooze is the coffee shop with most customer reviews which means it could be the most popular shop. So I picked this shop as an example for a deeper analysis.

As the table shows, they have business in 4 cities as we analyse before, Phoenix and Scottsdale are on the list. The star rating is around 4.0. And the business hours are from 6:30-15:00 which are earlier than other shops. It means they are focusing on breakfast and brunch.

Key words for Snooze

As we could see from the top words from Snooze reviews, breakfast and food are the customers' focusing point.

Place and service are also on the list which is important customer experience.

Price seems not that important in these analyze results.

Technical challenge

1. As the analysis is based on customer review text, semantic analysis to get the key words is the challenge for me.

2. Second the dataset is a large JSON file. Loading and transferring data is another difficult point. I chose to use spark SQL to operate big dataset.

```python
from pyspark.sql import SparkSession

spark = SparkSession \
    .builder \
    .appName("Python Spark SQL basic example") \
    .getOrCreate()

# df = spark.read.csv("/home/stp/test1.csv",header=True,sep="|");
spark
```

SparkSession - in-memory
SparkContext

Spark UI
Version
v2.4.5
Master
local[*]
AppName
Pi

```python
from collections import Counter

no_sw = topshop_analysis['text'].apply(lambda word: [word for word in word.split(' ') if word not in stopwords] )
corpus=[]
for ea in no_sw:
    corpus+=ea
wordDict=Counter(corpus)
[(k, v) for k, v in sorted(wordDict.items(), key=lambda item: item[1], reverse = True )][:20]
```

```python
from wordcloud import WordCloud

wordcloud = WordCloud().generate_from_frequencies(wordDict)
plt.imshow(wordcloud, interpolation = 'bilinear')
plt.axis('off')
plt.show()
```

Conclusion

As Coffeeking company is focusing on customers experience. This project has a deep analysis on customer reviews to provide clients a unique and targeted marketing plan.

From the different cities business analyse, we could find the best place for coffee shops such as **Phoenix, Las Vegas and Scottsdale.**

Operating hours could start early to get the breakfast market, the crowd hours seems to be big in the morning and brunch time.

From the key words analysing, the key factors ranking:

1.Food taste

2. Environment

3. Service

4. Price

Taste is the most important factor; price appears not quite much in customer reviews.