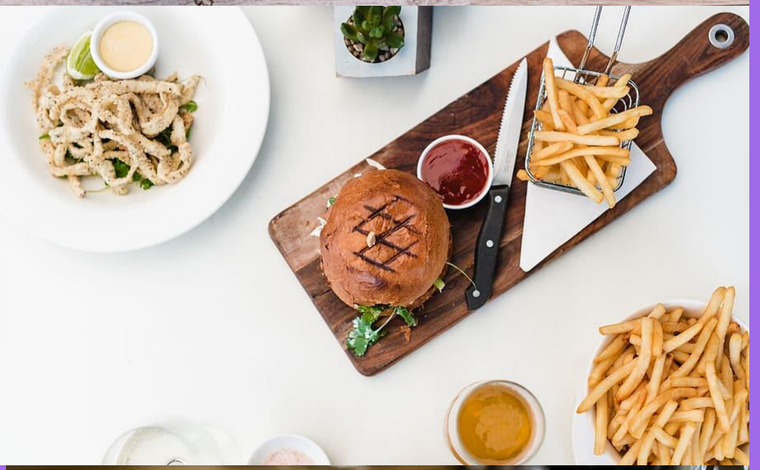
# Weekender: The SF Food Planner

# Overview

- There are over 4000 restaurants in the City of San Francisco and more than 15k in the larger Bay Area

- Picking the right place for where to eat when heading out with friends, family or even alone can be time-consuming and challenging with so many options available

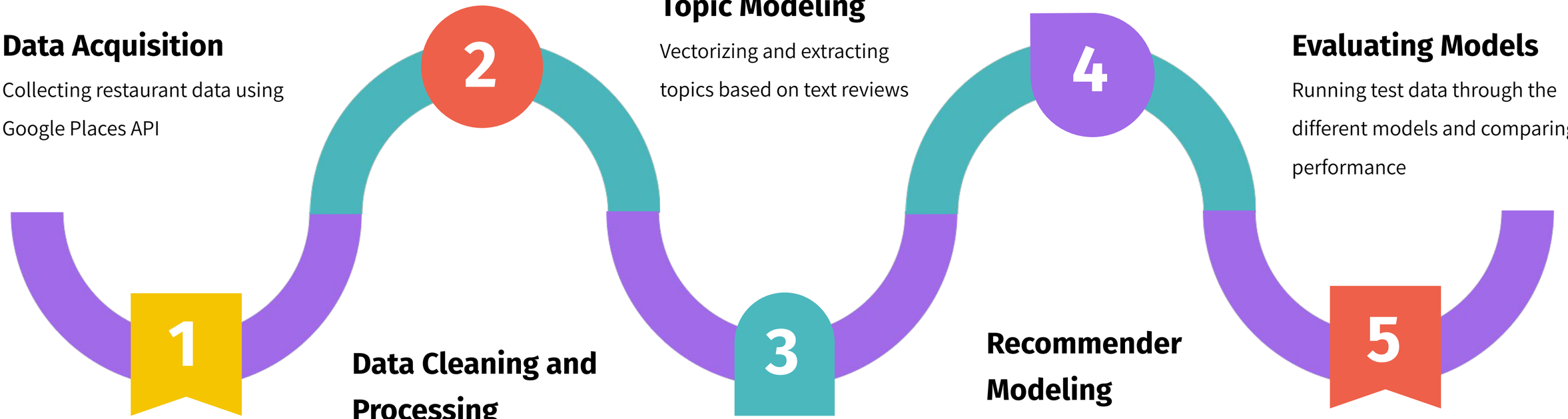
- **The Big Question:** Can we use the power of NLP to provide user-specific recommendations for restaurants in the SF Bay Area?



# Project Objective

Build a Recommender System using Google ratings and reviews to suggest restaurants around San Francisco Bay Area

# Methodology



### Data Acquisition

Collecting restaurant data using Google Places API

### Data Cleaning and Processing

Preparing the dataframes, cleaning text features and imputing missing values

### Vectorizing and Topic Modeling

Vectorizing and extracting topics based on text reviews

### Recommender Modeling

Elaborate on what you want to discuss.

### Evaluating Models

Running test data through the different models and comparing performance

# Data Acquisition

**No. of Rows/Docs:** 1301 entries
**No. of Reviews per Doc:** 5 text reviews
**No. of Columns/Features:** 10

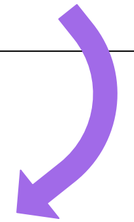
**Columns**

| | |
|---|---|
| 0 | Place |
| 1 | Address |
| 2 | Lat |
| 3 | Long |
| 4 | Business_Status |
| 5 | Dine-In |
| 6 | Price_Level |
| 7 | Rating |
| 8 | Total_Ratings |
| 9 | Reviews |
| 10 | URL |

- Queried using Google Places API
- Features are extracted from the JSON response
- Query provides first 20 responses based on Zip Code and Radius = 2000 meters

# Sample Row Entry

| Place | Address | Lat | Long | Business_ Status | Dine-In | Price_Level | Rating | Total_Ratings | Reviews | URL |
|---|---|---|---|---|---|---|---|---|---|---|
| La Mexicana | 3930 International Blvd, Oakland, CA 94601, USA | 37.774317 | -122.218381 | OPERATIONAL | TRUE | 1 | 4.5 | 171 | [{'author_name': 'Dianna Almonte', 'author_url... | https://maps.google.com/?cid=8520049234511722286 |



**5 Reviews per restaurant/entry**

{'author_name': 'Dianna Almonte',
'author_url': 'https://www.google.com/maps/contrib/102618579641291607894/reviews',
'language': 'en',
'original_language': 'en',
'profile_photo_url': 'https://lh3.googleusercontent.com/a-/ACNPEu-hOpLmi4trnVCJUQdXF1yk9xeKrfAGTZRAcQcgxA=s128-c0x00000000-cc-rp-mo',
'rating': 5,
'relative_time_description': '4 months ago',
'text': "Absolutely in love with this place! I've been here twice this month; Once with my family where we ordered a variety of items and I was amazed by how delicious every single plate was, I just had to come back! I recommend the Enchiladas, Menudo, and especially the quesadilla de maíz, they're amazing quality and a great choice for picky eaters as well.\n\nThe owner is incredibly nice and has been the reason my family has come here for years. This month was the first time I came here since i was really young, and I can truly see myself sticking to this spot for a taste of home. Definitely worth a try!",
'time': 1656228448,
'translated': False}

# Data Cleaning and Processing

## Cleaning

- Removing duplicate entries
- Converting Object type columns to numerical type “Float64”
- Removing entries with Business_Status != OPERATIONAL
- Imputing missing (NaN/None) values by using column aggregates for numerical features

## Aggregating Reviews

- Combining “text” values from all 5 reviews per entry into a new column “text_reviews”
- Using RegEx python package to clean the text:
  - Converting to lower-case
  - Removing spaces
  - Removing non-alphanumeric characters such as punctuations

## Lemmatization

- Using WordNetLemmatizer() to to pre-process and tokenize the review sentences into a list of root words
- Stop Words are also removed as part of tokeization

# Sample Final Review Data

## After Aggregation and Text Cleaning:

' this iconic eatery really lived up to its hype the team are really friendly and very attentive i felt well looked after and valued throughout the food from start to finish was amazing i wouldnt normally order fried figs to start or roast chicken for main but both were recommended and neither disappointed the chicken was the best id ever eaten i finished with a beautifully light berry sorbet for dessert which was the perfect palette cleanser and sweet treat to end on the cocktails were all outstanding i cant wait to go back went here for a brunch still many empty seat if you go slightly earlier didnt try the chicken because of an hour waiting time but the burger and pizza is good burger is juicy and flavorful pizza is thin and chewy the fries is very crispy but maybe its thin so soak more oil the cafe and cocktail is good for a brunch the cafe itself is nice large glass storefront allows sun bathing the table beautiful space with amazing natural light and great people watching oh  the food was great too caesar is a classic and lives ups while the tower of fries was also a hit  we learned something new that day  there are fries that can be eaten later as leftovers  slow paced restaurant perfect for our party and those who dont like to be rushed great food if a little bland for appetizers we had bread and butter tuscan melon and figs anchovies with celery and cheese all delicious appetizer choices for entrees we had a duck sugo and skirt steak served luke warm and a little too bland the duck easily could have worked better as pulled pork or lamb the wine bottle didnt make its way until after the appetizers had been served the service and staff were tops but our waiter was a novice or a little inexperienced the ambience was excellent and it more then made up for these small trifles we will definitely come back with friends for a birthday celebration or such enjoyed dessert better than i did with the entree  the dinning room exudes a warm wood fire feel twas cozy and invites intimate conversations  service overall was great  food here is good but technique was rather sloppy the sea bream needed less time the potato was slushy and needed more cream spinach over done unsure of the intention the steak was served as pieces it also seems that there was very little thought for presentation overall not impressed by the flavor  thankfully the eton mess and gateau victoire turned the experience around creamy with delicious fruit every bite had good balance of sweet and tart the gateau was not moist enough but the whipped cream was its redeeming quality  tldr dinner missed the mark but dessert service and atmosphere was well enjoyed'

## After Lemmatization and Tokenization:

['iconic',  'eatery', 'friendly', 'attentive', 'looked', 'valued', 'finish', 'amazing', 'chicken', 'recommended', 'disappointed', 'chicken', 'finished', 'beautifully', 'sorbet', 'dessert', 'perfect', 'palette', 'cleanser', 'cocktail', 'outstanding', 'brunch', 'slightly', 'earlier', 'chicken', 'waiting', 'burger', 'burger', 'flavorful', 'crispy', 'cocktail', 'brunch', 'storefront', 'bathing', 'beautiful', 'amazing', 'natural', 'watching', 'caesar', 'classic', 'learned', 'leftover', 'restaurant', 'perfect', 'rushed', 'appetizer', 'butter', 'tuscan', 'anchovy', 'celery', 'cheese', 'delicious', 'appetizer', 'choice', 'entree', 'served', 'easily', 'worked', 'pulled', 'bottle', 'appetizer', 'served', 'service', 'waiter', 'novice', 'inexperienced', 'ambience', 'excellent', 'trifle', 'definitely', 'friend', 'birthday', 'celebration', 'enjoyed', 'dessert', 'entree', 'dinning', 'exudes', 'invite', 'intimate', 'conversation', 'service', 'technique', 'sloppy', 'needed', 'potato', 'slushy', 'needed', 'spinach', 'unsure', 'intention', 'served', 'presentation', 'impressed', 'flavor', 'thankfully', 'gateau', 'victoire', 'turned', 'experience', 'creamy', 'delicious', 'balance', 'gateau', 'whipped', 'redeeming', 'quality', 'dinner', 'missed', 'dessert', 'service', 'atmosphere', 'enjoyed']

# Model 1: Content-Based Recommender using TF-IDF

- Each text_review is vectorized and transformed into a numerical **TF-IDF Matrix**

- **What is TF-IDF?**

  Tf-idf is designed to down weight the frequently occurring words in the feature vectors.

  It can be defined as the product of the term frequency (frequency of one word in a given document) and the inverse document frequency (occurrence of this word among all the documents).

  **Term frequency / Document frequency measures the relevance of a term in a given document.**

# Model 2: Content-Based Recommender using spaCy

- **spaCy** is an open-source NLP library in Python with pre-trained models and pipelines for vectorization and tokenization of string inputs
- Feature vector created after applying spaCy model: "en_core_web_sm”
- Original Text: this iconic eatery really lived up to its hyp...
  Spacy Conversion: ( , this, iconic, eatery, really, lived, up, t...

# Model 3: Topic Modeling using LDA

- **Latent Dirichlet Allocation Model** is used to convert text reviews into Topic features
- CountVectorizer() is used to create a Dictionary of all the words in text_reviews column
- The model is optimized to create 10 Topics from the unstructured text reviews
- Topic probabilities are added as feature columns to the dataframe

# Topics Generated

Topic 0


Topic 1


Topic 2


Topic 3


Topic 4


Topic 5


Topic 6


Topic 7


Topic 8

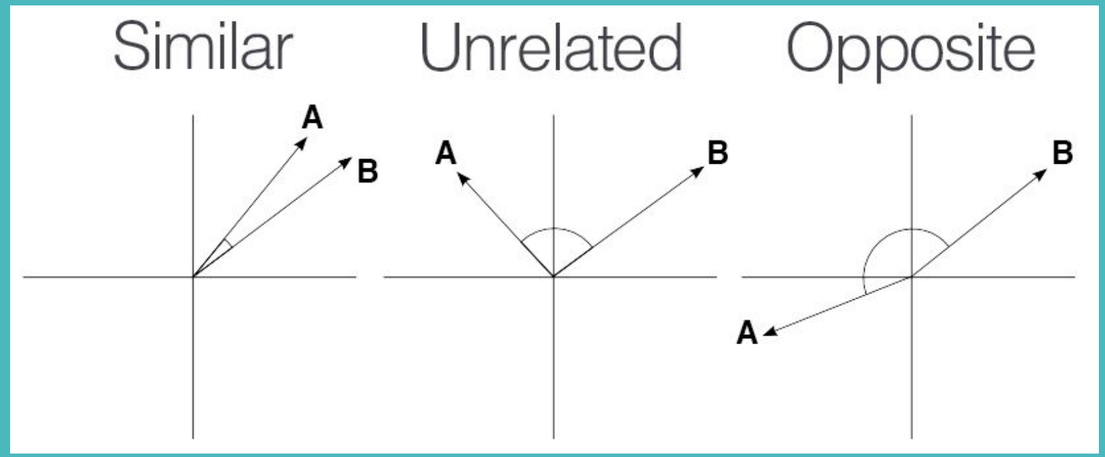

Topic 9

# Evaluating Recommender Models

Comparison Metric Used:

## Cosine Similarity

- Mathematical measurement of how similar the documents are
- Measures the cosine of the angle between the two vectors in a multi-dimensional space



# Model Recommendations

## Test “Liked” Restaurant

| | |
|---|---|
| **Name** | Lers Ros Thai |
| **Price_Level** | 2 |
| **Rating** | 4.2 |
| **Total_Ratings** | 683 |

## TF-IDF Recommender

| Place | Price_Level | Rating | Total_Ratings |
|---|---|---|---|
| Bella Saratoga | 3.0 | 4.3 | 475.0 |
| Servino Ristorante | 2.0 | 4.1 | 185.0 |
| New Delhi Indian Restaurant | 2.0 | 3.7 | 670.0 |
| Sam's Anchor Cafe | 2.0 | 4.3 | 1146.0 |
| Absinthe Brasserie & Bar | 3.0 | 4.4 | 1553.0 |

## SpaCy Recommender

| Place | Price_Level | Rating | Total_Ratings |
|---|---|---|---|
| Sipan Peruvian Restaurant & Bar | 3.0 | 4.2 | 494.0 |
| Fino | 2.0 | 4.6 | 597.0 |
| La Fondue | 4.0 | 4.4 | 868.0 |
| Plumed Horse | 4.0 | 4.4 | 674.0 |
| John's Grill | 3.0 | 4.3 | 3006.0 |

## LDA Recommender

| Place | Price_Level | Rating | Total_Ratings |
|---|---|---|---|
| La Boulangerie de San Francisco, Hayes | 2.0 | 4.2 | 725.0 |
| Marin Joe's Restaurant | 2.0 | 4.4 | 731.0 |
| New Delhi Indian Restaurant | 2.0 | 3.7 | 670.0 |
| Starbucks | 2.0 | 4.2 | 576.0 |
| Basil Thai Restaurant & Bar | 2.0 | 4.3 | 644.0 |

# Future Improvements

## Improve Data Retrieval Process

- Reduce duplicate response pulls due to Zip code and Radius overlap
- Reduce bias in the response data and add randomization
- Increase number of reviews queried per Place

## More Feature Engineering

- Factor in number of ratings for Aggregate rations using averaging techniques (Bayesian)
- Improve Text Tokenization: adding more common words to Stop Words dictionary (such as: food)

## Add More Features

Add  features such as Cuisine, Place Summary/Description

## Hybrid Collaborative Filtering

Add multiple user data to add collaborative-filtering to the Recommender

## Expand the Recommender

Add other places such as events/activities to recommender to create a holistic Weekend Planner

## Create a Web App

Create a user-friendly Web Application for public access