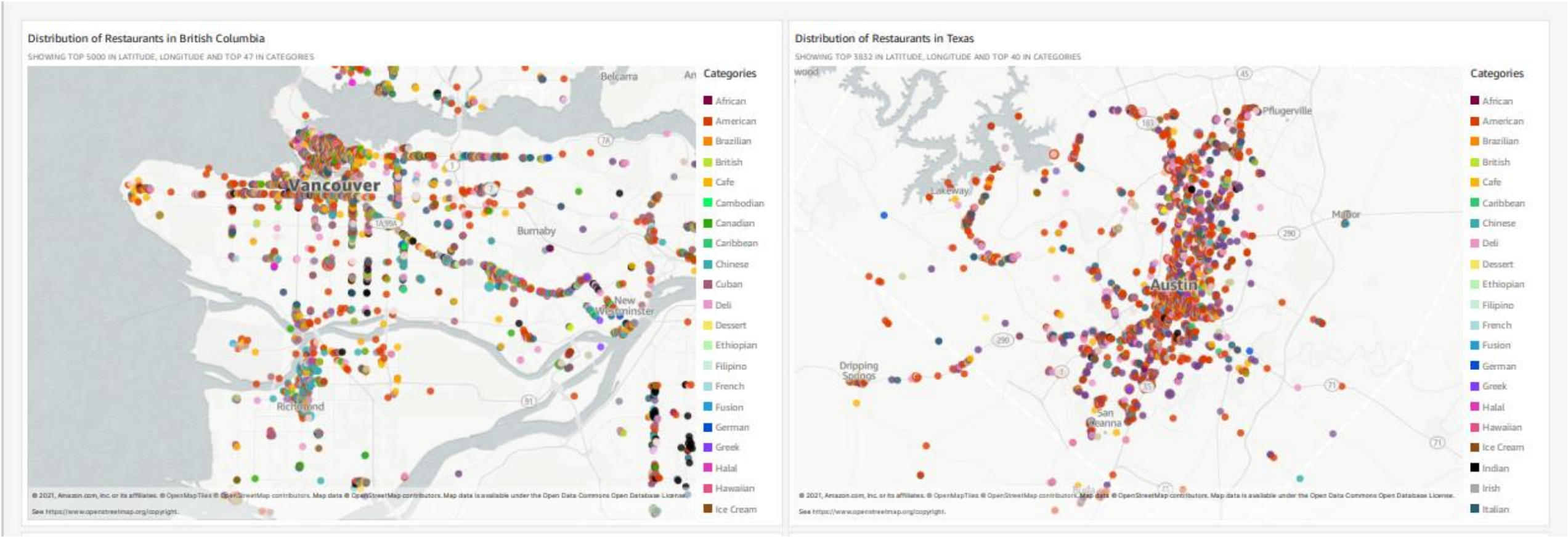
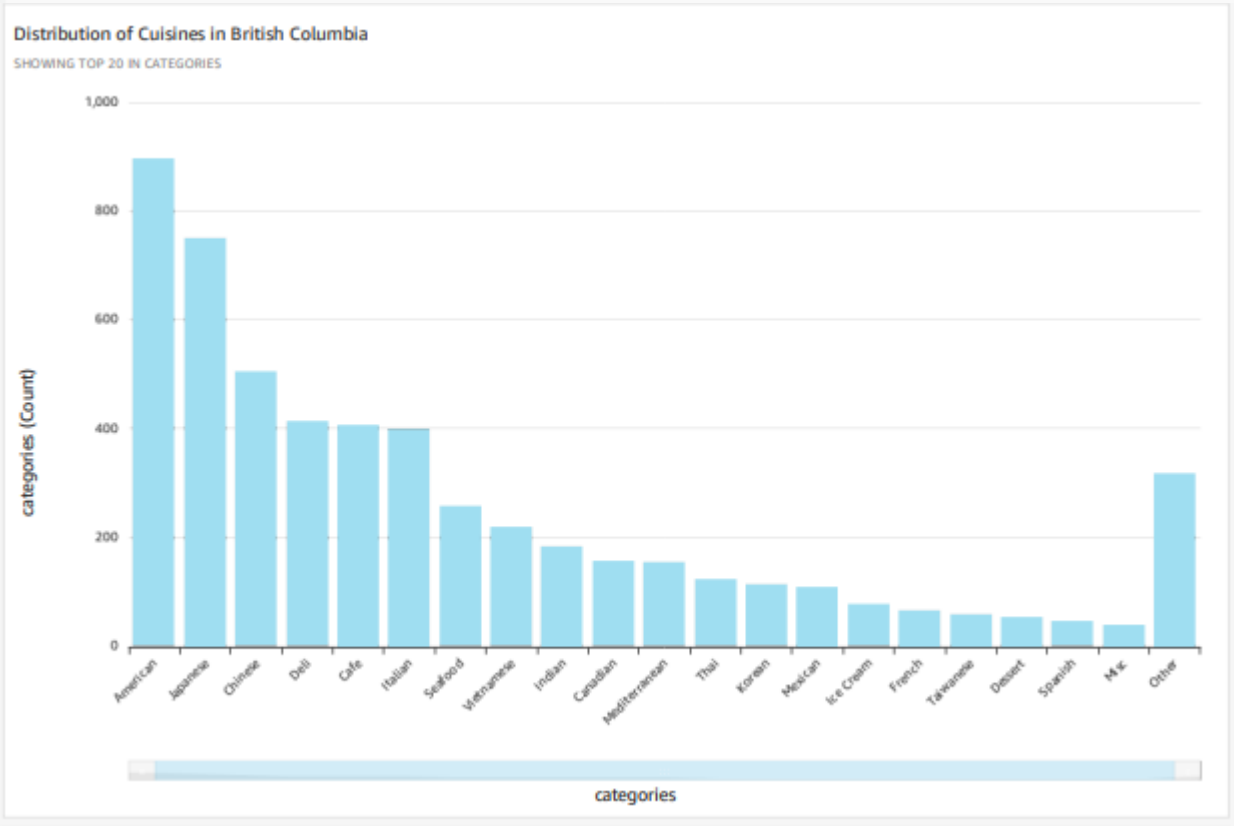
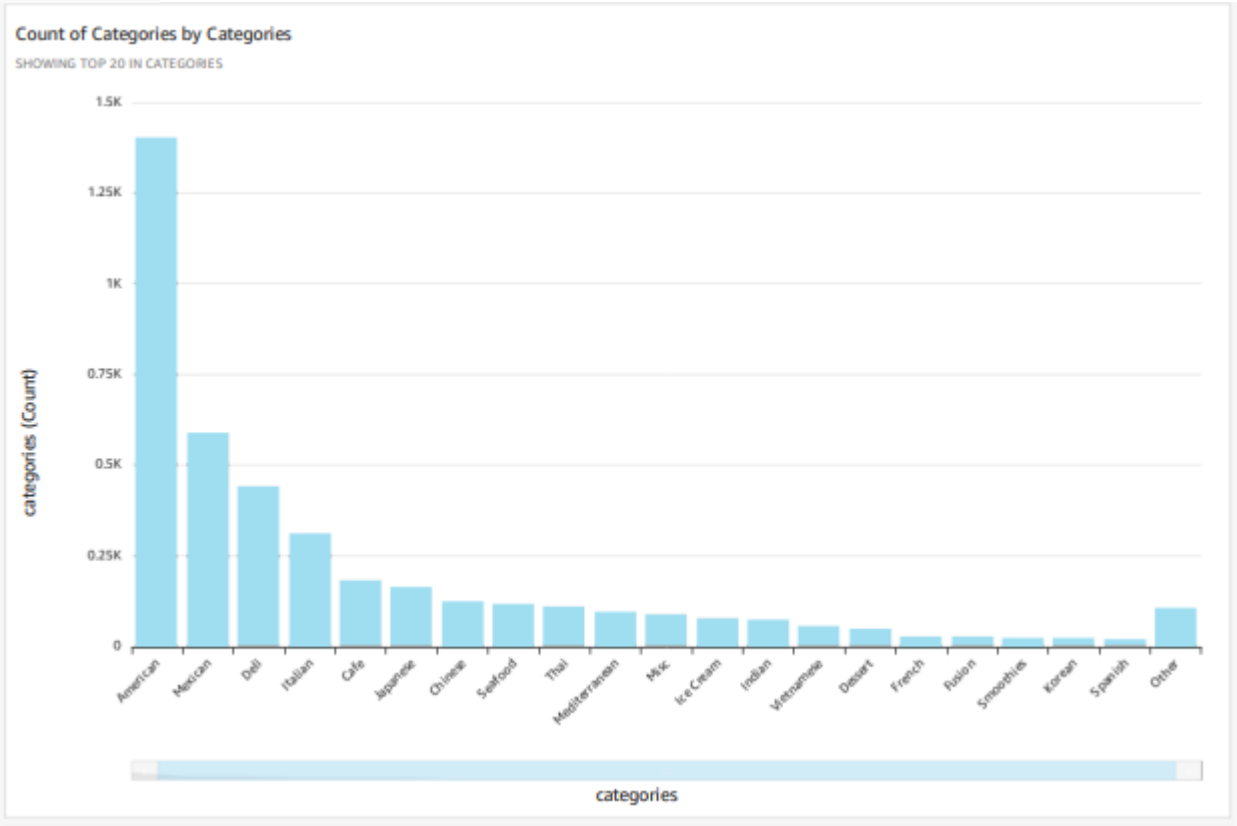


To study the impact of cuisine scarcity, 2 states with the largest distribution of restaurants were chosen for this case study. Plots above depict the distribution of restaurants in British Columbia and Texas, colored according to cuisine. The purpose of these plots is to portray a geographical overview for context

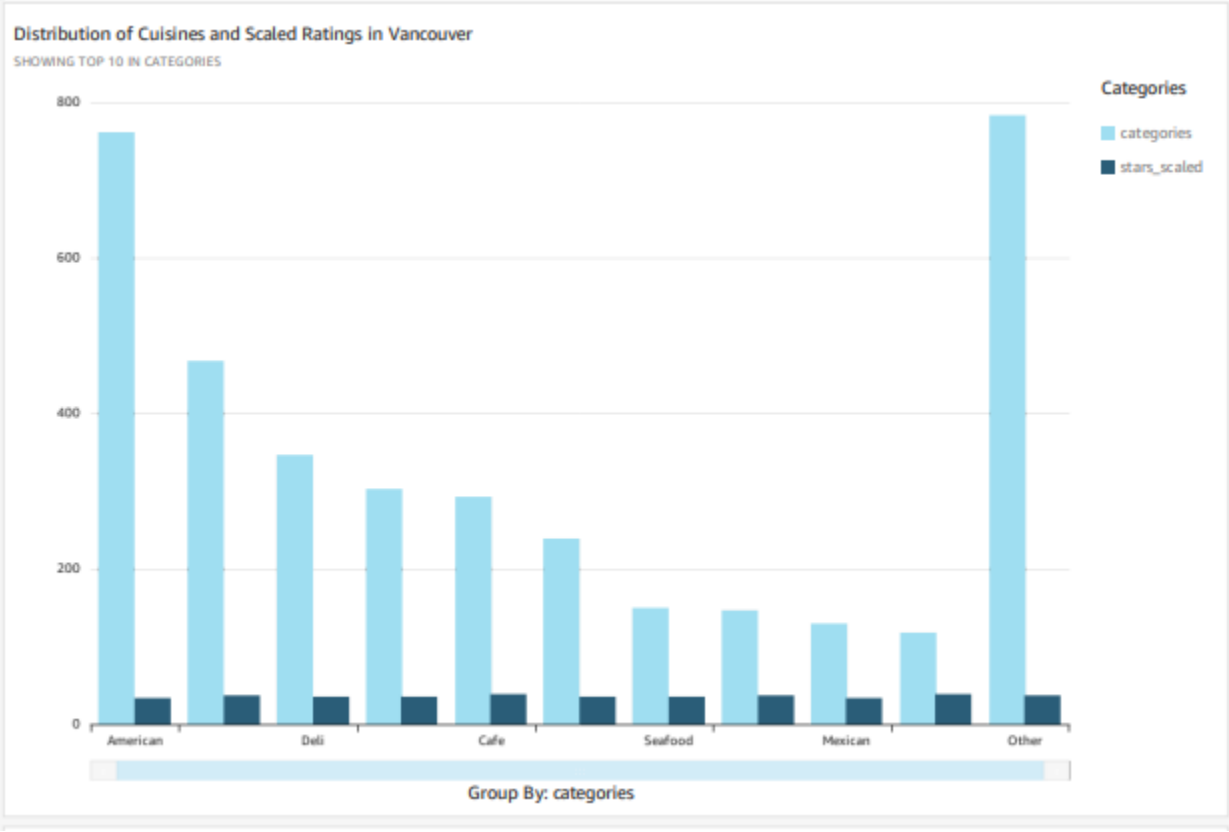
```
SELECT distinct categories, count(categories) as total from business
WHERE state LIKE 'BC'
GROUP BY categories
ORDER BY total DESC
```

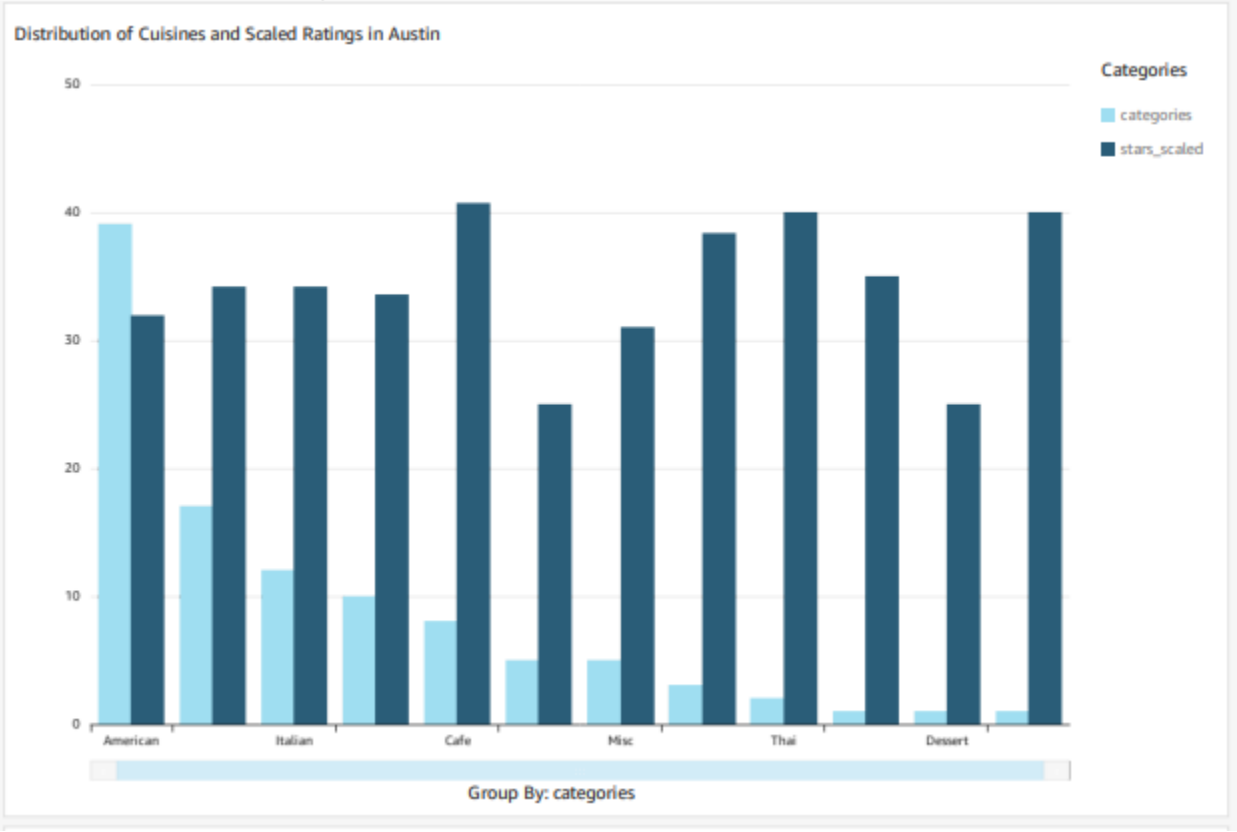




These charts depict the distribution of each cuisine in British Columbia (L) and Texas (R). These two regions states have diverse demographics and ethnic populations which is reflected in the population of restaurants. Top cuisines in BC are Japanese, Chinese, Delis and Cafes whereas in Texas the top cuisines are Mexican, Delis, Italian and Cafes.

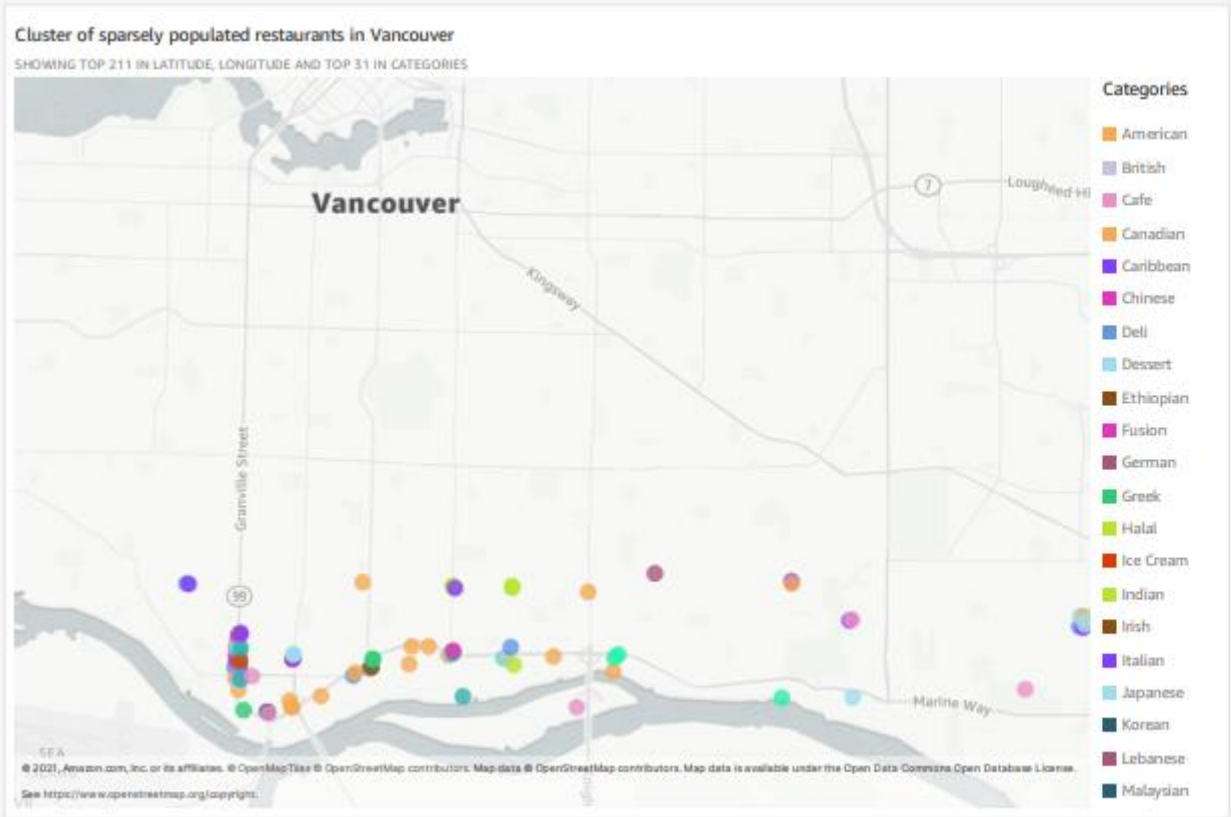
```
SELECT categories as cuisine, count(categories) as sum_cuisine, avg(stars) as avg_rating FROM business
WHERE city LIKE 'Vancouver'
GROUP BY categories
ORDER BY categories
```





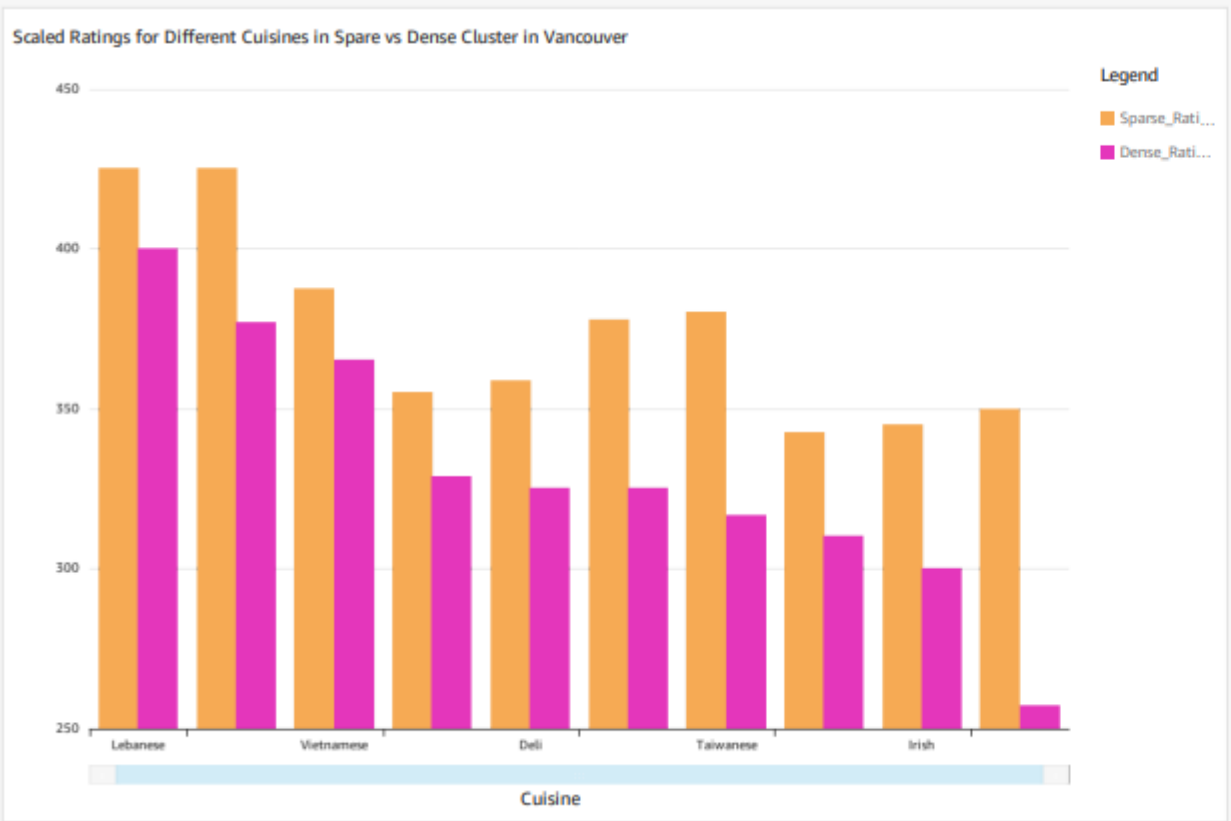
These plots depict the population of each cuisine with their average star rating for restaurants in Vancouver and Austin. For better representation on the combined y axis, the ratings were scaled 10x (3.5 -> 35 etc). Some interesting observations are the rating of Mexican restaurants in Austin [590 total] are on average 1 star lower than in Vancouver [72 total] whereas Japanese restaurants in Vancouver [432 total] are on average 0.4 stars lower than in Austin [162 total]. This is a good indication that scarcity/lower sample size ca inflate the rating of a particular cuisine.

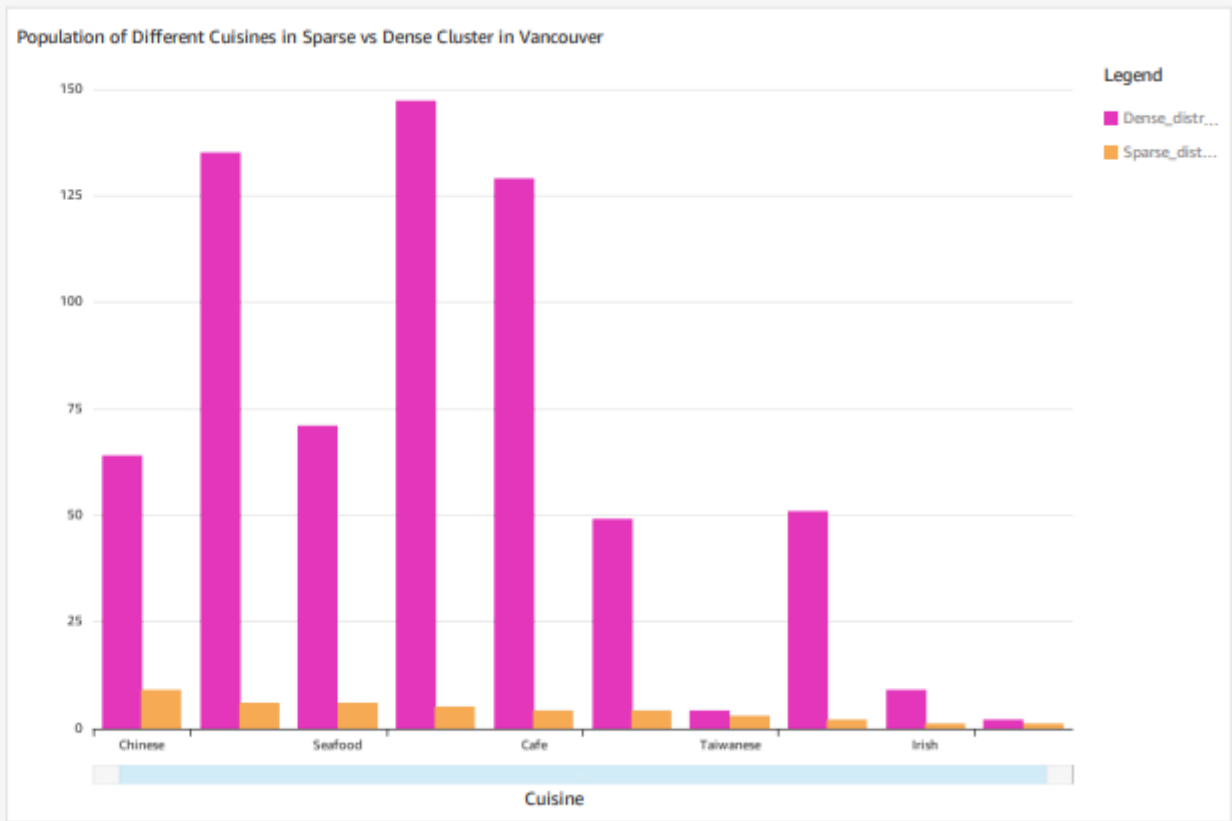
```
SELECT categories as cuisine, count(categories) as sum_cuisine, avg(stars) as avg_rating FROM query
WHERE latitude BETWEEN 40.001 AND 40.04
AND longitude BETWEEN -105.29 AND -105.20
GROUP BY categories
ORDER BY categories
```



Finally, zooming in on Vancouver, 2 different ‘clusters’ were compared to see if the rating of a cuisine was impacted if it had a smaller populated in a 2 km latitude spread.

```
SELECT s.cuisine, s.avg_rating as dense_cluster_rating, s.sum_cuisine as sparse_count, d.avg_rating as
    sparse_cluster_rating, d.sum_cuisine as dense_count from "dense_cluster_vancouver" as d
INNER JOIN sparse_cuisines_rating_count as s ON s.cuisine=d.cuisine;
```





The 2 bar graphs compare the star rating and population of restaurants in the sparely populated and densely populated areas in Vancouver. The number of restaurants of almost every cuisine is much higher (R) which inversely affect the star rating (L). Almost every cuisine in the sparse cluster has a much higher rating – for instance Italian restaurants are rated 1 star higher on average in the spare cluster with only 6 Italian restaurants compared to the dense cluster with 135.