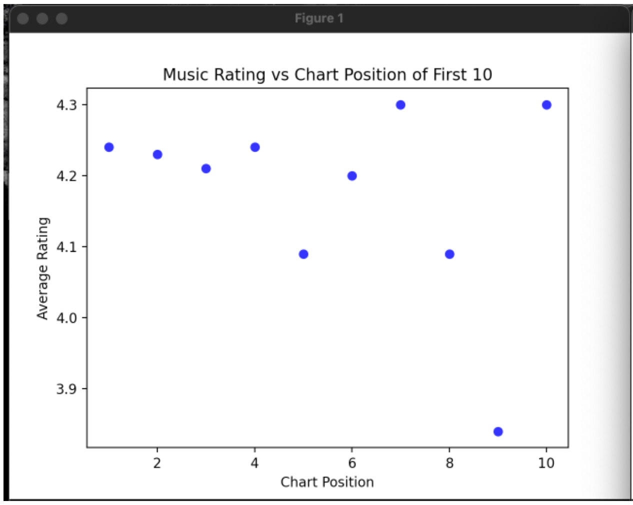
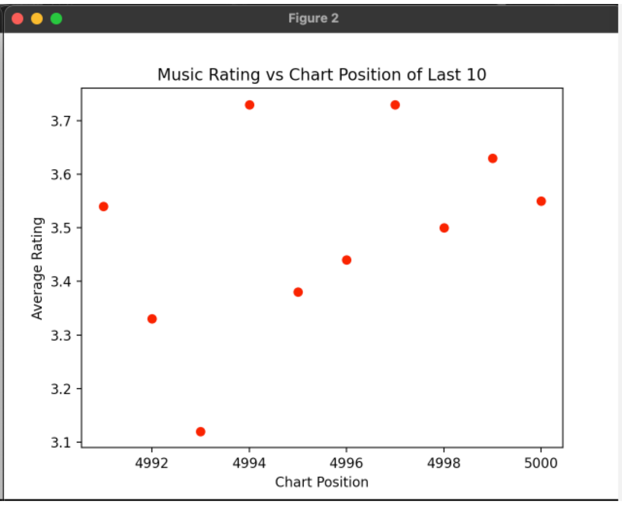
Rate Your Music: Top 5000 Albums Data Analysis

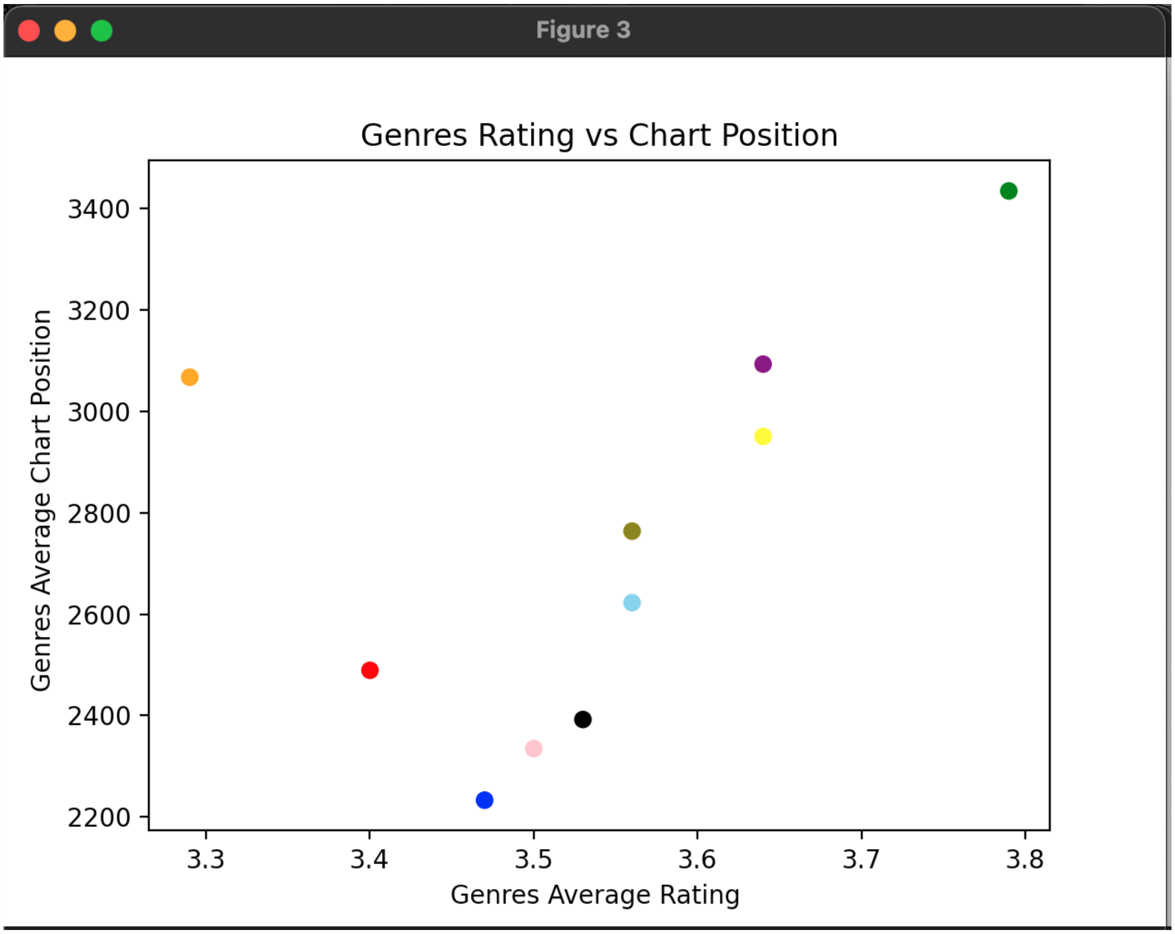
Introduction

This dataset was found on the website ‘Kaggle, but the data is from “Rate Your Music”. “Rate Your Music” is a website where they let users rate any released music whether it is a single, album and etc. My inspiration for choosing this dataset is because I’m a huge fan of a lot of music artists and wanted to learn more about music and their releases. What I wanted to find out was whether the rating of the album has any effect on where it is on the chart. Another thing I question is if the genre of music has any effect on where it would be on the chart.

Visuals







Analysis

The first question of this analysis is if the average rating of the album has anything to do with where it lands on the chart. From figure 1 and figure 2, we can see that there is a difference in the overall rating(y-axis). I took the first 10 data from the set and the last 10 to compare where they would place. If we compare figures 1 and 2 we can see that album rating does affect its chart position. All but 1 album in the top 10 had over a 4.0/5.0 rating. The highest rating in the last 10 of the top 5000 chart is a 3.7/5.0 rating. So from this, we can deduce that album rating does have an effect on chart position.

The second question I wanted to answer is if the genre has an effect on chart position. For this, I went through the list and took the data that was related to the top 10 music genres on Google to get the best votes since there would be a bigger sample size. This was graphed in figure 3 and from this, we can see that genre does have an effect on chart position. We can see starting at point blue(hip-hop), even though there’s an increase in rating, the chart position is averaging down. From this, I believe we can conclude that genre does have an effect on chart position. This chart also follows a similar trend to the data found on Google. The average rating for the genre might be lower but because there is more consumers for it, it’ll be higher on the chart.

Conclusion

By using the dataset and extracting the data, we can answer the questions in the introduction. The rating of the album does have an effect on where it would be on the top chart. If the album is rated highly by the consumers then it has a better chance of making it higher on the chart. The second thing we wanted to see is if the genre also has an effect on the chart position. From this dataset, we can conclude genre does play a part in chart position. Although the average rating for the albums in the genre got higher, the chart position averaged down. We can assume that the more popular the genre the better the chances of it going higher.

Further Study

What I would like to study more are the other components that might affect the rating such as sales and the popularity of the artist. I believe we don’t have the full picture since there are so many other things that could possibly impact the album's positioning. This dataset gave did give us a pretty good idea though of how to get higher on the chart. A good album that is received well by the consumer and in a popular music genre.

Bibliography

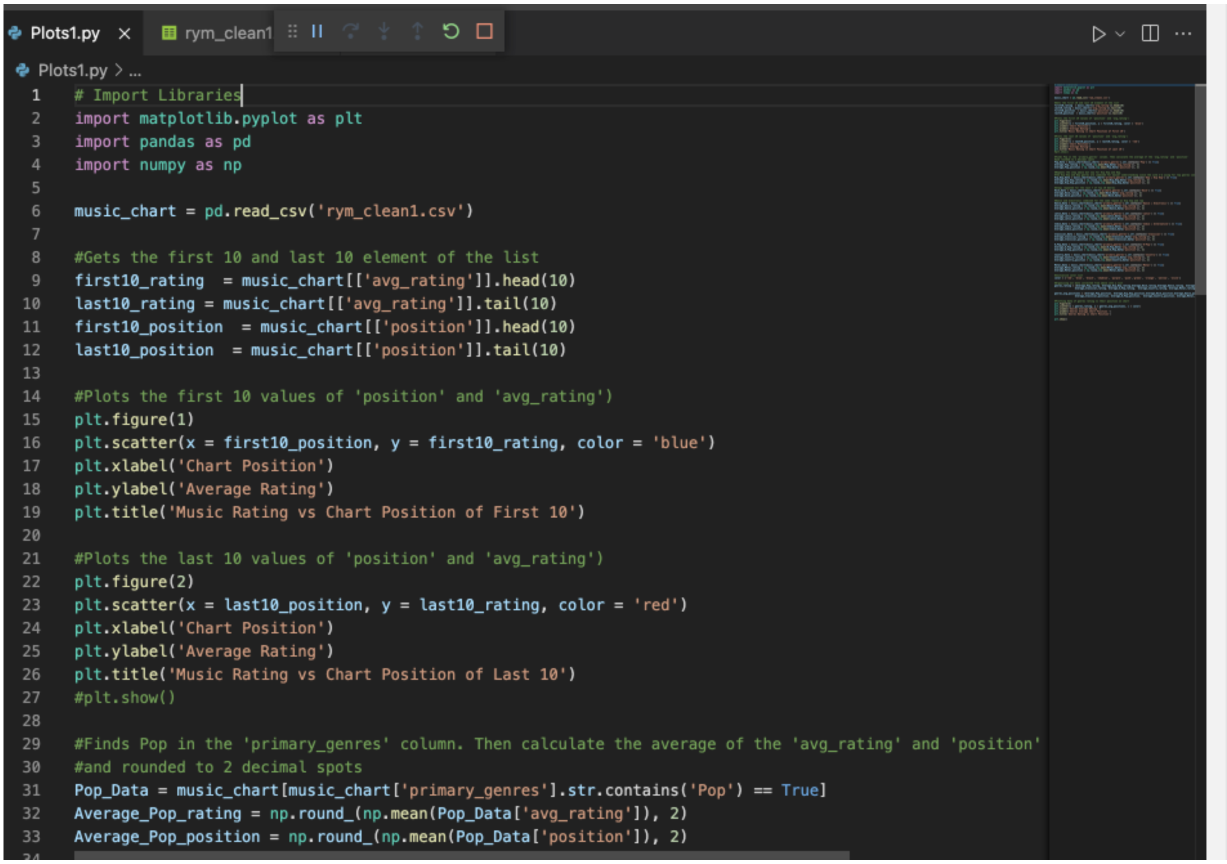
1. O., B. (2022, March 14). *Rate your music: The top 5,000 most popular albums*. Kaggle. Retrieved May 20, 2022, from https://www.kaggle.com/datasets/tobennao/rym-top-5000

2.*Where developers learn, share, & build careers*. Stack Overflow. (n.d.). Retrieved May 20, 2022, from https://stackoverflow.com/

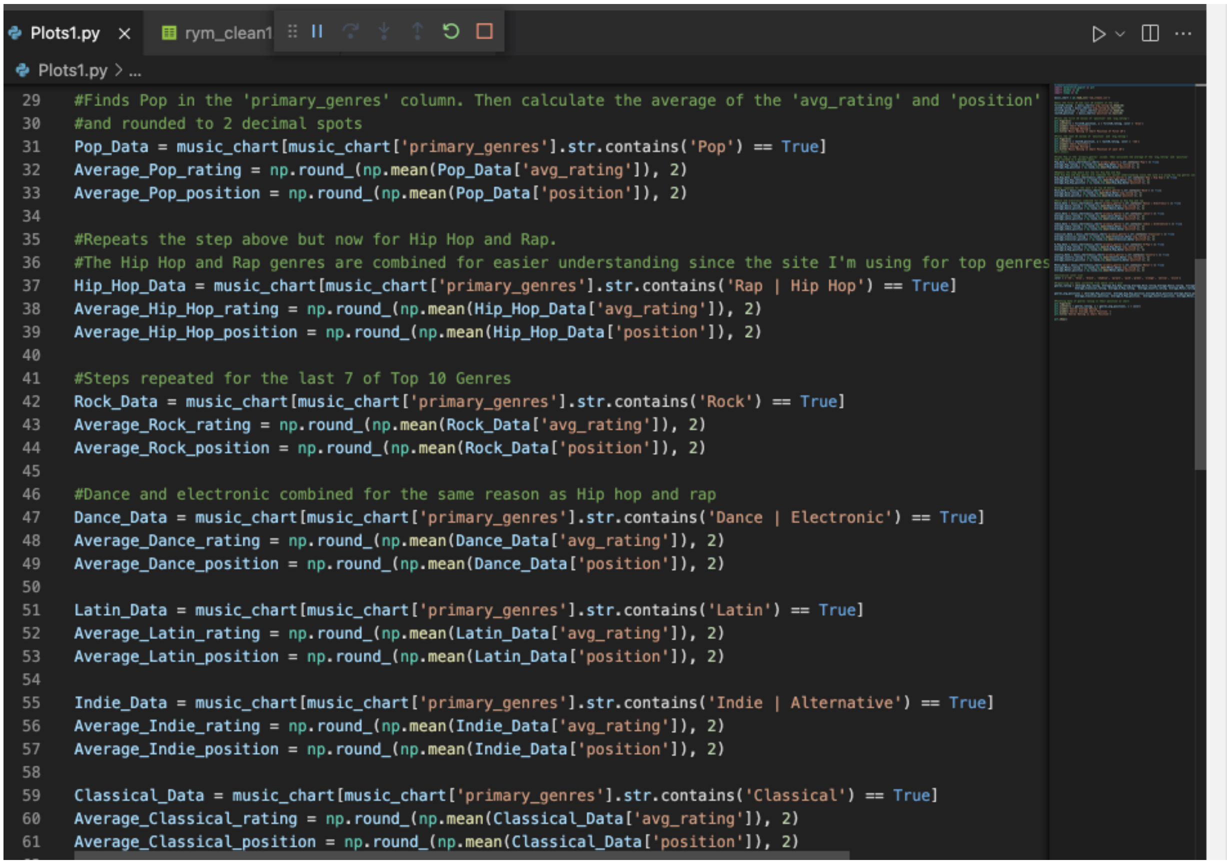
3.Dolon, B. (2021, February 7). *Check for a substring in a pandas DataFrame column*. Medium. Retrieved May 20, 2022, from https://towardsdatascience.com/check-for-a-substring-in-a-pandas-dataframe-column-4b949f64852#:~:text=Using%20%E2%80%9Ccontains%E2%80%9D%20to%20Find%20a,substring%20and%20False%20if%20not.

4.DataCamp. (n.d.). Retrieved May 20, 2022, from https://app.datacamp.com/learn

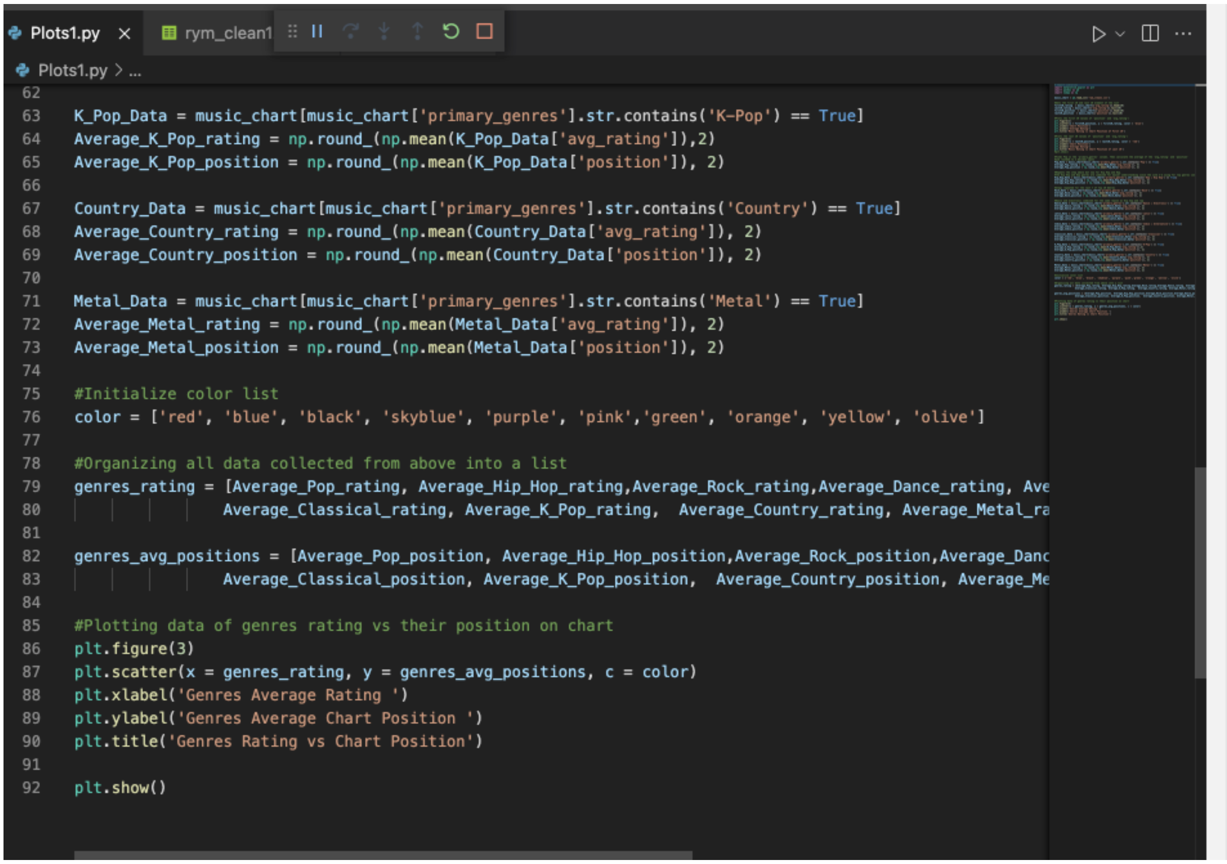
Appendix



```
# Import Libraries
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np

music_chart = pd.read_csv('rym_clean1.csv')

#Gets the first 10 and last 10 element of the list
first10_rating  = music_chart[['avg_rating']].head(10)
last10_rating = music_chart[['avg_rating']].tail(10)
first10_position  = music_chart[['position']].head(10)
last10_position  = music_chart[['position']].tail(10)

#Plots the first 10 values of 'position' and 'avg_rating')
plt.figure(1)
plt.scatter(x = first10_position, y = first10_rating, color = 'blue')
plt.xlabel('Chart Position')
plt.ylabel('Average Rating')
plt.title('Music Rating vs Chart Position of First 10')

#Plots the last 10 values of 'position' and 'avg_rating')
plt.figure(2)
plt.scatter(x = last10_position, y = last10_rating, color = 'red')
plt.xlabel('Chart Position')
plt.ylabel('Average Rating')
plt.title('Music Rating vs Chart Position of Last 10')
#plt.show()

#Finds Pop in the 'primary_genres' column. Then calculate the average of the 'avg_rating' and 'position'
#and rounded to 2 decimal spots
Pop_Data = music_chart[music_chart['primary_genres'].str.contains('Pop') == True]
Average_Pop_rating = np.round_(np.mean(Pop_Data['avg_rating']), 2)
Average_Pop_position = np.round_(np.mean(Pop_Data['position']), 2)
```



```python
#Finds Pop in the 'primary_genres' column. Then calculate the average of the 'avg_rating' and 'position'
#and rounded to 2 decimal spots
Pop_Data = music_chart[music_chart['primary_genres'].str.contains('Pop') == True]
Average_Pop_rating = np.round_(np.mean(Pop_Data['avg_rating']), 2)
Average_Pop_position = np.round_(np.mean(Pop_Data['position']), 2)

#Repeats the step above but now for Hip Hop and Rap.
#The Hip Hop and Rap genres are combined for easier understanding since the site I'm using for top genres
Hip_Hop_Data = music_chart[music_chart['primary_genres'].str.contains('Rap | Hip Hop') == True]
Average_Hip_Hop_rating = np.round_(np.mean(Hip_Hop_Data['avg_rating']), 2)
Average_Hip_Hop_position = np.round_(np.mean(Hip_Hop_Data['position']), 2)

#Steps repeated for the last 7 of Top 10 Genres
Rock_Data = music_chart[music_chart['primary_genres'].str.contains('Rock') == True]
Average_Rock_rating = np.round_(np.mean(Rock_Data['avg_rating']), 2)
Average_Rock_position = np.round_(np.mean(Rock_Data['position']), 2)

#Dance and electronic combined for the same reason as Hip hop and rap
Dance_Data = music_chart[music_chart['primary_genres'].str.contains('Dance | Electronic') == True]
Average_Dance_rating = np.round_(np.mean(Dance_Data['avg_rating']), 2)
Average_Dance_position = np.round_(np.mean(Dance_Data['position']), 2)

Latin_Data = music_chart[music_chart['primary_genres'].str.contains('Latin') == True]
Average_Latin_rating = np.round_(np.mean(Latin_Data['avg_rating']), 2)
Average_Latin_position = np.round_(np.mean(Latin_Data['position']), 2)

Indie_Data = music_chart[music_chart['primary_genres'].str.contains('Indie | Alternative') == True]
Average_Indie_rating = np.round_(np.mean(Indie_Data['avg_rating']), 2)
Average_Indie_position = np.round_(np.mean(Indie_Data['position']), 2)

Classical_Data = music_chart[music_chart['primary_genres'].str.contains('Classical') == True]
Average_Classical_rating = np.round_(np.mean(Classical_Data['avg_rating']), 2)
Average_Classical_position = np.round_(np.mean(Classical_Data['position']), 2)
```



```python

K_Pop_Data = music_chart[music_chart['primary_genres'].str.contains('K-Pop') == True]
Average_K_Pop_rating = np.round_(np.mean(K_Pop_Data['avg_rating']),2)
Average_K_Pop_position = np.round_(np.mean(K_Pop_Data['position']), 2)

Country_Data = music_chart[music_chart['primary_genres'].str.contains('Country') == True]
Average_Country_rating = np.round_(np.mean(Country_Data['avg_rating']), 2)
Average_Country_position = np.round_(np.mean(Country_Data['position']), 2)

Metal_Data = music_chart[music_chart['primary_genres'].str.contains('Metal') == True]
Average_Metal_rating = np.round_(np.mean(Metal_Data['avg_rating']), 2)
Average_Metal_position = np.round_(np.mean(Metal_Data['position']), 2)

#Initialize color list
color = ['red', 'blue', 'black', 'skyblue', 'purple', 'pink','green', 'orange', 'yellow', 'olive']

#Organizing all data collected from above into a list
genres_rating = [Average_Pop_rating, Average_Hip_Hop_rating,Average_Rock_rating,Average_Dance_rating, Ave
                Average_Classical_rating, Average_K_Pop_rating,  Average_Country_rating, Average_Metal_ra

genres_avg_positions = [Average_Pop_position, Average_Hip_Hop_position,Average_Rock_position,Average_Danc
                Average_Classical_position, Average_K_Pop_position,  Average_Country_position, Average_Me

#Plotting data of genres rating vs their position on chart
plt.figure(3)
plt.scatter(x = genres_rating, y = genres_avg_positions, c = color)
plt.xlabel('Genres Average Rating ')
plt.ylabel('Genres Average Chart Position ')
plt.title('Genres Rating vs Chart Position')

plt.show()
```