Marsha Curtis
Michela Ziemer
Erin Parker
Tuyet Do

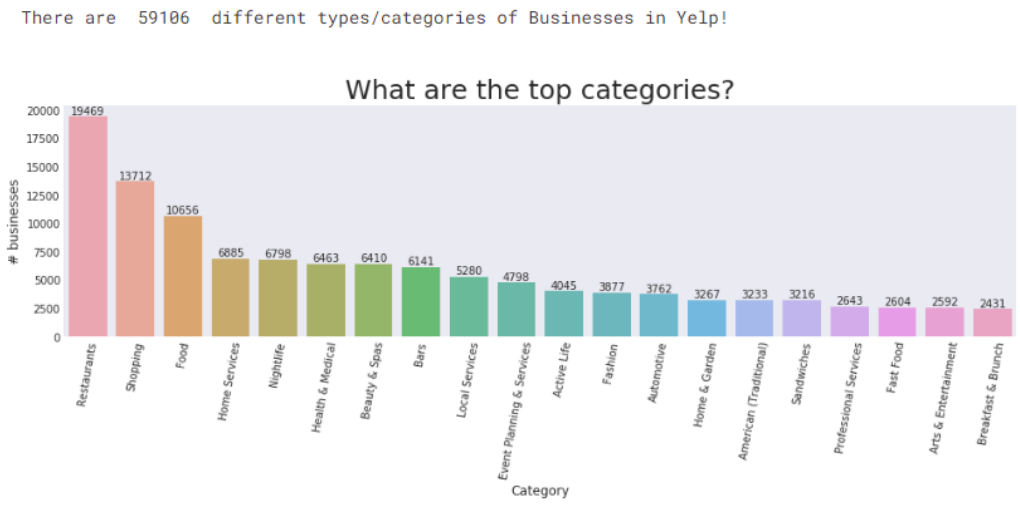
_ August 2022

Snow Bird Taco Tour

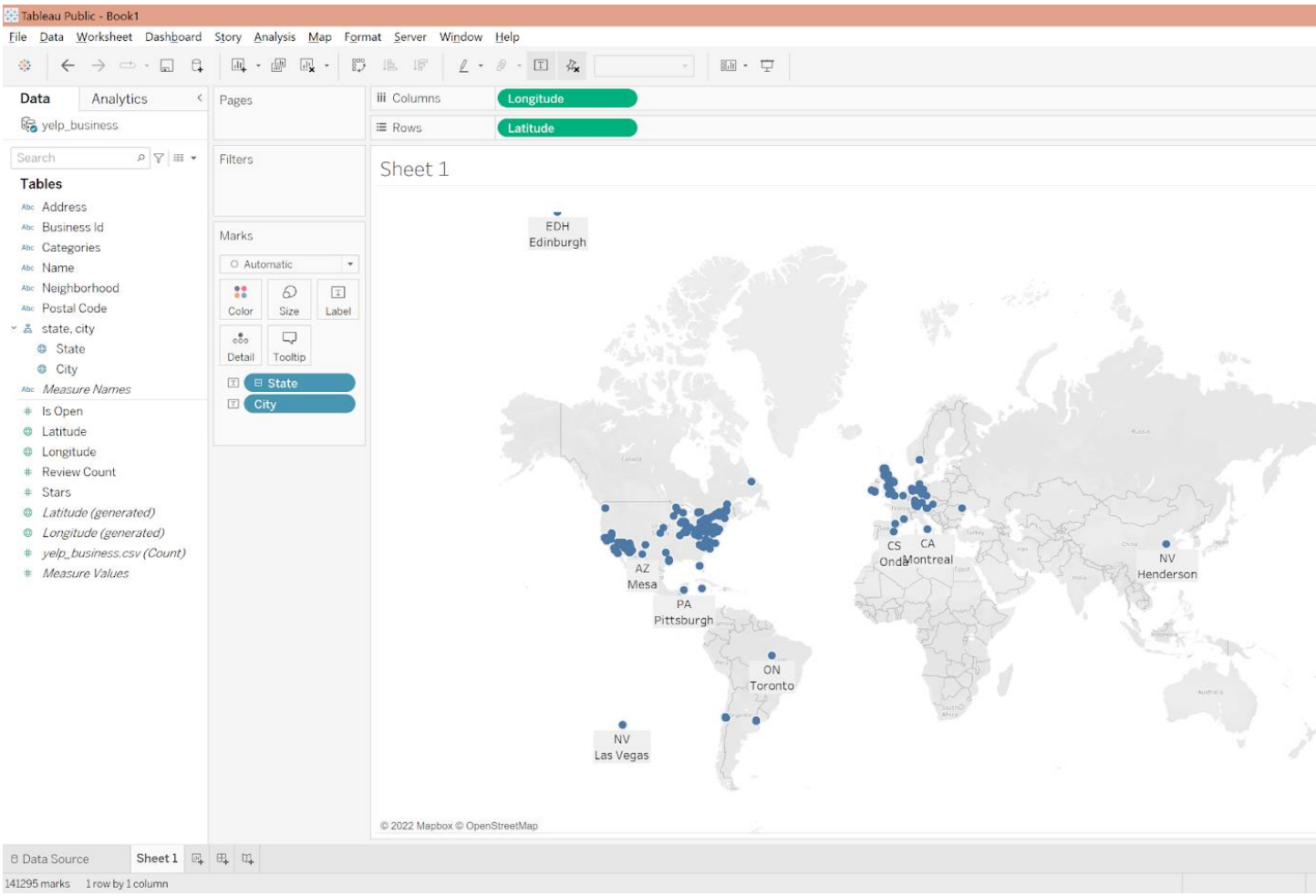
Our initial goal and plan was to do a “Great North American Taco Tour” utilizing the Yelp data found on Kaggle in order to analyze taco and burrito locations and help others find them too in the US and Canada. Our inspiration was rooted in the deep and unabiding love nearly all Americans have for the Mexican (or maybe not-so-Mexican depending on how close you are to Canada) favorite. The plan was to provide a resource for others to input their desired location(s) and attribute(s) and then provide them with recommendations. We wanted to use supervised machine learning to predict restaurant ratings based on various business attributes provided in the Kaggle dataset as well. These attributes included elements spanning a wide range of topics like parking availability, ambiance, music and dancing, menu availability for special diets, and whether our four-legged friends are welcome as well.

We felt like this was a fun topic because of its relatability, and also we could put a really spicy twist on our dashboards. Our initial expectations were that given the popularity of tacos and burritos, there would be a substantial amount of data and perhaps some concentrations along the coastlines and states bordering Mexico. But with all visions, we found that revisions needed to be made after we did a deep dive into our data. Turning what we were hoping as the “Great North American Taco Tour” into the “Snow Bird Taco Tour”.

During the exploratory data analysis (EDA), we decided on the data filters to first be only restaurants (19,469) and then to those that serve tacos and burritos (1,083).



But shortly after mapping the data based on geocodes in Tableau, it was apparent that the geocodes were not the most accurate and needed to be replaced.



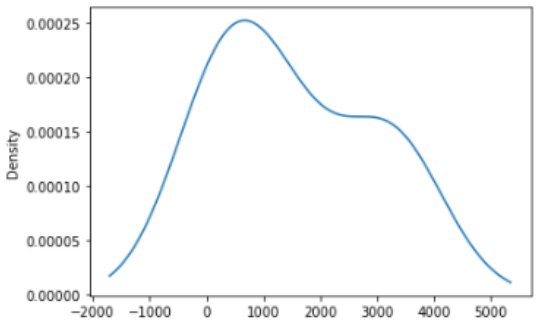
As seen above and to our surprise, the geocodes were not being plotted correctly. For example, there is a plot point for Toronto, Canada in Brazil, South America and even Henderson, Nevada in China to name only a couple. An additional EDA we found was during the machine learning process and it was looking at the visualized density/shape of the ratings. Although it is a fairly even distribution, there could be possible speculations of it being bimodal.

```python
# Visualize the value counts of stars
stars_counts = restaurants_df.stars.value_counts()
print(stars_counts)

stars_counts.plot.density()
```

```
3.5    3581
4.0    3194
3.0    2950
2.5    1485
4.5    1363
2.0     678
5.0     333
1.5     209
1.0      55
Name: stars, dtype: int64
```

```
<AxesSubplot:ylabel='Density'>
```



During our database process, below is the data model