

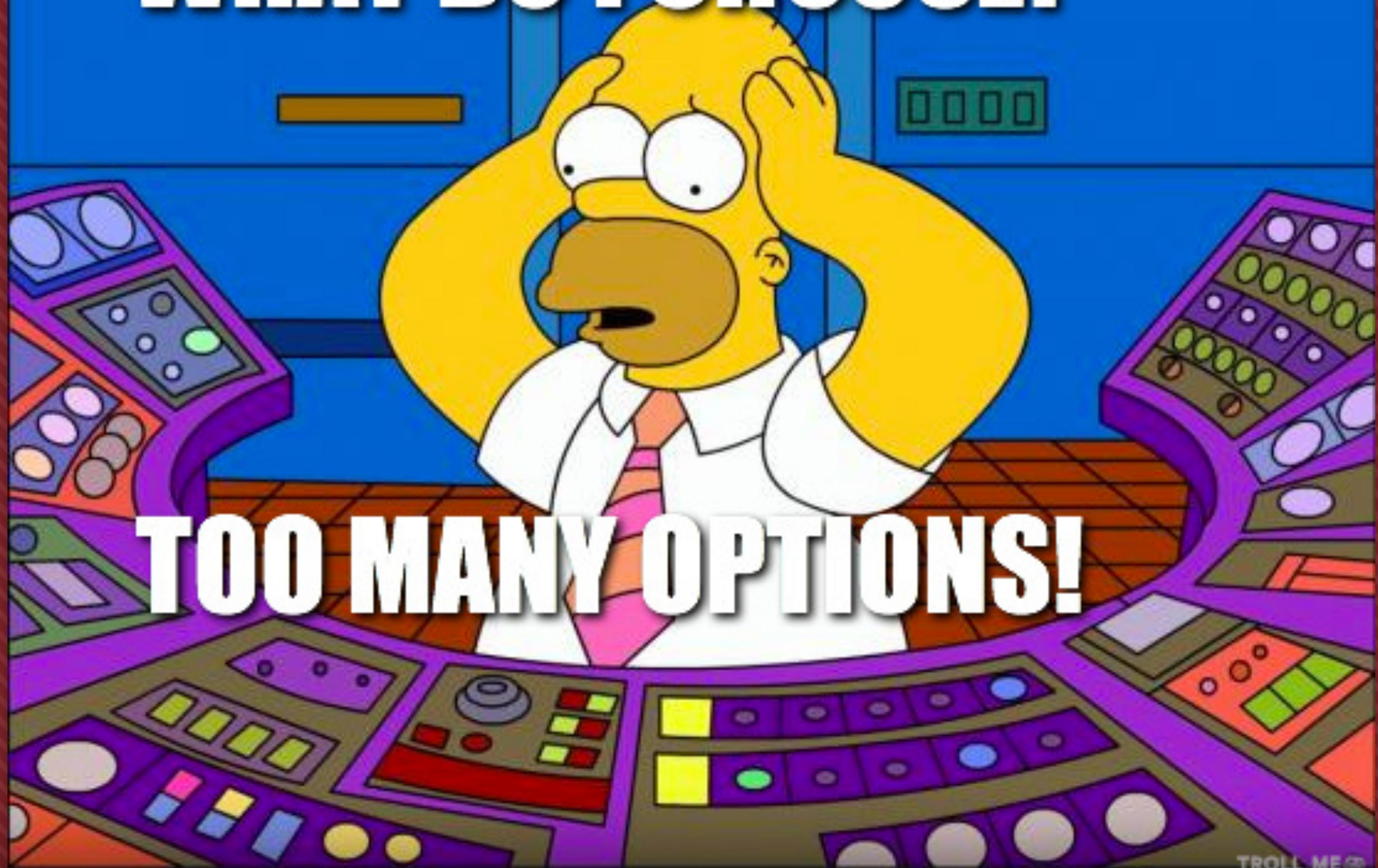


METIS PASSION PROJECT:  
**YELP RECOMMENDATION SYSTEM**

JENNIFER PHAN

# WHAT DO I CHOOSE?

## TOO MANY OPTIONS!



Objective

Data

Models

Flask Demo



When you finally decide where you want to eat dinner ....



It's been 84 years.

Objective

Data

Models

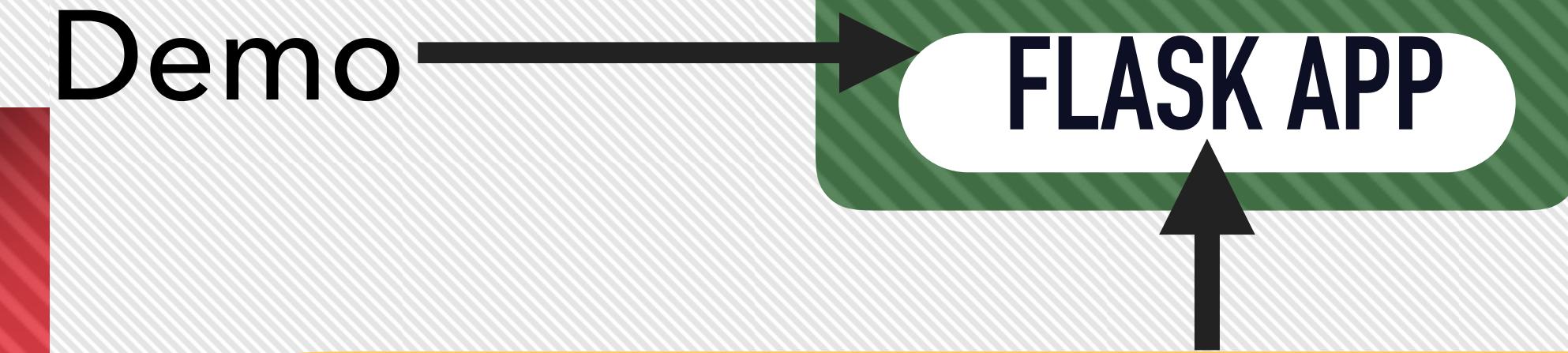
Flask Demo

yelp®

# AGENDA

Content with NLP  
and Collaborative  
Filtering

Data



PySpark Machine Learning & Spark SQL

Recommendation System

DISTANCE

CONTENT

COLLAB

SOCIAL

Preprocess Text, Prepare Model Inputs

Stored on AWS S3 Bucket

REVIEWS

USERS

BUSINESS

Objective

Data

Models

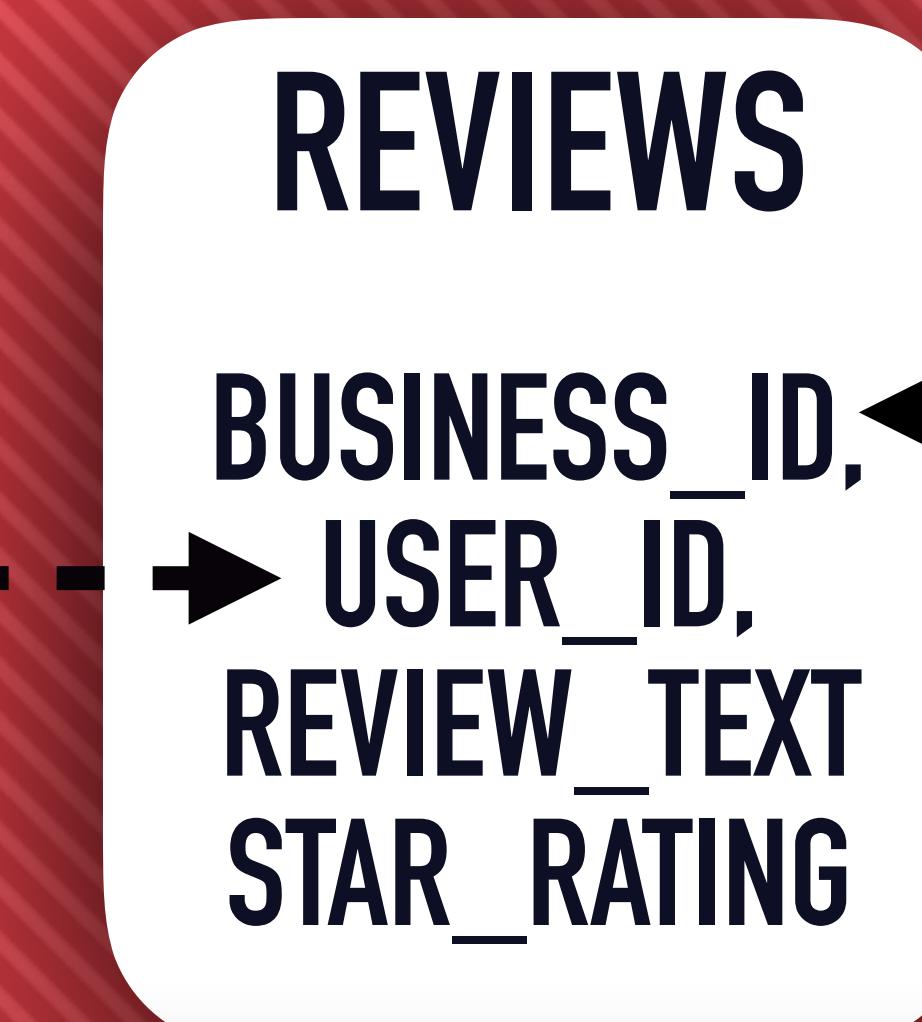
Flask Demo



# DATA SCHEMA

- ▶ 2016 Yelp Challenge Dataset
- ▶ All types of businesses worldwide

- ▶ Toronto
- ▶ 17,206 Businesses
- ▶ 93,106 Users
- ▶ 430,985 Reviews



Objective

Data

Models

Flask Demo



# DISTANCE FILTERING

- ▶ Haversine Formula
- ▶ Calculates the distance on a sphere

$\alpha$  Latitude

$\beta$  Longitude

$r$  Radius of Earth ~3,959 mi

$$d = 2 r \sin^{-1} \sqrt{\sin^2 \left( \frac{\alpha_2 - \alpha_1}{2} \right) + \cos(\alpha_1) \cos \alpha_2 \sin^2 \left( \frac{\beta_2 - \beta_1}{2} \right)}$$

Objective

Data

Models

Flask Demo



# CONTENT FILTERING WITH NLP



Objective

Data

Models

Flask Demo



# CONTENT FILTERING WITH NLP

## ▶ Cosine similarity

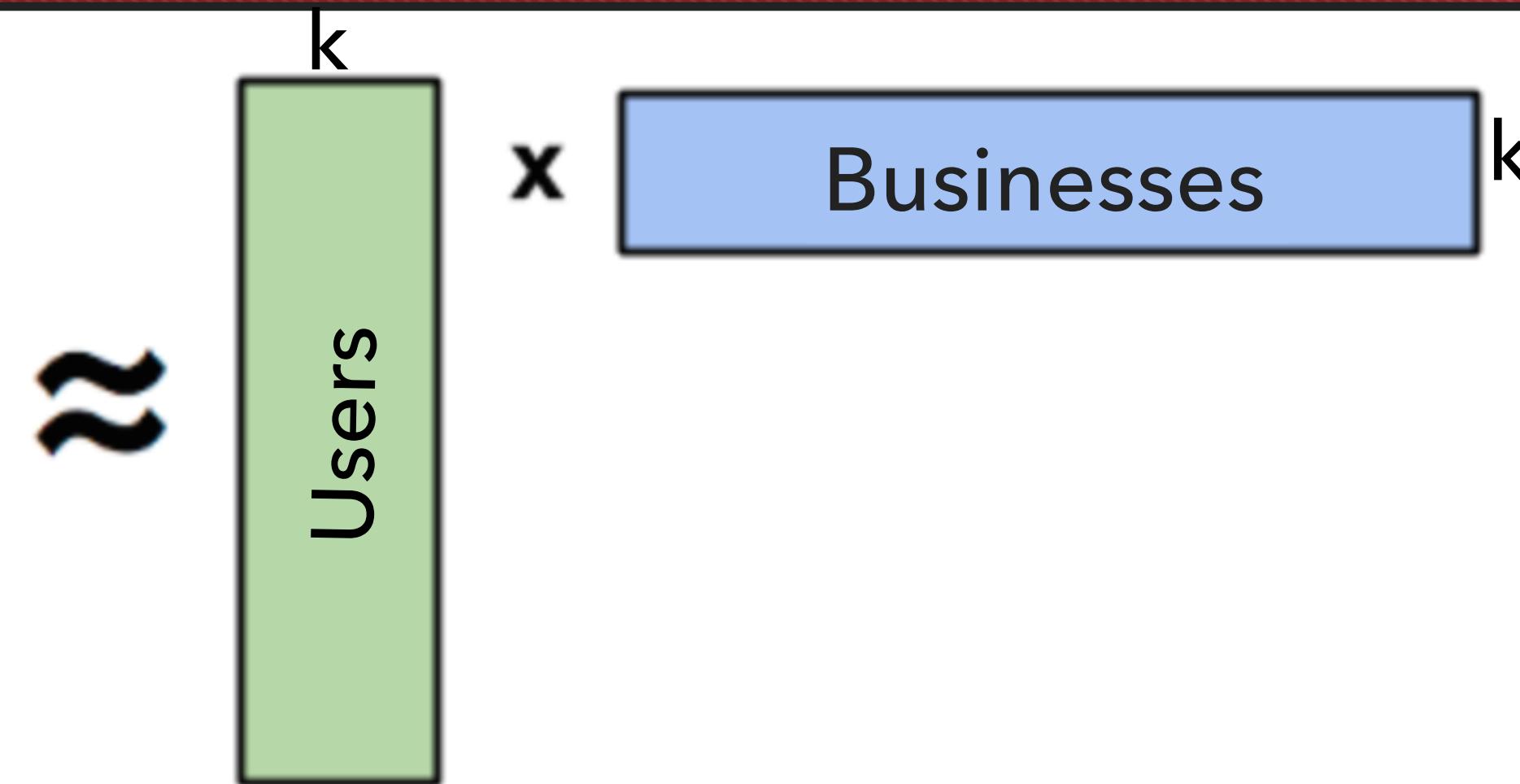
Recommend\_ByName("Viet House", Lat, Long, miles)

Name	Similarity	Stars	Lat	Long
Mi Mi Restaurant	0.96	4.0	43.66	-79.35
Que Ling Restaurant	0.96	4.0	43.66	-79.35
Pho Linh	0.95	4.0	43.65	-79.43

Recommend\_ByWord('sweet', Lat, Long, miles)

Name	Similarity	Stars	Lat	Long
Bakery Gateau	0.60	4.5	43.75	-79.35
The Mad Italian	0.56	4.5	43.67	-79.35
Brett's Ice Cream	0.56	4.5	43.67	-79.31

# COLLABORATIVE FILTERING



$$\min_{x,y} \sum_{\substack{r_{u,i} \text{ is known}}} (r_{u,i} - x_u^\top y_i)^2 + \lambda \left( \sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right)$$

- ▶ Uses existing user rating relationships
- ▶ Decompose to find latent features
- ▶ Alternating Least Squares can be easily parallelized
- ▶ RMSE = 1.64

Objective

Data

Models

Flask Demo



# SOCIAL NETWORK

- ▶ Recommend only what your friends have liked
- ▶ Lead to more trustworthy ratings

Name	Avg Stars	# Friends	Lat	Long
Northern Belle	5.0	4	43.70	-79.38
Byblos	5.0	4	43.66	-79.45
Lee Restaurant	5.0	3	43.65	-79.40
Cactus Club Cafe	5.0	3	43.66	-79.31
Uber	5.0	3	43.78	-79.38

Objective

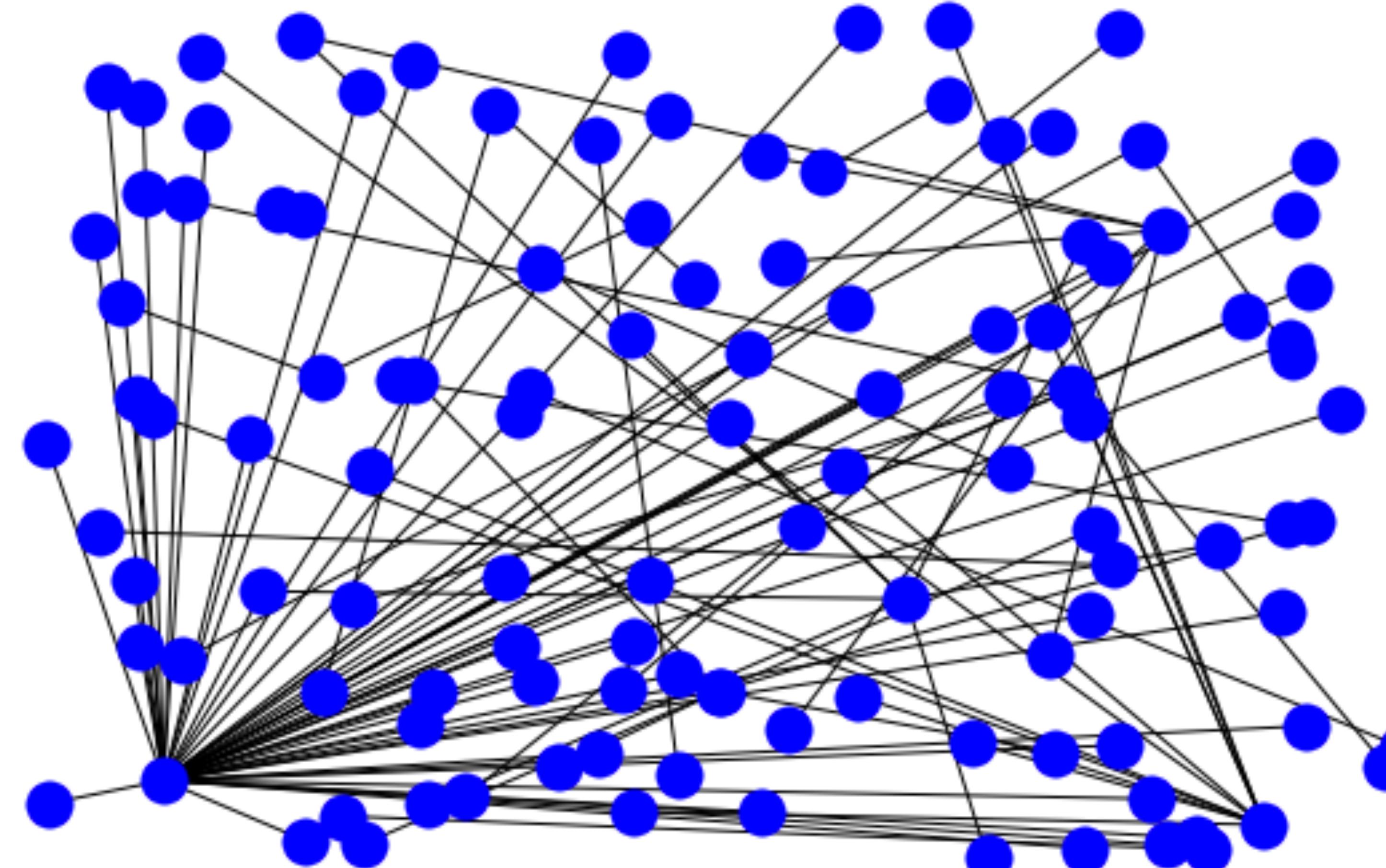
Data

Models

Flask Demo



# SOCIAL NETWORK



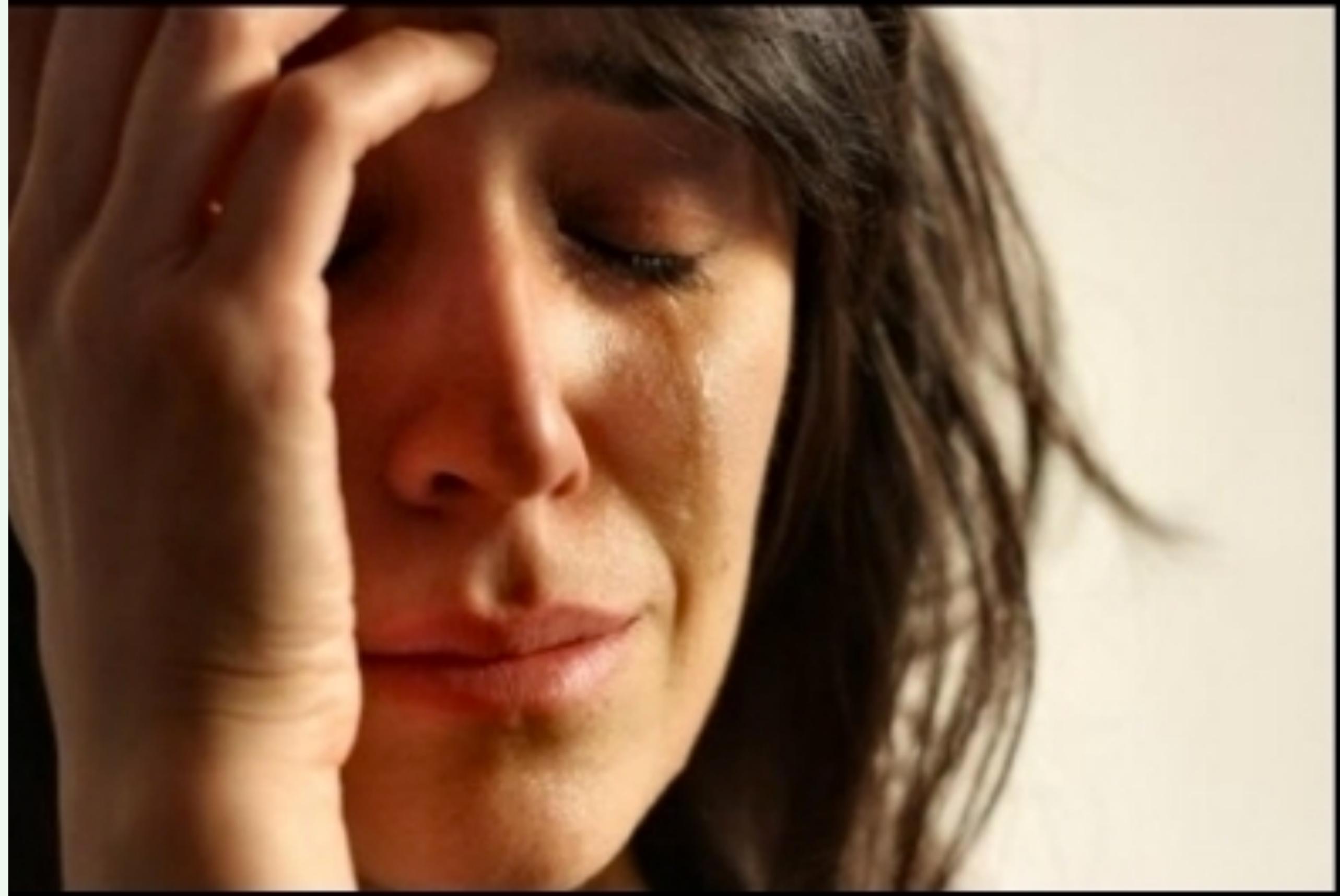
Objective

Data

Models

Flask Demo

**THERE ARE SO MANY  
RESTAURANTS**



**CAN'T DECIDE WHERE TO EAT**

Objective

Data

Models

Flask Demo

yelp

# Welcome to Yelp Recommender!

**Find me a similar business**

to:



**Suggest something relating**

to:



**My Yelp ID:**



**Select Category:**



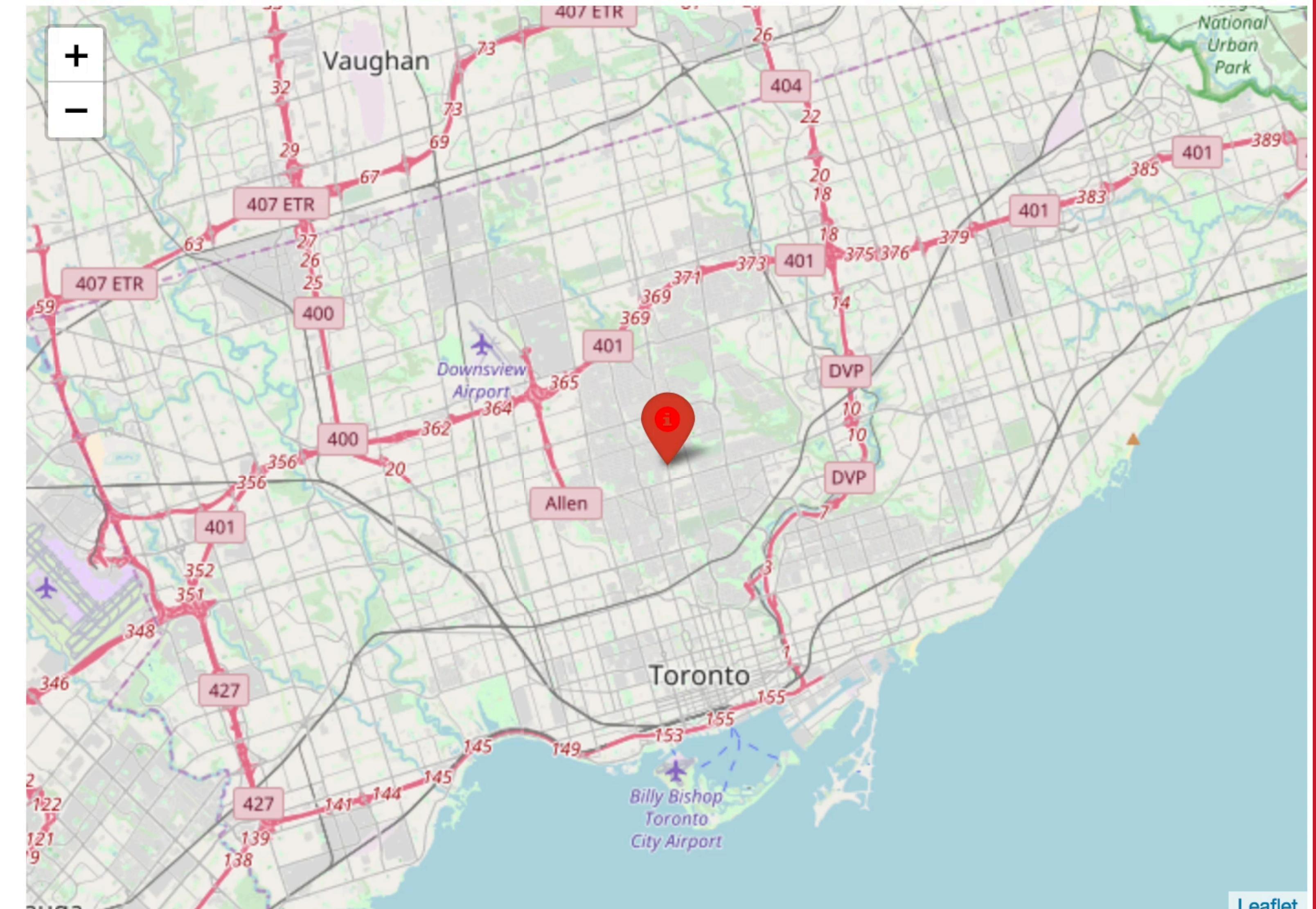
Choose how many miles you're willing to travel.



Number of miles: 5

Just show me places my friends have been.

I'm ready!



# Welcome to Yelp Recommender!

**Find me a similar business**

to:

**Suggest something relating**

to:

**My Yelp ID:**

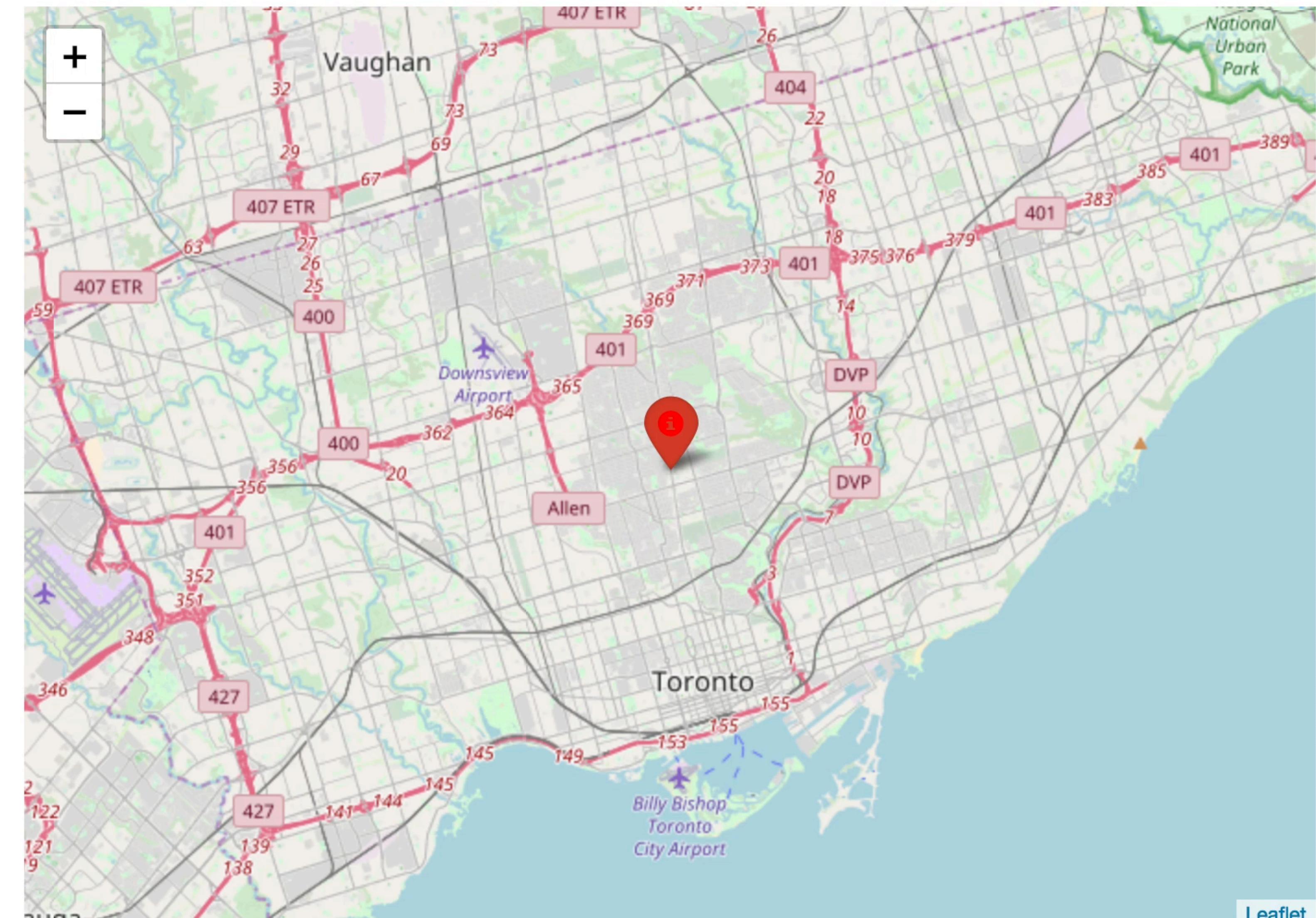
**Select Category:**

Choose how many miles you're willing to travel.

Number of miles: 5

Just show me places my friends have been.

I'm ready!



Objective

Data

Models

Flask Demo



# THANK YOU!



# JENNIFER-L-PHAN



# NEXT STEPS

- ▶ Create a filter to capture users who can best judge the authenticity of a restaurant - explore the origins of their last names
- ▶ Incorporate deep learning - analyze visual similarity of photo to recommend restaurants serving something similar to it

Objective

Data

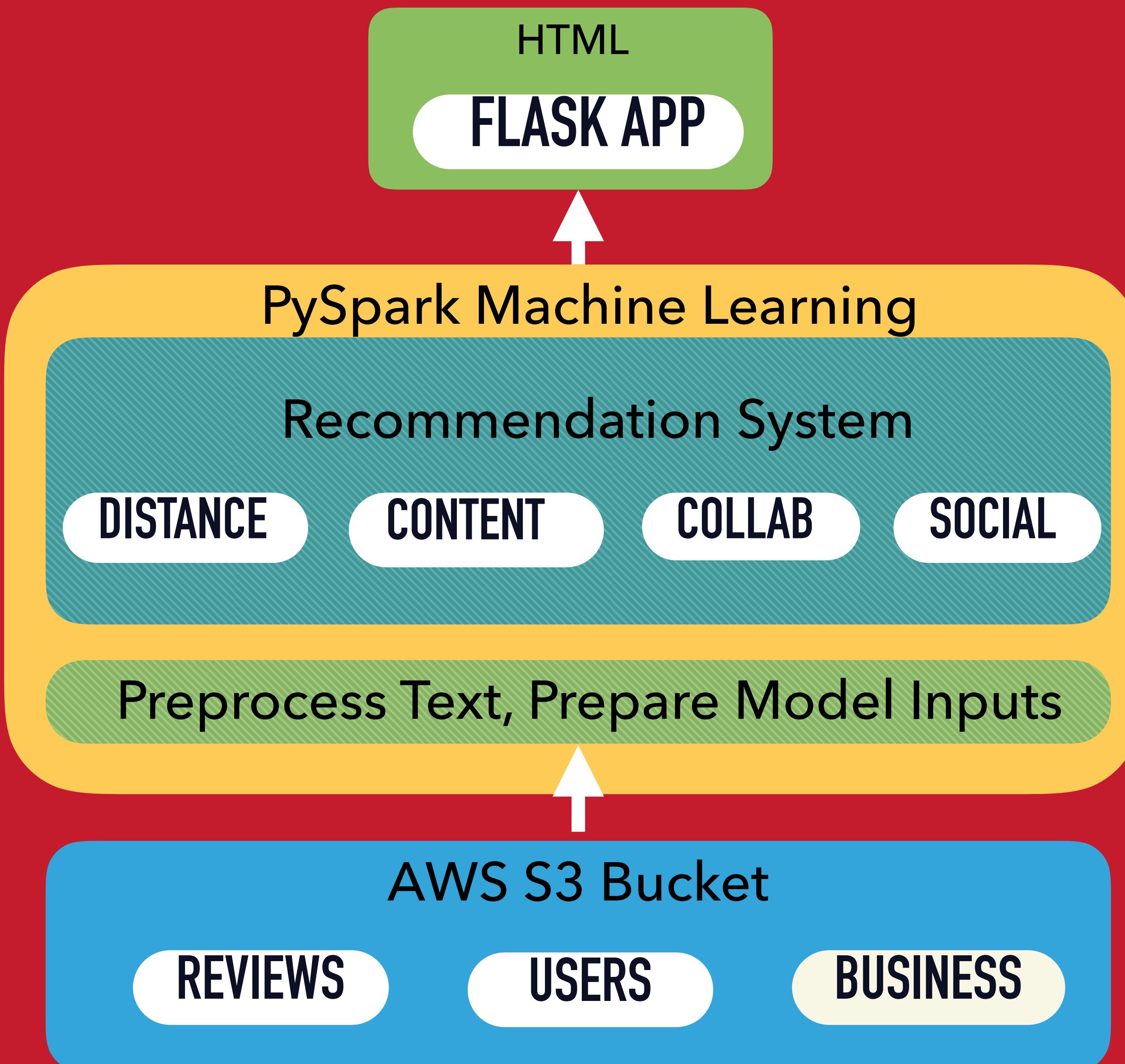
Models

Flask Demo



# 5 MINUTE AGENDA

- ▶ Data
- ▶ Recommendation Systems
  - ▶ 1. Content Filtering: Natural Language Processing
  - ▶ 2. Collaborative Filtering
  - ▶ Social Network
- ▶ Flask App Demo



When you're trying to  
decide where to eat



# COLLABORATIVE FILTERING

► Collaborative\_Filtering(userid,category)

UserID = 940681, category = "Shopping"

Name	Prediction	Stars	Lat	Long
Eye Studio	7.99	5.0	43.70	-79.38
Urban Gardener	7.95	4.5	43.66	-79.45
Authentic Couture	7.84	3.5	43.65	-79.40
Bead Heaven	7.75	5.0	43.66	-79.31
Tom's Gifts	7.63	5.0	43.78	-79.38

Objective

Data

Models

Flask Demo

