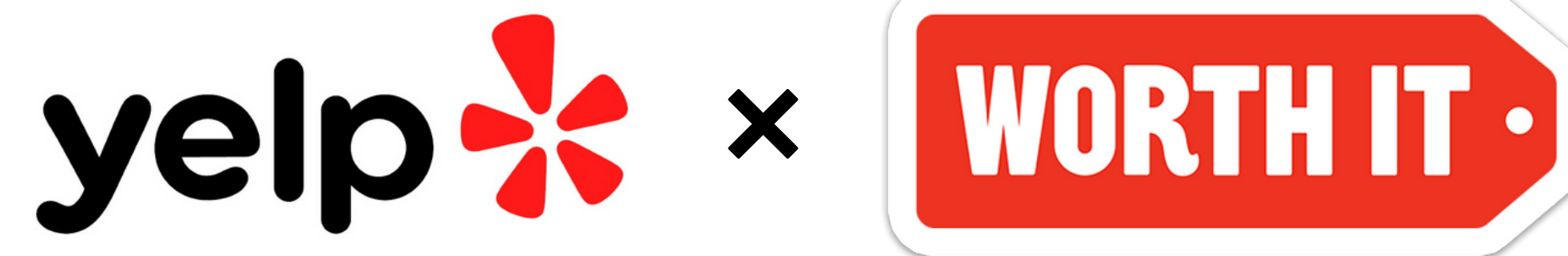




PROJECT 2: ETL



GUARDIANS OF ETL



Negin Djalali
Shaun Gutierrez
Trevor Jones



PURPOSE & APPLICATIONS

Our group aims to extract key information, such as ratings, price ranges, and more from Yelp to better understand performance, customer engagement, and uncover how the information gathered may correlate.



- 01** Gain insight to customer feedback by investigating star ratings
- 02** Understand product value, spending habits through price ranges
- 03** Make knowledgeable and well informed decisions with information on direct market competition

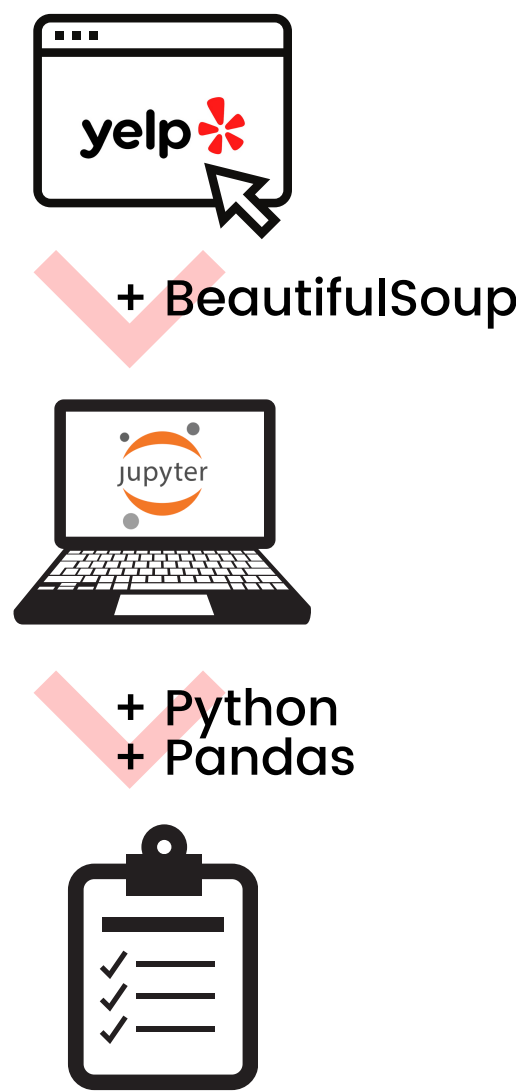
- 01** Is there a correlation between pricing and customer rating?
- 02** Is there a correlation between the volume of ratings and customer rating?
- 03** Which establishment has the most reviews? highest rated?
- 04** What does the landscape of competitors look like in terms of cuisine? price range?

DECOMPOSE THE ASK:



OVERVIEW

EXTRACT WORKFLOW



TRANSFORM WORKFLOW

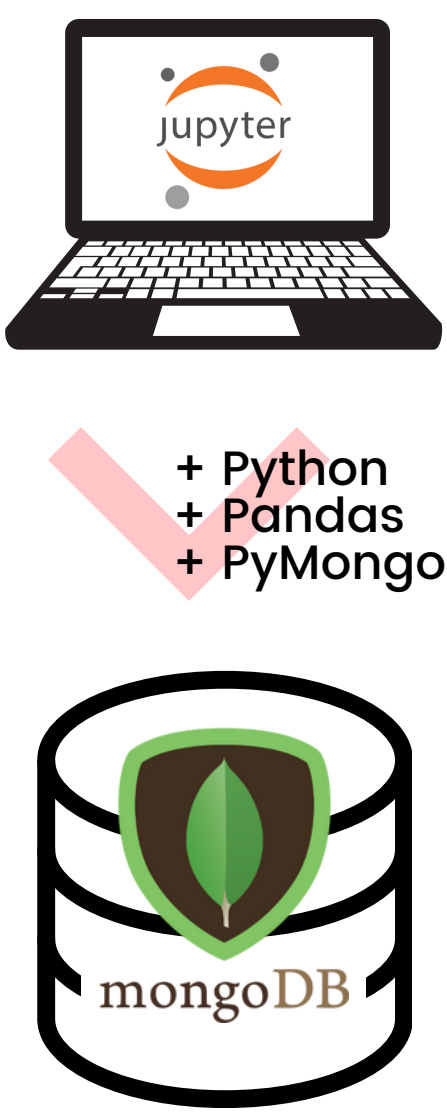
The Transform Workflow diagram shows two tables. The top table has columns: name, stars, price, and cuisine. The bottom table has columns: name, cuisine, price, and rating. Both tables list various restaurants and their attributes.

| name | stars | price | cuisine |
|---------------------------|-----------------|--------|--------------------|
| GRANVILLE | 4.5 star rating | \$\$ | American (New) |
| Soulmate. | 4 star rating | \$\$\$ | Mediterranean |
| Employees Only | 4 star rating | \$\$\$ | Cocktail Bars |
| Norah | 4 star rating | \$\$ | American (New) |
| Tu Madre - West Hollywood | 4 star rating | \$\$ | Mexican |
| ... | ... | ... | ... |
| Phorage WeHo | 4 star rating | \$\$ | Vietnamese |
| Angelini Osteria | 4 star rating | \$\$\$ | Italian |
| Citizen Public Market | 4.5 star rating | \$\$ | Food Court |
| Chibiscus Asian Cafe | 4.5 star rating | \$\$ | Asian Fusion |
| Eggslut | 3.5 star rating | \$ | Breakfast & Brunch |

| name | cuisine | price | rating |
|---------------------------|--------------------|-------|--------|
| GRANVILLE | American (New) | 2.0 | 4.5 |
| Soulmate. | Mediterranean | 3.0 | 4.0 |
| Employees Only | Cocktail Bars | 3.0 | 4.0 |
| Norah | American (New) | 2.0 | 4.0 |
| Tu Madre - West Hollywood | Mexican | 2.0 | 4.0 |
| ... | ... | ... | ... |
| Phorage WeHo | Vietnamese | 2.0 | 4.0 |
| Angelini Osteria | Italian | 3.0 | 4.0 |
| Citizen Public Market | Food Court | 2.0 | 4.5 |
| Chibiscus Asian Cafe | Asian Fusion | 2.0 | 4.5 |
| Eggslut | Breakfast & Brunch | 1.0 | 3.5 |

- + Jupyter
- + Pandas

LOAD WORKFLOW



EXTRACTION:



+ BeautifulSoup



+ Python
+ Pandas



- 01 We sourced our data from the desktop version of Yelp
 - 02 **BeautifulSoup** was used as our HTML Parser to load the web components into our Jupyter notebook
 - 03 We identified business name, rating, price, and cuisine as the pieces of information we wanted to retrieve
 - 04 Using **JupyterLab**, we set up a loop to extract and store our data into a list of dictionaries using **Python** and **Pandas**
 - 05 The resulting dataframe is now ready for the next step: Transform
-

TRANSFORMATION:

- 01
- With our Jupyter notebook we renamed and reordered the columns using Python to what we felt was most intuitive
- 02
- Removed star rating from the output within the stars column and converted from str to float
- 03
- Replaced the \$ price rating system to int-based to be able to manipulate and visualize our data
- 04
- We stored our cleaned dataframe as a .csv file to analyze and visualize. For now, we will proceed with the next step: Load

| name | stars | price | cuisine |
|---------------------------|-----------------|--------|--------------------|
| GRANVILLE | 4.5 star rating | \$\$ | American (New) |
| Soulmate. | 4 star rating | \$\$\$ | Mediterranean |
| Employees Only | 4 star rating | \$\$\$ | Cocktail Bars |
| Norah | 4 star rating | \$\$ | American (New) |
| Tu Madre - West Hollywood | 4 star rating | \$\$ | Mexican |
| ... | ... | ... | ... |
| Phorage WeHo | 4 star rating | \$\$ | Vietnamese |
| Angelini Osteria | 4 star rating | \$\$\$ | Italian |
| Citizen Public Market | 4.5 star rating | \$\$ | Food Court |
| Chibiscus Asian Cafe | 4.5 star rating | \$\$ | Asian Fusion |
| Eggslut | 3.5 star rating | \$ | Breakfast & Brunch |

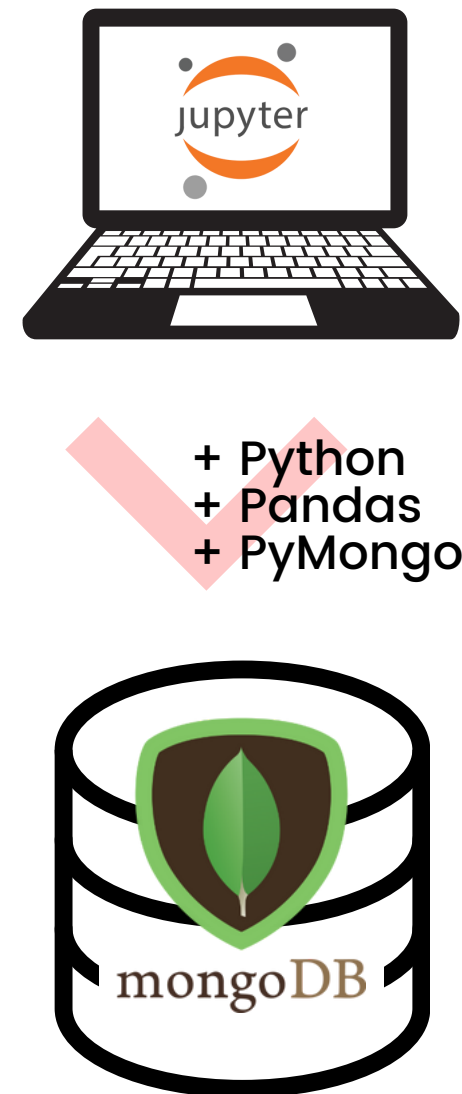
Fig. 01 – Raw Dataframe

+ Jupyter
+ Python

| name | cuisine | price | rating |
|---------------------------|--------------------|-------|--------|
| GRANVILLE | American (New) | 2.0 | 4.5 |
| Soulmate. | Mediterranean | 3.0 | 4.0 |
| Employees Only | Cocktail Bars | 3.0 | 4.0 |
| Norah | American (New) | 2.0 | 4.0 |
| Tu Madre - West Hollywood | Mexican | 2.0 | 4.0 |
| ... | ... | ... | ... |
| Phorage WeHo | Vietnamese | 2.0 | 4.0 |
| Angelini Osteria | Italian | 3.0 | 4.0 |
| Citizen Public Market | Food Court | 2.0 | 4.5 |
| Chibiscus Asian Cafe | Asian Fusion | 2.0 | 4.5 |
| Eggslut | Breakfast & Brunch | 1.0 | 3.5 |

Fig. 02 – Cleaned Dataframe

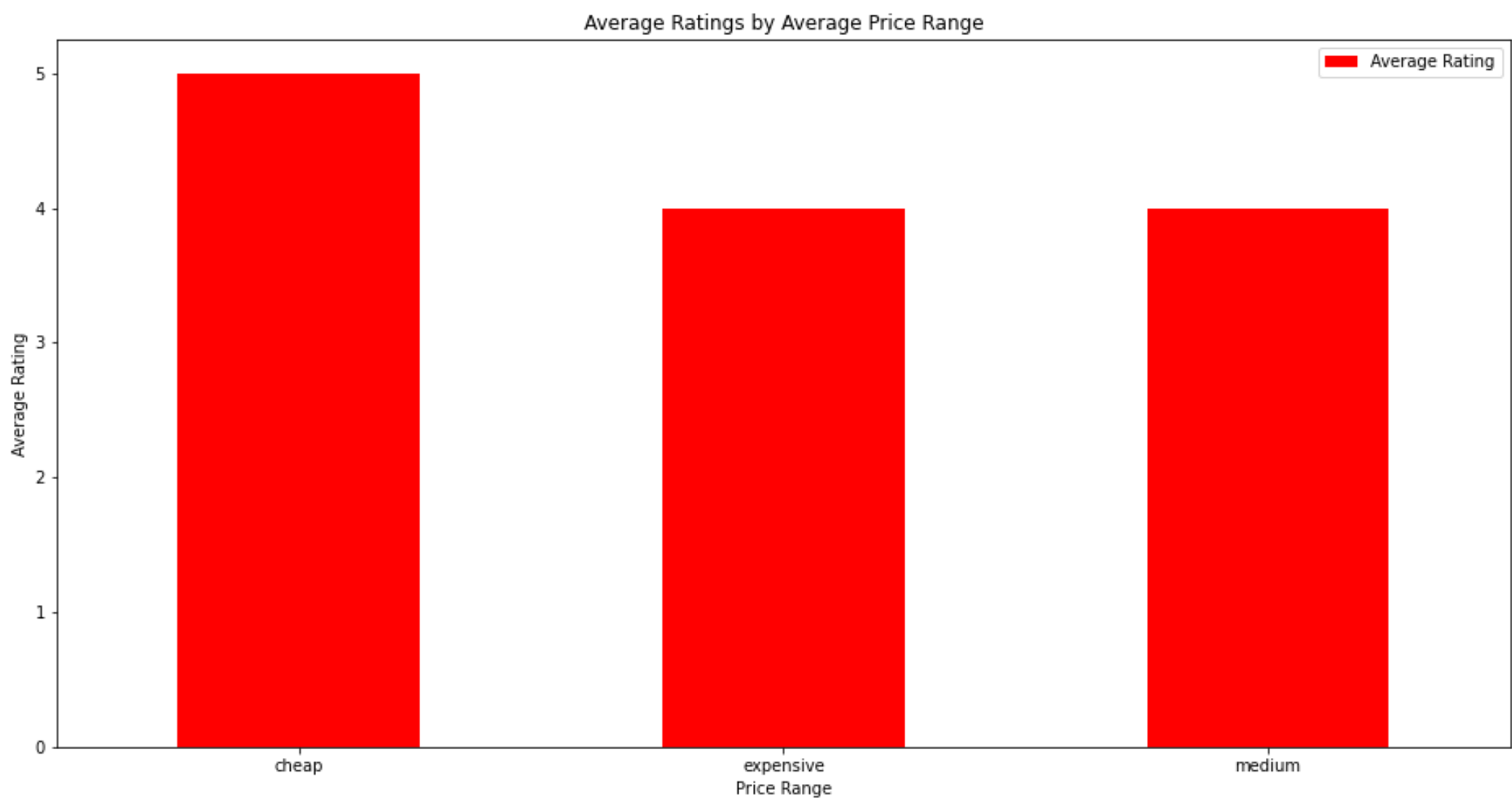
LOADING:



- 01** Using our **Jupyter notebook**, we established a connection to **MongoDB**
- 02** Then established a database within **MongoDB** and created a collection to house our dataframe
- 03** Before storing the dataframe, we needed to covert it to a dictionary using **Pandas** (.to_dict)
- 04** Once that was completed, our data was loaded into **MongoDB** using **PyMongo**

ANALYSIS:

Is there a correlation between pricing and customer rating?



CHEAP

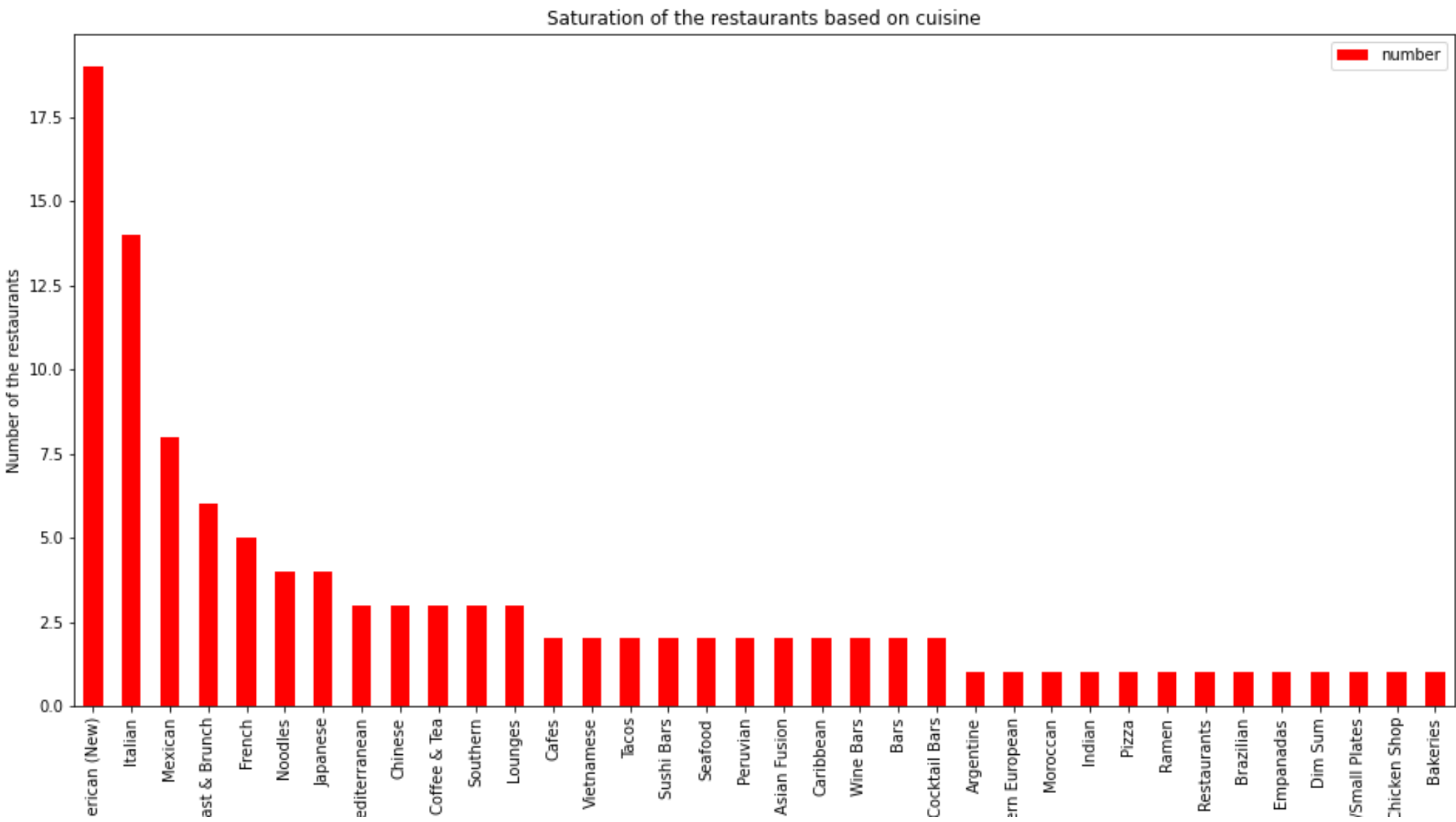
From our data we've uncovered that cheaper restaurants generated, on average, a 5-star rating. We believe this to be due to lower expectations to cheaper costing food.

NO

Our graph showed us that the city of West Hollywood has generally high ratings across the different price ranges.

ANALYSIS:

What does the landscape of competitors look like in terms of cuisine?



AMERICAN

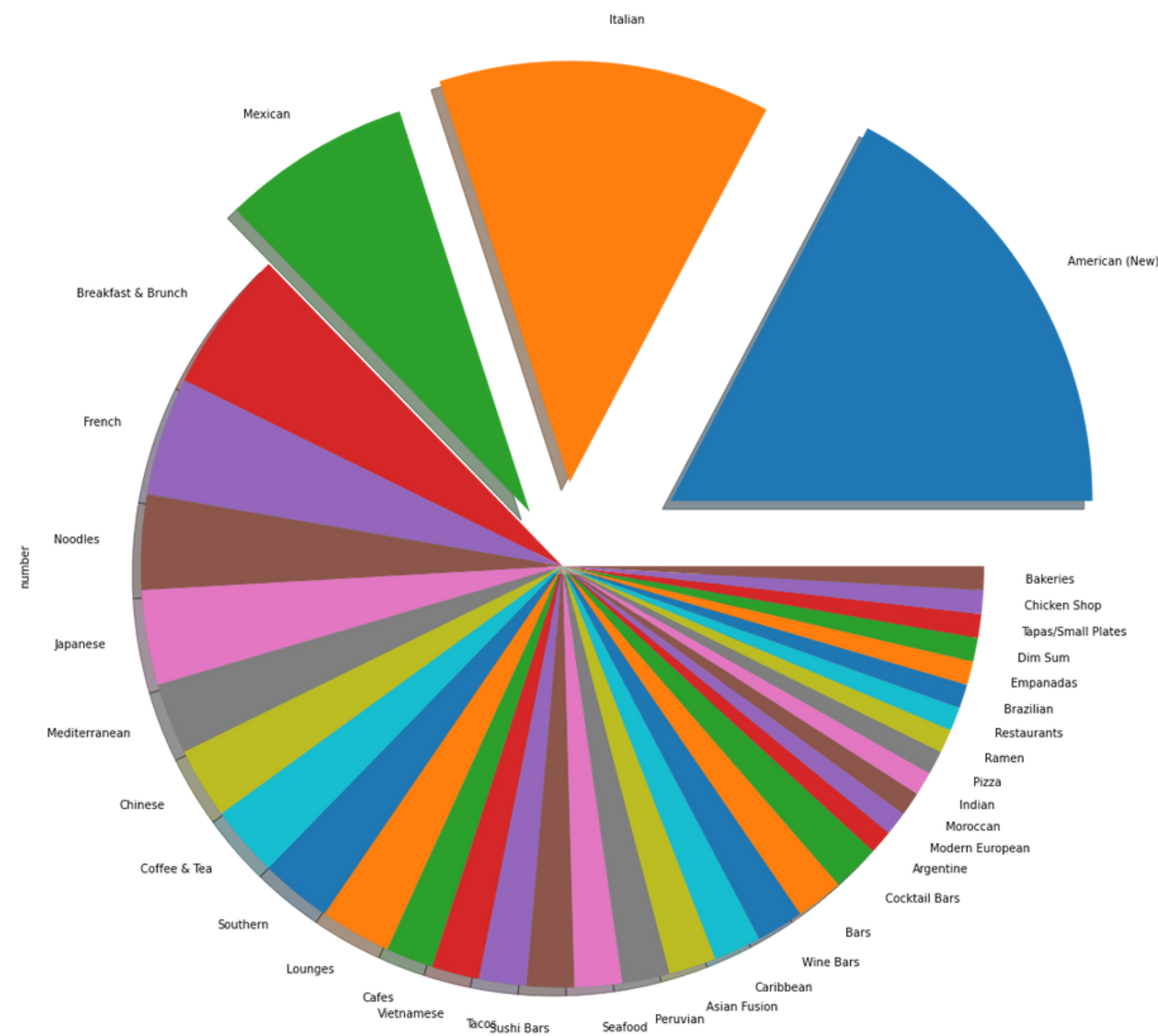
American cuisine was found to be the most prevalent in West Hollywood with a total of 19/110 establishments.

VARIETY

We found that West Hollywood has a total of 36 different cuisines. We also noticed that Argentine, Indian and Korean cuisine have very little to no representation.

ANALYSIS:

What does the landscape of competitors look like in terms of cuisine? (Part 2)

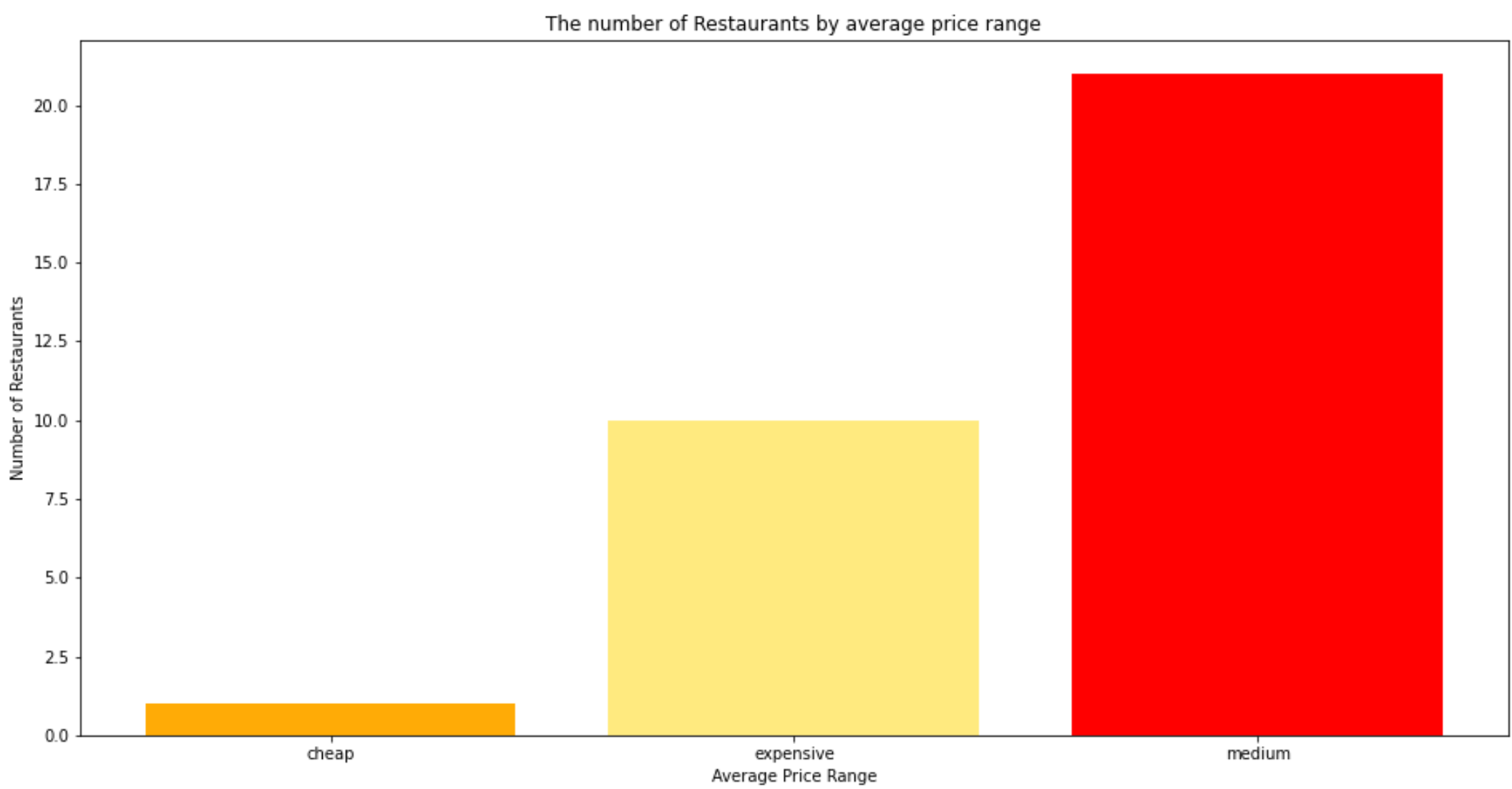


17%

We found restaurants that specialize in American cuisine dominate West Hollywood, comprising 17% of all restaurants in the area. Followed by Italian (12%), and Mexican (7%).

ANALYSIS:

What does the landscape of competitors look like in terms of price range?



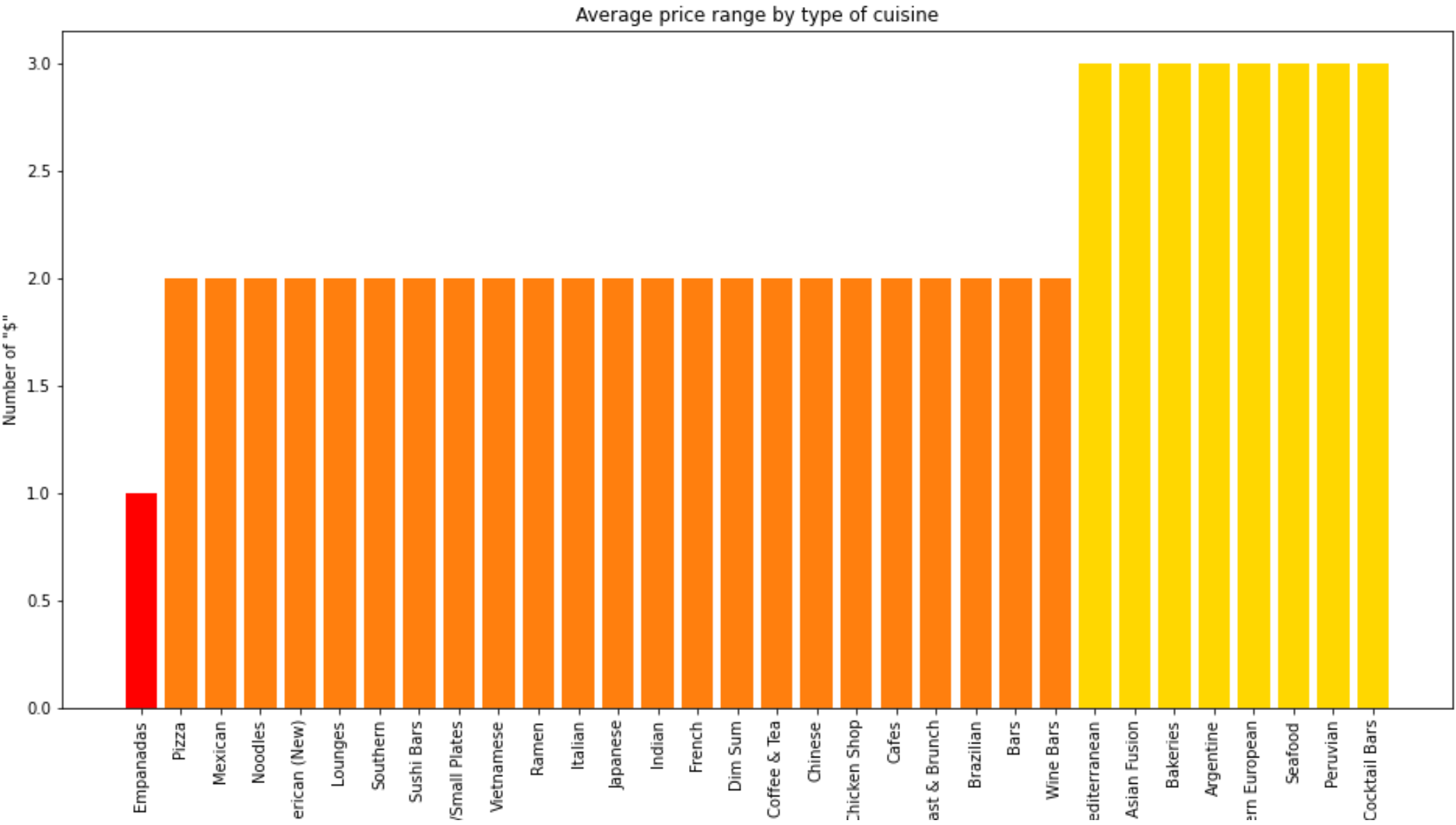
MEDIUM

From our analysis, 65% of restaurants fall within the medium price range, with Expensive (32%) and cheap (3%).

Generally speaking, we believe that medium tiered restaurants represent the best of both the cheap and expensive. Straddling the middle will allow customers to experience the quality and service that expensive restaurants have to offer. While offering a pricing structure that is more accessible to a wider audience.

ANALYSIS:

What is the average price range of each cuisine?



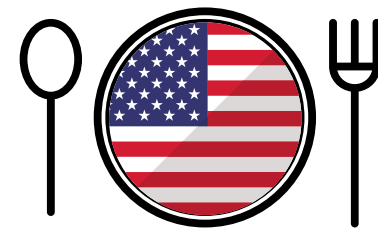
MEDIUM

We determined cuisines that fall within the medium price range account for 71% of all cuisines.

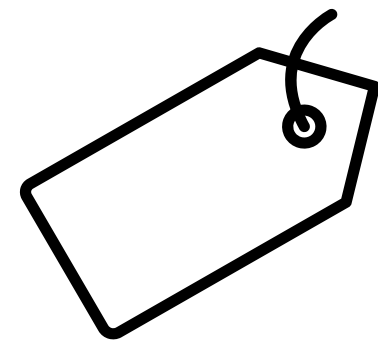
Next was the expensive tier making up 25% of the cuisine landscape.

Lastly, the cheap price range represented 3% of total cuisines.

LET'S RECAP!



AMERICAN CUISINE WAS THE MOST PREVALENT, REPRESENTING 17% OF ALL RESTAURANTS IN WEST HOLLYWOOD



IN TERMS OF PRICING, MEDIUM-TIERED ESTABLISHMENTS REPRESENTED 65% OF ALL RESTAURANTS BUT REPRESENTED 71% OF ALL CUISINES.



CHEAP-TIERED RESTAURANTS HAD THE HIGHEST AVERAGE RATINGS AT 5 STARS

LIMITATIONS/ IMPROVEMENTS

With more time we would explore the following items to improve on our project and analysis:

- 01** Expand our search area from West Hollywood to Los Angeles
- 02** Explore the number of reviews as a metric to investigate
- 03** Filter results and reclassify cuisine types to avoid overlap or unnecessary granularity



**THANK
YOU**

Negin Djalali
Shaun Gutierrez
Trevor Jones