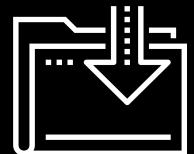




# Data Analytics & Visualization Boot Camp

Data Boot Camp  
Online—Welcome Session



**WELCOME!**



# Virtual Class Best Practices

# Best Practice 1: Always Mute

Please keep your microphone muted, unless you are called on or your host asks for feedback/communication.



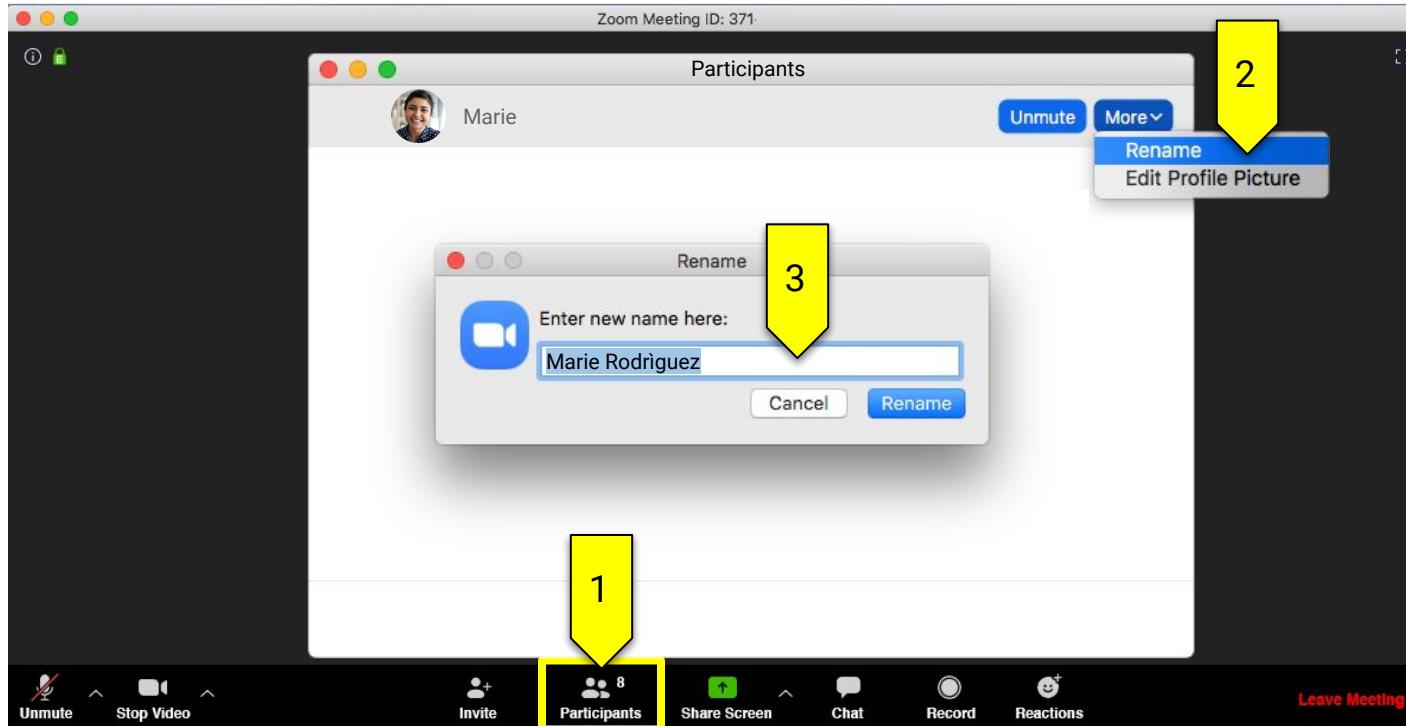
# Best Practice 2: Turn Your Video On

Kindly turn your video on so we can see you.



# Best Practice 3: Update Your Name

Please update your name under the Zoom *Participants* tab to your full name.



## Best Practice 3: Actively participate

We ask that you actively participate in the session by answering questions and being a part of the conversation in your class's Slack Workspace or Zoom chat.



# Orientation Overview

# What Will I learn?

---

By the end of this session, you will:



Get to know your instructional and support staff members



See how to successfully navigate your boot camp portal



See how virtual classes work and successfully complete this first session



Configure multiple technologies that will play a vital role in your toolkit of success



Be aware of the career services available to you and how to access them



Understand the minimum requirements in order to successfully graduate from this boot camp



*Enriching Lives. Transforming Careers*  
University of Toronto, School of Continuing Studies

# Kyla Dhanraj

## *Student Success Manager*

---

Welcome! I am here to assist you while you complete this program. I want to make your journey as easy as possible and help you reach your goal of completing the program. I cannot wait to get to know you all and connect with you.

In my free time I spend my time with my dog Logan, my friends and family. I love watching hockey and volunteering with Kids Help Phone.

You can contact me at: [kdhanraj@bootcampspot.com](mailto:kdhanraj@bootcampspot.com) or LSM-Kyla on slack!



# **Anthony Inthavong**

## **Teaching Assistant**

---

Professional and educational background  
Consultant at Jarvis Consultant

Education  
Graduate from the University of Toronto (UofT)  
Specialization in  
Machine Learning and Data Science (ML & DS)

General teaching style  
Patient, take your time

Hobbies and interests  
Netflix and League of Legends



# Laurel Lobo

## Teaching Assistant

---

Laurel is currently working as a Data Engineer, Ecommerce Analytics at the Home Depot. His main tasks include building pipelines between multiple data-sources, solving data quality issues, building API data layers, and automating traditional analytics reporting.

In 2018, after getting his feet wet doing some data work in Medical Safety, Laurel enrolled in the first cohort of the Data Analytics Bootcamp at the University of Toronto. For his final project, He ‘borrowed’ all the wine listings from the LCBO, and created a website that classified each wine based on a handful of parameters which he extracted using a light Natural Language Processing algorithm.

Laurel enjoyed his time at the bootcamp so much, he’s taking additional learning in order to do a Masters in Computer Science in the future.



# **Thais Bittencourt**

## **Assistant Instructor**

---

### Background

YorkU, BA, Psychology  
U of T, MI, Information Systems Design | Analytics

TELUS Digital, Sr. Strategy Manager & Analytics Lead  
RoseRocket, Sr. Analyst  
LoyaltyOne, Sr. Analyst, Digital Analytics

### General Teaching style

Facilitator - I'd rather facilitate learning by helping students troubleshoot and find the answers

### Hobbies and interests

Learning, reading, snowboarding, board games



## **YASH SHAH**

### **Teaching Assistant**

---

I work full-time with RBC as a SRE and also worked there as a Data Analyst. I have pursued my Master in Applied Computing from University of Windsor

I like to know about the student first before suggesting the approach to a particular problem. I would encourage all your questions and always appreciate curiosity

I like to read, write, and run. I have a blog channel where I write my thoughts.  
Fun Fact: I came in a national TV episode once



## **Sunil Khambaita**

### **Teaching Assistant**

---

Sunil is a Full Stack Engineer at CBC. He graduated from University of Toronto 2 years ago, with an honors in Computer Science and a major in Statistics.

Sunil has a love for data, especially in the backend and databases. His main interests are in architecting software solutions for products with high throughput and low latency requirements.

Outside of work, Sunil is a big soccer fan and loves to play physically and virtually.



## **Sal Hobbi**

### **Instructor**

---

Professional and educational background:  
Decades of industry experience in Software  
Development, delivery, support & training..

MS in Computer Information Systems & BS  
in Computer Science.

General teaching style:  
Interactive. Collaborative. Hands-on. Caring.  
What's in it for me. **Lots of fun.**

Hobbies and interests:  
Tennis, Hiking, BBQ, Programmable  
technologies...etc.



# Getting to Know Your Cohort

---

To complete this activity, you need to perform the following steps:

1. Head to the slack channel #first\_night\_introductions
2. **Break the ice!**
3. Please share with us ( you will see an example posted by LSM-Kyla):

**Your Name**

**Where you Live**

**Your Background**

**One Fun Fact About Yourself**

It is highly recommended that you find a “study buddy” that you can connect with for partnership during this journey! This is a great opportunity to find someone with common interests, or lives close to you. By the end of this Session, you can connect through Chat or over Slack, and introduce yourself!

**Suggested Time:** 10 Minutes



# Boot Camp Experience

# What does “for working professionals” mean?

## What it means:

- 2 required class, instead of 3 required classes
- More optional than required activities
- Support from instructional staff on weekends and evenings
- Curriculum built around common real-world scenarios



## What it DOESN'T mean:

- You can commit less time than our in-person offering
- You'll need less discipline to complete the course
- You can restart the course again at a later date
- You won't be challenged!



# Life in the Boot Camp

---

01

## Online Lessons

Every week will present new lessons to install and learn new tools, all **anchored on a real-world project.**

02

## Live Support

We'll come together weekly to cement the concepts from your self-paced learning.

Structured **optional** classes, office hours, and study halls provide **additional support** throughout the week.

03

## Weekly Challenge

Each week, you'll submit a challenge assignment around the module's technologies.

The last 5 weeks are built around a **comprehensive team project** that integrates all the skills of the course and provides a valuable work product.



We will teach you how to  
fish, but we won't *give* you  
the fish.

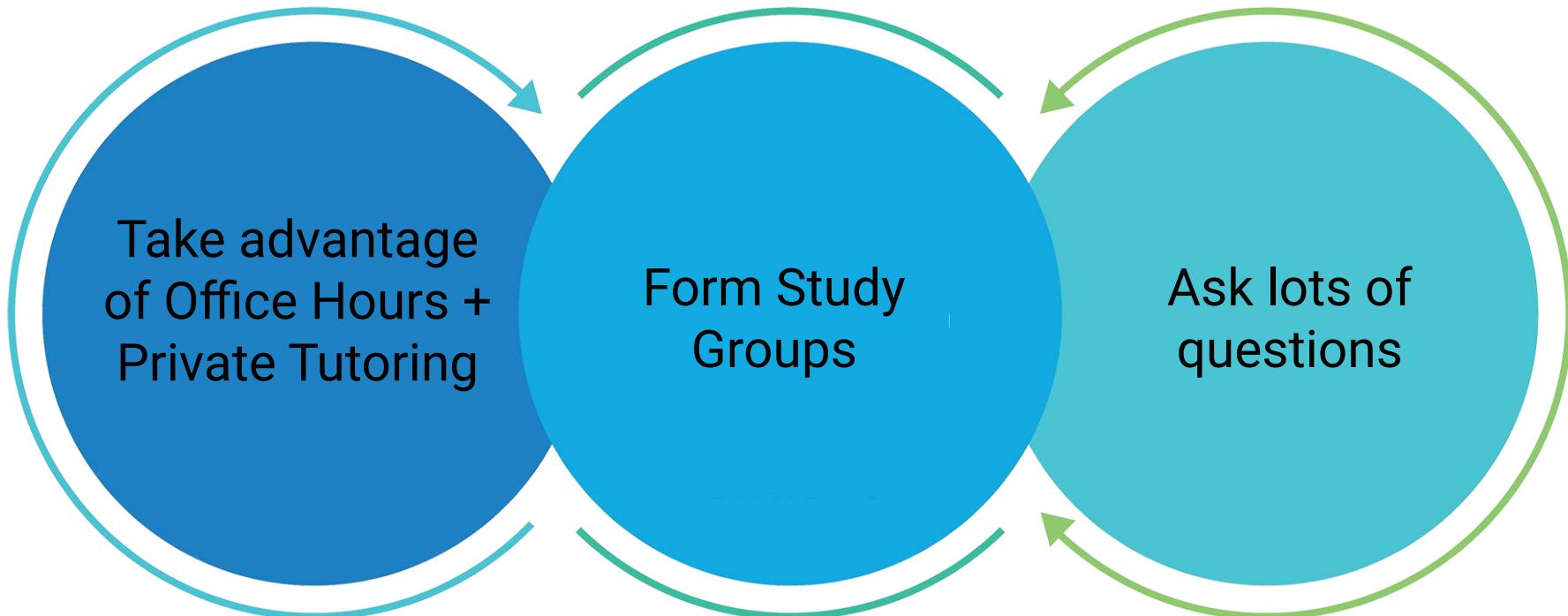
# Tips for Success

Use a calendar (digital or physical) to plan out your weeks

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
7:00 AM							
7:30 AM							
8:00 AM							
8:30 AM							
9:00 AM							
9:30 AM							
10:00 AM							
10:30 AM							
11:00 AM							
11:30 AM							
Noon							
12:30 PM							
1:00 PM							
1:30 PM							
2:00 PM							
2:30 PM							
3:00 PM							
3:30 PM							
4:00 PM							
4:30 PM							
5:00 PM							
5:30 PM							
6:00 PM							
6:30 PM	Office Hours		Office Hours				
7:00 PM							
7:30 PM	Instructor-led virtual class	Homework (2 hours)	Instructor-led virtual class	Homework (3 hours)	Homework (2 hours)	Homework (2 hours)	Homework (2 hours)
8:00 PM							
8:30 PM							
9:00 PM							
9:30 PM							

# Tips for Success

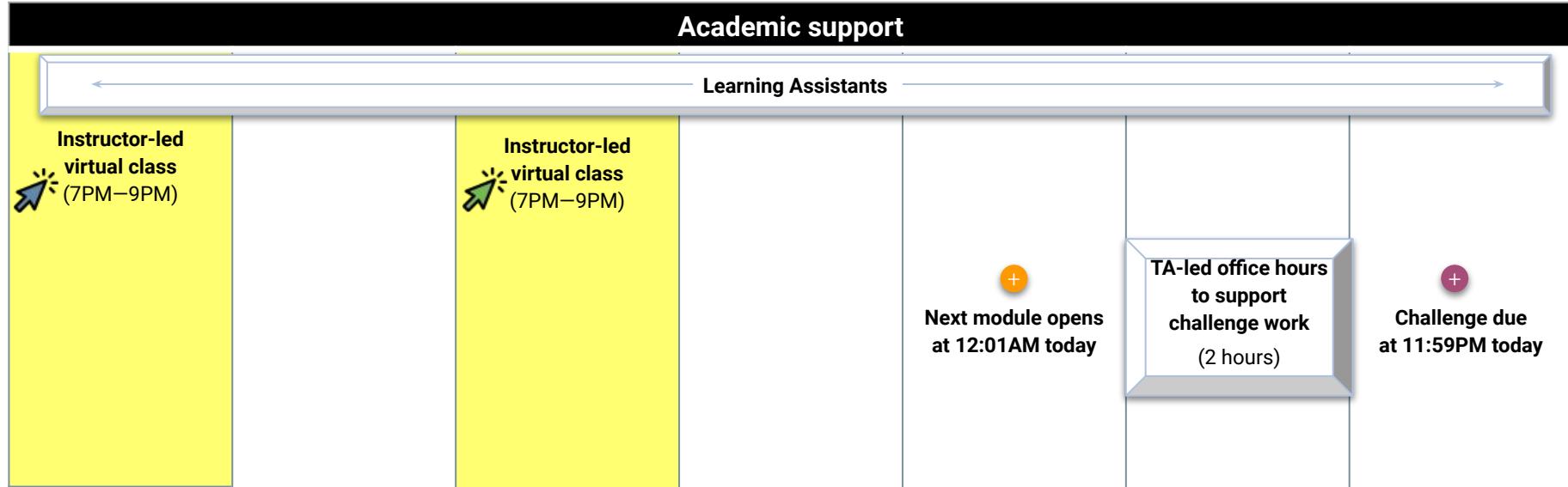
---



# Weekly Calendar [KEEP FOR MW]



## Academic support



# Minimum Graduation Requirements

---

These are the minimum graduation requirements.



Miss no more than 8 Required Virtual Classes



Have no more than 2 incomplete Challenge assignments



Participate and present in all class projects



Tuition Paid in Full

# Holiday Accommodations

---

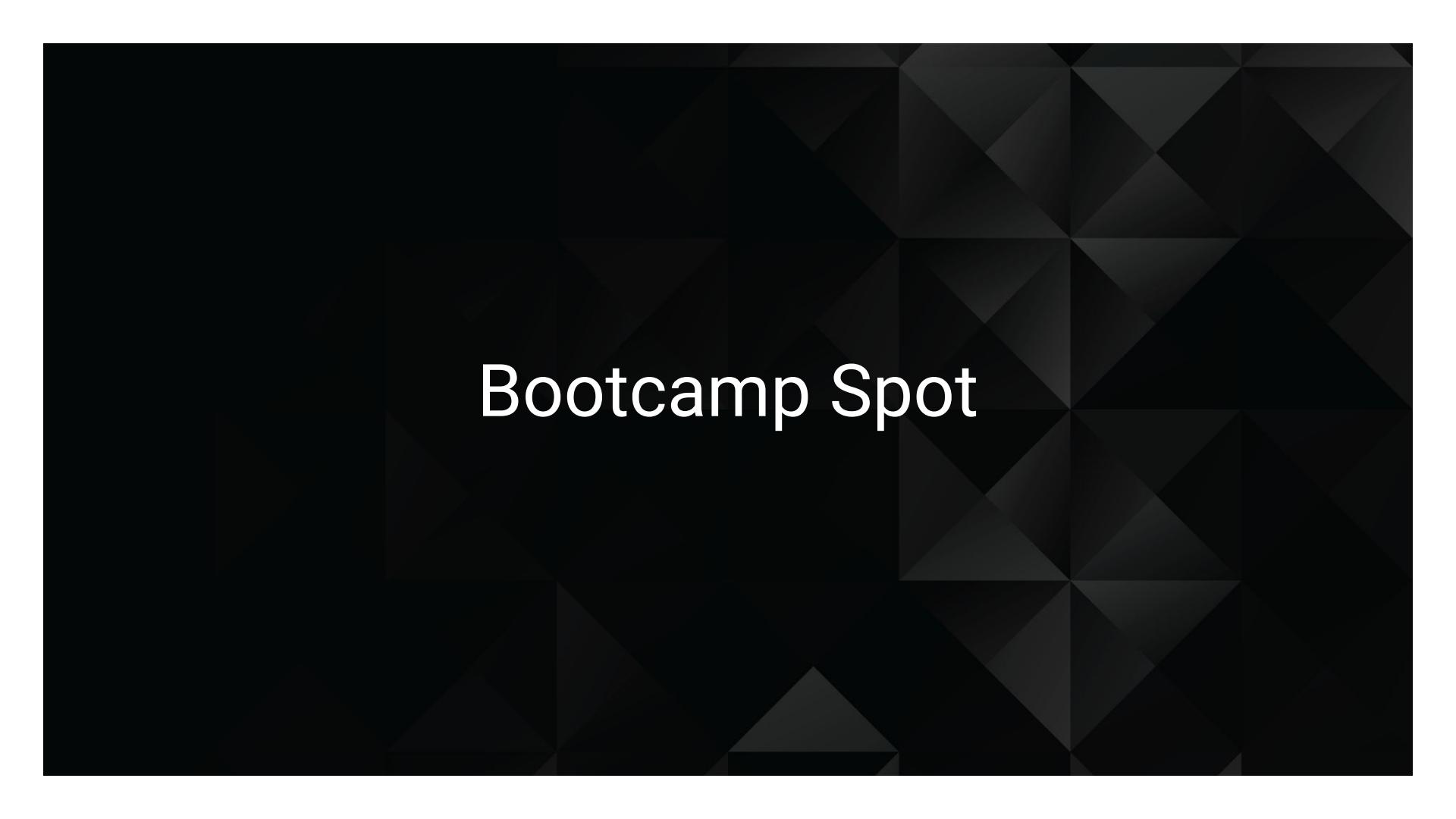
Our Boot Camp recognizes the following Holidays:

- Labour Day
- Thanksgiving

Any classes scheduled for these days will be adjusted appropriately in the course calendar. On Labour Day and Thanksgiving class will be moved from Monday to Tuesday. I will make announcements in slack as a reminder.



# Policies

The background of the slide features a dark, almost black, abstract pattern composed of numerous small, semi-transparent triangles. These triangles are arranged in a way that creates a sense of depth and perspective, resembling a star or a complex geometric design. The overall effect is modern and minimalist.

# Bootcamp Spot

# What will I use Bootcamp Spot (Canvas) for?

---

Here's what you'll do in Bootcamp Spot:



View course content and lecture videos



Submit Challenge assignments as files, text or links to GitHub



Take quizzes and assessments



Connect and participate in your Virtual Classes with Zoom



View your grades and module progress

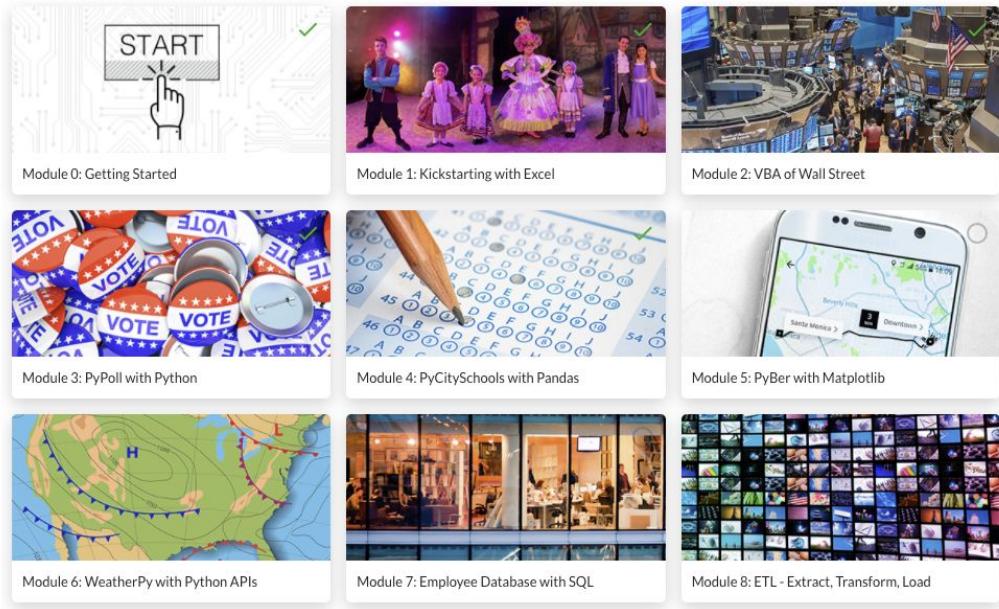


Find information on and connect with Career Services

# My Course in Bootcamp Spot

Your Bootcamp Spot homepage will look similar to this:

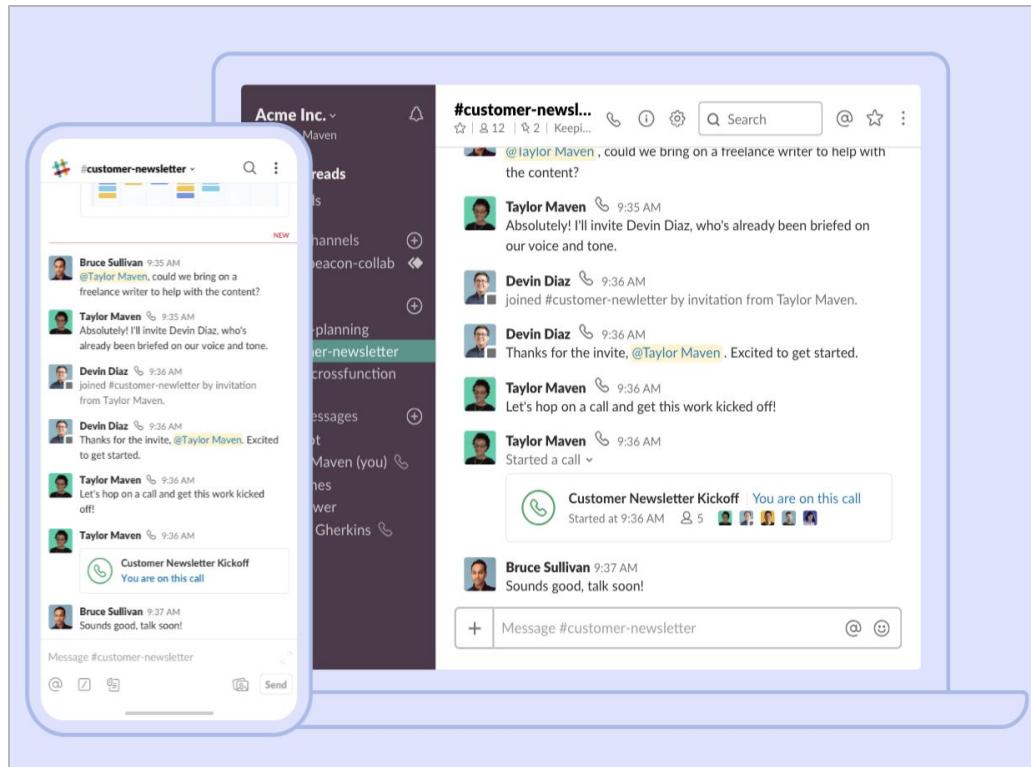
Today, we'll show you the layout of your course, and go over the process of navigating through your first module and talk about how to submit your first Challenge assignment.



# Collaboration on Slack

Slack is the shared virtual workspace where your class will interact during and outside of class hours:

- This class requires a lot of collaboration, so being able to share messages and files quickly is crucial for our success!
- After creating an account, be sure to bookmark the URL in your browser for easy access.



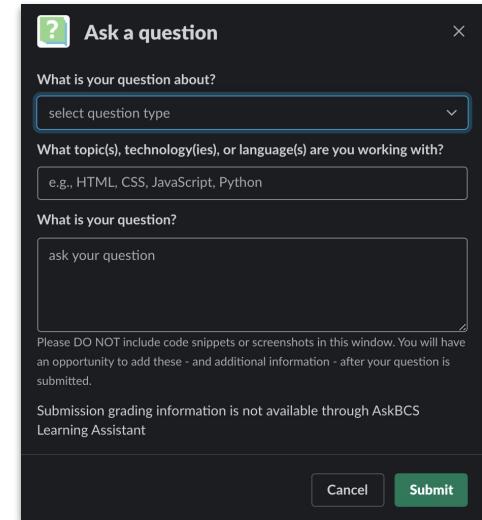
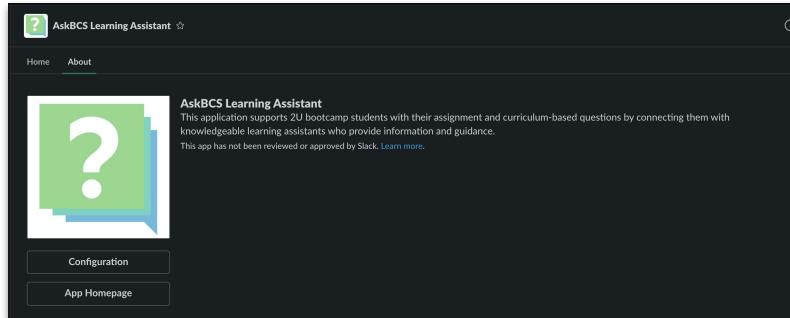
# AskBCS Learning Assistant

Within the Slack workspace, you should see an App called [AskBCS Learning Assistant](#)

Our team of LAs work  
during the daytime to help  
you get unstuck with your  
self-paced lessons.

**They're available:**

Monday—Sunday:  
24 hours, 7 days a week



# TA Office Hours

---

Our TAs will hold Office Hours to provide support outside of normal class times.

This is an unstructured, “drop-in” session where you can bring any question to the table -- from help on your homework, to personal projects!

For our first set of Office Hours, we will be reviewing the downloads and installs from the prework to make sure you’re ready for class! Office hours are accessed through Canvas on the class calendar and are held from 10am to 12pm Saturday mornings.

This weekend's office hours are **highly recommended** we want to make sure you have all your downloads and installs ready for next week!

# GitLab & GitHub

---

Git is software that lets you save a lot of different versions of a file or project. You'll hear Git as a prefix to a few things in the course.



- Your own portfolio
- Store your homework here, and submit a link to Bootcamp Spot
- Use it beyond the boot camp



- Virtual Class exercises and resources
- Shared by the class
- Won't use it beyond the boot camp



**Note:** Git, GitHub, and GitLab are common sticking points for new students. The first few TA Office Hours sessions will focus on setting these up properly.

# Activity: Systems Check

You should have access to the following systems. Follow the steps below to check, and type “ready” in the Group Chat if you’re good!

01

## Bootcamp Spot

Navigate to  
[courses.bootcampspot.com/](https://courses.bootcampspot.com/)  
and you should see the  
full course!

02

## Slack

React to the post in your  
[#01-live class](#) channel



03

## GitLab

Navigate to [GIT LAB](#) and  
login. You should get a 404 if  
you made an account.



# Career Services

# Becoming Employer-Competitive

# Employer-Ready and Employer-Competitive

---

What does employer-ready and employer-competitive mean?

01

Employer-Ready

An employer-ready candidate has created strong professional materials that meet the basic criteria for the typical job application process.

02

Employer-Competitive

Once created, these materials should be adapted to target specific career goals and opportunities, elevating them to employer-competitive.

# Milestones to Career Success

Home  
Navigator  
Modules  
Syllabus  
Grades  
Zoom  
Billing  
**Career Services**  
Career Events  
Student Support

## All Career Milestones



### Intro to Career Services: Employer-Ready vs Employer-Competitive

This brief overview provides an introduction to Career Services and how we support you in becoming Employer-Competitive.

[Assignment Details >](#)



### Milestone: Pave Your Pathway

For this submission, you will learn how to research target roles based on your background, interests, and skills, and you'll develop a professional brand statement that you can use for a variety of purposes (e.g. your resume, LinkedIn and other online profiles). Once you've developed your statement, you'll submit it to a Career Material Advisor for feedback.

[Assignment Details >](#)



### Milestone: Develop Your Resume

For this submission, you will develop a competitive resume, and submit it to a Career Material Advisor for feedback.

[Assignment Details >](#)



### Milestone: Polish Your Online Presence

For this submission, you will update your GitHub and LinkedIn profiles to meet employer-competitive standards, and submit links to your Career Material Advisor for feedback.

[Assignment Details >](#)

# Submitting Milestones to Career Services



## Milestone: Develop Your Resume

To meet this milestone, please complete the following: 1) Review details on what makes a competitive resume along with templates here: <https://mycshub.co/Data-Resume> 2) Develop your resume, and submit it as a Google Doc link with 'Anyone Can Edit' access. (Watch this short video on how to set edit permissions in your Google Doc: <http://bit.ly/2LgLAww>)

### Description

For this submission, you will develop a competitive resume, and submit it to a Career Material Advisor for feedback.

## Submit Milestone

### Add URLs to Your Work

+ Add Another URL

### Notes for Your Career Material Advisor

Submit



# Questions

# Add / Drop Date

August 25th, 2021

Have questions after this session?

Phone: 866-948-9237

Email: support@bootcampspot.com

# Data Bootcamp Warm-Up

---



# Thought Experiment:

## The Great Debate

As an entire class, ponder the following question.

Suggested Time:  
20 minutes



# The Great Debate

---

Which do Americans prefer:  
Italian or Mexican food?



# The Great Debate

---

We need to develop a strategy for answering this question with as much confidence possible. Specifically, answer questions like:

-  What data will you attempt to gather?
-  What relationships will you be looking for?
-  How will you ensure your answer is most likely “true”?

## **Assumptions:**

You are given 5 hours and a budget of \$10 to accomplish this.

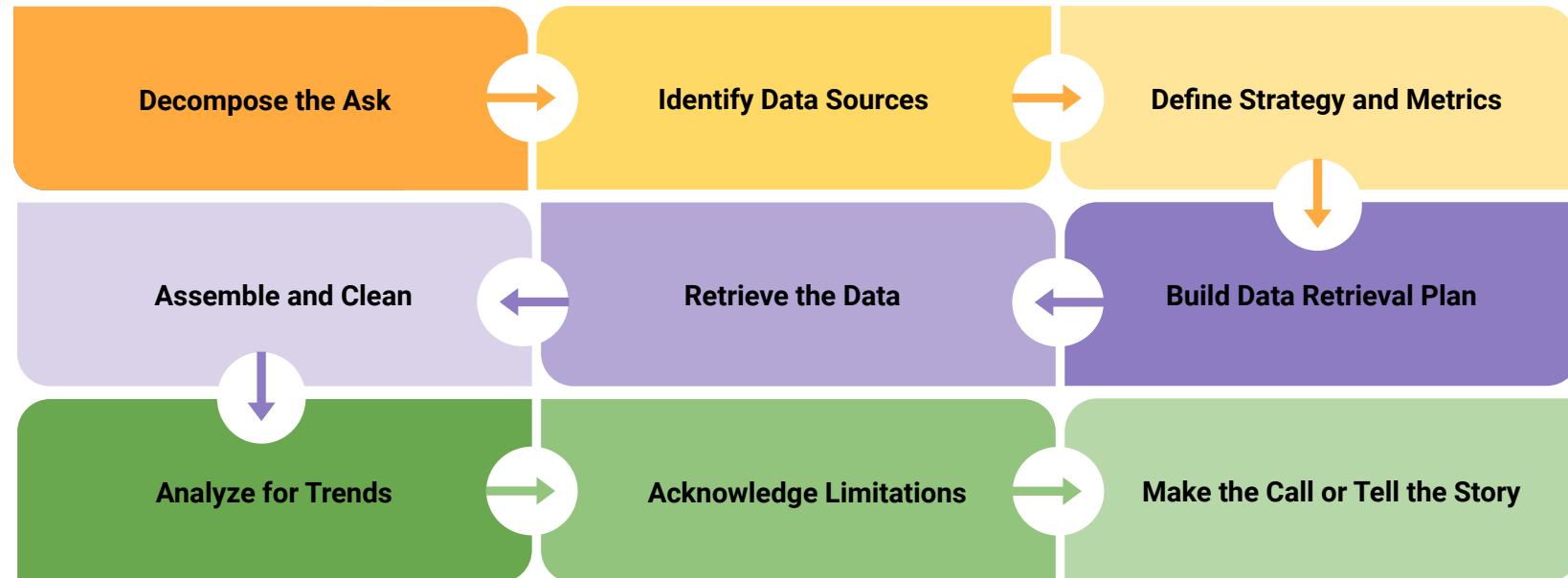
Your answer will be tested by randomly selecting 9 Americans who will each be asked the question.



# The Great Debate (Analyzed)

# Analytics Paradigm

Regardless of type or industry, this paradigm provides a repeatable pathway for effective data problem solving.



# Step 1: Decompose the “Ask”

## Step 1: Decompose the “Ask”

---

Which do **Americans** prefer:  
Italian or Mexican food?



## Step 1: Decompose the “Ask”

---

Which do **Americans** prefer: Italian or Mexican food?



Who exactly is an **American**?



Are **Americans** just homeowners?



Do **Americans** just live in big cities?



Are **Americans** just millennials?



How can we get a  
representative sample  
of Americans?

## Step 1: Decompose the “Ask”

---

Which do Americans **prefer**:  
Italian or Mexican food?



## Step 1: Decompose the “Ask”

---

Which do Americans **prefer**: Italian or Mexican food?



How do we define “preference”?



Do people prefer the foods they eat most frequently?



Do people prefer the foods they wish they could eat if cost was not an issue?



How uniform is the preference? Is it regionalized? Is it different by demographic?



Inherently, preference is **subjective**. We are going to need to make it **objective**.

## Step 1: Decompose the “Ask”

---

Which do Americans prefer:  
**Italian or Mexican food?**



## Step 1: Decompose the “Ask”

---

Which do Americans prefer: **Italian or Mexican food?**

01

How do we categorize foods? Is pizza Italian? Is Taco Bell Mexican?

02

How do we categorize food? Does making pasta at home constitute Italian? Or are we just talking about restaurants?

03

Are we just talking about “best experiences”? Or are we including poorer renditions of these foods?

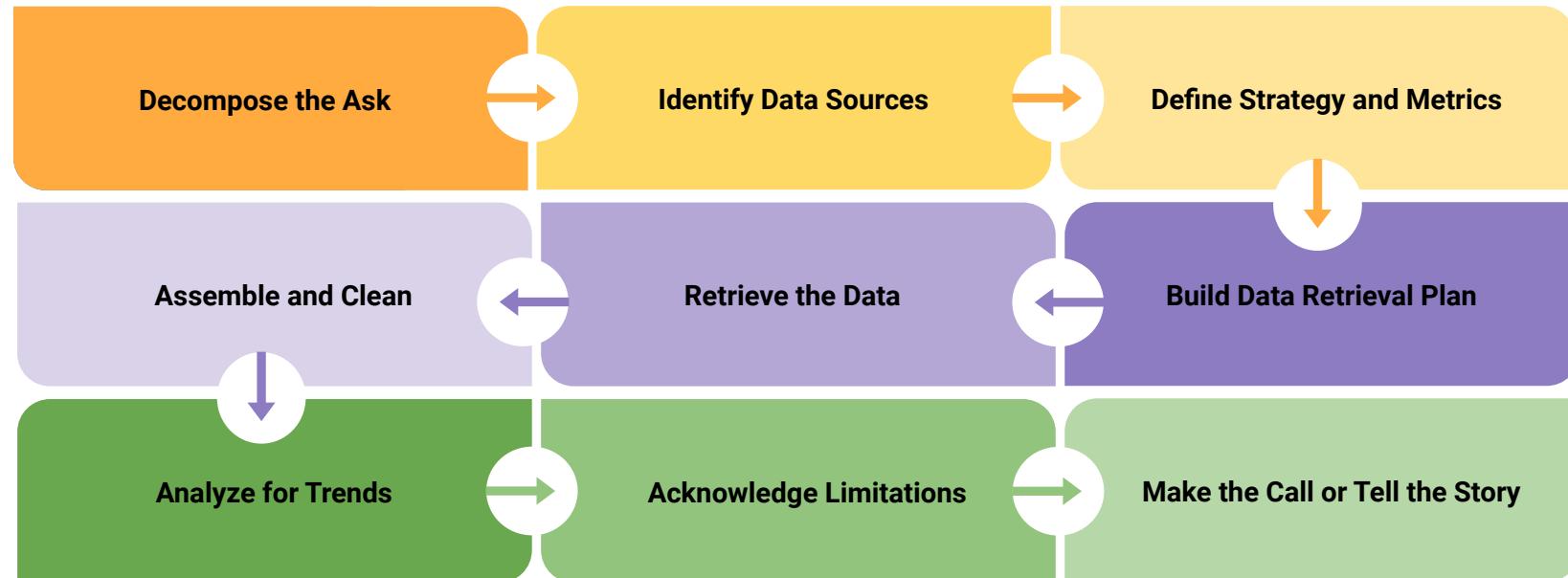


Italian and Mexican are  
**broad categories** we are  
pursuing. We will have to  
narrow the scope.

# Step 2: Identify Data Sources

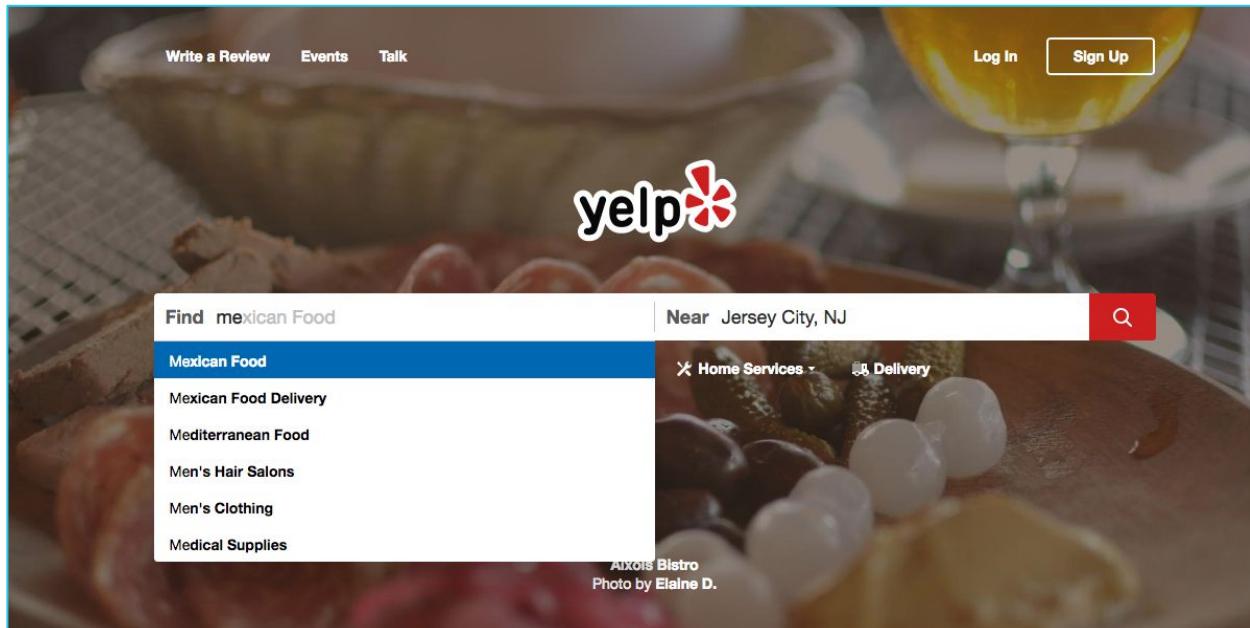
# Analytics Paradigm

Regardless of type or industry, this paradigm provides a repeatable pathway for effective data problem solving.



## Step 2: Identify Data Sources

As everyday consumers, we are **regularly** getting a pulse of everyday American food preferences to inform our own decisions. Perhaps we can make use of the same approach.



# Step 2: Identify Data Sources

Web services like Yelp provide an almost encyclopedic amount of information about the eating preferences of Americans.

The screenshot shows the Yelp website interface for the establishment "Mi Mariachi Taqueria". The top navigation bar includes the Yelp logo, a search bar with the query "Find tacos, cheap dinner, Max's Near Jersey City, NJ", and buttons for "Log In" and "Sign Up". Below the search bar are category filters: "Home Services", "Restaurants", "Auto Services", and "More". A "Write a Review" button is also present. The main content area features the restaurant's name, "Mi Mariachi Taqueria", with an "Unclaimed" status, a 4-star rating from 230 reviews, and a "Write a Review" button. The restaurant is categorized as "\$ Mexican". A map shows its location at 213 Sip Ave, Jersey City, NJ 07306. Below the map are links for "Get Directions", phone number "(201) 222-1998", email "mimariachi.letsseat.at", and "Send to your Phone". There are three images: a sandwich, a dining room interior, and a menu item. A testimonial at the bottom left reads: "Love their Al Pastor and carnitas tacos, shredded lamb, pork ribs with salsa verde and their tamales!" in 13 reviews. At the bottom right, there is a "Full menu" link and a price range indicator of "\$\$\$\$ Price range Under \$10".

## Step 2: Identify Data Sources

---

Why poll an audience when there already exist enormous databases of information about Americans' food preferences—readily available online?



zomato

# Step 2: Identify Data Sources

**Food Type** 

**Review Count** 

**Rating** 

**Yelp Search Results:** Best Italian Food Jersey City, NJ

Showing 1-30 of 3356

**Lorad Nia's Family Pizzeria**  
108 reviews  
Italian Pizza  
Offers takeout and delivery

**Zero Otto Uno Cafe**  
54 reviews  
Italian

**Map:** Mo' Map showing locations of Italian restaurants in New York and surrounding areas.



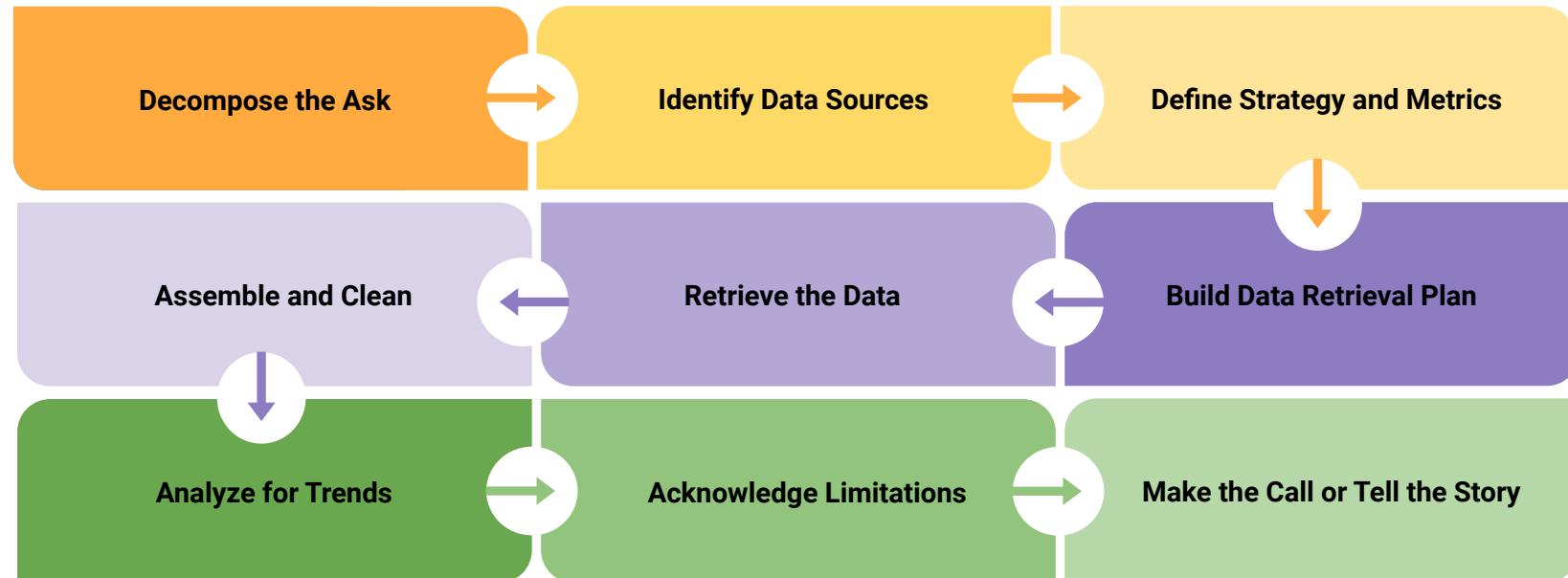
**Lots of Data!**

**Locations**

# Step 3: Define Strategy and Metrics

# Analytics Paradigm

Regardless of type or industry, this paradigm provides a repeatable pathway for effective data problem solving.



## Step 3: Define Strategy and Metrics

---

Here we created a blueprint for what we're targeting:

### **Americans:**

- Ideally, we need thousands of records from Americans in hundreds of different cities. (Large samples)

### **Preference:**

- Number of Yelp Reviews (More = Preference)
- Average Aggregated Ratings (Higher = Preference)

### **Italian and Mexican Food:**

- Top 20 Italian and Mexican restaurants in every city

# Step 3: Define Strategy and Metrics

---

Repeat this analysis for as many cities as possible.

New York, NY	
Italian	Mexican
Restaurant	Restaurant
Restaurant	Restaurant
Restaurant	VS.
Restaurant	Restaurant
Restaurant	Restaurant

Tucson, AZ	
Italian	Mexican
Restaurant	Restaurant
Restaurant	Restaurant
Restaurant	VS.
Restaurant	Restaurant
Restaurant	Restaurant

Washington, D.C.	
Italian	Mexican
Restaurant	Restaurant
Restaurant	Restaurant
Restaurant	VS.
Restaurant	Restaurant
Restaurant	Restaurant

Omaha, NE	
Italian	Mexican
Restaurant	Restaurant
Restaurant	Restaurant
Restaurant	VS.
Restaurant	Restaurant
Restaurant	Restaurant

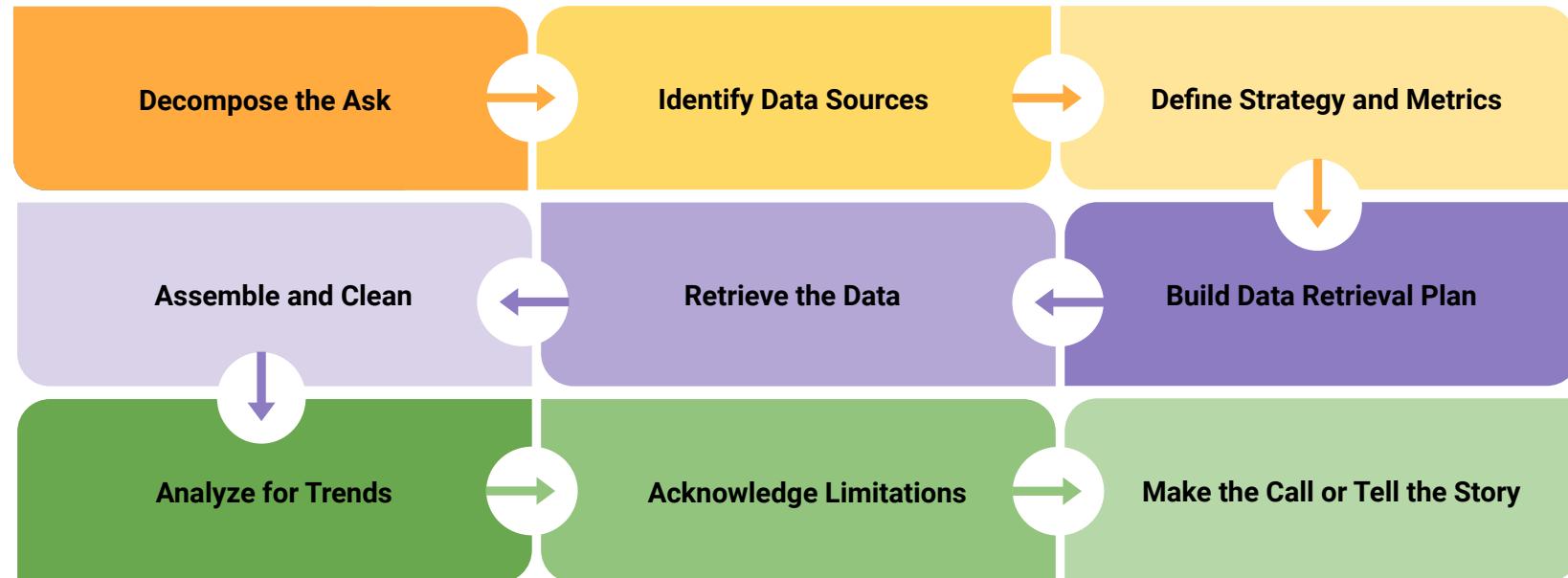
San Diego, CA	
Italian	Mexican
Restaurant	Restaurant
Restaurant	Restaurant
Restaurant	VS.
Restaurant	Restaurant
Restaurant	Restaurant

Atlanta, GA	
Italian	Mexican
Restaurant	Restaurant
Restaurant	Restaurant
Restaurant	VS.
Restaurant	Restaurant
Restaurant	Restaurant

# Step 4: Build Data Retrieval Plan

# Analytics Paradigm

Regardless of type or industry, this paradigm provides a repeatable pathway for effective data problem solving.



## Step 4: Build Data Retrieval Plan

---

We could retrieve this data by brute force, but it would be:

- Extremely time consuming
- Skewed by our city familiarity
- Labor intensive

The image displays three separate instances of a red-themed Yelp search interface. Each instance includes the Yelp logo, a search bar with the text "Find Mexican", a location input field with "Near [City, State]" (e.g., "Near New York, NY"), and a search button. The three instances represent different geographic locations: New York, NY; Denver, CO; and Oklahoma City, OK.

## Step 4: Build Data Retrieval Plan

---

Basically, it would be nearly impossible.

The image displays three separate instances of a red-themed Yelp search interface. Each instance includes the Yelp logo, a search bar with the text "Find Mexican", a location selector with "Near" followed by a city and state (e.g., "New York, NY"), and a search button. The three instances represent different geographical locations: New York, NY; Denver, CO; and Oklahoma City, OK.

# Thank You, Yelp!

Thankfully, we can take advantage of the **Yelp Fusion API** to programmatically run our queries. (#ThankGoodnessForProgramming)

The screenshot shows the Yelp Fusion API documentation page. The left sidebar has sections for General (Create App, Email / Notifications, Display Requirements, Terms of Use, FAQ), Yelp Fusion (Introduction, Business Endpoints, Business Search, Phone Search), and a search bar. The main content area is for the **/businesses/search** endpoint. It includes a description of the endpoint, a note about returning businesses without reviews, a Request section with the URL `GET https://api.yelp.com/v3/businesses/search`, and a Parameters section with two entries: `term` (string) and `location` (string).

**General**

[Create App](#)

[Email / Notifications](#)

[Display Requirements](#)

[Terms of Use](#)

[FAQ](#)

**Yelp Fusion**

[Introduction](#)

[Business Endpoints](#)

[Business Search](#)

[Phone Search](#)

**/businesses/search**

This endpoint returns up to 1000 businesses based on the provided search criteria. It has some basic information about the business. To get detailed information and reviews, please use the Business ID returned here and refer to [/businesses/{id}](#) and [/businesses/{id}/reviews](#) endpoints.

Note: at this time, the API does not return businesses without any reviews.

**Request**

```
GET https://api.yelp.com/v3/businesses/search
```

**Parameters**

These parameters should be in the query string.

Name	Type	Description
term	string	Optional. Search term, for example "food" or "restaurants". The term may also be business names, such as "Starbucks". If term is not included the endpoint will default to searching across businesses from a small number of popular categories.
location	string	Required if either latitude or longitude is not provided. This string indicates the geographic area to be used when searching for businesses. Examples: "New York City", "NYC", "350 5th Ave, New York, NY 10118". Businesses returned in the response may not be strictly within the specified location.

# Thank You, Yelp!

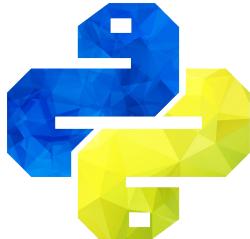
## Response Body

```
{  
  "total": 8228,  
  "businesses": [  
    {  
      "rating": 4,  
      "price": "$",  
      "phone": "+14152520800",  
      "id": "four-barrel-coffee-san-francisco",  
      "is_closed": false,  
      "categories": [  
        {  
          "alias": "coffee",  
          "title": "Coffee & Tea"  
        }  
      ],  
      "review_count": 1738,  
      "name": "Four Barrel Coffee",  
      "url": "https://www.yelp.com/biz/four-barrel-coffee-san-francisco",  
      "coordinates": {  
        "latitude": 37.7670169511878,  
        "longitude": -122.42184275  
      },  
      "image_url": "http://s3-media2.fl.yelpcdn.com/bphoto/HmgtaSP31_t4tPCL1iAsCg/o.jpg",  
      "location": {  
        "city": "San Francisco",  
        "country": "US",  
        "address2": "",  
        "address3": "",  
        "state": "CA",  
        "address1": "375 Valencia St",  
        "zip_code": "94103"  
      },  
      "distance": 1604.23,  
      "transactions": ["pickup", "delivery"]  
    },  
    // ...  
  ],  
  "region": {  
    "center": {  
      "latitude": 37.767413217936834,  
      "longitude": -122.42828739746094  
    }  
  }  
}
```



# Step 4: Build Data Retrieval Plan

We will build a Python script to randomly select over 700 zip codes from the U.S. Census, and then acquire review data from the top 20 Mexican and Italian restaurants for each zip code using the Yelp API.



11101	
Italian	Mexican
Restaurant	Restaurant

07360	
Italian	Mexican
Restaurant	Restaurant

20001	
Italian	Mexican
Restaurant	Restaurant

68007	
Italian	Mexican
Restaurant	Restaurant

22434	
Italian	Mexican
Restaurant	Restaurant

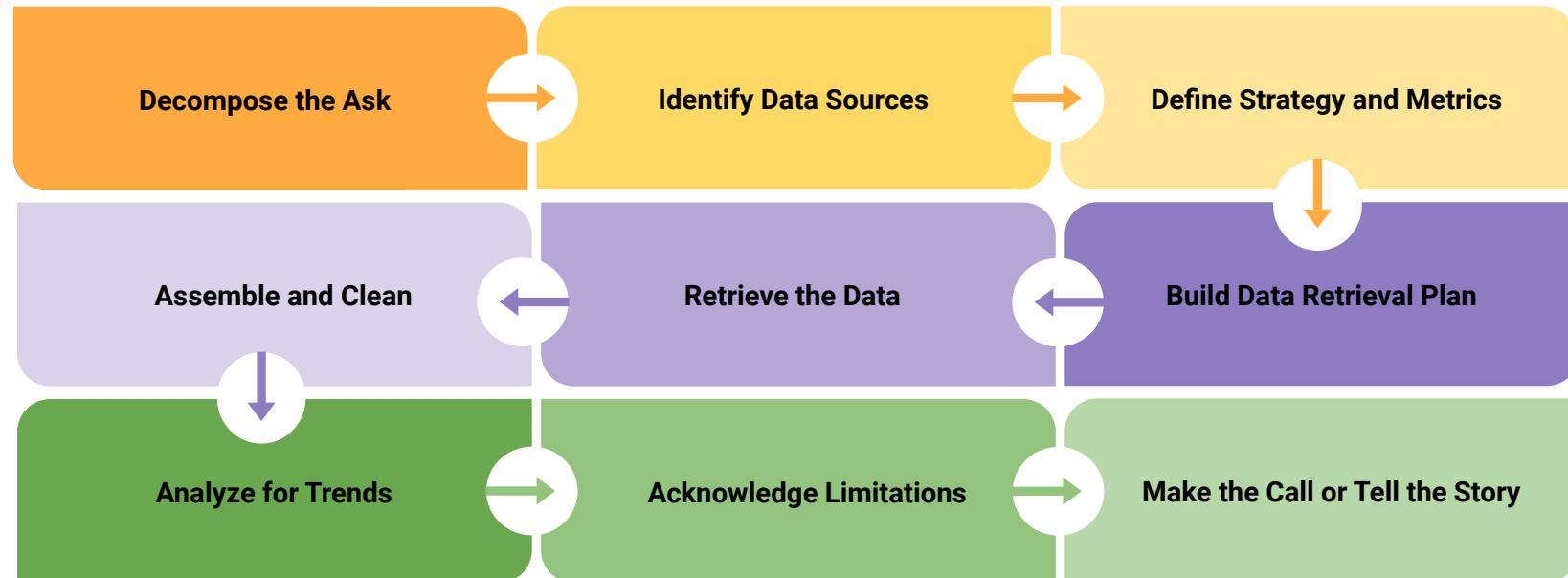
30301	
Italian	Mexican
Restaurant	Restaurant

A diagram illustrating the data retrieval process. It starts with the "United States Census Bureau" logo, which has a yellow arrow pointing down to a red header bar of a web browser. The browser window displays the "yelp fusion" logo and an illustration of three industrial robots or data processing units. The red header bar contains the yelp logo, "Fusion", "Fusion API", "GraphQL", and "Manage App". Below the header bar, there are social sharing icons (Share, f, t, in) and a user profile icon.

# Step 5: Retrieve the Data

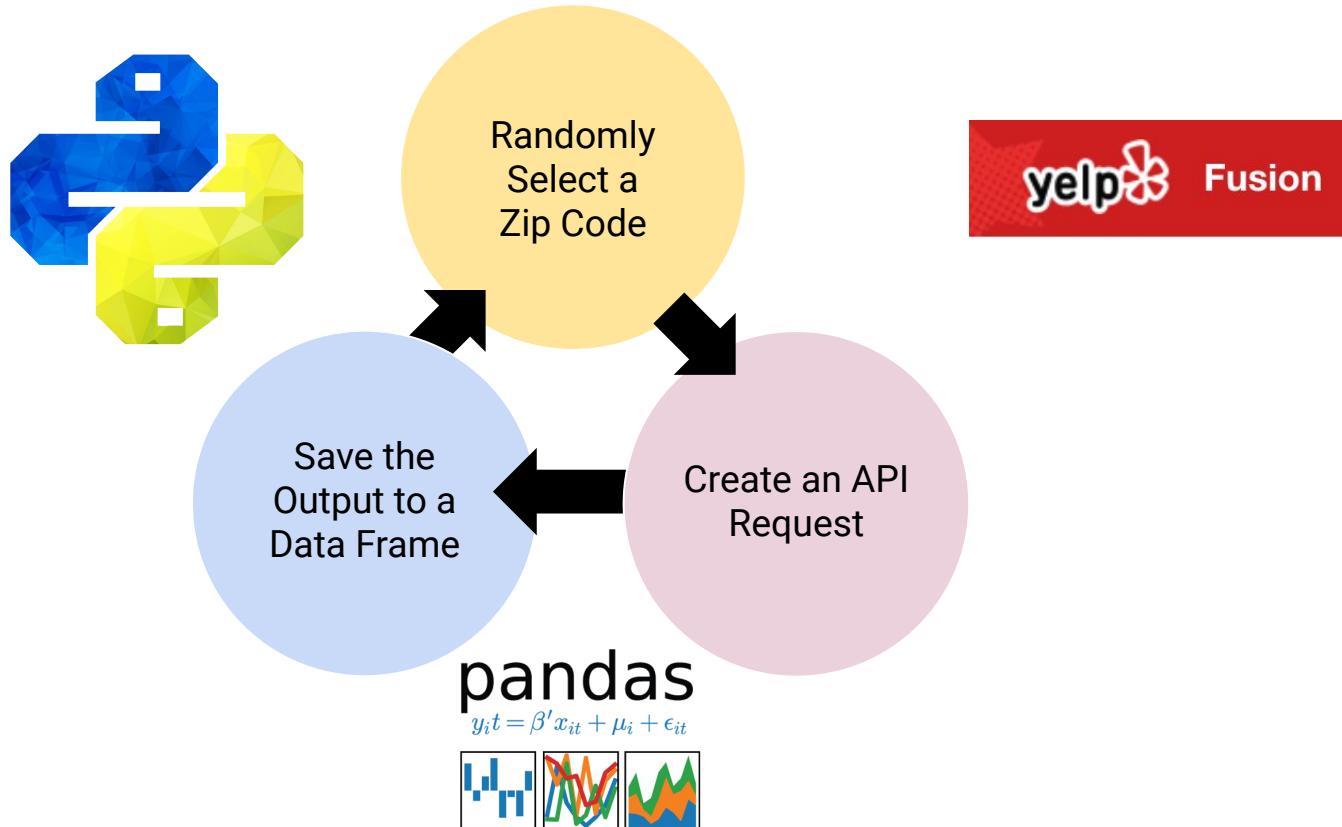
# Analytics Paradigm

Regardless of type or industry, this paradigm provides a repeatable pathway for effective data problem solving.



# Pulling with Python

---



# Pulling with Python

---

```
# Use Try-Except to handle errors
try:

    # Loop through all records to calculate the review count and weighted review value
    for business in yelp_reviews_italian["businesses"]:

        italian_review_count = italian_review_count + business["review_count"]
        italian_weighted_review = italian_weighted_review + business["review_count"] * business["rating"]

    for business in yelp_reviews_mexican["businesses"]:
        mexican_review_count = mexican_review_count + business["review_count"]
        mexican_weighted_review = mexican_weighted_review + business["review_count"] * business["rating"]

    # Append the data to the appropriate column of the data frames
    italian_data.set_value(index, "Zip Code", row["Zipcode"])
    italian_data.set_value(index, "Italian Review Count", italian_review_count)
    italian_data.set_value(index, "Italian Average Rating", italian_weighted_review / italian_review_count)
    italian_data.set_value(index, "Italian Weighted Rating", italian_weighted_review)

    mexican_data.set_value(index, "Zip Code", row["Zipcode"])
    mexican_data.set_value(index, "Mexican Review Count", mexican_review_count)
    mexican_data.set_value(index, "Mexican Average Rating", mexican_weighted_review / mexican_review_count)
    mexican_data.set_value(index, "Mexican Weighted Rating", mexican_weighted_review)

except:
    print("Uh oh")
```



This funky code...

# Pulling with Python

---

```
1  
https://api.yelp.com/v3/businesses/search?term=Italian&location=76556  
https://api.yelp.com/v3/businesses/search?term=Mexican&location=76556  
2  
https://api.yelp.com/v3/businesses/search?term=Italian&location=72039  
https://api.yelp.com/v3/businesses/search?term=Mexican&location=72039  
3  
https://api.yelp.com/v3/businesses/search?term=Italian&location=61606  
https://api.yelp.com/v3/businesses/search?term=Mexican&location=61606  
4  
https://api.yelp.com/v3/businesses/search?term=Italian&location=47232  
https://api.yelp.com/v3/businesses/search?term=Mexican&location=47232  
5  
https://api.yelp.com/v3/businesses/search?term=Italian&location=60565  
https://api.yelp.com/v3/businesses/search?term=Mexican&location=60565  
6  
https://api.yelp.com/v3/businesses/search?term=Italian&location=20634  
https://api.yelp.com/v3/businesses/search?term=Mexican&location=20634  
7  
https://api.yelp.com/v3/businesses/search?term=Italian&location=71046  
https://api.yelp.com/v3/businesses/search?term=Mexican&location=71046
```



**...will make all of  
these URLs.**

# Pulling with Python

GET https://api.yelp.com/v3/businesses/search?term=Italian&location=37764...

Headers (1)

Key	Value	Description	...	Bulk Edit	Presets
Authorization	Bearer gl6k6JmewUhjMVBv0I2x4Bz_NRIEggSjIjGbTaejmzbvBJXg 36F...				
New key	Value	Description			

Body

Pretty Raw Preview JSON

```
1 {  
2   "businesses": [  
3     {  
4       "id": "two-brothers-italian-pizza-kodak",  
5       "name": "Two Brothers Italian Pizza",  
6       "image_url": "https://s3-media3.fl.yelpcdn.com/bphoto/364BqQt0qtVHV1f0t_xznA/o.jpg",  
7       "is_closed": false,  
8       "url": "https://www.yelp.com/biz/two-brothers-italian-pizza-kodak?adjust_creative=1GwZyE0zIjSujpHtlMnodQ&utm_campaign=yelp_api_v3&utm_medium=  
9         _api_v3_business_search&utm_source=1GwZyE0zIjSujpHtlMnodQ",  
10      "review_count": 8,  
11      "categories": [  
12        {  
13          "alias": "pizza",  
14          "title": "Pizza"  
15        },  
16        {  
17          "alias": "italian",  
18          "title": "Italian"  
19        },  
20        {  
21          "alias": "pastashops",  
22          "title": "Pasta Shops"  
23        },  
24      ],  
25      "rating": 2,  
26      "coordinates":  
27        {  
28          "latitude": 35.9638662447754,  
29          "longitude": -83.5926620147413  
30        },  
31      "transactions": [],  
32      "location": {  
33        "address1": "1000 W Broad St",  
34        "address2": null,  
35        "city": "Columbus",  
36        "state": "OH",  
37        "zip_code": "43228",  
38        "country": "US",  
39        "cross_streets": "Highland Ave & 10th St",  
40        "display_address": ["1000 W Broad St", "Columbus, OH 43228"],  
41        "distance": 1000, "lat": 35.9638662447754, "lon": -83.5926620147413  
42      }  
43    }  
44  ]  
45}  
46
```

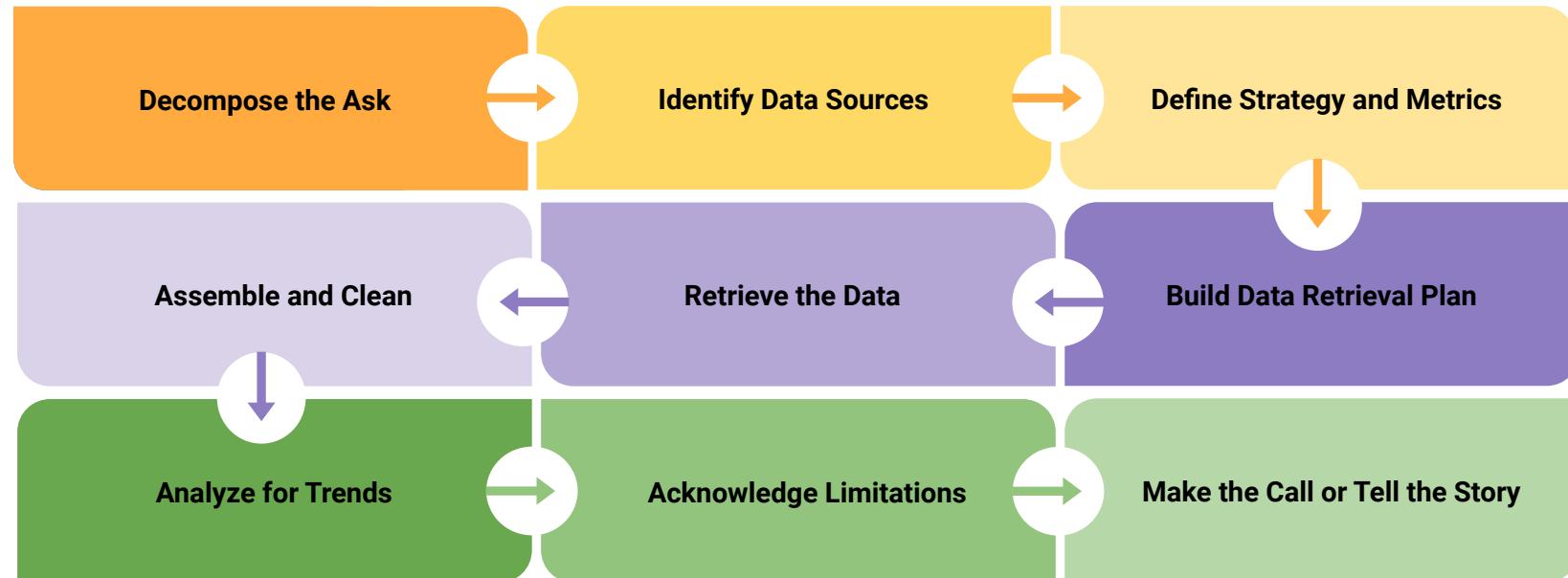


Each of these URLs holds a piece of our answer.

# Step 6: Assemble and Clean the Data

# Analytics Paradigm

Regardless of type or industry, this paradigm provides a repeatable pathway for effective data problem solving.



# Cleaning with Pandas

---

No data comes out intrinsically the way you want it to.

In our case, we needed multiple steps to aggregate the data along our channels of interest.

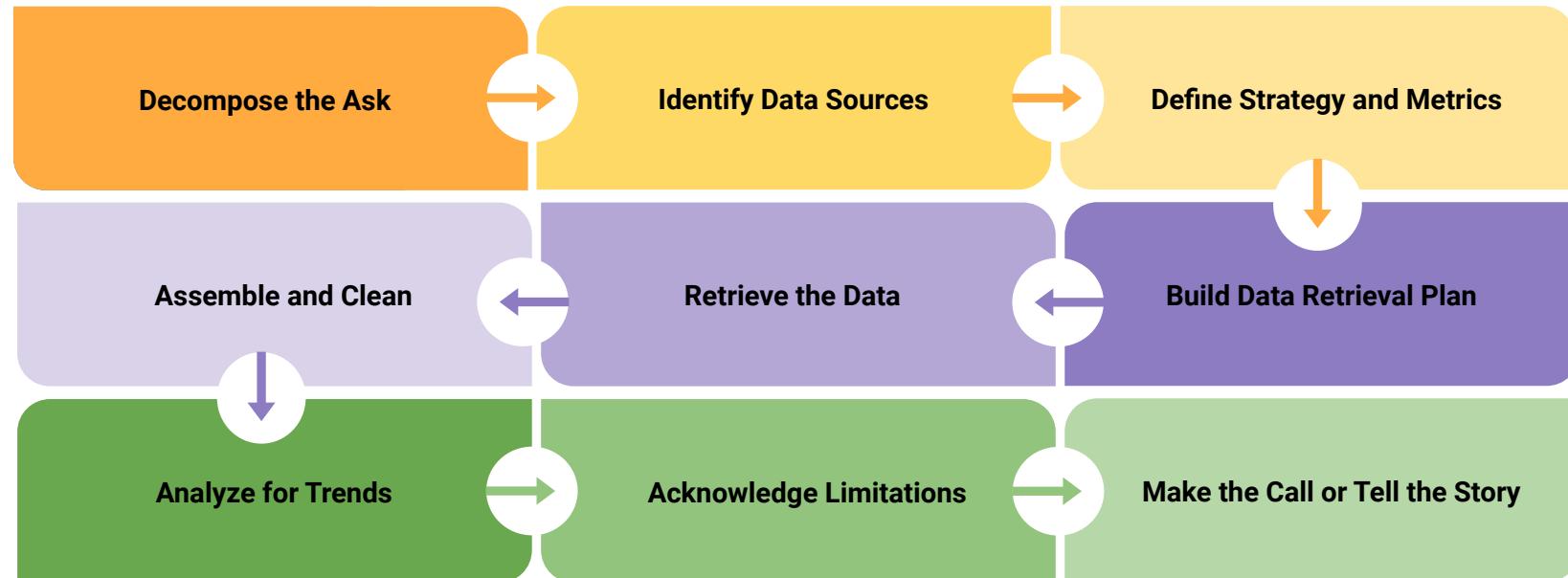
```
# Combine DataFrames into a single DataFrame  
combined_data = pd.merge(mexican_data, italian_data, on="Zip Code")  
combined_data.head()
```

	Zip Code	Mexican Review Count	Mexican Average Rating	Mexican Weighted Rating	Italian Review Count	Italian Average Rating	Italian Weighted Rating
0	76556	97	4.1134	399	63	3.78571	238.5
1	72039	256	4.11133	1052.2	266	3.81955	1016
2	61606	378	3.64286	1377	66	3.2197	212.5
3	47232	222	4.16892	925.5	420	3.77857	1587
4	60565	2842	3.94053	11199	2829	3.92824	11113

# Step 7: Analyze for Trends

# Analytics Paradigm

Regardless of type or industry, this paradigm provides a repeatable pathway for effective data problem solving.



# Analyze for Trends (Table)

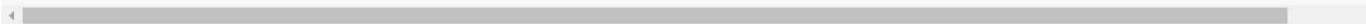
It's Close:

## Display Summary of Results

```
# Model 1: Head-to-Head Review Counts
italian_summary = pd.DataFrame({"Review Counts": italian_data["Italian Review Count"].sum(),
                                 "Rating Average": italian_data["Italian Average Rating"].mean(),
                                 "Review Count Wins": combined_data["Review Count Wins"].value_counts()["Italian"],
                                 "Rating Wins": combined_data["Rating Wins"].value_counts()["Italian"], index=["Italian"]})

mexican_summary = pd.DataFrame({"Review Counts": mexican_data["Mexican Review Count"].sum(),
                                 "Rating Average": mexican_data["Mexican Average Rating"].mean(),
                                 "Review Count Wins": combined_data["Review Count Wins"].value_counts()["Mexican"],
                                 "Rating Wins": combined_data["Rating Wins"].value_counts()["Mexican"], index=["Mexican"]})

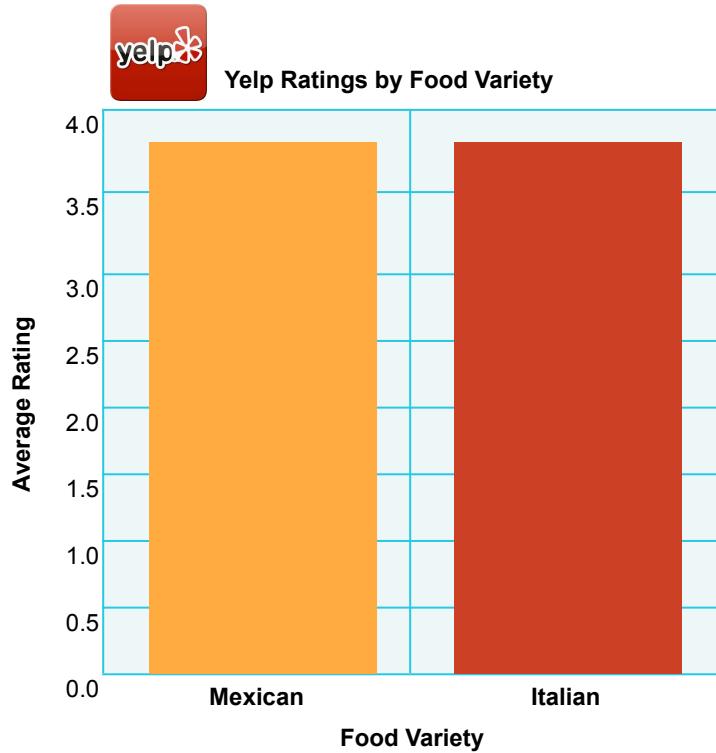
final_summary = pd.concat([mexican_summary, italian_summary])
final_summary
```



	Rating Average	Rating Wins	Review Count Wins	Review Counts	
Mexican	3.826588	273	220	476889	
Italian	3.806869	245	298	573733	

# Analyze for Trends (Ratings)

Yelpers rate Italian and Mexican relatively **equally**.

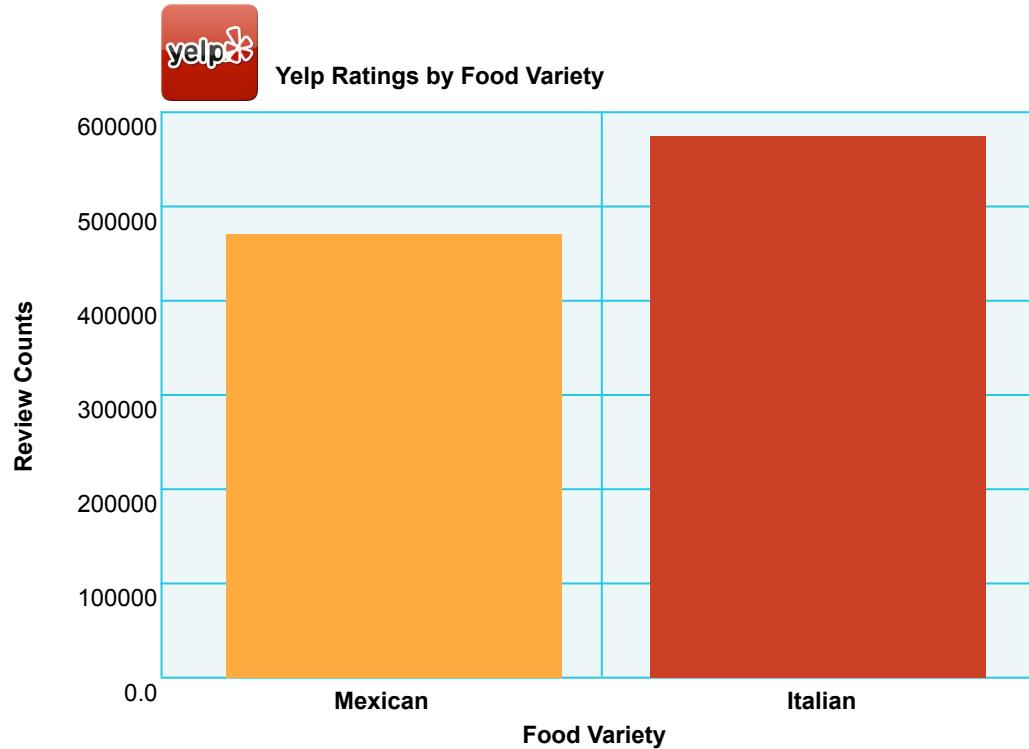


=



# Analyze for Trends (Ratings)

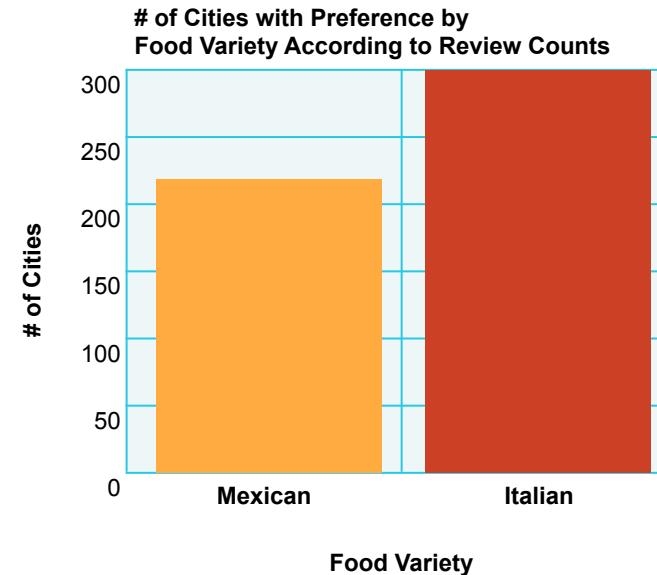
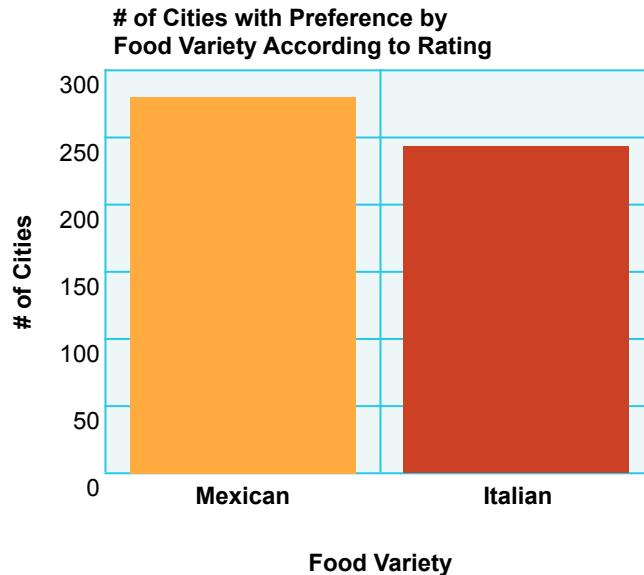
Yelpers seem to significantly **review more Italian** restaurants.



# Analyze for Trends (Winner Take All)

Just for kicks, let's throw in an analysis that aggregates the data from all cities using a winner-take-all approach.

**It's sort of a wash.**



# Analyze for Trends (Statistical Analysis)

---

Because of how close the numbers appear, we utilized a Student's t-test to quickly assess if the perceived differences are not statistically significant but could be considered substantial.

Metric	Italian	Mexican	p-Value (t-test)
Average Rating	3.806	3.826	0.284
Review Counts	573k	476k	0.057

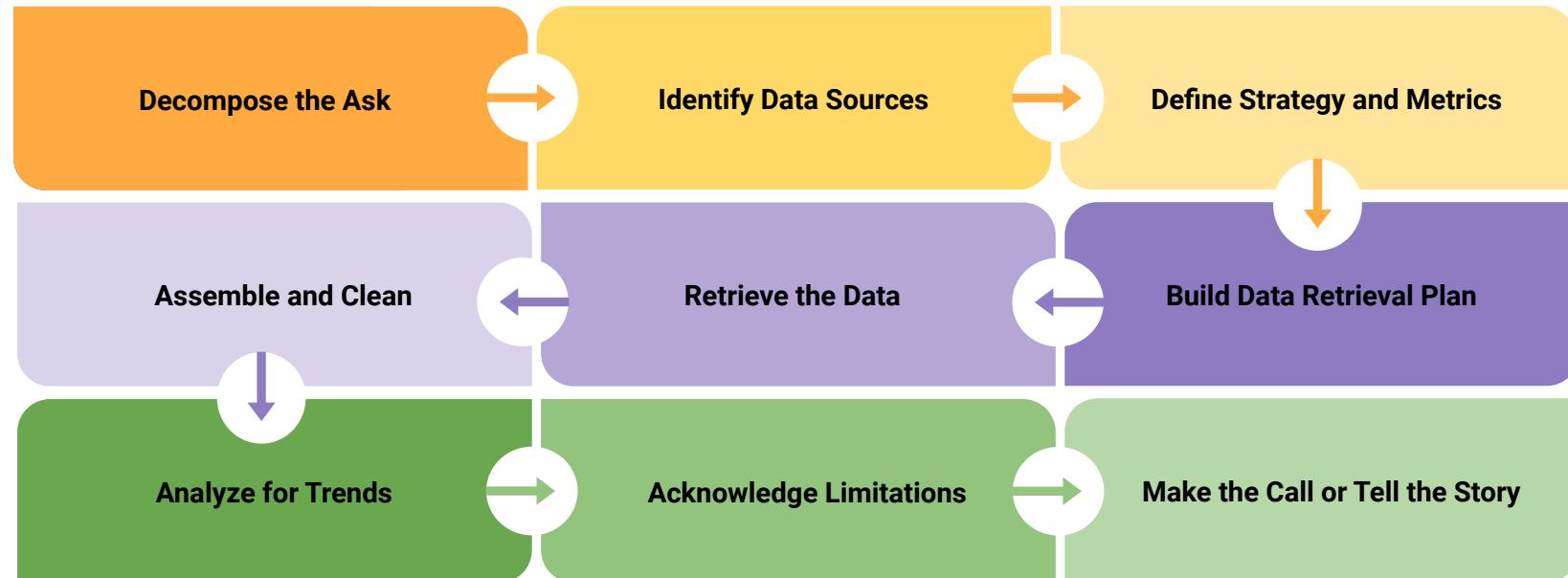


The difference in review count is **not statistically significant**.

# Step 8: Acknowledge Limitations

# Analytics Paradigm

Regardless of type or industry, this paradigm provides a repeatable pathway for effective data problem solving.



# Limitations of Analysis

Yelp demographics may not match the American demographic.



# Limitations of Analysis

---

Restaurant experiences do not equate to home-cooked meals.



# Limitations of Analysis

---

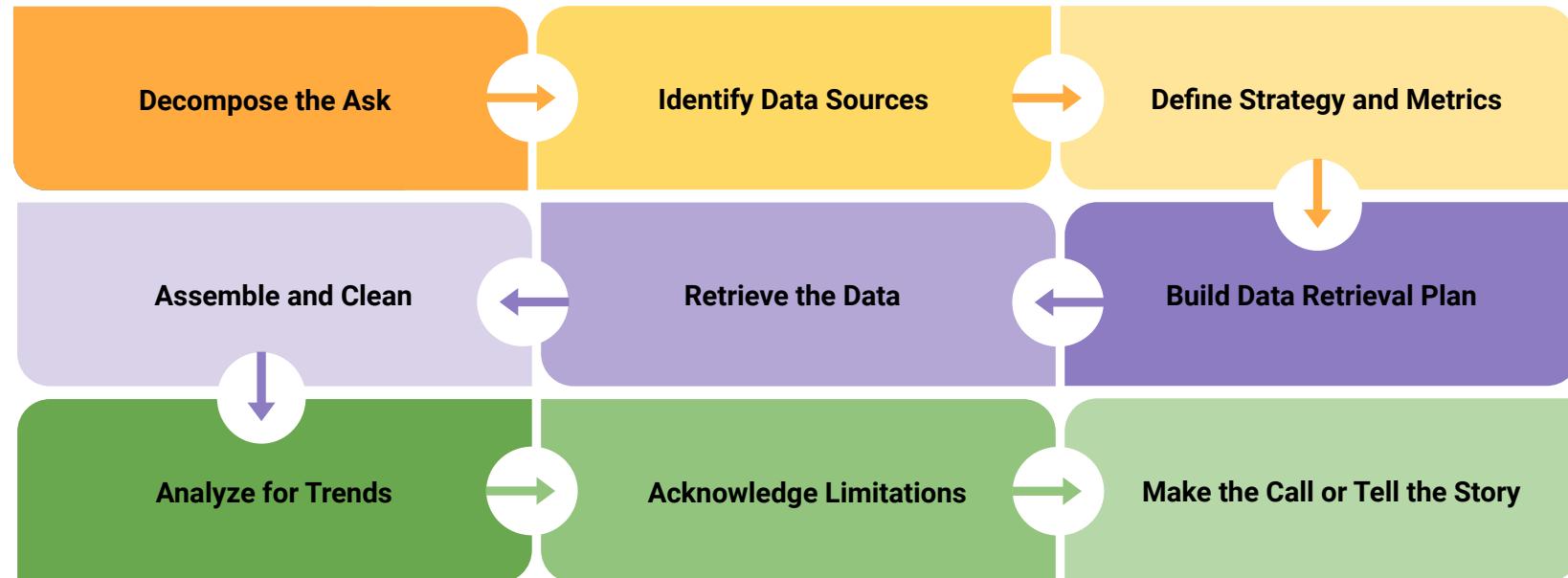
Fine-dining effect?



# Step 9: Make the Call

# Analytics Paradigm

Regardless of type or industry, this paradigm provides a repeatable pathway for effective data problem solving.



# Making the Call

---

## The “Proper” Conclusion:

Based on our analysis, it's clear that Americans' preferences for Italian and Mexican food are similar in nature. As a whole, Americans rate Mexican and Italian restaurants at non-statistically similar scores (avg. score: 3.8, p-value: 0.285). Although there are more reviews for Italian restaurants, we have shown that the difference is not statistically significant (+96k, p-value: 0.057).



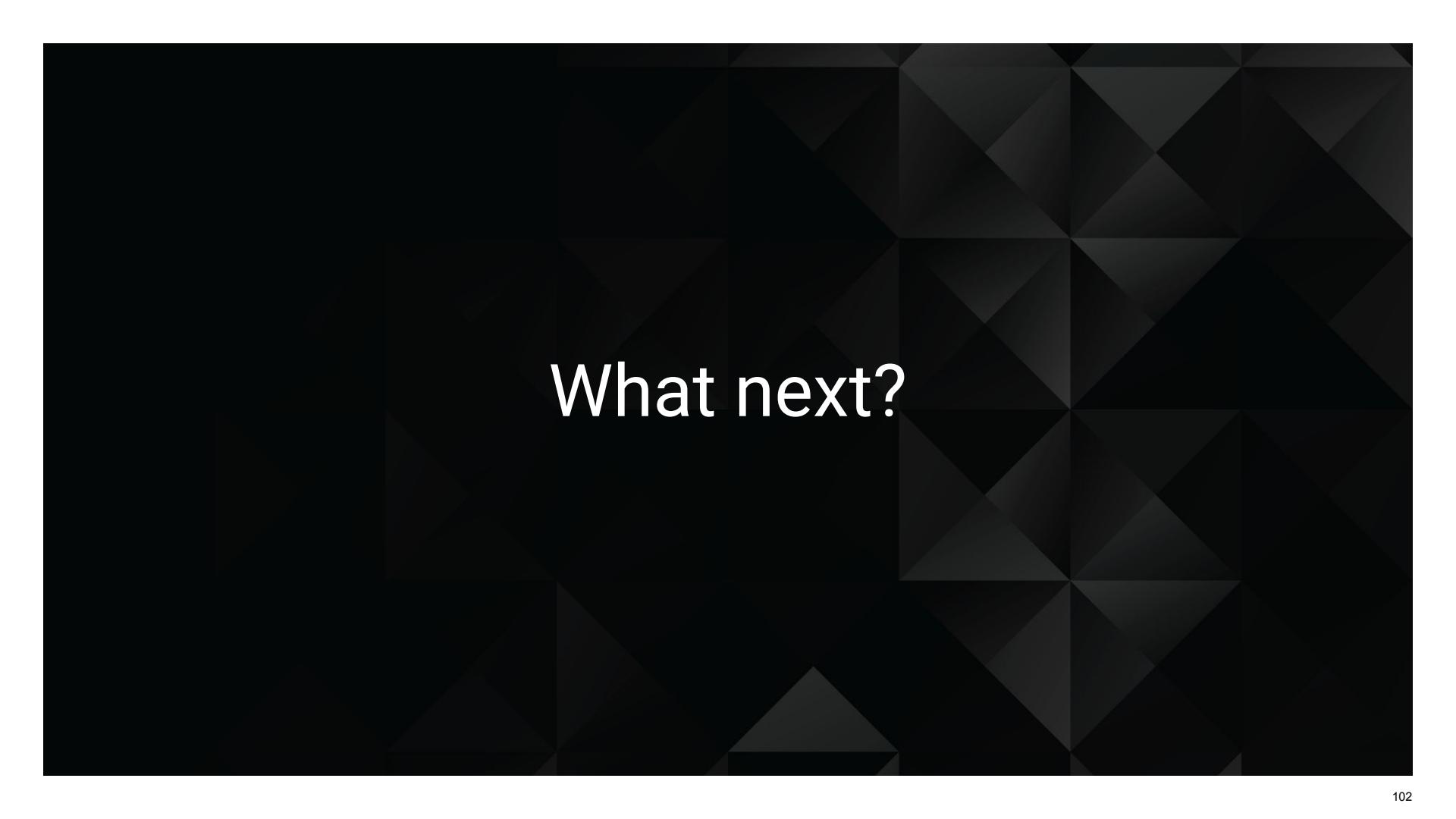
This may indicate there is an increased interest in visiting Italian restaurants at an experiential level. Or it may merely suggest that Yelp users enjoy writing reviews of Italian restaurants more than Mexican restaurants.

# Making the Call

---

The “Let’s Be Real” Conclusion: Italian (but it’s going to be close).





What next?

# Next Steps

---

1. Review the Welcome Session Recap email you'll receive tomorrow:
  - a. Github/Gitlab instructions
  - b. Slack invite link
  - c. Recording from tonight
  - d. This presentation deck
2. Get started with the Onboarding Survey and Module 1 Lessons on Canvas
3. Attend Install Weekend with the TAs starting Friday evening
4. Come to your Virtual Classes next week. The next class will be Tuesday February 23, 2021.