Group 4 Project Report

Amba, Amina, Becca, Bianca, Lisa, Nikki, Yrina

Project Title: Spotify Listening Trends and the Economic Climate

Introduction:
Primary Project Question: How do Spotify users' genre preferences vary with macroeconomic factors in the UK, including measures like unemployment, the rate of inflation and GDP?

Background: Media streaming platforms like Spotify collect vast amounts of behavioural data, offering valuable insights into public mood and cultural consumption. However, there is limited research examining how changes in the economic landscape influence listening habits among the UK public.

This study aims to explore whether shifts in macroeconomic indicators are reflected in Spotify genre preferences, with a focus on understanding music consumption as a potential real-time cultural barometer of economic sentiment. If a link can be established — such as a tendency to listen to relaxing music during recessions or energetic music during periods of growth — it may offer a novel way of measuring collective emotional responses to economic changes. This research fills a key gap at the intersection of economics, media studies, and behavioural science.

Methodologies:
Data sources: A combination of APIs and publicly available datasets were used to collect the relevant economic and music data.
Music Data APIs & Repositories:
- **Spotify**, **AudioDB** & **LastFM** **APIs** – Used to extract genre classification for tracks from historical dataset.
- **Spotify UK Top 200** (2016–2022) – GitHub repository of weekly top tracks, used for historical data on top tracks.

Economic Datasets (from the UK Office for National Statistics – ONS):
- **GDP (Quarterly National Accounts)**
- **Inflation (Consumer Price Indices)**
- **Unemployment Rate**
- Consumer Confidence data was initially included but excluded from final analysis due to high multicollinearity with GDP.

Data Processing and Cleaning: Following extraction, the datasets were cleaned and standardised for analysis:
- **Economic Data:** Group members Bianca, Becca, Amina, and Lisa processed ONS data by aligning it to quarterly intervals, removing null entries, and applying consistent formatting. Aggregates for each quarter were calculated for GDP,

inflation, and unemployment.
- **Music Data:** Historical charts dataset was first correctly mapped to corresponding economic quarters by Nikki. Yrina was then responsible for generating genre tags by calling three different APIs, and after the dataset was cleaned and processed.

Tools and Libraries Used:
- Data Analysis & Visualisation: Pandas, NumPy, Matplotlib, Seaborn
- Machine Learning/Statistical Analysis: Scikit-learn, SciPy, Statsmodels (VAR regression)
- Natural Language & Sentiment Analysis: NLTK (VADER), Spotify's audio features (valence/tempo)
- Environment & Collaboration: GitHub, Jupyter Notebook, Notion

Implementation Process: The team divided the work into thematic and technical streams:

Exploratory Data Analysis & Regression:
- **Amina** – Genre streams vs inflation
- **Bianca** – Genre streams vs GDP and unemployment
- **Becca** – Vector Autoregression (VAR) analysis
- **Lisa** – General exploratory data analysis (EDA) and data visualisation

Sentiment Analysis:
- **Yrina** – Sentiment Analysis - Natural Language Processing on song titles
- **Amba** – Valence and tempo-based sentiment profiling

Improvement of Data Quality and Non-Technical Requirements:
- **Nikki** - Spotify Data Web Scraping and Team Organisation

Agile Elements and Collaboration:
We started working in weekly sprints, dividing up tasks and reviewing progress in weekly Teams meetings. Closer to the deadline we increased meeting frequency to daily check-ins. Additionally, we carried out code reviews where and when we deemed beneficial.
To increase collaboration we used Notion, where we especially in the beginning gathered all our ideas and documented individual progress and challenges.

Challenges:
Initially, we expected to extract genre tags directly from the Spotify API at track/song level. However upon implementation, we discovered that genre classifications were primarily assigned at artist level, which limited the precision for our Analysis. As a result, we supplemented Spotify data with additional data LastFM & AudioDB.

For sentiment analysis, we focused on song titles rather than full song lyrics due to time and resource constraints. Conducting sentiment analysis on lyrics would have required integration with multiple APIs and more complex data processing pipelines which was beyond the scope and timescale of this project.

Another challenge we encountered was the inconsistency in the temporal frequency of our datasets. The Spotify data was available in weekly format, whereas the economic indicators from ONS were reported on a quarterly basis. To enable a meaningful comparison and statistical analysis across these sources we needed to aggregate the datasets by aligning them to a common quarterly time frame to ensure consistency and analytical validity.

Results & Key Takeaways:

Chill Music: Streams in the Chill genre showed a strong positive correlation with unemployment. This implies that during times of economic stress, listeners may turn to more calming music, potentially as a form of emotional self-regulation.

Electronic Music: Showed a moderate negative correlation with unemployment, suggesting that upbeat or energetic genres may decline in popularity during downturns, possibly reflecting reduced social activity or mood.

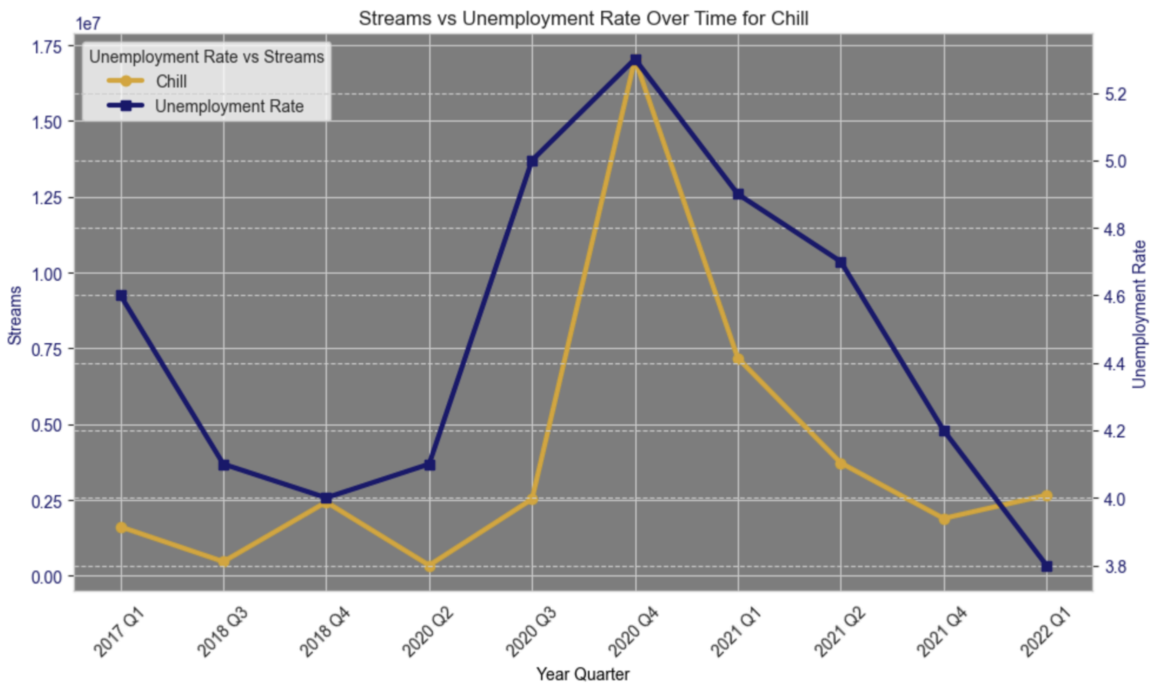

Figure: Streams vs Unemployment Rate Over Time for Chill — This graph shows how streams in the Chill genre and the unemployment rate moved together over time. Both peaked in late 2020, suggesting listeners turned to relaxing music during heightened economic stress, and declined as the economy recovered.

	GDP	Inflation	Unemployment
Jazz	0.64	0.42	−0.55
Chill	−0.07	−0.34	0.71
Rap/Hip-Hop	−0.24	−0.49	0.02
Rock/Punk/Metal	0.09	−0.01	−0.12
Dance/House	−0.10	−0.26	0.16
Electronic	0.15	−0.34	−0.36

Figure: Correlations between genre streams and economic indicators. Chill shows a strong positive correlation with unemployment (0.71). Electronic is negatively correlated with unemployment (-0.36). Other genres show weaker or mixed relationships.

Vector Autoregression (VAR) Analysis: Economic Factors and Genre Listening Trends

We conducted Vector Autoregression (VAR) analyses to explore whether changes in unemployment and inflation have lagged effects on genre listening behaviour over time. Each economic factor was modelled separately alongside selected genre variables. Only variables that were stationary (or transformed to be stationary) and not highly collinear were included in the models.

In the unemployment-focused model, VAR results, Granger causality tests and impulse response functions (IRFs) suggested that rising unemployment was associated with a short-term increase in the share of streams made up by chill and indie music, followed by a longer-term decline. A possible explanation is that listeners turn to these genres for emotional support or stress relief in the short term. However, over time, prolonged unemployment may shift focus away from introspective genres, possibly due to changing routines or priorities.

In the inflation-focused model, we found that inflation predicted a decline in chill music listening after one and three quarters, although these effects were weak and within IRF confidence intervals. This suggests a potential, but subtle, link between inflation and listening behaviour, perhaps reflecting reduced emotional bandwidth or lifestyle changes during times of economic pressure.

Finally, we observed that increases in K-pop listening appeared to predict rising inflation one to four quarters later. Again, this effect was weak and within confidence bounds, but it may suggest that K-pop acts as a proxy for broader consumer mood or optimism, rather than being causally related to inflation itself.

Impulse Responses:

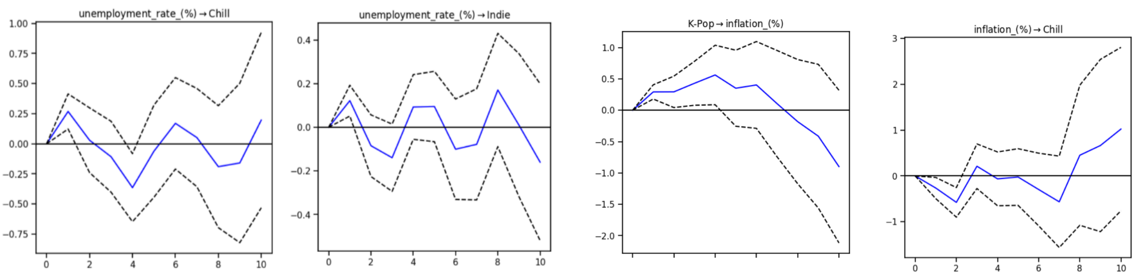

Figure: Impulse response plots showing how economic shocks impact genre streams. Unemployment shocks have a positive short-term effect on Chill streams and on Indie followed by a negative effect long-term. Inflation shocks negatively affect Chill in the short-term, and K-Pop positively predicts inflation over time.

Correlation vs Causation:

We see that Christmas music correlates positively with GDP and inflation. However, this is likely due to a seasonality factor as GDP tends to increase in Q4 of most years and Christmas music is almost exclusively listened to in Q4 of each year as well.

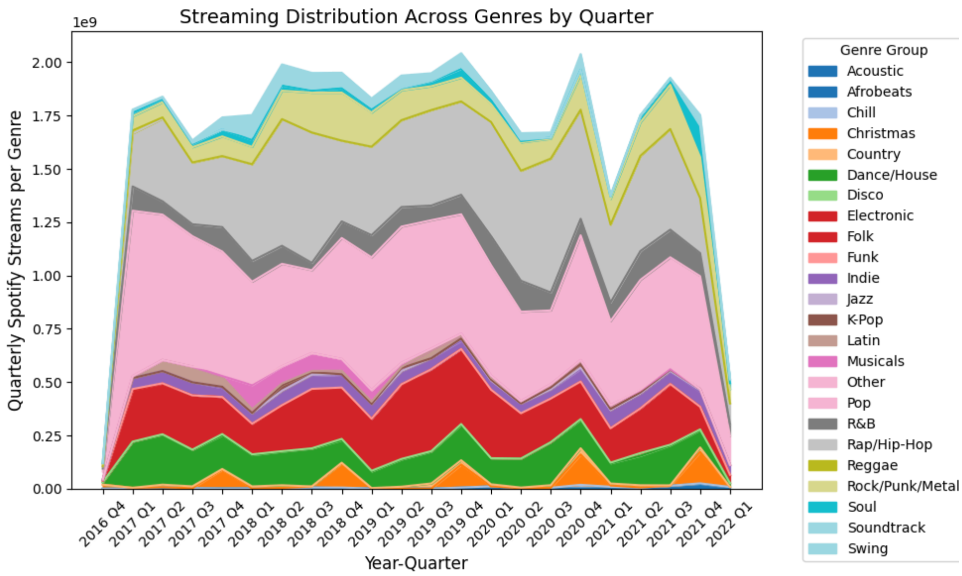

Figure: Quarterly distribution of Spotify streams by genre (Q4 2016–Q1 2022). Clear spikes in Christmas music streams appear every Q4, reflecting seasonal listening habits. This pattern aligns with regular Q4 GDP increases, illustrating how seasonality, not causation, explains the positive correlation between Christmas music and economic indicators.

Sentiment Analysis:

As an extension to the main study, we explored whether musical sentiment exhibits any patterns across genres and time, and whether it correlates with economic changes. To quantify sentiment, we applied VADER pre-trained sentiment analysis model to song titles.

We found that genres such as Soul, Dance/House, Electronic, Pop and Christmas showed a clear tendency towards positive sentiment. In contrast, Rap/Hip-Hop, Rock/Punk/Metal, R&B, and Indie emerged as the most negatively scored genres.

We then compared the overall sentiment scores to key economic indicators: GDP, inflation, and unemployment rate. Across the full period, we found no consistent or strong correlation between sentiment and these economic measures. However, two prominent dips in sentiment occurred just before and during the COVID-19 lockdown period. This coincided with a sharp fall in GDP, reflecting the economic recession triggered by the pandemic. While not definitive, this alignment suggests a possible influence of economic uncertainty on the sentiment of popular music during that time.

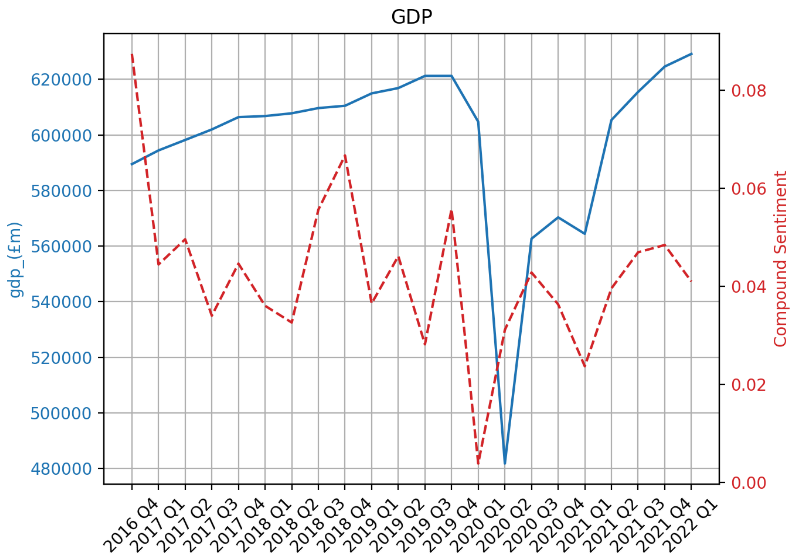

Figure: Time series plot showing UK GDP (blue solid line, left axis) alongside the compound sentiment of song titles (red dashed line, right axis). While GDP shows a sharp decline during the 2020 recession and a subsequent recovery, sentiment fluctuates over time without a clear long-term trend. No obvious relationship is observed between GDP and aggregate musical sentiment in song titles over the studied period.

Conclusion:
In summary, we can see that this project provides initial evidence that the listening trends for individual music genres show some correlation with economic measures. Most notably, we found a consistent positive correlation between chill music consumption and rising unemployment rates. Suggesting that listeners may gravitate toward calming or introspective genres during economically challenging periods. This aligns with the hypothesis that music serves as a form of emotional regulation or escapism in times of financial stress.

However, while these correlations are compelling, they do not necessarily imply causation. A more definitive understanding of how economic factors influence listening behaviour would require controlling for additional variables such as seasonality, time of day, and individual mood states—each of which can significantly affect music choices. Moreover, expanding the dataset beyond the current 2016–2022 timeframe would increase the robustness of the analysis, offering a longer historical context to validate trends and improve predictive reliability.

To deepen the emotional and cultural insights derived from this work, future research could incorporate sentiment analysis based on full song lyrics rather than just titles. Lyrics may reveal more nuanced emotional content that resonates with listeners depending on their lived economic realities regardless of genre. Since lyrics can cut across stylistic boundaries, analysing them could uncover emotional themes not easily captured through genre alone.

For our target audiences, these insights offer distinct applications. For stakeholders in the music and streaming industry (including platforms like Spotify), record labels, artist managers and music marketers could use such insights to fine-tune recommendation systems, plan genre-based campaigns, or time releases to match shifts in public sentiment.

For academics across sociology, cultural/media studies, and behavioural economics, this work provides a foundation for exploring music as a socio-economic barometer. Meanwhile, data enthusiasts and mid-level data science practitioners can view this project as an example of how interdisciplinary data analysis—combined with transparent methodology and thoughtful visualisation—can yield impactful insights.

Overall, our findings highlight the value of examining music consumption alongside economic conditions. Incorporating machine learning, sentiment analysis of lyrics, additional behavioural variables, and extended datasets could greatly enhance the utility of such research, offering the music industry powerful tools for more emotionally attuned, data-driven engagement strategies.