Lift Co. Data Challenge

Overview

The purpose of this data challenge is to get a sense of what it would be like to work with you on a day-to-day basis. The completed challenge gives us a body of work to go over together in a project review session. We will evaluate you on your implementation, design choices, choice of technologies, etc. Please communicate a present day as soon as you can. The deadline is 5 business days(1 week) after you get this assignment.

Technology Stack

You have the option to use a technology stack that you're most comfortable with. You could implement this challenge using any mainstream programming languages, libraries, frameworks and storage systems.

Requirements

Section 1: Mandatory for both permanent and internship applicants

1. Ingest and perform analytics over Yelp open dataset (Json format only NOT SQL format) from https://www.yelp.com/dataset to a data store (NoSQL or RDBMS)
2. Top 10 restaurants in Toronto with highest popularity. You are free to define your 'popularity', as long as it makes sense.
3. How many Canadian residents reviewed the business "Mon Ami Gabi" in last 1 year?
4. Top 10 most common words in the reviews of the business "Chipotle Mexican Grill"
5. What's the percentage of users, who reviewed "Mon Ami Gabi", and also reviewed at least 10 other restaurants located in Ontario?
6. Please think about 2 more analytics, which provide insights and help existing/future Business Owners, to make important decisions regarding new business or business expansion.

Section 2:
Mandatory for permanent applicants to finish at least one of below two questions.
Optional for internship applicants, if you are interested in working on Machine Learning projects.

1. For Yelpers living in Toronto, which restaurants do you recommend them, when they travel to California, US? We want a recommendation model, which can recommend restaurant for travelers, based on their previous actions in Yelp platform.

2. Given a review text, could you predict how many stars the reviewer will rate a restaurant? We want a prediction model, which can predict the rating(stars) based the review text and the reviewers' behaviors.

Deliverables

Share your assignments with us via GitHub, in the meanwhile you can always polish the assignments up until you present it to us on your interview day. It would be preferable if you could bring in your own laptop to review the assignments with us.

If you have any questions regarding requirements do not hesitate to get clarifications from us (ajin@lift.co).