GreenEats: Navigating Top Vegan & Veggie Cuisine

STAT628, Module 3, Group 13

Kanishk Saxena, Shan Leng, Sreeja Kodati

Introduction

- Taking Philadelphia as a sample city, we are curious to explore those popular and successful restaurants with vegan or vegetarian options.

- **Target businesses**: 245 restaurants(#reviews≥10) with vegetarian options in Philadelphia , and 47,991 reviews associated these restaurants.

- **Goals**
  - Extract popular menu items from reviews of these restaurants; provide suggestion on menu design for new business owners and make recommendations to vegetarian customers.
  - Summarize key attributes and common features contributing to the popularity and success of these restaurants; provide valuable information for new business owners.

Discovery of Popular Vegetarian Dishes

- **Extraction**
  Extract all mentions of potential menu items from businesses' reviews, which is very similar to the Named Entity Recognition (NER) process. We use "`flair`", a powerful natural language processing library in python, to achieve this .

Original Text of Review

"Good food, reasonably priced, and nice staff. Not ideal for eating in, but since they are going for a fast food approach to vegan food, that makes a lot of sense. I know the **buffalo bella** is the dish that they push the hardest, but for my money the **Cheesesteak** is the way to go. And pro-tip: Get the **sweet potato fries**."
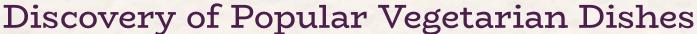
Tagged Entities by **Ontonotes Model**

["**Cheesesteak**"/PRODUCT]

["Good food"/NP, "reasonably priced"/ADJP, "nice staff"/NP, "Not ideal"/ADJP, "for"/PP, "eating"/VP, "in"/PRT, "since"/SBAR, "they"/NP, "are going"/VP, "for"/PP, "a fast food approach"/NP, "to"/PP, "vegan food"/NP, "that"/NP, "makes"/VP, "a lot"/NP, "of"/PP, "sense"/NP, "I"/NP, "know"/VP, **"the buffalo bella"/NP**, "is"/VP, "the dish"/NP, "that"/SBAR, "they"/NP, "push"/VP, "the hardest"/NP, "for"/PP, "my money"/NP, **"the Cheesesteak"/NP**, "is"/VP, "the way"/NP, "to go"/VP, "pro-tip"/NP, "Get"/VP, **"the sweet potato fries"/NP**]
Tagged Phrases by **Syntactic Chunking Model**

Discovery of Popular Vegetarian Dishes

- **Extraction**

Tagged Phrases by Syntactic Chunking Model

["Good food"/NP, "reasonably priced"/ADJP, "nice staff"/NP, "Not ideal"/ADJP, "for"/PP, "eating"/VP, "in"/PRT, "since"/SBAR, "they"/NP, "are going"/VP, "for"/PP, "a fast food approach"/NP, "to"/PP, "vegan food"/NP, "that"/NP, "makes"/VP, "a lot"/NP, "of"/PP, "sense"/NP, "I"/NP, "know"/VP, **"the buffalo bella"/NP**, "is"/VP, "the dish"/NP, "that"/SBAR, "they"/NP, "push"/VP, "the hardest"/NP, "for"/PP, "my money"/NP, **"the Cheesesteak"/NP**, "is"/VP, "the way"/NP, "to go"/VP, "pro-tip"/NP, "Get"/VP, **"the sweet potato fries"/NP**]

Vegetarian Food List

acorn, ale, almond, anchovy, anise, appetizer, apple, apricot, artichoke, asparagus, aspic, ate avocado, bagel, bamboo, banana … …
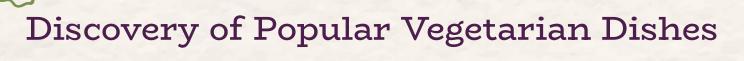
Food Vocabulary Word List
Food | Lists by Theme

EnchantedLearning.com

Some overlapping

330 high-frequency common words

None phrases for vegetarian dishes

Discovery of Popular Vegetarian Dishes

- **Extraction**

Post-processing

- Remove three types of words:
    - Stopwords (e.g., "the", "a", etc.);
    - General words (e.g., "dish", "brunch", etc.);
    - Non-vegan items (e.g., "fried chicken", "Cheesesteak", etc.).
- Standardize the expressions
  (e.g., "mac & cheese" and "Mac n cheese" are unified as "Mac&Cheese").
- Only keep products and noun phrases that have appeared more than 10 and 20 times, respectively, in all reviews.

⇒ 82 products and 395 noun phrases for vegetarian dishes

Discovery of Popular Vegetarian Dishes

- **Summarization**

  Popular vegetarian dishes: extracted menu items that
        1) are repeatedly mentioned a lot by customers;
        2) have an average star score of 4 or higher from reviews associated with them.

  We use "`stars`" of a review as the star rating for <u>all</u> the vegetarian dishes extracted from the corresponding text content.

  - For each restaurant, calculate the average star ratings (ave_star) for every vegetarian dish served by the restaurant and retain those with an ave_star ≥ 4.

  - "**Star Dishes**": 1) Extract vegetarian dishes served by more than 10 restaurants;
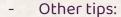   2) further filtering these common dishes to have an ave_star ≥ 4.5.

  ⇒ 48 "star dishes" that are popular and common for both business owners and vegetarian customers.

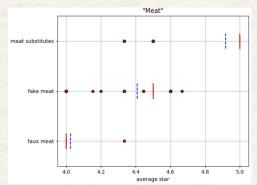# Discovery of Popular Vegetarian Dishes

- **Key Findings About "Star Vegetarian Dishes"**

- **Salad**: Compared with the very broad "salad" dish (ave_star = 4.40, same below), more specific ones such as spinach(5.00) and potato(4.95) salads are more popular, while cucumber(4.27) and taco(4.00) salads are less favored.

- **"Meat"**: Meat substitutes(4.92) seem to be easier to be accepted by vegetarians and can be considered by new business owners. Fake(4.41) and faux(4.03) meat are less popular.

- **Drink**: Low-lactose drinks like almond milk(4.95) and coffee(4.87) are more popular than milk(4.00) and milk derivatives (e.g., milkshake–4.01, frozen yogurt–4.00).

- Other tips:
    - Safe items: burrito(5.00), spring roll(4.97), bagel(4.96), vegpizza(4.94).
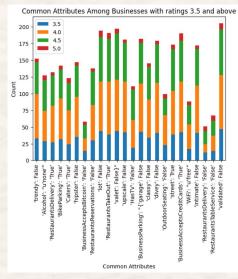   - Be careful with seasoner: curry(4.09), spices(4.07), sauce(4.06), garlic(4.00), ginger(4.00).

Summarizing Business Attributes

- **Main goal:** Generate common attributes among the successful businesses in the city of Philadelphia.

- Only successful restaurants in the city of Philadelphia were considered for the analysis.

- A successful restaurant defined in this case a restaurant with:

    - Yelp star rating of 3.5 and above.

    - Has more than 10 reviews at least.

    ⇒ Total number of restaurants initially: **245**

    ⇒ Number of restaurants after filtering: **195**

Summarizing Business Attributes

- The bar plot generated illustrates common attributes with count among businesses with star ratings 3.5 and above.

- Each attribute can be interpreted to make generalizations about the attributes contributing to a restaurant's success.

- The analysis provides key insights into attributes contributing to a successful restaurant in the city emphasizing the nature of success in the restaurant industry.

The other 2 datasets

- Linear model to predict the stars based on the trips and median income.

- Create 3 data frames from 3 datasets and merge all based on the common states.

- X contains the predictors, which are 'Trips' and 'MeanIncome' and y contains the target variable, which is 'Ratings'.

- Fit an Ordinary Least Squares (OLS) regression model - *sm.OLS(y, X).fit()*

- The model finds the best linear relationship that minimizes the difference between the actual 'Ratings' (y) and the predicted values (X).

- model.predict(X) function is used to generate predicted values for the target variable 'Ratings'.

- These predicted values are generated by applying the trained model to the same data (X) it was trained on, resulting in predicted ratings for each state based on their respective number of trips and mean income.
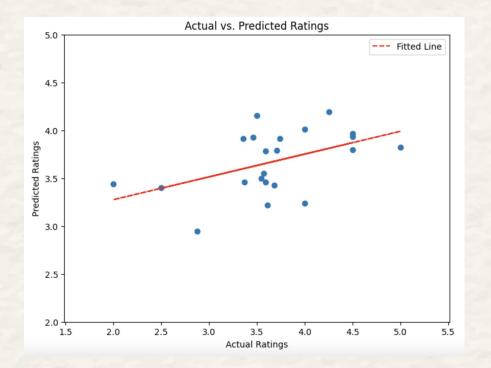
The other 2 datasets

- Findings
  - Linear Model : *stars ~ median_income + trips*

Shiny App

- This app, "GreenEats" helps you discover vegetarian and vegan restaurants in the city of Philadelphia.
- Using a map, one can filter restaurants by postal code, click on markers to view more details and explore the popular dishes.

Thank You!