

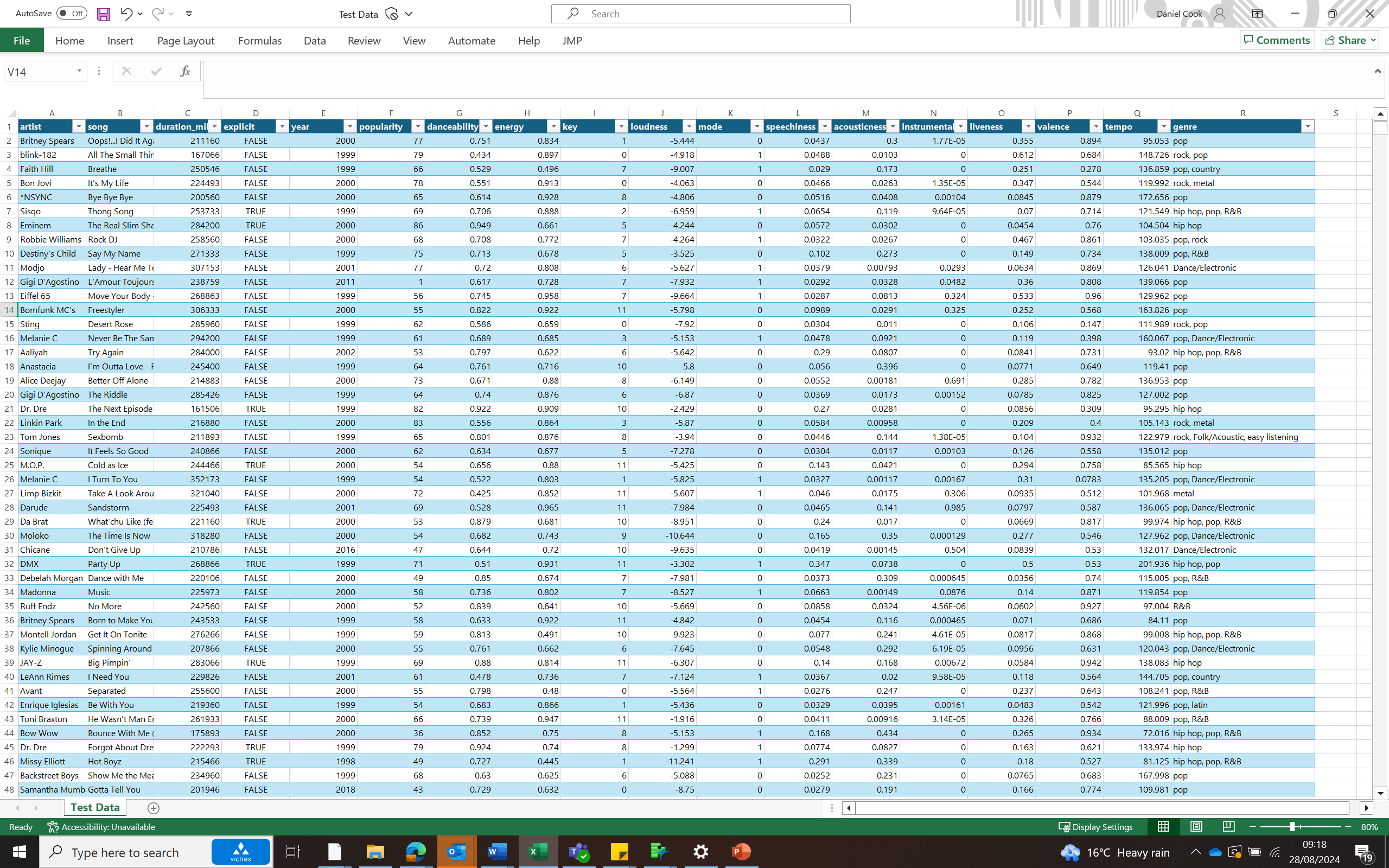
**Methodology**

The challenge I’ve chosen where the application of data analytics can contribute towards addressing is looking to understand the properties of music over time and to predict what type of music will be popular in the future. The statistical technique I have chosen to use to approach this challenge is Time series analysis as enables me to look at outliers, cyclical patterns and use which overall should allow me to suggest possibles explanations for the results and offer insights as to what music will be popular in the near future.

**Data Collection**

To start this off I found a data set on Kagle listing the Top Hits Spotify from 2000-2019 (as shown in figure 1). This data set contains 18 columns each describing the track and it's qualities with a ranking system for each, for example they have given each song a ranking of popularity from 1 to 100, they’ve given the duration of the songs, the loudness in decibels etc.. This data set is good for the challenge I had chosen to address as there any many different qualities that have been measured which will in fact increase the accuracy of my results as I will be able to better the change in types in music over time and I’ll have a better forecast for the future of music. To get this data in the right format that I would need to move on to exploratory data analysis, transferred the data set into a table format (as the data was in a raw form when I opening it). I then copied it into the JMP software in which I changed the types of all the columns to the correct data type e.g. I changed the popularity figure to a numeric and continuous data types. I then found an average figure for each year, for each quality. This gave me a grasp of what the music in that year was like. It also means I can look at the trends over the years without have to many figures clogging up my graphs, making it uninterpretable. To do this I created a summary table using the JMP software which created neat table.

Figure 1 (screenshot of a portion of the data set)



**Exploratory data Analysis**

This is an approach to analysing my data by using visual methods. It helps me to see what the data can tell us beyond the formal modelling. The Exploratory data analysis gives me a better understanding of my data. To start my EDA, I created a summary table on JMP software of the average for all the qualities of the songs (all the numeric ones) for each year to gauge a sense of how the trends have changed over time. To do this I created graphs with the average figures on the Y axis and the Time on the X. On this graph I am seeing an outlier in 1998. This would make sense as there is only 1 data point for 1998 which is significantly lower than the other years as they range from 38-111 data points. I can also see that a lot of the qualities have very stable figures over time, as there is some small fluctuations however the correct themselves and basically stay stable over time. For example energy dipped after 5 years around 2003 and then rose for 3-4 years and now has dropped again. This could suggest a cyclical pattern however the data doesn’t span back long enough to make that assumption as there is only one cycle in this pattern. Other than that there is no stand out assumptions that can be made about the data.

**R2 Metric/Linear Regression**

I put a Linear regression on the graphs and calculate an R2 value. The R2 metric takes the difference between each point and the trend line (error) and adds them together and divides them by the sum of square of each point compared to the average/mean, which shows the correlation between the two axis (in this case the average qualities of the songs and time) to see if the is any clear patterns. As you can see in figure 2 the only two qualities with any reasonable correlation to time is the valence and duration (milliseconds). Valence is the emotional positivity or negativity shown by a song: tracks with high valence sound happier while tracks with low valence sound more negative. The duration of songs against time had an R2 Value of 0.59 and valence against time had 0.5. An R² of 0.5-0.6 means that these models are explaining some of the variance, but there is still a significant amount of variation in the data that it doesn't explain.

**Outliers**

A data point representing a year with too few observations may not be statistically reliable or representative of the true trend. Such a point could introduce noise rather than meaningful information, leading to misleading interpretations. This is why I chose to remove the 1998 data from our graphs (as shown on Figure 5 and 6). This boosts the R2 value to 0.85 (duration) and 0.65 (valence) which is significantly higher than before. This suggests that there is a very strong trend that songs have continuously getting shorter since 1999. An article (Anon, (2023). *Are Songs Getting Shorter? A Comprehensive Analysis - Big Time Musicians*) states “one of the primary drivers behind the decreasing song lengths is the rise of digital platforms and streaming services. Today’s music consumption is largely dictated by platforms like Spotify, Apple Music, and YouTube. These platforms reward frequency, as artists are paid per play. Therefore, shorter songs mean more plays, and more plays translate to more revenue.” This hence shows that there has been less incentive to make longer songs, in fact they are probably more incentivised to make short catchier songs to get better value for their time. Another possible reason for reduced song duration is because of short attention spans. Another article (Tran, K. (2023). *Why Songs Are Getting Shorter - Songwriters Guide - Songwriting beginners*. ) states “In a world where we’re constantly bombarded with information, attention spans are getting shorter. Many people have reported losing interest in a song after just a few seconds, which might explain why artists are trimming down their tracks”. This further supports the trends we are seeing in the data.

Figure 2 and 3 – Picture of the Duration of songs against time and picture of Valence against time.

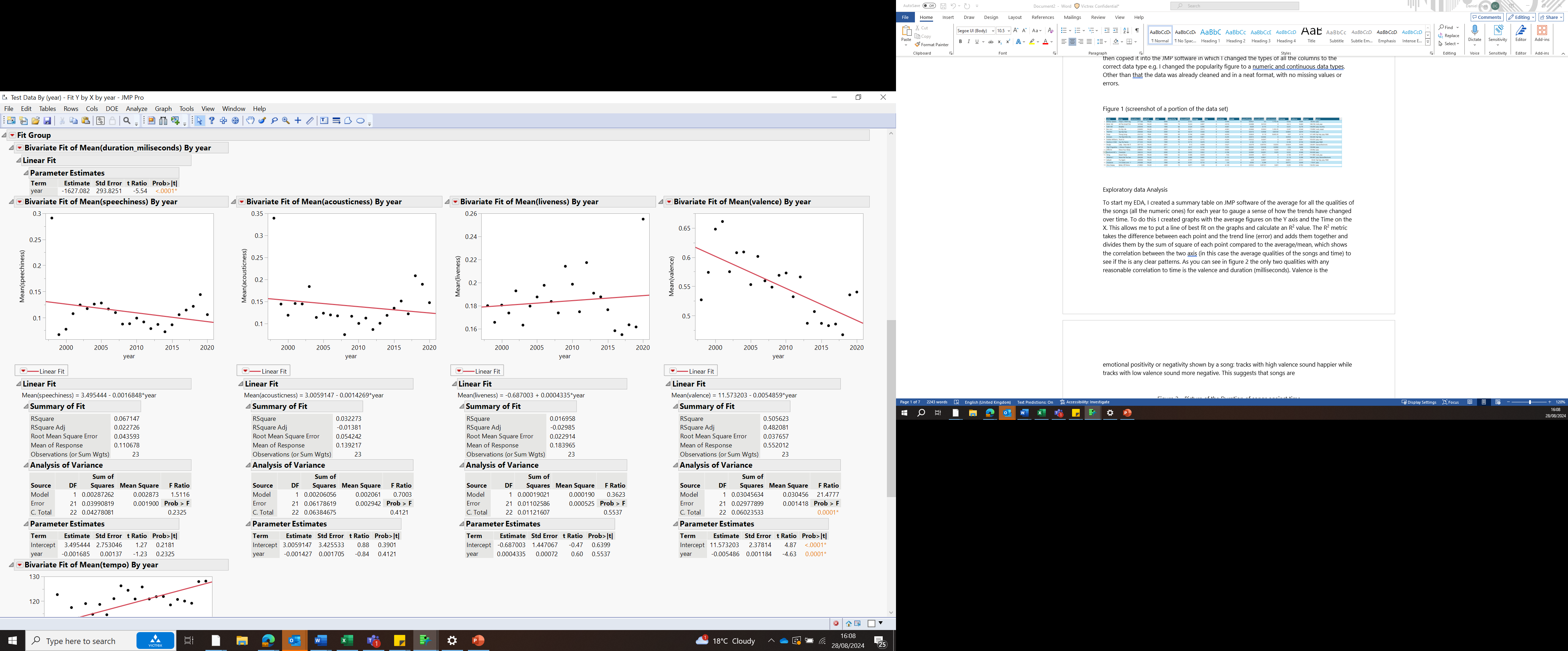
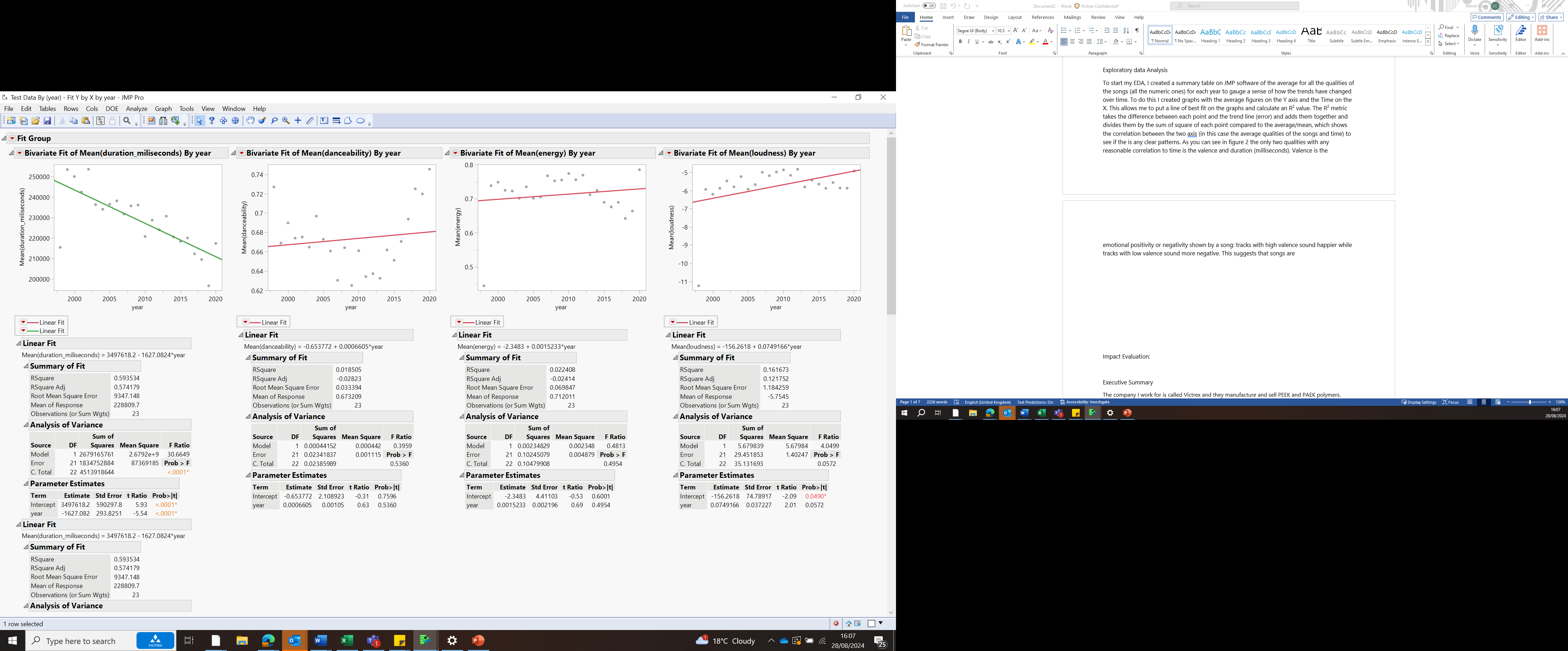
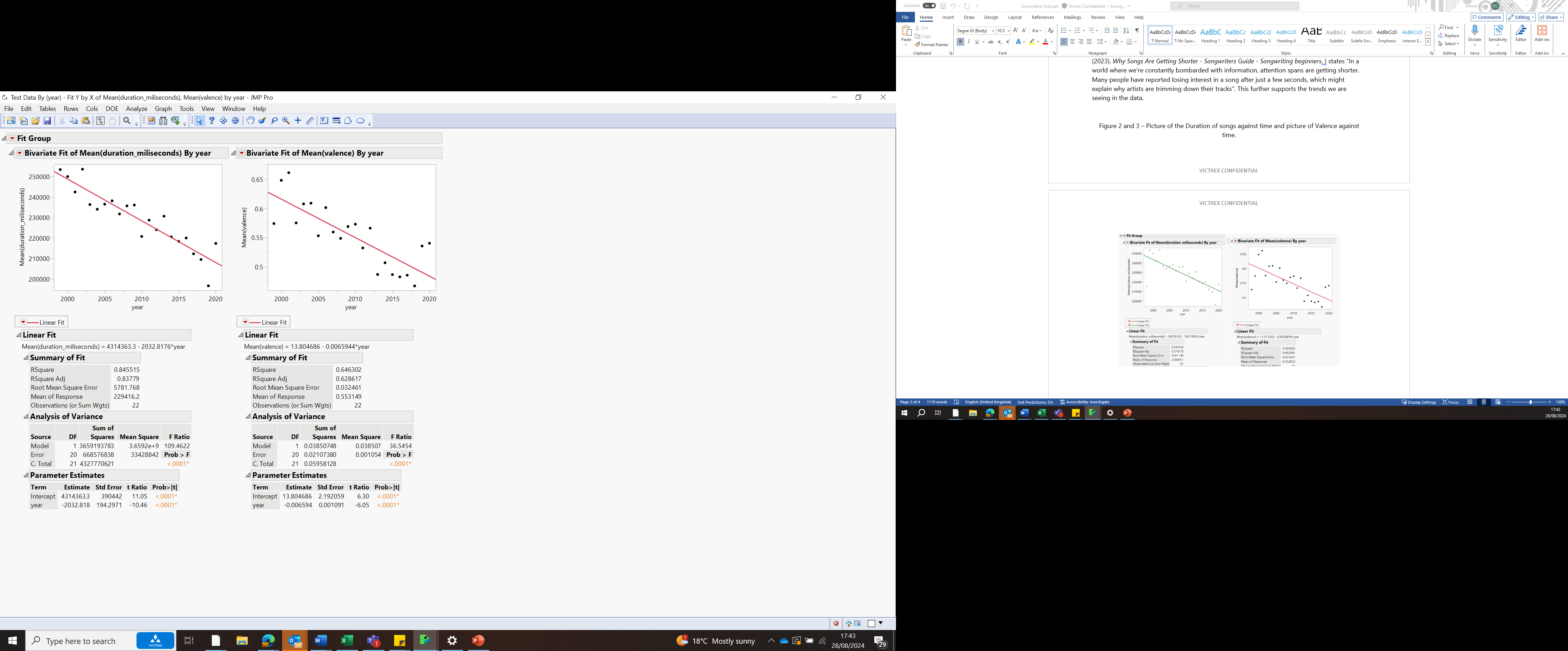


Figure 5 and 6 – These graphs show figure 3 and 4’s graphs without the data point in 1998



Bibliography:

Anon, (2023). *Are Songs Getting Shorter? A Comprehensive Analysis - Big Time Musicians*. [online] Available at: <https://bigtimemusicians.com/are-songs-getting-shorter/>.

Tran, K. (2023). *Why Songs Are Getting Shorter - Songwriters Guide - Songwriting beginners*. [online] Songwriters Guide. Available at: https://blog.lyricstudio.net/2023/05/27/why-songs-are-getting-shorter/ [Accessed 28 Aug. 2024].