Hypothesis

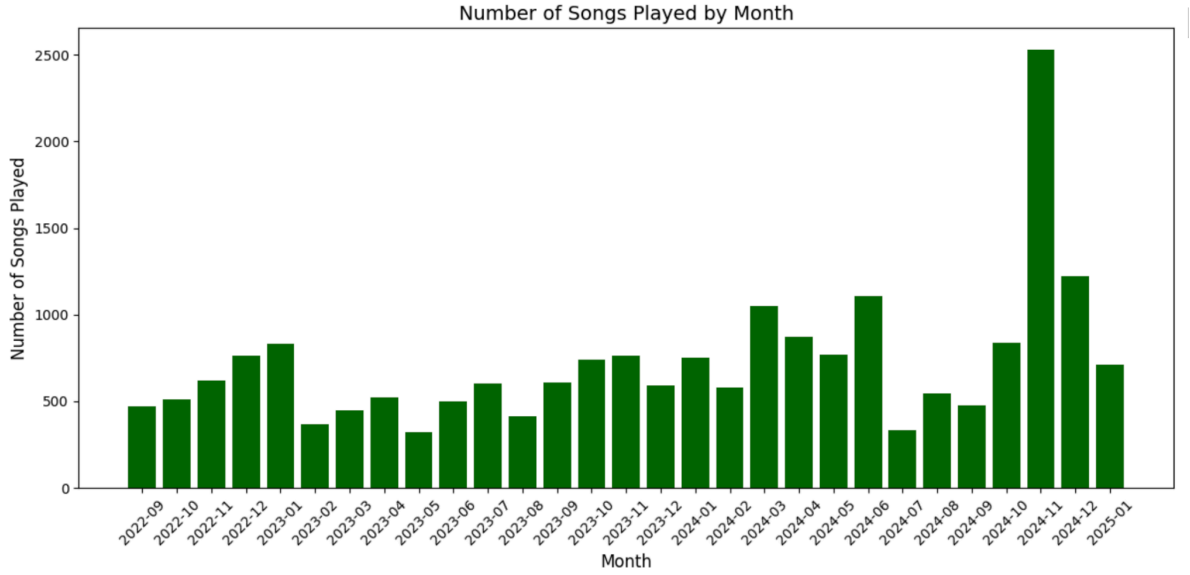
When the weather is warm I spend less time listening to music. On the contrary, when the weather is cold I spend more time listening to music.

Spotify Data

Spotify's personal data request feature is used. The data is sent in .json format. It includes a time stamp, track name, artist name, album name, and other metadata. The files are uploaded into dataframes, concatenated into a single dataframe and columns that will not be used such as "skipped","shuffle" information are dropped. The "ms_played" (millisecond played) column is not meaningful, so using the information in ms_played, minutes played is calculated for each song and the new column minutes_played is added to the dataframe. Rows with missing spotify_track_uri's are also dropped. The column "ts" (timestamp) is not very meaningful as well, so a new column called date is created and timestamp information is converted to dates.

Number of Songs Played by Month

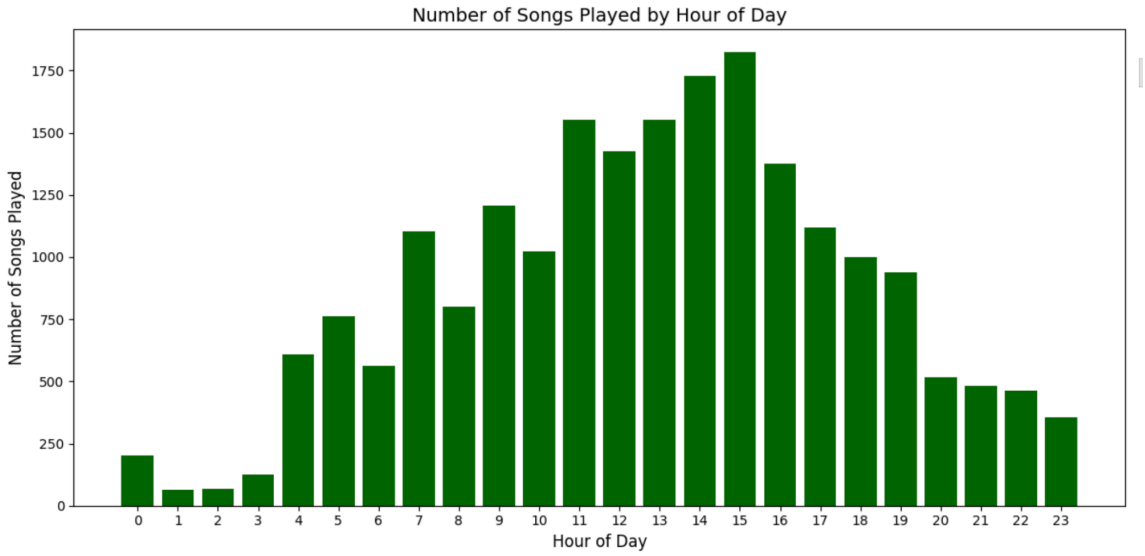
The data is grouped by months using the .groupby method, and the total number of rows (=total number of songs played) is found. Note that months from different years are not concatenated to see the overall trend over the years. A bar chart is plotted:



The bar chart depicts the variation in the number of songs played across different months over a span of several years. The x-axis represents the months in chronological order, while the y-axis indicates the total number of songs played. The chart shows noticeable fluctuations, with some months having significantly higher activity, such as November 2024, which peaks at over 2,500 songs played.

Number of Songs Played by Hour of Day

The hour information is extracted from the date column and data is grouped by hours. The total number of songs streamed for each hour is found. A bar chart is plotted:



The bar chart titled "Number of Songs Played by Hour of Day" illustrates the distribution of song playback activity across the 24-hour day. The x-axis represents each hour of the day, while the y-axis shows the corresponding number of songs played. The chart reveals a clear trend, with a gradual increase in activity during the early morning hours, peaking significantly between 14:00 and 15:00, where the number of songs played exceeds 1,750. Following this peak, there is a steady decline, with minimal activity observed during late-night hours, particularly between midnight and 4:00.

Getting Historical Weather Data

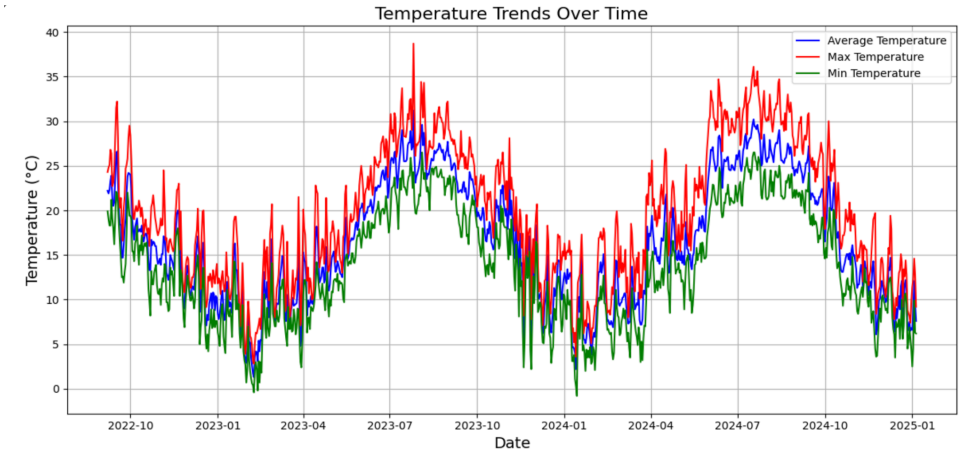
The historical weather information is from an api. First api url and parameters are defined. For this, visualcrossing.com is used. It requires opening an account, and using the api key given for that account in order to fetch weather information. The url includes location information and time period. So data is specifically fetched for a location (istanbul in my case), and a specific date period.

Note that there is a limit to the number of days you can request weather data for, requesting more requires paying for it. In order to not pay, and get the data for free, I opened multiple accounts and used the api keys for each to extract data for various days. This is done to cover all the dates in the spotify dataset, so that each spotify data row can have weather information.

Minimum temperature, maximum temperature, average temperature, precipitation data is fetched from the api. Data is then saved as a .csv file locally. After this, all weather data collected is concatenated into a single .csv file and saved locally. This data is sorted by date.

Temperature Trends Visualisation

The weather data collected is visualised. Minimum, maximum, average temperatures are shown in different colors. It is shown that collected data is correct.

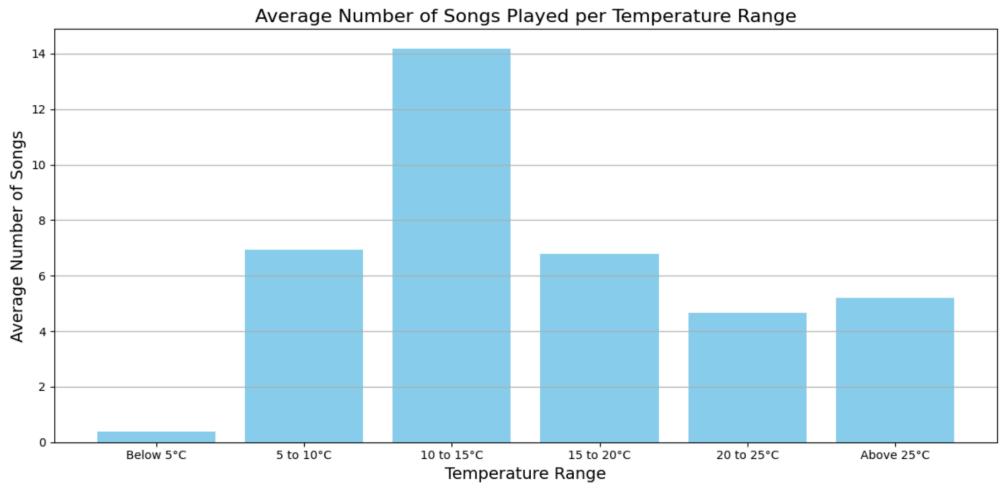


Merging Spotify Data and Weather Date (Temperature)

To the spotify data a new column (average temperature) is added from the weather data we fetched. Date columns for each are made sure to match since for both the .to_datetime method is used. Average temperature information is added for each spotify data, for each song. The newly created .csv file is also saved locally.

Average Number of Songs Played by Temperature

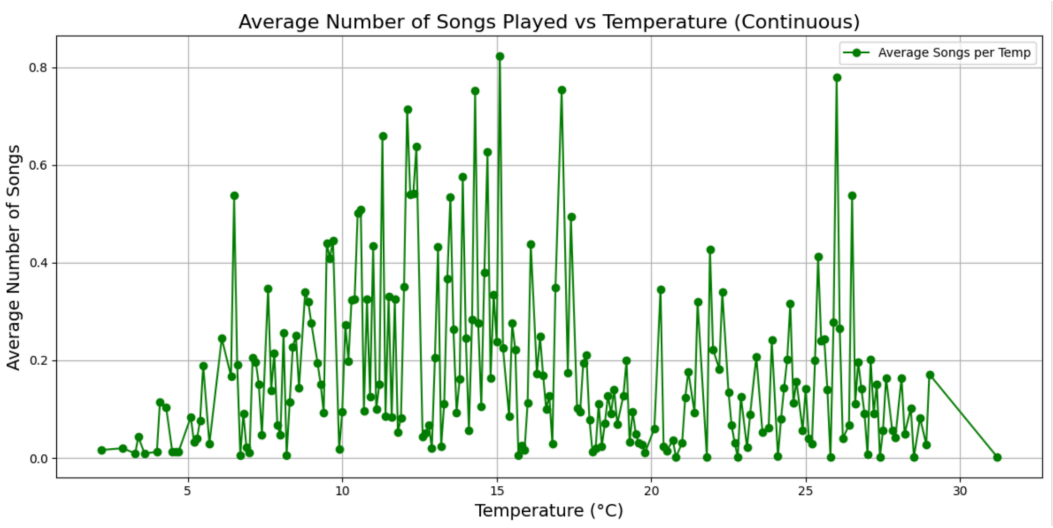
Bins are used to group temperatures like Below 5°C', '5 to 10°C' etc. Merged data (spotify and temperature) is grouped by these bins. Then the total number of songs listened is summed up for each bin, and the average is taken. A bar chart is plotted:

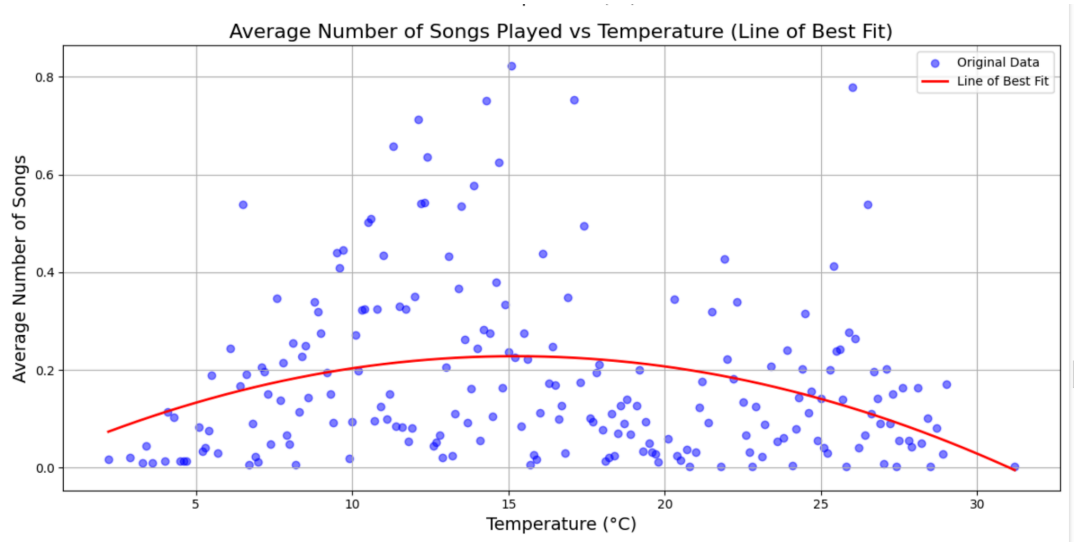


This illustrates the relationship between temperature ranges and the average number of songs played. The x-axis categorizes temperatures into ranges (e.g., below 5°C, 5 to 10°C, etc.), while the y-axis represents the average number of songs played within each range. The data shows a distinct peak in listening activity for the 10 to 15°C range suggesting that this temperature range might align with my listening activities. This trend suggests that listening habits may correlate with specific weather conditions.

Average Number of Songs Played by Temperature (Continuous and with a line of best fit)

The same relationship is observed with a continuous manner, instead of this. In addition to this a polynomial regression of degree 2 is fitted to the data of average temperature x and average songs played y, and a smooth line of best fit is generated using the polynomial model p over a specified range of x values.





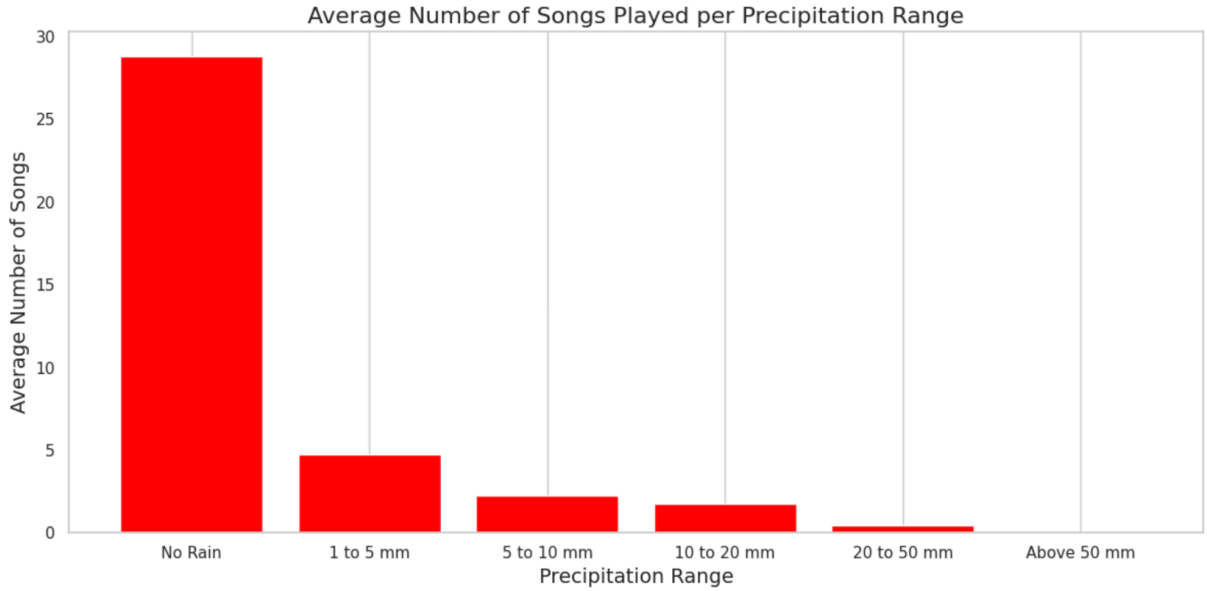
The two plots visualize the relationship between temperature and the average number of songs played. The first plot presents a detailed, point-by-point representation of the data, showing significant variability and scattered peaks across the temperature range. This variability suggests a lack of a simple linear trend, making it challenging to identify a clear pattern. The second plot simplifies the data by overlaying a second-degree polynomial regression line on a scatter plot of the original data. The curve reveals a subtle trend: song activity increases as temperatures rise up to around 15°C but decreases beyond that point. This suggests a possible optimal temperature range for increased listening activity.

Merging Spotify Data and Weather Data (Precipitation)

To the spotify data a new column (precipitation) is added from the weather data we fetched. Precipitation information is added for each spotify data, for each song. The newly created .csv file is also saved locally.

Average Number of Songs Played by Precipitation

Once again bins are used to group precipitation data. Then, the number of songs played for each group is found. A bar chart is plotted showing the relationship between average number of songs played for precipitation intervals.

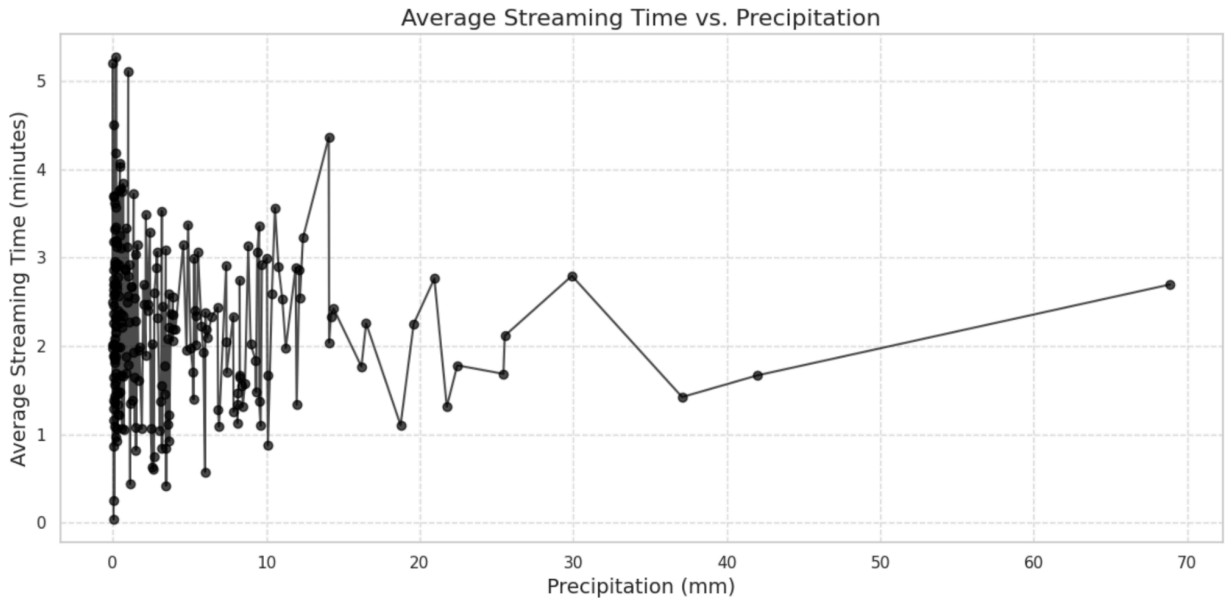


The bar chart demonstrates the relationship between rainfall intensity and song activity. The x-axis represents different precipitation ranges, while the y-axis shows the corresponding average number of songs played. The data reveals a stark contrast: the "No Rain" category significantly outpaces all others, with an average of around 30 songs played, whereas categories with any measurable precipitation exhibit much lower activity. As precipitation increases, from 1–5 mm to above 50 mm, the average number of songs played steadily decreases, suggesting that

rain may discourage listening activity. This sharp drop-off highlights a strong preference for music listening during dry weather conditions.

Average Streaming Time versus Precipitation

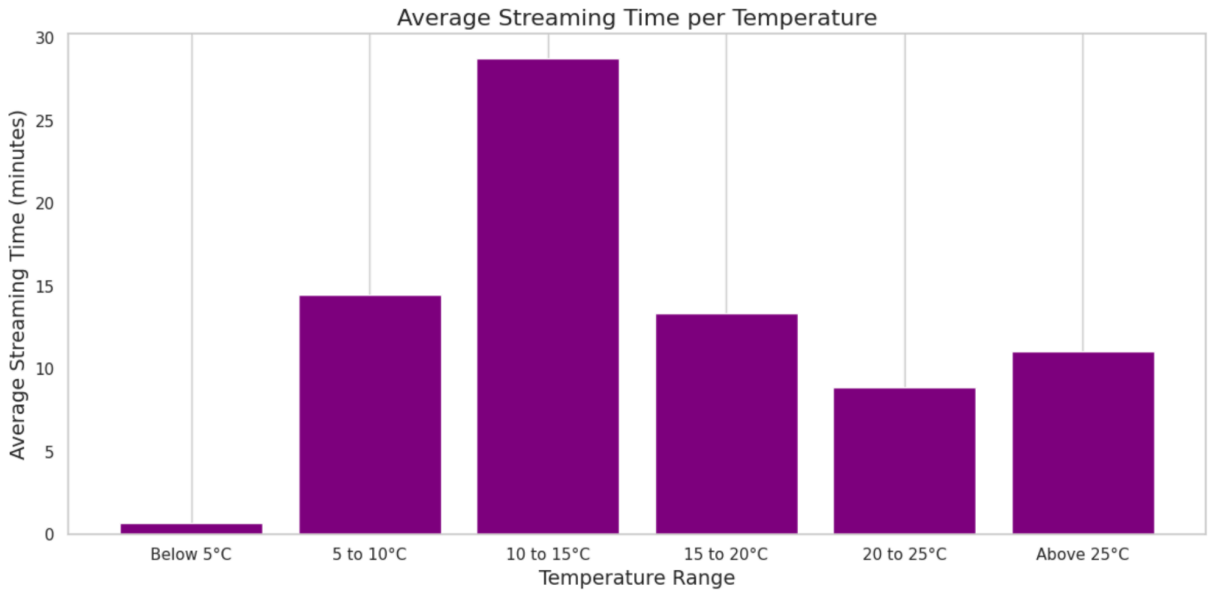
The relationship observed is the same, but this time instead of average number of songs played, average streaming time is observed.



The plot shows a dense cluster of data points at lower precipitation levels (0–10 mm), indicating a high variability in streaming time during light rain or dry conditions. As precipitation increases beyond 10 mm, the data becomes sparser, with a generally increasing trend in average streaming time, particularly at extreme precipitation levels above 50 mm. This suggests that heavier rainfall may encourage longer streaming sessions, possibly due to staying indoors.

Average Streaming Time versus Temperature

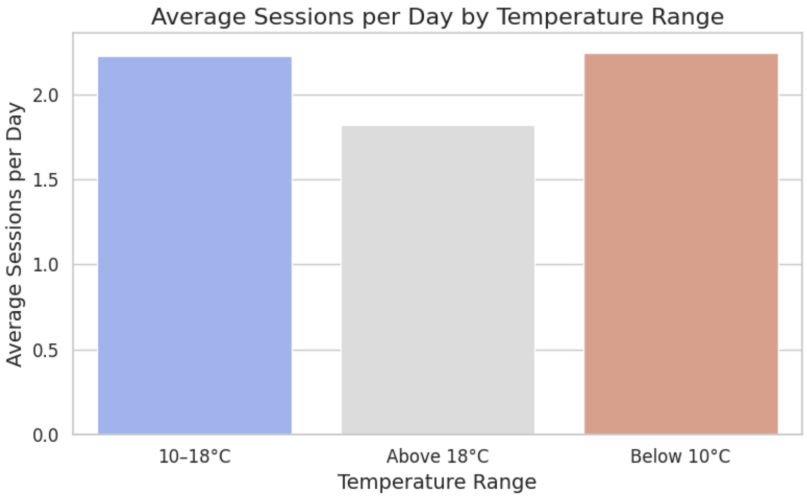
Streaming time and temperature is observed. The temperature information is grouped in bins as previously talked about.



The data reveals that streaming activity is most prominent in the 10 to 15°C range, where the average streaming time peaks at nearly 30 minutes. Lower temperatures, such as those below 5°C, show minimal streaming activity, while moderate to warmer temperatures (15 to 25°C and above 25°C) exhibit relatively consistent but lower average streaming times compared to the peak.

Average Session Counts per Day by Temperature

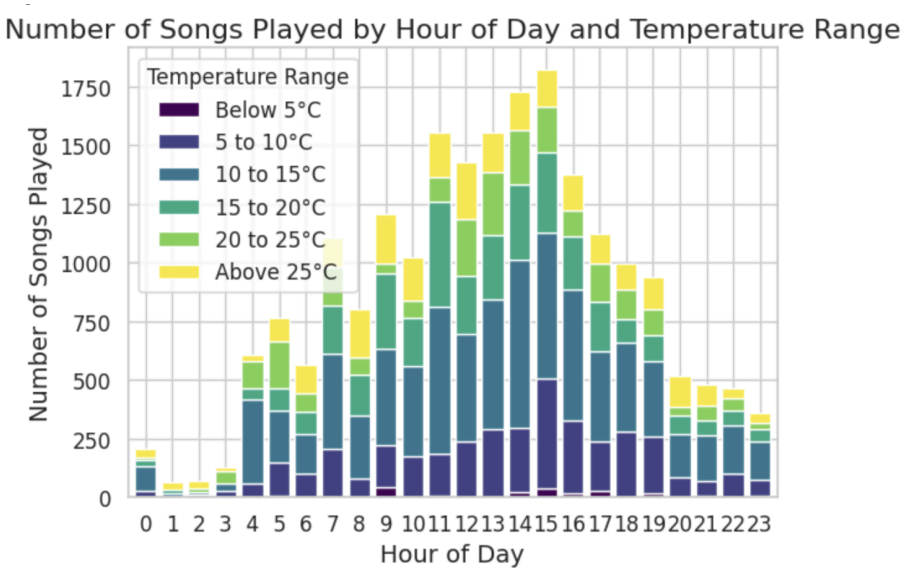
Average session count refers to how many times in a day I spent listening to music. In this sense streaming and stopping streaming and then streaming again results in two sessions. A session ends after 30 minutes of not streaming. Temperature is grouped using bins and for each group session count is calculated.



The data shows similar engagement levels for "Below 10°C" and "10–18°C," both averaging just over two sessions per day. In contrast, the "Above 18°C" category shows a slight decline in the number of sessions per day, indicating reduced activity during warmer conditions. This pattern suggests that cooler or moderate temperatures may encourage more frequent streaming sessions for me.

Number of Songs Played by Hour of Day and Temperature Range

To analyze how song playback varies with temperature, custom temperature ranges (or bins) are defined, such as "Below 5°C" and "5 to 10°C." Each data point is categorized into one of these ranges based on its temperature. The data is then grouped by both the hour of the day and temperature range, calculating the number of songs played for each combination.



The stacked bar chart visualizes how song playback varies throughout the day across different temperature ranges. The x-axis represents the hours of the day (0–23), and the y-axis shows the total number of songs played. Each bar is divided into segments corresponding to six temperature ranges, from "Below 5°C" to "Above 25°C," as indicated in the legend. The chart highlights that song playback peaks between 12:00 and 16:00 across most temperature ranges, with "10 to 15°C" and "15 to 20°C" contributing significantly during these peak hours. In the early morning and late evening, playback activity decreases, with minimal contributions from colder temperatures ("Below 5°C").

Hypothesis Evaluation

From the observed and visualized data, the hypothesis turns out to be not true. I spend much more time listening to music when the weather is warm, particularly around 15°C.

Limitations

The aim was to observe spotify listening habits with relation to weather (temperature, precipitation). To do this the initial plan was to also get tempo, danceability, energy, acousticness information for each track from spotify api. These would've given a more meaningful and extensive analysis. Temperature versus tempo etc would be observed which is more meaningful. Normally, from developer.spotify.com with the account information and upon a creation of an app, id and key information is given and from these audio features api is accessible. However, Spotify has restricted access to some of its api's and these audio features are no longer accessible. This was not the case while planning and writing a proposal for this project. Unfortunately, there is nothing to be done for these restrictions. The apps created prior to restrictions can only access these api's and there is no way for apps that are created after the restrictions to access these data.