

A top-down view of a white plate of food on a grey and white checkered napkin, which is placed on a light-colored wooden table. The food consists of a sunny-side-up fried egg, a portion of yellow potatoes, sliced red tomatoes, and sliced white radishes. A silver fork is on the left and a silver knife is on the right of the plate.

Restaurant WordCloud

Based on Potential Reviewers and Star Rating Prediction

Group 18



Introduction

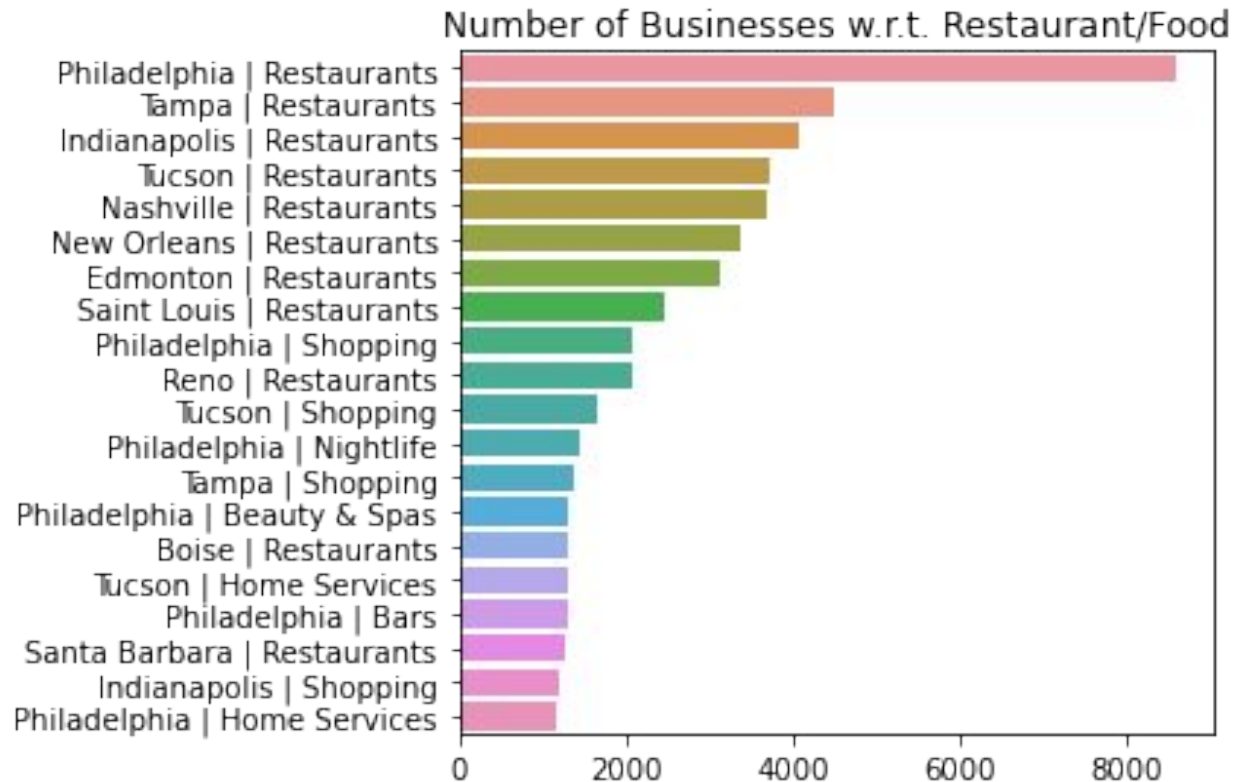
- **Model:**

- a) Define potential customers of a restaurant as people who do not rate that restaurant now.
- b) Generate personalized wordclouds for restaurant owners in Philadelphia by predicting potential customers' star rating to the restaurants and analyzing their previous reviews.

- **Process:**

- a) Narrow down our research to restaurants in Philadelphia in EDA & Preprocessing part.
- b) Use Matrix Factorization model and Neural Network model to build customers' star ratings prediction system.
- c) Extract previous reviews of potential customers to the similar restaurants to generate wordclouds for each restaurant.
- d) Build up Shiny App to display the personalized wordclouds.

EDA & Preprocessing

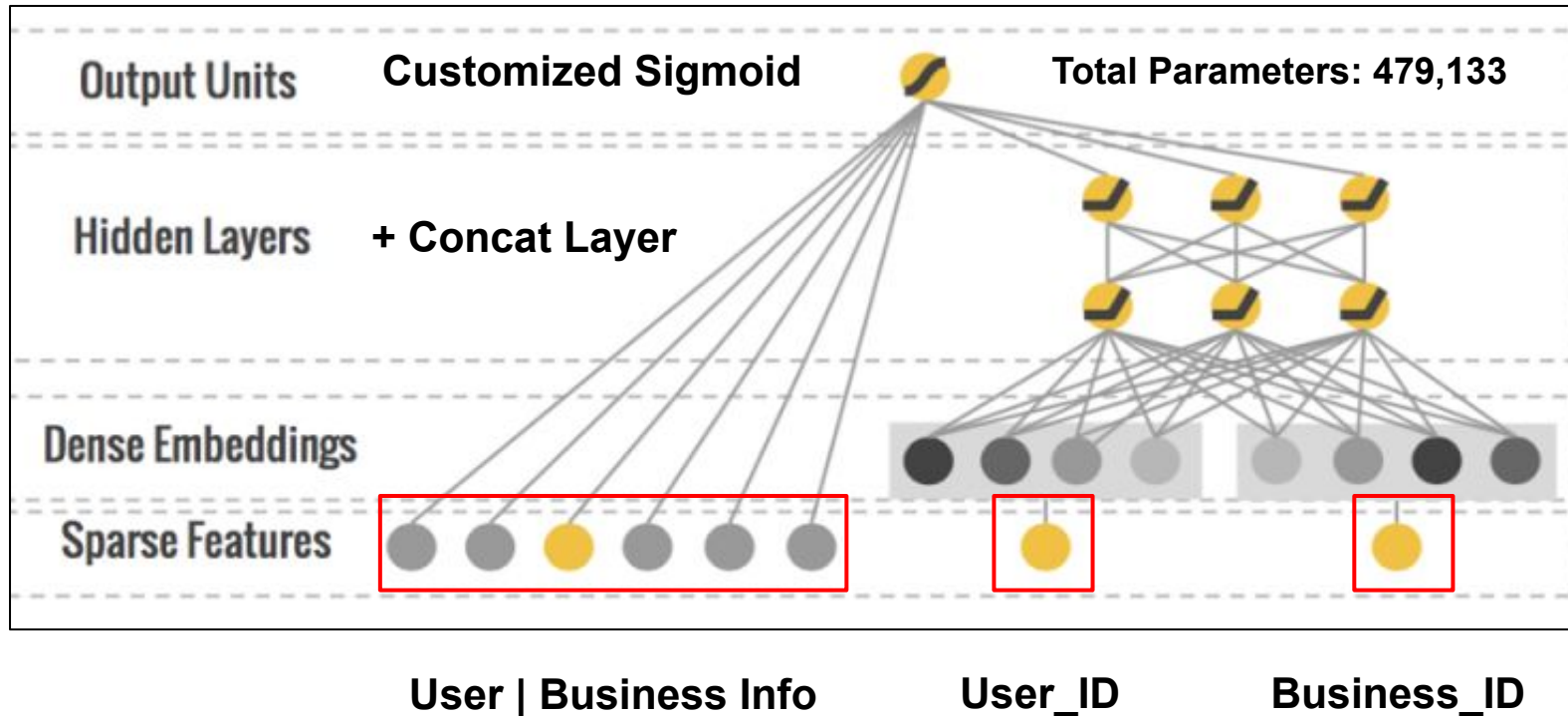


- Philadelphia has the largest number of restaurant and food entities, which will lead to a more robust model.
- Users and restaurants with more than 20 reviews are chosen to ensure an accurate model.

Recommendation System (Star Rating Prediction)

- **Neural Network Architecture**
- **Regression Problem (1 to 5 stars)**

Google's Wide & Deep Learning Architecture (Heng-Tze Cheng, 2016)



Recommendation System (Star Rating Prediction)

- **Train/Val/Test Set Splitting**

Test: Most Recent Review for each user

Val: Second Recent

Train: Remaining

Keras functional API

- **Evaluation Metrics - RMSE**

Regression Problem

Small Difference Preferred More

RMSE 1.125

- **Neural Net vs Matrix Factorization**

RMSE 1.125 vs 1.134

Neural Net model Selected

How Do We Find Similar Restaurants?



	B1	B2	B3	B4	B5
					
U1	3	2	?	2	?
U2	?	?	5	1	3
U3	?	3	1	?	2
U4	?	4	?	5	?
U5	?	?	1	5	?

Potential positive reviewer's **negative** review history

Word Embedding






1. **Select 'categories' Attribute**
 - comma separated attribute
 - 150,346 entities
2. **Build 'word2vec' model with 'gensim' library**
 - category embedding vectors
3. **Extract top 5 categories for each category**
 - cosine similarity score

Category	Similarity
Brazilian	1.000000
Argentine	0.662890
Peruvian	0.532551
Basque	0.531147
Spanish	0.501794

TOP 5 Most Similar To 'Brazilian'

4. Analyze Purple cells' Textual Reviews!

Similar to B1

	B1	B2	B3	B4	B5
					
U1	3	2	3	2	3
U2	5	2	5	1	3
U3	5	3	1	5	2
U4	1	4	2	5	2
U5	5	1	1	5	3

Potential positive reviewer's negative review history

WordCloud & Shiny App

4. Analyze Purple cells' Textual Reviews!

- Remove accents, white spaces, stopwords
- Build N-Gram Model
- etc.

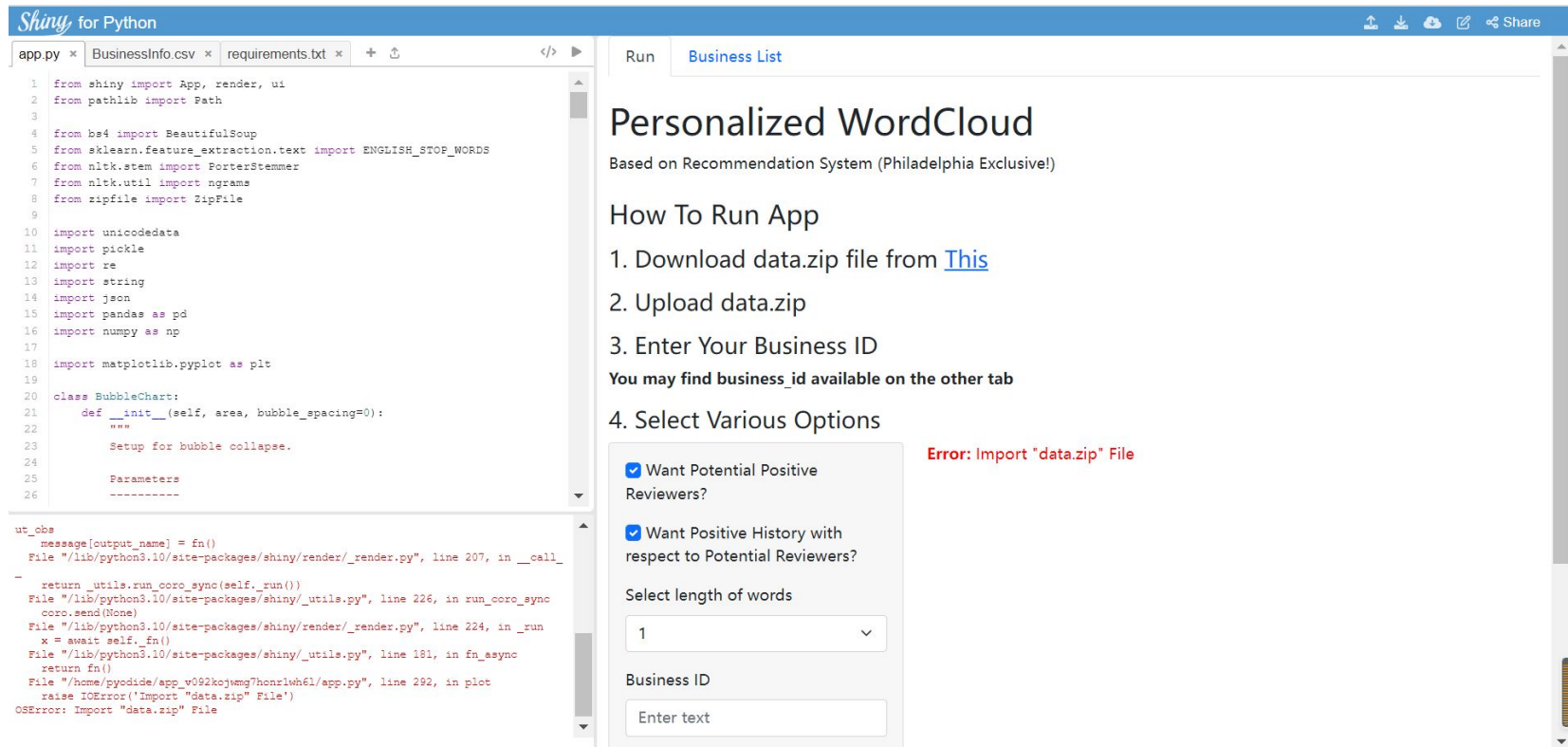
Similar to B1

	B1	B2	B3	B4	B5
U1	3	2	3	2	3
U2	5	2	5	1	3
U3	5	3	1	5	2
U4	1	4	2	5	2
U5	5	1	1	5	3

Potential positive reviewer's negative review history

WordCloud & Shiny App

Here is the entrance interface to our shiny app:



The screenshot displays the Shiny for Python web interface. On the left, a code editor shows the Python code for the application. On the right, the user interface is visible, featuring a title 'Personalized WordCloud', a subtitle 'Based on Recommendation System (Philadelphia Exclusive!)', and a section 'How To Run App' with four steps. Below this, there are two checked checkboxes for 'Want Potential Positive Reviewers?' and 'Want Positive History with respect to Potential Reviewers?'. A dropdown menu for 'Select length of words' is set to '1'. A text input field for 'Business ID' contains the placeholder 'Enter text'. A red error message 'Error: Import "data.zip" File' is displayed on the right side of the interface.

```
1 from shiny import App, render, ui
2 from pathlib import Path
3
4 from bs4 import BeautifulSoup
5 from sklearn.feature_extraction.text import ENGLISH_STOP_WORDS
6 from nltk.stem import PorterStemmer
7 from nltk.util import ngrams
8 from zipfile import ZipFile
9
10 import unicodedata
11 import pickle
12 import re
13 import string
14 import json
15 import pandas as pd
16 import numpy as np
17
18 import matplotlib.pyplot as plt
19
20 class BubbleChart:
21     def __init__(self, area, bubble_spacing=0):
22         """
23         Setup for bubble collapse.
24
25         Parameters
26         """
27
28 ut_obs
29 message[output_name] = fn()
30 File "/lib/python3.10/site-packages/shiny/render/_render.py", line 207, in __call__
31
32 return _utils.run_coro_sync(self._run())
33 File "/lib/python3.10/site-packages/shiny/_utils.py", line 226, in run_coro_sync
34 coro.send(None)
35 File "/lib/python3.10/site-packages/shiny/render/_render.py", line 224, in _run
36 x = await self._fn()
37 File "/lib/python3.10/site-packages/shiny/_utils.py", line 181, in fn_async
38 return fn()
39 File "/home/pyodide/app_v092kojwmg7honzlwh6l/app.py", line 292, in plot
40 raise IOError('Import "data.zip" File')
41 OSError: Import "data.zip" File
```

Run Business List

Personalized WordCloud

Based on Recommendation System (Philadelphia Exclusive!)

How To Run App

1. Download data.zip file from [This](#)
2. Upload data.zip
3. Enter Your Business ID
You may find business_id available on the other tab
4. Select Various Options

☒ Want Potential Positive Reviewers?

☒ Want Positive History with respect to Potential Reviewers?

Select length of words

1

Business ID

Enter text

Error: Import "data.zip" File

It needs some time to initialize and upload the full data set.
Then, by entering a business ID from the business list, we can get the word cloud.

WordCloud & Shiny App

You may find `business_id` available on the other tab

4. Select Various Options

☒ Want Potential Positive Reviewers?

☒ Want Positive History with respect to Potential Reviewers?

Select length of words

1

Business ID

MUTTqe8uqyMdB186RmNeA

Choose a file to upload:

Browse... data.zip

Upload complete

Frequently Used Words For Potential Positive/Negative Reviewers



Business Name: Tuna Bar

Business Address: 205 Race St, Philadelphia, PA, 19106

Business Stars: 4.0

Business Categories: Sushi Bars, Restaurants, Japanese

Here is a example with one word length

WordCloud & Shiny App

You may find `business_id` available on the other tab

4. Select Various Options

☒ Want Potential Positive Reviewers?

☒ Want Positive History with respect to Potential Reviewers?

Select length of words

3

Business ID

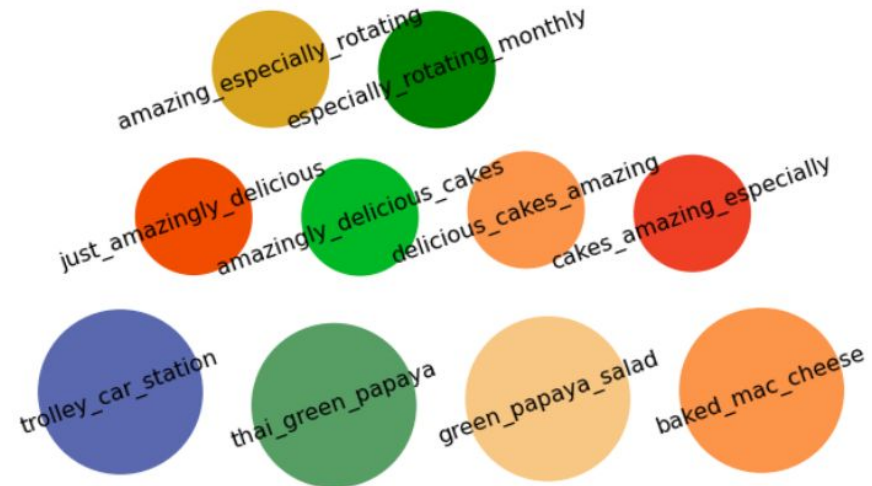
MUTTqe8uqyMdB1186RmNeA

Choose a file to upload:

Browse... data.zip

Upload complete

Frequently Used Words For Potential Positive/Negative Reviewers



Business Name: Tuna Bar

Business Address: 205 Race St, Philadelphia, PA, 19106

Business Stars: 4.0

Business Categories: Sushi Bars, Restaurants, Japanese

Change the word length to 3

WordCloud & Shiny App

You may find `business_id` available on the other tab

4. Select Various Options

☒ Want Potential Positive Reviewers?

☒ Want Positive History with respect to Potential Reviewers?

Select length of words

3

Business ID

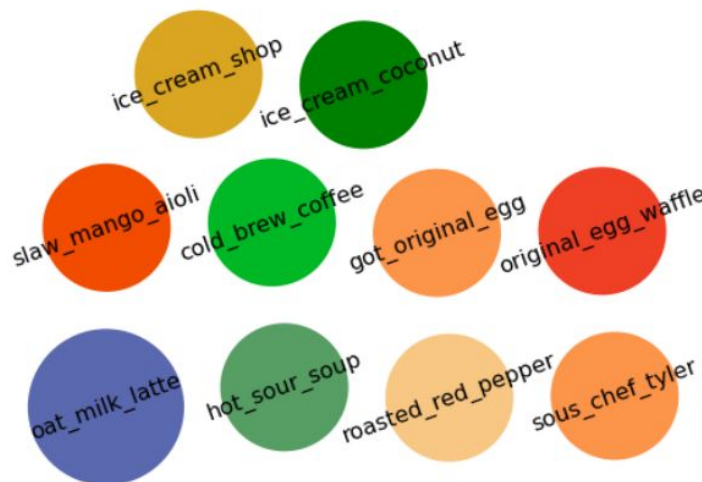
GOSVLhjUT6TnrPhkkmDOPw

Choose a file to upload:

Browse... data.zip

Upload complete

Frequently Used Words For Potential Positive/Negative Reviewers



Business Name: Vineyards Cafe

Business Address: 847 N 16th St, Philadelphia, PA, 19130

Business Stars: 4.5

Business Categories: Coffee & Tea, Cafes, Food, Creperies, Internet Cafes, Restaurants

Conclusion

- Final model: Neural Network model
- Recommendation system: predicted unrated star reviews
- Word embedding technique: similar business categories
- RMSE: 1.125
- NLP: Generating the wordcloud
- Shiny visualization: Wordcloud for every business owner
- Next step: business with few reviews?
hyperparameter tuning with Keras-Tuner?

Thank You