1 Method's Current Usage

1.1 Suicide Trends in the Early Months of the COVID-19 Pandemic: An Interrupted Time-Series Analysis

Authors: Jane Pirkis et al.
Published: The Lancet Psychiatry, 2021
Summary: This study uses interrupted time-series analysis to assess the impact of COVID-19 on suicide rates in 21 countries. Monthly suicide trends from pre-pandemic (2019–March 2020) and early-pandemic (April–July 2020) periods were compared. The study found no overall increase in suicide rates, with significant decreases in 12 regions (e.g., New Zealand, Alberta). The findings suggest that community support and government interventions mitigated early pandemic-related risks, though long-term effects remain uncertain.
Takeaway: This demonstrates how time-series analysis can reveal temporal trends and evaluate external event impacts, inspiring my project to analyze event-driven changes in music innovation and identify significant shifts over time.

1.2 Time Series Facebook Prophet Model and Python for COVID-19 Outbreak Prediction

Authors: Mashael Khayyat et al.
Published: Computer Modeling in Engineering & Sciences, 2021
Summary: This study uses the Facebook Prophet model for time series analysis to predict the spread of COVID-19 in Saudi Arabia. The model forecasts daily and weekly trends for confirmed, recovered, and death cases. With preprocessing and cleaning of the dataset, the Prophet model generates predictions that align closely with observed values for deaths but shows limited accuracy for recovered cases. Results indicate that while the model is effective for certain metrics, additional data and refinement are needed for more reliable forecasts.
Takeaway: This application highlights the utility of Prophet for capturing temporal patterns, inspiring its use in my project to analyze and predict time series trends in music innovation over time.

1.3 Aggregating Prophet and Seasonal Trend Decomposition for Time Series Forecasting of Italian Electricity Spot Prices

Authors: Stefano F. Stefenon et al.
Published: Energies, 2023

Summary: This study integrates Facebook Prophet and Seasonal and Trend Decomposition using LOESS (STL) to enhance electricity price forecasting. STL pre-processes data by isolating trend, seasonal, and residual components, reducing noise, while Prophet forecasts using these components. The method improves forecast accuracy, particularly for trend analysis, with metrics showing an 18% reduction in mean absolute percentage error (MAPE). The approach captures complex market dynamics influenced by seasonality and external factors like holidays.
Takeaway: This hybrid methodology demonstrates the value of combining decomposition and forecasting tools, inspiring a potential application in my project to analyze and predict temporal trends in music features with improved accuracy.

2 My Experimentation

2.1 Goal

I conducted **interrupted time series analysis** with *Prophet* to quantitively identify the impact of the streaming platform on the global music innovation.
Background: Previously I've found that the number of tracks of different innovation levels show similar trend before 2000, but diverge after 2000. With some domain knowledge, I hypothesize that the streaming platform is the key factor that causes this significant change in innovation trending.

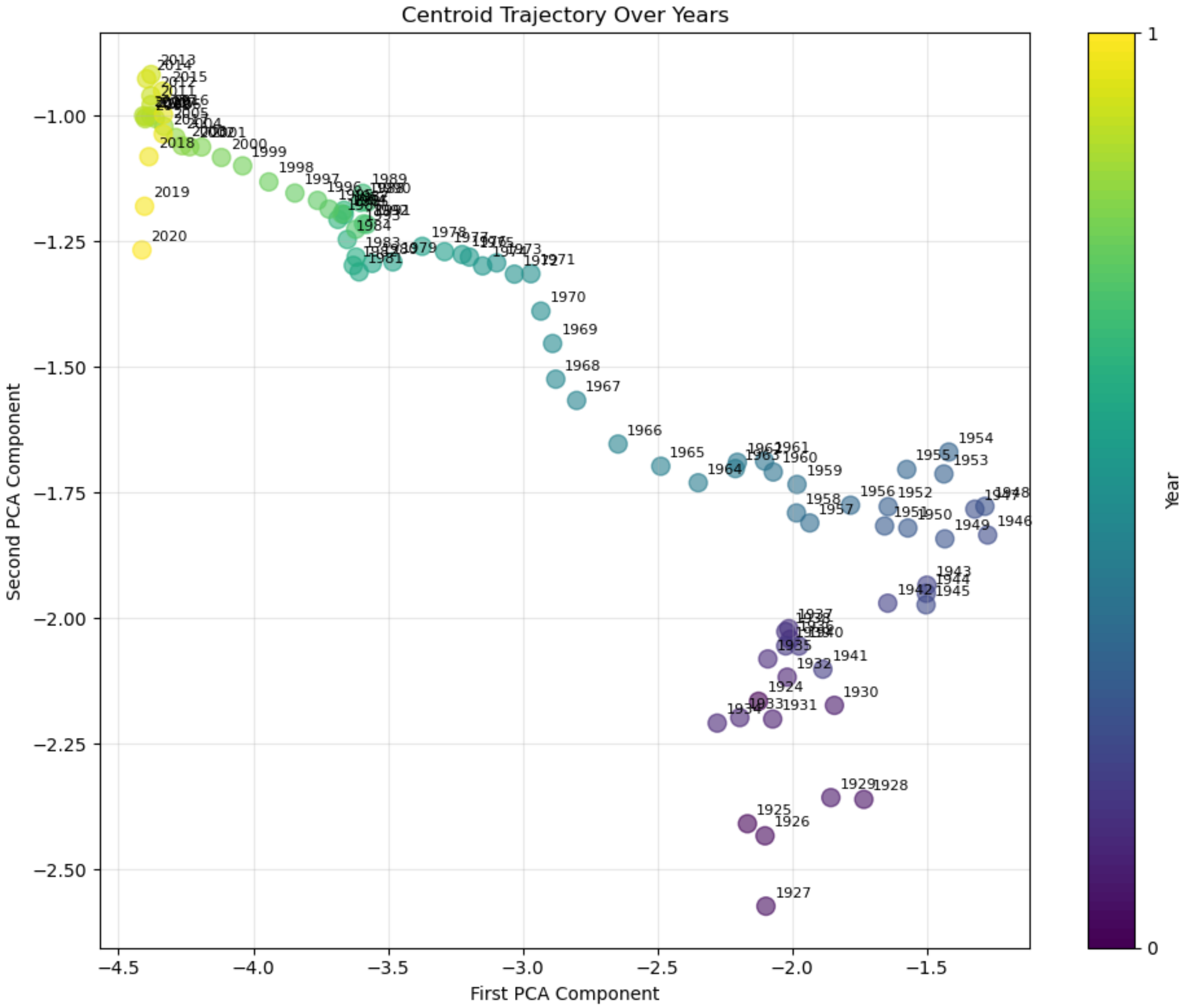
2.2 Challenges

1. **Parameterization**: Prophet has many parameters, like trend, seasonality and changepoints, and they are sensitive to the dataset. I need to tune then to get a good fit of my data.
2. **Model Validation**: I need to validate the model to avoid overfitting.
3. **Visualization**: There are different innovation levels of music, and different innovation levels of music have different trajectories. I need to visualize the results in a comparative way that can clearly show the impact of the streaming platform on the music composition across different innovation levels.

2.3 Process

1. This analysis is based on data derived from *scalable computation*: the number of tracks at different innovation levels across years.
    - As in the original dataset, many released dates only include the year instead of full date, to maintain consistency, all the analysis are based at the **year level**.
    - Innovation level is calculated based on the distance of one track to the centroid of all tracks in the last 3 years before its released year (inclusive) in the PCA space. This time span choice was suggested by professors at the workshop, to better describe

innnovation —— more likely to be influenced by recent history.



- The distance is then binned to set levels.

2. I first tuned Prophet to fit the data in terms of overall time series patterns:
   - Since the analysis is based on year, I didn't include any seasonality on shorter time scale, for example, `weekly.seasonality`, in the model.
   - By observing the trajectories, I found that they do have some similar patterns in general:
     - They showed two downward "spikes" around 1966 and 1993: the number of tracks suddenly declined, then quickly rebounded to the original trend in the next year. I included these two spikes as `holidays` in the model.
     - The trajectories also showed some periodic patterns across longer time scale, for example, around 40-year cycle. I included these patterns as self-defined `seasonality` in the model.
3. Then I try to describe the impact of the streaming platform on the music composition across different innovation levels through the **interrupted time series analysis**:
   - Clarify period of event impact: According to the information I have found on internet, the first streaming platform was born in 1999, and many new platforms have emerged since then. Therefore, I consider the period after 2000 as the period of streaming

influence. Methodologically.
I introduce a new event variable to the Prophet regressor: Before 2000, its value is 0, and after that, its value is 1.

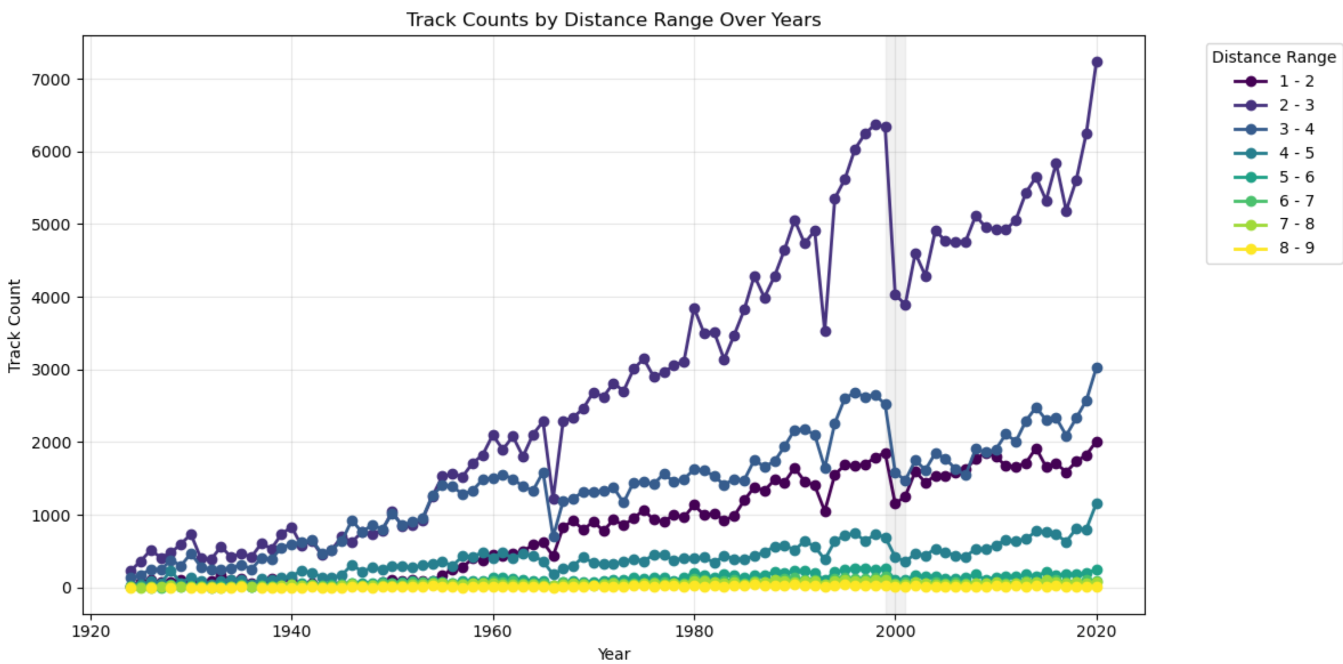
- Through observation, I found that the streaming event might influence the music composition in both slope and intercept. Therefore, I built two variables based on the event variable:
    - `intercept_change`: this variable is equal to the event variable. It's coefficient reflects the influence of the streaming platform on the **number** of tracks.
    - `slope_change`: this variable is the product of the event variable and the year distance to 2000. For example, in 1995, it's 0; in 2005, it's 5, and in 2010, it's 10. The coefficient reflects the influence of the streaming platform on the **growth rate** of the number of tracks.

4. Train the model:
    - For each year, I binned the innovation distance into 1-length intervals (for example, 1~2: distance to centroid from 1 to 2), and calculated the number of tracks at each innovation level, yielding several by-year sequences. Note that the 0~1 interval or interval larger than 10 has too small number of tracks, so I didn't include them in the future analysis.
    - Then I trained separate models for trajectories of each innovation level. I set the `mcmc_samples` = 500 to get more stable results and able to get the confidence interval of the results.
5. Visualization:
    - **Streaming effect on innovation level:** I exacted the slope and intercept of streaming effect from each model, and visualized them together: the x axis is the innovation level, and the y axis is the coefficient of streaming effect. Orange line is the slope, and blue line is the intercept. With shading, I visualized the confidence interval of the results.

3 Findings and Implications

This analysis demonstrates how innovation in music composition has evolved over time and how technological shifts, such as the emergence of streaming platforms, have influenced this evolution.

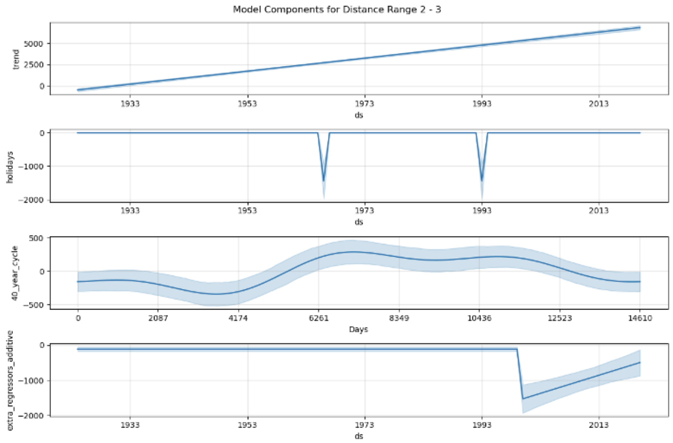
3.1 General pattern （especially before 2000）

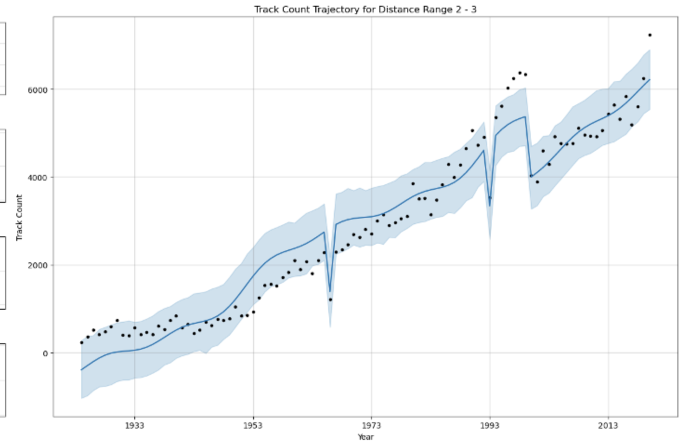


Generally, the higher innovation level, the lower number of tracks. The exception is the innovation level 1.0~2.0, where the number of tracks is less than 2.0~3.0 interval. The number of tracks in this closest central interval is relatively small, which may indicate that music creation is not a pure exploitation of existing music styles but more likely includes some minor creativity. Across time, the number of tracks of different innovation levels increase over time (positive trend), with some seasonal fluctuations, with period of 40 years.

Notably, here are two downward "spikes" around 1966 and 1993, where the number of tracks suddenly declined, then quickly rebounded to the original trend in the next year. According to the information I have found on internet, this might be due to the technological revolution in the music industry, such as the invention of the electric guitar and the development of the recording technology.

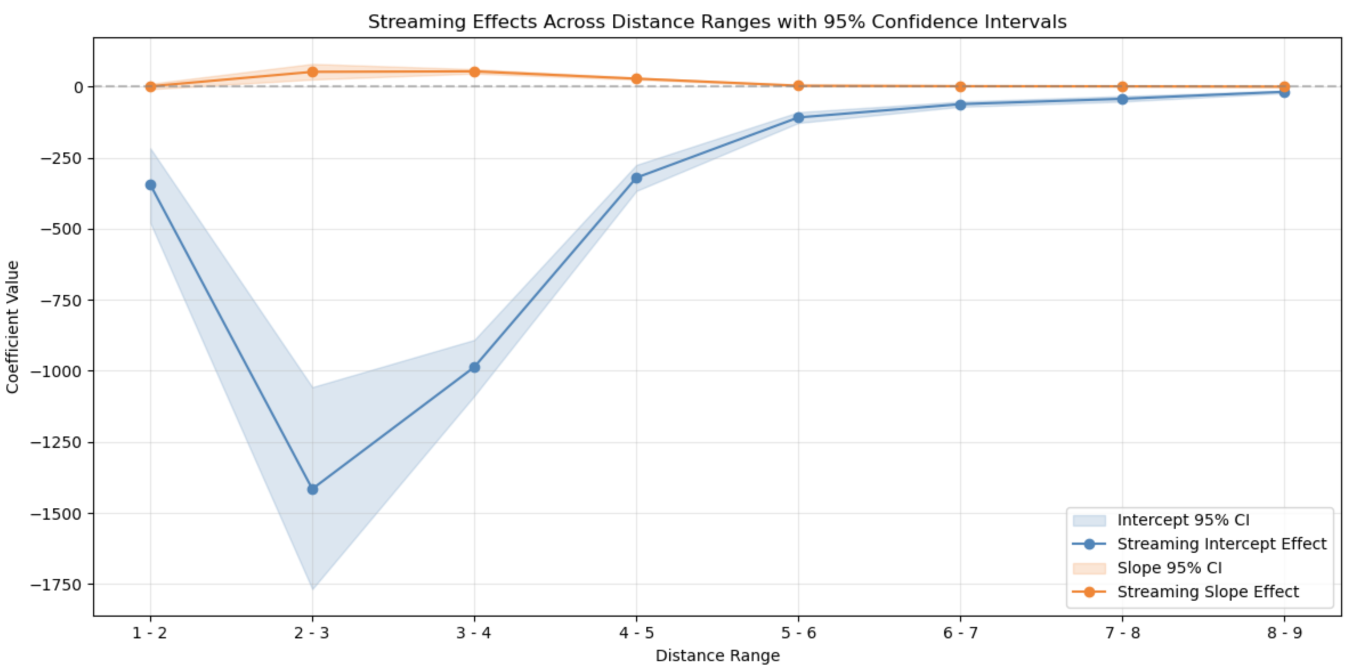
Those observations are confirmed by the time series analysis. For example, here is the model results of innovation level 2.0~3.0:





Interestingly, the number of tracks of different innovation levels show divergent pattern after 2000. The number of low innovation tracks show a explosive increase while high innovation tracks get stagnating. This observation inspire the following analysis.

3.2 Streaming platform impact



Interrupted time-series analysis highlights significant effects of streaming platforms on both the intercept (the baseline number of tracks) and the slope (the rate of growth in production).
By investigating the coefficients of streaming effect at different innovation levels, we can find that the coefficient of intercept is negative for most of the innovation levels. This is reasonable, as the emergence of streaming platform rised some copyright issues, leading to the drop down of track composition. Briefly, the lower innovation level is, the more negative the coefficient is. This might be because the lower innovation has higher baseline number of tracks.
However, if we look at the coefficient of slope, we can find that the coefficient of slope is positive only for low innovation tracks (distance to centroid from 2 to 5), while not significant from zero or even negative for medium and high innovation tracks. This suggests that the streaming

platform has a significant promotion on the number of low innovation tracks, but not on the medium and high innovation tracks.

Also, one exception is the innovation level 1.0~2.0, where the coefficient of slope is not significant from zero, and the coefficient of intercept is smaller than 2.0~3.0 interval in the sense of absolute value. This might be due to its small number of tracks, as shown in the previous trajectory figure.

Together, the analysis suggests that when streaming platforms emerged, they indeed setback the music composition in general; but later they selectively promote mass production of tracks at low innovation levels, reinforcing mainstream trends.

3.3 Summary

The findings shed light on the broader cultural dynamics of technological transformation in the music industry.

Streaming platforms have accelerated the homogenization of music composition. Streaming platforms have democratized access to music, allowing independent artists to reach wider audiences as well as other artists' works. However, such information accessibility may inadvertently lead artists to be more likely inspired by popular music, resulting in a gradual convergence in their creations.

In conclusion, this research deepens our understanding of the interplay between technology, culture, and creativity.