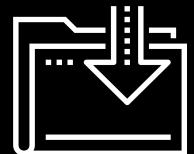




Data Analytics & Visualization Boot Camp

Data Boot Camp
Online—Welcome Session



WELCOME!

Virtual Class Best Practices



- Please keep your microphone muted, unless you are called upon or your host asks for feedback/communication
- Kindly turn your video on so we can see you
- Please update your name under the Zoom *Participants* tab to your full name
- 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.
- Please hold all questions until the designated Q&A section!

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

Chavon Rodriguez

Student Success Manager

Professional and educational background

- BA in Business Administration & Management from Purdue University.
- I have worked in Education for over 18 years as a Academic Advising Manager and as a Associate Director of Academics for Military students. I've been in my Current role as a Student Success Manager for 4 years.

Fun Fact:

I'm a twin! My family & friends call me the "Fairy Godmother".



Emily Elia

Teaching Assistant

BS in Computer Science

I like to ask a lot of questions to learn how students learn and use that info to help them arrive at an answer

Music, photography, yoga



Kulraj Singh Kohli

Teaching Assistant

Education & Background:

- Business Intelligence Analyst 2
- Masters in Information Management
University of Washington - Seattle

Hobbies & Interests:

- Coding, Cooking, Sports

I can move my ears! =P



Jamie Miller

Instructor

Education & Background:

- Masters in Management
- Graduate Studies in Operations Research
- Retired - Army Operations Research Analyst
- Data Viz Bootcamp Graduate (2019)
- Data Viz Bootcamp Instructor & Tutor

Hobbies & Interests:

- Mountain biking
- Trail running
- Life-long learner

Fun Fact

- I completed the Vermont 100 mile endurance run in just over 24 ½ hours (2018)



Boot Camp Experience

What does “for working professionals” mean?

What it means:

- Only 2 virtual classes per week
- 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

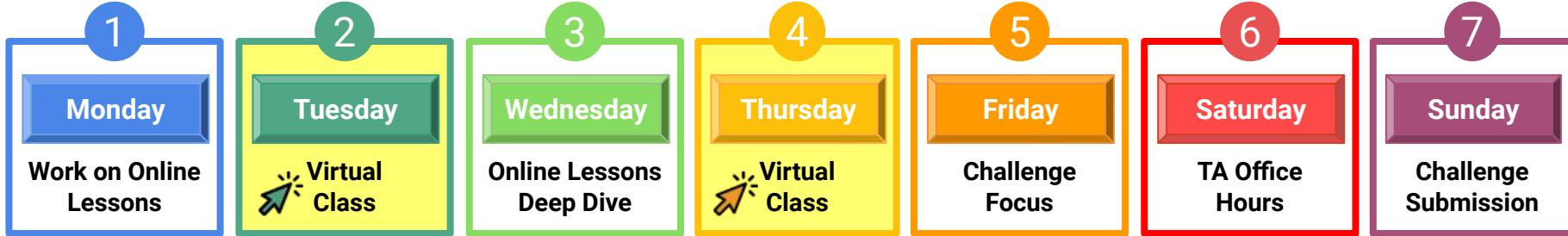
Each week, you'll submit a challenge assignment around the module's technologies.

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.



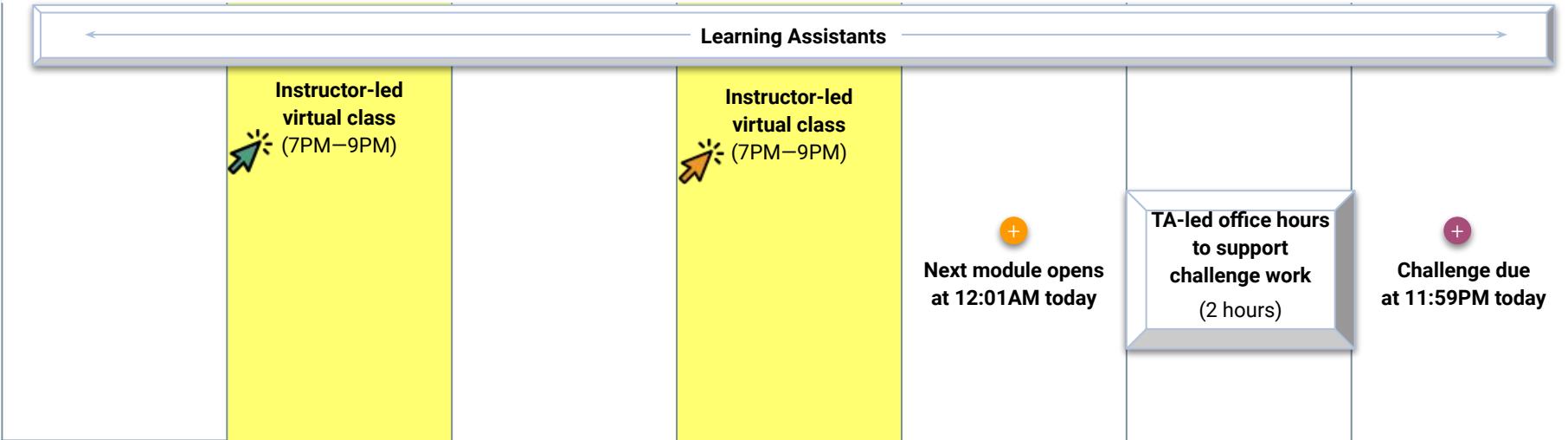
We will teach you how to
fish, but we won't *give* you
the fish.

Weekly Calendar [KEEP FOR TTH]



Academic support

Learning Assistants



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 in all class projects



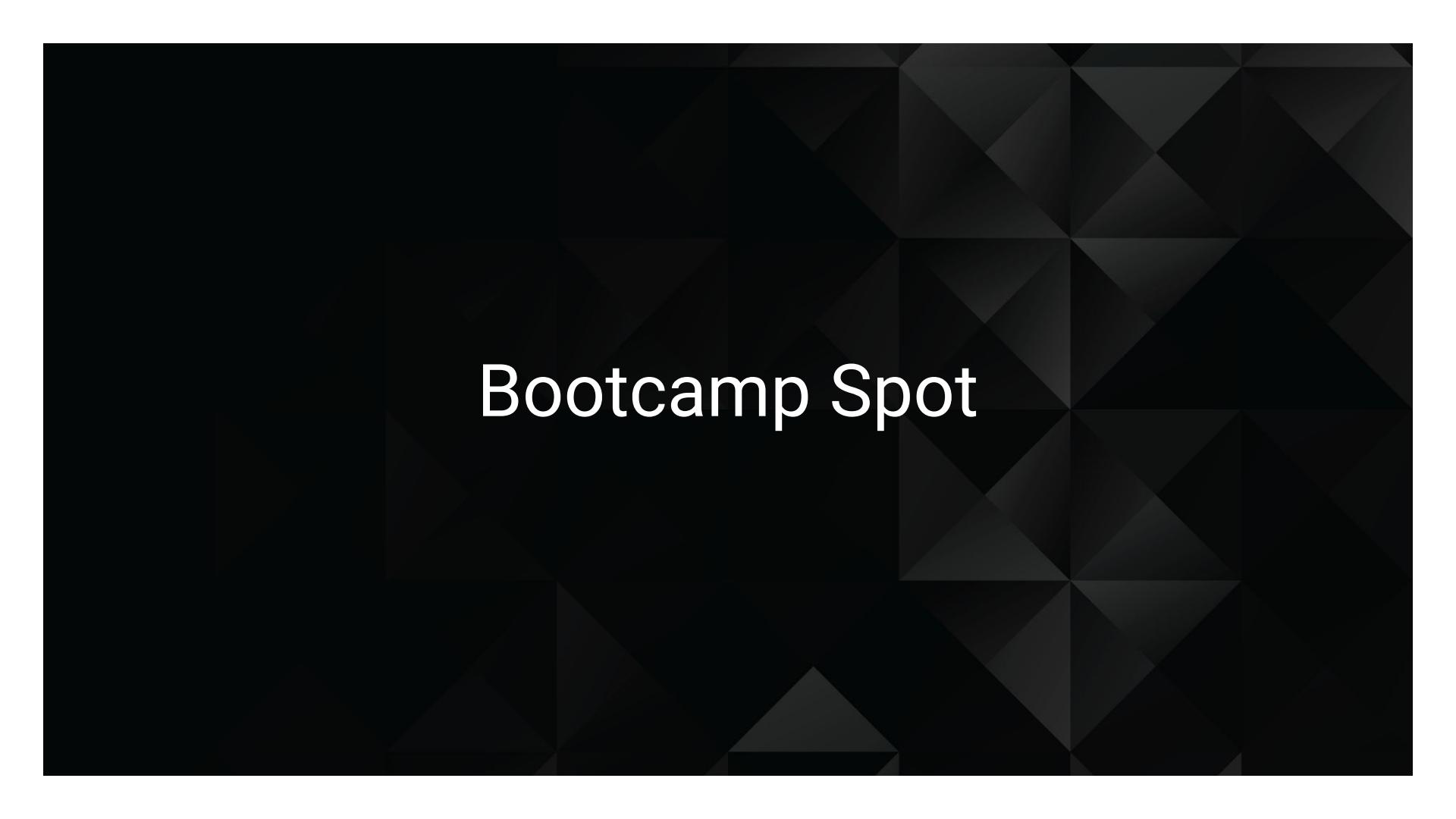
Tuition Paid in Full

Holiday Accommodations

Our Boot Camp recognizes the following Holidays:

- Thanksgiving - Class schedule change - We will not have class on Nov. 25th, which is Thanksgiving day. Class will be held on Thursday, Nov 24th instead.
- Winter Break from 12/24 - 1/1

Any classes scheduled for these days will be adjusted appropriately in the course calendar.

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 the first module and walking through how to submit your first Challenge assignment.



Home

Navigator

Syllabus

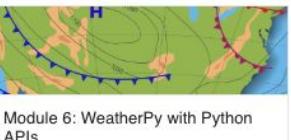
Slack

Grades

People

Zoom

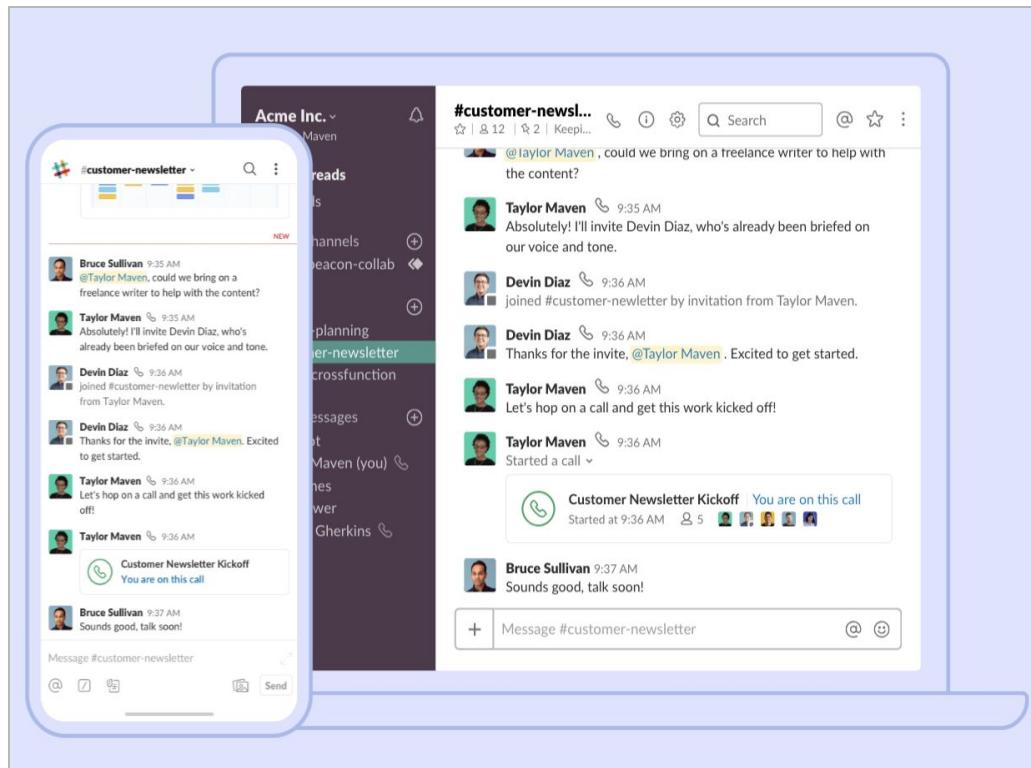
Live Session
Playback

 Module 1: Kickstarting with Excel	 Module 2: VBA of Wall Street	 Module 3: PyPoll with Python
 Module 4: PyCitySchools with Pandas	 Module 5: Pyber with Matplotlib	 Module 6: WeatherPy with Python APIs
 Module 7: Employee Database with SQL	 Module 8: ETL - Extract, Transform, Load	 Module 9: Surf's Up with Advanced Data Storage and Retrieval
 Module 10: Mission to Mars - Web Scraping with HTML/CSS	 Module 11: UFO Sightings with JavaScript	 Module 12: Plotly & Belly Button Biodiversity
 Module 13: Mapping Earthquakes with JS & APIs	 Module 14: NY Citibike with Tableau	 Module 15: R + Statistics

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.



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!

Install Weekend [KEEP FOR TTH]

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!

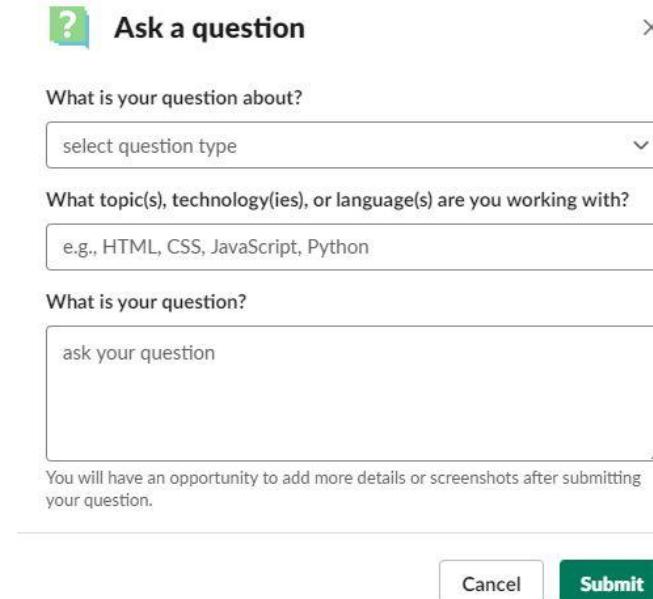
Office Hours will take place every **Saturday** from **12pm to 2pm**.

AskBCS Learning Assistant

Within the Slack Workspace, you should see an application called **AskBCS Learning Assistant**.

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

They are available 24/7



The image shows a screenshot of a Slack application window titled "Ask a question". The window includes fields for asking a question, selecting a question type, specifying topics, and a large text area for the actual question. It also features a note about adding details after submission and two buttons at the bottom: "Cancel" and "Submit".

What is your question about?

select question type

What topic(s), technology(ies), or language(s) are you working with?

e.g., HTML, CSS, JavaScript, Python

What is your question?

ask your question

You will have an opportunity to add more details or screenshots after submitting your question.

Cancel Submit

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/
and you should see the
full course!

02

Slack

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



03

GitHub

Navigate to
<https://github.com/UCF-Coding-Boot-Camp/UCF-VIRT-DATA-PT-10-2021-U-B> and login. You should get a 404 if you have not made an account.



GitHub

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.

Add / Drop Date

October 28th, 2021

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

As your Student Success Manager, I am your POC for ALL things Bootcamp related.



This includes any questions about your enrollment, your class, or even if you just want to chat. If I cannot help you, I know the person who can! Do not hesitate to reach out to me, I am here to help!

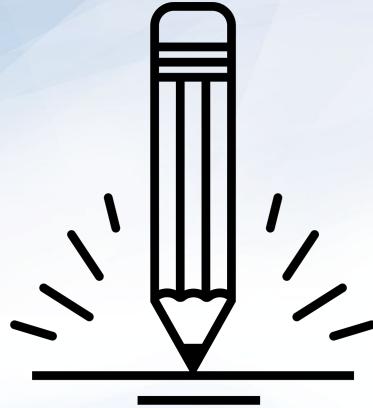


Questions

Have questions after this session?

Email: crodriguez@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—with 0 qualifiers.

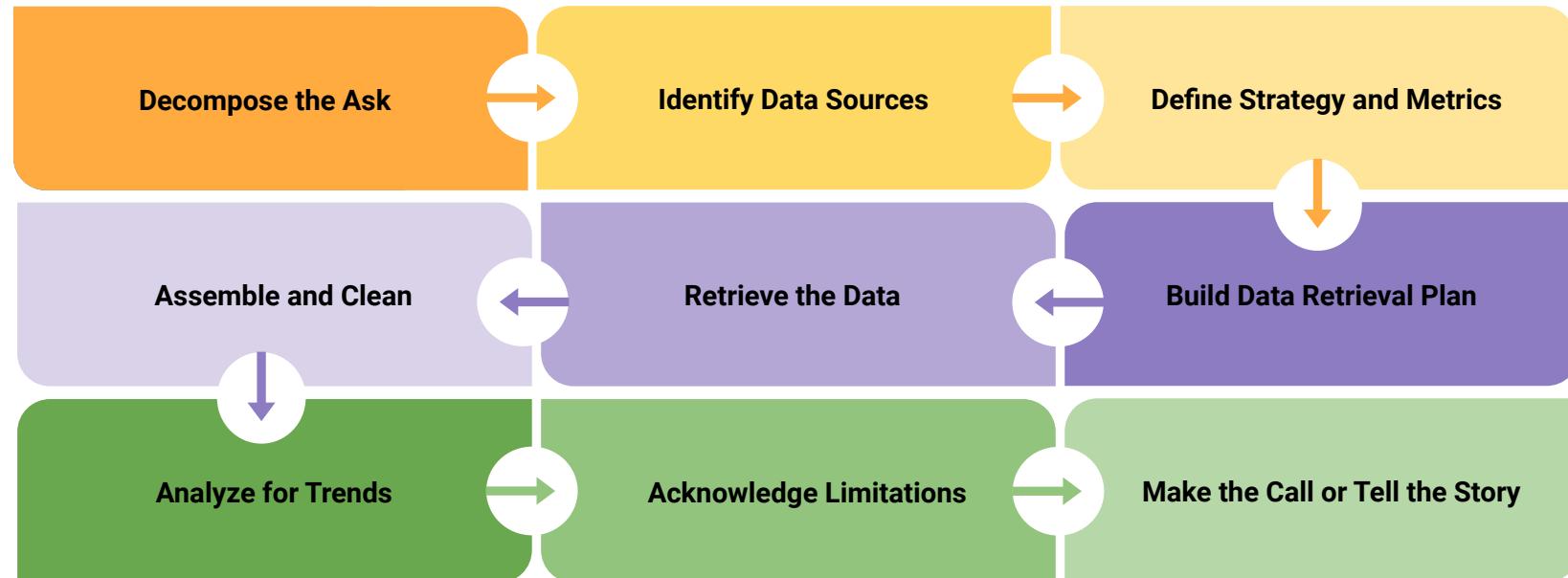
You only have your team.



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?

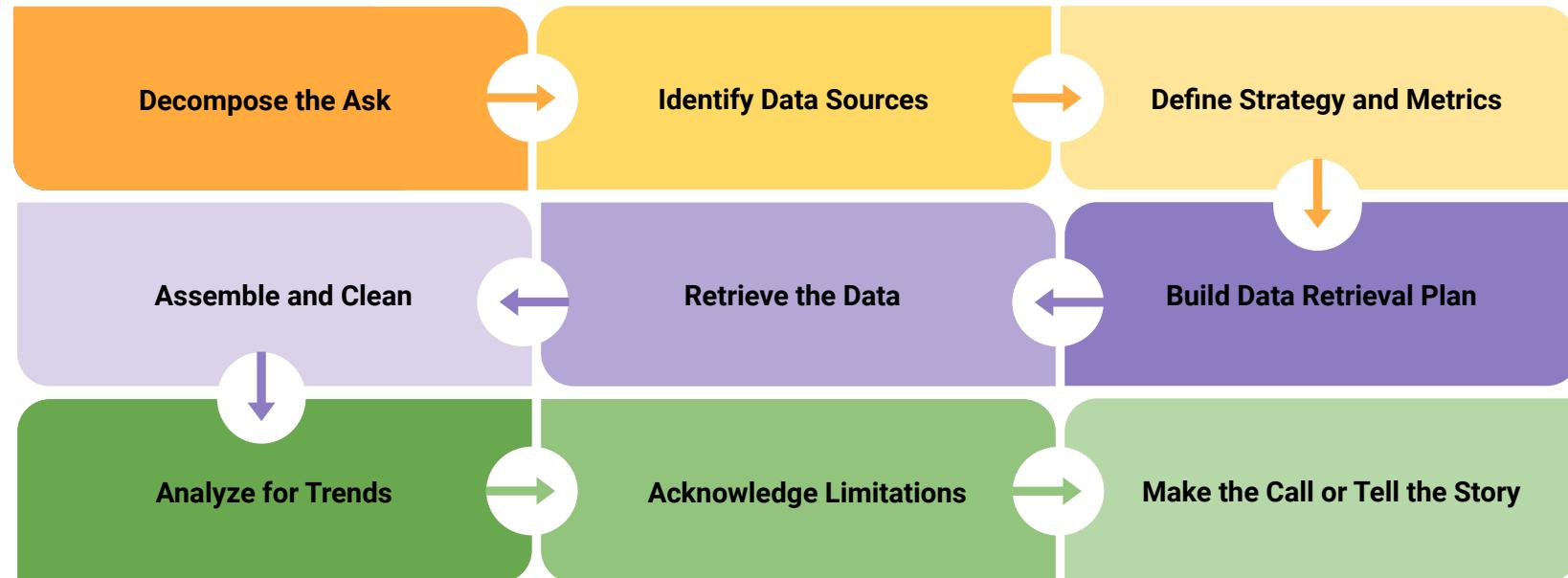


Italian and Mexican are
broad categories we are
pursuing. We will have to
narrow the scope.

Step 2: Identify Data Sources

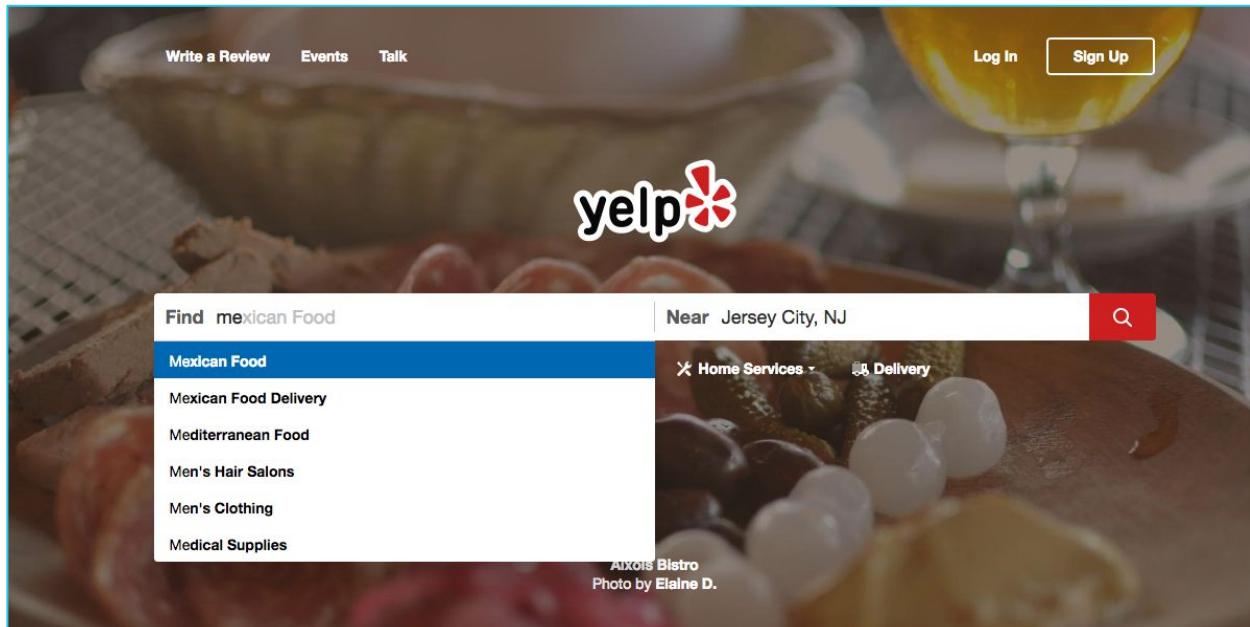
Analytics Paradigm

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?



Step 2: Identify Data Sources

The screenshot shows a Yelp search results page for "Best Italian Food Jersey City, NJ". The search bar at the top has "Near Jersey City, NJ" entered. Below the search bar, there are filters for Home Services, Restaurants, Auto Services, and More. A yellow arrow labeled "Food Type" points to the "Restaurants" filter. The main search results show two entries:

- Lorad Nia's Family Pizzeria**: Ad, 108 reviews, \$\$. Offers takeout and delivery. Description: "One of the best tasting pizza around Jersey City. Perfect sauce seasoned perfectly (not sweet out of the can taste) thin crust.... Finest cheese you can even order the Whole wheat..." Read more. A yellow arrow labeled "Review Count" points to the review count.
- Zero Otto Uno Cafe**: Ad, 54 reviews, \$\$ - Pizza, Italian. Offers takeout and delivery. Description: "Really good staff, neither too intrusive nor too dismissive, Brought our two year old and they handled it well. Ordered the pizza, met expectations. Nice to write a good review." Read more.

A yellow arrow labeled "Rating" points to the star rating icon for Zero Otto Uno Cafe. To the right of the search results is a map of New York City and surrounding areas, with numbered pins indicating various locations. A yellow arrow labeled "Locations" points to the map.



Lots of Data!

Food Type

Review Count

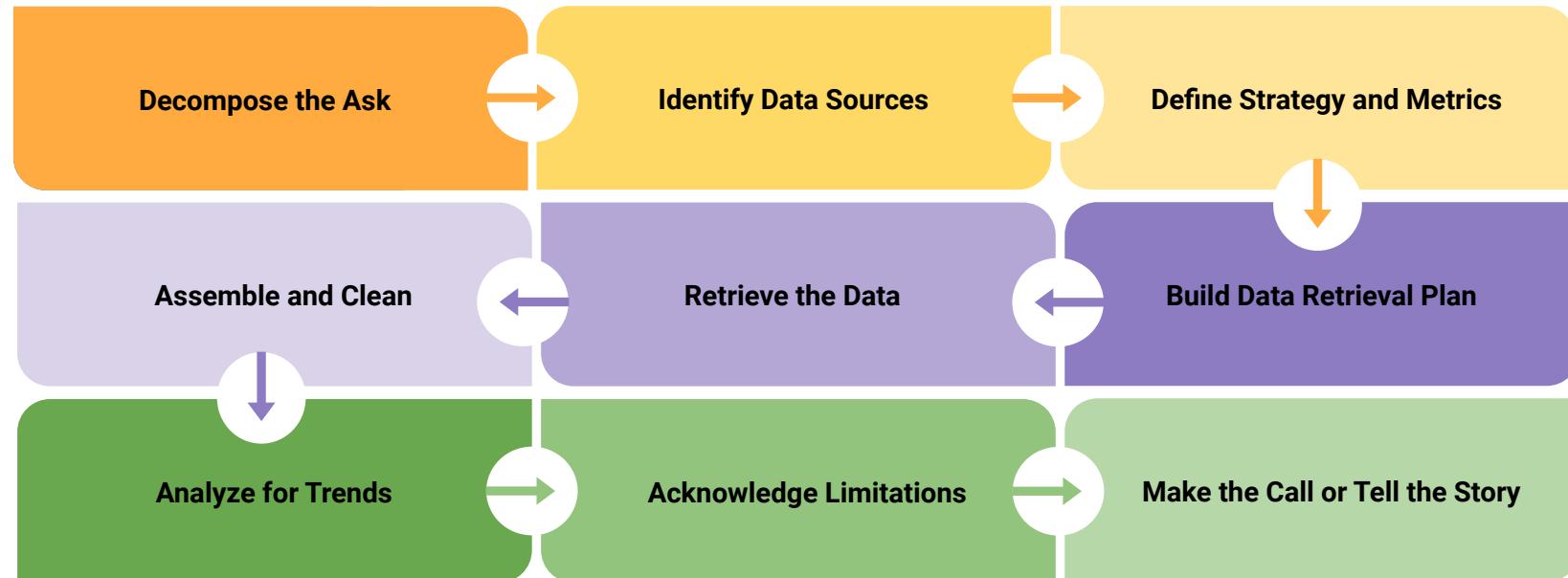
Rating

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

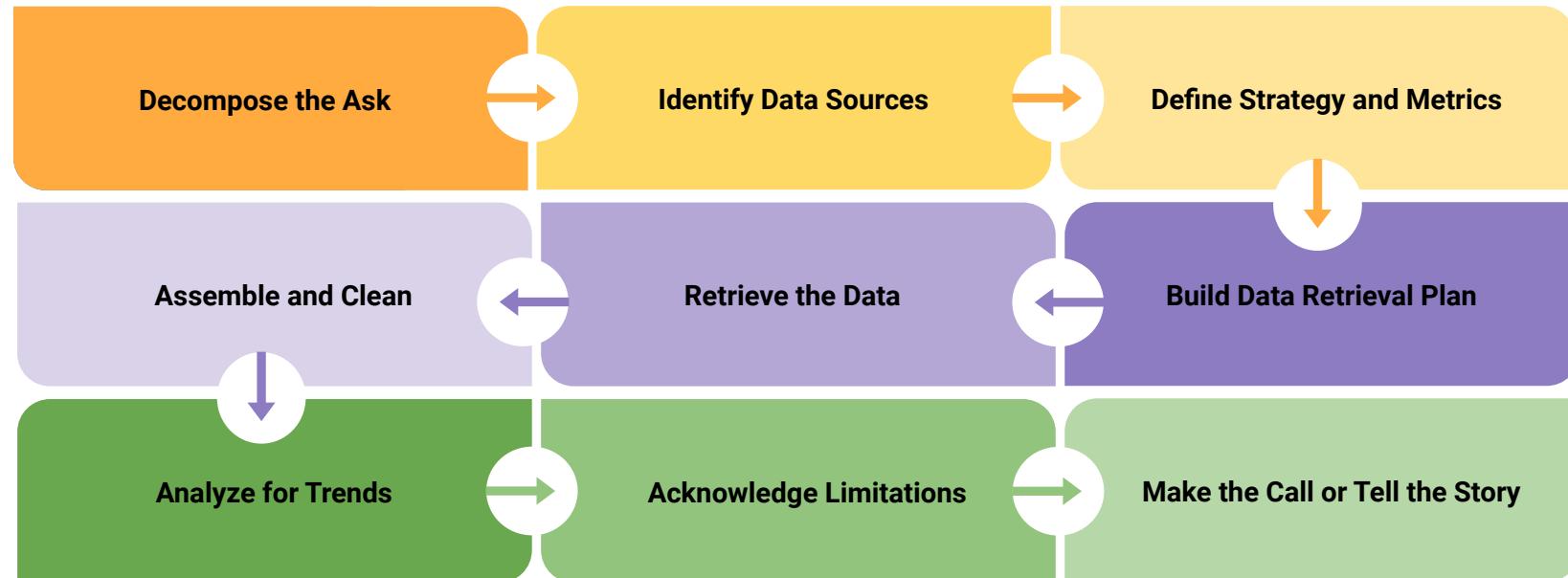
San Diego, CA	
Italian	Mexican
Restaurant	Restaurant
Restaurant	Restaurant
Restaurant	VS.
Restaurant	Restaurant
Restaurant	Restaurant

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 selector with "Near" and a dropdown menu showing "New York, NY", and a search button with a magnifying glass icon. The three instances are vertically aligned, representing the same search being performed in three different locations: New York, Denver, and Oklahoma City.

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 top navigation bar includes the Yelp logo, 'Fusion', 'Fusion API', 'GraphQL', and 'Manage App' buttons, along with user profile and notification icons.

General

/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!

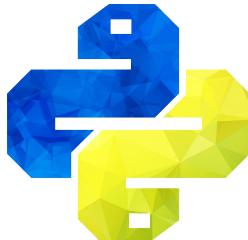
Response Body

```
{  
  "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/MmgtaSP3l_t4tPCLiiAsCg/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": -122.42820739746094  
    }  
  }  
}
```



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

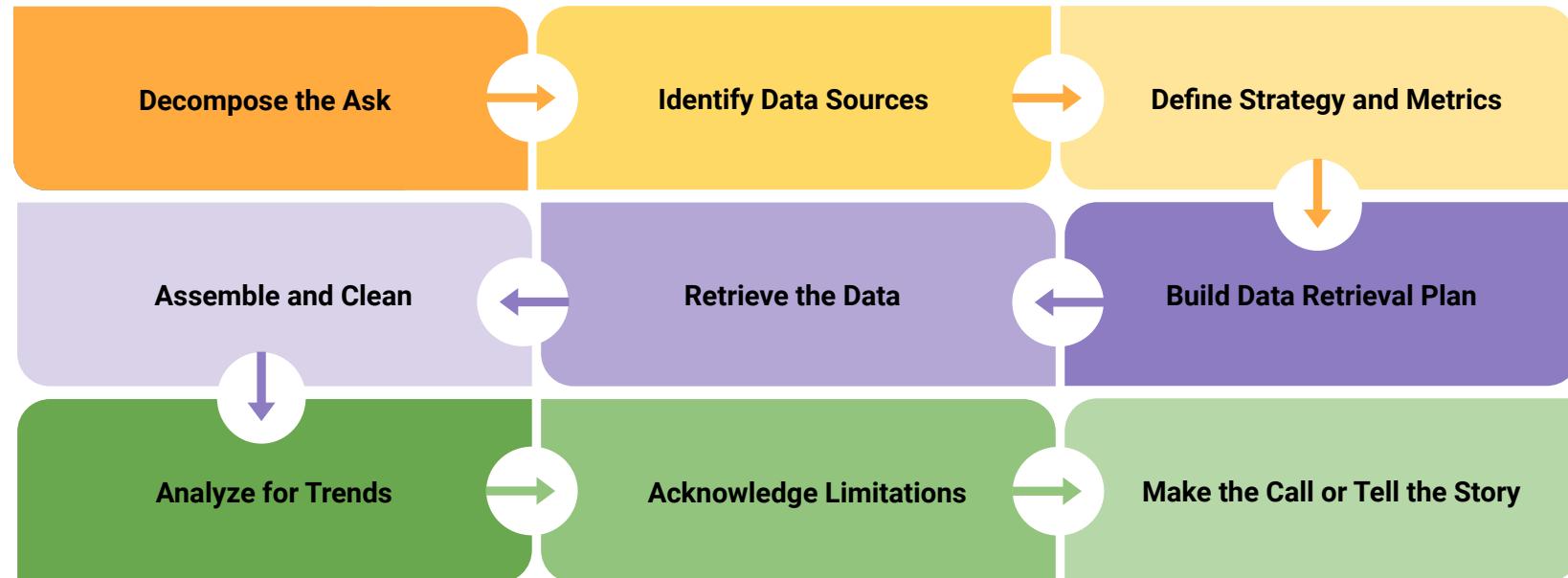
30301	
Italian	Mexican
Restaurant	Restaurant

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 computer monitors with various icons. The red header bar contains the yelp logo, "Fusion", "Fusion API", "GraphQL", and "Manage App". Below the browser are social sharing icons: Share, f, t, in, and a user profile icon.

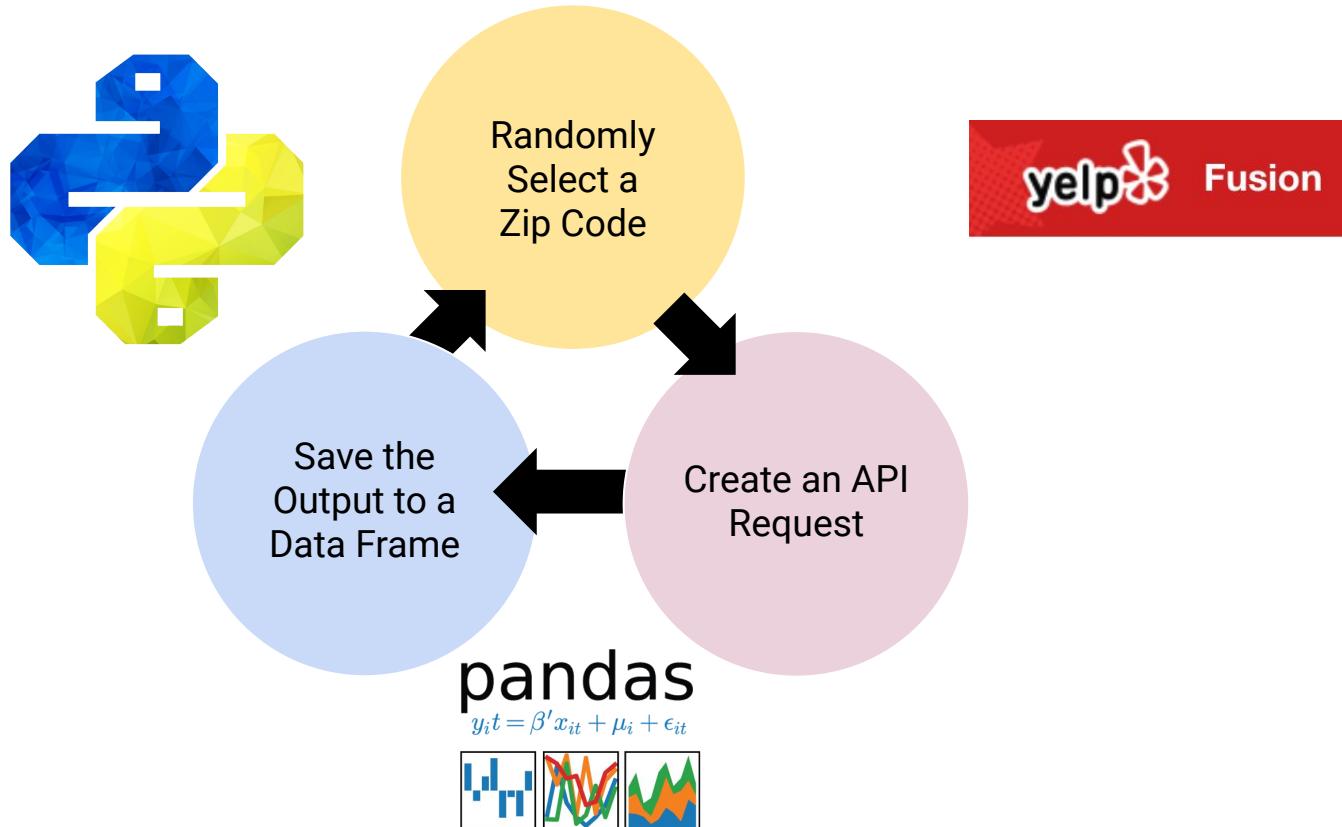
Step 5: Retrieve the Data

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... Params Send Save

Headers (1) Authorization Body Pre-request Script Tests Code

Key	Value	Description	...	Bulk Edit	Presets ▾
<input checked="" type="checkbox"/> Authorization	Bearer gl6k6JmewUhjMVbvoI2x4Bz_NRIEggSjIGbTaejmzbvBJXg 36F...				
New key	Value	Description			

Body Cookies Headers (10) Tests Status: 200 OK Time: 665 ms

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": "43210",  
38        "country": "US",  
39        "display_address": ["1000 W Broad St", "Columbus, OH 43210"]  
40      }  
41    }  
42  ]  
43}  
44
```

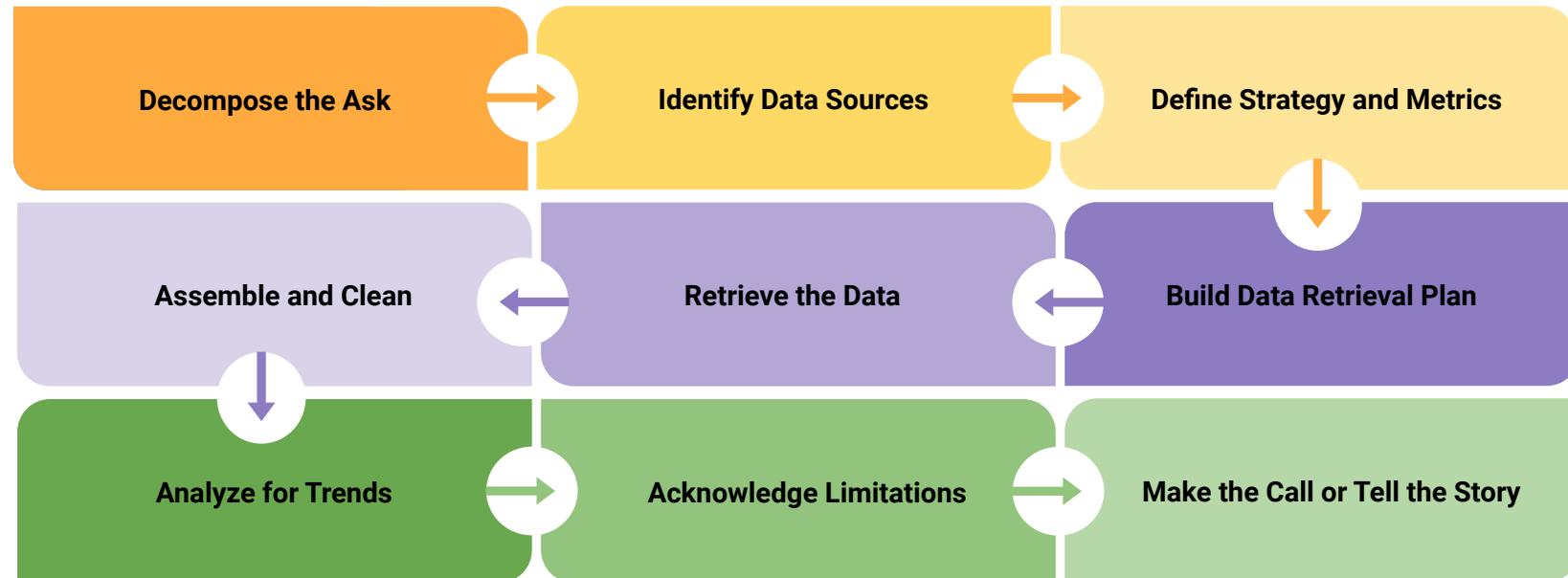


Each of these URLs holds a piece of our answer.

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.

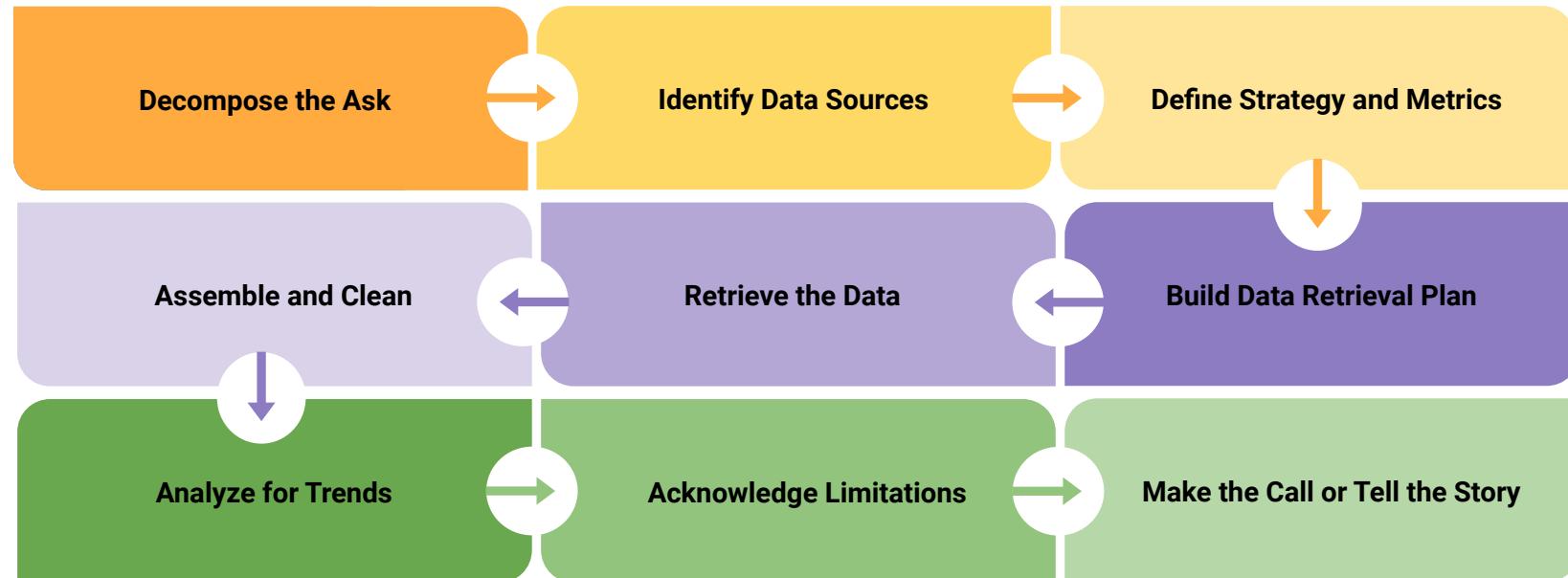
```
# Combine DataFrames into a single DataFrame  
combined_data = pd.merge(mexican_data, italian_data, on="Zip Code")  
combined_data.head()
```

	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)

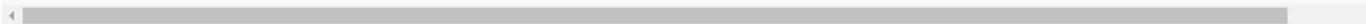
It's Close:

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"]})

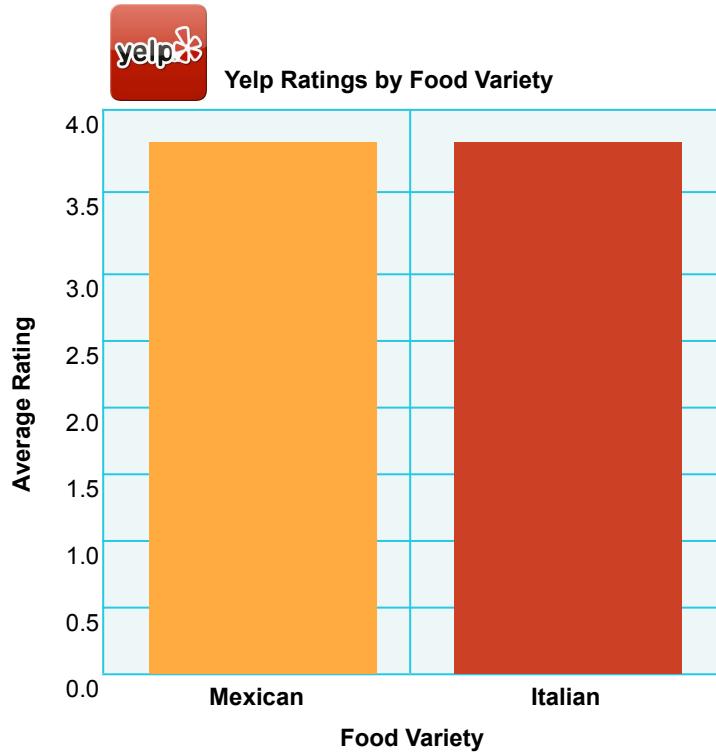
final_summary = pd.concat([mexican_summary, italian_summary])
final_summary
```



	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**.

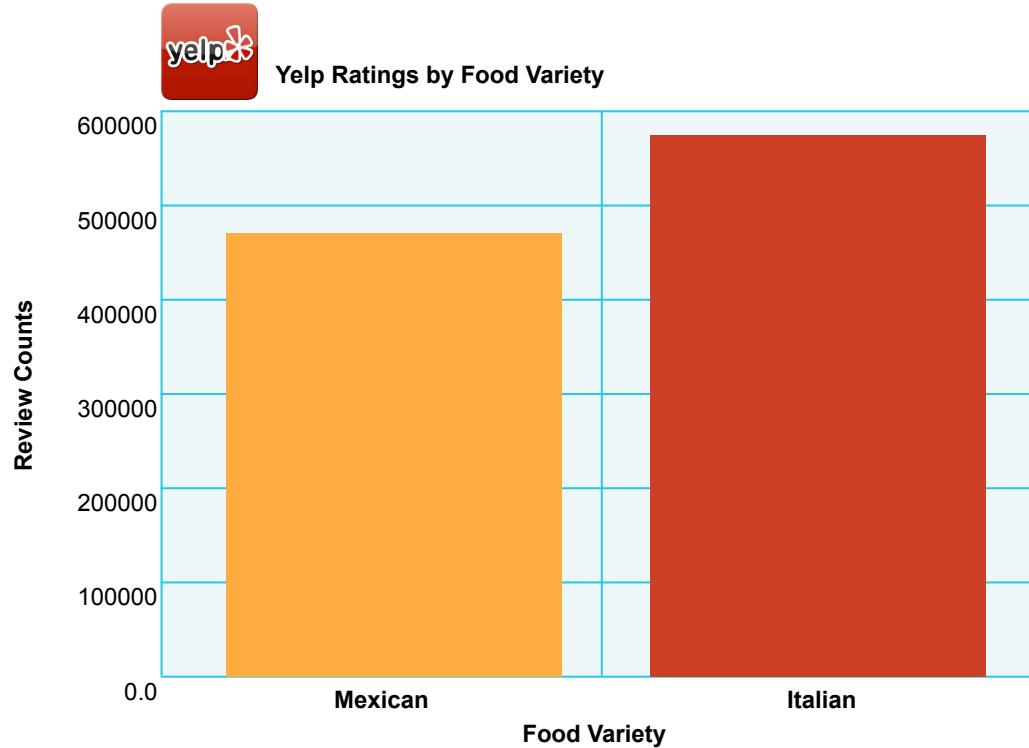


=



Analyze for Trends (Ratings)

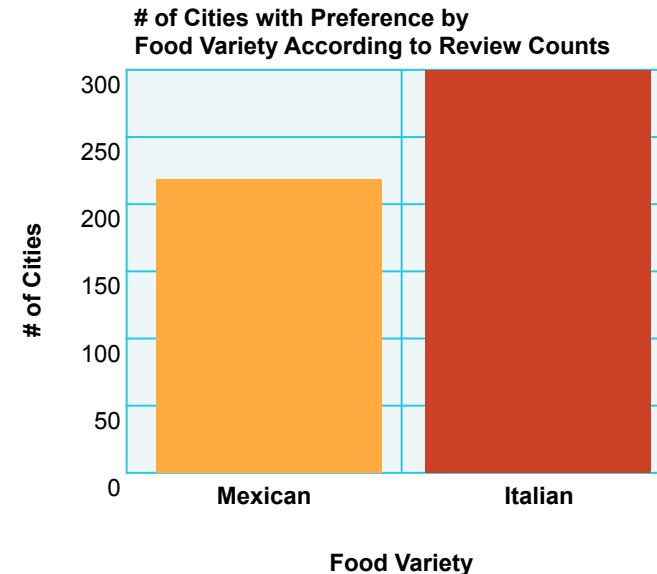
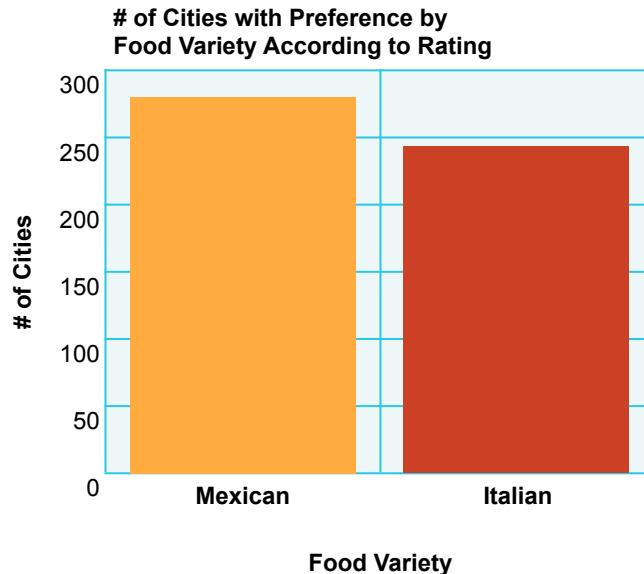
Yelpers seem to significantly **review more Italian** restaurants.



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

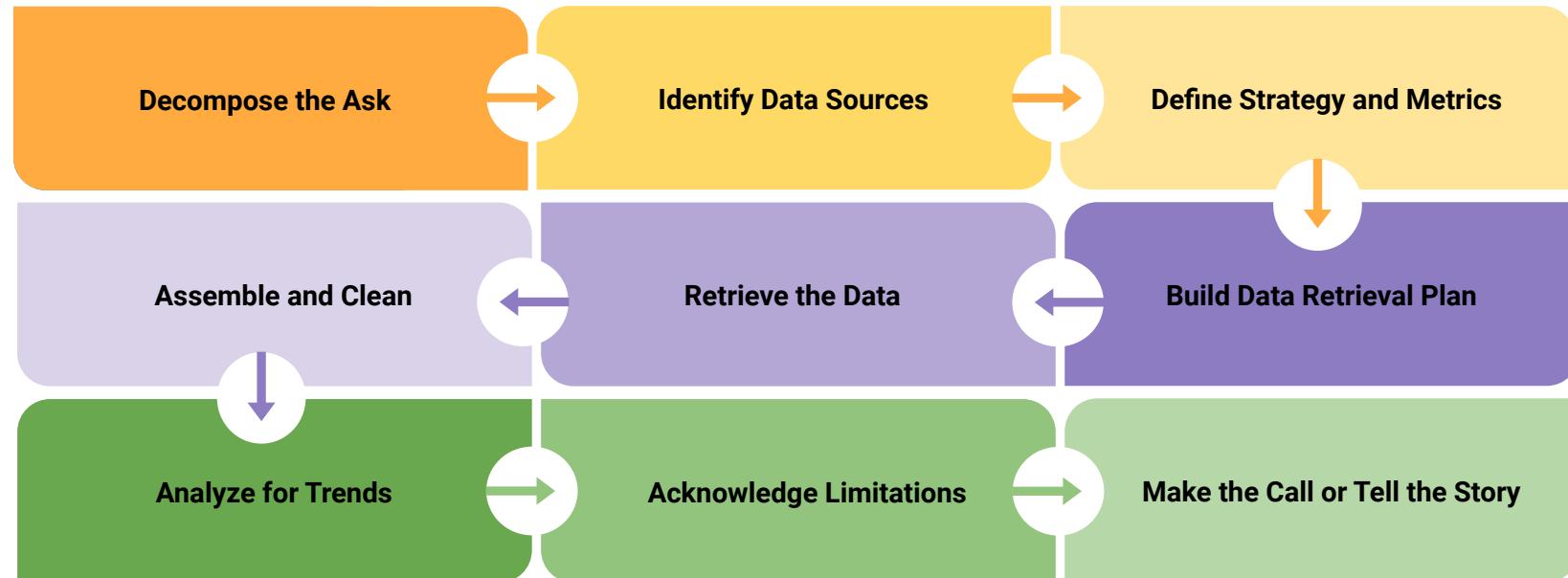


The difference in review count is **not statistically significant**.

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.



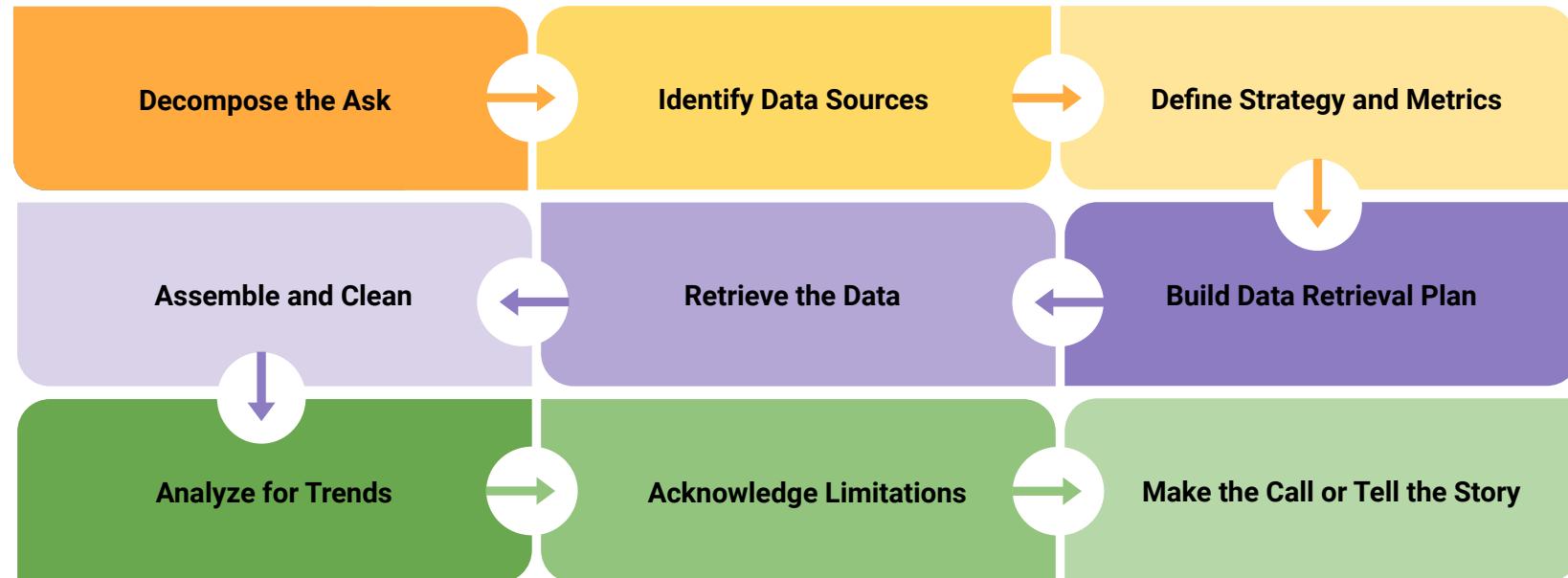
Limitations of Analysis



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).

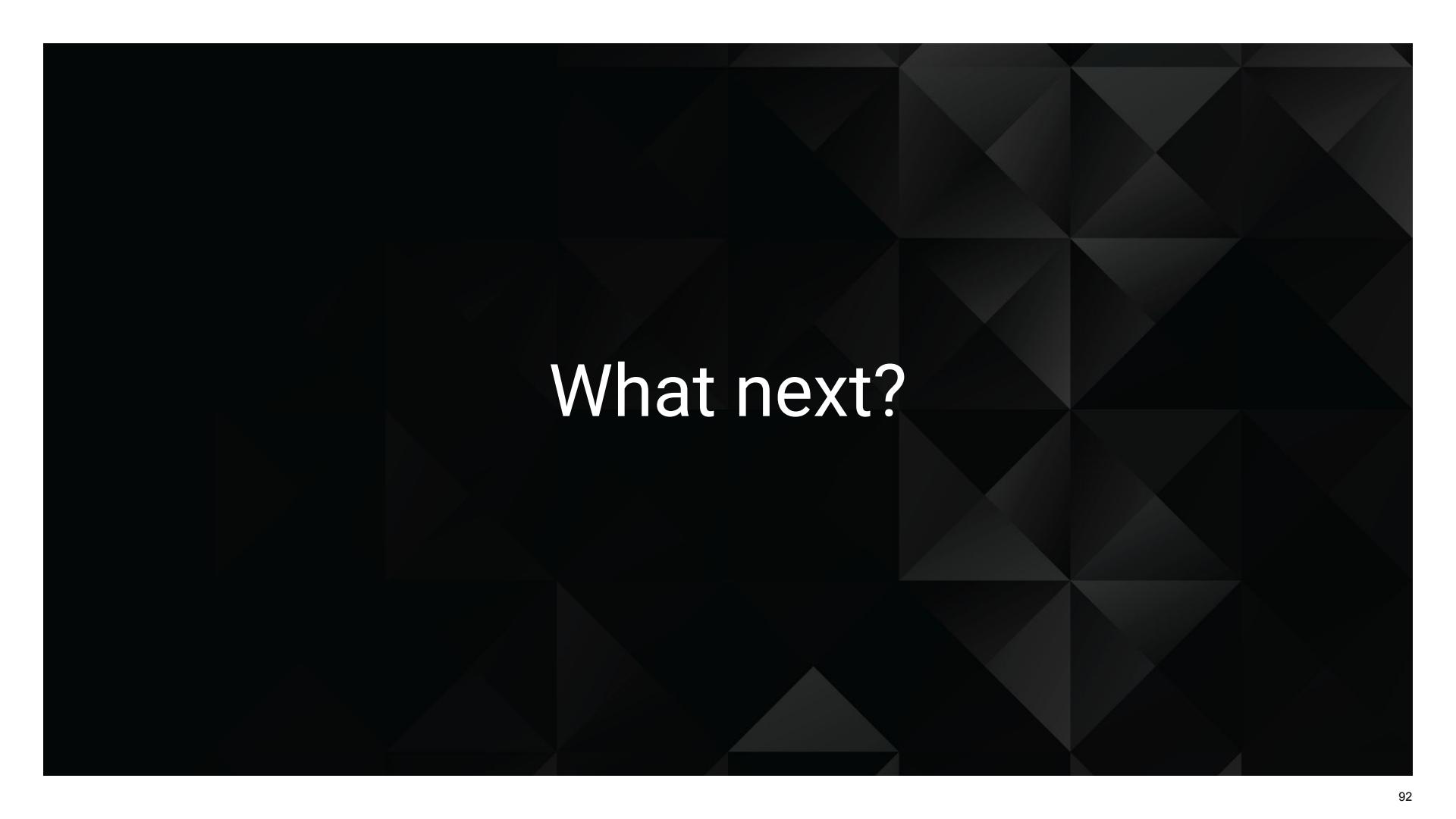


This may indicate there is an increased interest in visiting Italian restaurants at an experiential level. Or it may merely suggest that Yelp users enjoy writing reviews of Italian restaurants more than Mexican restaurants.

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 on **Tuesday** and **Thursday**