

NYC Restaurant Inspection

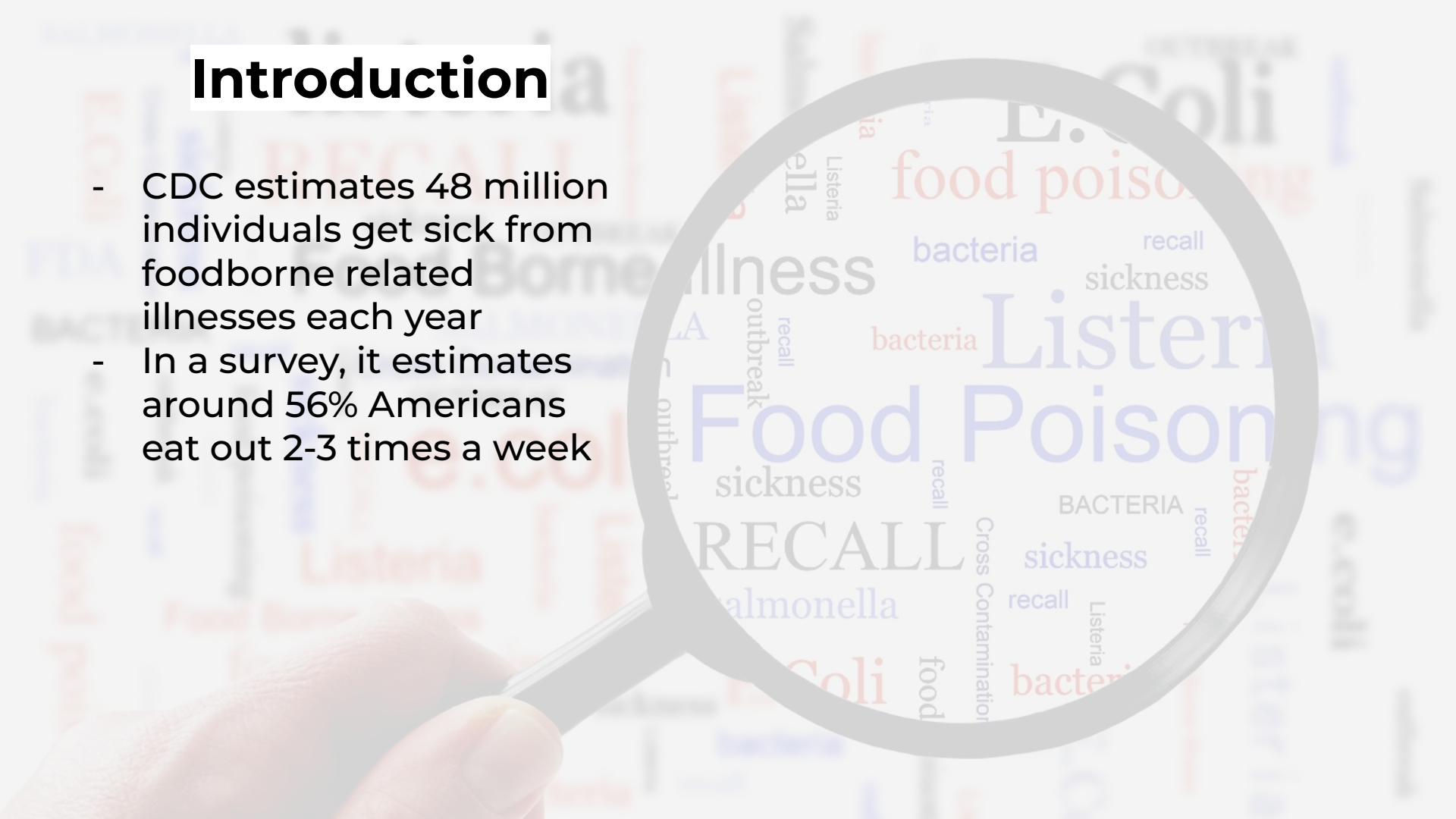


Hannah Kim

Data Engineering Project

Introduction

- CDC estimates 48 million individuals get sick from foodborne related illnesses each year
- In a survey, it estimates around 56% Americans eat out 2-3 times a week



Objective

- To bring awareness on potential risks when eating at restaurants with a history of violations
- Foodborne illness outbreaks occur not only from manufactured products, but also from restaurants

FOOD AND DRINKS · Published April 21, 2020 5:13pm EDT

Chipotle to pay record \$25M fine over foodborne illness outbreaks

Outbreaks at Chipotle restaurants sickened more than 1,100 people from 2015 to 2018

By Thomas Barrabi | FOXBusiness



[Chipotle Mexican Grill](#) will pay a \$25 million fine to [settle criminal charges](#) related to its role in outbreaks of foodborne illnesses at its restaurants beginning in 2015, the [U.S. Department of Justice](#) said Tuesday.

Markets

Quote Lookup 

DOW JONES FUTURES (I:DJF)

31,018.00
▼ -148.00 (-0.48%)

NASDAQ FUTURES (I:COMPX)

11,955.50
▼ -63.50 (-0.53%)

S&P 500 FUTURES (IND)

3,891.00
▼ -19.50 (-0.50%)

Market Update sponsored by

 **Ameritrade** Discover true value 

← Ad by **CRITEO**

[Report this ad](#)

Workflow Overview

NYC OpenData

Google Maps Platform

GeoPy

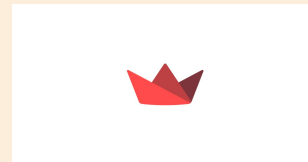


pandas

NumPy



SQLAlchemy



Data Pipeline

Data Ingestion

- API requests from SODAPY, Google Map Platform, Geopy

Pre-Processing

- Using pandas cleaned data

Processing

- Store into relational database using SQLAlchemy

Processing

- Load SQL database
- Read data into pandas
- Queries for analysis
- Save into csv files

Deployment

- Streamlit web application
- Interactive

1. Data Ingestion

DOHMH New York City Restaurant Inspection
Results data from NYC Open Data Source

- Obtained with SODAPY API
 - 248,020 restaurant violations
 - 26 features
 - 2015-2022
 - Used Google Map Platform API + Geopy
 - (latitude, longitude) to map in web app
-

2. Pre-Processing Stage

- pandas and numpy to data clean
 - Removed data where Inspection date : 1/1/1900-
 - indicates establishments that have not been inspected yet
 - Drop duplicates
 - 66174 rows of data (reduced 70% of data)
-

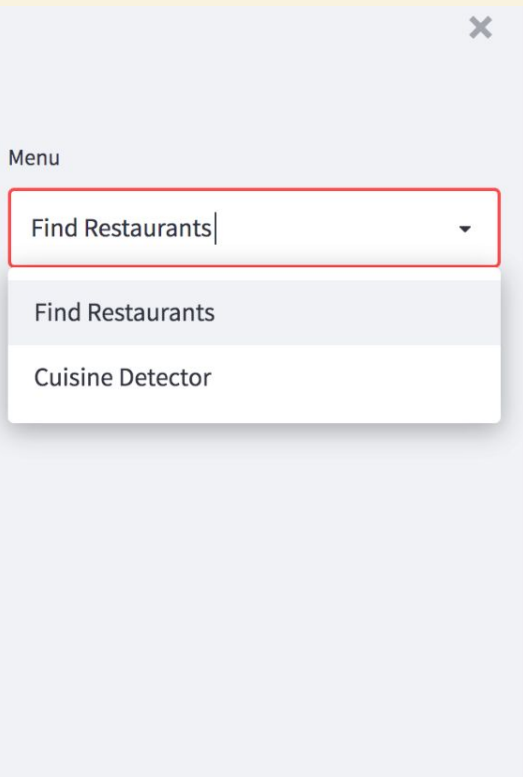
3. Processing Stage

- Store data into SQL database
 - Load by reading into pandas
 - SQL queries for data analysis
 - Save as CSV file
-
- From 2015- 2022: American cuisine had the most cumulative scores
 - From 2020-2021: Number of inspections low due to the pandemic + Covid
-

4. Deployment

- Web application using Streamlit
- Folium for mapping
- Features:
 - Grade level color coded
 - Popup marker lists: name, cuisine, address, grade, score
 - Number of loaded results indicated

4. Deployment



NYC Restaurant Inspection

Introduction

Click to expand

An average of 56% Americans eat out about 2-3 times a week revealed in a [survey](#).

Statistics released by [CDC](#) indicates roughly every year around "48 million individuals get sick" from foodborne illnesses. This dashboard aims to give awareness to the public on how restaurants in NYC are scored and graded in terms of food safety.

4. Deployment

×

Click to expand

+

Menu

Find Restaurants

Restaurant Viewer

Inspection Year

2018

Cuisine

American

Grade

B


C

Score Range

0.0

104.0

Submit



Leaflet | Data by © OpenStreetMap, under ODbL

4. Deployment

×

Click to expand

+

Menu

Find Restaurants

Restaurant Viewer

Inspection Year

2018

Cuisine

American X

Grade

B x C x

Score Range

Submit

+

– Name: PROHIBITION

Cuisine: American

Address: 503 COLUMBUS AVENUE NYC

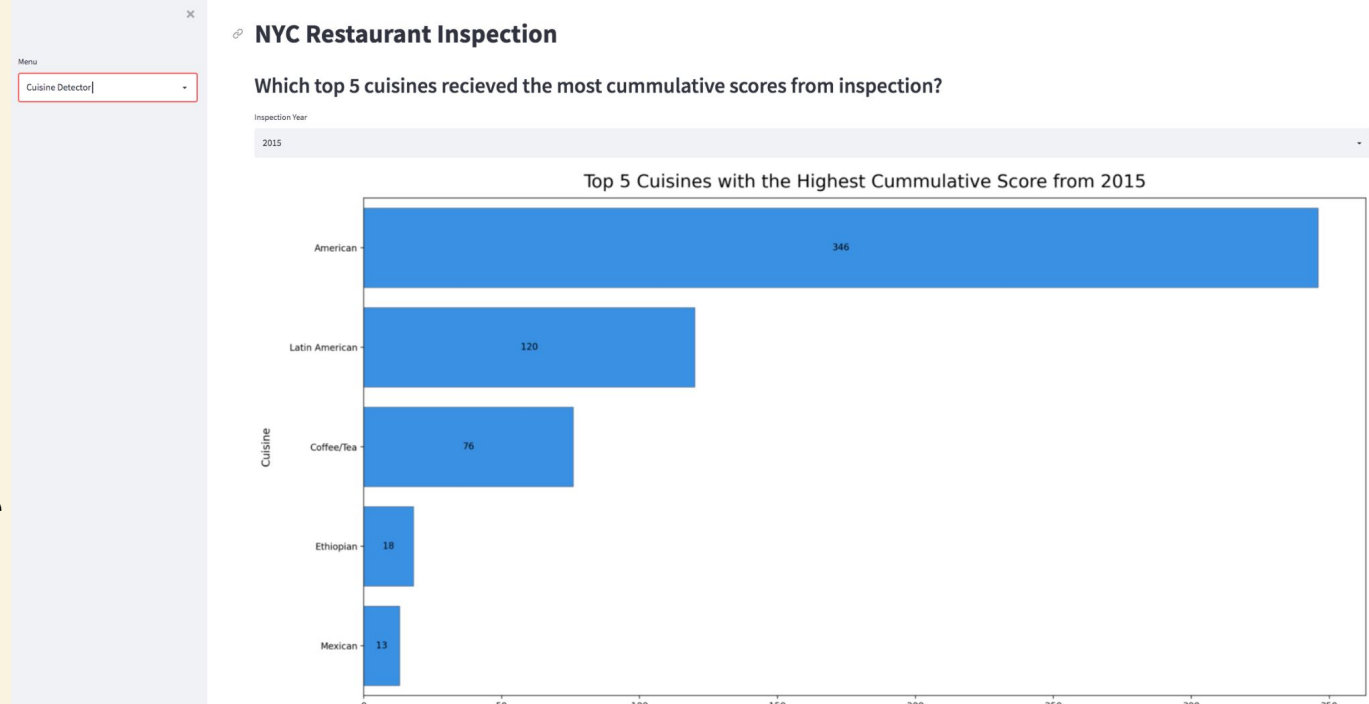
Grade: C

Score: 19.0

Loaded 70 results.

4. Deployment

- Seaborn + Matplotlib for barchart visualization
- User selects year to update barchart with the top 5 cuisines with highest cumulative scores



Future Work

- Automate API that updates the data on a daily basis
 - Create search bar engine where user can input restaurant name
 - Implement nearby location search where top 5 results of nearby restaurants will appear on web application
 - Deploy app either on Heroku or Github
-