Capstone Project 1: Project Proposal

Experience Coffee Searcher

1. **What is the problem you want to solve?**

Improve the approach of recommendations when you are looking for a specific experience, in this particular case, coffee shops. This methodology could be extrapolated to other spectrums of cuisine experiences.

2. **Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?**

Every day, customers of coffee shops are asking for more quality, more variety and care in details. Comfortable, calm, noisy, dynamic, minimalist, colorful or over-decorated spaces contribute to building diverse experiences. Some customers want to be surprised, and others wish to find exactly they expect. It's possible to identify different types of coffee drinkers and with all of the information out there, we can find the best recommendation for each one.

The following analysis allows giving suggestions matching personal likes with shops that can satisfy them. In this way, customers can improve their experiences and avoid disappointing about their choices. Every coffee shop has interesting features agree with a determined public. It's not about which is better than others, but which one fits better with individual hopes.

Additionally, we can find a business opportunity in experiences with less coverage to interested clients.

3. **What data are you using? How will you acquire the data?**

- It will do an API request to Yelp for getting coffee shops in San Francisco (names, address, rating score, number of reviews, reviews and attending hours).
- It will use web scraping in Python to get names and description of coffee shops in a couple of blogs of coffee shops in San Francisco, written by experts (local baristas and recognized writers).

To build features vectors we need general information about coffee shops and special attributes obtained from reviews in Yelp and blogs.

We are looking for some keywords to build the feature vector. Some examples of words/phrases are listed below:

a. Coffee quality: **amazing, delicious, good, nice, wrong, disgusting, disappointed, own roaster,** the **variety** of **beans**
b. Coffee drinks headliners: hot drinks (**americano, latte, cappuccino, mochaccino, ristretto, macchiato, espresso**) and cold drinks (**iced coffee, frappe, cold brew coffee, affogato**)
c. Food headliners (**pastries, breakfast, lunch**)
d. Place: quiet, noisy, minimalist, music, type of decoration (art, books, office, bar)
e. About the experience ("I went for **studying**, **reading**, **buy** and **leaving**, **staying**, **share with friends** and It was **…**")

4. **Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.**

The diagram below (Figure 1) shows the flow of data and all the processes included in the project.

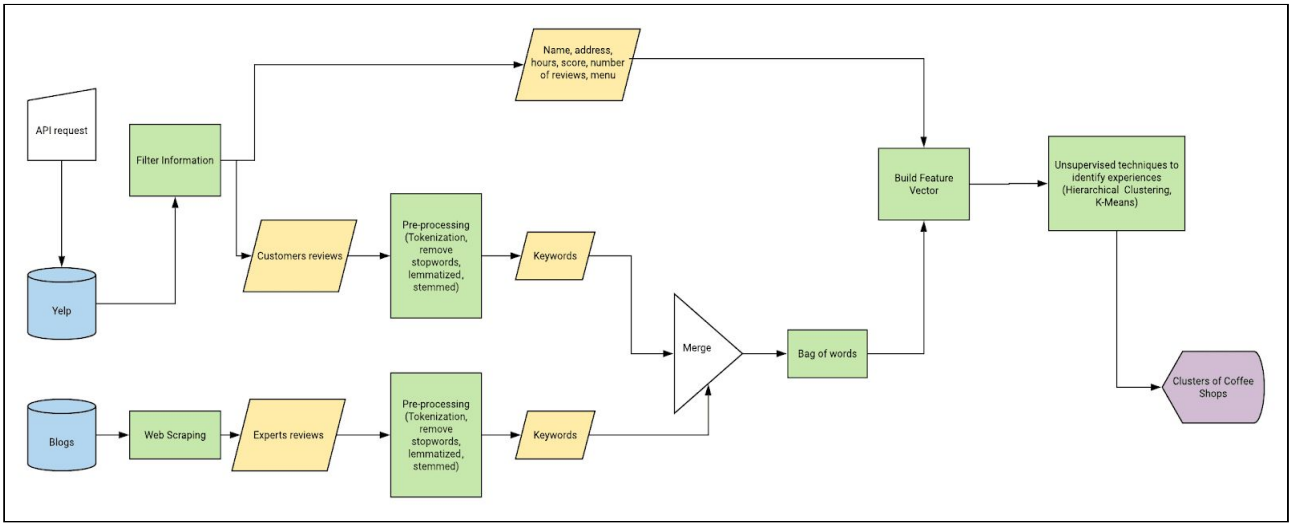

Figure 1: Flow of information and processes involved to find coffee experiences clusters

5. What are your deliverables? Typically, this includes code, a paper, or a slide deck.

It will be delivered in slides deck and in a repository on GitHub.