QUESTION 1

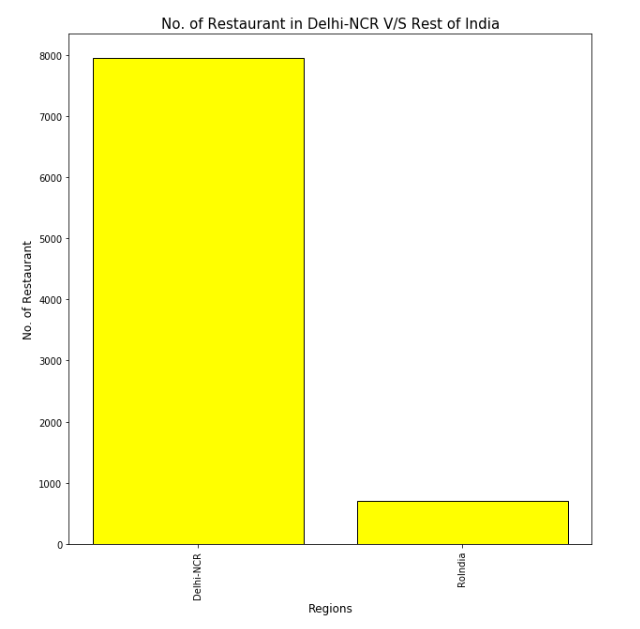
1. The dataset is highly skewed toward the cities included in Delhi-NCR. So, we will summarise all the other cities in Rest of India while those in New Delhi, Ghaziabad, Noida, Gurgaon, Faridabad to Delhi-NCR. Doing this would make our analysis turn toward Delhi-NCR v Rest of India.
   1. Plot the bar graph of number of restaurants present in Delhi NCR vs Rest of India.
   2. Find the cuisines which are not present in restaurant of Delhi NCR but present in rest of India.Check using Zomato API whether this cuisines are actually not served in restaurants of Delhi-NCR or just it due to incomplete dataset.
   3. Find the top 10 cuisines served by maximum number of restaurants in Delhi NCR and rest of India.
   4. Write a short detailed analysis of how cuisine served is different from Delhi NCR to Rest of India. Plot the suitable graph to explain your inference.

I a) Plot the bar graph of number of restaurants present in Delhi NCR vs Rest of India.

Code:



Graph:



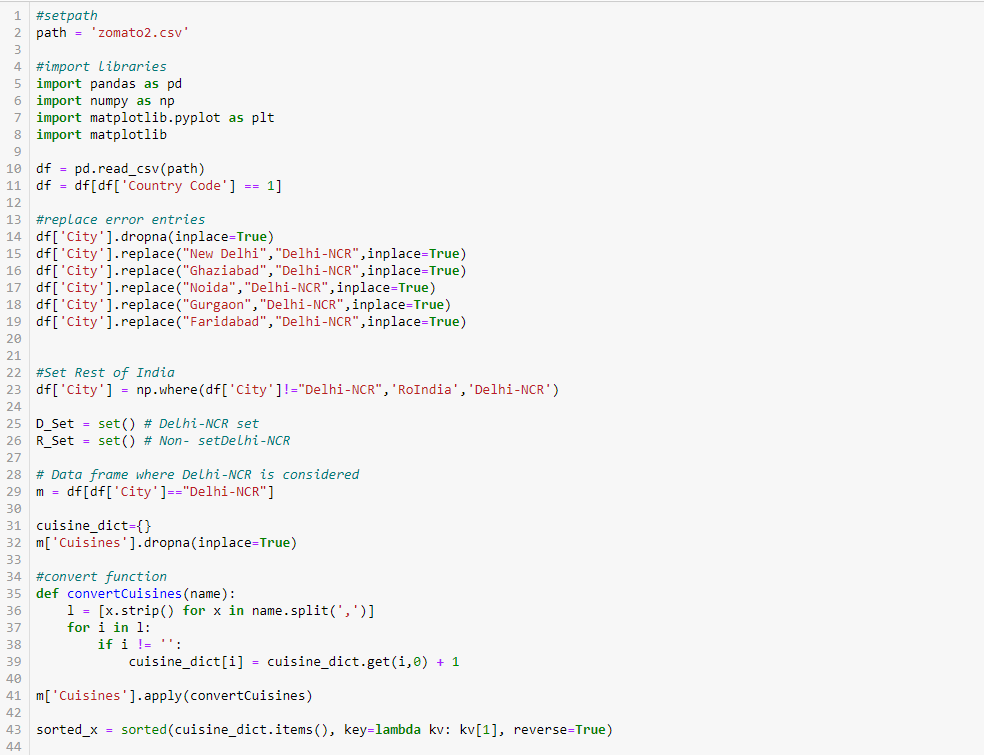
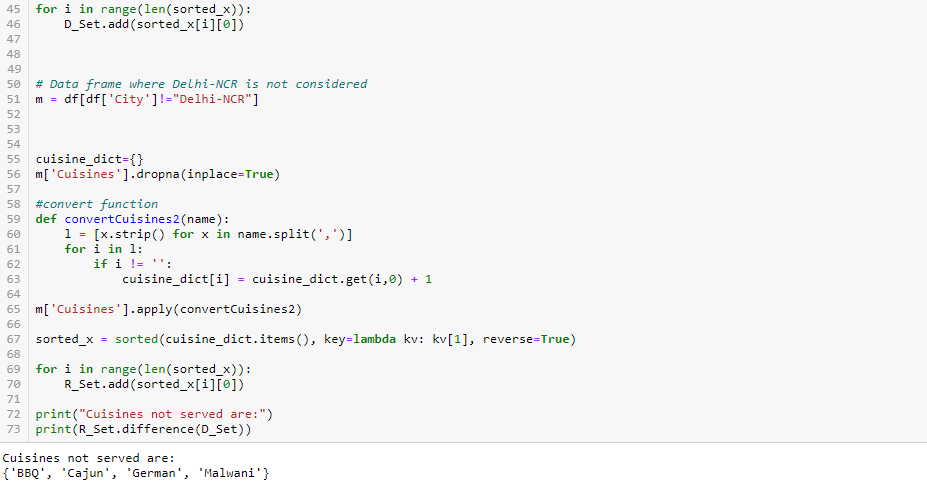
Justification:

All the cities i.e. New Delhi, Ghaziabad, Noida, Gurgaon, Faridabad have been renamed to Delhi-NCR. Also all other cities name are converted to RestofIndia. A simple function of pandas value\_counts() give us expected results. Plot is easily implemented on the output.

Libraries used:

* import pandas as pd
* import numpy as np
* import matplotlib.pyplot as plt
* import matplotlib

I b) Find the cuisines which are not present in restaurant of Delhi NCR but present in rest of India. Code:

Justification:

In the output it can be seen cuisines not served are:

BBQ, Cajun, German, Malwani

Check using Zomato API whether this cuisines are actually not served in restaurants of Delhi-NCR or just it due to incomplete dataset.

Code:



Justification:

In the output:

Code for Delhi is used to find the cuisine\_id for Malwani.

Checking if any Malwani cuisine restaurant is present in Delhi on Zomato API.

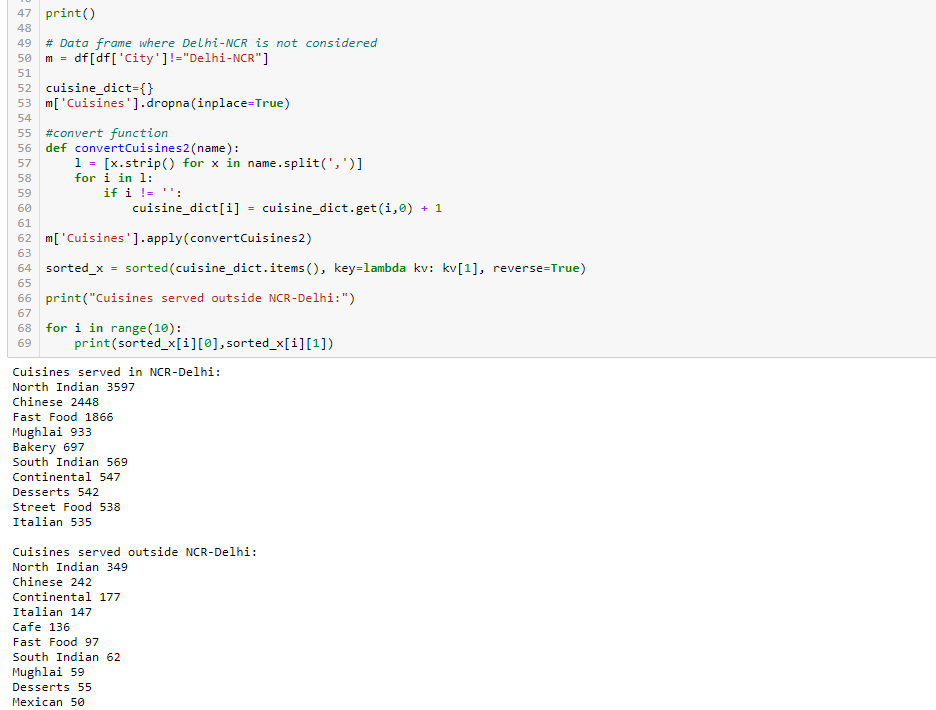
Yes a restaurant is present in Delhi serving Malwani food.

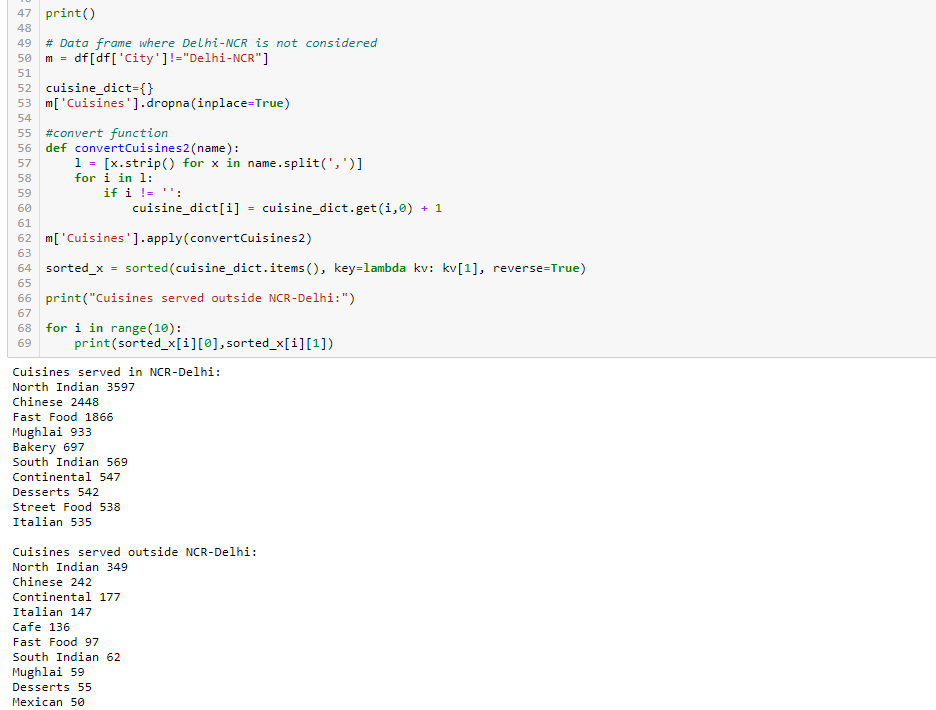
Checked using Zomato API the cuisines are served but error is due to incomplete dataset

I c) Find the top 10 cuisines served by maximum number of restaurants in Delhi NCR and rest of India.

Code:







Justification:

In the output:

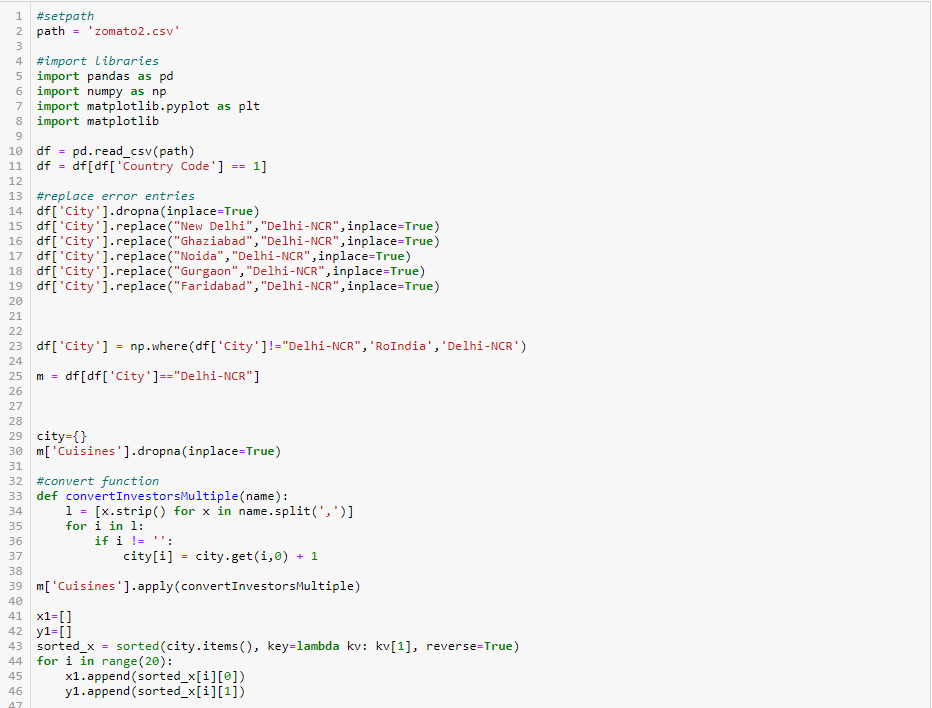
All the cities i.e. New Delhi, Ghaziabad, Noida, Gurgaon, Faridabad have been renamed to Delhi-NCR. Also all other cities name are converted to RestofIndia. A simple function of pandas value\_counts() give us expected results.

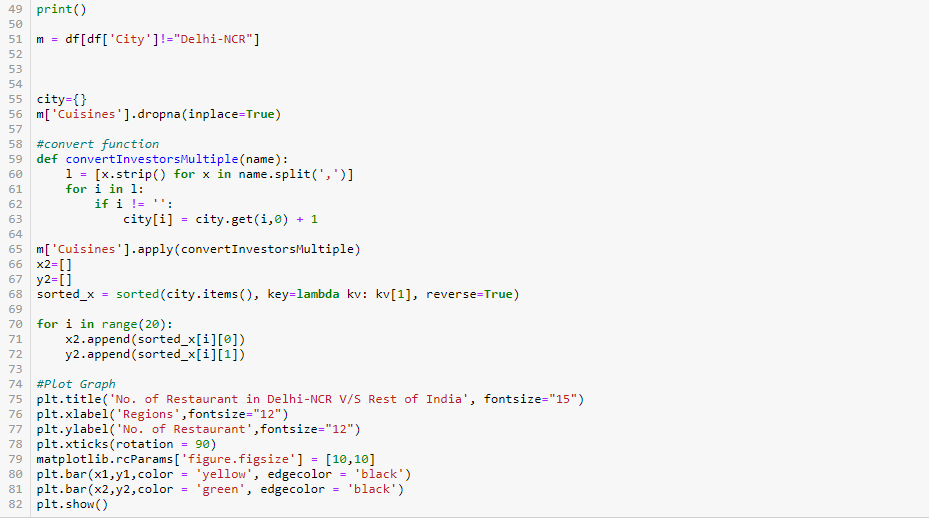
A dictionary cuisine\_dict counts the cuisines present in Delhi NCR.

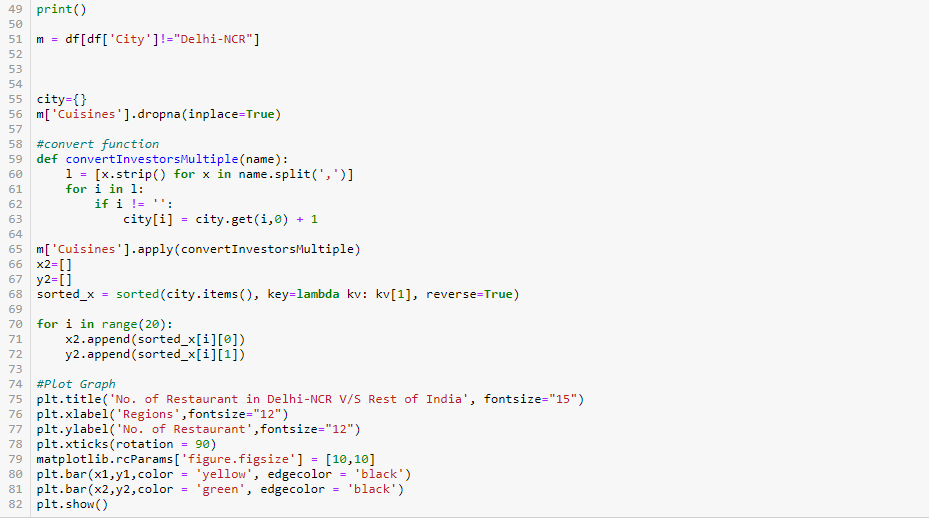
A simple convert function is applied on column ‘Cuisines’ to get desired output for Delhi NCR and Rest of India.

I d) Write a short detailed analysis of how cuisine served is different from Delhi NCR to Rest of India. Plot the suitable graph to explain your inference.

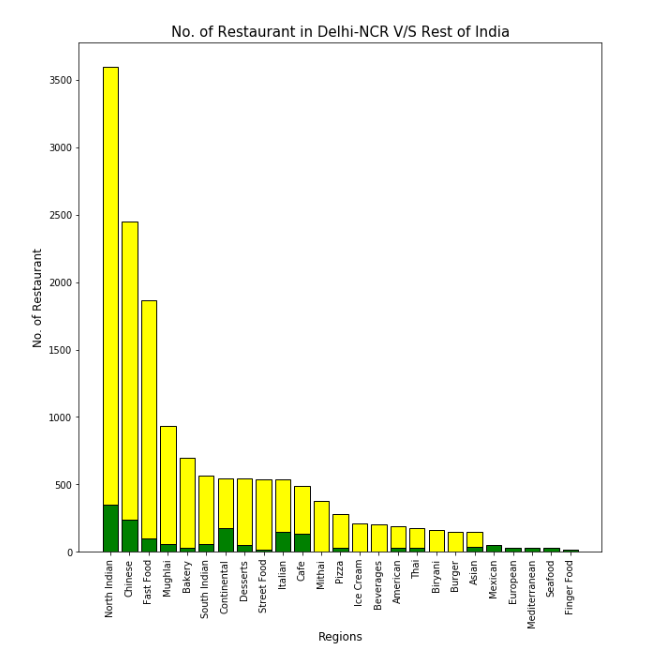
Code:







Plot:



Justification:

**Prominent Cuisnes served are [Top 20]:**

Cuisines served in NCR-Delhi:

North Indian, Chinese, Fast Food, Mughlai, Bakery, South Indian, Continental, Desserts, Street Food, Italian, Cafe, Mithai, Pizza, Ice Cream, Beverages, American, Thai, Biryani, Burger, Asian

Cuisines served outside NCR-Delhi:

North Indian, Chinese, Continental, Italian, Cafe, Fast Food, South Indian, Mughlai, Desserts, Mexican, Asian, European, American, Pizza, Mediterranean, Thai, Bakery, Seafood, Finger Food, Street Food

More rare and exotic food don’t make it to top 20 in Delhi-NCR.:

Mexican, European, Mediterranean, Seafood, Finger Food aren’t served.

While people of Delhi are mostly fond of:

Mughali, Biryani and Burger Joints

QUESTION 2

1. User Rating of a restaurant plays a crucial role in selecting a restaurant or ordering the food from the restaurant.
   1. Write a short detail analysis of how the rating is affected by restaurant due following features: Plot a suitable graph to explain your inference.
      1. Number of Votes given Restaurant
      2. Restaurant serving more number of cuisines.
      3. Average Cost of Restaurant
      4. Restaurant serving some specific cuisines.
   2. Find the weighted restaurant rating of each locality and find out the top 10 localities with more weighted restaurant rating?
      1. Weighted Restaurant Rating=Σ (number of votes \* rating) / Σ (number of votes) .

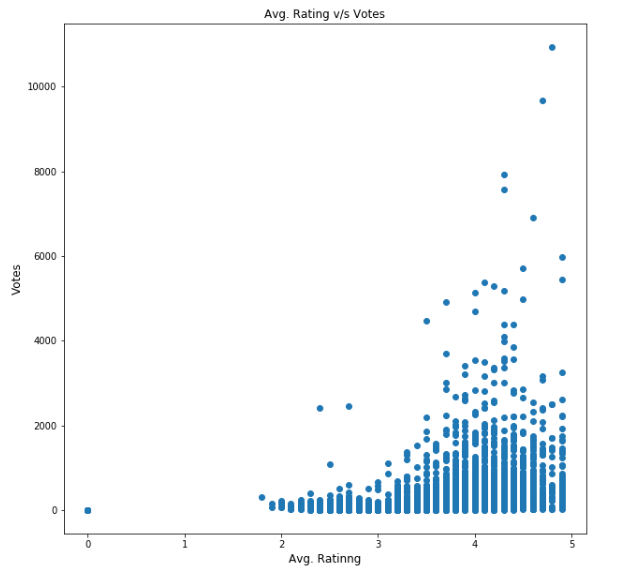
Write a short detail analysis of how the rating is affected by restaurant due following features: Plot a suitable graph to explain your inference.

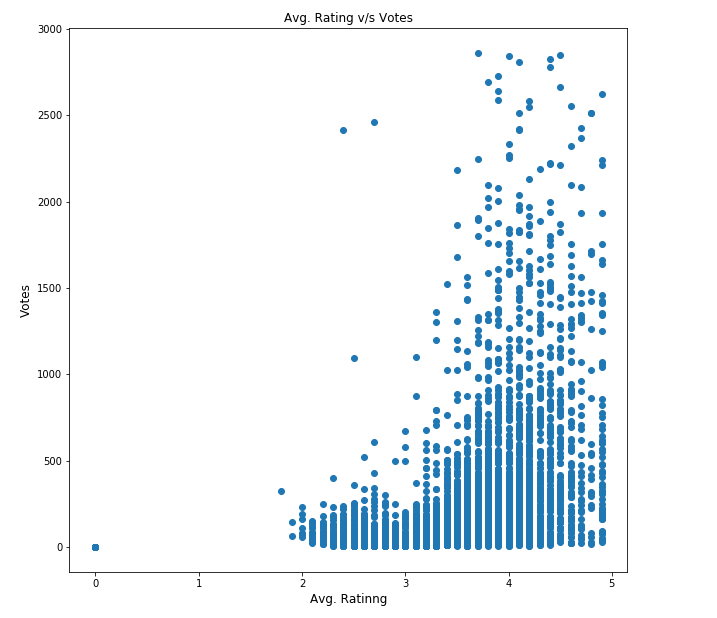
II a) i) Number of Votes given Restaurant

Code:



Plot:





Justification:

As can be seen in the output:

Most votes are concentrated around 4.0 Rating. In general more the number of votes higher the rating.

It can be easily seen that as votes increase there is a concertation around high numbered Rating.

There are dots around (0,0) i.e. no votes no rating.

Libraries used:

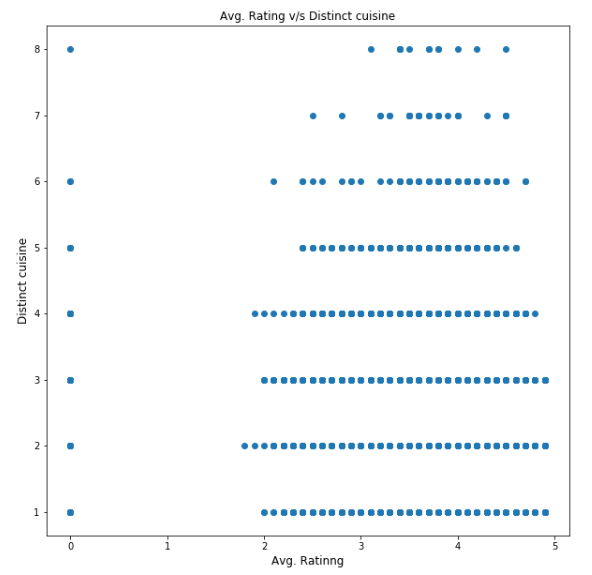
* import pandas as pd
* import numpy as np
* import matplotlib.pyplot as plt
* import matplotlib

II a) ii) Restaurant serving more number of cuisines.

Code:



Plot:



Justification:

As can be seen in the output:

Most Ratings are concentrated around 2.0 - 5.0 Rating. In general more the number of cuisines ratings lie in higher range 3.0 – 5.0.

It can be easily seen that as cuisines increase there is a concertation around high numbered Rating.

Libraries used:

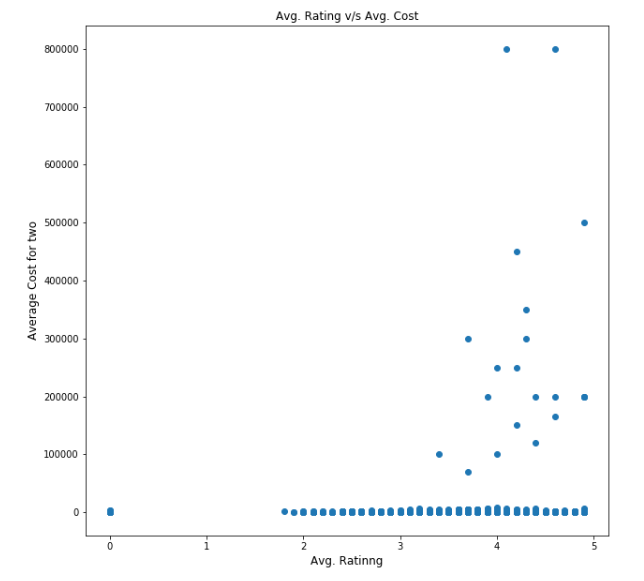
* import pandas as pd
* import numpy as np
* import matplotlib.pyplot as plt

II a) iii) Average Cost of Restaurant

Code:



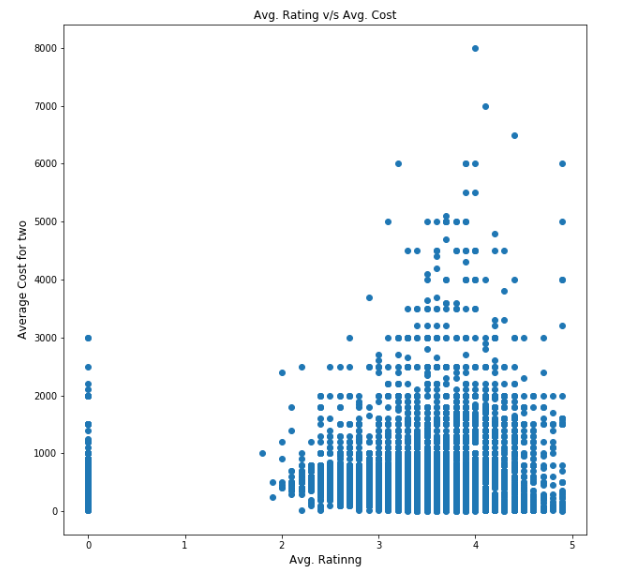
Plot:



Code:



Plot:



Justification:

As can be seen in the output:

Most Ratings are concentrated around 2.0 - 5.0 Rating. In general more Avg. cost ratings lie in higher range 3.0 – 5.0 . There is higher concentration around 4.0.

It can be easily seen that as Avg. cost increase there is a concertation around high numbered Rating.

There are dots around (:,0) i.e. no votes no rating.

Libraries used:

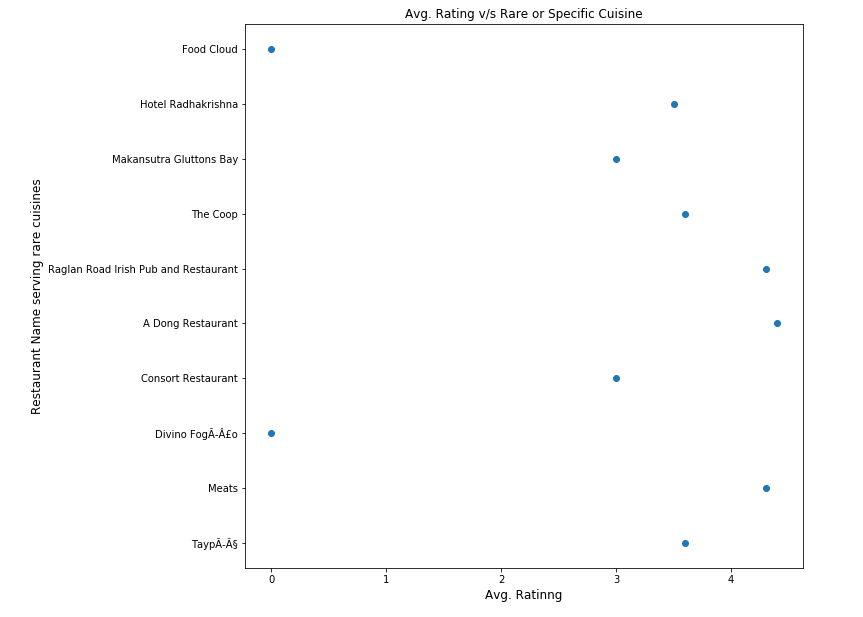
* import pandas as pd
* import numpy as np
* import matplotlib.pyplot as plt
* import matplotlib

II a) iv) Restaurant serving some specific cuisines.

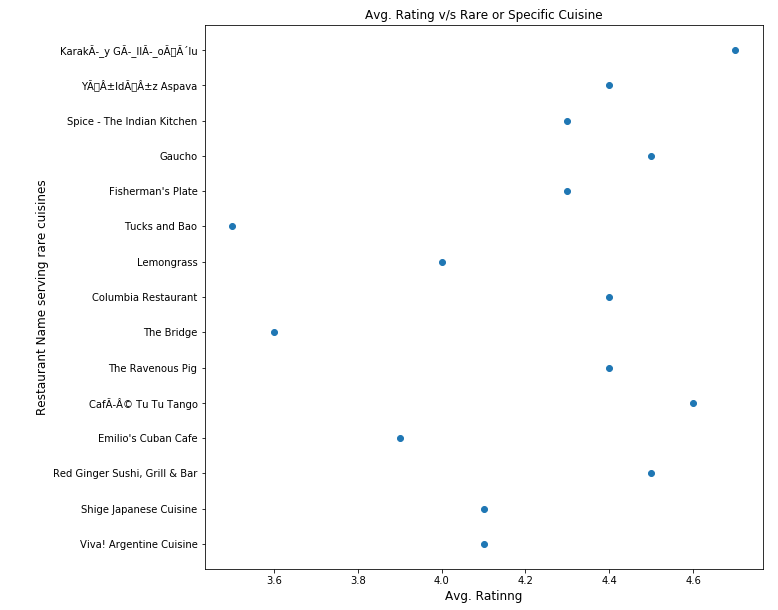
Code:



Plot: (Top 10 Rare Cuisines)



Plot: (Top 10-20 Rare Cuisines)



Justification:

As can be seen in the output:

Most Ratings are concentrated around 3.0 - 5.0 Rating. In general more Rare Cuisines ratings lie in higher range 3.6 – 4.9 rating. There is higher concentration around 4.0. for top 10-20 rare cuisines.

It can be easily seen that for Rare Cuisines there is a concertation around high numbered Rating.

There are dots around (:,0) i.e. no votes no rating.

Libraries used:

* import pandas as pd
* import numpy as np
* import matplotlib.pyplot as plt
* import matplotlib

Summary:

Number of Votes given Restaurant

Most votes are concentrated around 4.0 Rating. In general more the number of votes higher the rating.

It can be easily seen that as votes increase there is a concertation around high numbered Rating.

There are dots around (0,0) i.e. no votes no rating.

Restaurant serving more number of cuisines.

Most Ratings are concentrated around 2.0 - 5.0 Rating. In general more the number of cuisines ratings lie in higher range 3.0 – 5.0.

It can be easily seen that as cuisines increase there is a concertation around high numbered Rating.

Average Cost of Restaurant

Most Ratings are concentrated around 2.0 - 5.0 Rating. In general more Avg. cost ratings lie in higher range 3.0 – 5.0 . There is higher concentration around 4.0.

It can be easily seen that as Avg. cost increase there is a concertation around high numbered Rating.

There are dots around (:,0) i.e. no votes no rating.

Restaurant serving some specific cuisines.

Most Ratings are concentrated around 3.0 - 5.0 Rating. In general more Rare Cuisines ratings lie in higher range 3.6 – 4.9 rating. There is higher concentration around 4.0. for top 10-20 rare cuisines.

It can be easily seen that for Rare Cuisines there is a concertation around high numbered Rating.

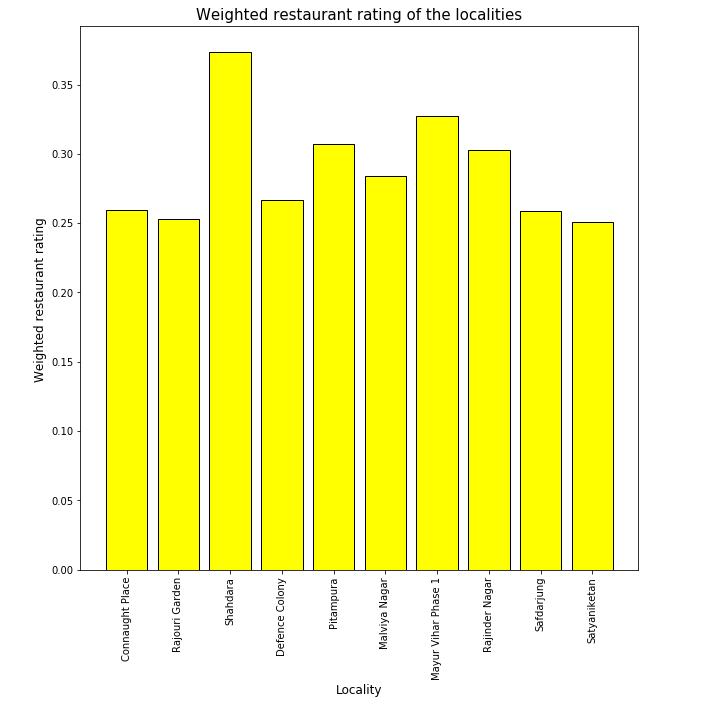
There are dots around (:,0) i.e. no votes no rating.

II b) i) a. Find the weighted restaurant rating of each locality and find out the top 10 localities with more weighted restaurant rating?

Code:



Plot:



Justification:

Weighted Restaurant Rating can be found using:

**With Pandas multiplication is simple due to Vectorization.**

Weighted Restaurant Rating =Σ (number of votes \* rating) / Σ (number of votes) .

Weighted restaurant ratings lie between 0.25 to 0.35 for localities.

Libraries used:

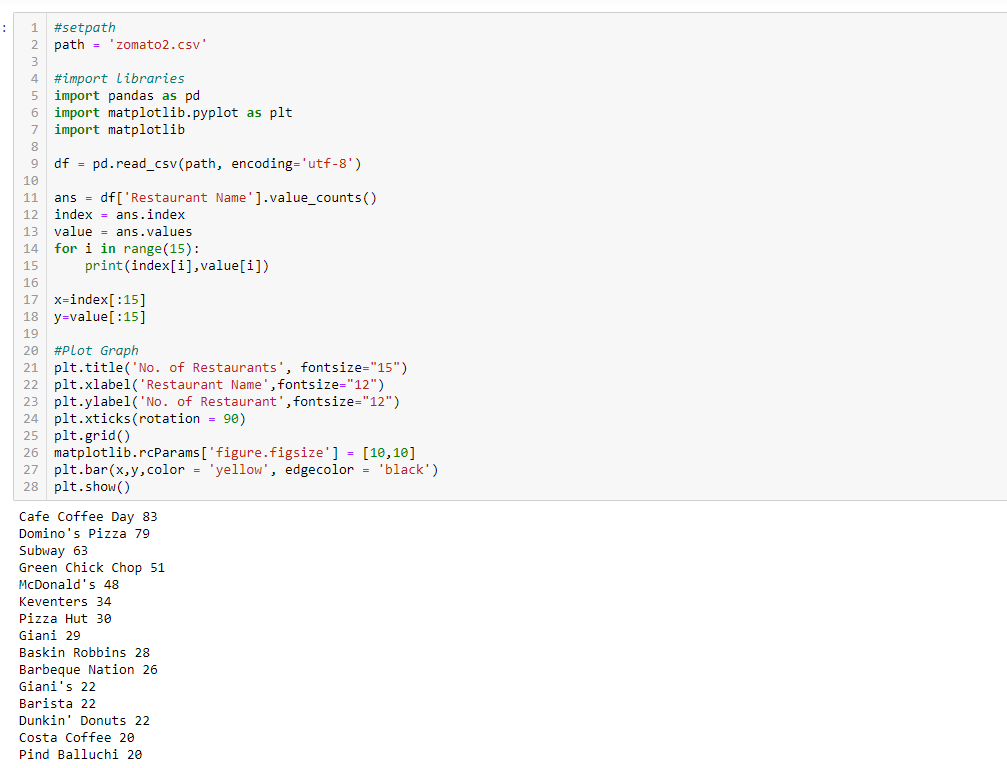
* import pandas as pd
* import numpy as np
* import matplotlib.pyplot as plt
* import matplotlib

QUESTION 3

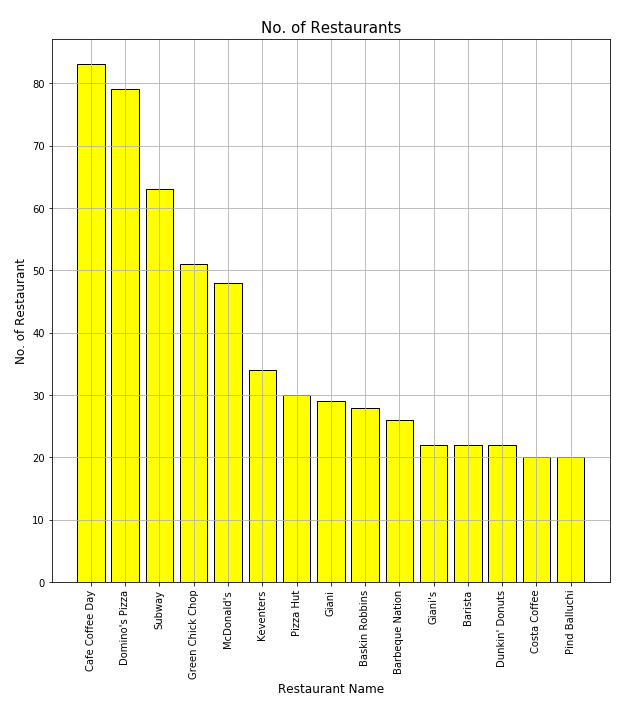
1. Visualization
   1. Plot the bar graph top 15 restaurants have a maximum number of outlets.
   2. Plot the histogram of aggregate rating of restaurant( drop the unrated restaurant).
   3. Plot the bar graph top 10 restaurants in the data with the highest number of votes.
   4. Plot the pie graph of top 10 cuisines present in restaurants in the USA.
   5. Plot the bubble graph of a number of Restaurants present in the city of India and keeping the weighted restaurant rating of the city in a bubble.

III a) Plot the bar graph top 15 restaurants have a maximum number of outlets.

Code:



Plot:



Justification:

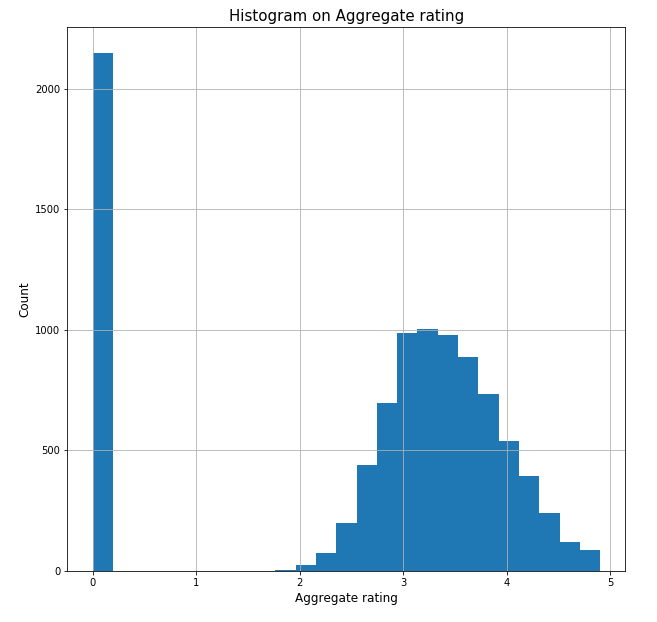
A simple function of pandas value\_counts() give us expected results. Plot is easily implemented on the Restaurant Name to find maximum outlets. Plot is easily implemented on the output using to\_list() funcotionaliy of Pandas.

III b) Plot the histogram of aggregate rating of restaurant( drop the unrated restaurant).

Code:



Plot:



Justification:

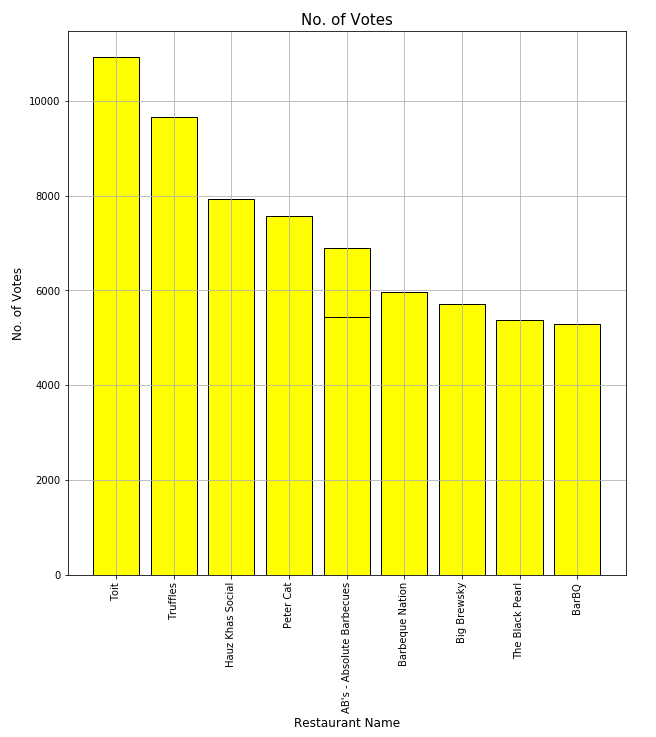
Plot is easily implemented on the output using to\_list() funcotionaliy of Pandas on Aggregate Ratings column. Histogram is aggregated around 3.5.

III c) Plot the bar graph top 10 restaurants in the data with the highest number of votes.

Code:



Plot:



Justification:

Plot is easily implemented on the output on column ‘Votes’ on basis of maximum no. of votes to find top 10 restaraunts.

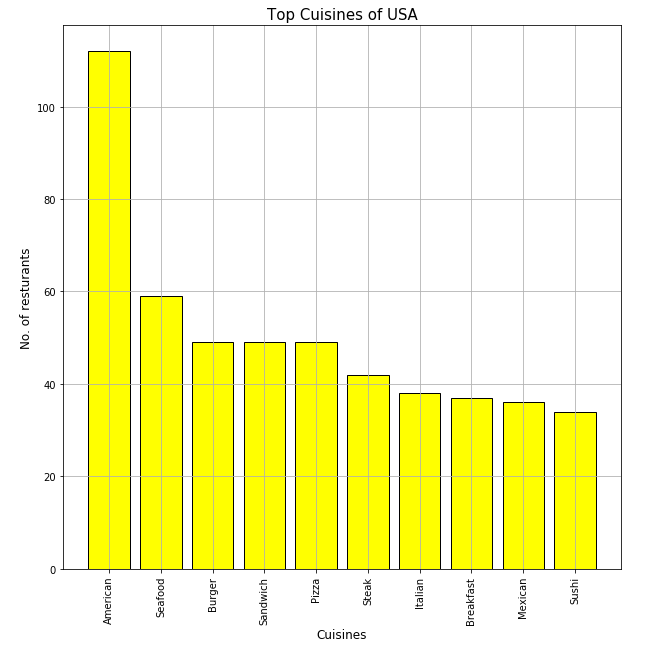
III d) Plot the pie graph of top 10 cuisines present in restaurants in the USA.

Code:





Plot:



Justification:

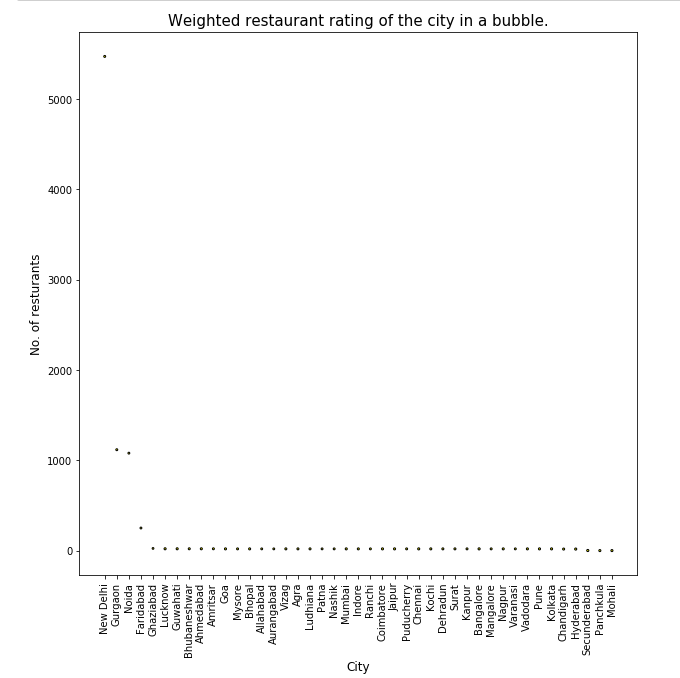
Plot is easily implemented on the output on column ‘Cuisines’ to find No. of Restaurants per cuisines.

II e) Plot the bubble graph of a number of Restaurants present in the city of India and keeping the weighted restaurant rating of the city in a bubble.

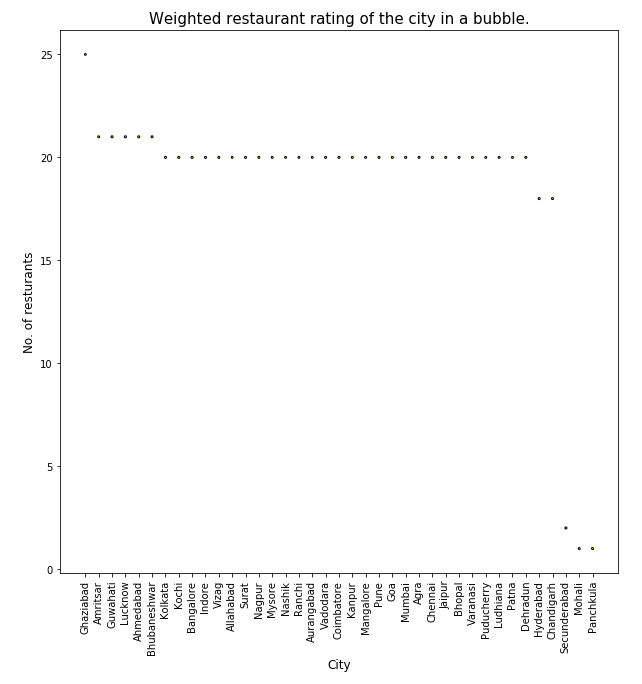
Code:

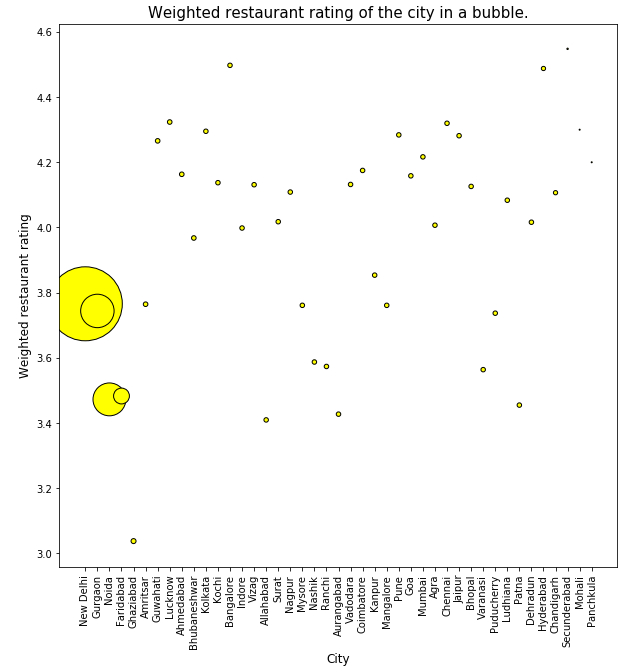


Plot:



No of outlets < 50:





Justification:

A simple function of pandas value\_counts() give us expected results on column City. Plot is easily implemented on the Weighted Restaurant Rating to find maximum outlets.

Weighted Restaurant Rating can be found using:

**With Pandas multiplication is simple due to Vectorization.**

Weighted Restaurant Rating =Σ (number of votes \* rating) / Σ (number of votes) .

In bubble Graph if weighted rating is in bubble. Graph can’t give any inshight. But with Weighted Rating on y axis give highly insightful graph.