

DSCI 510: Principles of Programming for Informatics

Professor: Gleb Satyukov

Final Project Report

December 7, 2023

Chia-chi, Lee

SafeBite Insights- A Comprehensive Analysis of Restaurant Quality and Safety

1. Introduction

"SafeBite Insights" is an innovative data analysis project that focuses on Los Angeles City and aims to provide a holistic understanding of restaurant quality and safety by merging three crucial datasets: restaurant basic information, reviews dataset from Yelp API, and health inspection dataset from "the County of Los Angeles Open Data". This project delves into the wealth of information available on Yelp, exploring users' reviews and ratings. By combining this user-generated content with local health inspection data, the analysis aims to observe correlations and insights about restaurant quality and safety. In addition, utilize different visualization methods such as Word Cloud and charts to offer comprehensive and precise results for customers. In conclusion, SafeBite Insights bridges the gap between user-generated sentiments on Yelp and official health inspection data and sheds light on the relationship between these variables in the context of Los Angeles City.

2. Data Collection and Processed

a. Data Sources

The project has three datasets: basic restaurant information, reviews dataset from Yelp API, and health inspection dataset from the County of Los Angeles Open Data. In order to gather the restaurant's basic information in Los Angeles City from Yelp API (<https://api.yelp.com/v3/businesses/search>), I created my account and requested an authorization key. Then, I started a new JSON file to store the restaurant's basic information and used the restaurant's primary ID to find and keep their reviews from another Yelp API ([https://api.yelp.com/v3/businesses/+restaurant_id+"/reviews](https://api.yelp.com/v3/businesses/+restaurant_id+)) that provided customer reviews and ratings for the restaurant (Figures 1 and 2). The third dataset I downloaded the restaurant inspection CSV file from "the County of Los Angeles Open Data" (<https://data.lacounty.gov/datasets/environmental-health-restaurant-and-market-inventory-09-30-2023-1/about>) that included scores, grades, descriptions, and other results in the inspection (Figure 3). There are 50 restaurants with 150 reviews selected in the project.

```

> xlsx2csv(2015-ssocsv)
# reviews:
  # 0:
    id:
      "5e90f1e6030b30c00e"
    url:
      "https://www.yelp.com/biz/evishkashvinsky-18666672"
    text:
      "Hello, everyone! Evishka is truly cares about..."
    rating:
      5
    time_created:
      "2015-11-18 00:13:15"
    user:
      (-)
  # 1:
    id:
      "c0e978701b5e9ec9ec"
    url:
      "https://www.yelp.com/biz/evishkashvinsky-18666672"
    text:
      "It goes without saying t. Fun, and warm. what..."
    rating:
      5
    time_created:
      "2015-10-27 17:38:13"
    user:
      (-)
  # 2:
    id:
      "b9e700843030c4f6a9"
    url:
      "https://www.yelp.com/biz/evishkashvinsky-18666672"
    text:
      "seeing establishment to begin to celebrate my..."
    rating:
      5
    time_created:
      "2015-11-13 16:57:56"
    user:
      (-)
total:
  2589
# possible_languages:
  # 0:
    1:
      "en"
    2:
      "es"
    3:
      "it"
    4:
      "fr"
# 769b0c1a931031a2040:
# reviews:
  # 0:
    id:
      "976_g7e10e030b30c00e"
    url:
      "https://www.yelp.com/biz/evishkashvinsky-18666672"
    text:
      "Excellent food! Great L.fond, warmware, etc..."
    rating:
      5
    time_created:
      "2015-10-26 13:00:20"
    user:
      (-)

```

Figure 2 (Right): Reviews information JSON file

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
Serial number	Activity/Location	Score	Grade	Service	Converse	Decorative	Facility	Facility	Facility	Facility	Facility	Facility	Facility	Facility	Facility	Facility	Facility	Facility	Facility
1	DAIR0078	12/25/HABITAT COFFEE SHO	95	A	ROUTINE000002 378 N ELOS AN FM041746C	90025 00W01701 ASSESS RESTAU	90014 HABITAT COFFEE SHO	90014 HABITAT COFFEE SHO	90014 HABITAT COFFEE SHO	90014 HABITAT COFFEE SHO	90014 HABITAT COFFEE SHO	90014 HABITAT COFFEE SHO	90014 HABITAT COFFEE SHO	90014 HABITAT COFFEE SHO	90014 HABITAT COFFEE SHO	90014 HABITAT COFFEE SHO	90014 HABITAT COFFEE SHO	90014 HABITAT COFFEE SHO	90014 HABITAT COFFEE SHO
2	DAIC0RFB	12/25/RELLY'S	92	A	ROUTINE000002 100 WORLDS AN FM02144C	90045 00W0084 AREAS RESTAU	90014 RELLY'S ACTIVITY	90014 RELLY'S ACTIVITY	90014 RELLY'S ACTIVITY	90014 RELLY'S ACTIVITY	90014 RELLY'S ACTIVITY	90014 RELLY'S ACTIVITY	90014 RELLY'S ACTIVITY	90014 RELLY'S ACTIVITY	90014 RELLY'S ACTIVITY	90014 RELLY'S ACTIVITY	90014 RELLY'S ACTIVITY	90014 RELLY'S ACTIVITY	90014 RELLY'S ACTIVITY
3	DAISILUR	12/25/STREET CHURROS	93	A	ROUTINE000003 6601 HOLLOS AN FM02144C	90028 00W0239 STREET RESTAU	90014 STREET CHURROS	90014 STREET CHURROS	90014 STREET CHURROS	90014 STREET CHURROS	90014 STREET CHURROS	90014 STREET CHURROS	90014 STREET CHURROS	90014 STREET CHURROS	90014 STREET CHURROS	90014 STREET CHURROS	90014 STREET CHURROS	90014 STREET CHURROS	90014 STREET CHURROS
4	DAILUSAT	12/25/TRINIT BCHO PARK	94	A	ROUTINE000002 1814 S SLOS AN FM02252C	90032 00W0426 AMERIC RESTAU	90014 TRINIT ACTIVITY	90014 TRINIT ACTIVITY	90014 TRINIT ACTIVITY	90014 TRINIT ACTIVITY	90014 TRINIT ACTIVITY	90014 TRINIT ACTIVITY	90014 TRINIT ACTIVITY	90014 TRINIT ACTIVITY	90014 TRINIT ACTIVITY	90014 TRINIT ACTIVITY	90014 TRINIT ACTIVITY	90014 TRINIT ACTIVITY	90014 TRINIT ACTIVITY
5	DAKYZEAT	12/25/POLLEN	94	A	ROUTINE000002 2100 ECHLOS AN FM02272C	90026 00W0400 POLLN RESTAU	90014 POLLEN ACTIVITY	90014 POLLEN ACTIVITY	90014 POLLEN ACTIVITY	90014 POLLEN ACTIVITY	90014 POLLEN ACTIVITY	90014 POLLEN ACTIVITY	90014 POLLEN ACTIVITY	90014 POLLEN ACTIVITY	90014 POLLEN ACTIVITY	90014 POLLEN ACTIVITY	90014 POLLEN ACTIVITY	90014 POLLEN ACTIVITY	90014 POLLEN ACTIVITY
6	DAKZORLL	12/25/STREET ORILL	93	A	ROUTINE000002 3300 ECHLOS AN FM02272C	90024 00W0400 POLLN RESTAU	90014 STREET ORILL	90014 STREET ORILL	90014 STREET ORILL	90014 STREET ORILL	90014 STREET ORILL	90014 STREET ORILL	90014 STREET ORILL	90014 STREET ORILL	90014 STREET ORILL	90014 STREET ORILL	90014 STREET ORILL	90014 STREET ORILL	90014 STREET ORILL
7	DAKSBTMC	12/25/LEVEE 47215A	93	A	ROUTINE000003 3330 S FLOS AN FM043747C	90043 47215A 47215A IN FOOD M	90014 LEVEE ACTIVITY	90014 LEVEE ACTIVITY	90014 LEVEE ACTIVITY	90014 LEVEE ACTIVITY	90014 LEVEE ACTIVITY	90014 LEVEE ACTIVITY	90014 LEVEE ACTIVITY	90014 LEVEE ACTIVITY	90014 LEVEE ACTIVITY	90014 LEVEE ACTIVITY	90014 LEVEE ACTIVITY	90014 LEVEE ACTIVITY	90014 LEVEE ACTIVITY
8	DAACLNFD	12/25/CFK	91	A	ROUTINE000003 100 WORLDS AN FM02252C	90045 00W0084 AREAS RESTAU	90014 CFK ACTIVITY	90014 CFK ACTIVITY	90014 CFK ACTIVITY	90014 CFK ACTIVITY	90014 CFK ACTIVITY	90014 CFK ACTIVITY	90014 CFK ACTIVITY	90014 CFK ACTIVITY	90014 CFK ACTIVITY	90014 CFK ACTIVITY	90014 CFK ACTIVITY	90014 CFK ACTIVITY	90014 CFK ACTIVITY
9	DAACRFNS	12/25/PHO LALA	92	A	ROUTINE000009 3500 W 6 LOS AN FM02489C	90020 00W0432 VANILL RESTAU	90014 PHO LALA ACTIVITY	90014 PHO LALA ACTIVITY	90014 PHO LALA ACTIVITY	90014 PHO LALA ACTIVITY	90014 PHO LALA ACTIVITY	90014 PHO LALA ACTIVITY	90014 PHO LALA ACTIVITY	90014 PHO LALA ACTIVITY	90014 PHO LALA ACTIVITY	90014 PHO LALA ACTIVITY	90014 PHO LALA ACTIVITY	90014 PHO LALA ACTIVITY	90014 PHO LALA ACTIVITY
10	DAEIXAWW	12/25/ABC DONUTS	91	A	ROUTINE000002 3037 N SLOS AN FM02027C	90005 00W0038 PUNJO RESTAU	90014 ABC DONUTS ACTIVITY	90014 ABC DONUTS ACTIVITY	90014 ABC DONUTS ACTIVITY	90014 ABC DONUTS ACTIVITY	90014 ABC DONUTS ACTIVITY	90014 ABC DONUTS ACTIVITY	90014 ABC DONUTS ACTIVITY	90014 ABC DONUTS ACTIVITY	90014 ABC DONUTS ACTIVITY	90014 ABC DONUTS ACTIVITY	90014 ABC DONUTS ACTIVITY	90014 ABC DONUTS ACTIVITY	90014 ABC DONUTS ACTIVITY
11	DAEISG02	12/25/ELLO LOCO	91	A	ROUTINE000003 5330 S SLOS AN FM002125C	90027 00W0108B POLLN RESTAU	90014 ELLO LOCO ACTIVITY	90014 ELLO LOCO ACTIVITY	90014 ELLO LOCO ACTIVITY	90014 ELLO LOCO ACTIVITY	90014 ELLO LOCO ACTIVITY	90014 ELLO LOCO ACTIVITY	90014 ELLO LOCO ACTIVITY	90014 ELLO LOCO ACTIVITY	90014 ELLO LOCO ACTIVITY	90014 ELLO LOCO ACTIVITY	90014 ELLO LOCO ACTIVITY	90014 ELLO LOCO ACTIVITY	90014 ELLO LOCO ACTIVITY
12	DAEYOUTBU	12/25/POCONO NTM BAKER'S	91	A	ROUTINE000005 928 S WELOS AN FM040047C	90006 00W0193 CDM, JRN RESTAU	90014 POCONO ACTIVITY	90014 POCONO ACTIVITY	90014 POCONO ACTIVITY	90014 POCONO ACTIVITY	90014 POCONO ACTIVITY	90014 POCONO ACTIVITY	90014 POCONO ACTIVITY	90014 POCONO ACTIVITY	90014 POCONO ACTIVITY	90014 POCONO ACTIVITY	90014 POCONO ACTIVITY	90014 POCONO ACTIVITY	90014 POCONO ACTIVITY
13	DAE7HMBU	12/25/EMC RESTAURANT GR	90	A	ROUTINE000009 3500 W 6 LOS AN FM041631C	90020 00W0127 EMC RESTAU	90014 EMC RESTAURANT ACTIVITY	90014 EMC RESTAURANT ACTIVITY	90014 EMC RESTAURANT ACTIVITY	90014 EMC RESTAURANT ACTIVITY	90014 EMC RESTAURANT ACTIVITY	90014 EMC RESTAURANT ACTIVITY	90014 EMC RESTAURANT ACTIVITY	90014 EMC RESTAURANT ACTIVITY	90014 EMC RESTAURANT ACTIVITY	90014 EMC RESTAURANT ACTIVITY	90014 EMC RESTAURANT ACTIVITY	90014 EMC RESTAURANT ACTIVITY	90014 EMC RESTAURANT ACTIVITY
14	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY
15	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY
16	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY
17	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY
18	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY
19	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY
20	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY
21	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY
22	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY
23	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY
24	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY
25	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY
26	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY
27	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY
28	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY
29	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY
30	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY
31	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY
32	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY
33	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY
34	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY
35	DAE30PFB	12/25/POPULAR DONUTS	90	A	ROUTINE000002 3614 W SLOS AN FM040444C	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS ACTIVITY	90014 POPULAR DONUTS	

Figure 3: Restaurant inspection information CSV file

There are four steps that each dataset needs to process: handling missing values, removing duplicates, converting data into a structured format, and converting string to lowercase. In order to complete the steps, utilize the “replace_nan_value”, “drop_duplicate()”, “to_numeric()”, and “lower()” functions to those JSON and CSV files. For the “replace_nan_value” function, I built it as an individual function, for both two JSON files have list and dictionary values. The “Replace_nan_value” function can identify the value and eventually set the value to the empty value (Figure 4).

```
# Handling Missing Values
def replace_nan_value(obj):
    # if it is a dictionary
    if isinstance(obj, dict):
        return {key: replace_nan_value(value) for key, value in obj.items()}
    # if it is a list
    elif isinstance(obj, list):
        return [replace_nan_value(element) for element in obj]
    elif pd.isna(obj) or obj == 'null':
        return ''
    else:
        return obj
```

Figure 4: The “Replace nan value” function in clean_data.py

c. Challenges

One of the challenges in the project is that it is difficult to read the JSON file. To deal with this problem, I downloaded Firefox web browser to open all JSON files, and it works. Another challenge is that Yelp API only offers three reviews in one restaurant, so that the reviews samples are less and not enough. Unfortunately, I cannot solve this problem, but I gather more restaurants detailed information and refrain from running analyses that will be impacted by this reason. Although facing these challenges, I did not change my original plan: analyze the correlation and insights about restaurant quality and safety.

3. Data analysis and Visualization

a. Analysis Techniques and Findings

(1) Sentiment Analysis of Reviews

The sentiment analysis technique uses TextBlob to identify customer reviews as positive, negative, or neutral and provide the sentiment score. The finding in sentiment analysis of reviews is that positive sentiment reviews were prevalent, contributing to a high overall positive sentiment score. The Word Cloud visualization showcased frequently mentioned words, emphasizing aspects like 'service,' 'ambiance,' and 'cleanliness.'

(2) Correlation Analysis - Yelp Ratings vs. Sentiment Scores

The correlation analysis uses the matplotlib library and calculates the Pearson correlation coefficient between restaurant ratings and sentiment scores. The finding in correlation analysis is that a moderate positive correlation (0.357) was observed, indicating a connection between user satisfaction and sentiment. The scatter plot with a fitted linear regression line illustrated the relationship between numerical ratings and sentiment scores.

(3) Correlation Analysis - Inspection Scores vs. Yelp Ratings

The correlation analysis techniques use the seaborn library to plot the regression results and calculate the Pearson correlation coefficient between inspection scores and Yelp ratings. The finding in the regression plot demonstrated a positive correlation between restaurant ratings and inspection scores. The correlation coefficient was found to be 0.346, indicating a moderate positive relationship.

(4) Language Distribution Analysis

The language distribution analysis techniques use the matplotlib library to display the distribution of possible languages among restaurant reviews. The finding in the pie chart illustrated the prevalence of English in restaurant reviews, with a diverse representation of other languages. This insight is crucial for understanding the linguistic diversity of customer feedback.

(5) Comprehensive Analysis - Merged DataFrames

The comprehensive analysis techniques use the Pandas library to merge the datasets and utilize the seaborn library to create a pair plot and a heatmap. The finding in the merged DataFrame provided an integrated view, facilitating the examination of relationships between various attributes. The pair plot and correlation matrix heatmap revealed correlations between numerical variables, offering a deeper understanding of interdependencies.

b. Description of Figures

(1) Word cloud for restaurant reviews versus inspection

The restaurant reviews word cloud is set by customers' review text, and descriptions are set from the organization of the inspection word cloud. Utilize the Word Cloud library; all of the word sizes correspond to word frequency and provide the concern of reviews and inspection descriptions (Figures 5 and 6).



Figure 5 (Left): Word Cloud of reviews

Figure 6 (Right): Word Cloud of inspection description

(2) Sentiment scores versus Numerical scores (customers' ratings) plot

Utilized a scatter plot with a fitted linear regression line to showcase the relationship between sentiment scores and numerical ratings. Each point represents a restaurant, and the regression line indicates the trend in sentiment concerning numerical ratings (Figure 7).

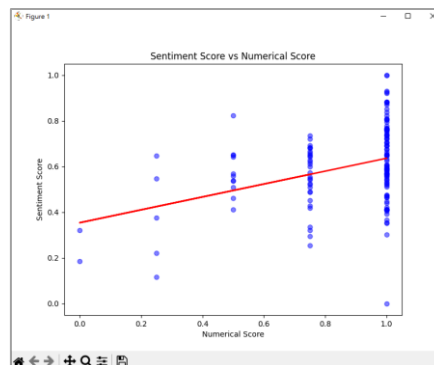


Figure 7: Sentiment Score vs Numerical Score

(3) Regression Plot of Inspection Score vs. Restaurant Score

Presented a regression plot to visualize the correlation between health inspection scores and restaurant ratings. Scatter points represent individual restaurants, and the plotted regression line illustrates the overall trend (Figure 8). Due to the restaurant's ratings only having 4.0 and 4.5, the distribution of points range is extreme.

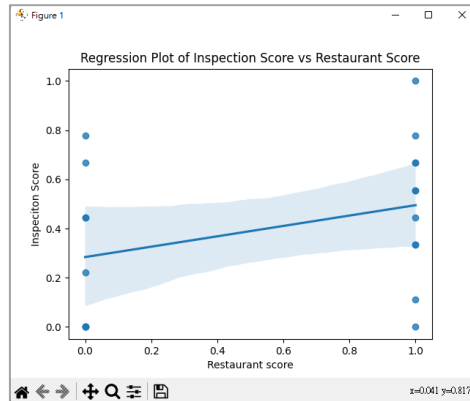


Figure 8: Regression Plot of Inspection Score vs Restaurant Score

(4) Language Distribution Pie Chart

Pie chart showcasing the distribution of possible languages among restaurant reviews. Slices represent different languages, with accompanying labels indicating the percentage of reviews in each language (Figure 9). For instance, ‘en’ means English, and ‘fr’ means French.

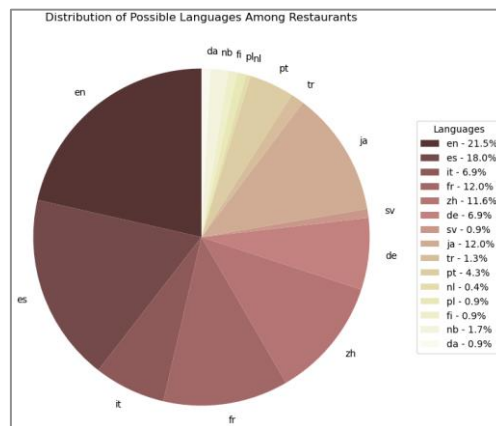


Figure 9: Distribution of Possible Languages among Restaurants

(5) Pair Plot and Correlation Matrix Heatmap

I utilized pair plots and a correlation matrix Heatmap in the merged DataFrame to explore relationships between numerical variables such as inspection score, restaurant zip code, and rating. Diagonal histograms show the distribution of individual variables, while scatter plots and the Heatmap visualize correlations between pairs of variables (Figures 10 and 11).

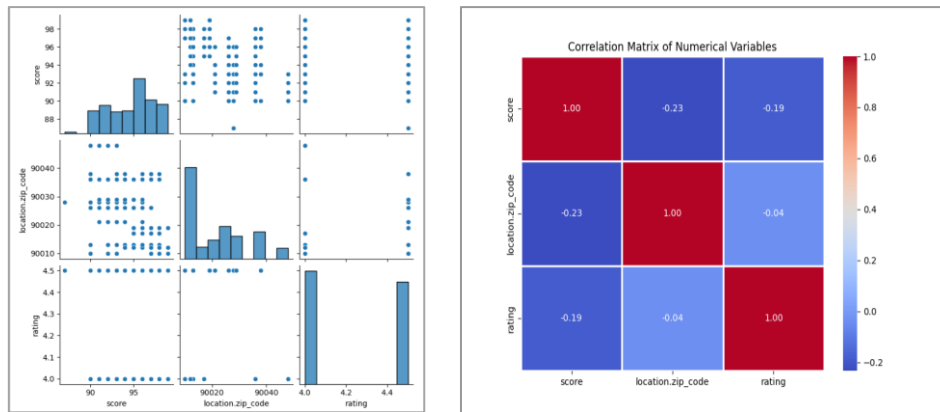


Figure 10 (Left): Pair Plot for inspection score, restaurant zip code, and rating

Figure 11 (Right): Heatmap for inspection score, restaurant zip code, and rating

c. Observations and Conclusion

Four observations are displayed in the analyses. First, the word Cloud in inspection descriptions and restaurant reviews shows customers care more about the food, service, and place. In contrast, an inspection organization focuses on the restaurant's environment, such as the quality of the seats. Second, the positive correlation between Yelp ratings and sentiment scores suggests that higher numerical ratings align with more positive sentiments in reviews. Third, the correlation between health inspection scores and Yelp ratings indicates a moderate positive relationship, emphasizing the potential connection between public perception and regulatory compliance. Lastly, the language distribution analysis highlights the linguistic diversity of customer feedback, with English being the predominant language.

d. Impact of Findings

Three findings that can be found in the analysis results are consumer empowerment, regulatory awareness, and diverse feedback understanding. Consumer empowerment means that consumers can gain insights into the relationships between user satisfaction, sentiment, and objective health inspection scores, aiding informed decision-making. Regulatory awareness means restaurants can better understand the correlation between customer satisfaction and inspection outcomes, promoting proactive measures to enhance both. Diverse feedback understanding means that the language distribution analysis provides a nuanced understanding of the linguistic diversity in customer reviews, facilitating targeted engagement strategies. In conclusion, these analyses contribute to a holistic understanding of the Los Angeles restaurant landscape, bridging user sentiments, health inspection outcomes, and linguistic diversity to benefit consumers, restaurateurs, and regulatory authorities.

4. Future Work

For the data aspect, I will seek more restaurant reviews and inspection information, for there are not many restaurants match in both restaurant information from Yelp API and inspections, which makes the regression plot of inspection scores and restaurant ratings does not have many samples to calculate and might impact their results. In addition, I will incorporate external factors, such as economic conditions and public events, on restaurant quality and safety, offering a more comprehensive understanding for customers. Eventually, I will develop an interactive user interface to allow users to dynamically explore and visualize the data.