Web Scraping Report

Topic: Restaurants in Baltimore TripAdvisor (https://www.tripadvisor.com/Restaurants-g60811-Baltimore_Maryland.html)

Project Idea: Web scraping of restaurant data of a certain city can help future restaurant owners to know things like what type of cuisine is most preferred, deciding on food prices, etc. Furthermore, it also helps current restaurant owners to understand recent trend and choices of customers. This information is fruitful to increase the business and make changes based on the need.
Future Scope of this data is to perform sentiment analysis on customer reviews to understand positive and negative scores in reviews. Also, to understand what describes restaurant the best.

Execution of the project Idea: This can be found in Web_Scraping_Project.ipynb file.

The steps used for execution of project are as follows:

- Import Dependencies
- Specifying headers which will be used for get request.
- Initialize all the lists which will be used for visualizations.
- For the restaurants in Baltimore, scrape the TripAdvisor website to fetch restaurant names, ratings, number of reviews, price, cuisine. In two loops for each page request and for each restaurant- container in it, scraping the above-mentioned content.
- Convert these lists of scraped data into a dataframe which can be used for visualizations.
- Use regular expressions to clean and process the raw data.
- Perform data analysis and visualization

Results and Visualizations:

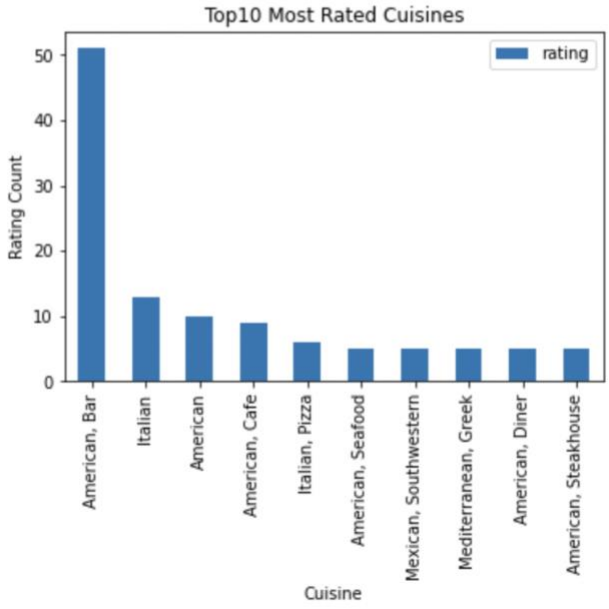
1. Cleaned and Pre-processed Data

	Unnamed: 0	restaurant	cuisine	price	rating	total review count
0	0	DiPasqualesItalianMarketplaceDeli	Italian, Pizza	$$ - $$$	5.0	694
1	1	JoeBennysFocacceria	Italian, Pizza	$$ - $$$	5.0	215
2	2	CharlestonRestaurant	American	$$$$	4.5	943
3	3	LaTavola	Seafood, Italian	$$ - $$$	4.5	682
4	4	ThamesStreetOysterHouse	American, Bar	$$ - $$$	4.5	2105

2. Top10 Most Rated Cuisines

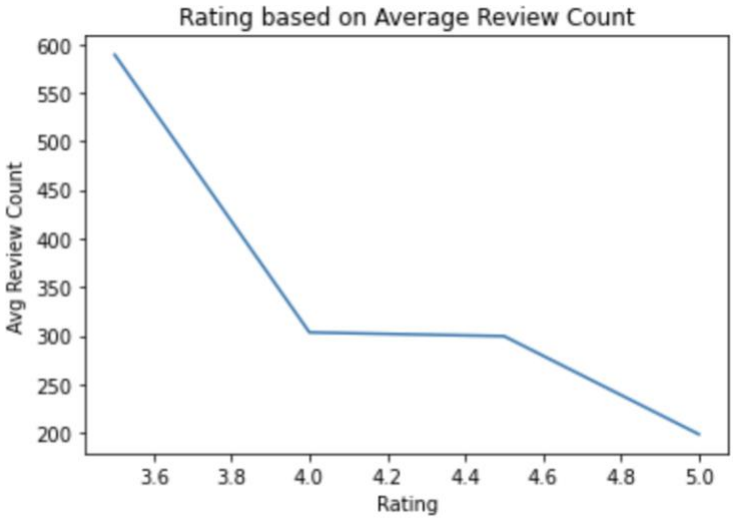
In [24]:
```
# Bar Chart for 10 top most ratings
plot = topRating.plot(kind='bar')
plot.set_title('Top10 Most Rated Cuisines')
plot.set_xlabel("Cuisine")
plot.set_ylabel("Rating Count")
#plt.savefig('Best Reviewed.png')
#plt.show()
```

Out[24]:
```
Text(0, 0.5, 'Rating Count')
```



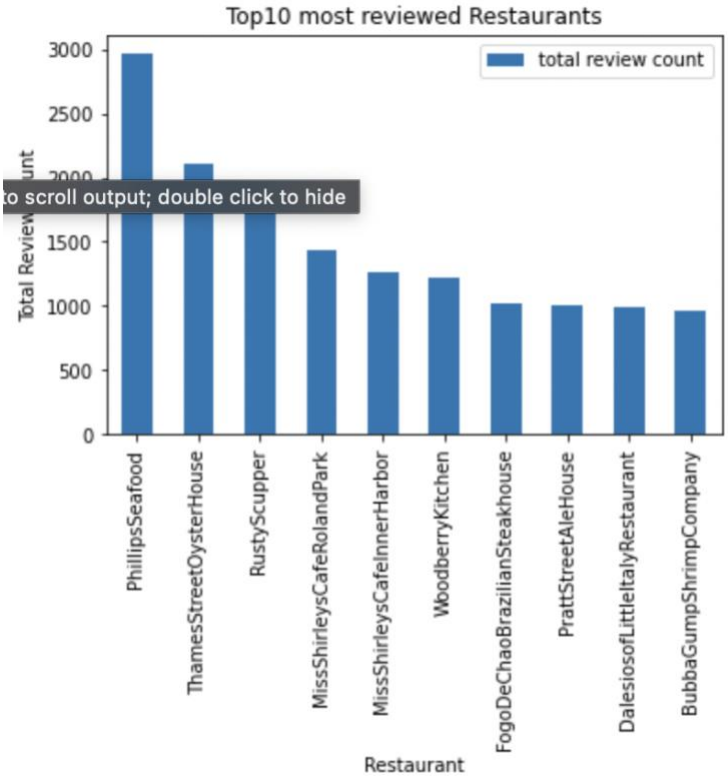
American Bar was rated by almost 50 customers. This can also conclude that people visit American Bar more often than any other cuisine type.

3. Rating based on Average Review Count



People who reviewed most restaurants had rating of 3.6 while restaurants with 5.0 rating had the least reviews. The difference between these two is more than double. It will be interesting to understand nature of these reviews i.e., whether they fall in good or bad category.

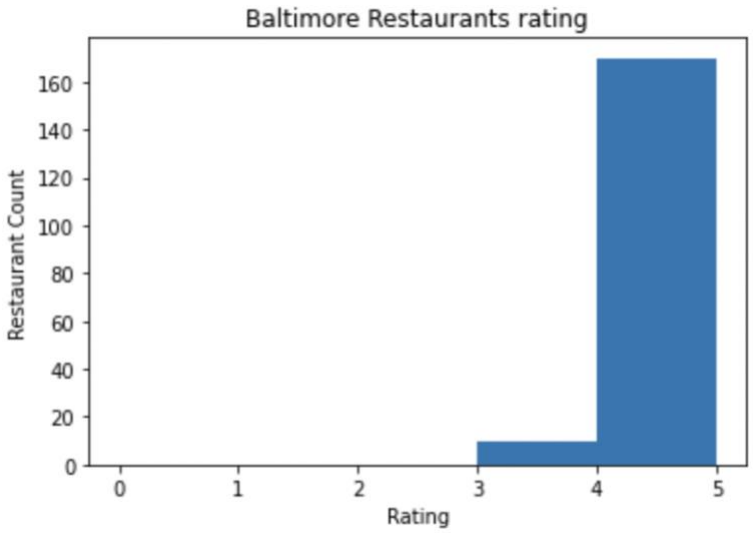
4. Top10 most reviewed Restaurants



	Unnamed: 0	restaurant	cuisine	price	rating	total review count
63	63	PhillipsSeafood	American, Bar	$	4.0	2966
4	4	ThamesStreetOysterHouse	American, Bar	$	4.5	2105
28	28	RustyScupper	American, Seafood		4.0	1967
13	13	MissShirleysCafeRolandPark	American	$	4.5	1434
17	17	MissShirleysCafeInnerHarbor	American	$	4.5	1267
29	29	WoodberryKitchen	American		4.5	1217
26	26	FogoDeChaoBrazilianSteakhouse	Steakhouse, Brazilian		4.5	1018
54	54	PrattStreetAleHouse	American, Bar	$	4.0	1010
5	5	DalesiosofLittleItalyRestaurant	Italian	$	4.5	993
96	96	BubbaGumpShrimpCompany	American, Bar	$	3.5	969

Phillips Seafood is the most reviewed restaurant and has 4.0 rating. If we look at the dataframe shown above, then we can see that most of the restaurants serve American cuisine and have rating between 4 - 4.5.

5. **Baltimore Restaurants rating**



The scraped restaurants mostly fall into rating ranging between 4 - 5.

Conclusion: Thus, web scraping of restaurants in Baltimore from TripAdvisor website was executed successfully. The analysis and visualization of the scraped data is also implemented.