An Exploratory Analysis of Spotify Quarterly Revenue and Costs in R

MariCarmen Mosso

Northeastern University

MKTG 2602: Quantitative Analysis of Consumer Data

Dr. Mirna Rossi

12 December 2023

Ever since the first recorded song, there has been the question of how consumers should be able to access these songs. Before the internet, vinyls and the radio were consumers main access to music, but with the creation of the internet this all changed. At first, there was no good solution for how music should be distributed to consumers on an internet scale. Piracy ran rampant because of this and peer-to-peer platforms like Napster and Limewire slashed U.S. music revenues by over half, plummeting from $14.6 billion in 1999 to $6.7 billion in 2014 and 2015. The shift from physical sales to easily shared sound files fostered rampant piracy. iTunes briefly stabilized the freefall, yet a generation grew up thinking music should be free. Then came Spotify, born in 2006 from the minds of Daniel Ek and Martin Lorentzon in a bid to tackle piracy. Offering vast music libraries for a monthly fee or ad-supported listening, it revolutionized streaming, arguably leading to Apple Music's debut in 2015 after Spotify's U.S. launch in 2011 (Robinson, 2021). The Swedish born company of Spotify is now the leading music streaming service, with over 515 million active users across the globe in 2023, a cumulative revenue of over 11.7 billion euros, and over 9,000 employees worldwide (Shepard, 2023). What is not only intriguing about Spotify, but also sets it aside from its competitors is that it offers two versions of the service: one free and one premium. The free version of Spotify is very limited, with ad breaks and the restriction of not being able to play a certain song whenever you want, but you do have access to all the songs on the platform to add to your own playlists. On the other hand, with the premium version, users have access to all songs on the platform however they want with no ad breaks. It is as if premium users, in a sense, own all of the songs on Spotify. Of their 515 million active users, a surprising 317 million are ad- supported, meaning they have the free version, while 210 million pay for the premium subscription (Shepard, 2023). Apple music, however, offers no free version and is subscription only. Furthermore, in terms of demographics, Spotify's user base is 56% male, 29% are millennials, and the US is the country contributing the most to Spotfy's overall revenue (Saxena, 2022). This insight is important to note because now looking at Apple Music, 59% of their users are from the continent North America (Wise, 2023). This means that both of Spotify and Apple Music's biggest bases are in North America, where the competition between them is becoming increasingly distinctive in the music streaming industry.

However, even though Spotify is the biggest service in the music streaming industry, it is not as profitable as it, in a way, should be. Recently, Spotify's CEO Daniel Ek shared on Monday December 4th 2023, that the company's global workforce would be cut by 17%, about 1,500 jobs. This would be Spotify's third round of layoffs this year, including a 6% reduction this past January, which ended up affecting close to 600 staffers, and then a trimming of its podcast division this past June that reduced its workforce by about 2%, or around 200 jobs, meaning Spotify has let go of over 2,000 workers this year alone (Schneider, 2023). While Covid did not change how Spotify distributed its product, they did have an increase in workforce size. CEO Daniel Ek said: "To understand this decision, I think it is important to assess Spotify with a clear, objective lens. In 2020 and 2021, we took advantage of the opportunity presented by lower-cost capital and invested significantly in team expansion, content enhancement, marketing,

and new verticals. These investments generally worked, contributing to Spotify's increased output and the platform's robust growth this past year. However, we now find ourselves in a very different environment. And despite our efforts to reduce costs this past year, our cost structure for where we need to be is still too big" (Schneider, 2023). Earlier this year, Spotify raised the prices of its subscription plan, increasing the monthly bill for users about $1 to $2, depending on the plan. Spotify, therefore, said that "the early effects of price increases" were partially responsible for the 11% year-over-year revenue growth (Richardson, 2023). Spotify is still not as profitable as it could be, and even though they have seen a steady incline of income and profit since their start in 2011, they are still searching for ways to effectively cut costs while still being able to deliver the best quality service to their consumers.

Since Apple Music entered the streaming scene, it has become prevalent to analyze the two services in order to understand not only their positions in the market, but also the music streaming market in general. Analyzing Spotify's revenue holds paramount importance in understanding the music streaming industry's landscape and Spotify's market positioning. With the rise of Apple Music reshaping the music streaming market with over $8.3 billion in revenue in 2022, it has become increasingly beneficial to analyze Spotify's costs and revenues in order to strategize next steps in their market competition with Apple Music (Curry, 2023). In order to analyze Spotify's positioning in the music streaming market, I found a data set entitled Spotify_Quarterly from kaggle.com that provides insights to Total Revenue, Cost of Revenue, Gross Profit, Premium Revenue, Premium Cost Revenue, Premium Gross Profit, Ad Revenue, Ad Cost of revenue, Ad gross Profit, MAUs, Premium MAUs, Ad MAUs, Premium ARPU, and Sales and Marketing Cost (Maurya, 2023). As a customer of Spotify, it was very intriguing for me to be able to analyze this dataset because I have always been interested in the music streaming world. This year, I listened to over 90,000 minutes of music on Spotify with my premium subscription, and was eager to see how my subscription and listening habits profits Spotify, as Spotify is in technicality the middle man between artist and consumer, delivering the music from the artists to the listeners. Spotify has always been ahead of the game with personalized playlists, artist messages, yearly listening wraps, and notifications that draw, me, the listener back in consistently to primarily listen to music on Spotify rather than on physical CDs, vinyls, or other streaming platforms. However, with Apple Music becoming more and more popular, I was even more interested to see how this has affected Spotify sales, as even though I am an avid Spotify fan, I see the perks of Apple Music, and about half of my friends are now Apple Music listeners, rather than Spotify listeners.

Data set from:
https://www.kaggle.com/datasets/mauryansshivam/spotify-revenue-expenses-and-its-premium-users

I learned many things about the patterns of Spotify's revenue over time through my analysis of this dataset. For starters, I was surprised by how gross profit has only increased from 105 million euros to 776 million euros from 2017 to 2023. This was a large drop off from total revenue to total profit. I had always heard murmurs in the news about how Spotify is not as

profitable as people would anticipate it being, but it was still shocking to see the evidence right in front of me, and I did not expect this when I was working on Interim project 1. Going off of that, I was also surprised with how small the gap between revenue and cost over time was when shown on a graph. As revenue went up, cost went up right along with it, and there was only a small gap between the number of revenue and the number of costs, meaning that even though Spotify is making more money, costs are going up at the same time, halting overall profit even though Spotify is gaining more and more users every year. This now becomes a problem with Apple Music's increasing popularity because Spotify is already not as profitable as it could be as the leading streaming service, so with Apple Music's increasing popularity these numbers will start to see a drop off. Overall, I believe I found a fair amount of observations, and even though I was shocked by some of the insights I found, I believed that this was good because it was great practice to be able to put these analyses from R into common words. Even though in Interim project 2 I decided to have a more of a keen focus on gross profit, total revenue, and a breakdown of costs, all variables were useful and insightful. The only part of the dataset that was missing was the information from 2016, which was all blank. I believe that this was just a mistake on the dataset's part and just started my analysis from 2017. The only problem I ran into with this particular dataset was that the values were not originally numeric, even though they were all numbers, but with help from my professor, I was able to change this in R in order to make graphs and perform the proper analysis needed. In Interim project 1 I was interested to not only analyze this dataset, but also look for some solutions to how to market Spotify more, so even though this dataset was not missing any information, I wanted to find something else to analyze as well, so I found the Top 50 global songs playlist on Spotify and analyzed this as well. Even though I did not need more data, I felt as though my analysis of just revenue with no idea of some sort of solution would be incomplete, and analyzing the playlist gave me more of an outlook into how Spotify can use their most popular songs and artists to make a more personalized experience for every users, setting them aside in that manner from Apple Music.

Analysis of Code

In order to read my data set, I used this code to read it in R:
Spotify_Quarterly <- read_csv("Spotify Quarterly.csv")
#Data was taken from
https://www.kaggle.com/datasets/mauryansshivam/spotify-revenue-expenses-and-its-premium-users/

DIM
Run Dim Function: > dim(Spotify_Quarterly)
[1] 26 17

##The dim() function serves to extract or define the dimensions of an array or matrix. When employed on an object, it yields a vector that signifies the rows and columns (and potentially more dimensions for arrays) within that particular object.

Variables and Observations

In this data set, there are 17 variables and 26 observations. In order to find this information, I ran the code ncol() and nrow():

```
> ncol(Spotify_Quarterly)
[1] 17
> nrow(Spotify_Quarterly) #Number of observations
[1] 26
```

Colnames

```
> colnames(Spotify_Quarterly) #For Variables
 [1] "Date"                        "Total Revenue"
 [3] "Cost of Revenue"             "Gross Profit"
 [5] "Premium Revenue"             "Premium Cost Revenue"
 [7] "Premium Gross Profit"        "Ad Revenue"
 [9] "Ad Cost of revenue"          "Ad gross Profit"
[11] "MAUs"                        "Premium MAUs"
[13] "Ad MAUs"                     "Premium ARPU"
[15] "Sales and Marketing Cost"    "Research and Development Cost"
[17] "General and Administrative Cost"
```

Str, Glimpse, and Summary

In R, str(), summary(), and glimpse() are codes that give a condensed view of data frames. Str() reveals column names, data types, and a snippet of values from each column. Summary() generates statistical summaries for numeric variables within a dataset, outputting minimum, first quartile, median, mean, third quartile, and maximum values for each numeric column. Glimpse() provides information about the data types and the first few observations of each column.

Visualized here is the Spotify_Quarterly Output:

```
> str(Spotify_Quarterly)
spc_tbl_ [26 × 17] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ Date                       : chr [1:26] "31-03-2023" "31-12-2022" "30-09-2022" "30-06-2022" ...
 $ Total Revenue              : num [1:26] 3042 3166 3036 2864 2661 ...
 $ Cost of Revenue            : num [1:26] 2276 2365 2286 2160 1990 ...
 $ Gross Profit               : num [1:26] 776 801 750 704 671 712 668 663 548 575 ...
 $ Premium Revenue            : num [1:26] 2713 2717 2651 2504 2379 ...
 $ Premium Cost Revenue       : num [1:26] 1937 1939 1908 1804 1704 ...
 $ Premium Gross Profit       : num [1:26] 776 778 743 700 675 670 633 633 538 545 ...
```

```
$ Ad Revenue                : num [1:26] 329 449 385 360 282 394 323 275 216 281 ...
$ Ad Cost of revenue        : num [1:26] 339 426 378 356 286 352 288 245 206 251 ...
$ Ad gross Profit           : num [1:26] -10 23 7 4 -4 42 35 30 10 30 ...
$ MAUs                      : num [1:26] 515 489 456 433 422 406 381 365 356 345 ...
$ Premium MAUs              : num [1:26] 210 205 195 188 182 180 172 165 158 155 ...
$ Ad MAUs                   : num [1:26] 317 295 273 265 252 236 220 210 208 199 ...
$ Premium ARPU              : num [1:26] 4.32 4.55 4.63 4.54 4.38 4.4 4.34 4.29 4.12 4.26 ...
$ Sales and Marketing Cost  : num [1:26] 347 453 432 391 296 340 280 279 236 294 ...
$ Research and Development Cost: num [1:26] 435 415 386 336 250 253 208 255 196 232 ...
$ Genreal and Adminstraive Cost: num [1:26] 140 164 160 171 131 126 105 117 102 118 ...
- attr(*, "spec")=
 .. cols(
 ..   Date = col_character(),
 ..   `Total Revenue` = col_double(),
 ..   `Cost of Revenue` = col_double(),
 ..   `Gross Profit` = col_double(),
 ..   `Premium Revenue` = col_double(),
 ..   `Premium Cost Revenue` = col_double(),
 ..   `Premium Gross Profit` = col_double(),
 ..   `Ad Revenue` = col_double(),
 ..   `Ad Cost of revenue` = col_double(),
 ..   `Ad gross Profit` = col_double(),
 ..   MAUs = col_double(),
 ..   `Premium MAUs` = col_double(),
 ..   `Ad MAUs` = col_double(),
 ..   `Premium ARPU` = col_double(),
 ..   `Sales and Marketing Cost` = col_double(),
 ..   `Research and Development Cost` = col_double(),
 ..   `Genreal and Adminstraive Cost` = col_double()
 .. )
 - attr(*, "problems")=<externalptr>
> glimpse(Spotify_Quarterly)
Rows: 26
Columns: 17
$ Date                      <chr> "31-03-2023", "31-12-2022", "30-09-2022", "30…
$ `Total Revenue`           <dbl> 3042, 3166, 3036, 2864, 2661, 2689, 2501, 233…
$ `Cost of Revenue`         <dbl> 2276, 2365, 2286, 2160, 1990, 1977, 1833, 166…
$ `Gross Profit`            <dbl> 776, 801, 750, 704, 671, 712, 668, 663, 548, …
$ `Premium Revenue`         <dbl> 2713, 2717, 2651, 2504, 2379, 2295, 2178, 205…
$ `Premium Cost Revenue`    <dbl> 1937, 1939, 1908, 1804, 1704, 1625, 1545, 142…
```

```
$ `Premium Gross Profit`        <dbl> 776, 778, 743, 700, 675, 670, 633, 633, 538, …
$ `Ad Revenue`                  <dbl> 329, 449, 385, 360, 282, 394, 323, 275, 216, …
$ `Ad Cost of revenue`          <dbl> 339, 426, 378, 356, 286, 352, 288, 245, 206, …
$ `Ad gross Profit`             <dbl> -10, 23, 7, 4, -4, 42, 35, 30, 10, 30, 1, -16…
$ MAUs                          <dbl> 515, 489, 456, 433, 422, 406, 381, 365, 356, …
$ `Premium MAUs`                <dbl> 210, 205, 195, 188, 182, 180, 172, 165, 158, …
$ `Ad MAUs`                     <dbl> 317, 295, 273, 265, 252, 236, 220, 210, 208, …
$ `Premium ARPU`                <dbl> 4.32, 4.55, 4.63, 4.54, 4.38, 4.40, 4.34, 4.2…
$ `Sales and Marketing Cost`    <dbl> 347, 453, 432, 391, 296, 340, 280, 279, 236, …
$ `Research and Development Cost` <dbl> 435, 415, 386, 336, 250, 253, 208, 255, 196, …
$ `Genreal and Adminstraive Cost` <dbl> 140, 164, 160, 171, 131, 126, 105, 117, 102, …
> glimpse(Spotify_Quarterly)
Rows: 26
Columns: 17
$ Date                          <chr> "31-03-2023", "31-12-2022", "30-09-2022", "30…
$ `Total Revenue`               <dbl> 3042, 3166, 3036, 2864, 2661, 2689, 2501, 233…
$ `Cost of Revenue`             <dbl> 2276, 2365, 2286, 2160, 1990, 1977, 1833, 166…
$ `Gross Profit`                <dbl> 776, 801, 750, 704, 671, 712, 668, 663, 548, …
$ `Premium Revenue`             <dbl> 2713, 2717, 2651, 2504, 2379, 2295, 2178, 205…
$ `Premium Cost Revenue`        <dbl> 1937, 1939, 1908, 1804, 1704, 1625, 1545, 142…
$ `Premium Gross Profit`        <dbl> 776, 778, 743, 700, 675, 670, 633, 633, 538, …
$ `Ad Revenue`                  <dbl> 329, 449, 385, 360, 282, 394, 323, 275, 216, …
$ `Ad Cost of revenue`          <dbl> 339, 426, 378, 356, 286, 352, 288, 245, 206, …
$ `Ad gross Profit`             <dbl> -10, 23, 7, 4, -4, 42, 35, 30, 10, 30, 1, -16…
$ MAUs                          <dbl> 515, 489, 456, 433, 422, 406, 381, 365, 356, …
$ `Premium MAUs`                <dbl> 210, 205, 195, 188, 182, 180, 172, 165, 158, …
$ `Ad MAUs`                     <dbl> 317, 295, 273, 265, 252, 236, 220, 210, 208, …
$ `Premium ARPU`                <dbl> 4.32, 4.55, 4.63, 4.54, 4.38, 4.40, 4.34, 4.2…
$ `Sales and Marketing Cost`    <dbl> 347, 453, 432, 391, 296, 340, 280, 279, 236, …
$ `Research and Development Cost` <dbl> 435, 415, 386, 336, 250, 253, 208, 255, 196, …
$ `Genreal and Adminstraive Cost` <dbl> 140, 164, 160, 171, 131, 126, 105, 117, 102, …
```

The Skim Function

Skim() delivers an extensive overview encompassing the structure, statistical summaries, and missing data presence within a dataset. Within my dataset, only 1 variable is missing within every subset of variables. This is because the year 2016 is included in the dataset, but there are no values attributed to this year, so for every variable, 1 missing variable is noted because there are no values for 2016. This was actually not a problem in my analysis because it was the first year noted, so I could just start my analysis from 2017 and look at a 6 year period, rather than 2016. If there were missing values for a year in the middle such as 2020, this would have served

to be more of a problem because there would have been a hole in the middle of my analysis that would have been hard to work around.

Table()

The table function provides information to the frequency of a certain variable. For example, in my analysis I used the table() function with the variable "cost of revenue." This provided me with the information for how frequent the costs of revenue were. I found that every year the cost of revenue was different, so all the frequencies were 1, and that the cost of revenue rose every year and never repeated.

Code Visualized:

```
> table(Spotify_Quarterly$`Cost of Revenue`) #Table of the frequency of Cost Revenue

 775  797  802  856  867  944 1010 1096 1138 1233 1290 1376 1381 1410 1486 1593 1599
   1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1
1668 1833 1977 1990 2160 2276 2286 2365
   1    1    1    1    1    1    1    1
```

The n_distinct() function serves to tally the count of unique or distinct values contained within a specific variable or column of a dataset. In my dataset, through this function I found that there were 26 distinct values within my dataset.

Code Visualized:

```
> n_distinct(Spotify_Quarterly)
[1] 26
```

Standard Deviation

Standard deviation quantifies the extent to which data points diverge from the mean value, computed as the square root of the variance. It gauges the dispersion within a dataset by assessing how each data point varies from the mean. As data points stray further from the mean, the deviation within the dataset increases, reflecting a greater spread in the data and consequently yielding a higher standard deviation (Hargrave, 2023). In sum, standard deviation is how spread out data is. Standard deviation is also similar to the variance and is calculated by taking the square root of the variance. Variance quantifies how much the numbers in a dataset differ from the mean, squared, while standard deviation is a more interpretable measure that shows how much the data typically deviates from the mean (Investopedia Team, 2023.)

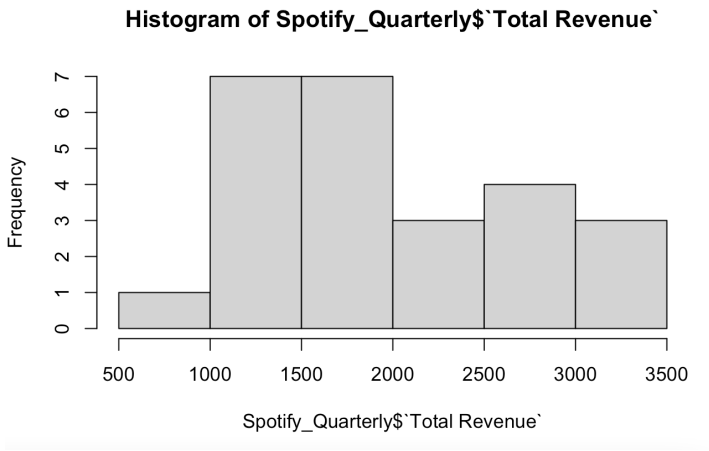
Mean, Median, Standard Deviation

Mean, median, and standard deviation functions provided some interesting insights into my dataset. First, I found the mean for the variables Total Revenue, Gross Profit, and Mean Profit. I

found that the mean for Total Revenue was 1,949.2, the mean for Gross Profit was 501.28 and that the mean for Ad Revenue was 215.76 (all in millions and euros). Then, using the same variables, I wanted to find if the median and standard deviation were different, and if so how different, and what did they say about these variables as a whole. Using the median() function, I found that the median total revenue was 1,855, the median gross profit was 479, and the median ad revenue was 175 (all in millions and euros). Now to analyze, the mean is the result of adding all values and dividing by the count, the average, while the median stands as the middle value when the data is arranged. Overall, extreme values have a bigger impact on the mean, while the median remains more stable and less influenced by outliers (Bastos, 2021). I was not surprised that the mean and median values were different, but around the same value for each variable. It made sense that the mean would be bigger numbers while the median saw more of a drop off due to how these values within each variable are different each year. Standard deviation, on the other hand, was a bit more interesting to see. The standard deviation for Total Revenue was 688.5715, for Gross Profit was 188.4935, and for Ad Revenue was 107.5788 (all in millions and euros). These numbers are interesting because Total Revenue had the highest standard deviation, meaning that it has the highest amount of variance, meaning that the Total Revenue values are more dispersed throughout the dataset. This makes sense because as Spotify has grown over the last 6 years Total Revenue has increased yearly, with the lowest value being 902 and the highest being 3,042, which is a large dispersion of data.

Histograms

For my analysis in R, I created a histogram with the variable Total Revenue to see how the values of this variable are distributed throughout the dataset. The histogram that was created showed more of a negative skewness, meaning that more of the values fell on the left side. Usually, a bell curve shows a curve right above the mean, but skewed histograms mean that the values are not all evenly distributed based on the mean, with more values falling specifically to the left or the right side. Based on finding the mean and median, I did expect the histogram to be negatively skewed because the mean and median were more on the lower side compared to the highest value. In other words, the mean and median did not represent the exact middle values between the lowest Total Revenue and the highest Total Revenue, so there were more values on the lower side than the higher side. In fact, there were only 3 times that the Total Revenue was over 3,000. Therefore, these observations make sense and were not a big shock to see after analyzing the mean and median.



Mutate, Filter, and Group_By

When I first selected this dataset for analyzing, I realized that all the values were not numeric in the dataset, meaning I was not able to produce graphs, or find the mean, median, mode, or standard deviation for any of the variables even though they were all values. In order to fix this problem, my professor helped me with a code using the mutate() function to manipulate the dataset so that all variables became numerical, except for the dates. This code was necessary to find insights in the dataset because I was able to find insights on changes to these numbers over time. I also used mutate and group_by together in my analysis to summarize Mean Revenue, Max Profit, and Median Cost by Year. I used these functions because the original date format is day, month, year, so I had to mutate this variable in order to only show years, and then group all of the data together within that year because there are 4 values within every year due to the fact that the data set is a summary of each quarter at Spotify. Using this code helped visualize how these values are changing every year and how max_profit is seeing the slowest increase.

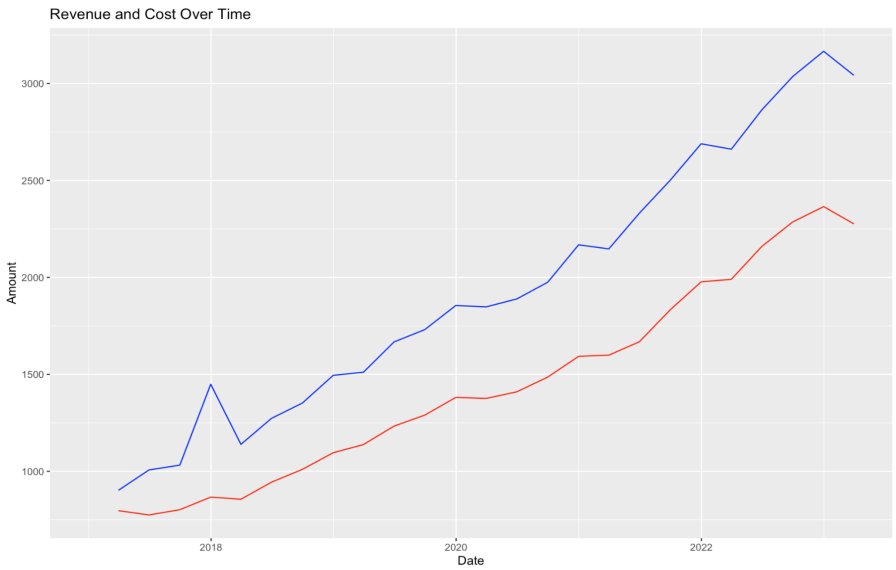
Year	mean_revenue	max_profit	median_cost
<dbl>	<dbl>	<dbl>	<dbl>
2016	NA	NA	NA
2017	1098.	582	800.
2018	1315.	399	977
2019	1691	474	1262.
2020	1970	575	1448
2021	2417	712	1750.
2022	2932.	801	2223
2023	3042	776	2276

In order to view how much Total Revenue has changed from 2017 to 2023, I used the filter() function to filter the year 2017 and the year 2023 in terms of Total Revenue. I found that in 2017 the Total Revenue was 902, while in 2023 it was 3,042. I also found that Gross Profit was 105 in 2017, while in 2023 it was 776 (all in millions of euros). Not only did filtering the year help me analyze the change of Total Revenue over time in the last 6 years, but also, when filtered by date, helped me see the stark drop off from Total Revenue to Gross Profit that is taking place. In the last 6 years, Total Revenue grew exponentially, while Gross Profit's growth was rather small in comparison with Total Revenue.

While filtering by year was insightful to see overall change from 2017 to 2023, using the groupby() function to see all the data summarized grouped together was insightful to see growth from quarter to quarter. Using this function in terms of the data set, I was able to see how Total Revenue, Cost of Revenue, Gross Profit, and Premium Revenue change every quarter. This is where I noticed that values for all variables are at their highest at the end of the year, and see a slight drop off at the beginning of the following year. For example Gross Profit was 801 at the end of 2022, but dropped down a bit to 776 the following quarter in March of 2023. This pattern is consistent with other variables as well, as Cost of Revenue was at 2,365 at the end of 2022, but

fell to 2,276 in March of 2023. These insights stay consistent to how all the variables seem to be connected with each other, as Revenue goes up so does Profit and Cost.

Relationship between Total Revenue and Cost of Revenue



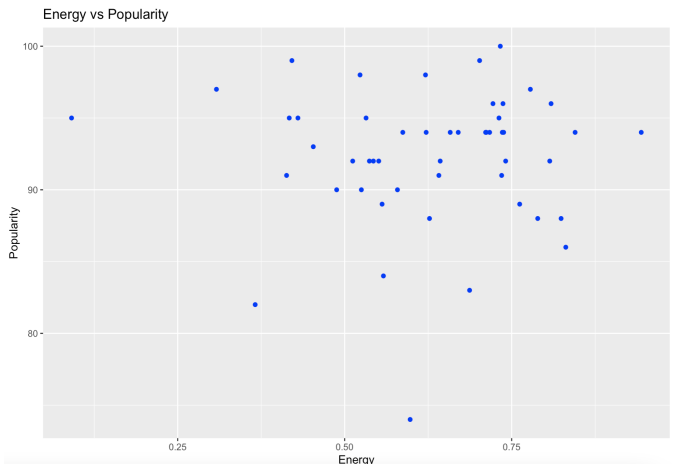
When I first started my analysis, I was very intrigued to see why Spotify's profit saw a significant drop off from Total Revenue. In order to visualize this, I decided it would be best to compare the two variables of Total Revenue and Cost of Revenue. I decided this over Total Revenue and Gross Profit because Cost of Revenue is Spotify's biggest cost and I wanted to see if there were any patterns that I could find between these two variables. When making a graph to visualize these two variables over time, I was right in my hypothesis that there was a pattern taking place. As Total Revenue (blue) goes up, Cost of Revenue goes up right along with it. This means that as Spotify is getting more popular and making more money, their cost of distributing their services to consumers is increasing as well. It is important to note that interestingly Spotify has not figured out a way to bring this cost down, as seen in the graph it remains increasing at a steady pace. However, even though this is seen in the above graph, it was important to still check whether it was statistically significant or not, so I ran a correlation test to check this. Through the cor() function, I found that the correlation coefficient between Cost of Revenue and Total Revenue in my dataset was 0.99 (rounded up). As the closer a correlation coefficient is to 1 the stronger it is, this proves what is seen in the above graph, that there is a very strong positive relationship between the two variables, meaning when one increases, the other increases as well.

Correlation Code:

```
> correlation <- cor(Spotify_Quarterly$`Cost of Revenue`, Spotify_Quarterly$`Total Revenue`, use = "complete.obs")
> correlation
[1] 0.9943471
```

Other observations

Even though Spotify has not found a solution to close the gap between Total Revenue and Total Cost, I think more profit could come from ads, especially with the rise of Apple Music on the music streaming scene. So what can Spotify do to combat the rise of Apple Music and continue to see a steady rise in subscribers and profit? The last part of my analysis was an attempt to explore a possible answer to this question. I decided to extract data from Spotify's Global Top 50 playlist (data exported through exportify.com), and was provided with the following variables: Spotify ID, Artist IDs, Track Name, Album Name, Artist Name(s), Release Date, Duration (ms), Popularity, Added By, Added At, Genres, Danceability, Energy, Key, Loudness, Mode, Speechiness, Acousticness, Instrumentalness, Liveness, Valence, Tempo, and Time Signature. I made this decision because music streaming data offers a unique insight into human behavior due to individualized choices that capture preferences and attitudes, emotional influence and brain activation, trends in culture, and a broader representation of personality, presenting a distinctive lens to comprehend human individuality, which all can be used in advertising and marketing plans (Anderson et al., 2021). I filtered top genres and artists, along with two graphs that visualized popularity among top artists, and popularity among how energetic the top 50 songs are. I found that Taylor Swift, pop music, and slightly above average energetic songs, are particularly dominating the top 50 songs in the world. Spotify should interpret this data for more pop-centric advertisements and, possibly, more Taylor Swift Spotify exclusive collaborations to account for the drop off in revenue after the Apple Music Super Bowl Halftime show, which will only continue with every Halftime show Apple Music sponsors. This secondary analysis provided me with conclusions in order to take the next step to try to come up with some ideas for how Spotify can solve the problems I found in my primary analysis of their quarterly revenue and costs.



	`Artist Name(s)`	n
	<chr>	<int>
1	Taylor Swift	6
2	Doja Cat	2
3	Dua Lipa	2
4	SZA	2
5	Arctic Monkeys	1
6	Bad Bunny	1
7	Bad Bunny,Feid	1
8	Billie Eilish	1
9	Calle 24,Chino Pacas,Fuerza Regida	1
10	Darell	1

Results.

Overall, this analysis of Spotify's quarterly revenue proved to be very insightful. I was genuinely surprised with how expensive Spotify's cost of revenue is. In one sense, I understand why it increases as revenue increases because with more subscribers comes more consumers to

service. However, I am shocked that with Spotify's popularity and advancement of technology they have not only not found a way to decrease the cost of revenue even as they are gaining more subscribers, but have also not been able to decrease the gap between cost of revenue and total revenue in other ways such as more ads because even though Spotify's premium gross profit is its largest subset of profit, as mentioned in my introduction, their largest amount of subscribers are actually ad supported subscribers, but their gross ad profit is shockingly low, ranging from -16 to 42. Spotify should find a balance between not having too many ads in their ad supported subscription, but enough ads for them to be far more profitable than they are. I believe that since Spotify feels like such an individual platform, unique to every user in the sense of customized playlists, if Spotify had more customized ads for users using their ad supported subscription, these ads could be more profitable and less annoying to their listeners. However, the dataset did not account for how much each ad cost, or clarify if the ads were in app ads, or external ads to promote Spotify. This would be the only limitation within the dataset, along with the missing variables from 2016, which even though they did affect or skew the overall analysis, they still would have been interesting to see, or even to have a complete dataset of Spotify revenue since 2011, to really understand where improvements have been made since their beginnings. On the topic of skewness, most of the variables are negatively skewed because of the vast difference of numbers over a 6 year period that is not evenly distributed, but constantly changing every quarter. For this reason, I would say that when doing an in depth analysis of any of these variables it would be important to find the median for all of them. Furthermore, I was surprised with the trend of having the highest total revenue and gross profit at the end of every year and then seeing a drop off at the beginning of the following year. This was surprising to me because I had thought that the beginning of the year would show to have more subscribers and users rather than at the end of every year. However, after reflecting upon this trend, conveniently also at the end of the year, I realized that this is when Spotify Wrapped comes out, which is Spotify's personalized data sent to every individual user letting them know their listening habits for the year. This includes their top artists, top songs, top genres, minutes listened, and even personalized messages from their top artists. Every year the Spotify Wrapped is highly anticipated and generates lots of free promotion and advertising for Spotify as users partake in posting and discussing their Spotify Wrapped on social media channels (Guzman, 2023). Therefore, after all of this discourse about Spotify at the end of the year, it makes sense that their highest quarterly revenue would be at the end of every year. Also, I was intrigued by how significant the relationships were between Total Revenue and Cost of Revenue, and Total Revenue and Gross Profit. Even though I expected these relationships to be statistically significant, I was more so surprised by how strong they were with both correlation coefficients being over 0.95. I was surprised because I had anticipated that maybe there were other outliers for this such as other costs or even outside events like trends in different economies or music tastes.

Reflection

From this assignment I was really able to take code that I have learned over the course of the semester and apply it to a real data set, being able to try to hypothesize and foresee different trends and conclusions from a marketing perspective. I learned how so many of the variables are connected within not only Spotify, but presumably other companies as well. As one value for a variable changes, this can affect all of the other variables and create a domino effect, having an impact on all of the other values within a company. In terms of the dataset specifically, I learned how high Spotify's Cost of Revenue is, and how this greatly affects their overall Gross Profit, resulting in smaller gross profit numbers than I had anticipated, especially when compared to the Total Revenue values. Before I had started, I wish I would have known how intertwined all the variables are, and perhaps I could have looked more in depth at other variables that I did not do an in depth analysis for because I wanted to find stronger connections and conclusions from the larger variables. Perhaps if I had known previously how much money Spotify spends in Cost of Revenue, I would have spent more time analyzing more possible solutions for a larger gross profit than establishing how connected Total Revenue, Cost of Revenue, and Gross Profit are to each other. I would look more deeper into all the variables about the Premium subscriptions and see what connections I could have found there to make Spotify more profitable from both premium subscriptions and ad based subscriptions.

What's Next?

For future students embarking on this project I would offer the advice of having specific insights that you would like to analyze before starting the coding analysis. Due to how versed I am in the Spotify scene, I was very intrigued and already interested in this dataset before I had even started analyzing it. I went into my analysis with some questions I had already had about Spotify's cost and revenues. Therefore, in my opinion, it is important to pick a dataset that you have a lot of interest in because it is not only more fulfilling, but also easier because there are certain variables and questions that you have to want to analyze and understand. If I had gone into this dataset with no prior knowledge of Spotify or no prior questions about their business and service, I would have had a more difficult time finding patterns and thinking of next steps for the streaming service. Due to my interest in Spotify, I was able to complete the next step of analyzing global listening patterns in order to apply these insights to a marketing and advertising sense in how Spotify could have more ads that would bring in more revenue and engagement. If I had the opportunity to continue this analysis, I would want to look at how Spotify users are using the app on a larger scale. I think it is so interesting how Spotify tries to individualize every user's experience, and I genuinely believe that they could do way more with that aspect of their service to fully set them apart from Apple Music. Apple Music now has their own "Wrapped" for users that competes with Spotify's. Spotify has to do more to individualize user experiences in a more unique way.

Overall, I really enjoyed my data analysis of Spotify's Quarterly Revenue, and when I first started the project, I had wished that the dataset was by year instead of by quarter because I thought that that would have made it easier to visualize and understand. However, after completing my analysis I came to see how it was actually better that it was in quarters rather than

year. For me, this ended up being the best part of the dataset because of the vast information it gave me. The dataset being in quarters gave way more insights for the analysis than if it had been in years. Through the dataset being in quarters, I was able to see the trend of the highest revenue being at the end of every year, showcasing how much users engage with personalized and individual experiences. Furthermore, my favorite function in R that I was able to use for this dataset was the table() function. The table() provides a frequency table of all the values within a certain variable. Through using this function, I was able to see all of the values within a certain variable, in ascending order, in a fast and easy manner. Not only that, but when the values are displayed in such a manner I was able to see how much these values are increasing by when they are increasing, if it is a little or a lot every quarter. Therefore, this function is, in my opinion, the quickest and simplest function to visualize change and frequency within a variable. To conclude, I found numerous intriguing insights from my analyzation of Spotify's quarterly cost and revenue including the significant relationship between Cost of Revenue and Total Revenue, the low profit of ads, the trend of highest revenue being at the end of every year, and conclusions and insights into more personalized ads that Spotify can implore based on what users are listening to on the app. To my team, I hope these insights provided a clear looking glass into how Spotify's business operates and how all their costs and revenues affect each other. I look forward to seeing how Spotify's business model changes with the advancement of technology and whether or not their cutting of jobs will help their gross profit in any way. While I do not believe it will because most of their costs come from revenue, and administrative costs are the least costly part of their business, it would be interesting to run this analysis with numbers from 2024 and 2025, seeing if the gap between Cost of Revenue and Total Revenue widened at all.

References

Anderson, I., Gil, S., Gibson, C., Wolf, S., Shapiro, W., Semerci, O., & Greenberg, D. M. (2021). "Just the Way You Are": Linking Music Listening on Spotify and Personality. *Social Psychological and Personality Science*, *12*(4), 561–572. https://doi.org/10.1177/1948550620923228

Bastos, R. (2021, July 14). *Statistics #01—Mean, Median, and Mode*. Medium. https://towardsdatascience.com/statistics-01-mean-median-and-mode-d6d8597ed9f1

Curry, D. (2023). Apple Music Revenue and Usage Statistics. (n.d.). *Business of Apps*. https://www.businessofapps.com/data/apple-music-statistics/

Hargrave, M (2023, May 11) Standard Deviation Formula and Uses vs. Variance. *Investopedia*. https://www.investopedia.com/terms/s/standarddeviation.asp

Maurya S. (2023). Spotify Revenue, Expenses and Its Premium Users. *Kaggle*. https://www.kaggle.com/datasets/mauryansshivam/spotify-revenue-expenses-and-its-premium-users

Richardson, D. (2023, October 24). *Spotify stock jumps 10% as cost-cutting boosts streamer to first quarterly profit in a year*. CNBC. https://www.cnbc.com/2023/10/24/spotify-spot-earnings-q3-2023.html

Robinson, K. (2021, April 13). 15 Years of Spotify: How the Streaming Giant Has Changed and Reinvented the Music Industry. *Variety*. https://variety.com/2021/music/news/spotify-turns-15-how-the-streaming-giant-has-changed-and-reinvented-the-music-industry-1234948299/

Saxena, P. (2022, December 8). 30+ Spotify Statistics, Facts and Trends to Know in 2022. *Appinventiv*. https://appinventiv.com/blog/spotify-statistics-facts/

Schneider, M. (2023, December 4). Spotify Slashes Global Workforce By 17% in Latest Cost-Cutting Effort. *Billboard*. https://www.billboard.com/business/streaming/spotify-layoffs-2023-restructure-job-cuts-1235531731/

Standard Deviation vs. Variance: What's the Difference? (n.d.). Investopedia. https://www.investopedia.com/ask/answers/021215/what-difference-between-standard-deviation-and-variance.asp

Wise, J (2023, November 6) How Many People Use Apple Music in 2024? (Demographics) - *EarthWeb*. https://earthweb.com/how-many-people-use-apple-music/

Guzman, C (2023, November 29) Your Complete Guide to Spotify Wrapped 2023. *TIME*. https://time.com/6340656/spotify-wrapped-guide-2023/

23 Essential Spotify Statistics You Need to Know in 2023. (n.d.). The Social Shepherd. https://thesocialshepherd.com/blog/spotify-statistics