Project Plan

Business / research question:

- How does daily retail foot traffic influence business competitiveness and popularity on Yelp, and does this change differ between business types and across seasons?

Hypotheses:

- Many businesses grew in popularity on yelp - meaning they received more reviews - as a result of increased local foot traffic during the spring and summer months.
- Due to higher leasing rates and increased competition, businesses closed faster on average in areas where local foot traffic was higher.
  - Which business types are affected most by changes in daily foot traffic?
- How does daily foot traffic influence business competitiveness, meaning are businesses opening and closing faster as result of changes in foot traffic? How does this influence differ across business types?
- What influence do seasons have on the amount of reviews a business receives? And which business types perform best in which seasons?
- Do specific seasons significantly affect closing rates (business "competitiveness") for specific types of businesses? For example, does winter have a greater negative impact on closure rates for businesses related to outdoor activities (such as tours, outdoor dining, etc.)?

Anticipated Data:

- Yelp dataset: Business table (including 'business_id', 'city', 'state', 'postal_code', 'stars', and 'review_count'), Category table (including 'business_id' and 'category_name'), and Review table (including 'business_id' and 'review_date')
- Additional dataset: Retail Foot Traffic Data (including 'state', 'foot_traffic_count_normalized', 'number_of_locations_description', 'timestamp')

Obtained Data:

- Yelp dataset: Yelp.sql from class
- Additional dataset: Retail Foot Traffic Data

Note: Season date ranges referenced using (corresponding to each year in the data): https://www.calendardate.com/year2021.php

Data Collection Plan:

- For data that we do not have, we will first look at snowflake marketplace to find extra resources. If we can not find strong sources, we will look for other data marketplaces.

Anticipated Challenges:

- The foot traffic data has one column, where each row is a json object containing all of the variables - rather than a column for every variable. This requires using json_data:*variable_name* when referencing a variable in the sql queries (i.e. “json_data:state” instead of “state”)
- Views are not allowed on Snowflake for the foot traffic shared database, which means the sql queries will be more complex.
- The yelp data is not uniform across states and cities, as some cities and states are more-represented than others in the dataset. Therefore, the conclusions should be based on metrics that keep this principle in mind.