Samantha Cabral

DSC 530 – Data Exploration and Analysis

Prof. Metzger

Due: November 17, 2024

Final Term Project

Song Debut Data Analysis of the Billboard Hot 100

Summarization

1. **Statistical/Hypothetical Question**

Do songs that debut in the top 10 stay on the Billboard Hot 100 chart longer than songs that debut between ranks 11 and 100?

1. **Outcome of your EDA**

There is statistical significance between the debut rank and the amount of weeks the song remains on the Hot 100 Chart(Board), although weak with a variance of 4.1%. The scatterplot comparing debut rank and weeks on board shows there is more data that proves something similar: as the debut rank (closer to 100) of songs increases, the time spent on the chart tends to decrease and stay on the chart for less time. This is a negative correlation suggesting that debut rank and time spent on the chart are inversely related.

The R^2 value of 0.041 in the multiple regression analysis shows that debut rank and having a rank of 1 is indicative of a weak relationship between debut rank and time spent on the chart. Other factors not captured by debut rank and rank of 1 also likely influence how long songs stay on the chart, so debut rank is not a strong predictor of the time spent on the chart.

1. **What do you feel was missed during the analysis?**

There are many factors that were not taken into account since no values were provided in the dataset to quantify those factors. Release dates of songs relative to the chart release is one such factor, as I believe it may explain why some songs debut closer to 100 (higher) and then peak at lower ranks.

1. **Were there any variables you felt could have helped in the analysis?**

Since I only considered debut ranks and final ranks, I wonder if rank changes throughout the longevity of songs on the Hot 100 may explain the correlation or causation of their longevity.

1. **Were there any assumptions made you felt were incorrect?**

I feel as though I’m assuming that data that shows the opposite of my statistical question is true also proves that my question is true or correct. Since songs that debut close to 100 have shorter longevity, then songs that debut close to 1 will have longer longevity. Is there a statistical or logical fallacy that can prove my assumption is incorrect?

1. **What challenges did you face, what did you not fully understand?**

While coding my assignment, I accidentally ended up changing my dataset into a Series to find statistical information. When I went back to try to improve my code, I caused major issues that set me back and wasted my time. If I understood more about how the PMF is calculated and visualized, I feel that I wouldn’t have been so frustrated with my assignment.