Zomato Reviews: Analyzing the Relationship between Price, Cuisine, and Rating Factors

Authors: Blaise (Michael) Mongiardo, Sisi Shao, Kenneth Yu

Emails: mbm485@stern.nyu.edu, ss9270@stern.nyu.edu, kly237@stern.nyu.edu

Background & Introduction

Founded in 2008, Zomato is a restaurant search service based in India. Much like Yelp, the company seeks to provide users with restaurant information like cuisine and general price and to allow them to leave reviews. Zomato operates in 24 countries and thus has global footprint.

We were interested in seeing the relationship specifically between a restaurant's price and cuisine and its rating. Certain cuisines tend to be considered "high-end" and therefore command a higher price point for a meal. Moreover, because price tends to signal quality, especially in the restaurant industry, the food and service should be relatively better.

Through analysis of Zomato data, we hope to test two hypotheses as follows: restaurants with cuisines that connotate fine dining (specifically French and Japanese) will have a higher price point and thus rating, and restaurants with a higher price will have higher ratings left by reviewers.

Data

The data we will use comes from Kaggle (https://www.kaggle.com/shrutimehta/zomato-restaurants-data (https://www.kaggle.com/shrutimehta/zomato-restaurants-data)). The two data sets that will be utilized from this source are the main zomato.csv file and the country-code.xlsx file.

The author of these data sets mentions that the main file is "Analyzing the best restaurants of the major cities".

Data Set at a Glance

Here, we begin to run our code.

In [1]:

```
import pandas as pd #analyze data
import matplotlib.pyplot as plt #visualize data
import numpy as np #facilitate mathematical operations

path = 'https://github.com/kenneth-yu19/Data_Bootcamp_Final_Project/blob/master/
zomato.xlsx?raw=true'
exc_data = pd.read_excel(path) #read the excel data from source

zom = pd.DataFrame(exc_data) #convert excel data into pandas dataframe
zom
```

Out[1]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality
0	6317637	Le Petit Souffle	162	Makati City	Third Floor, Century City Mall, Kalayaan Avenu...	Century City Mall, Poblacion, Makati City
1	6304287	Izakaya Kikufuji	162	Makati City	Little Tokyo, 2277 Chino Roces Avenue, Legaspi...	Little Tokyo, Legaspi Village, Makati City
2	6300002	Heat - Edsa Shangri-La	162	Mandaluyong City	Edsa Shangri-La, 1 Garden Way, Ortigas, Mandal...	Edsa Shangri-La, Ortigas, Mandaluyong City
3	6318506	Ooma	162	Mandaluyong City	Third Floor, Mega Fashion Hall, SM Megamall, O...	SM Megamall, Ortigas, Mandaluyong City
4	6314302	Sambo Kojin	162	Mandaluyong City	Third Floor, Mega Atrium, SM Megamall, Ortigas...	SM Megamall, Ortigas, Mandaluyong City
5	18189371	Din Tai Fung	162	Mandaluyong City	Ground Floor, Mega Fashion Hall, SM Megamall, ...	SM Megamall, Ortigas, Mandaluyong City

6	6300781	Buffet 101	162	Pasay City	Building K, SM By The Bay, Sunset Boulevard, M...	SM by the Bay, Mall of Asia Complex, Pasay City
7	6301290	Vikings	162	Pasay City	Building B, By The Bay, Seaside Boulevard, Mal...	SM by the Bay, Mall of Asia Complex, Pasay City
8	6300010	Spiral - Sofitel Philippine Plaza Manila	162	Pasay City	Plaza Level, Sofitel Philippine Plaza Manila, ...	Sofitel Philippine Plaza Manila, Pasay City
9	6314987	Locavore	162	Pasig City	Brixton Technology Center, 10 Brixton Street, ...	Kapitolyo
10	6309903	Silantro Fil-Mex	162	Pasig City	75 East Capitol Drive, Kapitolyo, Pasig City	Kapitolyo
11	6309455	Mad Mark's Creamery & Good Eats	162	Pasig City	23 East Capitol Drive, Kapitolyo, Pasig City	Kapitolyo
12	6318433	Silantro Fil-Mex	162	Quezon City	Second Floor, UP Town Center, Katipunan Avenue...	UP Town Center, Diliman, Quezon City
13	6310470	Guevarra's	162	San Juan City	387 P. Guevarra Corner Argonne Street, Additio...	Addition Hills
14	6314605	Sodam Korean Restaurant	162	San Juan City	17 J. Abad Santos Drive, Little Baguio, San Ju...	Little Baguio
15	18185059	Cafe Arabelle	162	Santa Rosa	Ayala Mall, Solenad, Nuvali, Santa Rosa - Taga...	Nuvali, Don Jose, Santa Rosa

16	18182702	Nonna's Pasta & Pizzeria	162	Santa Rosa	Ground Floor, Building G, Solenad 3, Nuvali, D...	Solenad 3, Don Jose, Santa Rosa
17	6318213	Balay Dako	162	Tagaytay City	Aguinaldo Highway, Tagaytay City	Tagaytay City
18	18255654	Hobing Korean Dessert Cafe	162	Taguig City	Third Floor, BGC Stopover Pavillion, Rizal Dri...	BGC Stopover Pavillion, Bonifacio Global City
19	6308205	Wildflour Cafe + Bakery	162	Taguig City	Ground Floor, Netlima Building, 4th Avenue Cor...	Bonifacio Global City
20	6315438	NIU by Vikings	162	Taguig City	Sixth Floor, SM Aura Premier, C5 Road Corner 2...	SM Aura Premier, Bonifacio Global City, Taguig...
21	6310406	The Food Hall by Todd English	162	Taguig City	Fifth Floor, SM Aura Premier, C5 Corner 26th S...	SM Aura Premier, Bonifacio Global City, Taguig...
22	6600681	Chez Michou	30	Brasí_lia	SCLN, 208, Bloco A, Loja 30, Asa Norte, Brasí_lia	Asa Norte
23	6601005	Cafí© Daniel Briand	30	Brasí_lia	SCLN 104, Bloco A, Loja 26, Asa Norte, Brasí_lia	Asa Norte
24	6600292	Casa do Biscoito Mineiro	30	Brasí_lia	SCLN 210, Bloco D, Loja 36/48, Asa Norte, Bras...	Asa Norte
25	6600441	Maori	30	Brasí_lia	CLN 110, Bloco D, Loja 28, Asa Norte, Brasí_lia	Asa Norte
					SCS 214, Bloco	

26	6600970	Pizza íæ Bessa	30	Brasí_lia	C, Loja 40, Asa Sul, Brasí_lia	Asa Sul
27	6600379	Sushi Loko	30	Brasí_lia	SCS 213, Bloco C, Loja 35, Asa Sul, Brasí_lia	Asa Sul
28	6600214	Beirute	30	Brasí_lia	CLS 109, Bloco A, Loja 2/6, Asa Sul, Brasí_lia	Asa Sul
29	6601218	New Koto	30	Brasí_lia	SCS 212, Bloco B, Loja 26, Asa Sul, Brasí_lia	Asa Sul
...	...	...	...	...	...	...
9521	6003426	Liva	208	Ankara	î_ukurambar Mahallesi, Muhsin YazÛ±cÛ±oÛôlu Ca...	î_ukurambar
9522	6000549	Meôhur TavacÛ± Recep Usta	208	Ankara	Gí_zeltepe Mahallesi, Dikmen Vadisi, Hoôdere ...	Dikmen
9523	6000871	î_ukuraÛôa SofrasÛ±	208	Ankara	Emek Mahallesi, Bosna Hersek Caddesi, No 22/C,...	Emek
9524	6004011	Gaga Manjero	208	Ankara	Gazi Osman Paôa Mahallesi, Filistin Caddesi, ...	Gazi Osman Paôa
9525	6000409	Cafemiz	208	Ankara	Gaziosmanpaôa Mahallesi, Arjantin Caddesi, No...	Gazi Osman Paôa
9526	6000019	Nusr-Et	208	Ankara	Gaziosmanpaôa Mahallesi, _ehit í_mer Haluk S...	Gazi Osman Paôa
9527	6001537	Kebap 49	208	Ankara	Remzi OÛôuz ArÛ±k Mahallesi, TunalÛ± Hilmi	KavaklÛ±dere

					Cad...	
9528	6003668	Timboo Cafe	208	Ankara	Kentpark AVM, Kat -1, Mustafa Kemal Mahallesi,...	Kentpark AVM, îiniversiteler, î‚ankaya
9529	6001748	Meôhur í_zí_elik Aspava	208	Ankara	Kí_í_í_k Esat Mahallesi, Esat Caddesi, No 110/...	Kí_í_í_k Esat
9530	6001757	YÛ±ldÛ±z Aspava	208	Ankara	Kí_í_í_k Esat Mahallesi, Esat Caddesi, No 110/...	Kí_í_í_k Esat
9531	6000747	The Bigos	208	Ankara	Mahallesi, Selanik 2 Caddesi, No 61/A, î‚ankay...	KÛ±zÛ±lay
9532	6002025	MasabaôÛ±	208	Ankara	Kocatepe Mahallesi, Mithatpaôa Caddesi, No 62...	KÛ±zÛ±lay
9533	6003879	Zigana Pide	208	Ankara	Macun Mahallesi, Erciyes ÛÁôyerleri Sitesi, 2...	Macunkí_y
9534	6004089	Dí_veroÛôlu	208	Ankara	Maltepe Mahallesi, Gení_lik Caddesi, No 28, î‚...	Maltepe
9535	6000921	Dí_veroÛôlu	208	Ankara	îimitkí_y Mahallesi, 2432. Cadde (8. Cadde), N...	îimitkí_y
9536	6004813	Pizza ÛÁl Forno	208	Ankara	YÛ±ldÛ±zevler Mahallesi, 720. Sokak, No 2/B, í...	YÛ±ldÛ±zevler
					AsmalÛ±mescit	

9537	5904116	J'adore Chocolatier	208	ÛÁstanbul	Mahallesi, ÛÁstiklal Caddesi, Em...	AsmalÛ±mescit
9538	5901782	Starbucks	208	ÛÁstanbul	Bebek Mahallesi, Cevdetpaôa Caddesi, No 30/A,...	Bebek
9539	5902117	Valonia	208	ÛÁstanbul	Tí_rkali Mahallesi, Ihlamurdere Caddesi, No 40...	Beôiktaô Merkez
9540	5927248	Draft Gastro Pub	208	ÛÁstanbul	Caddebostan Mahallesi, BaÛôdat Caddesi, No 349...	Caddebostan
9541	5905215	Emirgan Sí_tiô	208	ÛÁstanbul	Emirgan Mahallesi, SakÛ±p SabancÛ± Caddesi, No...	Emirgí¢n
9542	5926979	Leman Kí_ltí_r	208	ÛÁstanbul	CaferaÛôa Mahallesi, Neôet í_mer Sokak, No 9/...	KadÛ±kí_y Merkez
9543	5916085	Dem Karakí_y	208	ÛÁstanbul	Kemankeô Karamustafa Paôa Mahallesi, Hoca Ta...	Karakí_y
9544	5915547	Karakí_y Gí_llí_oÛôlu	208	ÛÁstanbul	Kemankeô Karamustafa Paôa Mahallesi, RÛ±htÛ±...	Karakí_y
9545	5915054	Baltazar	208	ÛÁstanbul	Kemankeô Karamustafa Paôa Mahallesi, KÛ±lÛ±í...	Karakí_y
		NamlÛ±			Kemankeô Karamustafa	

9546	5915730	Gurme	208	ÛÁstanbul	Paôa Mahallesi, RÛ±htÛ±...	Karakí_y
9547	5908749	Ceviz AÛôacÛ±	208	ÛÁstanbul	Koôuyolu Mahallesi, Muhittin îistí_ndaÛô Cadd...	Koôuyolu
9548	5915807	Huqqa	208	ÛÁstanbul	Kuruí_eôme Mahallesi, Muallim Naci Caddesi, N...	Kuruí_eôme
9549	5916112	Aôôk Kahve	208	ÛÁstanbul	Kuruí_eôme Mahallesi, Muallim Naci Caddesi, N...	Kuruí_eôme
9550	5927402	Walter's Coffee Roastery	208	ÛÁstanbul	CafeaÛôa Mahallesi, BademaltÛ± Sokak, No 21/B,...	Moda

9551 rows × 21 columns

In [2]:

```
zom.shape
```

Out[2]:

```
(9551, 21)
```

Here, we see that our DataFrame has 21 columns and 9551 rows. This means that the original data set analyzes 9,551 restaurants based on 21 different qualities. Looking through the columns, we see that not all of them are necessary for our data analysis to test our various hypotheses, so we would like to remove a few to clean up our DataFrame.

Cleaning Data

In [3]:

```
zom = zom.drop(columns=['Address','Locality','Locality Verbose','Longitude','Lat
itude','Has Table booking','Has Online delivery','Is delivering now','Switch to
order menu'])
#removing unnecessary columns from df

zom
```

Out[3]:

	Restaurant ID	Restaurant Name	Country Code	City	Cuisines	Average Cost for two	Curren
0	6317637	Le Petit Souffle	162	Makati City	French, Japanese, Desserts	1100	Botswa Pula(P)
1	6304287	Izakaya Kikufuji	162	Makati City	Japanese	1200	Botswa Pula(P)
2	6300002	Heat - Edsa Shangri-La	162	Mandaluyong City	Seafood, Asian, Filipino, Indian	4000	Botswa Pula(P)
3	6318506	Ooma	162	Mandaluyong City	Japanese, Sushi	1500	Botswa Pula(P)
4	6314302	Sambo Kojin	162	Mandaluyong City	Japanese, Korean	1500	Botswa Pula(P)
5	18189371	Din Tai Fung	162	Mandaluyong City	Chinese	1000	Botswa Pula(P)
6	6300781	Buffet 101	162	Pasay City	Asian, European	2000	Botswa Pula(P)
7	6301290	Vikings	162	Pasay City	Seafood, Filipino, Asian, European	2000	Botswa Pula(P)
8	6300010	Spiral - Sofitel Philippine Plaza Manila	162	Pasay City	European, Asian, Indian	6000	Botswa Pula(P)
9	6314987	Locavore	162	Pasig City	Filipino	1100	Botswa Pula(P)
		Silantro Fil-			Filipino,		Botswa

10	6309903	Mex	162	Pasig City	Mexican	800	Pula(P)
11	6309455	Mad Mark's Creamery & Good Eats	162	Pasig City	American, Ice Cream, Desserts	900	Botswa Pula(P)
12	6318433	Silantro Fil-Mex	162	Quezon City	Filipino, Mexican	800	Botswa Pula(P)
13	6310470	Guevarra's	162	San Juan City	Filipino	1000	Botswa Pula(P)
14	6314605	Sodam Korean Restaurant	162	San Juan City	Korean	700	Botswa Pula(P)
15	18185059	Cafe Arabelle	162	Santa Rosa	Cafe, American, Italian, Filipino	800	Botswa Pula(P)
16	18182702	Nonna's Pasta & Pizzeria	162	Santa Rosa	Italian, Pizza	850	Botswa Pula(P)
17	6318213	Balay Dako	162	Tagaytay City	Filipino	1200	Botswa Pula(P)
18	18255654	Hobing Korean Dessert Cafe	162	Taguig City	Cafe, Korean, Desserts	600	Botswa Pula(P)
19	6308205	Wildflour Cafe + Bakery	162	Taguig City	Cafe, Bakery, American, Italian	1500	Botswa Pula(P)
20	6315438	NIU by Vikings	162	Taguig City	Seafood, American, Mediterranean, Japanese	3000	Botswa Pula(P)
21	6310406	The Food Hall by Todd English	162	Taguig City	American, Asian, Italian, Seafood	1800	Botswa Pula(P)
22	6600681	Chez Michou	30	Brasí_lia	Fast Food, French	55	Brazilia Real(R$
23	6601005	Cafí© Daniel Briand	30	Brasí_lia	Cafe	30	Brazilia Real(R$

24	6600292	Casa do Biscoito Mineiro	30	Brasí_lia	Bakery	45	Braziliar Real(R$
25	6600441	Maori	30	Brasí_lia	Brazilian	60	Braziliar Real(R$
26	6600970	Pizza íæ Bessa	30	Brasí_lia	Pizza	50	Braziliar Real(R$
27	6600379	Sushi Loko	30	Brasí_lia	Japanese	80	Braziliar Real(R$
28	6600214	Beirute	30	Brasí_lia	Arabian	90	Braziliar Real(R$
29	6601218	New Koto	30	Brasí_lia	Japanese	200	Braziliar Real(R$
...	...	...	...	...	...	...	...
9521	6003426	Liva	208	Ankara	Patisserie, Coffee and Tea	50	Turkish Lira(TL)
9522	6000549	Meôhur TavacÛ± Recep Usta	208	Ankara	Kebab	100	Turkish Lira(TL)
9523	6000871	î_ukuraÛôa SofrasÛ±	208	Ankara	Kebab, Izgara	60	Turkish Lira(TL)
9524	6004011	Gaga Manjero	208	Ankara	World Cuisine	80	Turkish Lira(TL)
9525	6000409	Cafemiz	208	Ankara	World Cuisine, Mexican, Italian	150	Turkish Lira(TL)
9526	6000019	Nusr-Et	208	Ankara	Steak	400	Turkish Lira(TL)
9527	6001537	Kebap 49	208	Ankara	Kebab, Turkish Pizza	60	Turkish Lira(TL)
9528	6003668	Timboo Cafe	208	Ankara	Cafe	70	Turkish Lira(TL)
9529	6001748	Meôhur í_zí_elik Aspava	208	Ankara	Kebab, Turkish Pizza	40	Turkish Lira(TL)
9530	6001757	YÛ±ldÛ±z Aspava	208	Ankara	Kebab, Turkish Pizza, Dí_ner	50	Turkish Lira(TL)

9531	6000747	The Bigos	208	Ankara	Cafe	80	Turkish Lira(TL)
9532	6002025	MasabaôÛ±	208	Ankara	Kebab, Turkish Pizza	100	Turkish Lira(TL)
9533	6003879	Zigana Pide	208	Ankara	Turkish Pizza	50	Turkish Lira(TL)
9534	6004089	Dí_veroÛôlu	208	Ankara	Kebab, Desserts, Turkish Pizza	70	Turkish Lira(TL)
9535	6000921	Dí_veroÛôlu	208	Ankara	Kebab, Desserts, Turkish Pizza	70	Turkish Lira(TL)
9536	6004813	Pizza ÛÁl Forno	208	Ankara	Pizza	40	Turkish Lira(TL)
9537	5904116	J'adore Chocolatier	208	ÛÁstanbul	Desserts	50	Turkish Lira(TL)
9538	5901782	Starbucks	208	ÛÁstanbul	Cafe	30	Turkish Lira(TL)
9539	5902117	Valonia	208	ÛÁstanbul	Restaurant Cafe, Desserts	80	Turkish Lira(TL)
9540	5927248	Draft Gastro Pub	208	ÛÁstanbul	Bar Food	130	Turkish Lira(TL)
9541	5905215	Emirgan Sí_tiô	208	ÛÁstanbul	Restaurant Cafe, Turkish, Desserts	75	Turkish Lira(TL)
9542	5926979	Leman Kí_ltí_r	208	ÛÁstanbul	Restaurant Cafe	80	Turkish Lira(TL)
9543	5916085	Dem Karakí_y	208	ÛÁstanbul	Cafe	35	Turkish Lira(TL)
9544	5915547	Karakí_y Gí_llí_oÛôlu	208	ÛÁstanbul	Desserts, Bí_rek	40	Turkish Lira(TL)
9545	5915054	Baltazar	208	ÛÁstanbul	Burger, Izgara	90	Turkish Lira(TL)
9546	5915730	NamlÛ± Gurme	208	ÛÁstanbul	Turkish	80	Turkish Lira(TL)
9547	5908749	Ceviz AÛôacÛ±	208	ÛÁstanbul	World Cuisine, Patisserie, Cafe	105	Turkish Lira(TL)

9548	5915807	Huqqa	208	ÛÁstanbul	Italian, World Cuisine	170	Turkish Lira(TL)
9549	5916112	Aôôk Kahve	208	ÛÁstanbul	Restaurant Cafe	120	Turkish Lira(TL)
9550	5927402	Walter's Coffee Roastery	208	ÛÁstanbul	Cafe	55	Turkish Lira(TL)

9551 rows × 12 columns

We have now removed unnecessary columns from our DataFrame. Details like specific geographical aspects will not help in our hypothesis testing.

Now, looking closer at the contents of our DataFrame, we see that a few of the "Aggregate rating" entries actually contain "0.0", which essentially equates to a null value; this will not help us in our data analysis, so we will remove the restaurants with this value. Though first, we will try to see why these restaurants receive the "0.0" rating. We assume that the number of votes may not be sufficient, so we will sort the DataFrame by the last column in ascending order.

In [4]:

```
zom.sort_values(by=['Votes'], ascending=True)
```

Out[4]:

	Restaurant ID	Restaurant Name	Country Code	City	Cuisines	Average Cost for two	Curren
5799	18460302	Khalsa Eating Point	1	New Delhi	North Indian	300	Indian Rupees(R
7411	18431145	Radha Swami Chaat Bhandar	1	New Delhi	Street Food	100	Indian Rupees(R
7414	18430905	Ram Ram Ji Kachori Bhandar	1	New Delhi	Street Food	50	Indian Rupees(R
7415	18430907	Rana's Food Corner	1	New Delhi	North Indian	200	Indian Rupees(R
7416	18451597	Sanjay Chicken Shop	1	New Delhi	Raw Meats, Fast Food	350	Indian Rupees(R

7418	18492057	Shree Raja Ram	1	New Delhi	Fast Food	50	Indian Rupees(R
7420	18429149	Special Moradabadi Chicken Corner	1	New Delhi	Biryani	200	Indian Rupees(R
7422	305567	Sushil Punjabi Vaishno Dhaba	1	New Delhi	North Indian	150	Indian Rupees(R
7423	18455551	Variety of Shawarmas	1	New Delhi	Fast Food	150	Indian Rupees(R
7410	18492103	New Sindhi Chicken Corner	1	New Delhi	North Indian, Mughlai	500	Indian Rupees(R
6472	18398593	Hungry Folks	1	New Delhi	North Indian, Chinese	500	Indian Rupees(R
7424	18430902	Vrindavan Sweets	1	New Delhi	South Indian, Street Food, Mithai	300	Indian Rupees(R
6469	18345771	Foodieholic	1	New Delhi	Chinese, North Indian, South Indian	200	Indian Rupees(R
8745	18373828	Parul's Cooking Hub	1	Noida	North Indian	450	Indian Rupees(R
6872	18492960	Mouthmatics	1	New Delhi	North Indian	200	Indian Rupees(R
6869	18489523	Le Village Pastry Shop	1	New Delhi	Bakery, Fast Food	300	Indian Rupees(R
6868	18355137	Kalka's Food Centre	1	New Delhi	North Indian, Chinese	200	Indian Rupees(R
6867	18355145	Janta Canteen	1	New Delhi	North Indian	150	Indian Rupees(R
8743	18273432	Cupcakes & More	1	Noida	Bakery, Desserts	300	Indian Rupees(R
8746	18252364	Republic of Chicken	1	Noida	Raw Meats, Fast Food	400	Indian Rupees(R
		6 Packs					Indian

9044	18486858	Momos	1	Noida	Chinese	300	Rupees(R
7409	18312627	New Pishori Chicken Kabab	1	New Delhi	Fast Food	300	Indian Rupees(R
7408	18462257	Momo's King	1	New Delhi	Chinese	200	Indian Rupees(R
7395	18365986	Batra Chinese Food & Chaap	1	New Delhi	Chinese	250	Indian Rupees(R
6878	18465571	Smoke Trailer Grill	1	New Delhi	American	400	Indian Rupees(R
7396	18441532	Celebration Family Restaurant	1	New Delhi	North Indian, Chinese	800	Indian Rupees(R
7397	18492045	Ching Chinese	1	New Delhi	Chinese, Fast Food	400	Indian Rupees(R
7399	18451605	Delhi-27	1	New Delhi	Chinese, South Indian	250	Indian Rupees(R
7400	18455549	Desi Kukkad	1	New Delhi	North Indian	300	Indian Rupees(R
7401	18455545	Freezy	1	New Delhi	Ice Cream	50	Indian Rupees(R
...	...	...	...	...	...	...	...
2490	49003	SpiceKlub	1	Mumbai	North Indian	1500	Indian Rupees(R
2300	90744	Exotica	1	Hyderabad	Mughlai, North Indian, Chinese	1500	Indian Rupees(R
3083	309790	The Vault Cafe	1	New Delhi	North Indian, Mediterranean, Asian, Continental	1600	Indian Rupees(R
5007	1819	The All American Diner	1	New Delhi	American, Fast Food	1000	Indian Rupees(R
6848	1777	Rajinder Da Dhaba	1	New Delhi	North Indian, Mughlai,	800	Indian Rupees(R

					Chinese		
731	56464	Glen's Bakehouse	1	Bangalore	Bakery, Desserts, Cafe	800	Indian Rupees(R
1861	4959	Downtown - Diners & Living Beer Cafe	1	Gurgaon	North Indian, Chinese, Italian, Continental	1800	Indian Rupees(R
3119	4249	Wenger's	1	New Delhi	Bakery, Fast Food, Desserts	400	Indian Rupees(R
3992	5030	Out Of The Box	1	New Delhi	American, North Indian, European, Asian	1800	Indian Rupees(R
821	65055	Barbeque Nation	1	Chennai	North Indian, Continental	2000	Indian Rupees(R
7863	301700	Big Yellow Door	1	New Delhi	Cafe, Italian, Fast Food	600	Indian Rupees(R
3336	304262	Ricos	1	New Delhi	Cafe, Mexican, American, Italian, Lebanese, Co...	900	Indian Rupees(R
6144	799	Gulati	1	New Delhi	North Indian, Mughlai	1500	Indian Rupees(R
1252	308022	Farzi Cafe	1	Gurgaon	Modern Indian	2200	Indian Rupees(R
2410	20350	Mocambo	1	Kolkata	Continental, Italian, North Indian	1050	Indian Rupees(R
4178	463	Karim's	1	New Delhi	Mughlai, North Indian	800	Indian Rupees(R
3085	301605	Warehouse Cafe	1	New Delhi	American, Continental, Italian, North Indian, ...	1500	Indian Rupees(R
4638	1614	Big Chill	1	New Delhi	Italian, Continental, European,	1500	Indian Rupees(R

					Cafe		
2480	35217	Joey's Pizza	1	Mumbai	Pizza	800	Indian Rupees(R
3110	900	Saravana Bhavan	1	New Delhi	South Indian	500	Indian Rupees(R
2411	20870	BarBQ	1	Kolkata	Chinese, North Indian	900	Indian Rupees(R
736	54162	The Black Pearl	1	Bangalore	North Indian, European, Mediterranean	1400	Indian Rupees(R
2307	94286	AB's - Absolute Barbecues	1	Hyderabad	European, Mediterranean, North Indian	1500	Indian Rupees(R
743	58882	Big Brewsky	1	Bangalore	Finger Food, North Indian, Italian, Continenta...	1800	Indian Rupees(R
2414	20842	Barbeque Nation	1	Kolkata	North Indian, Chinese	1600	Indian Rupees(R
739	56618	AB's - Absolute Barbecues	1	Bangalore	European, Mediterranean, North Indian	1400	Indian Rupees(R
2412	20404	Peter Cat	1	Kolkata	Continental, North Indian	1000	Indian Rupees(R
3994	308322	Hauz Khas Social	1	New Delhi	Continental, American, Asian, North Indian	1600	Indian Rupees(R
735	51040	Truffles	1	Bangalore	American, Burger, Cafe	800	Indian Rupees(R
728	51705	Toit	1	Bangalore	Italian, American, Pizza	2000	Indian Rupees(R

9551 rows × 12 columns

From sorting the data, we draw three conclusions: a restaurant gets a null rating if the "Rating text" is "Not rated", the "Rating color" is "White", and if the number of "Votes" is less than "4". We now have to decide which of the columns to use as the parameter for removing the null-value rows. We choose the "Rating color" "White" designation, as it will be simplest to remove rows based on this value.

However, before we do this, we must now replace all spaces DataFrame's column headers with underscores. We will rename some columns completely to shorten the header.

In [5]:

```
zom = zom.rename(index=str, columns={"Restaurant ID": "ID","Restaurant Name": "N
ame","Country Code": "Country_Code","Average Cost for two": "Cost","Price range"
: "Price","Aggregate rating": "Rating","Rating color": "Color","Rating text": "T
ext"})
#renaming columns

zom = zom[zom.Color != 'White'] #utilizing boolean indexing to remove rows with
"White" as the "Color"

zom
```

Out[5]:

	ID	Name	Country_Code	City	Cuisines	Cost	Currei
0	6317637	Le Petit Souffle	162	Makati City	French, Japanese, Desserts	1100	Botsw: Pula(P
1	6304287	Izakaya Kikufuji	162	Makati City	Japanese	1200	Botsw: Pula(P
2	6300002	Heat - Edsa Shangri-La	162	Mandaluyong City	Seafood, Asian, Filipino, Indian	4000	Botsw: Pula(P
3	6318506	Ooma	162	Mandaluyong City	Japanese, Sushi	1500	Botsw: Pula(P
4	6314302	Sambo Kojin	162	Mandaluyong City	Japanese, Korean	1500	Botsw: Pula(P
5	18189371	Din Tai Fung	162	Mandaluyong City	Chinese	1000	Botsw: Pula(P
6	6300781	Buffet 101	162	Pasay City	Asian, European	2000	Botsw: Pula(P
7	6301290	Vikings	162	Pasay City	Seafood, Filipino, Asian, European	2000	Botsw: Pula(P

8	6300010	Spiral - Sofitel Philippine Plaza Manila	162	Pasay City	European, Asian, Indian	6000	Botsw Pula(P
9	6314987	Locavore	162	Pasig City	Filipino	1100	Botsw Pula(P
10	6309903	Silantro Fil-Mex	162	Pasig City	Filipino, Mexican	800	Botsw Pula(P
11	6309455	Mad Mark's Creamery & Good Eats	162	Pasig City	American, Ice Cream, Desserts	900	Botsw Pula(P
12	6318433	Silantro Fil-Mex	162	Quezon City	Filipino, Mexican	800	Botsw Pula(P
13	6310470	Guevarra's	162	San Juan City	Filipino	1000	Botsw Pula(P
14	6314605	Sodam Korean Restaurant	162	San Juan City	Korean	700	Botsw Pula(P
15	18185059	Cafe Arabelle	162	Santa Rosa	Cafe, American, Italian, Filipino	800	Botsw Pula(P
16	18182702	Nonna's Pasta & Pizzeria	162	Santa Rosa	Italian, Pizza	850	Botsw Pula(P
17	6318213	Balay Dako	162	Tagaytay City	Filipino	1200	Botsw Pula(P
18	18255654	Hobing Korean Dessert Cafe	162	Taguig City	Cafe, Korean, Desserts	600	Botsw Pula(P
19	6308205	Wildflour Cafe + Bakery	162	Taguig City	Cafe, Bakery, American, Italian	1500	Botsw Pula(P
20	6315438	NIU by Vikings	162	Taguig City	Seafood, American, Mediterranean, Japanese	3000	Botsw Pula(P
		The Food Hall by			American,		

21	6310406	Todd English	162	Taguig City	Asian, Italian, Seafood	1800	Botsw Pula(P
22	6600681	Chez Michou	30	Brasí_lia	Fast Food, French	55	Brazili Real(R
23	6601005	Cafí© Daniel Briand	30	Brasí_lia	Cafe	30	Brazili Real(R
24	6600292	Casa do Biscoito Mineiro	30	Brasí_lia	Bakery	45	Brazili Real(R
25	6600441	Maori	30	Brasí_lia	Brazilian	60	Brazili Real(R
26	6600970	Pizza íæ Bessa	30	Brasí_lia	Pizza	50	Brazili Real(R
27	6600379	Sushi Loko	30	Brasí_lia	Japanese	80	Brazili Real(R
28	6600214	Beirute	30	Brasí_lia	Arabian	90	Brazili Real(R
29	6601218	New Koto	30	Brasí_lia	Japanese	200	Brazili Real(R
...	...	...	...	...	...	...	...
9521	6003426	Liva	208	Ankara	Patisserie, Coffee and Tea	50	Turkish Lira(TL
9522	6000549	Meôhur TavacÛ± Recep Usta	208	Ankara	Kebab	100	Turkish Lira(TL
9523	6000871	î'ukuraÛôa SofrasÛ±	208	Ankara	Kebab, Izgara	60	Turkish Lira(TL
9524	6004011	Gaga Manjero	208	Ankara	World Cuisine	80	Turkish Lira(TL
9525	6000409	Cafemiz	208	Ankara	World Cuisine, Mexican, Italian	150	Turkish Lira(TL
9526	6000019	Nusr-Et	208	Ankara	Steak	400	Turkish Lira(TL
9527	6001537	Kebap 49	208	Ankara	Kebab, Turkish Pizza	60	Turkish Lira(TL

9528	6003668	Timboo Cafe	208	Ankara	Cafe	70	Turkish Lira(TL
9529	6001748	Meôhur í_zí_elik Aspava	208	Ankara	Kebab, Turkish Pizza	40	Turkish Lira(TL
9530	6001757	YÛ±ldÛ±z Aspava	208	Ankara	Kebab, Turkish Pizza, Dí_ner	50	Turkish Lira(TL
9531	6000747	The Bigos	208	Ankara	Cafe	80	Turkish Lira(TL
9532	6002025	MasabaôÛ±	208	Ankara	Kebab, Turkish Pizza	100	Turkish Lira(TL
9533	6003879	Zigana Pide	208	Ankara	Turkish Pizza	50	Turkish Lira(TL
9534	6004089	Dí_veroÛôlu	208	Ankara	Kebab, Desserts, Turkish Pizza	70	Turkish Lira(TL
9535	6000921	Dí_veroÛôlu	208	Ankara	Kebab, Desserts, Turkish Pizza	70	Turkish Lira(TL
9536	6004813	Pizza ÛÁl Forno	208	Ankara	Pizza	40	Turkish Lira(TL
9537	5904116	J'adore Chocolatier	208	ÛÁstanbul	Desserts	50	Turkish Lira(TL
9538	5901782	Starbucks	208	ÛÁstanbul	Cafe	30	Turkish Lira(TL
9539	5902117	Valonia	208	ÛÁstanbul	Restaurant Cafe, Desserts	80	Turkish Lira(TL
9540	5927248	Draft Gastro Pub	208	ÛÁstanbul	Bar Food	130	Turkish Lira(TL
9541	5905215	Emirgan Sí_tiô	208	ÛÁstanbul	Restaurant Cafe, Turkish, Desserts	75	Turkish Lira(TL
9542	5926979	Leman Kí_ltí_r	208	ÛÁstanbul	Restaurant Cafe	80	Turkish Lira(TL
9543	5916085	Dem Karakí_y	208	ÛÁstanbul	Cafe	35	Turkish Lira(TL
9544	5915547	Karakí_y	208	ÛÁstanbul	Desserts,	40	Turkish

		Gí_Ilí_oÛôlu			Bí_rek		Lira(TL
9545	5915054	Baltazar	208	ÛÁstanbul	Burger, Izgara	90	Turkish Lira(TL
9546	5915730	NamlÛ± Gurme	208	ÛÁstanbul	Turkish	80	Turkish Lira(TL
9547	5908749	Ceviz AÛôacÛ±	208	ÛÁstanbul	World Cuisine, Patisserie, Cafe	105	Turkish Lira(TL
9548	5915807	Huqqa	208	ÛÁstanbul	Italian, World Cuisine	170	Turkish Lira(TL
9549	5916112	Aôôk Kahve	208	ÛÁstanbul	Restaurant Cafe	120	Turkish Lira(TL
9550	5927402	Walter's Coffee Roastery	208	ÛÁstanbul	Cafe	55	Turkish Lira(TL

7403 rows × 12 columns

We have now removed the null vales for ratings.

We would now like to clean the data a little bit more. In terms of geography, we have "City" and "Country_Code". Our analysis doesn't need to be so granular to look at a restaurant's city, but we also can't directly tell the restaurant's country location by the "Country_Code" alone. Luckily, the Kaggle user from whom we got the data included an Excel workbook that shows what country pertains to each "Country_Code". We must thus import this new workbook as a dataframe and perform a sort of VLOOKUP -> replace into our original 'zom' DataFrame. The replacing part seems a little complicated, so we opted to use a merge/join operation then manually delete the "Country_Code" column.

In [6]:

```
path2 = 'https://github.com/kenneth-yu19/Data_Bootcamp_Final_Project/blob/master/country_code.xlsx?raw=true'
exc_data2 = pd.read_excel(path2)
concode = pd.DataFrame(exc_data2) #convert excel data into pandas dataframe
concode = concode.rename(index=str, columns={"Country Code": "Country_Code"}) #renaming column for easier analysis
concode
```

Out[6]:

	Country_Code	Country
0	1	India
1	14	Australia
2	30	Brazil
3	37	Canada
4	94	Indonesia
5	148	New Zealand
6	162	Phillipines
7	166	Qatar
8	184	Singapore
9	189	South Africa
10	191	Sri Lanka
11	208	Turkey
12	214	UAE
13	215	United Kingdom
14	216	United States

In [7]:

```
zom = zom.merge(concode, on='Country_Code', how='left') #left merge using "Country_Code" as joining column
zom = zom.drop(columns=['Country_Code','City']) #dropping unnecessary columns
zom = zom[['ID','Name','Country','Cuisines','Cost','Currency','Price','Rating','Color','Text','Votes']] #re-ordering columns
zom
```

Out[7]:

	ID	Name	Country	Cuisines	Cost	Currency	Price	Rating
0	6317637	Le Petit Souffle	Phillipines	French, Japanese, Desserts	1100	Botswana Pula(P)	3	4.8
1	6304287	Izakaya Kikufuji	Phillipines	Japanese	1200	Botswana Pula(P)	3	4.5
2	6300002	Heat - Edsa Shangri-La	Phillipines	Seafood, Asian, Filipino, Indian	4000	Botswana Pula(P)	4	4.4
3	6318506	Ooma	Phillipines	Japanese, Sushi	1500	Botswana Pula(P)	4	4.9
4	6314302	Sambo Kojin	Phillipines	Japanese, Korean	1500	Botswana Pula(P)	4	4.8
5	18189371	Din Tai Fung	Phillipines	Chinese	1000	Botswana Pula(P)	3	4.4
6	6300781	Buffet 101	Phillipines	Asian, European	2000	Botswana Pula(P)	4	4.0
7	6301290	Vikings	Phillipines	Seafood, Filipino, Asian, European	2000	Botswana Pula(P)	4	4.2
8	6300010	Spiral - Sofitel Philippine Plaza Manila	Phillipines	European, Asian, Indian	6000	Botswana Pula(P)	4	4.9
9	6314987	Locavore	Phillipines	Filipino	1100	Botswana Pula(P)	3	4.8
10	6309903	Silantro Fil-Mex	Phillipines	Filipino, Mexican	800	Botswana Pula(P)	3	4.9
11	6309455	Mad Mark's Creamery & Good Eats	Phillipines	American, Ice Cream, Desserts	900	Botswana Pula(P)	3	4.2
12	6318433	Silantro Fil-Mex	Phillipines	Filipino, Mexican	800	Botswana Pula(P)	3	4.8
13	6310470	Guevarra's	Phillipines	Filipino	1000	Botswana Pula(P)	3	4.2
14	6314605	Sodam Korean Restaurant	Phillipines	Korean	700	Botswana Pula(P)	3	4.3

15	18185059	Cafe Arabelle	Phillipines	Cafe, American, Italian, Filipino	800	Botswana Pula(P)	3	3.6
16	18182702	Nonna's Pasta & Pizzeria	Phillipines	Italian, Pizza	850	Botswana Pula(P)	3	4.0
17	6318213	Balay Dako	Phillipines	Filipino	1200	Botswana Pula(P)	3	4.5
18	18255654	Hobing Korean Dessert Cafe	Phillipines	Cafe, Korean, Desserts	600	Botswana Pula(P)	2	4.5
19	6308205	Wildflour Cafe + Bakery	Phillipines	Cafe, Bakery, American, Italian	1500	Botswana Pula(P)	4	4.4
20	6315438	NIU by Vikings	Phillipines	Seafood, American, Mediterranean, Japanese	3000	Botswana Pula(P)	4	4.7
21	6310406	The Food Hall by Todd English	Phillipines	American, Asian, Italian, Seafood	1800	Botswana Pula(P)	4	4.5
22	6600681	Chez Michou	Brazil	Fast Food, French	55	Brazilian Real(R$)	2	3.0
23	6601005	Cafí© Daniel Briand	Brazil	Cafe	30	Brazilian Real(R$)	1	3.8
24	6600292	Casa do Biscoito Mineiro	Brazil	Bakery	45	Brazilian Real(R$)	2	3.7
25	6600441	Maori	Brazil	Brazilian	60	Brazilian Real(R$)	3	3.8
26	6600970	Pizza íæ Bessa	Brazil	Pizza	50	Brazilian Real(R$)	2	3.2
27	6600379	Sushi Loko	Brazil	Japanese	80	Brazilian Real(R$)	3	3.1
28	6600214	Beirute	Brazil	Arabian	90	Brazilian Real(R$)	3	3.7

29	6601218	New Koto	Brazil	Japanese	200	Brazilian Real(R$)	4	3.7
...	...	...	...	...	...	...	...	...
7373	6003426	Liva	Turkey	Patisserie, Coffee and Tea	50	Turkish Lira(TL)	2	3.4
7374	6000549	Meôhur TavacÛ± Recep Usta	Turkey	Kebab	100	Turkish Lira(TL)	3	4.5
7375	6000871	íˆukuraÛôa SofrasÛ±	Turkey	Kebab, Izgara	60	Turkish Lira(TL)	3	4.4
7376	6004011	Gaga Manjero	Turkey	World Cuisine	80	Turkish Lira(TL)	3	4.9
7377	6000409	Cafemiz	Turkey	World Cuisine, Mexican, Italian	150	Turkish Lira(TL)	4	4.4
7378	6000019	Nusr-Et	Turkey	Steak	400	Turkish Lira(TL)	4	4.1
7379	6001537	Kebap 49	Turkey	Kebab, Turkish Pizza	60	Turkish Lira(TL)	3	4.3
7380	6003668	Timboo Cafe	Turkey	Cafe	70	Turkish Lira(TL)	3	4.2
7381	6001748	Meôhur í_zí_elik Aspava	Turkey	Kebab, Turkish Pizza	40	Turkish Lira(TL)	2	4.6
7382	6001757	YÛ±ldÛ±z Aspava	Turkey	Kebab, Turkish Pizza, Dí_ner	50	Turkish Lira(TL)	2	4.4
7383	6000747	The Bigos	Turkey	Cafe	80	Turkish Lira(TL)	3	3.8
7384	6002025	MasabaôÛ±	Turkey	Kebab, Turkish Pizza	100	Turkish Lira(TL)	3	4.2
7385	6003879	Zigana Pide	Turkey	Turkish Pizza	50	Turkish Lira(TL)	2	4.3
7386	6004089	Dí_veroÛôlu	Turkey	Kebab, Desserts, Turkish Pizza	70	Turkish Lira(TL)	3	4.4
7387	6000921	Dí_veroÛôlu	Turkey	Kebab, Desserts, Turkish Pizza	70	Turkish Lira(TL)	3	4.2

7388	6004813	Pizza ÛÁl Forno	Turkey	Pizza	40	Turkish Lira(TL)	2	4.7
7389	5904116	J'adore Chocolatier	Turkey	Desserts	50	Turkish Lira(TL)	2	4.7
7390	5901782	Starbucks	Turkey	Cafe	30	Turkish Lira(TL)	2	4.9
7391	5902117	Valonia	Turkey	Restaurant Cafe, Desserts	80	Turkish Lira(TL)	3	4.2
7392	5927248	Draft Gastro Pub	Turkey	Bar Food	130	Turkish Lira(TL)	4	4.9
7393	5905215	Emirgan Sí_tiô	Turkey	Restaurant Cafe, Turkish, Desserts	75	Turkish Lira(TL)	3	4.2
7394	5926979	Leman Kí_ltí_r	Turkey	Restaurant Cafe	80	Turkish Lira(TL)	3	3.7
7395	5916085	Dem Karakí_y	Turkey	Cafe	35	Turkish Lira(TL)	2	4.5
7396	5915547	Karakí_y Gí_llí_oÛôlu	Turkey	Desserts, Bí_rek	40	Turkish Lira(TL)	2	4.7
7397	5915054	Baltazar	Turkey	Burger, Izgara	90	Turkish Lira(TL)	3	4.3
7398	5915730	NamlÛ± Gurme	Turkey	Turkish	80	Turkish Lira(TL)	3	4.1
7399	5908749	Ceviz AÛôacÛ±	Turkey	World Cuisine, Patisserie, Cafe	105	Turkish Lira(TL)	3	4.2
7400	5915807	Huqqa	Turkey	Italian, World Cuisine	170	Turkish Lira(TL)	4	3.7
7401	5916112	Aôôk Kahve	Turkey	Restaurant Cafe	120	Turkish Lira(TL)	4	4.0
7402	5927402	Walter's Coffee Roastery	Turkey	Cafe	55	Turkish Lira(TL)	2	4.0

7403 rows × 11 columns

Above we remove the "Country_Code" and "City" columns and move over the "Country" column so that it's the third column.

Now, there is another column we need to address in our data cleaning: "Currency". We want to normalize the currency, or at least get a relative measure of priciness of a restaurant, as price is a factor in some of our hypotheses, and we think it would be useful to know the extent of priciness for restaurants. Luckily, we have a price rating in the "Price" column and a few restaurants in the U.S. We will eventually remove the "Currency" column, but we would first like to see what exactly are the prices for each of the price ratings (1 to 5). Thus, we create a new DataFrame 'uszom' to anaylze.

In [8]:

```
uszom = zom.loc[zom['Country'] == 'United States'] #finding restaurants with "Co
untry" value as "United States"
print(uszom.shape)
```

(431, 11)

We have 431 U.S. restaurants, which provides a pretty good sample for analyzing prices. However, some of the values in the "Cost" column are "0", which isn't helpful, so we will remove those rows.

In [9]:

```
uszom = uszom[uszom.Cost != 0] #utilizing boolean indexing to remove rows with "
0" as the "Cost"
uszom
```

Out[9]:

	ID	Name	Country	Cuisines	Cost	Currency	Price	Rating	C
77	17284404	Austin's BBQ and Oyster Bar	United States	BBQ, Burger, Seafood	25	Dollar($)	2	3.3	Ora
78	17284203	BJ's Country Buffet	United States	American, BBQ	10	Dollar($)	1	3.3	Ora
81	17284397	Elements Coffee Co - Northwest	United States	Coffee and Tea, Sandwich	10	Dollar($)	1	3.4	Ora
83	17284094	Chick-fil-A	United States	Fast Food	10	Dollar($)	1	3.5	Yell
84	17284409	Guang Zhou Chinese Restaurant	United States	Asian, Chinese, Vegetarian	10	Dollar($)	1	3.9	Yell

85	17284139	Harvest Moon	United States	Pizza, Bar Food, Sandwich	25	Dollar($)	2	3.7	Yell
86	17284403	Henry Campbell's Steakhouse	United States	Steak, Tapas, Bar Food	70	Dollar($)	4	3.5	Yell
87	17284145	Hong Kong Cafe	United States	Chinese, Seafood, Vegetarian	25	Dollar($)	2	3.6	Yell
88	17284150	House of China Restaurant II	United States	Chinese	10	Dollar($)	1	3.8	Yell
89	17284158	Jimmie's Hot Dogs	United States	NaN	10	Dollar($)	1	3.9	Yell
90	17284175	Locos Grill & Pub	United States	American, Burger, Sandwich	25	Dollar($)	2	3.5	Yell
91	17284179	Longhorn Steakhouse	United States	American, Steak	25	Dollar($)	2	3.5	Yell
92	17284197	Mikata Japanese Steakhouse	United States	Japanese, Steak, Sushi	40	Dollar($)	3	3.6	Yell
93	17284241	Shogun Japanese Steak House	United States	Japanese, Steak, Sushi	40	Dollar($)	3	3.5	Yell
94	17284390	The Catch Seafood Room & Oyster Bar	United States	Seafood, Tapas, Bar Food	40	Dollar($)	3	3.8	Yell
95	17284279	Villa Gargano	United States	Italian, Pizza	10	Dollar($)	1	3.7	Yell
96	17284364	3 Squares Diner	United States	American, Breakfast, Diner	10	Dollar($)	1	3.4	Ora
98	17293281	Last Resort Grill	United States	American, Southern, Southwestern	40	Dollar($)	3	4.5	Dar Gre
99	17293301	Mama's Boy Restaurant	United States	Southern	10	Dollar($)	1	4.5	Dar Gre

100	17293409	Sr. Sol 1	United States	Mexican	10	Dollar($)	1	4.6	Dar Gre
101	17293163	Choo Choo Eastside	United States	Japanese, Korean	10	Dollar($)	1	3.9	Yell
102	17293228	The Grill	United States	Breakfast, Burger, Sandwich	10	Dollar($)	1	3.7	Yell
103	17293880	Big City Bread Cafe	United States	Breakfast, Sandwich	10	Dollar($)	1	4.3	Gre
104	17293169	Clocked	United States	American, Burger, Sandwich	10	Dollar($)	1	4.2	Gre
105	17293186	DePalma's Italian Cafe - Downtown	United States	American, Italian, Pizza	25	Dollar($)	2	4.0	Gre
106	17293180	DePalma's Italian Cafe - East Side	United States	American, Italian, Pizza	25	Dollar($)	2	4.1	Gre
107	17293205	Five & Ten	United States	American	40	Dollar($)	3	4.3	Gre
108	17293229	Grit	United States	International, Southern, Vegetarian	10	Dollar($)	1	4.2	Gre
109	17293897	Ike & Jane	United States	Sandwich	10	Dollar($)	1	4.1	Gre
110	17293273	La Dolce Vita Ristorante	United States	Italian	40	Dollar($)	3	4.1	Gre
...	...	...	...	...	...	...	...	...	...
523	17678229	Masato Japanese	United States	Japanese, Sushi	25	Dollar($)	2	3.9	Yell
524	17678097	Mom & Dad's Italian Restaurant	United States	Italian	25	Dollar($)	2	3.7	Yell
525	17678216	Mori's Japanese Steakhouse & Sushi Bar	United States	Japanese, Steak, Sushi	70	Dollar($)	4	3.5	Yell

526	17678233	Passage 2 India	United States	Indian, Middle Eastern	25	Dollar($)	2	3.8	Yell
527	17678148	Rodeo Mexican Restaurant	United States	Mexican	25	Dollar($)	2	3.8	Yell
528	17678291	Steel Magnolias	United States	Southern	40	Dollar($)	3	3.8	Yell
529	17678218	Smok'n Pig B-B-Q	United States	BBQ	25	Dollar($)	2	4.1	Gre
530	17558738	Blue House Cafe	United States	Coffee and Tea, Mediterranean	10	Dollar($)	1	4.3	Gre
533	17697444	Masala Grill & Coffee House	United States	Indian, Middle Eastern	10	Dollar($)	1	3.2	Ora
534	17696871	Brown Bottle The Cedar Falls	United States	Italian, Pizza, Vegetarian	25	Dollar($)	2	3.7	Yell
535	17696891	Four Queens Dairy Cream	United States	Desserts	10	Dollar($)	1	3.9	Yell
536	17696901	Hong Kong Chinese Restaurant	United States	Chinese	10	Dollar($)	1	3.8	Yell
537	17697384	HuHot Mongolian Grill	United States	Asian, Chinese	25	Dollar($)	2	3.7	Yell
538	17697417	J's Homestyle Cooking	United States	American, Breakfast	10	Dollar($)	1	3.6	Yell
539	17696918	Montage	United States	American, Italian, Seafood	40	Dollar($)	3	3.6	Yell
540	17696920	Mulligan's Brick Oven Grill	United States	Burger, Pizza, Sandwich	25	Dollar($)	2	3.6	Yell
541	17697398	Sakura	United States	Japanese, Sushi	40	Dollar($)	3	3.8	Yell
				Coffee and					

542	17697406	Scratch	United States	Tea, Desserts, Beverages	10	Dollar($)	1	3.7	Yell
543	17696941	SOHO Sushi Bar & Deli	United States	Sandwich, Sushi, Tapas	25	Dollar($)	2	3.6	Yell
544	17696955	Texas Roadhouse	United States	American, BBQ, Steak	25	Dollar($)	2	3.6	Yell
545	17696957	Tony's La Pizzeria	United States	Pizza, Sandwich	25	Dollar($)	2	3.6	Yell
546	17697418	Chapala	United States	Mexican	25	Dollar($)	2	3.6	Yell
547	17697386	Galleria de Paco	United States	American	40	Dollar($)	3	3.6	Yell
548	17697224	Golden China	United States	Chinese	10	Dollar($)	1	3.7	Yell
549	17697304	Rudy's Tacos	United States	Mexican	25	Dollar($)	2	3.6	Yell
550	17697389	The Screaming Eagle	United States	American, Bar Food	10	Dollar($)	1	3.7	Yell
551	17697424	The Thai Bowl	United States	Thai	10	Dollar($)	1	3.5	Yell
552	17697332	Tokyo Japanese Steak House	United States	Japanese, Steak, Sushi	25	Dollar($)	2	3.9	Yell
553	17694056	Theo Yianni's Authentic Greek Restaurant	United States	Burger, Greek, Sandwich	25	Dollar($)	2	3.9	Yell
554	17559793	Fishpatrick's Crabby Cafe	United States	Burger, Seafood, Steak	25	Dollar($)	2	3.2	Ora

422 rows × 11 columns

We still have 422 restaurants, which is good. To find the price range of each price rating, we will create a new DataFrame with the price rating, minimum cost for a meal for two, and the maximum cost as the columns. This will utilize the "Cost" and "Price" columns of our 'uszom' DataFrame.

In [10]:

```
print(uszom.loc[uszom['Price'] == 1, 'Cost'].min(),uszom.loc[uszom['Price'] == 1
, 'Cost'].max()) #print min/max cost for price 1
print(uszom.loc[uszom['Price'] == 2, 'Cost'].min(),uszom.loc[uszom['Price'] == 2
, 'Cost'].max()) #print min/max cost for price 2
print(uszom.loc[uszom['Price'] == 3, 'Cost'].min(),uszom.loc[uszom['Price'] == 3
, 'Cost'].max()) #print min/max cost for price 3
print(uszom.loc[uszom['Price'] == 4, 'Cost'].min(),uszom.loc[uszom['Price'] == 4
, 'Cost'].max()) #print min/max cost for price 4
print(uszom.loc[uszom['Price'] == 5, 'Cost'].min(),uszom.loc[uszom['Price'] == 5
, 'Cost'].max()) #print min/max cost for price 5
```

```
10 10
25 25
30 45
50 100
nan nan
```

Interestingly, the price range is 1 to 4, not 1 to 5 like the overall restaurant ranking. We see here that a restaurant is ranked 1 if the average price is USD10, 2 if USD25, 3 if between USD30 and USD45, and 4 if between USD50 and USD100. This calculation just serves as a contextual reference for price; these numbers won't necessarily be used in our graphing, but we wanted to have them to put more color to our analysis. We can now remove the "Cost" and "Currency" columns of our 'zom' DataFrame.

In [11]:

```
zom = zom.drop(columns=['Cost','Currency']) #dropping unnecessary columns
zom
```

Out[11]:

	ID	Name	Country	Cuisines	Price	Rating	Color	Text
0	6317637	Le Petit Souffle	Phillipines	French, Japanese, Desserts	3	4.8	Dark Green	Excellent
1	6304287	Izakaya Kikufuji	Phillipines	Japanese	3	4.5	Dark Green	Excellent
2	6300002	Heat - Edsa Shangri-La	Phillipines	Seafood, Asian, Filipino, Indian	4	4.4	Green	Very Good
				Japanese,			Dark	

3	6318506	Ooma	Phillipines	Sushi	4	4.9	Green	Excellent
4	6314302	Sambo Kojin	Phillipines	Japanese, Korean	4	4.8	Dark Green	Excellent
5	18189371	Din Tai Fung	Phillipines	Chinese	3	4.4	Green	Very Good
6	6300781	Buffet 101	Phillipines	Asian, European	4	4.0	Green	Very Good
7	6301290	Vikings	Phillipines	Seafood, Filipino, Asian, European	4	4.2	Green	Very Good
8	6300010	Spiral - Sofitel Philippine Plaza Manila	Phillipines	European, Asian, Indian	4	4.9	Dark Green	Excellent
9	6314987	Locavore	Phillipines	Filipino	3	4.8	Dark Green	Excellent
10	6309903	Silantro Fil-Mex	Phillipines	Filipino, Mexican	3	4.9	Dark Green	Excellent
11	6309455	Mad Mark's Creamery & Good Eats	Phillipines	American, Ice Cream, Desserts	3	4.2	Green	Very Good
12	6318433	Silantro Fil-Mex	Phillipines	Filipino, Mexican	3	4.8	Dark Green	Excellent
13	6310470	Guevarra's	Phillipines	Filipino	3	4.2	Green	Very Good
14	6314605	Sodam Korean Restaurant	Phillipines	Korean	3	4.3	Green	Very Good
15	18185059	Cafe Arabelle	Phillipines	Cafe, American, Italian, Filipino	3	3.6	Yellow	Good
16	18182702	Nonna's Pasta & Pizzeria	Phillipines	Italian, Pizza	3	4.0	Green	Very Good
17	6318213	Balay Dako	Phillipines	Filipino	3	4.5	Dark Green	Excellent
		Hobing						

18	18255654	Korean Dessert Cafe	Phillipines	Cafe, Korean, Desserts	2	4.5	Dark Green	Excellent
19	6308205	Wildflour Cafe + Bakery	Phillipines	Cafe, Bakery, American, Italian	4	4.4	Green	Very Good
20	6315438	NIU by Vikings	Phillipines	Seafood, American, Mediterranean, Japanese	4	4.7	Dark Green	Excellent
21	6310406	The Food Hall by Todd English	Phillipines	American, Asian, Italian, Seafood	4	4.5	Dark Green	Excellent
22	6600681	Chez Michou	Brazil	Fast Food, French	2	3.0	Orange	Average
23	6601005	Cafí© Daniel Briand	Brazil	Cafe	1	3.8	Yellow	Good
24	6600292	Casa do Biscoito Mineiro	Brazil	Bakery	2	3.7	Yellow	Good
25	6600441	Maori	Brazil	Brazilian	3	3.8	Yellow	Good
26	6600970	Pizza íæ Bessa	Brazil	Pizza	2	3.2	Orange	Average
27	6600379	Sushi Loko	Brazil	Japanese	3	3.1	Orange	Average
28	6600214	Beirute	Brazil	Arabian	3	3.7	Yellow	Good
29	6601218	New Koto	Brazil	Japanese	4	3.7	Yellow	Good
...	...	...	...	...	...	...	...	...
7373	6003426	Liva	Turkey	Patisserie, Coffee and Tea	2	3.4	Orange	Average
7374	6000549	Meôhur TavacÛ± Recep Usta	Turkey	Kebab	3	4.5	Dark Green	Excellent
7375	6000871	î̂ukuraÛôa SofrasÛ±	Turkey	Kebab, Izgara	3	4.4	Green	Very Good
7376	6004011	Gaga Manjero	Turkey	World Cuisine	3	4.9	Dark Green	Excellent

7377	6000409	Cafemiz	Turkey	World Cuisine, Mexican, Italian	4	4.4	Green	Very Good
7378	6000019	Nusr-Et	Turkey	Steak	4	4.1	Green	Very Good
7379	6001537	Kebap 49	Turkey	Kebab, Turkish Pizza	3	4.3	Green	Very Good
7380	6003668	Timboo Cafe	Turkey	Cafe	3	4.2	Green	Very Good
7381	6001748	Meôhur í_zí_elik Aspava	Turkey	Kebab, Turkish Pizza	2	4.6	Dark Green	Excellent
7382	6001757	YÛ±ldÛ±z Aspava	Turkey	Kebab, Turkish Pizza, Dí_ner	2	4.4	Green	Very Good
7383	6000747	The Bigos	Turkey	Cafe	3	3.8	Yellow	Good
7384	6002025	MasabaôÛ±	Turkey	Kebab, Turkish Pizza	3	4.2	Green	Very Good
7385	6003879	Zigana Pide	Turkey	Turkish Pizza	2	4.3	Green	Very Good
7386	6004089	Dí_veroÛôlu	Turkey	Kebab, Desserts, Turkish Pizza	3	4.4	Green	Very Good
7387	6000921	Dí_veroÛôlu	Turkey	Kebab, Desserts, Turkish Pizza	3	4.2	Green	Very Good
7388	6004813	Pizza ÛÁl Forno	Turkey	Pizza	2	4.7	Dark Green	Excellent
7389	5904116	J'adore Chocolatier	Turkey	Desserts	2	4.7	Dark Green	Excellent
7390	5901782	Starbucks	Turkey	Cafe	2	4.9	Dark Green	Excellent
7391	5902117	Valonia	Turkey	Restaurant Cafe, Desserts	3	4.2	Green	Very Good
7392	5927248	Draft Gastro Pub	Turkey	Bar Food	4	4.9	Dark Green	Excellent
7393	5905215	Emirgan Sí_tiô	Turkey	Restaurant Cafe, Turkish, Desserts	3	4.2	Green	Very Good

7394	5926979	Leman Kí_ltí_r	Turkey	Restaurant Cafe	3	3.7	Yellow	Good
7395	5916085	Dem Karakí_y	Turkey	Cafe	2	4.5	Dark Green	Excellent
7396	5915547	Karakí_y Gí_llí_oÛôlu	Turkey	Desserts, Bí_rek	2	4.7	Dark Green	Excellent
7397	5915054	Baltazar	Turkey	Burger, Izgara	3	4.3	Green	Very Good
7398	5915730	NamlÛ± Gurme	Turkey	Turkish	3	4.1	Green	Very Good
7399	5908749	Ceviz AÛôacÛ±	Turkey	World Cuisine, Patisserie, Cafe	3	4.2	Green	Very Good
7400	5915807	Huqqa	Turkey	Italian, World Cuisine	4	3.7	Yellow	Good
7401	5916112	Aôôk Kahve	Turkey	Restaurant Cafe	4	4.0	Green	Very Good
7402	5927402	Walter's Coffee Roastery	Turkey	Cafe	2	4.0	Green	Very Good

7403 rows × 9 columns

Now, we have one final step in data cleaning! We want to look at individual cuisines in our hypotheses, and we see that the "Cuisines" column has up to 4 cuisines/types of food listed. First, we will do a text-to-columns operation to separate the cuisines, then we will look at the unique cuisines, then we will choose the cuisines we want to analyze, then we will attribute those respective cuisines as the main cuisine.

In [12]:

```
cuisine = zom["Cuisines"].str.split(", ", n = 3, expand = True) #create 'cuisine
' df with 4 columns separating cuisines
zom["Cuisine1"]= cuisine[0] #append first column to original 'zom' df
zom["Cuisine2"]= cuisine[1] #append second column to original 'zom' df
zom["Cuisine3"]= cuisine[2] #append third column to original 'zom' df
zom["Cuisine4"]= cuisine[2] #append fourth column to original 'zom' df

cuis = pd.unique(zom[['Cuisine1','Cuisine2','Cuisine3','Cuisine4']].values.ravel
('K')) #list unique cuisines from cuisine columns
cuis = cuis.astype(str) #convert from object to string
cuis.sort() #sort array alphabetically
print(cuis)
print(len(cuis)) #finding the number of unique cuisines that show up
```

```
['Afghani' 'African' 'American' 'Andhra' 'Arabian' 'Argentine' 'Asia
n'
 'Asian Fusion' 'Assamese' 'Australian' 'Awadhi' 'BBQ' 'Bakery' 'Bar
Food'
 'Belgian' 'Bengali' 'Beverages' 'Bihari' 'Biryani' 'Brazilian'
 'Breakfast' 'British' 'Burger' 'Burmese' 'Bí_rek' 'Cafe' 'Cajun'
 'Canadian' 'Cantonese' 'Caribbean' 'Charcoal Grill' 'Chettinad' 'Ch
inese'
 'Coffee and Tea' 'Contemporary' 'Continental' 'Cuban' 'Curry' 'Deli
'
 'Desserts' 'Dim Sum' 'Diner' 'Drinks Only' 'Durban' 'Dí_ner' 'Europ
ean'
 'Fast Food' 'Filipino' 'Finger Food' 'Fish and Chips' 'French' 'Fus
ion'
 'German' 'Goan' 'Gourmet Fast Food' 'Greek' 'Grill' 'Gujarati' 'Haw
aiian'
 'Healthy Food' 'Hyderabadi' 'Ice Cream' 'Indian' 'Indonesian'
 'International' 'Iranian' 'Irish' 'Italian' 'Izgara' 'Japanese' 'Ju
ices'
 'Kashmiri' 'Kebab' 'Kerala' 'Kiwi' 'Korean' 'Latin American' 'Leban
ese'
 'Lucknowi' 'Maharashtrian' 'Malaysian' 'Malwani' 'Mangalorean'
 'Mediterranean' 'Mexican' 'Middle Eastern' 'Mithai' 'Modern Austral
ian'
 'Modern Indian' 'Moroccan' 'Mughlai' 'Naga' 'Nepalese' 'New America
n'
 'None' 'North Eastern' 'North Indian' 'Oriya' 'Pakistani' 'Parsi'
 'Patisserie' 'Peranakan' 'Persian' 'Peruvian' 'Pizza' 'Portuguese'
 'Pub Food' 'Rajasthani' 'Ramen' 'Raw Meats' 'Restaurant Cafe' 'Sala
d'
 'Sandwich' 'Scottish' 'Seafood' 'Singaporean' 'Soul Food' 'South Af
rican'
 'South American' 'South Indian' 'Southern' 'Southwestern' 'Spanish'
 'Sri Lankan' 'Steak' 'Street Food' 'Sunda' 'Sushi' 'Taiwanese' 'Tap
as'
 'Tea' 'Teriyaki' 'Tex-Mex' 'Thai' 'Tibetan' 'Turkish' 'Turkish Pizz
a'
 'Vegetarian' 'Vietnamese' 'Western' 'World Cuisine' 'nan']
142
```

Here, we have a decision to make. We have all unique cuisines/types of food that show up in the DataFrame, but we need to choose which one will represent the restaurant. It's clear that the cuisines aren't listed in alphabetical order if the restaurant has multiple, so we assume that the cuisines are listed by saliency. Thus, we will use the cuisine in the "Cuisine1" column as the representative cuisine.

In [13]:

```
zom = zom.drop(columns=['Cuisines','Cuisine2','Cuisine3','Cuisine4']) #dropping
unnecessary columns
zom = zom.rename(index=str, columns={"Cuisine1": "Cuisine"}) #renaming column
zom = zom[['ID','Name','Country','Cuisine','Price','Rating','Color','Text','Vote
s']] #re-ordering columns
zom
```

Out[13]:

	ID	Name	Country	Cuisine	Price	Rating	Color	Text	Vo
0	6317637	Le Petit Souffle	Phillipines	French	3	4.8	Dark Green	Excellent	31
1	6304287	Izakaya Kikufuji	Phillipines	Japanese	3	4.5	Dark Green	Excellent	59
2	6300002	Heat - Edsa Shangri-La	Phillipines	Seafood	4	4.4	Green	Very Good	27
3	6318506	Ooma	Phillipines	Japanese	4	4.9	Dark Green	Excellent	36
4	6314302	Sambo Kojin	Phillipines	Japanese	4	4.8	Dark Green	Excellent	22
5	18189371	Din Tai Fung	Phillipines	Chinese	3	4.4	Green	Very Good	33
6	6300781	Buffet 101	Phillipines	Asian	4	4.0	Green	Very Good	52
7	6301290	Vikings	Phillipines	Seafood	4	4.2	Green	Very Good	67
8	6300010	Spiral - Sofitel Philippine Plaza Manila	Phillipines	European	4	4.9	Dark Green	Excellent	62
9	6314987	Locavore	Phillipines	Filipino	3	4.8	Dark Green	Excellent	53
10	6309903	Silantro Fil-Mex	Phillipines	Filipino	3	4.9	Dark Green	Excellent	10
11	6309455	Mad Mark's Creamery & Good Eats	Phillipines	American	3	4.2	Green	Very Good	48

12	6318433	Silantro Fil-Mex	Phillipines	Filipino	3	4.8	Dark Green	Excellent	29
13	6310470	Guevarra's	Phillipines	Filipino	3	4.2	Green	Very Good	45
14	6314605	Sodam Korean Restaurant	Phillipines	Korean	3	4.3	Green	Very Good	22
15	18185059	Cafe Arabelle	Phillipines	Cafe	3	3.6	Yellow	Good	29
16	18182702	Nonna's Pasta & Pizzeria	Phillipines	Italian	3	4.0	Green	Very Good	72
17	6318213	Balay Dako	Phillipines	Filipino	3	4.5	Dark Green	Excellent	21
18	18255654	Hobing Korean Dessert Cafe	Phillipines	Cafe	2	4.5	Dark Green	Excellent	11
19	6308205	Wildflour Cafe + Bakery	Phillipines	Cafe	4	4.4	Green	Very Good	39
20	6315438	NIU by Vikings	Phillipines	Seafood	4	4.7	Dark Green	Excellent	53
21	6310406	The Food Hall by Todd English	Phillipines	American	4	4.5	Dark Green	Excellent	61
22	6600681	Chez Michou	Brazil	Fast Food	2	3.0	Orange	Average	6
23	6601005	Cafí© Daniel Briand	Brazil	Cafe	1	3.8	Yellow	Good	9
24	6600292	Casa do Biscoito Mineiro	Brazil	Bakery	2	3.7	Yellow	Good	11
25	6600441	Maori	Brazil	Brazilian	3	3.8	Yellow	Good	11
26	6600970	Pizza íæ Bessa	Brazil	Pizza	2	3.2	Orange	Average	11
27	6600379	Sushi Loko	Brazil	Japanese	3	3.1	Orange	Average	10

28	6600214	Beirute	Brazil	Arabian	3	3.7	Yellow	Good	8
29	6601218	New Koto	Brazil	Japanese	4	3.7	Yellow	Good	5
...	...	...	...	...	...	...	...	...	...
7373	6003426	Liva	Turkey	Patisserie	2	3.4	Orange	Average	11
7374	6000549	Meôhur TavacÛ± Recep Usta	Turkey	Kebab	3	4.5	Dark Green	Excellent	23
7375	6000871	î_ukuraÛôa SofrasÛ±	Turkey	Kebab	3	4.4	Green	Very Good	29
7376	6004011	Gaga Manjero	Turkey	World Cuisine	3	4.9	Dark Green	Excellent	95
7377	6000409	Cafemiz	Turkey	World Cuisine	4	4.4	Green	Very Good	11
7378	6000019	Nusr-Et	Turkey	Steak	4	4.1	Green	Very Good	97
7379	6001537	Kebap 49	Turkey	Kebab	3	4.3	Green	Very Good	10
7380	6003668	Timboo Cafe	Turkey	Cafe	3	4.2	Green	Very Good	79
7381	6001748	Meôhur í_zí_elik Aspava	Turkey	Kebab	2	4.6	Dark Green	Excellent	10
7382	6001757	YÛ±ldÛ±z Aspava	Turkey	Kebab	2	4.4	Green	Very Good	72
7383	6000747	The Bigos	Turkey	Cafe	3	3.8	Yellow	Good	12
7384	6002025	MasabaôÛ±	Turkey	Kebab	3	4.2	Green	Very Good	10
7385	6003879	Zigana Pide	Turkey	Turkish Pizza	2	4.3	Green	Very Good	10
7386	6004089	Dí_veroÛôlu	Turkey	Kebab	3	4.4	Green	Very Good	13
7387	6000921	Dí_veroÛôlu	Turkey	Kebab	3	4.2	Green	Very Good	15
7388	6004813	Pizza ÛÁl Forno	Turkey	Pizza	2	4.7	Dark Green	Excellent	10
		J'adore					Dark		

7389	5904116	Chocolatier	Turkey	Desserts	2	4.7	Green	Excellent	13
7390	5901782	Starbucks	Turkey	Cafe	2	4.9	Dark Green	Excellent	10
7391	5902117	Valonia	Turkey	Restaurant Cafe	3	4.2	Green	Very Good	87
7392	5927248	Draft Gastro Pub	Turkey	Bar Food	4	4.9	Dark Green	Excellent	52
7393	5905215	Emirgan Sí_tiô	Turkey	Restaurant Cafe	3	4.2	Green	Very Good	87
7394	5926979	Leman Kí_ltí_r	Turkey	Restaurant Cafe	3	3.7	Yellow	Good	50
7395	5916085	Dem Karakí_y	Turkey	Cafe	2	4.5	Dark Green	Excellent	76
7396	5915547	Karakí_y Gí_llí_oÛôlu	Turkey	Desserts	2	4.7	Dark Green	Excellent	13
7397	5915054	Baltazar	Turkey	Burger	3	4.3	Green	Very Good	87
7398	5915730	NamlÛ± Gurme	Turkey	Turkish	3	4.1	Green	Very Good	78
7399	5908749	Ceviz AÛôacÛ±	Turkey	World Cuisine	3	4.2	Green	Very Good	10
7400	5915807	Huqqa	Turkey	Italian	4	3.7	Yellow	Good	66
7401	5916112	Aôôk Kahve	Turkey	Restaurant Cafe	4	4.0	Green	Very Good	90
7402	5927402	Walter's Coffee Roastery	Turkey	Cafe	2	4.0	Green	Very Good	59

In [14]:

```
zom.dtypes #checking the column types to ensure they're able to be graphed
```

Out[14]:

```
ID            int64
Name         object
Country      object
Cuisine      object
Price         int64
Rating      float64
Color        object
Text         object
Votes         int64
dtype: object
```

And with that, we've cleaned all our data! Above, we checked to see that the "Price", "Rating", and "Votes" columns were either integers or floats so that we can plot them and perform calculations on them. We're now ready to analyze our DataFrame to look at some basic statistics and test our hypotheses through data analysis. We still have 7,403 restaurants, which is quite a sizeable data set to use for our analysis.

Pivot Table Analysis

Here, we want to profile our data, so we use pivot tables to garner a few insights. We can utilize pivot tables to test our hypotheses.

In [15]:

```
countryvotes = pd.pivot_table(zom, values='Votes', index='Country', aggfunc=np.s
um, margins=True)
countryvotes.sort_values(['Votes'], ascending=[True], inplace=True) #sort countr
y based on total votes
countryvotes
```

Out[15]:

	Votes
Country	
Canada	412
Singapore	638
Brazil	1170
Australia	2674
Sri Lanka	2929
Qatar	3276
Phillipines	8963
New Zealand	9721
Turkey	14670
Indonesia	16214
United Kingdom	16436
South Africa	18910
UAE	29611
United States	185842
India	1185310
All	1496776

This pivot table shows the total number of reviews by country. Of course, India has the most votes at 1,185,310 reviews left, as Zomato is based in that country. Zomato operates in 24 countries, but we only have 15 countries represented in the data. Therefore, we must keep in mind that this data set of 1,496,776 votes isn't completely representative of the Zomato population.

In [16]:

```
cuisdata_p = pd.pivot_table(zom, index='Cuisine', values=['Price','Rating','Vote
s'], aggfunc=np.average)
cuisdata_p.sort_values(['Price', 'Rating', 'Votes'], ascending=[True, False, Fal
se], inplace=True) #sort cuisine based on price
cuisdata_p.tail(10) #display first 10 rows
```

Out[16]:

	Price	Rating	Votes
Cuisine			
French	3.785714	4.178571	176.0
Argentine	4.000000	4.500000	602.0
African	4.000000	4.450000	319.0
Irish	4.000000	4.300000	782.0
Gourmet Fast Food	4.000000	4.300000	68.0
Kiwi	4.000000	4.200000	271.0
Asian Fusion	4.000000	3.850000	73.5
Peruvian	4.000000	3.600000	5.0
South American	4.000000	3.500000	72.5
Drinks Only	4.000000	3.500000	45.0

We see here that French cuisine is the 10th most expensive cuisine in the data set. This confirms the first part of our first hypothesis that French cuisine, which is usually related to fine dining, does usually command a high price. Another thing to note is that a few of the top cuisines have an average less than 50 votes, and with 1.5 million votes, such a figure is quite insignificant and could present misleading skew.

In [17]:

```
cuisdata_r = pd.pivot_table(zom, index='Cuisine', values=['Rating','Price','Vote
s'], aggfunc=np.average)
cuisdata_r.sort_values(['Rating', 'Price', 'Votes'], ascending=[True, False, Fal
se], inplace=True) #sort cuisine based on rating
cuisdata_r.tail(10) #display first 10 rows
```

Out[17]:

	Price	Rating	Votes
Cuisine			
Hawaiian	3.400000	4.480000	1149.800000
Argentine	4.000000	4.500000	602.000000
World Cuisine	3.333333	4.500000	414.666667
Ramen	3.000000	4.600000	418.000000
Persian	3.000000	4.600000	177.000000
Filipino	3.000000	4.616667	454.500000
Taiwanese	2.500000	4.650000	192.000000
Scottish	3.000000	4.700000	163.000000
Cajun	2.000000	4.700000	1412.000000
Sunda	3.000000	4.900000	1838.000000

Here we see the top 10 cuisines based on average rating. The second part of our first hypothesis and our second hypothesis do not hold: cuisine is not necessarily a signal for a high rating. In fact, Sunda cuisine, a particular type of Indonesian food, has both the highest average rating and number of votes, while maintaining an average price level of 3. Only Argentine food has the highest price level of 4 as well as a high rating. However, the rest of the cuisines have an average price level under 3.50 while maintaining a top 10 average rating.

In [18]:

```
cuisdata_c = pd.pivot_table(zom, index=['Cuisine','Price'],values=['Rating','Vot
es'], aggfunc=np.average, fill_value=0)
cuisdata_c
```

Out[18]:

		Rating	Votes
Cuisine	**Price**		

Afghani	**2**	2.900000	39.000000
African	**4**	4.450000	319.000000
American	**1**	3.847826	214.956522
	2	3.428873	276.485915
	3	3.947761	590.507463
	4	3.910526	570.210526
Andhra	**1**	3.400000	139.000000
	2	3.550000	139.000000
Arabian	**1**	3.500000	95.000000
	2	3.000000	44.500000
	3	3.950000	226.000000
Argentine	**4**	4.500000	602.000000
Asian	**1**	3.930769	373.538462
	2	3.805263	179.473684
	3	3.984615	483.346154
	4	4.007143	385.857143
Asian Fusion	**4**	3.850000	73.500000
Assamese	**1**	2.800000	45.000000
	2	3.500000	218.000000
Australian	**2**	4.100000	87.000000
Awadhi	**2**	3.800000	86.000000
BBQ	**1**	4.116667	167.166667
	2	4.060000	445.600000
	3	4.500000	674.000000
	4	4.300000	39.000000
Bakery	**1**	3.285993	77.817590
	2	3.513514	152.342342
	3	3.469231	78.769231
	4	4.400000	13.000000
Bar Food	**1**	4.400000	546.000000
...	**...**	...	...
Sushi	**2**	3.675000	44.000000

	3	4.157143	203.428571
	4	3.700000	10.500000
Taiwanese	2	4.900000	161.000000
	3	4.400000	223.000000
Tapas	2	3.900000	502.000000
	4	4.900000	194.000000
Tea	1	3.040000	125.600000
	2	3.400000	18.000000
	3	3.500000	37.000000
	4	4.100000	68.000000
Tex-Mex	3	4.000000	911.000000
Thai	1	4.050000	207.500000
	2	3.760000	152.800000
	3	4.011111	205.555556
	4	4.137500	183.625000
Tibetan	1	2.971429	69.428571
	2	3.375000	232.750000
	3	3.700000	807.000000
Turkish	2	3.450000	133.750000
	3	4.100000	788.000000
	4	4.300000	43.000000
Turkish Pizza	2	4.300000	103.000000
Vietnamese	1	4.100000	270.000000
	2	4.300000	96.000000
	3	4.000000	83.000000
Western	3	4.200000	259.000000
	4	3.200000	32.000000
World Cuisine	3	4.550000	564.500000
	4	4.400000	115.000000

289 rows × 2 columns

Here, we further explore the relationship between price level and rating. We break down each cuisine by the restaurants in each price level and find the average rating and number of votes that correspond to them. From eying the pivot table and looking at a few cuisines that have restaurants in all 4 price levels, it's clear that our second hypothesis definitely does not hold. American restaurants at price level 1 had a rating (3.85) almost the same as those with price level 4 (3.91). The same goes for Asian and Thai cuisine. The average rating at price level 1 is very close to that at price level 4. If our hypothesis had held, we would have seen the rating increase as price level increases. but that is not the case in the data set. Fine dining and higher prices do not equate to a better dining experience for Zomato users.

Data Visualization

Though we have just tested our hypotheses using pivot tables, we'd like to visualize our data to analyze our hypotheses in a different lens. First we'll look at the relationship between price and rating.

In [21]:

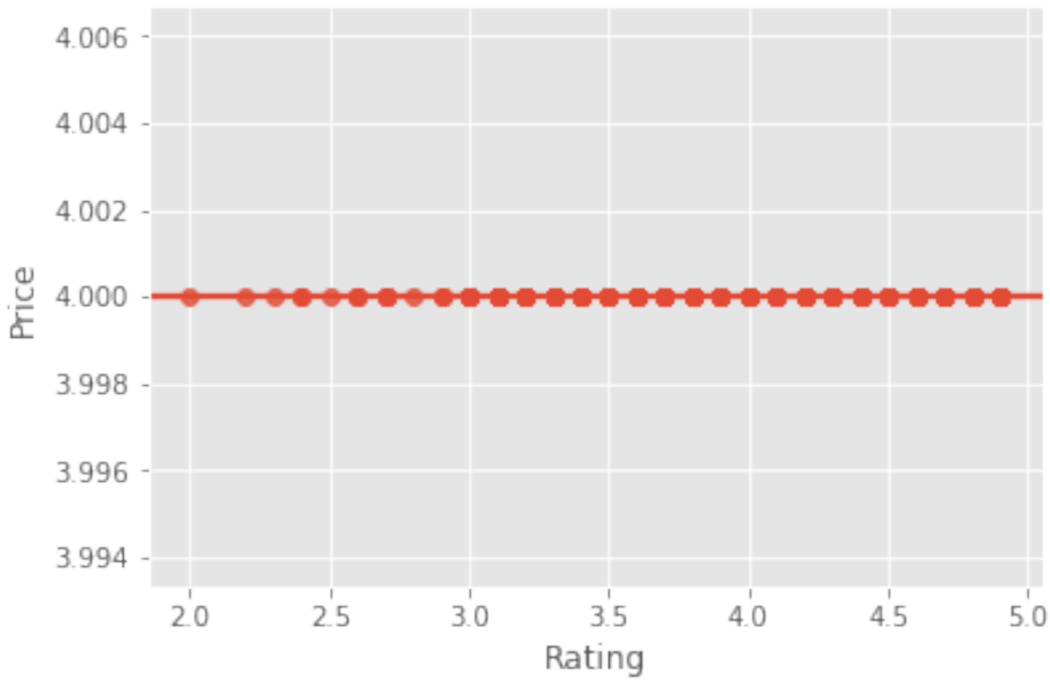
```
#setup for data visualization

import sys
import matplotlib.pyplot as plt #plotting visualizations
from math import pi
import datetime
from scipy.stats.stats import linregress
import seaborn as sns #statistical graphs

plt.style.use("ggplot") #data vizualization

%matplotlib inline

import warnings
warnings.filterwarnings('ignore') #ignore warnings that may pop up in plotting
```

In [22]:

```
high_price = zom.loc[zom['Price']==zom['Price'].max(),:]
sns.regplot(x = high_price['Rating'], y=high_price['Price'])
```

Out[22]:

```
<matplotlib.axes._subplots.AxesSubplot at 0xab61c4b240>
```



Here, we find all restaurants with a price level of 4 and plot their ratings on the x-axis. Under our second assumption, we should see a very heavy right skew in the dots, but the graph clearly shows that these restaurants have ratings that are widely distributed; thus, our second hypothesis does not hold.

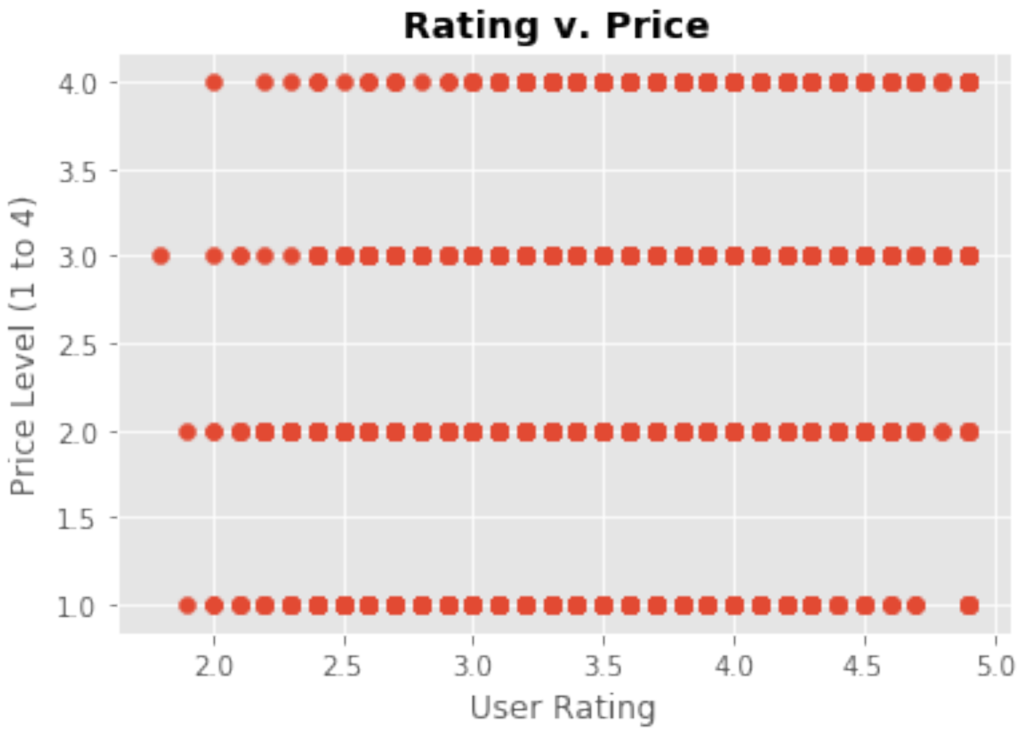
Now, let's compare the other price levels to their respective ratings.

In [23]:

```
fig, ax = plt.subplots()
ax.scatter(zom["Rating"], zom["Price"])

ax.set_title('Rating v. Price ', loc='center', fontsize=14, fontweight = "bold")


ax.set_xlabel("User Rating")
ax.set_ylabel("Price Level (1 to 4)")
```

Out[23]:

```
Text(0,0.5,'Price Level (1 to 4)')
```



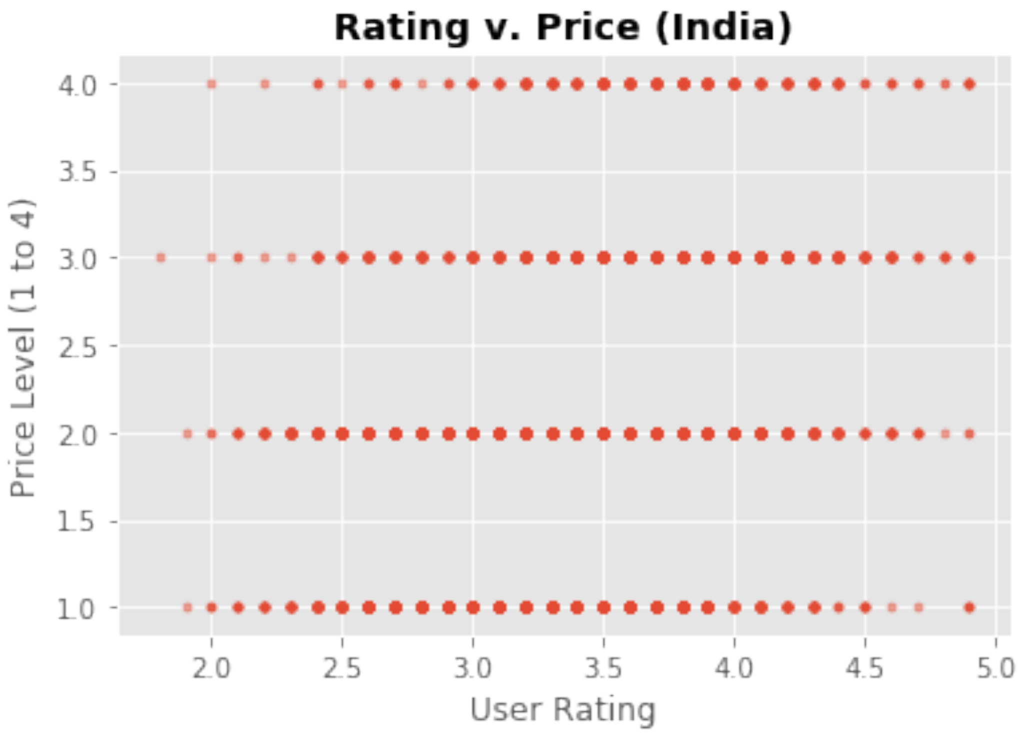
Again, we see that price level really has no correlation with overall ratings. All price levels exhibit no skew to a low or high rating.

In [25]:

```
in_rst = zom.loc[zom['Country'] == 'India',:]

fig, ax = plt.subplots()
ax.scatter(in_rst["Rating"], in_rst["Price"], alpha= 0.50,
           s=10)

ax.set_title('Rating v. Price (India)', loc='center', fontsize=14, fontweight =
"bold")


ax.set_xlabel("User Rating")
ax.set_ylabel("Price Level (1 to 4)")
```

Out[25]:

```
Text(0,0.5,'Price Level (1 to 4)')
```



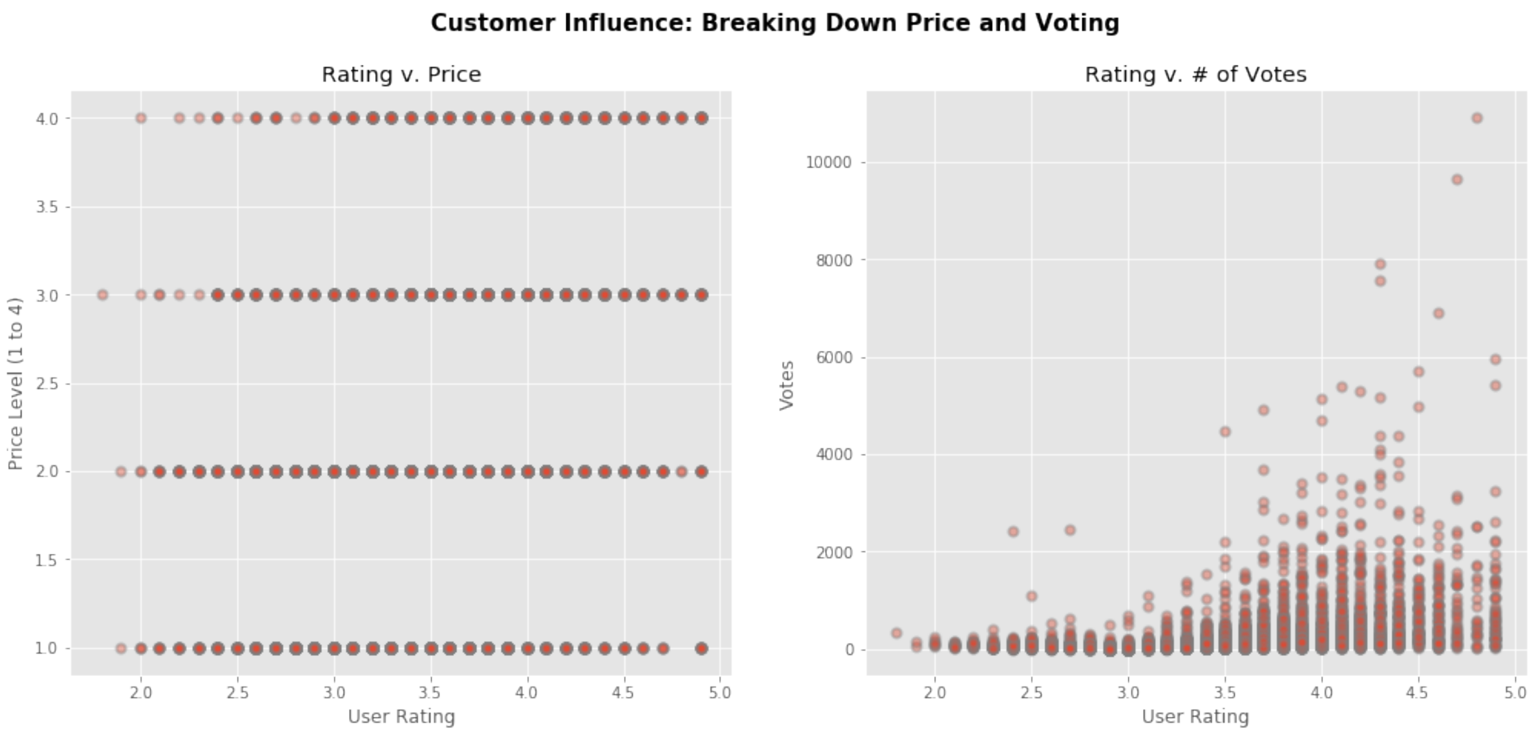
Now, we look at only India's restaurants (as it has the greatest share of the sample size) to see if the same result occurs: it does. We can garner no clear correlation between the price level of the restaurant and its user rating.

In fact, all ratings at each price level seem to concentrate between 3 and 4. This is actually an interesting phenomenon in platforms that allow consumers to leave reviews (eBay, Yelp, etc.). All factors (price, cuisine, etc.) notwithstanding, the probability a user will leave a certain rating is the same; that rating varies from platform to platform, but the pattern remains the same.

We can plot a histogram comparing user ratings on restaurants and the number of votes that begot the rating.

In [26]:

```
fig, (ax1, ax2) = plt.subplots(1, 2, sharex ="col", figsize = (17, 7))
plt.xlabel("the X axis")
plt.ylabel("the Y axis")
plt.suptitle("Customer Influence: Breaking Down Price and Voting", fontsize = 15
, fontweight = "bold")

ar = np.float64(zom["Rating"])
pr = np.float64(zom["Price"])
v = np.float64(zom["Votes"])

ax1.scatter(ar, pr, cmap="Blues", alpha=0.4, edgecolors="grey", linewidth=2) #ad
d titles (main and on axis)

ax1.set_title("Rating v. Price")
ax1.set_xlabel('User Rating')
ax1.set_ylabel('Price Level (1 to 4)')

ax2.scatter(ar, v, cmap="Blues", alpha=0.4, edgecolors="grey", linewidth=2)
ax2.set_title("Rating v. # of Votes")
ax2.set_xlabel('User Rating')
ax2.set_ylabel('Votes')

plt.show()
```

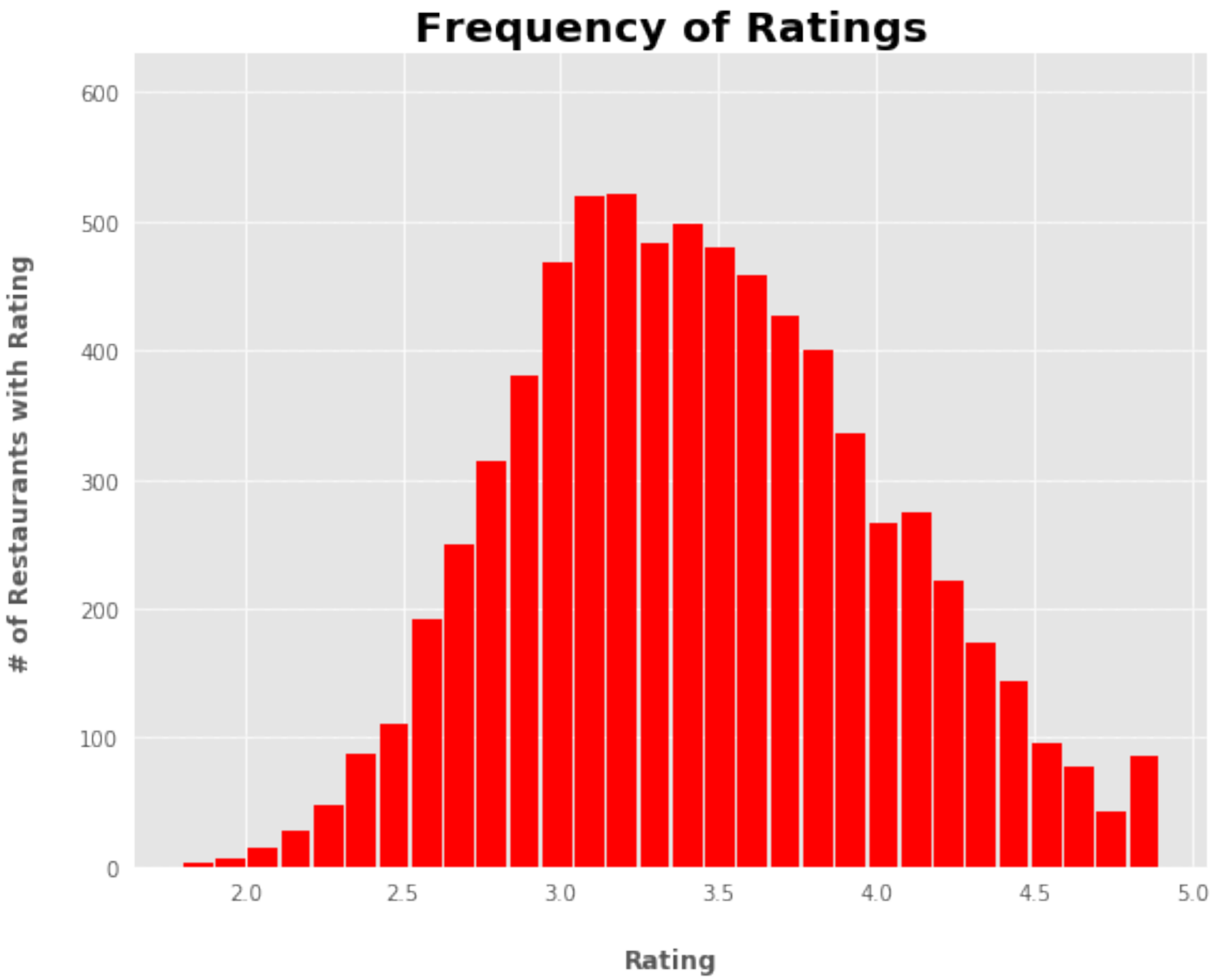


Here we actually see that restaurants that received more than 2,000 votes tend to achieve overall ratings between 4 and 5, not 3 and 4. Below the 2,000 vote mark, it's a little hard to tell, but it's clear that restaurants with ratings less than 3.5 do not get a large number of votes. From the right graph, we can conclude that a restaurant with more than 2,000 votes is more likely to have a higher rating.

Now, we will make a histogram showing the frequency with which an individual restaurant achives a certain rating.

In [27]:

```
ax = zom.hist(column='Rating', bins=30, grid=True, figsize=(9,7), color='r', zor
der=2, rwidth=0.9)

ax = ax[0]
for x in ax:

    #despine
    x.spines['right'].set_visible(False)
    x.spines['top'].set_visible(False)
    x.spines['left'].set_visible(False)

    #switch off ticks
    x.tick_params(axis="both", which="both", bottom="off", top="off", labelbotto
m="on", left="off", right="off", labelleft="on")

    #draw horizontal axis lines
    vals = x.get_yticks()
    for tick in vals:
        x.axhline(y=tick, linestyle='dashed', alpha=0.4, color='#eeeeee', zorder
=1)

    #set main title
    x.set_title("Frequency of Ratings", weight='bold', size=20)

    #set x-axis label
    x.set_xlabel("Rating", labelpad=20, weight='bold', size=12)

    #set y-axis label
    x.set_ylabel("# of Restaurants with Rating", labelpad=20, weight='bold', siz
e=12)
```



In [28]:

```
zom[['Price','Rating']].corr()
```

Out[28]:

	Price	Rating
Price	1.000000	0.403169
Rating	0.403169	1.000000

This is more in line with our previous observation that all ratings seemed to concentrate between a rating of 3 and 4 (more accurately, between 2.9 and 3.8). Very few restaurants received a rating below 2.50; there is definitely a skew towards the higher ratings.

Above we also explicitly calculate the correlation between price and rating, and the conclusions from our graphs match the result. There is no convincing correlation, but restaurants with a higher price probably won't have a relatively low rating.

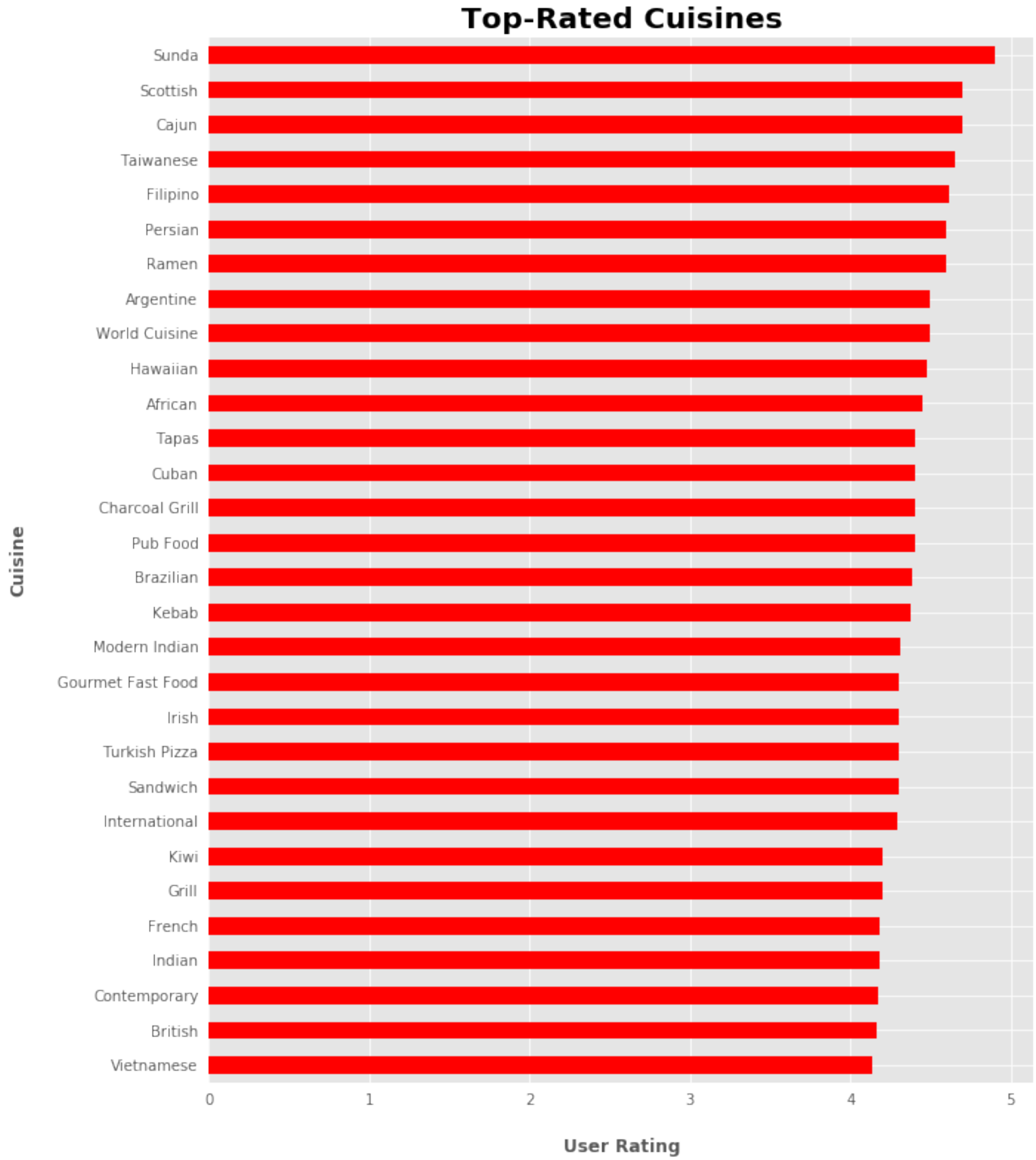
Now we want to visualize cuisine and ratings.

In [29]:

```
x1 = zom.groupby('Cuisine')['Rating'].mean().sort_values().tail(30)

ax = x1.plot(kind='barh', figsize=(10, 13), color='r', zorder=2, width=0.5)

#despine
ax.spines['right'].set_visible(False)
ax.spines['top'].set_visible(False)
ax.spines['left'].set_visible(False)
ax.spines['bottom'].set_visible(False)

#switch off ticks
ax.tick_params(axis="both", which="both", bottom="off", top="off", labelbottom="
on", left="off", right="off", labelleft="on")

#set main title
ax.set_title("Top-Rated Cuisines", weight='bold', size=20)

#set x-axis label
ax.set_xlabel("User Rating", labelpad=20, weight='bold', size=12)

#set y-axis label
ax.set_ylabel("Cuisine", labelpad=20, weight='bold', size=12)
```

Out[29]:

```
Text(0,0.5,'Cuisine')
```



This is the shows the same information as cuisdata_r pivot table. We expanded it to show the top 30 cuisines so that we could find French cuisine, which ranks at 26th. This is quite low, as there over 100 cuisines. Again, we cannot find convincing data that the connotation of French cuisine to fine dining and good dining experiences holds true.

Conclusion & Next Steps

Through this project, we have conclude that cuisines that are usually considered high-dining experiences like French and Japanese are not necessarily the most expensive and are definitely not the highest-rated. Moreover, from the data set, the cuisines with highest average rating were not the most expensive. This data sample from Zomato effectively disproves our two original hypotheses that we inferred based on prior knowledge and experiences.

We do note a few limitations in the data:

- Zomato is based in India, and this data set is highly concentrated on restaurants in the country
- India's restaurant ecosystem should not be taken as representative of the global restaurant industry
- Haute cuisines like French and Japanese may not be fully developed in the Indian market (or there may not be a big enough market for them yet)
- This data set by no mean represents the population of Zomato votes; it's merely a sample
- We did cull quite a bit of the data when cleaning it, so the sample became further constrained
- The cuisines we chose as the representative cuisine for the restaurant may not have been truly representative

To further explore this topic, we propose these next steps:

- Compare data sets from other restaurant search platforms like TripAdvisor and Yelp
- Find a representative sample for the global restaurant industry
- Further research the implicit biases and platform dynamics that occur when leaving review/ratings

Ultimately, this project is quite constrained, and the analysis performed was elementary. Further insight can be gained by expanding the data set and using more complex data analysis tools to further test relationships between restaurant qualities and ratings.

GitHub Link

https://github.com/ss9270/Data_Bootcamp_Final_Project
(https://github.com/ss9270/Data_Bootcamp_Final_Project)