

SnowBird Taco Tour

Our concept, Snowbird Taco Tour, is based on the notion that it's common for people to flee the cold weather in the northern states, like Minnesota, for warmer weather in Phoenix, Arizona. When we head south we have to find a great spot to eat and hang out with friends and family, and why not make it easy: everybody loves tacos!

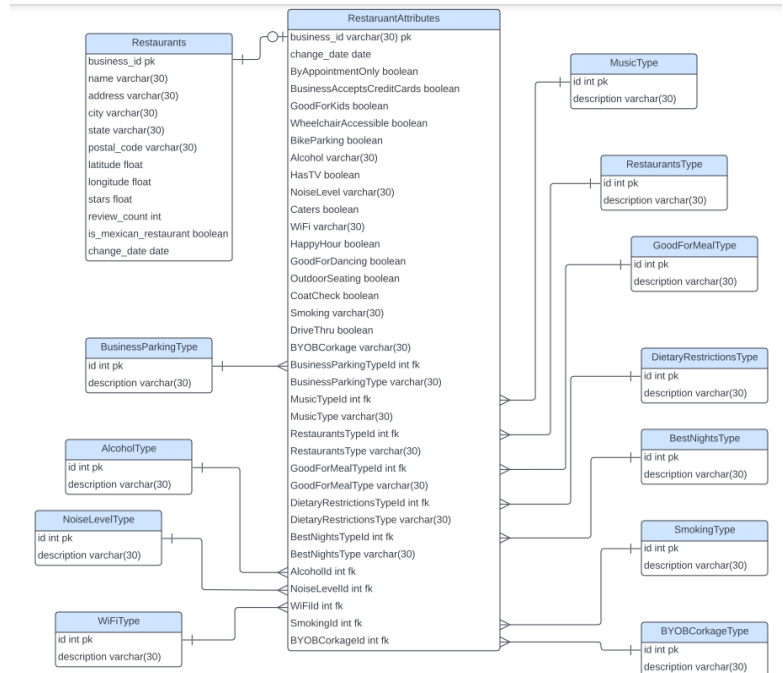
For our Exploratory Data Analysis (EDA), we used the Yelp data from Kaggle and filtered to only restaurants (19,469) and then to those that serve tacos and burritos (1,083). When we created our data model we considered both BI and ML use cases when developing our data model. Our database contains two main tables: Restaurants and RestaurantAttributes. It also includes 11 categorical tables derived using distinct values from categorical columns or from ranges of boolean columns containing similar attributes.

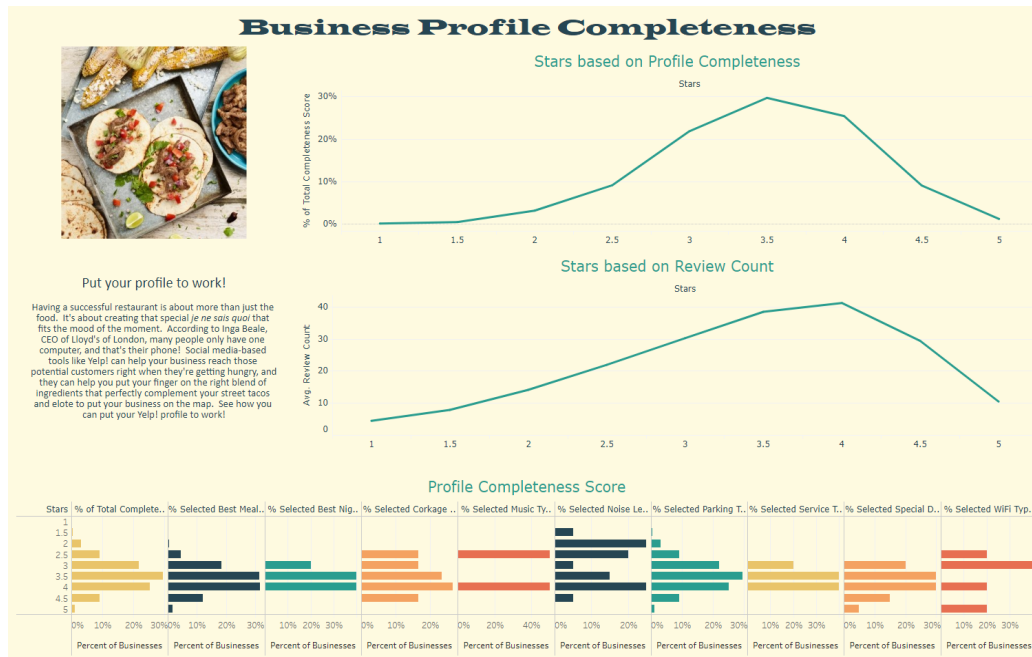
For our ETL process, we developed a number of helper functions for reusable pieces of code. These functions performed tasks like handling nulls, applying geocoding corrections, and dynamically creating lookup tables for various categories of data. Once cleaning and transformation is complete the data is loaded into a SQLite database, which is accessed for both machine learning and right here on the website!

Our goal with our machine learning project was to use restaurant attributes to predict how a restaurant would be rated by customers. After testing multiple models, we saw that all the models were extremely overfit in the training phase, and XGBoost and RandomForest were the worst in that regard. All models yielded accuracy scores around 40-50% in the testing set. Also, in all models we observed that 4-star ratings had the highest F1 scores around 60-65%, 1 and 5-star reviews were the lowest with F1 scores between 0 and 3%.

In the end we dropped all attributes except the 9 most complete attributes and re-trained the model, which yielded very similar results. Finally, we selected the LightGBM model as our final model as it seems least overfit and performed similarly to the others.

Our first dashboard is customer focused, with the goal of making it easy to find a great locale for a tasty south-of-the-border style plate of goodness. Our dashboard is organized to be read in two columns. The left panel groups restaurants by how many stars they've received in the top graph, and in the bottom left panel lets them filter down to a single "hot spot" if they desire. Clicking the dots applies a consistent filter across the board, so to see all 4-star reviews the user clicks a dot on the top panel and all 4-star reviews are revealed on the map as well as in the "hot spots" panel on the bottom left. From there, the user can pick a restaurant based on location on the map or they can further explore options on the Hot Spots pane. In Hot Spots, the dot is sized based on the count of reviews that contributed to that restaurant's scoring.





Our second dashboard is business focused. More customers in the door means more customers to leave reviews. The better the reviews of course the more customers will come in, but that doesn't appear to be the end of the story. Here, we highlight how important a complete profile is to businesses by comparing how complete a business profile is to how many reviews they receive.

The color palette for the webapp and tableau dashboards was inspired by the colors of popular fiestas like Cinco de Mayo and Dia de los Muertos.

The UI for our webpage was loosely inspired by a webflow template: Delice - Restaurant Website Template, however we decided to not use a template for the HTML or CSS. Bootstraps, W3Schools and Stack Overflow were referenced often when we had roadblocks. We also used an AOS plugin to add scrolling animations to make information on our webpages more easy to navigate.



For site navigation, we used a fixed top navigation bar so users can navigate to and from other pages of the site without having to scroll up to the top. Our webapp ended up having seven pages, so to prevent the navigation bar from looking too busy we added a "Project" drop down menu.

The data engineering page includes a plotly bubble chart of the average ratings of restaurants that shows the distribution of the ratings based on the attributes the user filtered by. There is also a table of the restaurant information based on the query.

With some creative thinking we overcame many of the limitations of data. As mentioned during our second dashboard it shows there is a correlation between the business profile completeness and higher stars rating, which implies the more complete their profile is, the more likely they are getting more traffic through the door and could bring more visibility to future customers for these restaurants. So, to all restaurants out there selling tacos and burritos, complete your online profile and make it a hotspot destination for all on a taco tour!