

RESTAURANT RECOMMENDING CHATBOT

TEAM

MURALI KRISHNA

SWATHI GURIJILA

VARADHARAJAN

VIGNESH KUMAR

PRODUCT SUMMARY

To search for and recommend best restaurants in a city for different kinds of cuisines based on reviews given by customers



OBJECTIVES






- To predict rating of restaurants listed in the Yelp dataset based on the reviews given by the users. Classification techniques such logistic regression are used.
- Recommending restaurants to the users using the predicted stars and sentiment polarity values.
- Chatbot GUI which particularly takes two inputs from user or customers to predict 5 best restaurants available in a particular city for a particular cuisine provided by customer








DATASET OVERVIEW

- The dataset consists of a set of JSON files that include business information, reviews, tips (shorter reviews), user information and check-ins. Business objects list name, location, opening hours, category, average star rating, the number of reviews about the business
- Review objects list a star rating, the review text, the review date, and the number of votes that the review has received

DATASET DESCRIPTION

 yelp_academic_dataset_business.json	2021-01-28 2:06 PM	JSON File	121,466 KB
 yelp_academic_dataset_checkin.json	2021-01-28 2:11 PM	JSON File	388,938 KB
 yelp_academic_dataset_review.json	2021-01-28 2:29 PM	JSON File	6,774,100 KB
 yelp_academic_dataset_tip.json	2021-01-28 2:13 PM	JSON File	224,910 KB
 yelp_academic_dataset_user.json	2021-01-28 2:11 PM	JSON File	3,598,150 KB

MONGO DB

Collection Name ^	Documents	Avg. Document Size	Total Document Size	Num. Indexes	Total Index Size	Properties
checkin	138,876	2.8 KB	382.7 MB	1	1.1 MB	
rawdata	160,585	808.7 B	123.8 MB	1	2.2 MB	
reviews	8,635,403	827.4 B	6.7 GB	1	93.4 MB	
tips	1,162,119	223.1 B	247.2 MB	1	12.4 MB	
userdata	0	-	0.0 B	1	4.0 KB	

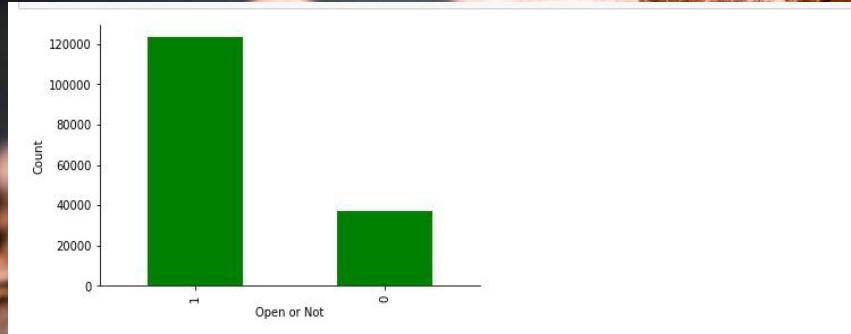
DATA COLLECTION - BUSINESS

	business_id	name	address	city	state	postal_code	latitude	longitude	stars	review_count	is_open	attrib
0	6iYb2HFDywm3zjuRg0shjw	Oskar Blues Taproom	921 Pearl St	Boulder	CO	80302	40.017544	-105.283348	4.0	86	1	{'RestaurantsTableServ 'True', 'WiFi':
1	tCbdrRPZA0oilYSmHG3J0w	Flying Elephants at PDX	7000 NE Airport Way	Portland	OR	97218	45.588906	-122.593331	4.0	126	1	{'RestaurantsTakeOut': 'T 'Restaurants,
2	bvN78flM8NLprQ1a1y5dRg	The Reclaimory	4720 Hawthorne Ave	Portland	OR	97214	45.511907	-122.613693	4.5	13	1	{'BusinessAcceptsCreditCa 'True', 'Rest
3	oaepsyvc0J17qwi8cfrOWg	Great Clips	2566 Enterprise Rd	Orange City	FL	32763	28.914482	-81.295979	3.0	8	1	{'RestaurantsPriceRange2' 'BusinessAcc
4	PE9uqAjdW0E4-8mjGl3wVA	Crossfit Terminus	1046 Memorial Dr SE	Atlanta	GA	30316	33.747027	-84.353424	4.0	14	1	{'GoodForKids': 'Fa 'BusinessParking'

DATA COLLECTION - REVIEW

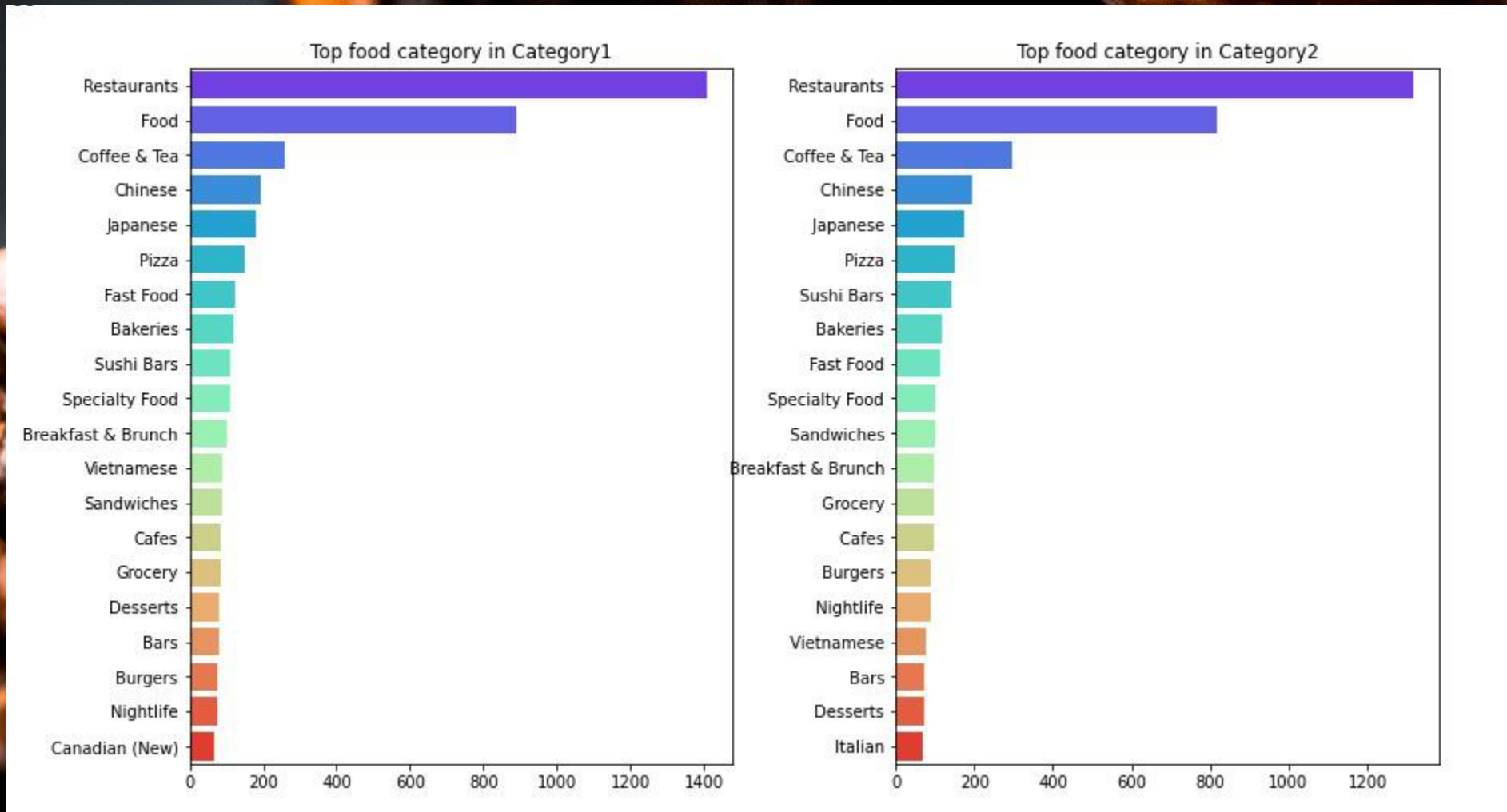
	review_id	user_id	business_id	stars	useful	funny	cool	text	date
0	IWC-xP3rd6obsecCYsGZRg	ak0TdVmGKo4pwqdJSTLwWw	buF9druCkbuXLX526sGELQ	4	3	1	1	Apparently Prides Osteria had a rough summer a...	2014-10-11 03:34:02
1	8bFej1QE5LXp4O05qjGqXA	YoVfDbnlSIW0f7abNQAClg	RA4V8pr014UyUbDvl-LW2A	4	1	0	0	This store is pretty good. Not as great as Wal...	2015-07-03 20:38:25
2	NDhkzczKjLshODbqDoNLSg	eC5evKn1TWDyHCyQAawguUw	_sS2LBIGNT5NQb6PD1Vtjw	5	0	0	0	I called WVM on the recommendation of a couple...	2013-05-28 20:38:06
3	T5fAqjjFooT4V0OeZyuk1w	SFQ1jcnGguO0LYWnbbftAA	0AzLzHfOJgL7ROwhdww2ew	2	1	1	1	I've stayed at many Marriott and Renaissance M...	2010-01-08 02:29:15
4	sjm_uUcQVxab_EeLCqsYLg	0kA0PAJ8QFMeveQWHFqz2A	8zehGz9jnxPqXtOc7KaJxA	4	0	0	0	The food is always great here. The service fro...	2011-07-28 18:05:01

EXPLORATORY DATA ANALYSIS



	category1	category2	category3	category4	category5	category6	category7	category8	category9
0	Gastropubs	Food	Beer Gardens	Restaurants	Bars	American (Traditional)	Beer Bar	Nightlife	Breweries
1	Salad	Soup	Sandwiches	Delis	Restaurants	Cafes	Vegetarian	None	None
2	Antiques	Fashion	Used	Vintage & Consignment	Shopping	Furniture Stores	Home & Garden	None	None
3	Beauty & Spas	Hair Salons	None	None	None	None	None	None	None
4	Gyms	Active Life	Interval Training Gyms	Fitness & Instruction	None	None	None	None	None
...
123243	Automotive	Tires	Towing	Auto Repair	None	None	None	None	None
123244	Automotive	Auto Detailing	None	None	None	None	None	None	None
123245	Real Estate	Real Estate Services	Home Services	Real Estate Agents	None	None	None	None	None
123246	Health Markets	Food	Specialty Food	Grocery	None	None	None	None	None
123247	Cuban	Sandwiches	Restaurants	Cafes	None	None	None	None	None
123248 rows x 10 columns									

EXPLORATORY DATA ANALYSIS



EXPLORATORY DATA ANALYSIS

Over all food categories:

- 'Food', ' Restaurants', 'Pizza', 'Mexican', 'American (Traditional)', 'American (New)', 'Italian', 'Indian', ' Pakistani', 'Thai', ' Japanese', 'French', ' Canadian (New)', ' Middle Eastern', 'German', 'Vietnamese', 'Chinese', 'Hungarian'

Cuisines:

- Indian
- Chinese
- Thai
- Italian
- Japanese

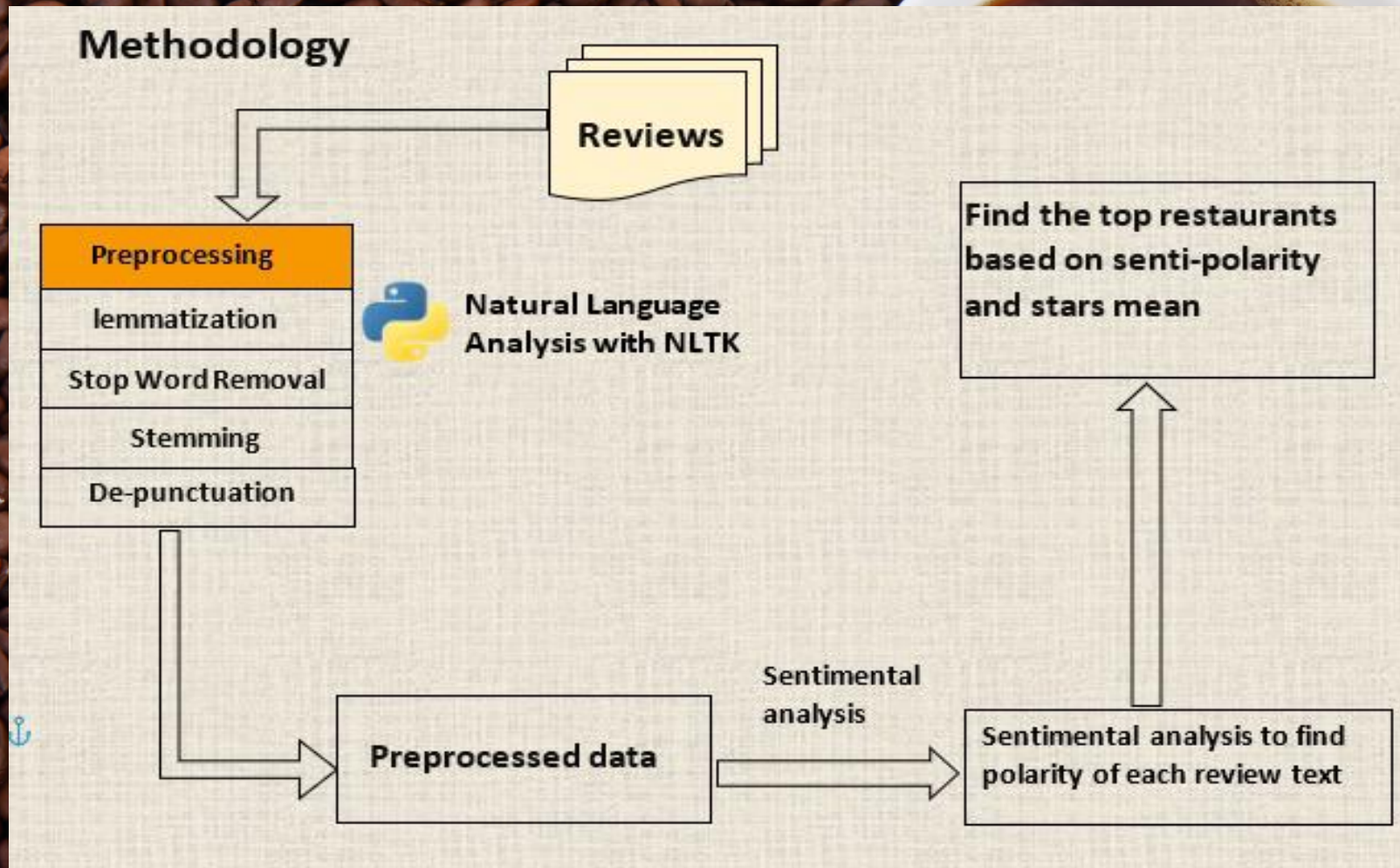
EXPLORATORY DATA ANALYSIS

Let us display the city names with their total review count and its average stars

```
restaurant_reviews = restaurant_ON[['city', 'review_count', 'stars']].groupby(['city']).  
restaurant_reviews.head(50)
```

	review_count	stars
city		
Vancouver	253711	3.597698
Richmond	34941	3.350968
Burnaby	25679	3.433434
North Vancouver	12572	3.594595
Coquitlam	8849	3.386447
New Westminster	8523	3.492424
Surrey	6433	3.370504
West Vancouver	3060	3.540816
Port Moody	2589	3.612676
Delta	2340	3.438889
Port Coquitlam	2137	3.642857
Bowen Island	300	3.928571
Richmond	87	3.500000
Ladner	64	2.833333
RICHMOND	50	3.000000
Steveston	47	4.000000
N Vancouver	34	3.500000
BURNABY	31	2.500000
Downtown Vancouver	31	4.000000
PORT COQUITLAM	24	1.500000
New Westminster	22	2.250000
Lions Bay	20	4.000000

METHODOLOGY



MODELLING - TRAINING THE DATASET

```
# Train all models
for name, model in models:
    now = datetime.now()
    print(f'{name} training started at {now.strftime("%H:%M:%S")}')
    model.fit(train_data, train_label)
    now = datetime.now()
    print(f'{name} training completed at {now.strftime("%H:%M:%S")}')
```

```
LR training started at 19:39:10
LR training completed at 19:40:59
KNN training started at 19:40:59
KNN training completed at 19:40:59
XGB training started at 19:40:59
XGB training completed at 19:44:19
NB training started at 19:44:19
NB training completed at 19:44:19
Decision Tree training started at 19:44:19
Decision Tree training completed at 19:48:32
Random Forest training started at 19:48:32
Random Forest training completed at 19:59:13
```


TRAINING SCORE

	Name	F1 Mean	F1 STD	Accuracy Mean	Accuracy STD
0	LR	0.782199	0.002640	0.799316	0.002195
3	NB	0.772098	0.003115	0.774238	0.003062
2	XGB	0.711712	0.003057	0.751095	0.002098
5	Random Forest	0.679467	0.002069	0.748741	0.001590
4	Decision Tree	0.669946	0.004161	0.674849	0.004558
1	KNN	0.613156	0.003652	0.657535	0.003702

TEST DATA EVALUATION

HYPERPARAMETER TUNING

```
=====
best params: {'solver': 'newton-cg', 'C': 0.1}
best score: 0.7886700289213612
=====
```

MODEL WITH TEST DATA

```
1 #Test data Evaluation
2 from sklearn.metrics import accuracy_score
3 from sklearn.metrics import f1_score
4
5 accuracy = accuracy_score(test_y,predicted_labels)
6 f1_score = f1_score(test_y,predicted_labels, average='weighted')
7
8 print(f'Accuracy for the test set is {accuracy}')
9 print(f'F1 score for the test set is {f1_score}')
```

```
Accuracy for the test set is 0.7341304823761221
F1 score for the test set is 0.7669307112489586
```


PREDICTION RESULTS FOR CHATBOT

	name	address	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	RestaurantsPriceRange2	spm
4658	dee	2013 yonge st	11:30-22:0	11:30-22:0	11:30-22:0	11:30-22:0	11:30-22:0	12:0-22:0	12:0-22:0	NaN	0.414553
2478	basil box	105 the pond road, unit r30	11:0-22:0	11:0-22:0	11:0-22:0	11:0-22:0	11:0-22:0	12:0-21:0	12:0-21:0	NaN	0.371071
2081	river tai restaurant	92 harbord street	NaN	NaN	NaN	NaN	NaN	NaN	NaN	\$	0.346048
4762	bach yen	738 gerrard street e	12:0-21:30	12:0-21:30	12:0-21:30	12:0-21:30	12:0-21:30	12:0-21:30	12:0-21:30	\$	0.337062
2722	silk restaurant & bar	446 parliament street	11:0-22:0	11:0-22:0	11:0-22:0	11:0-22:0	11:0-22:0	11:0-22:0	11:0-22:0	\$\$	0.333452
2685	thai mango	641 dupont street	11:0-23:0	11:0-23:0	11:0-23:0	11:0-23:0	11:0-23:0	11:0-23:0	11:0-23:0	NaN	0.329821
4155	saigon lotus	6 st andrew street	11:0-22:0	11:0-22:0	11:0-22:0	11:0-22:0	11:0-22:0	11:0-22:0	11:0-22:0	\$\$	0.323116
2359	sala modern thai kitchen & bar	1262 danforth avenue	11:0-22:0	16:30-22:0	16:30-22:0	16:30-22:0	16:30-22:30	12:0-22:30	12:0-22:0	\$\$	0.322934
3923	bolan thai cuisine	709 mount pleasant road	11:30-22:0	11:30-22:0	11:30-22:0	11:30-22:0	11:30-22:0	12:0-22:0	12:0-22:0	\$\$	0.311523
3499	kumo japanese restaurant	562 kipling avenue	11:0-22:0	11:0-22:0	11:0-22:0	11:0-22:0	11:0-23:0	11:0-23:0	12:0-22:0	\$\$	0.308854

AZURE CLOUD SERVICES

- Q&A Maker
- Knowledge base Chat Bot cognitive service.

Azure services

Create a resource User settings QnA makers Bot Services Virtual machines Quickstart Center App Services Storage accounts SQL databases More services

Recent resources

Name	Type	Last Viewed
yelpbot	App Service	8 hours ago
yelpbot	App Service plan	8 hours ago
yelpbot	Resource group	8 hours ago
yelpbot	QnA maker	8 hours ago
CAPSTONE	Virtual machine	3 weeks ago
capstone	Resource group	3 weeks ago

Name	Type	Last Viewed
restaurantQnA	QnA maker	3 hours ago
restaurantQnA	Resource group	3 hours ago
Azure subscription 1	Subscription	2 days ago

AZURE COGNITIVE Q&A BOT SERVICE- TRAINING DATA.

The screenshot displays the Azure Cognitive Q&A Bot Service training interface. At the top, a dark navigation bar contains the URL `https://www.yelp-support.com/FAQs?l...`, and buttons for **EDIT**, **PUBLISH**, **SETTINGS**, **Save and train**, and **← Test**. The **Save and train** and **← Test** buttons are highlighted with green boxes. Below the navigation bar, the interface is divided into two sections, each with a source header and a list of training data items.

Source: Editorial

- Who Created you ×
 - who owns you ×
 - + Add alternative phrasing
- I am created by GROUP3*
 - + Add follow-up prompt

Source: `https://www.yelp-support.com/FAQs?l=en_US`

- How does billing work? ×
 - + Add alternative phrasing

The training data items are displayed in a list. The first item, "Who Created you", has a sub-list of "who owns you" and "Add alternative phrasing". The second item, "I am created by GROUP3", has a sub-list of "Add follow-up prompt". The third item, "How does billing work?", has a sub-list of "Add alternative phrasing". The text "I am created by GROUP3" is italicized. The text "How does billing work?" is followed by a paragraph of text: "Depending on your program, you'll be billed on the first of the month or every 30 days from the date your Ads go live. Refer to your Billing section for your specific billing date. If you have a billing threshold, you may also be billed at that time. A billing threshold is an amount of advertising spend that triggers Yelp to charge your debit or credit card. You can see the billing date on previous statements by visiting the billing section of Yelp for Business. If you haven't been billed yet, the Ads page".

SAMPLE Q&A BOT

☐ Published KB ? [Start over](#)

package

primarily focused on supporting independent, local restaurant and nightlife businesses. Please select from the list below to see ways we can help you. [Restaurants and bars with Yelp Ads campaigns](#)[Restaurants and bars without active Yelp Ads campaigns](#)[Businesses beyond restaurants and bars that advertise on Yelp](#)[Frequently asked questions](#)

https://www.yelp-support.com/FAQs?l=en_US (Test) at 3:51 PM

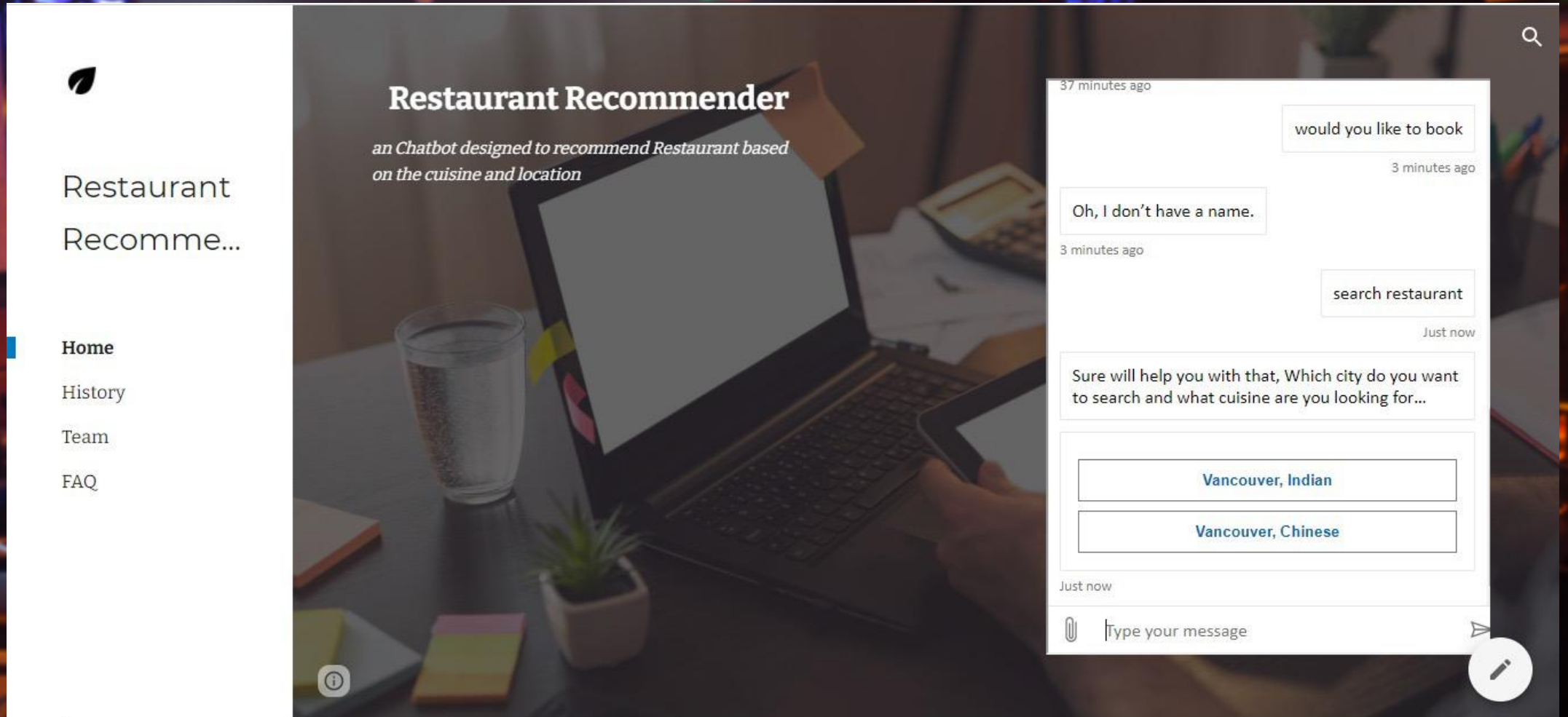
created

[Inspect](#) [You](#)

I am created by GROUP3

Type your message here ...

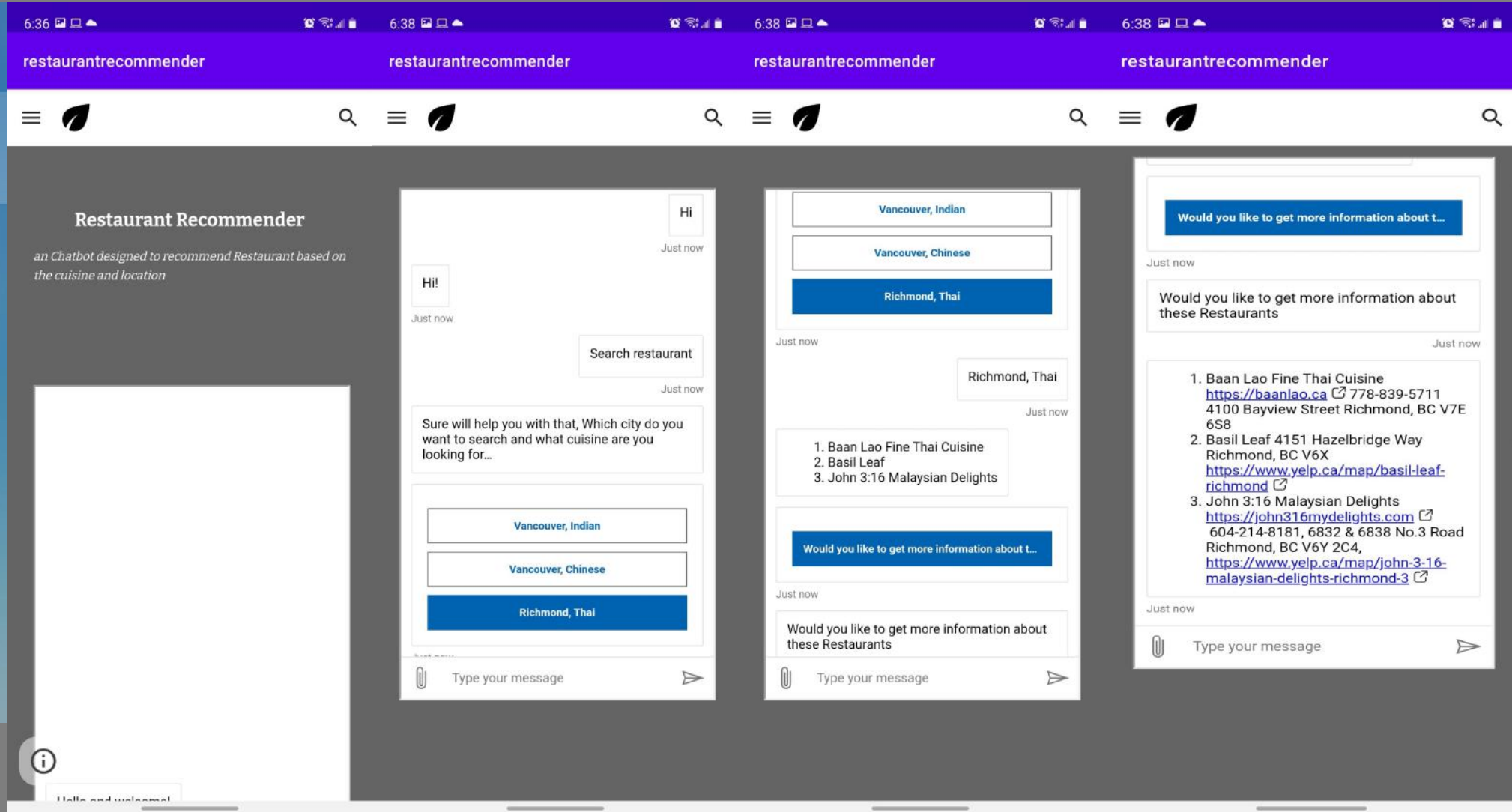
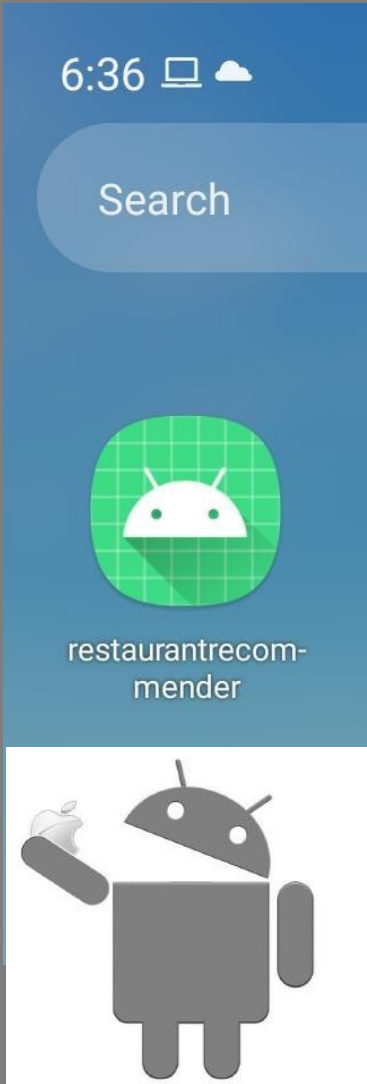
CHATBOT UI



ANDROID APP CREATION

A stylized graphic of an Android robot's skeleton, rendered in light blue lines against a dark teal background. The robot is centered, with its head at the top, torso in the middle, and legs at the bottom. The skeleton is composed of various bone-like shapes, including a brain-like structure in the head, a ribcage in the torso, and long, segmented limbs. The entire graphic is framed by a dark teal border that resembles a tablet screen. The text "ANDROID APP CREATION" is overlaid on the robot's chest area in a white, bold, sans-serif font.

ANDROID APP UI



PRODUCT DEMO



