Toe Jam Listener Data Analysis

ToeJam (Group 41) Uday B, Quang H, Kai F, Garett (Gayle) R

Hypothesis 1

Users have a preference in genre.

Process

In order to determine genre preference we calculated the proportion of genres on the platform, and then calculated the proportion of each genre listened to by each user.
After that we did a summation for each genre, of proportion genre listened by user/proportion genres on platform, which was then divided by the amount of genres.

If completely random the mean of this value will be 1, or $\mu_0 = 1$. Our found mean, x, was 1.11655…, and our found population standard deviation was 0.23559…, the amount of users in this data or n was 1549. So doing the Z score calculation, where $Z = (x - \mu_0)/(\sigma/\sqrt{n})$, we got a Z of 19.47… which is a very high Z score, without a doubt proving that users have genre preferences.

Hypothesis 2

Number of releases increase over time, however newer genres grow faster than older ones.

Process

To determine if the growth of music releases was uniform across genres, we calculated the total number of songs released per year, we calculated the specific rease counts for each genre. Then we performed a linear regression analysis to find the slope of the total dataset and each genre.

If the growth was uniform across all genres, then each would show a similar steady increase over time. However our data found the slopes of some genres, like Rock and Electronic, had extremely high slopes (over 4.5) while others like Classical and Blues had very low slopes (near 0). This difference shows that although the number of releases increase over time, it is heavily dependent on the genre as well. Older genres would include; Classical, Folk, Ethnic, Soul, Blues, Punk, Metal, Jazz. While “newer” genres would include the rest.

Charts

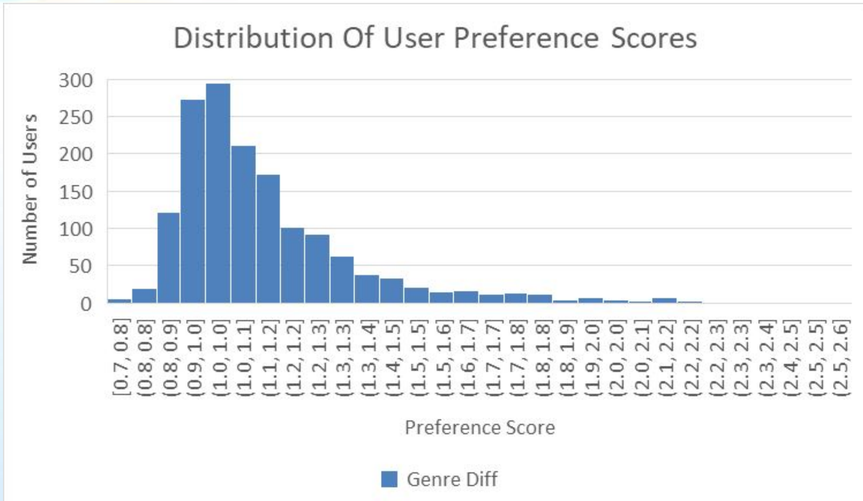

Fig. 1: Distribution of User Preference Scores. Shows a mean that is shifted to the right, which indicates that users actively select specific genres rather than randomly listening to music.

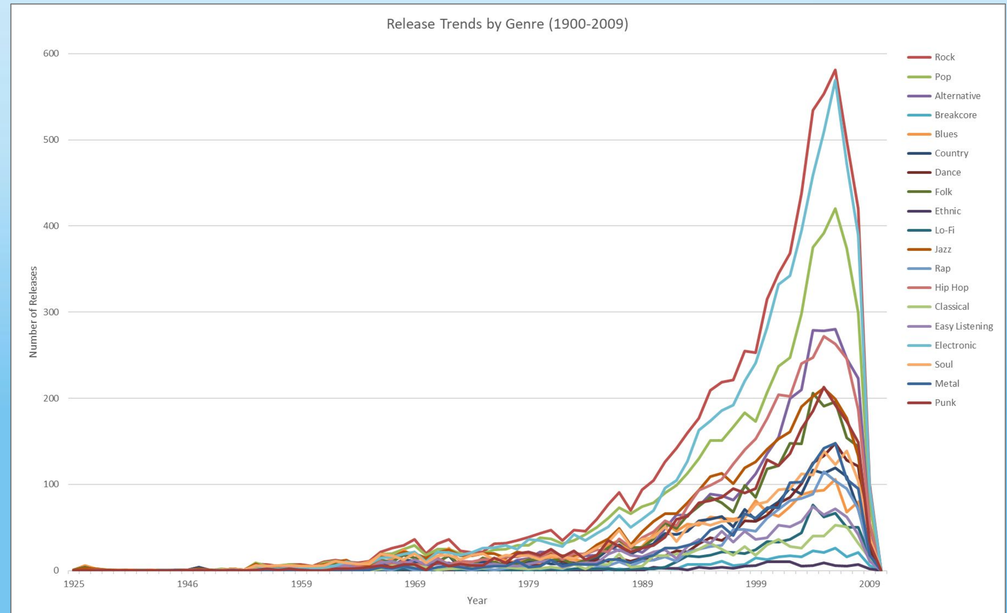

Fig. 2: Release trends by genre from 1900-2009. The chart shows a increase in releases with sharp difference between release rates between modern and older genres over the years. Showing Rock, Electronic, and Pop as the top 3 growers.

Technologies Used

For Data: Python, SQLite3, OpenPyXL
For Charts: Excel