Yelp Dataset Analysis Report

CS686 Big Data Processing Personal Project

Reporter: Zini Zhu

Date: 2020 May 1st

General Information

Business #

192,609

Users #

1,637,138

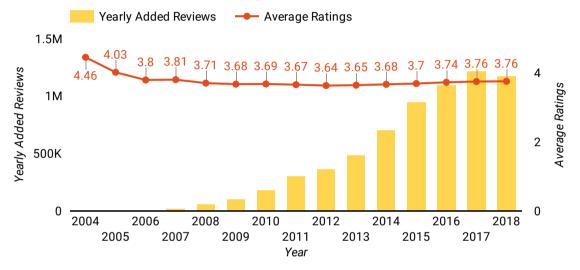
Reviews #

6,685,900

Tips #

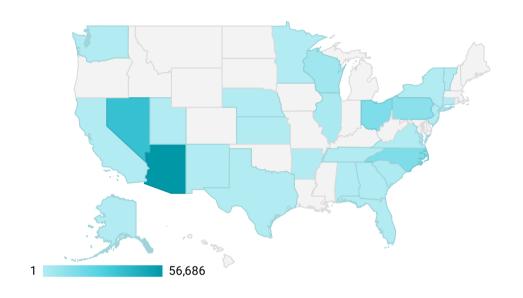
1,223,094

Total reviews & average stars for each business

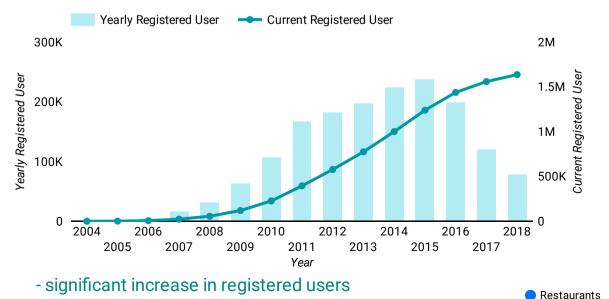


- significant increase in reviews each year
- average stars stabilize at about 3.7

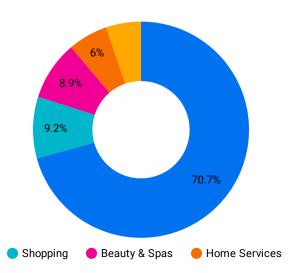
Business distribution



New registered users & total registered users each year



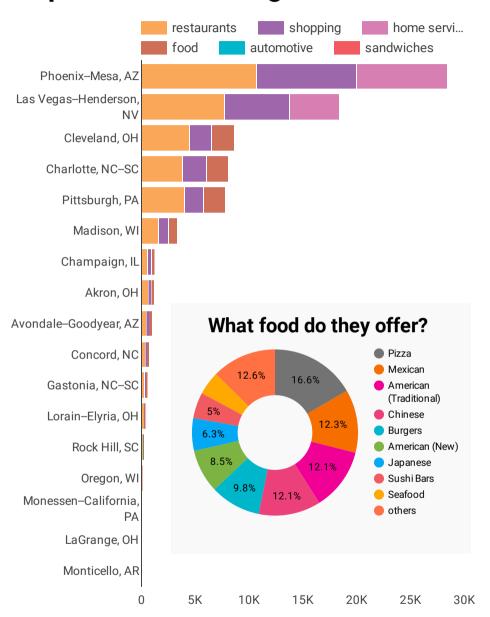
Business categories distribution





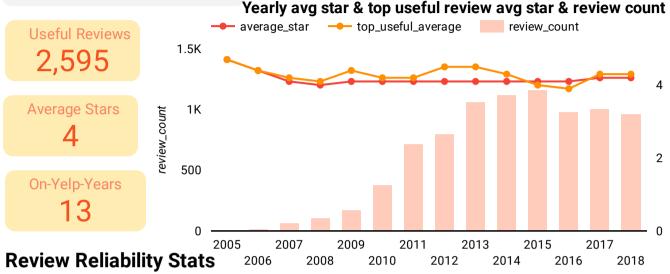
Business Analysis

Top 3 business categories in cities



*In large cities(like Las Vegas), different businesses tend to distribute evenly, while in smaller cities, food-related businesses take a big portion.





On-Yelp Yrs ▼	Reviews #	Users #	Useful per Revi	Star Std Devi	Suspicious User
13	8,570	8,349	0.8	1	1

	Review	stars	useful	month	year
1	I . I really love the Las Vegas Strip and I have eaten at Mon Ami Gabi so many times!	5	55	October	2018

This is really an awesome restaurant at many levels!

I love the atmosphere in the morning for a nice breakfast watching the Bellagio famous fountains directly across Las Vegas Blvd with a great view from the patio part of the restaurant! Talk about Ambience. The atmosphere sitting on the patio is awesome enjoying the morning sun and on the inside very beautiful for fine dining!

The staff is very professional, the restaurant is very clean, the tables are set of high class, the prices reasonable for being on the Las Vegas Strip and totally a 5 star service from the servers.

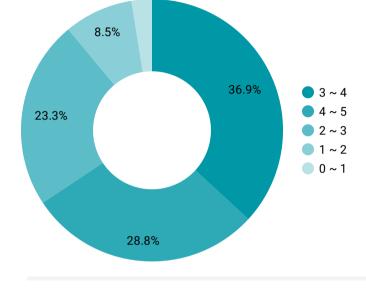
The food is delicious "YUM" and food comes to the table with great presentation "warm or cold" depending on what you ordered.

I even had an awesome holiday dinner there once "Thanksgiving" and it was awesome from start to finish! So filling in a Yum and Delicious Way! :)

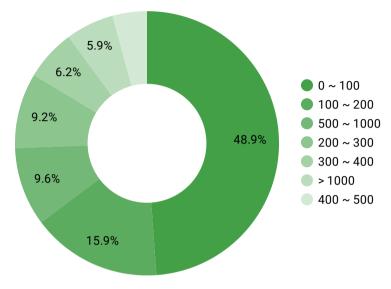
I highly recommend this place for all your friends

User Analysis

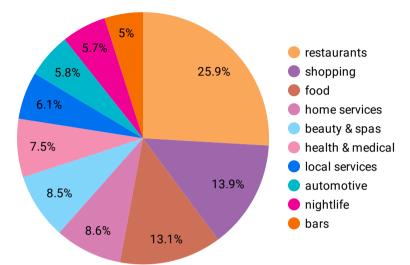
Users Distribution by Average Stars Given



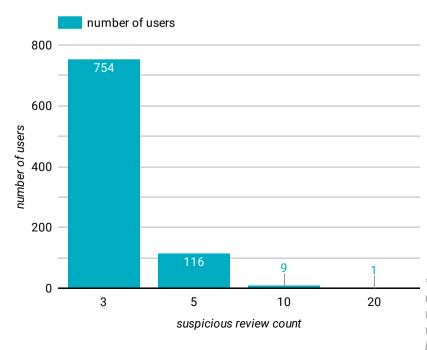
Users Distribution by Number of Reviews Made



Top 10 Categories Users Frequently Review On



Suspicious Users Identification



watch out!!

0.046% of all users are identified as suspicious users with a threshold number of similar reviews of **3**.

* Suspicious users are identified as users that made more than thethreshold number of reviews with the same rating on the same business within one year.

Popular Users Count

Average Fans 1.47

Average Reviews 22.29

Average Useful 40.53

Review #	Fan #	AUPR	User Count ▼
1000	100	5	384
2000	200	6	66
3000	300	8	6
5000	500	10	1

^{*} UPR: Average Useful Per Review

384 popular users out of 1637138(0.02%) total users reviewed on **179** popular businesses out of **348(51.4%)** total popular businesses.

Their impacts are huge!

^{*} Popular users are identified as users with number of reviews, number of fans and average useful per review all exceeding threshold value.