FOOD HABIT ANALYSIS OF NORTH AMERICANS

Name: Md Rashel Uzzaman

Student ID: 23099317

M.Sc. in Artificial Intelligence

Friedrich-Alexander-Universität Erlangen-Nürnberg

Research Question:

■ Is there a pattern in food habits in different regions of North America?

Objective:

To identify regional food preferences and trends across the USA and Canada, aiding businesses in market entry strategies.

Data Overview

■ Dataset Source: Kaggle

■ Dataset Details: 1,500 restaurant entries

■ Attributes: Name Cuisine(s), Rating, City, State, and more

■ Challenges Identified: Multiple cuisine names with inconsistencies (e.g., "burger" vs. "burgers"), Right-skewed restaurant ratings.

Data Pipeline

- Step 1: Download Data: Used httpx library to fetch data from Kaggle.
- Step 2: Data Extraction: Utilized zip file library to extract CSV from ZIP archive.
- Step 3: Data Cleaning and Normalization: Standardized cuisine names using Gensim's glove2word2vec model to reduce unique cuisine entries from 21,521 to 572.
- Step 4: Data Storage: Saved cleaned data in SQLite database for easy querying using SQL Alchemy.

Data Cleaning

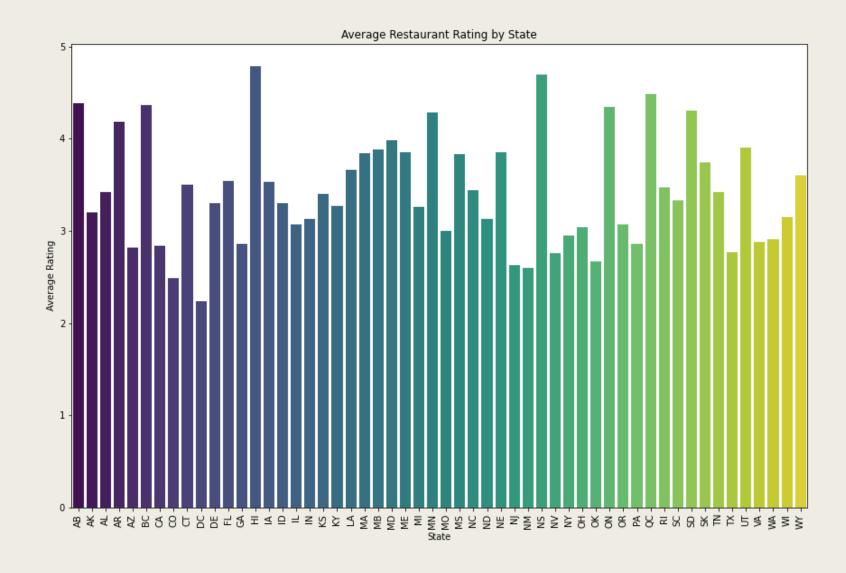
■ Duplicate and inconsistent entries in the cuisine column, such as: "burgers" vs. "burger", "Coffee and Tea" vs. "Coffee & Tea".

***** Result:

■ Reduced unique cuisine names from 21,521 to 572.

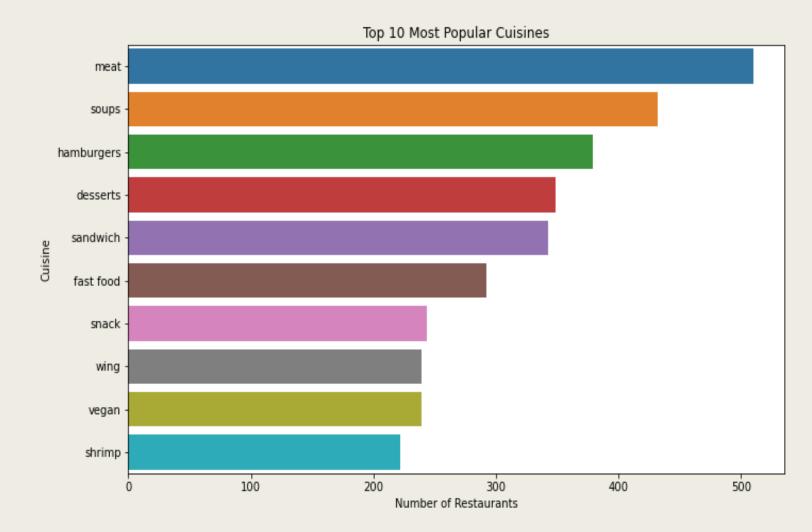
Findings

- States like Hawaii (HI) and Quebec (QC) have higher average ratings.
- States such as Connecticut (CT), California (CA) and New Jersey (NJ) have lower ratings.
- Indicates potential variation in restaurant quality.



Top 10 Most Popular Cuisines

- "Meat," "Soup," and "Hamburgers" are the top three most frequent cuisines.
- "Fast Food" and "Vegan" also feature prominently, reflecting evolving dietary preferences.



Heatmap of Average Cuisine

Ratings

- Heatmap Showing Cuisine Ratings by State.
- Maryland (MD): High ratings for "Soup" and "Vegan."
- Georgia (GA): Low ratings for "Soup" and "Vegan."
- Michigan (MI): Low ratings across top 10 cuisines, indicating unique regional preferences.
- Consistent high ratings for "Hamburgers" and "Fast Food".

Heatmap of Average Cuisine Ratings by State (Normalized)										
AB -	4.04	4.45	3.92	4.52	4.20	4.20	4.49	4.12	3.95	3.38
AK -	3.20	0.00	3.20	0.00	0.00	3.20	0.00	0.00	0.00	0.00
AL -	3.55	3.27	0.00	3.30	3.47	3.15	3.50	3.20	0.00	2.50
AR -	4.23	4.56	4.07	3.15	4.46	4.23	3.00	3.83	3.83	4.40
AZ -	2.83	3.10	3.78	2.60	3.63	3.80	0.00	3.40	4.20	0.00
BC -	4.40	4.25	4.41	4.17	4.60	4.34	4.19	4.51	4.30	4.11
CA -	2.77	2.81	2.36	3.01	2.64	2.87	3.43	2.85	2.56	2.64
CO -	2.83	0.00	3.12	3.13	2.54	3.13	0.00	0.00	2.46	1.00
CT -	0.00	0.00	3.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
DC -	2.70	0.00	2.70	0.00	2.70	2.70	0.00	0.00	0.00	0.00
FL -	3.94	3.97	3.27	3.22	3.60	3.47	2.80	3.45	4.32	3.52
GA -	3.08	2.54	3.37	3.33	2.69	2.95	2.05	2.80	2.80	2.98
HI -	4.70	4.83	4.40	5.00	4.80	0.00	4.40	0.00	0.00	4.80
IA -	3.60	4.30	3.50	3.73	3.60	3.58	4.70	0.00	4.30	0.00
ID -	3.30	0.00	3.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00
IL -	2.85	2.90	3.70	2.80	3.83	2.85	0.00	3.38	3.60	2.45
IN -	3.12	0.00	2.83	0.00	2.10	2.98	0.00	2.57	0.00	3.30
KS -	3.25	0.00	3.20	3.70	3.25	3.35	0.00	2.80	0.00	0.00
KY -	0.00	0.00	3.00	0.00	2.90	3.20	0.00	0.00	0.00	0.00
LA -	3.84	3.93	3.07	3.65	3.66	3.77	0.00	3.84	0.00	4.20
MA -	3.80	2.70	0.00	0.00	3.75	3.65	2.70	3.80	3.50	0.00
MB -	3.80	0.00	3.80	0.00	0.00	0.00	0.00	0.00	0.00	3.80
MD -	4.00	5.00	3.55	0.00	3.60	3.60	0.00		5.00	3.75
ME -	0.00	0.00	0.00	0.00	0.00	3.30	0.00	3.30	0.00	0.00
a. MI-	2.00	0.00	2.95	2.00	3.40	2.73	0.00	3.00	2.00	2.90
.≝ MN -	4.15	4.39	4.25	4.00	4.37	3.85	0.00	3.92	4.00	4.32
Ouisine No W	2.60	2.53	3.12	2.00	2.72	3.05	2.00	0.00	2.00	0.00
1112	3.20	3.00	4.05	0.00	3.70	4.40	0.00	0.00	3.00	3.33
NC -	3.10	2.70	2.90	2.70	2.65	3.50	2.70	2.70	2.70	3.30
ND -	0.00	0.00	0.00	0.00	0.00	2.00	0.00	0.00	0.00	0.00
NE -	0.00	0.00	0.00	0.00	0.00	3.20	0.00	0.00	0.00	0.00
NJ -	2.58	2.59	2.70	2.39	2.55	2.67	3.40	2.45	2.50	2.54
NM -	2.50	0.00	2.70	0.00	2.70	0.00	0.00	2.50	0.00	0.00
NS -	0.00	4.70	4.70	0.00	0.00	0.00	0.00	0.00	4.70	0.00
NV -	2.00	2.80	0.00	3.40	2.44	2.35	0.00	0.00	0.00	1.80
NY -	3.48	3.18	3.60	2.43	2.73	3.22	2.00	2.77	2.00	4.14
OH -	2.98	3.20	3.10	2.70	3.11	2.89	0.00	2.83	2.40	3.22
OK -	0.00	2.33	0.00	3.00	2.33	0.00	0.00	0.00	0.00	0.00
ON -	4.33	4.28	4.33	4.35	4.21	4.22	4.29	4.36	4.42	4.39
OR -	2.77	2.50	0.00	0.00	2.50	3.15	0.00	2.50	3.90	0.00
PA -	2.92	2.16	2.96	3.00	2.70	2.87	0.00	2.61	2.70	2.80
QC -	4.44	4.54	4.38	4.55	4.30	4.30	4.52	4.20	4.48	4.40
RI -	0.00	2.90	0.00	0.00	0.00	3.20	0.00	0.00	0.00	2.90
SC - SK -	3.73	4.24 3.60	2.93	3.00	3.28	2.78 0.00	0.00 3.60	3.00	3.88 0.00	3.08
	3.85		3.98	3.60	3.60			3.60		3.60
TN -	3.25 3.45	3.88	3.71	3.37	3.56	3.13	2.70	3.41	3.47	4.70
TX -		2.92	2.76	2.10	2.92	3.80	2.30	3.32	1.73	0.00
UT - VA -	0.00 2.53	0.00 2.53	3.70 2.40	0.00 2.00	0.00 2.37	3.70 0.00	0.00 0.00	3.70 2.60	0.00 0.00	0.00 2.00
						2.77	0.00			
WA - WI -	3.35 2.70	2.47 3.40	4.00 2.87	2.67 2.88	3.50 3.28	3.04	3.75	4.30 2.85	1.90 3.77	2.08 2.77
WY -	0.00	3.60	0.00	0.00	3.60	3.60	0.00	0.00	0.00	0.00
WI -		-	-	-	-			1	- 1	-
	meat	soups	hamburgers	desserts	sandwich	fast food	snack	wing	vegan	shrimp

Challenges and Limitations

- Only 1,500 restaurants.
- Simplification of cuisine names may overlook nuanced distinctions.
- Certain states show no data for some cuisines due to data sparsity.

Conclusion

- Clear regional patterns in food preferences.
- High ratings for "Soup" and "Vegan" in some regions; consistent preference for "Hamburgers" and "Fast Food." Variability in restaurant quality across states.

***** Applications:

■ Insights can guide businesses in menu optimization and regional marketing strategies.

Future Work

- Expand dataset to include more restaurants.
- Incorporate additional variables such as price range and customer demographics.
- Use advanced machine learning models for deeper insights.

Thank You!

