OpenTable™ RESTAURANT ANALYSIS

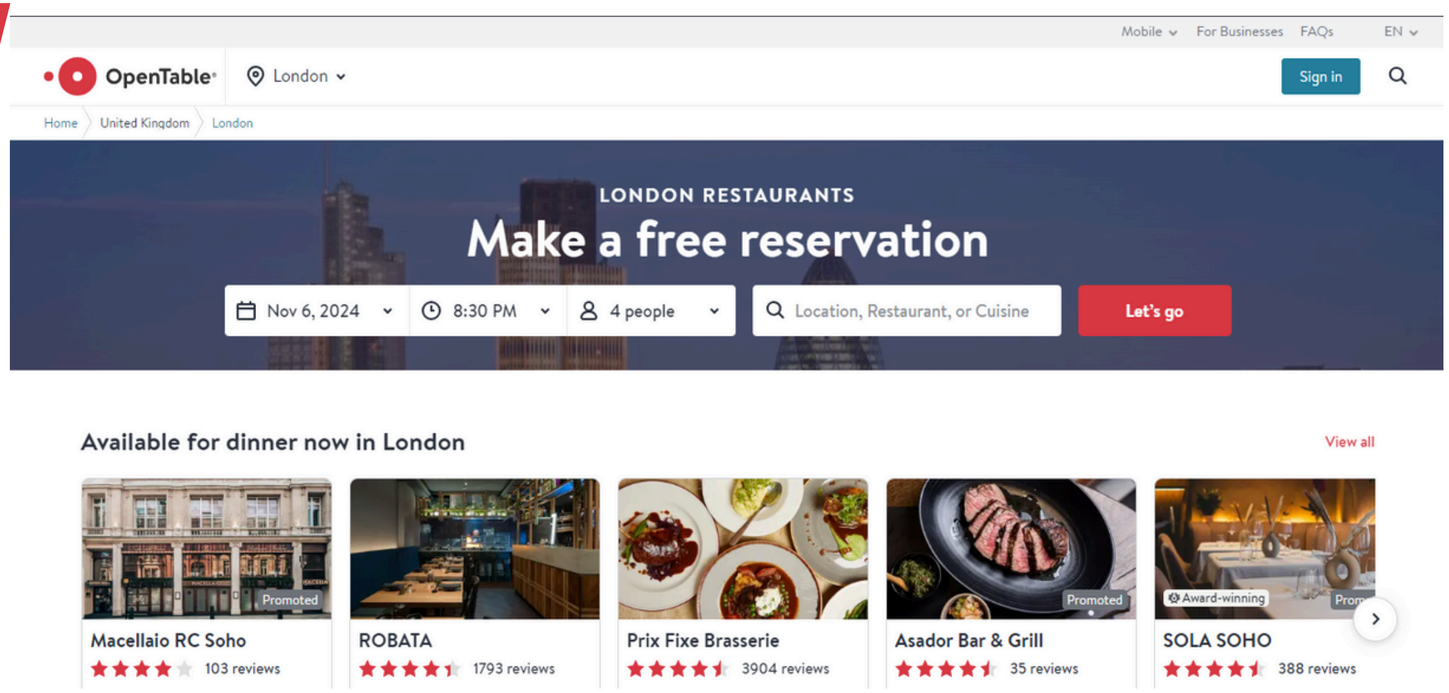
RAISA ALEKSANYAN

Agenda

1. SCRAPPING
2. DATA CLEANING
3. EDA
4. CSV_TO_POSTGRESQL
5. POSTGRESQL
6. TESTING
7. LLM_DASH
8. POWER BI

Scrapping

```python
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.common.exceptions
import NoSuchElementException,
ElementClickInterceptedException,
ElementNotInteractableException,
StaleElementReferenceException,
TimeoutException
from selenium.webdriver.support.ui
import WebDriverWait
from selenium.webdriver.support
import expected_conditions as EC
import pandas as pd
import time
import random
```



- url
- rest_name (restaurant name)
- number_of_reviews (number of reviews of each restaurant)
- rating (customer rating)
- food_type (type of cuisine offered)
- coupon ()
- food (rating of food)
- ambience (rating of atmosphere)
- service (rating of provided service)
- value (rating of price/quality ratio)
- about_rest (brief information about restaurant)
- comments (comments of users regarding restaurant)
- image_url (image of the restaurant)

OpenTable™

AS A RESULT OF SCRAPPING, I HAVE THE FOLLOWING DATASETS:

- ALBERTA
- MANITOBA
- ONTARIO
- QUEBEC
- VANCOUVER

Data Cleaning

- CHECK FOR EMPTY CELLS
- CONVERT DATA TYPES
- REMOVE DUPLICATE ROWS
- STANDARDIZE TEXT

```
df['rest_name'].is_unique
```

```
True
```

```
duplicates = df['rest_name'][df['rest_name'].duplicated()].unique()
print("Non-unique restaurant names:")
print(duplicates)
```

```
Non-unique restaurant names:
[]
```

```
df = df.drop_duplicates(subset=['rest_name'], keep='first')
```

Database Connection

```python
import pandas as pd
from sqlalchemy import create_engine
```

Connect to PostgreSQL database

```python
# Database connection details
db_username = 'postgres'
db_password = '********'
db_host = 'localhost'
db_port = '5432'
db_name = 'final_project'
```

- Tables (5)
  - alberta
  - manitoba
  - ontario
  - quebec
  - vancouver

PostgreSQL

PostgreSQL

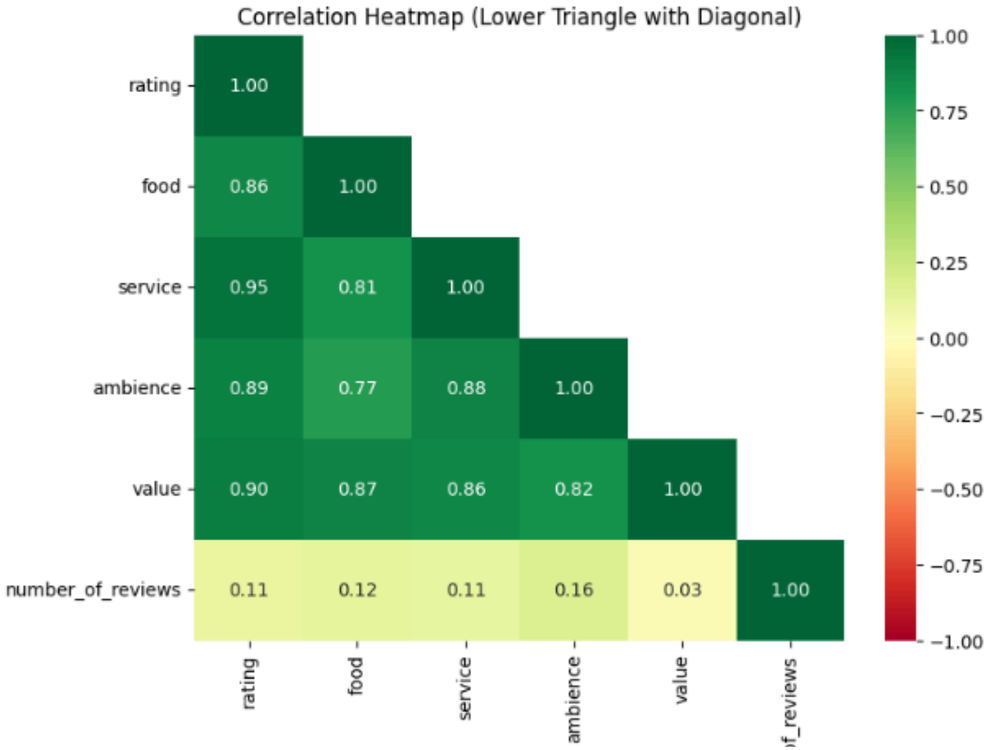
```sql
--18 Using CTE to Rank Restaurants by Number of Reviews
WITH RankedRestaurants AS (
    SELECT rest_name, number_of_reviews,
           RANK() OVER (ORDER BY number_of_reviews DESC) AS review_rank
    FROM alberta
)
SELECT * FROM RankedRestaurants
WHERE review_rank <= 5;
```

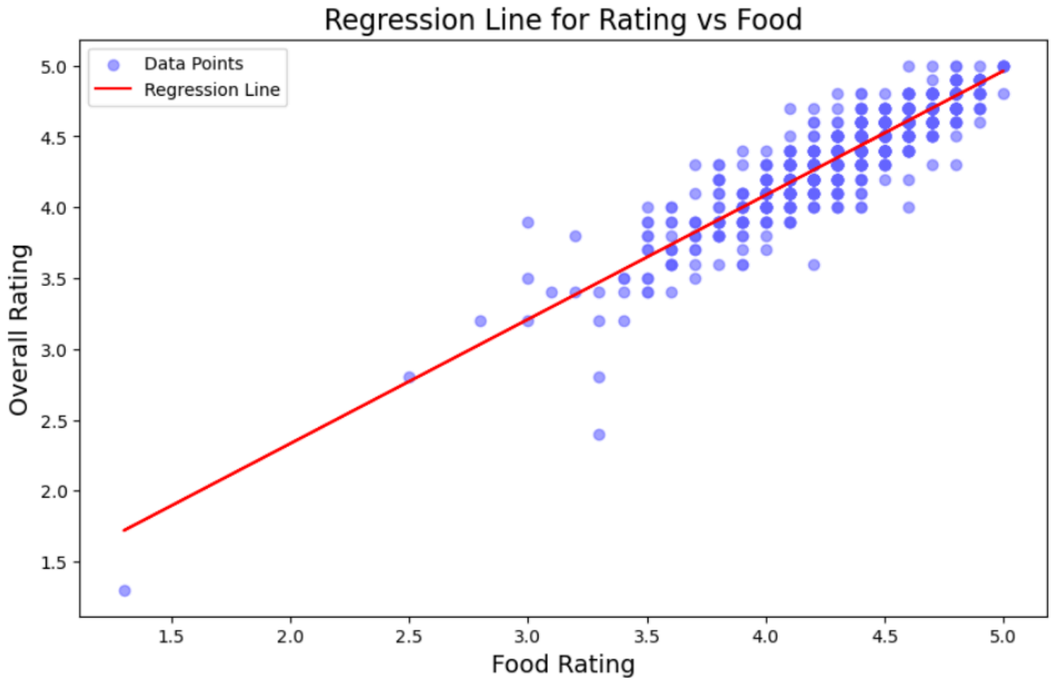
THE RESULT OF A QUERY

	rest_name text	number_of_reviews bigint	review_rank bigint
1	Sabor Restaurant	4688	1
2	The Keg Steakhouse + Bar - West Edmonton	4309	2
3	The Keg Steakhouse + Bar - South Edmonton Com...	3891	3
4	The Keg Steakhouse + Bar - Edmonton - Windermere	3091	4
5	The Melting Pot - Edmonton	2855	5

EDA

- DISTRIBUTION OF RATINGS
- CORRELATION HEATMAP
- REGRESSION LINE
- BAR PLOT
- HISTOGRAM
- SCATTER PLOT





T-TEST and Anova-Test

```python
import pandas as pd
from scipy import stats
```

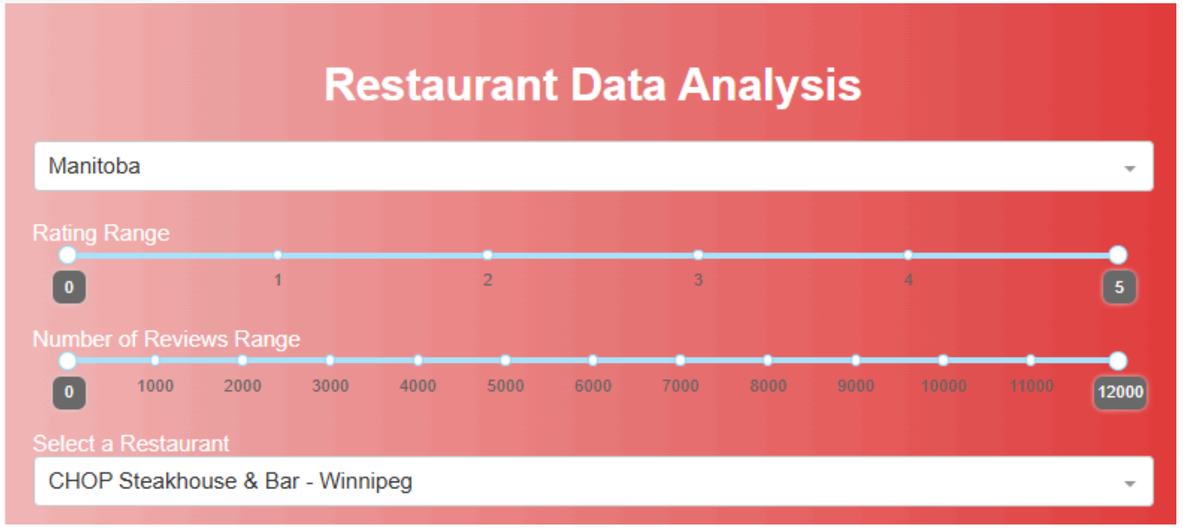
```python
# Statistical Testing: T-tests for pairwise comparisons and ANOVA for all groups
def perform_tests(dataframes, column):
    # Conduct a one-way ANOVA to check for differences between all groups
    data = [df[column].dropna() for df in dataframes.values()]
    anova_result = stats.f_oneway(*data)
    print(f"ANOVA for '{column}': F-statistic = {anova_result.statistic:.3f}, p-value = {anova_result.pvalue:.3e}")

    # Pairwise T-tests between datasets
    dataset_names = list(dataframes.keys())
    for i in range(len(dataset_names)):
        for j in range(i + 1, len(dataset_names)):
            group1 = dataframes[dataset_names[i]][column].dropna()
            group2 = dataframes[dataset_names[j]][column].dropna()
            t_test_result = stats.ttest_ind(group1, group2, equal_var=False)  # Welch's t-test
            print(f"T-test between '{dataset_names[i]}' and '{dataset_names[j]}' for '{column}': "
                  f"T-statistic = {t_test_result.statistic:.3f}, p-value = {t_test_result.pvalue:.3e}")
```

```
=== Testing for column: 'value' ===
ANOVA for 'value': F-statistic = 1.505, p-value = 1.985e-01
T-test between 'Alberta' and 'Manitoba' for 'value': T-statistic = -1.255, p-value = 2.110e-01
T-test between 'Alberta' and 'Ontario' for 'value': T-statistic = -0.337, p-value = 7.362e-01
T-test between 'Alberta' and 'Quebec' for 'value': T-statistic = -1.023, p-value = 3.073e-01
T-test between 'Alberta' and 'Vancouver' for 'value': T-statistic = 0.572, p-value = 5.677e-01
T-test between 'Manitoba' and 'Ontario' for 'value': T-statistic = 1.415, p-value = 1.609e-01
T-test between 'Manitoba' and 'Quebec' for 'value': T-statistic = 0.349, p-value = 7.278e-01
T-test between 'Manitoba' and 'Vancouver' for 'value': T-statistic = 2.156, p-value = 3.263e-02
T-test between 'Ontario' and 'Quebec' for 'value': T-statistic = -1.135, p-value = 2.571e-01
T-test between 'Ontario' and 'Vancouver' for 'value': T-statistic = 1.354, p-value = 1.766e-01
T-test between 'Quebec' and 'Vancouver' for 'value': T-statistic = 1.972, p-value = 4.920e-02
```

LLM and Dash

```python
import os
import pandas as pd
from dash import Dash, dcc, html, Input, Output
import openai
```

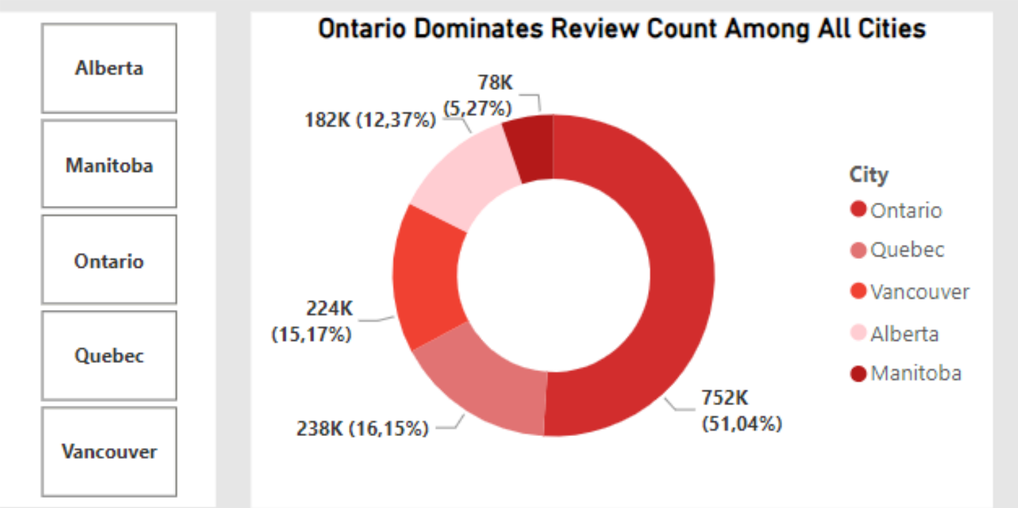
```python
chat_completion = openai.ChatCompletion.create(
    model="gpt-3.5-turbo",
    messages=[
        {"role": "system", "content": "You are a helpful assistant. Summarize the following restaurant reviews."},
        {"role": "user", "content": comments}
    ]
)
summary = chat_completion.choices[0].message.content
```

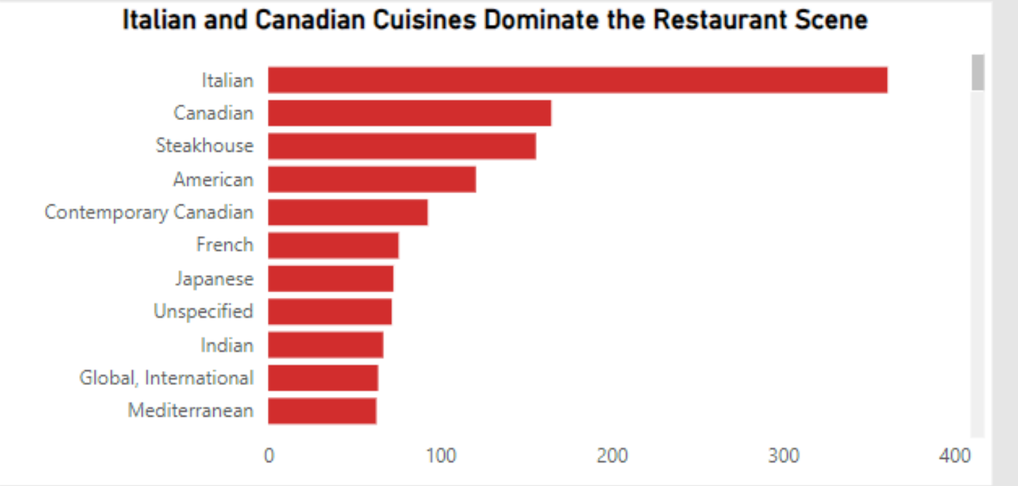


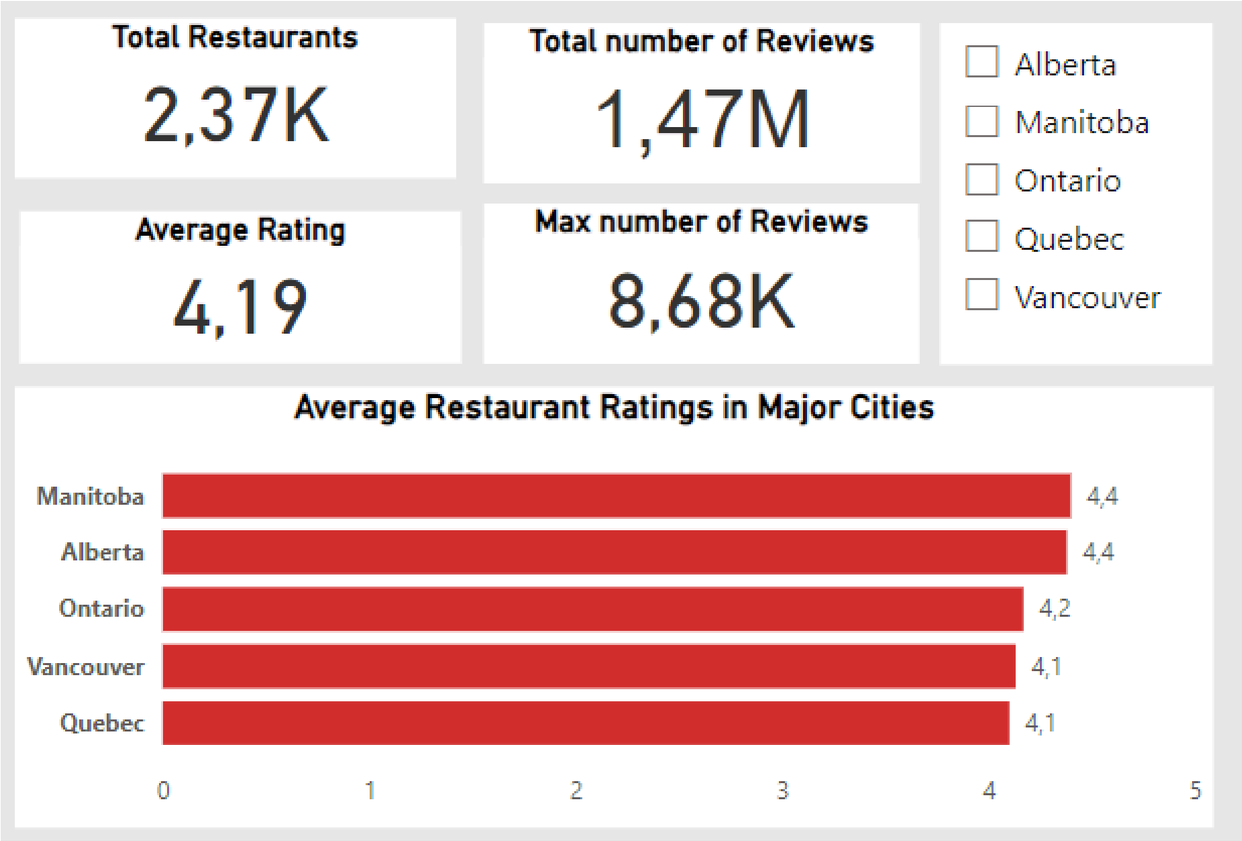
Summary of Comments

The first review mentions delicious food, great ambience, and great service at CHOP, but the steaks were undercooked and had to be sent back twice. The second review comments on the great service and food even though the restaurant was busy. The third review notes a positive experience at CHOP, with a revamped menu, excellent cocktails, delicious appetizers, and entrees cooked to perfection. The reviewer praises the attentive and knowledgeable server. The fourth review also recommends CHOP, praising the food and service, although some dishes were just okay. It mentions the staff's professionalism and courtesy.

Power BI


Alberta
Manitoba
Ontario
Quebec
Vancouver
Ontario Dominates Review Count Among All Cities
78K (5,27%)
182K (12,37%)
224K (15,17%)
238K (16,15%)
752K (51,04%)
City
Ontario
Quebec
Vancouver
Alberta
Manitoba


Italian and Canadian Cuisines Dominate the Restaurant Scene
Italian
Canadian
Steakhouse
American
Contemporary Canadian
French
Japanese
Unspecified
Indian
Global, International
Mediterranean
0
100
200
300
400


Total Restaurants
2,37K
Total number of Reviews
1,47M
Average Rating
4,19
Max number of Reviews
8,68K
Alberta
Manitoba
Ontario
Quebec
Vancouver
Average Restaurant Ratings in Major Cities
Manitoba 4,4
Alberta 4,4
Ontario 4,2
Vancouver 4,1
Quebec 4,1
0
1
2
3
4
5

THANK YOU!