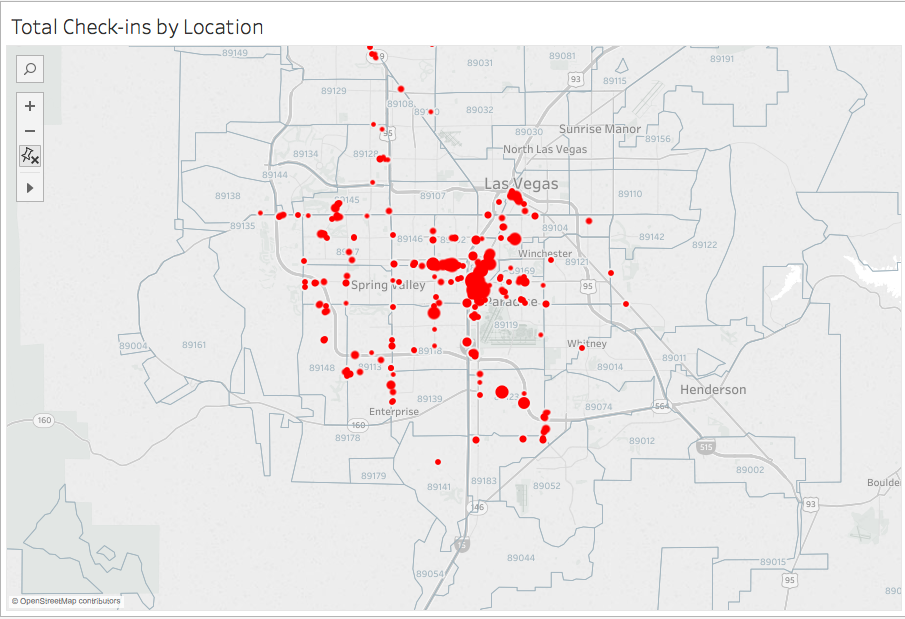
**DATASET:**

***Link:*** [***https://www.yelp.com/dataset\_challenge***](https://www.yelp.com/dataset_challenge)

* We acquired Yelp database from the Yelp website, which was in “JSON” format.
* We converted these JSON files into CSV files using python
* We then filtered the Business categories in Excel to focus on just restaurants
* We worked on the following:
  + Business: This contained all the information of the restaurant such as:
    - Location
    - Whether it is good for lunch, dinner, breakfast, drinks etc.
    - Ambience
    - Price Range
    - Alcohol information
    - Hours
    - Stars on Yelp
  + Check-in: This contained the check-in data for all the restaurants based on the day and time of the day.
  + User: This contained the reviewer data such as number of followers, how many people voted that reviewer to be useful etc.

# **INSIGHT 2a**

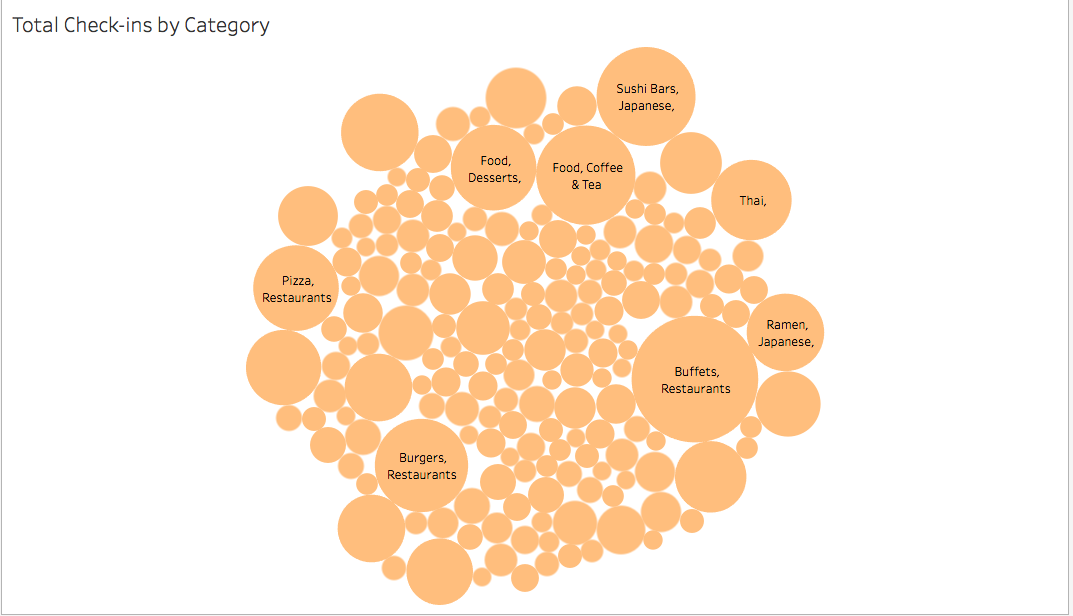
* As explained above, we decided to focus on Las Vegas. This made sense to us as well, because Las Vegas has a lot of influx of tourists and it would have great potential to open a restaurant.
* We decided to focus on check-ins in the city of Las Vegas.
* We wanted to focus on areas with good restaurants and ones that were popular.
* Therefore, we filtered with the Average stars to be at least 3. And minimum number of check-ins to be 1000.
* We created a calculation that calculated the total check-ins of the restaurants after adding the hour wise check-in information.
* The result is the map below.



* Quite visibly there are certain areas that have more check-ins and better rated restaurants.
* This area has the Zip code 89109, which is the Las Vegas strip.

# **INSIGHT 2b**

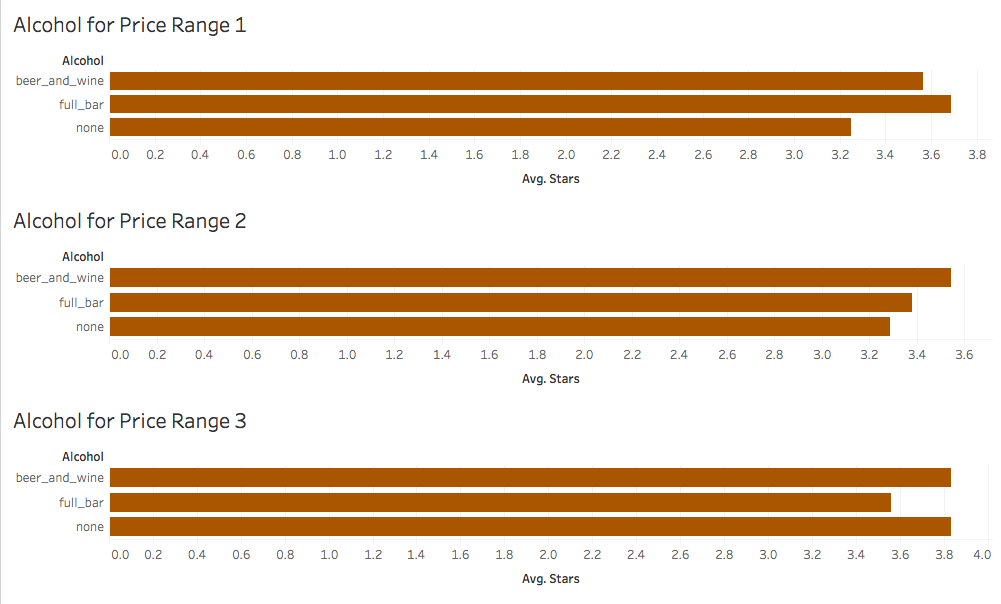
* We then wanted to focus on what categories of restaurants do better than the others.
* Again, we used the same filters as mentioned above (Average stars = 3 and Minimum check-ins = 1000)
* The following is the resulting bubble chart.

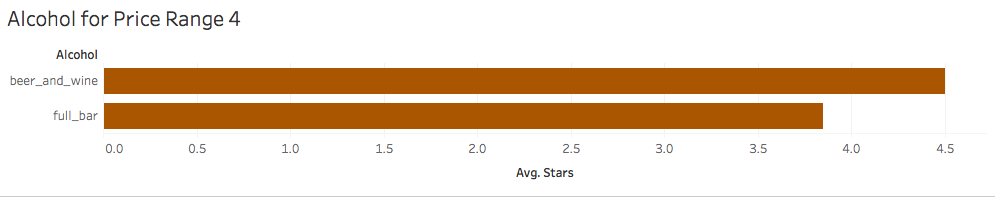


* According to the visualization above, it was clear that Restaurants with category “Buffet” were the most popular and highly rated. Followed by Cafes and Sushi Bars.

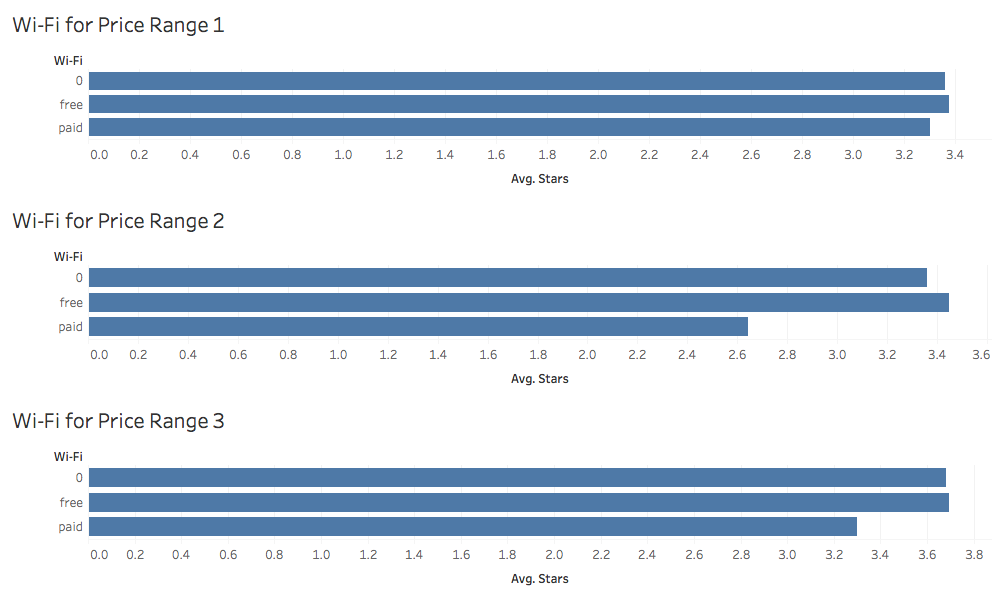
# **INSIGHT 3**

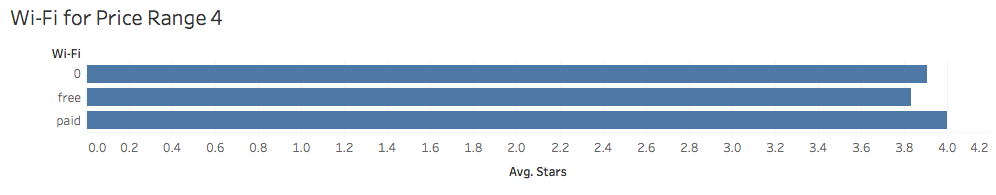
* Another important and interesting visualization was to compare different factors that a restaurant in different price ranges need to be successful.
* Therefore, we visualized these factors according to the Average stars they received.
* We filtered those restaurants that did not have this information available.
* The first factor we compared was Alcohol



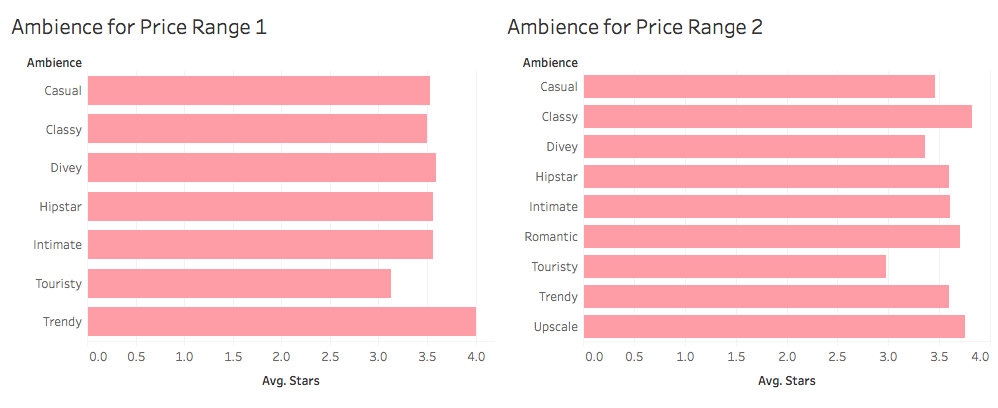


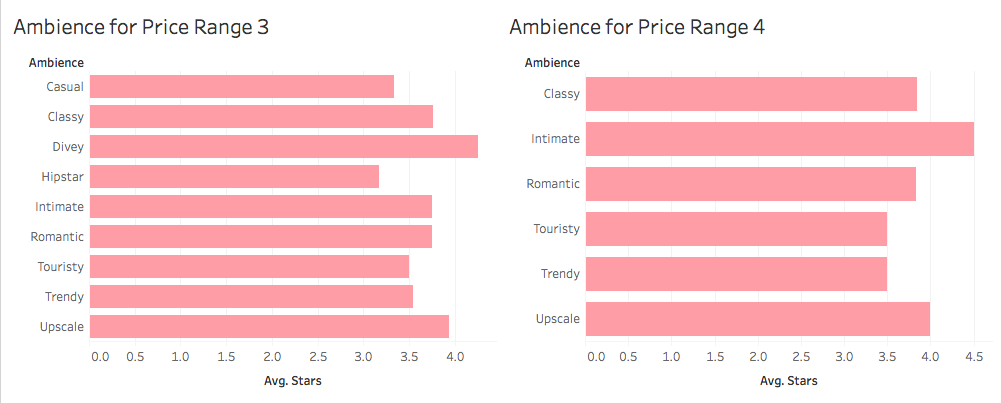
* The findings of this visualization suggested that for cheaper restaurants had better ratings with a Full bar (Price range 1) or Beer and Wine (Price range 2). Price range 3 had equally good ratings for No alcohol and Beer and Wine. While, If it is very expensive restaurant, it was expected to serve alcohol, and just beer and wine restaurants did better.
* The second factor we compared was Wi-Fi



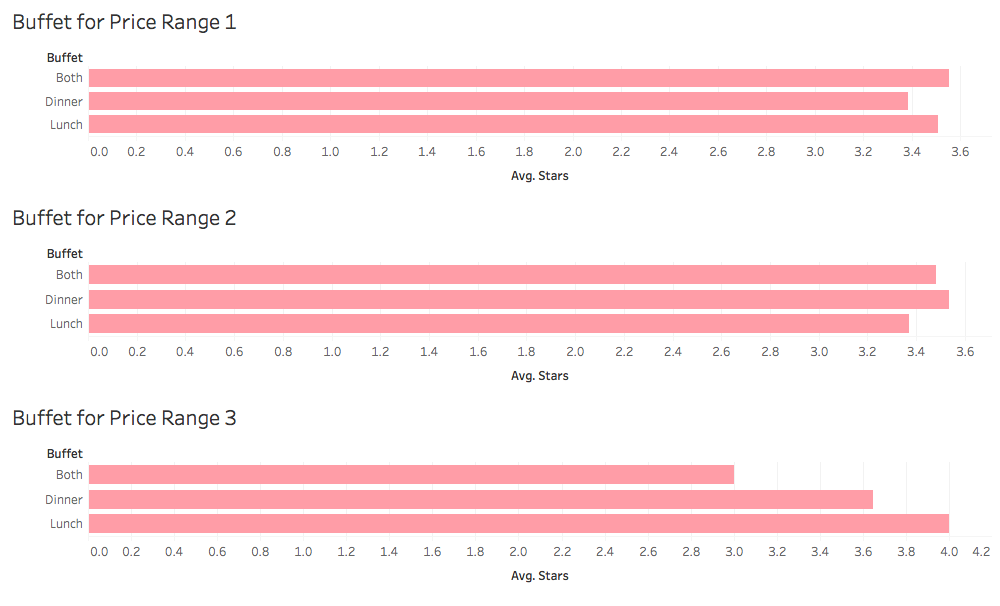


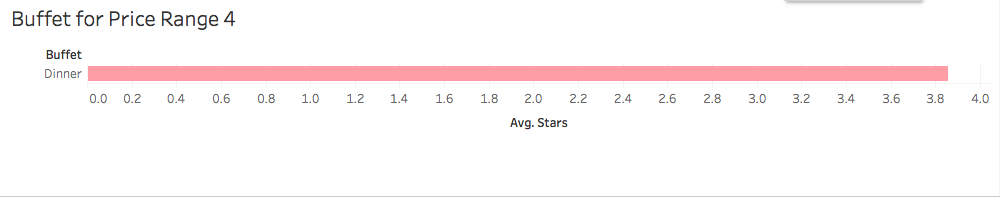
* It was clear that restaurants with no or free Wi-Fi performed better up to price range 3. Which means, it was better to have no Wi-Fi at all as compared to Paid Wi-Fi.
* But, for the really expensive restaurants, Paid Wi-Fi did better than the other two. Which meant if it was a really expensive place, people did not mind paying for Wi-Fi.
* The third factor was Ambience





* The Ambience that did best with cheaper restaurants were “Trendy” (Price Range 1) and “Classy” (Price Range 2). The Ambience for Price Range 4 that did best was “Intimate”
* The fourth and last factor was Buffet.

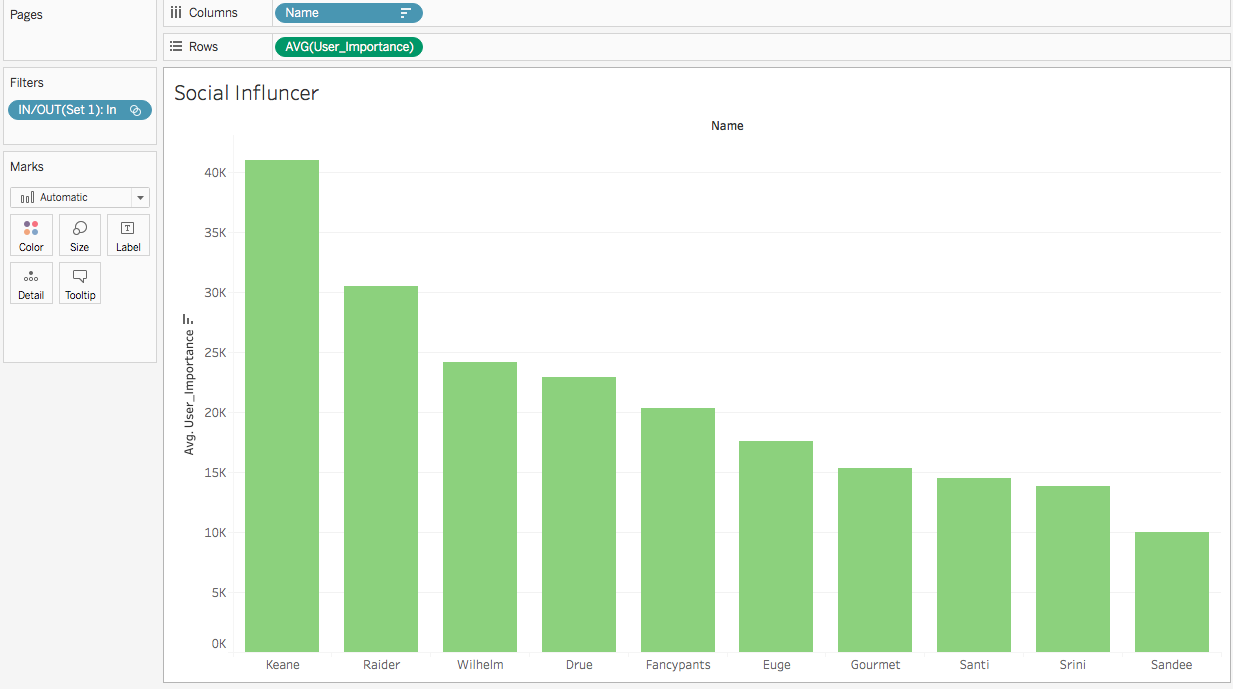




* Restaurants with Price range one performed best if they served both Lunch and dinner Buffets. Price Range 3 restaurants did best for Lunch buffets. And very expensive restaurants served buffet only for dinner.

# **INSIGHT 4**

* We then wanted to determine who would be best reviewers to get restaurants reviewed by such as to create the most buzz.
* We ran Regression using R on the various factors that cause the user to be important.
* Using this regression result we created a calculated field for user importance attaching a weight to the factors:
  + Number of Fans
  + Number of times reviewer was voted useful
  + Number of times reviewer was voted funny
  + Number of times reviewer was voted cool
* We then created a set that included only the 10 reviewers based on their User importance.



* Therefore, we got a list of the top 10 reviewers that should be contacted and invited to review a restaurant.