Yelp Dataset Analysis

Preliminary

STAT628, Module 3, Group 13

Kanishk Saxena, Shan Leng, Sreeja Kodati

Introduction

- **Motivation**: Vegetarians always have a difficult time finding an appropriate restaurant or looking for recommended dishes when eating outside.

- **Question**:
  - (Business owners) In (busy areas of) Philadelphia, what are the popular vegetarian/vegan options among consumers, and what key attributes contribute to the popularity?
  - (Consumers) Same as above, what are the popular vegetarian/vegan options available?

- **Goal**:
  - (Business owners) Provide valuable information to new business owners to strategically position themselves and differentiate their vegetarian/vegan options in the competitive market of Philadelphia.
  - (Consumers) Recommend popular vegetarian restaurants in a certain area together with popular menu items.

Preliminary Analysis

Data Processing: Vegan Restaurants

- In 'business.json', select businesses containing both 'Restaurants' and 'Vegan/Vegetarian' in the 'categories' column. ⇒ 1,629 businesses.
- Group these businesses by 'state' and 'city' respectively, to decide an ideal subset of restaurants for further analysis(e.g., sample size & geographical scope).

State	PA	FL	LA	TN	MO
#Vegan Restaurants	480	267	132	129	106
City	**Philadelphia**	Tampa	New Orleans	Nashville	Tucson
#Vegan Restaurants	**277**	124	117	89	75

Preliminary Analysis

Data Processing: Vegan Restaurant Reviews

- Exclude vegan restaurants with no more than 10 reviews ('`review_count`')
  ⇒ 245 businesses with 47,991 reviews.
- Use corresponding '`business_id`' to extract associated reviews from 'review.json'
- For these reviews:
  - Segment the sentences into words and remove stopwords and punctuations.
  - Categorize words into nouns, verbs, adjectives(adj.), adverbs(adv.).
  - Stem verbs and nouns.
    The exact formats of adj. and adv. could show emotional tendencies.
  - Remove words with low frequency(≤ 10) across the reviews for each category.
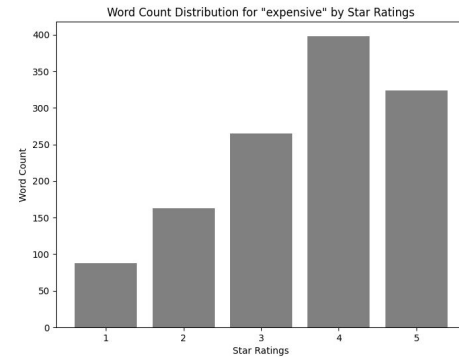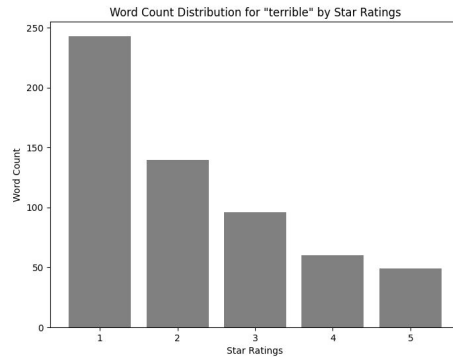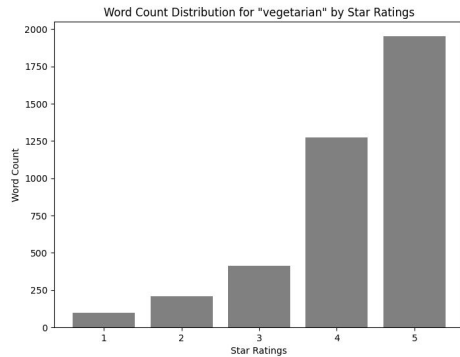
Preliminary Analysis

Processed data

	business_id	text	stars	adj	adv	noun	verb
0	Dv6RfXLYe1atjgz3Xf4GGw	Good food, reasonably priced, and nice staff. ...	4.0	[good, nice]	[reasonably, not]	[approach, way, dish, bella, ideal, fry, food,...	[price, get, know, go, eat, make]
1	AY1SyySRyTnBwBdf7qfGxQ	From the second you walk in the door, you're g...	5.0	[great, recent, adorable, incredible, vegetari...	[well, really, always, actually, particularly,...	[sausage, flavorl, right, bar, candy, variety,...	[price, come, greet, seem, know, amaze, enjoy,...
2	Dv6RfXLYe1atjgz3Xf4GGw	Super lunch option\nWatch out for the yoga mat...	4.0	[great, hard, difficult, last, green, bad, sma...	[nearly, long, yet, really, especially, also, ...	[hipcity, tend, suffice, stand, Super, time, f...	[pair, come, see, take, get, open, rectify, am...
3	Dv6RfXLYe1atjgz3Xf4GGw	Wow. I am not a vegetarian, nor am I a health-...	5.0	[vegetarian, good, delicious]	[well]	[random, place, time, co-worker, company, peop...	[talk, put, create, get, greet, go, wait, try]
4	bTve2mwLk5Zc01vRKqc2KQ	This might be my favorite coffee shop in the c...	5.0	[great, enjoyable, unbelievable, fresh, nice, ...	[definitely, well, especially]	[seat, edgy, want, food, place, favorite, comf...	[seat, check, order]

- Bar plots using Matplotlib are generated to visualize the word count distribution of the different adjectives across different star ratings.
- These plots will help us observe if there's a correlation between the use of certain adjectives and higher or lower star ratings.

Preliminary Analysis

- These plots below can be used to see how the frequency of positive or negative adjectives correlates with star ratings.
- For example, adjectives that are frequently associated with positive reviews, may indicate strengths of the businesses.

Word Count Distribution for "vegetarian" by Star Ratings

Word Count Distribution for "terrible" by Star Ratings

Word Count Distribution for "expensive" by Star Ratings

Further Analysis

Vegan Restaurants & Reviews

- Summarize common features w.r.t price, hours/location, ambience, wifi, etc.
- To get the popular menu options that vegetarians prefer, we need to look at the reviews containing 'vegan/vegetarian'
- Recognize menu items using Named Entity Recognition (NER) -
  - Recognizing menu items is a crucial step in understanding the preferences and trends among consumers in the vegetarian/vegan restaurant domain.
  - Utilize NER models or techniques specialized in recognizing food-related entities within the text. These models are trained to identify and classify entities like dish names, ingredients, or food categories.
  - Apply the NER model to the reviews focusing on 'vegan' or 'vegetarian' to identify and extract mentions of specific dishes, food items, or menu categories.
  - Organize the extracted menu items into categories or types of dishes (e.g., appetizers, main courses, desserts) for a more structured analysis.
  - We can use Chi-square test to identify items that could bring high ratings.

Further Analysis

Neighborhood Wealth & Mobility

- Restaurants located in "Busy Areas"

- Linear model: star ~ median_income + #trips (using dataset 2 and 3)

- Consumer Behavior Patterns:
  - Analyze how consumer behavior and spending patterns differ across neighborhoods with varying income levels. This can offer insights into preferences and affordability concerning vegan dining options.

Further Analysis

Neighborhood Wealth & Mobility

- Price Sensitivity Analysis:
    - Investigate whether there's a correlation between median income in an area and the pricing strategies of successful vegan restaurants. Determine whether restaurants in higher-income neighborhoods tend to charge higher prices for their offerings.
- Behavioral Analysis based on Trips Data:
    - Investigate consumer behavior patterns derived from trip data. Determine if there's a correlation between the frequency of trips made to vegan restaurants and the overall ratings or popularity of those establishments.

Expected Results

- **Suggestions:**

  For successful vegan restaurants, based on analysis conducted.

  - Menu Diversity and Specialization:
    - Encourage the inclusion of diverse and innovative vegan dishes to cater to a wide range of tastes and dietary preferences.
  - Allergen-Friendly Options:
    - Ensure the availability of allergen-friendly options to accommodate customers with various dietary restrictions or allergies.

Expected Result

- **Shiny-app:**

  Given a region, show good vegan restaurants and recommended dishes.

  - Search and Filter Options:
    - Allow users to input their region or specific location within Philadelphia.
    - Provide filters for dietary preferences, price range, ambience, and more.
  - Restaurant Recommendations:
    - Display a list of recommended vegan restaurants in the chosen region.
    - Include restaurant names, ratings, proximity, and brief descriptions.
  - Menu Item Recommendations:
    - Showcase popular or recommended dishes from the selected restaurants.
    - Include images, descriptions, and any special features.

Thank You for Listening.