Business Advisor

---Solution to support business decision with up-to-date customer feedback

#### Team 11: Lingyun Yi, Ankit Phaterpekar, Jui Saba Guram, Yupeng Su, Xiaoting cui

## Business Background

People have come to yelp to leave reviews and ranking for the services they have taken since 2004. There are already more than 100 million reviews in yelp dataset and it is still growing. On the other hand, these reviews are affecting the future customers make their service decisions. If the business get relatively good rankings and positive reviews in the local area, it’s more likely to get more customers. With these two reasons, the big yelp data are extremely critical for the business success, which could help business get the direct feedbacks from their customers and also understand the need from their customers. The information getting from the data could help business optimize their services and also make the business decision. However, the business owners do not have a business intelligence tool present their up-to-date customer feedback and analysis to help them make business decision. This is the business problem we are trying to solve with our business intelligence product: Business Advisor.

Business Advisor is a business intelligence tool provide up-to date customer feedback data visualization and analysis. We are getting yelp dataset to build up our business intelligence product. In the future we are aiming to incorporate more data sources to this product. It could be Facebook, Twitter or other type of social media tool datasets. It will give more comprehensive information of the customer feedback for the business and also the understanding of the customers. Within the demo, we are using restaurant “Stripsteak” as an user sample to present our product: Business Advisor. Stripsteak restaurant is a Las vegas based Steakhouse restaurant. It has 4 stars and 985 reviews on Yelp.

## Data Collection/Storage:

The dataset contains :

* 2.7M reviews and 649K tips by 687K users for 86K businesses
* 566K business attributes, e.g., hours, parking availability, ambience.
* Social network of 687K users for a total of 4.2M social edges.
* Aggregated check-ins over time for each of the 86K businesses
* 200,000 pictures from the included businesses

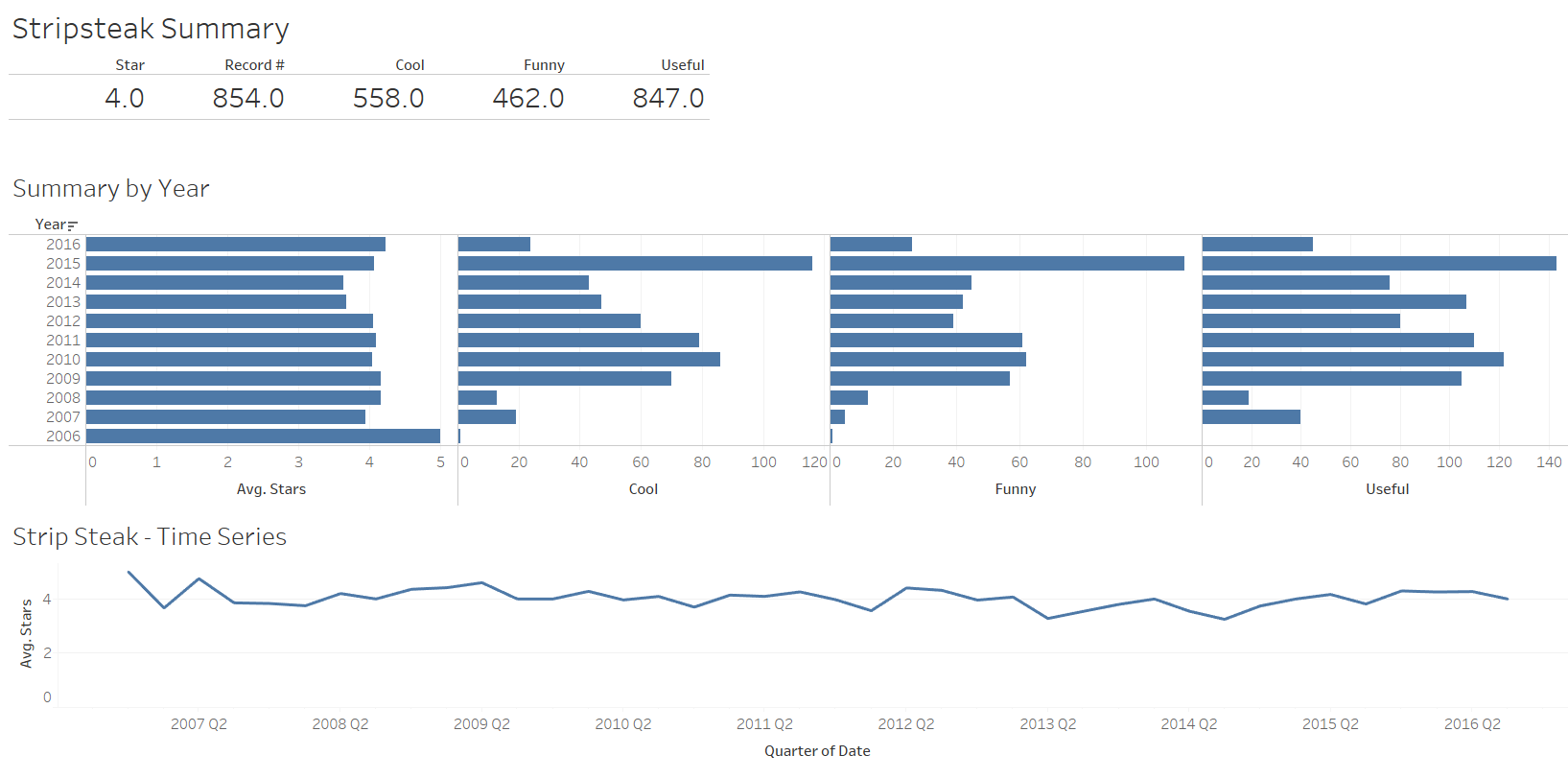
It is composed of JSON files for Yelp Business, Review and User data.

* First we set up a Yelp database in a MongoDB cluster.
* Next, we loaded the JSON files in MongoDB using the bulkimport feature of MongoDB.
* Next we used an MongoDB ODBC connector to leverage our MongoDB database into Tableau Dekstop for real-time analytics.
* Using an automatic schema generator inprovided by our ODBC connector, we created the data sources the database in Tableau. We created a schema using the connector to extract the data from Yelp database in MongoDB for running our queries.
* We also used Tableau Online to publish our worksheets for realtime code sharing and analysis.

## Business Intelligence Product Dashboard

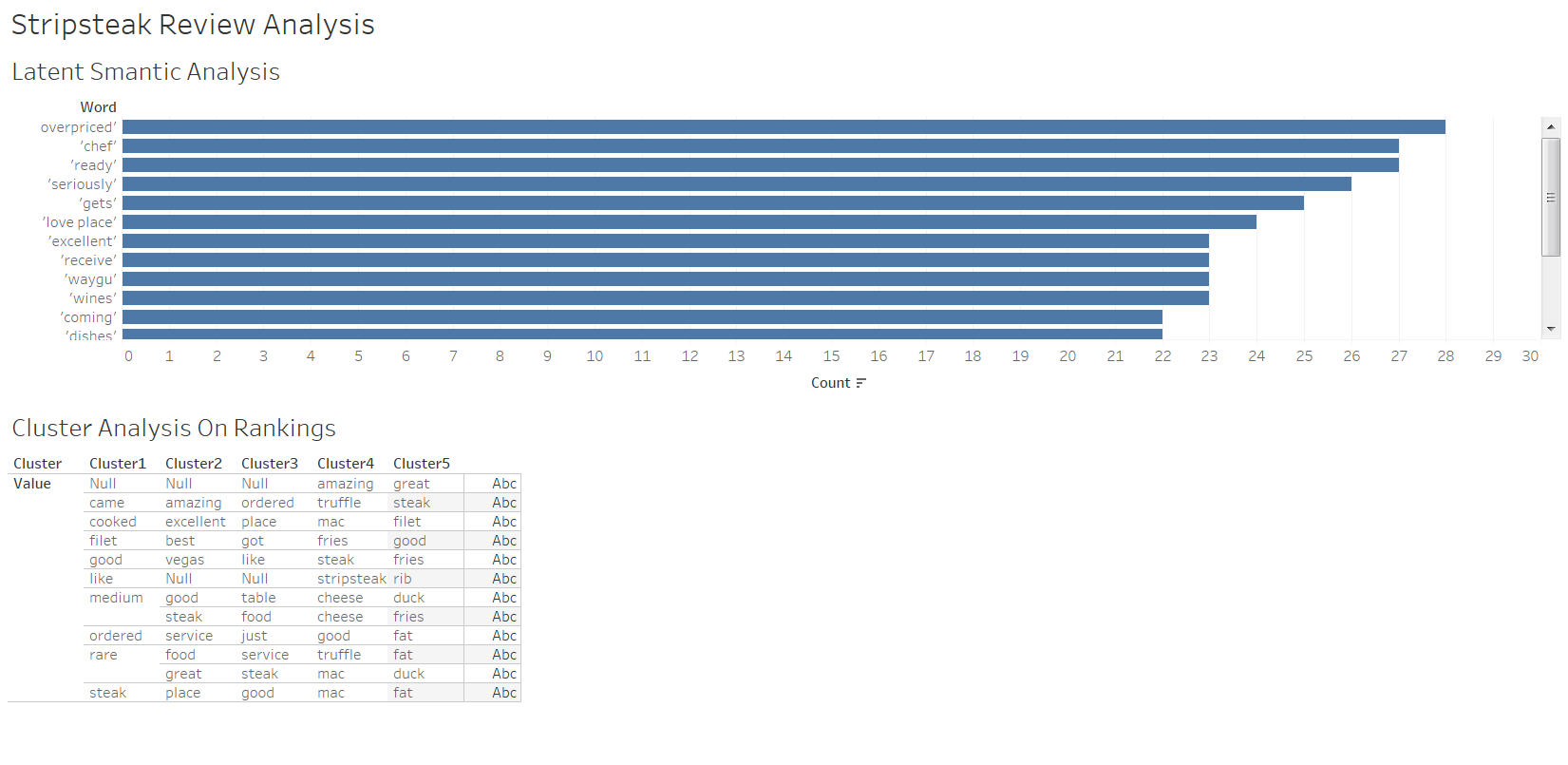
### Business Summary

On the Business Summary page, we are aiming to provide an overview sight of the business review. It displays the review measures for the business, like stars, number of review records, count of cool etc. Also, the time series charts provide insights on the ranks through times, like Year or Year-Month.



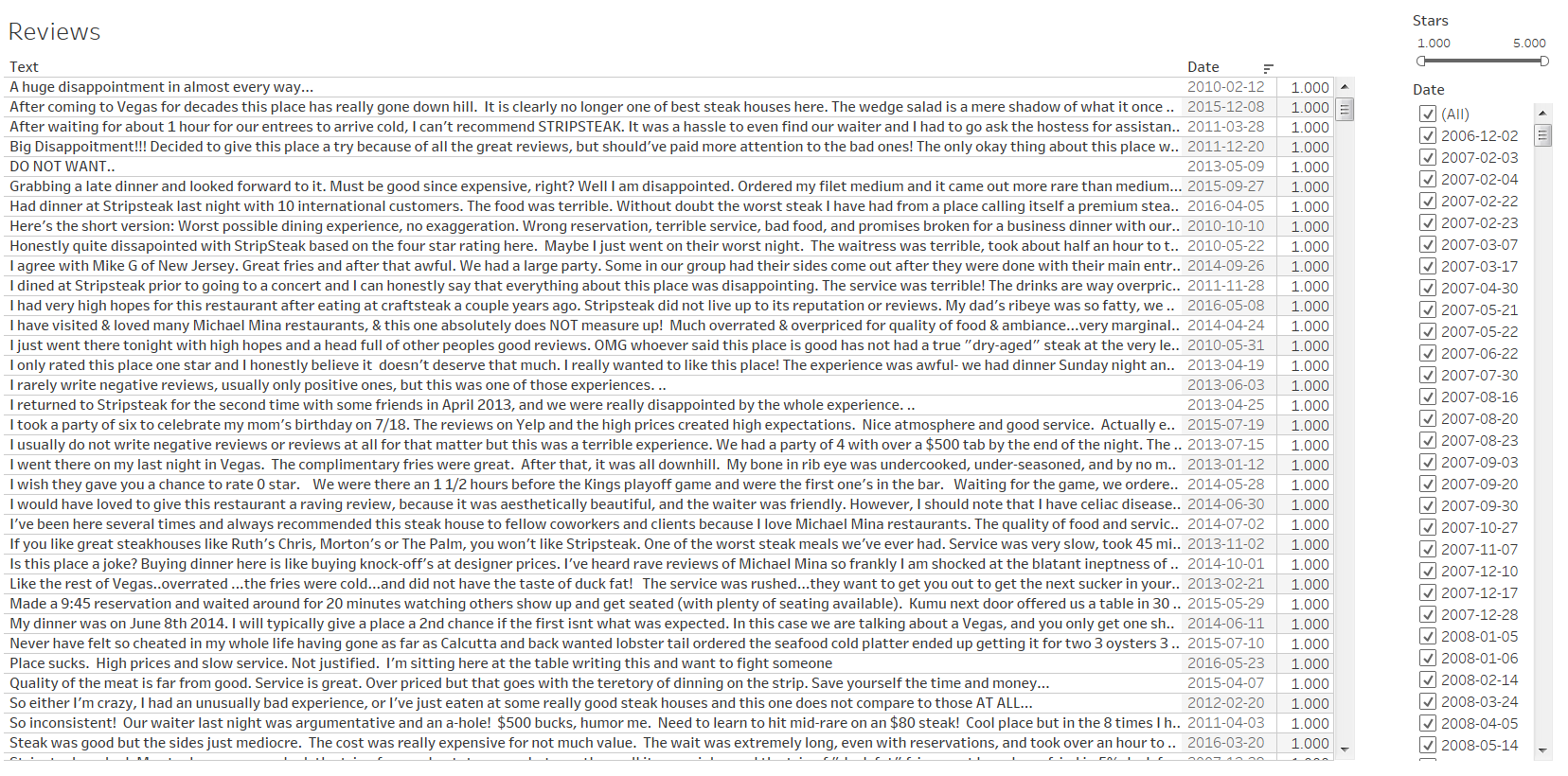
### Business Review Analysis

On Business Review Analysis page, we display analysis results for latent semantic analysis and cluster analysis on rankings. We will have more detail explain on these 2 analysis later. For the restaurant stripsteak, we could see that the word “Overpriced” shows up the most on the reviews.



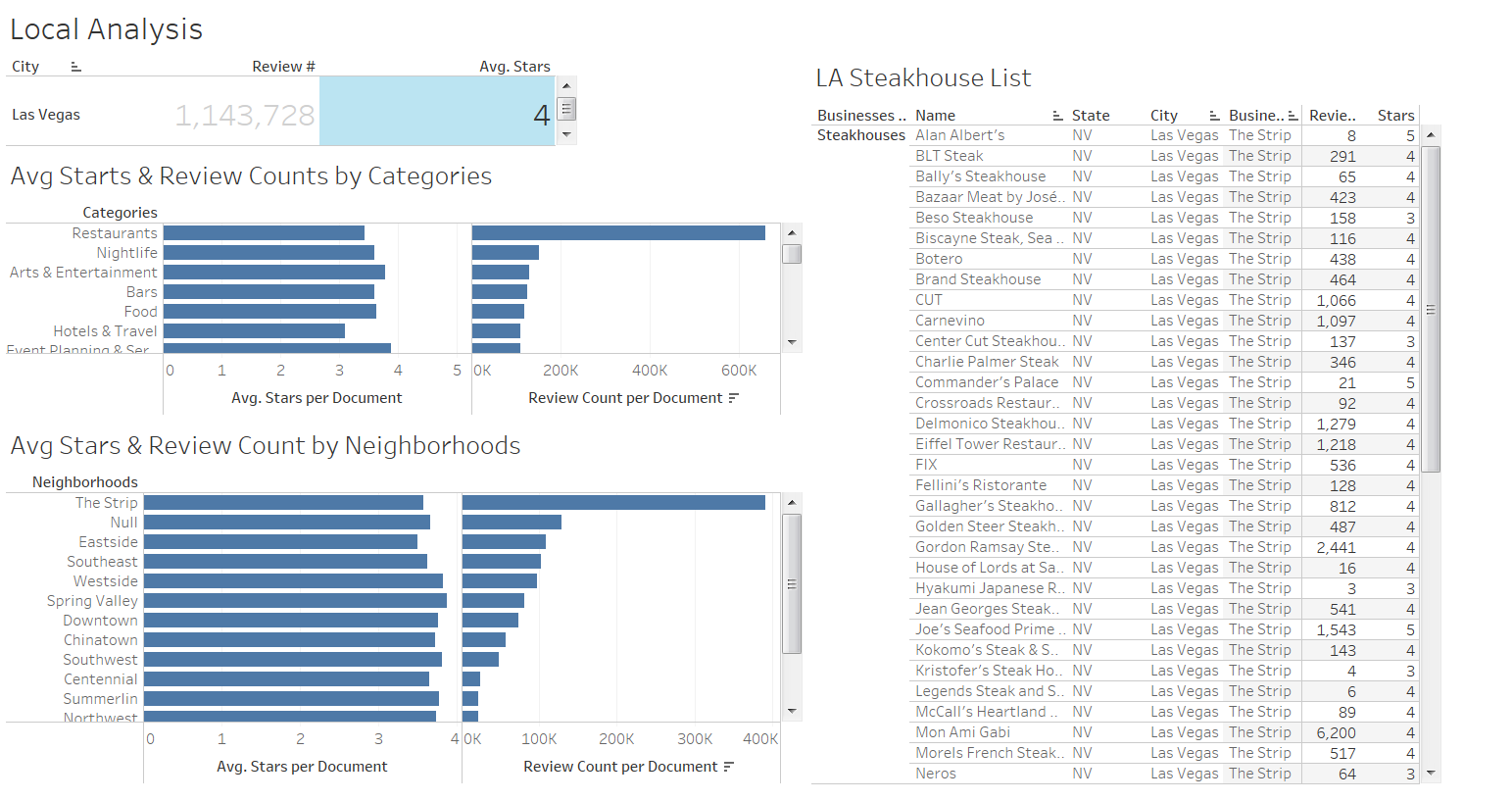
### Reviews

Within Reviews page, we are aiming to provide ability to dive on each review with 2 filters: Date & Stars.



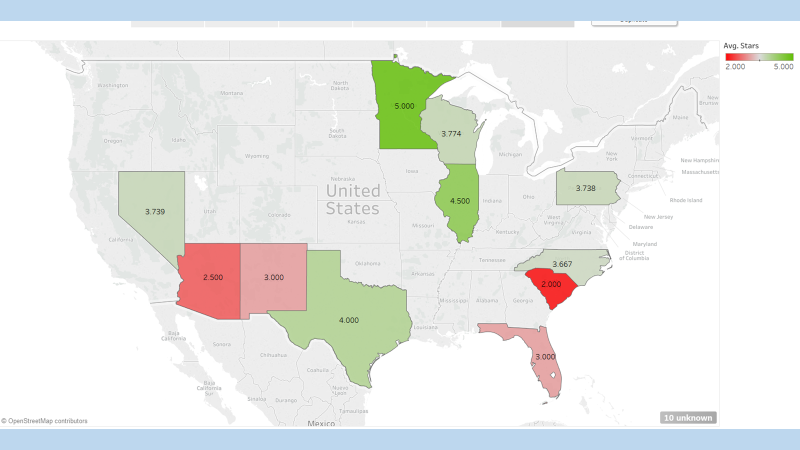
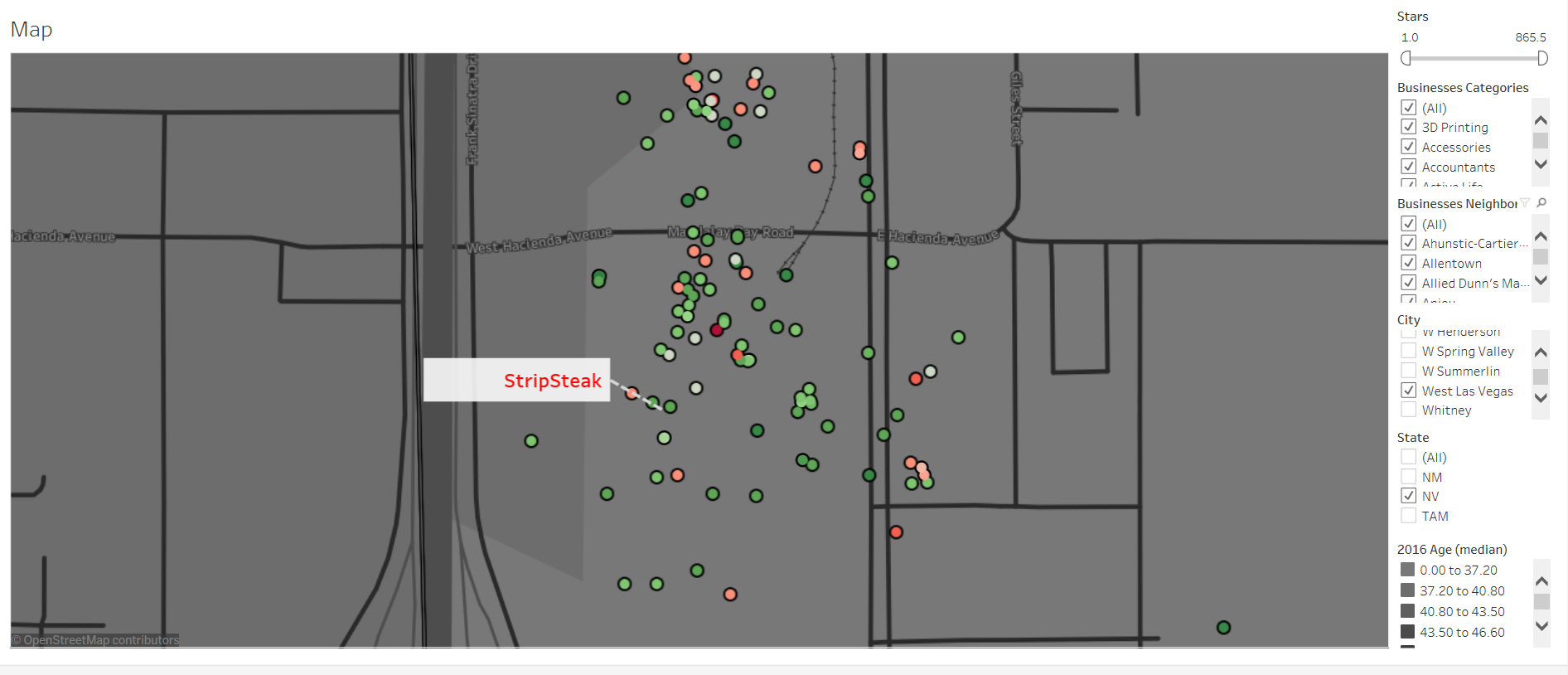
### Local Analysis

On local Analysis page, we are providing insights for the review measures of the local area, for restaurant stripsteak, the local area is Las Vegas. It displays review summary and the summary by dimension Category and Neighborhoods. On the right hand side of the page, we also provide a list of business with the same business type as stripsteak: steakhouse.



### Map

2 maps within dashboard: analysis map on the local area and review map for US states. The analysis maps provide an interactive feature to allow use to view the nearby business’s review summaries with the filter selections. The US state maps provide a bold view for US state. It will be good for business owner to make decision if they want to explore business on the other states.

## Latent semantic analysis

The Stristeak restaurant has 854 reviews, and since each review has the label stars to represent the positive or negative status, so we do only care about the review contents. That is where the semantic analysis comes into picture.

Latent semantic analysis is a process to detect the latent relation between the words. It assumes that if a document has the point, the words in that document should be related with each other to represent the point. For example, here is a review that “I dislike the hot dog, especially the sauce in it.” From this review, we can see that hot dog with dislike and sauce are come up together. If there is several more review about the hot dogs, it will be captured by the tfidf vectorization process to detect and extract as the feature, and the feature can be applied to SVD to extract several concepts which may include the main point about the hot dog.

According to the theory introduced above, we generate 854 concepts to represent 854 reviews. the part of results are as following.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| concept 0: | concept 1: | concept 2: | concept 3: | concept 4: | concept 5: | concept 6: |
| steak | duck fat | duck fat | rib | mac | happy | rib |
| good | duck | steak | cap | cheese | hour | eye |
| fries | fat fries | fat fries | steak | mac cheese | happy hour | rib eye |
| great | fat | duck | rib cap | best | steak | great |
| food | fries | fat | kobe | steak | michael | rib cap |
| cheese | cheese | great | wagyu | best steak | stripsteak | cap |
| ordered | mac | fries | eye | truffle | mina | happy |
| place | mac cheese | best | rib eye | amazing | michael mina | happy hour |
| just | truffle | vegas | best steak | truffle mac | best steak | hour |
| service | truffle mac | best steak | beef | vegas | strip steak | dinner |
| duck | trio | awesome | american | ve | kobe | staff |
| fat | truffled | happy hour | best | excellent | like | food |
| mac | truffled mac | happy | american kobe | strip | best | vegas |
| like | sauce | food | aged | great | medium | wonderful |
| mac cheese | filet | place | foie | simply | ordered | experience |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| concept 7: | concept 8: | concept 9: | concept 10: | concept 11: | concept 12: | concept 13: | concept 14: |
| foie | rib | michael | kobe | foie | great | best | strip |
| foie gras | steak | mina | american | gras | medium | ve | strip steak |
| gras | eye | michael mina | american kobe | foie gras | rare | good | excellent |
| michael | rib eye | dry | food | best | strip | cap | foie |
| mina | hour | steaks | burger | ve | wine | rib cap | foie gras |
| michael mina | mac cheese | aged | cap | filet | kobe | best steak | gras |
| happy hour | mac | dry aged | rib cap | rare | mina | wagyu | rib cap |
| good | happy hour | eye | beef | medium | michael | wine | cap |
| hour | cheese | rib eye | kobe beef | steaks | medium rare | server | rib |
| happy | medium | medium | kobe rib | meal | michael mina | food | food |
| sliders | rib cap | bone | restaurant | wine | happy hour | hour | wagyu |
| aged | happy | fries | fat fries | medium rare | strip steak | rib | steak |
| dry | cap | fat fries | time | sliders | excellent | happy hour | wagyu rib |
| menu | rare | table | certificate | best steak | beef | filet | bay |
| kobe | dry aged | duck fat | table | marrow | happy | simply | staff |

Each concept represent the corresponding review’ main point. So to have a general idea about the whole concepts, we do some data exploration to count the words frequency it shows that from the reviews, lots of them talk about the price, and food. More people think that the it is overpriced, others think that it is worth. From this analysis we can conclude that the food is yummy, service is so good that can offer them a happy hour, however price should be modified slightly may make more profits.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Word | Count | Word | Count | |
| overpriced' | 28 | 'coming' | | 22 |
| 'ready' | 27 | 'dishes' | | 22 |
| 'chef' | 27 | 'sea' | | 22 |
| 'seriously' | 26 | 'yummy' | | 22 |
| 'gets' | 25 | 'worth' | | 21 |
| 'love place' | 24 | 'cocktails' | | 21 |
| 'receive' | 23 | 'absolutely' | | 21 |
| 'excellent' | 23 | 'wow' | | 21 |
| 'wines' | 23 | 'happy hour' | | 20 |
| 'waygu' | 23 | 'service great' | | 19 |

**How is my restaurant doing?**

The above question can be simply answered by looking at the number of 5 star reviews ,4 star reviews received for the restaurant against the total number of reviews received. Since the total number of reviews are 854, here is the table for review count and percentage.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 5 Star | 4 Star | 3 Star | 2 Star | 1 Star |
| Reviews | 396 | 229 | 103 | 79 | 47 |
| Pct Reviews | 0.4637 | 0.26815 | 0.120609 | 0.092506 | 0.055035 |

**Semantic analysis of Reviews**

We trained our data on the reviews obtained using SGDC classifier so that we can predict our future reviews to which category of stars it belongs to. We used count vectorizer and tfidf transformer to transform the words into numbers so that machine learning algorithm can accept it.

Here are the results obtained from it

61.2% accuracy

array([[ 0, 1, 2, 0, 1],

[ 1, 11, 2, 3, 0],

[ 0, 10, 7, 6, 9],

[ 0, 3, 4, 12, 28],

[ 2, 5, 1, 5, 101]])

**"loved the resturant,really good chicken.The servers are friendly" Predicted : 4 stars**

**"Hated the resturant,very bad food.The service sucks” Predicted :1 stars**