Università degli Studi di Padova

Analysis of an emergent artist's growth on Spotify

Business, Economic and Financial Data Project

Chiarello Federico (ID: 2058163)
Marinelli Andrea (ID: 2091700)
Rinaldi Giorgia (ID: 2092226)

Outline

1. Analysis Objectives
2. Introduction to Music Data
3. Artist Profile
4. Exploratory Data Analysis
5. Modeling
6. Conclusions

Analysis Objectives

Analysis **Objectives**

- Examine the **growth** of an emerging artist to understand the **factors** influencing it.
- Highlight the artist's **strengths** and **weaknesses**, emphasizing **areas to leverage** for further growth.

Introduction to Music Data

How to measure an artist's growth?

Streams and **Followers** are two candidate

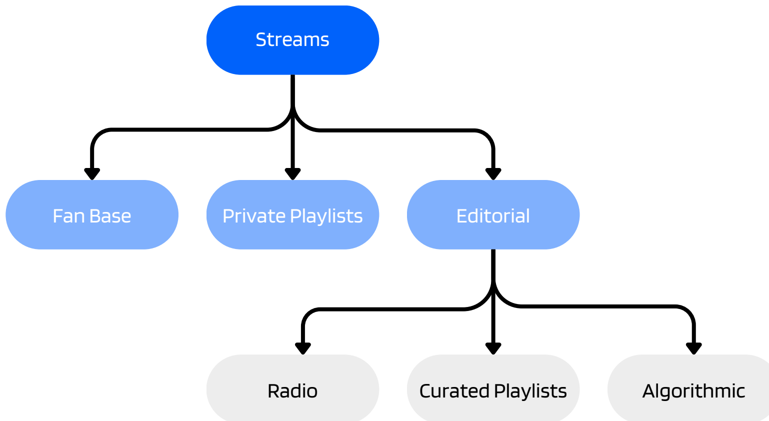
For an emerging artist, the **evolution of streams** represents the primary factor for measuring growth

How does Spotify count streams?

- Track stream: it is counted when a user listens for at least 30 seconds

Sources of the streams

Where do the streams come from?



Growth over time

Different strategies for different time horizons

Short term

- **Aim**: Increase visibility to expand audience.
- **Tools**: Private playlists, sponsorships, events, featuring.

Long term

- **Aim**: Build a solid fan base.
- **Tools**: Establish a strong identity, regular publications, getting featured in editorial sources.

Artist Profile

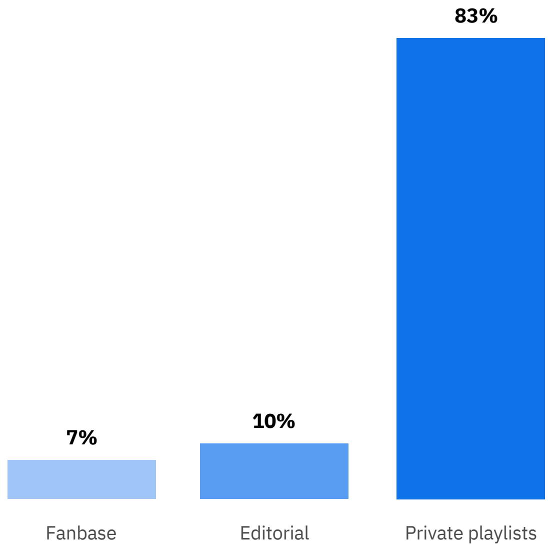
Artist Profile

The artist at the focus of the proposed analysis is a Ligurian producer and mix/master engineer, originally from the city of Genoa.

He mainly deals with the production and mix/mastering of rap/pop tracks aimed at a young audience.

Sources of Streams

- **Private Playlist** are the main source of streams
- Unestablished **Fan Base**
- Limited Presence in **Editorial**



Exploratory Data Analysis

Gethering Data

Data Source:

Spotify for Artists is an online platform provided by Spotify that allows musical artists to manage and promote their presence on the Spotify. It provides some useful **Statistics and Analytics.**

Spotify for Artists

Data set:

Our dataset contains information related to 5 variables within the time range from 01-01-2021 to 15-10-2023..

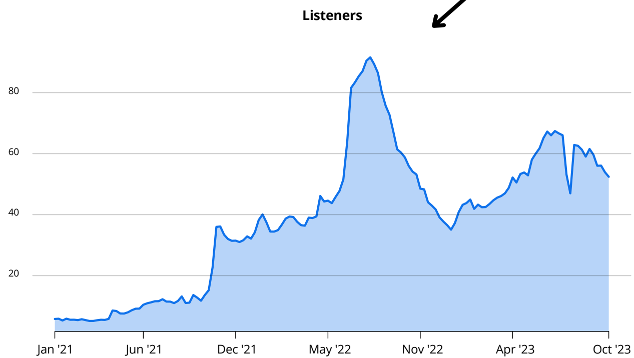
Listeners	Average number of streams in the last 30 days
Streams	Total amount of daily listens of all artist's songs
Playlist count	Number of private playlists in which the artist is featured
Playlist reach	Total number of followers of all playlists that feature a song from that artist
Followers	Number of users following the artist's page

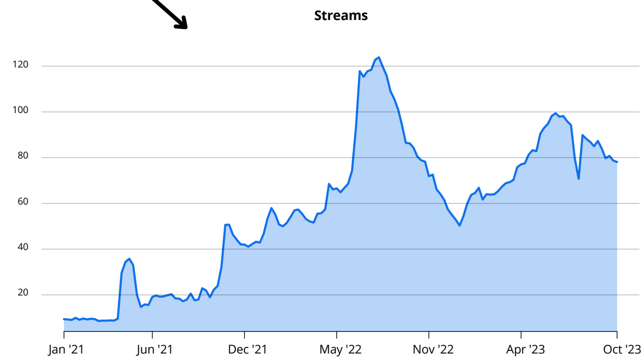
Streams vs Listeners

The variable **listeners** can be interpreted as the moving average of an artist's daily **streams** over a month.

Since we need to choose one between them, we decide to analyze the variable **listeners** as it provides a more stable and representative view of the artist's audience, smoothing out fluctuations.

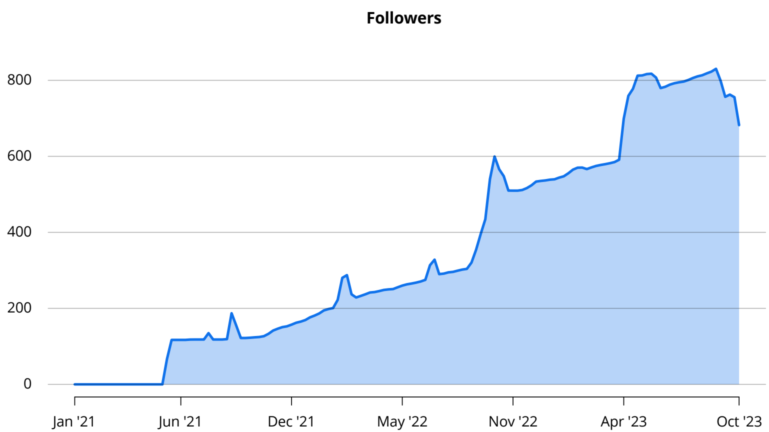
Correlation about 99%





Followers

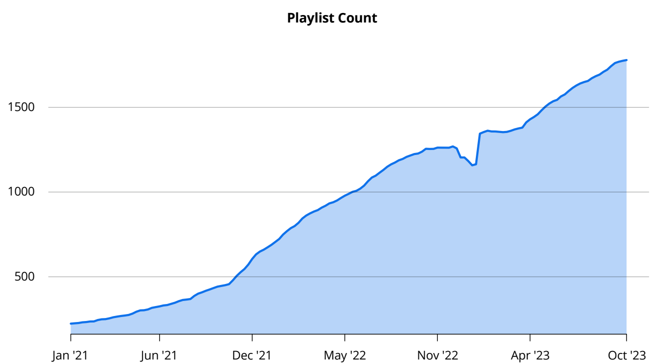
Considering that the variable **followers** introduces multicollinearity issues and represents a source of streams that is not particularly significant (only 7% of streams come from the fan base, i.e., 'followers'), we choose to exclude it from our analysis.



Playlists

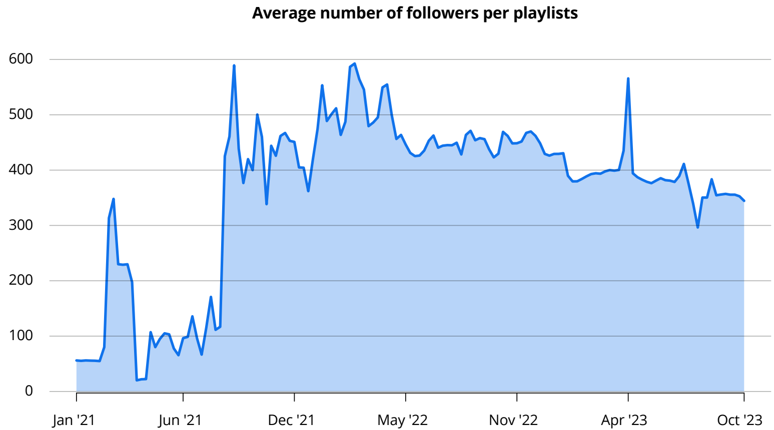
The variables **playlist count** and **playlist reach** exhibit a high correlation, suggesting a mutual explanation.

This correlation introduces challenges of multicollinearity, prompting us to consider defining a new variable through a transformation of these two.



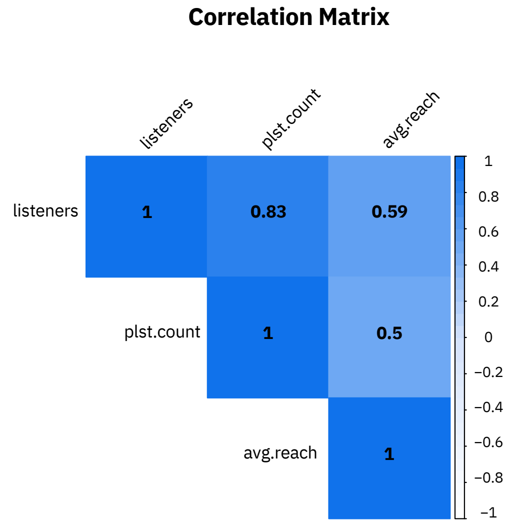
Average Reach

Defined as the ratio between **playlist reach** and **playlist count**, **avg.reach** represents the average number of followers per playlist.

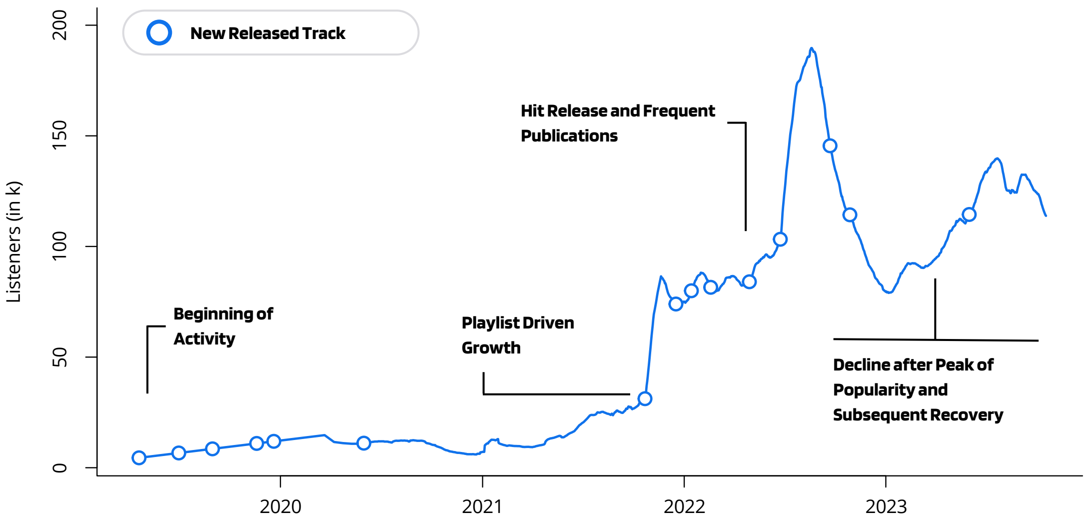


Final **Dataset**

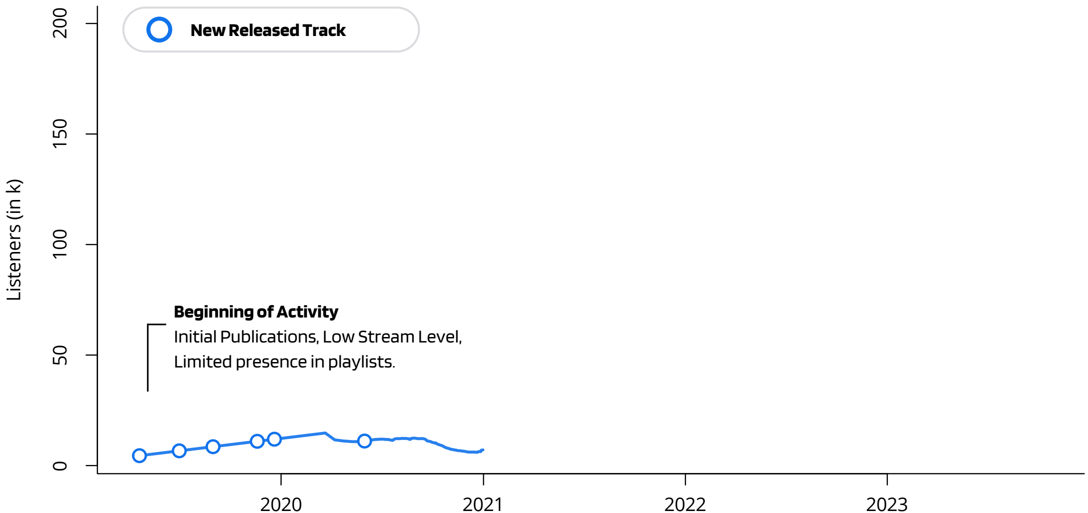
- The final dataset comprises three variables: **listeners, playlist count** and **average reach**
- The data has been transformed from **daily to weekly** to reduce variability



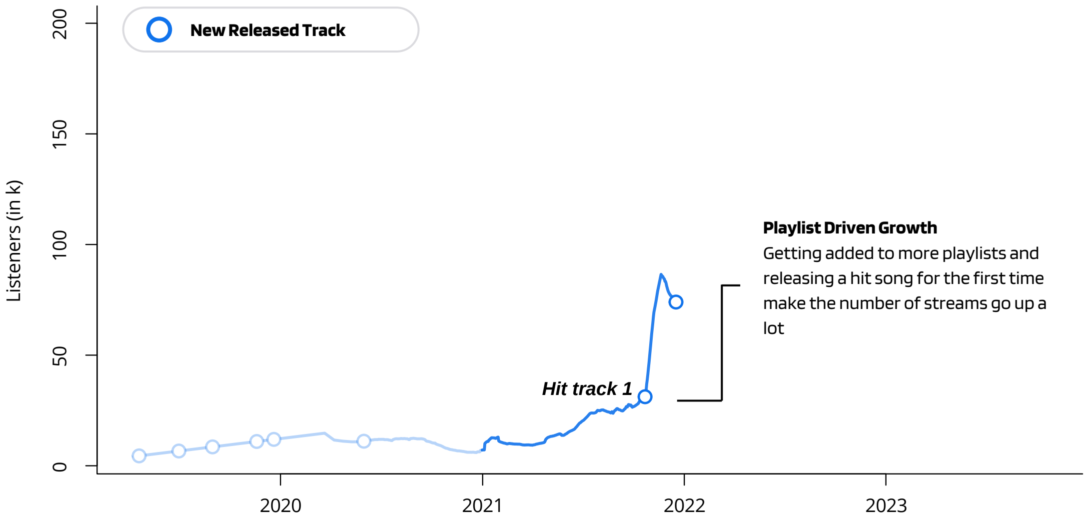
Overview of the artist's activity



Overview of the artist's activity

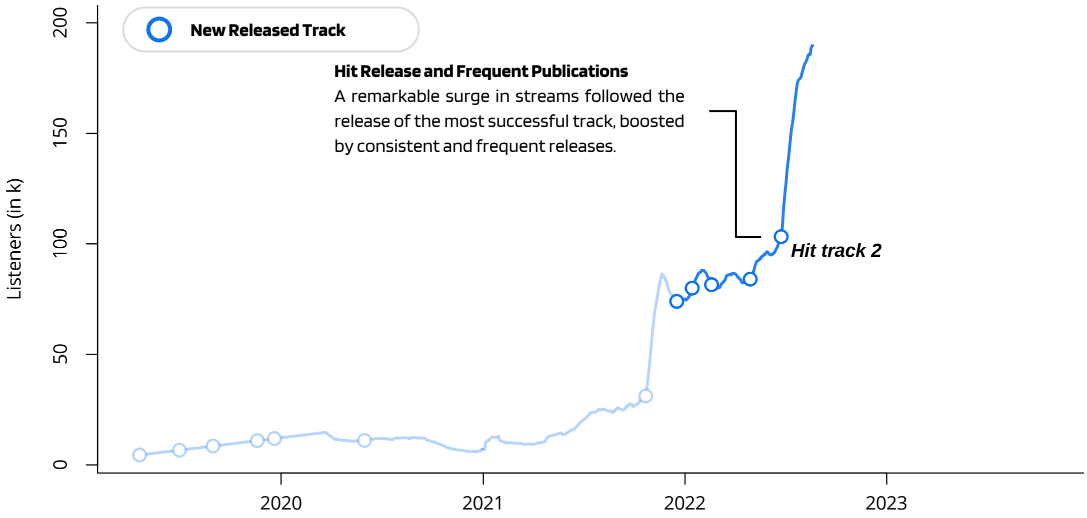


Overview of the artist's activity

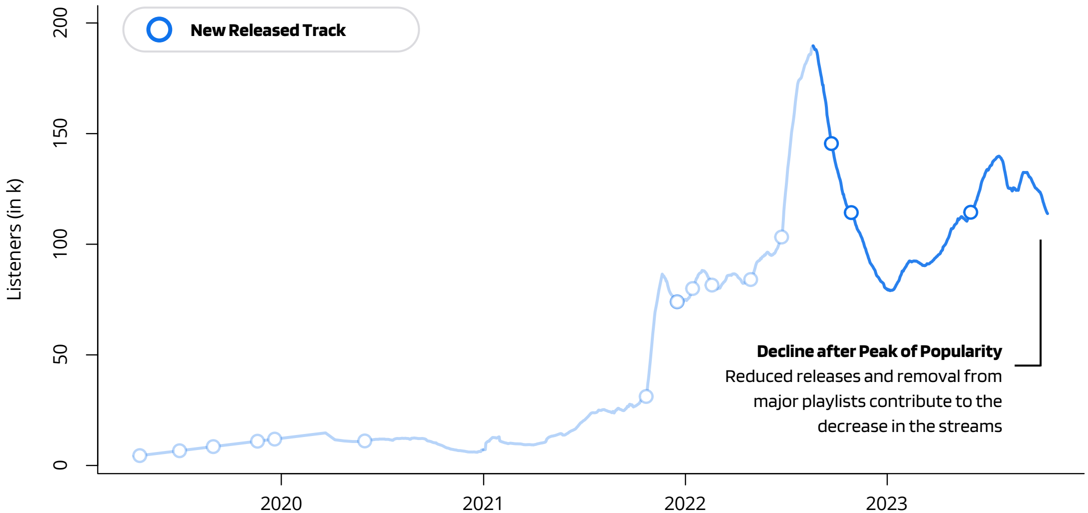


Playlist Driven Growth
Getting added to more playlists and releasing a hit song for the first time make the number of streams go up a lot

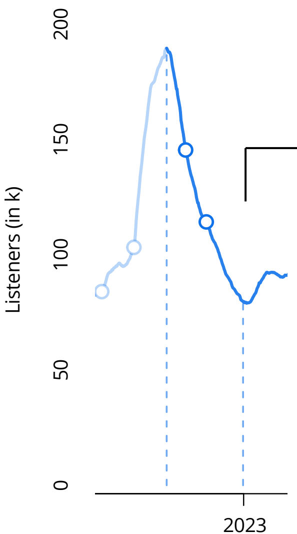
Overview of the artist's activity



Overview of the artist's activity

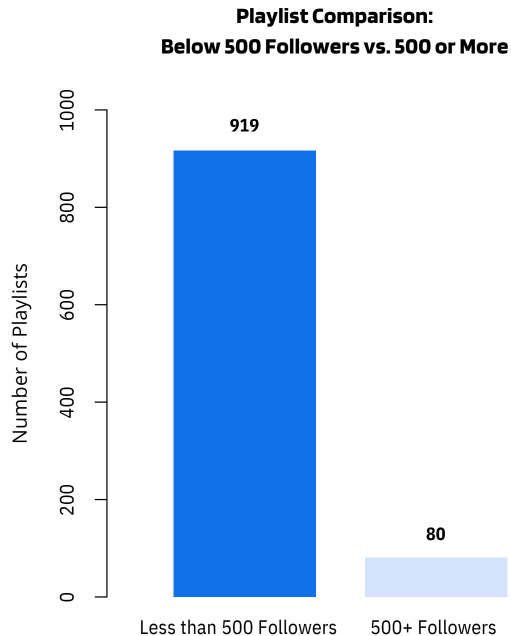


Overview of the artist's activity



The artist's tracks were **removed** from 15 of the **most popular** playlists.

Approximately half of these playlists have more than **10k** followers.



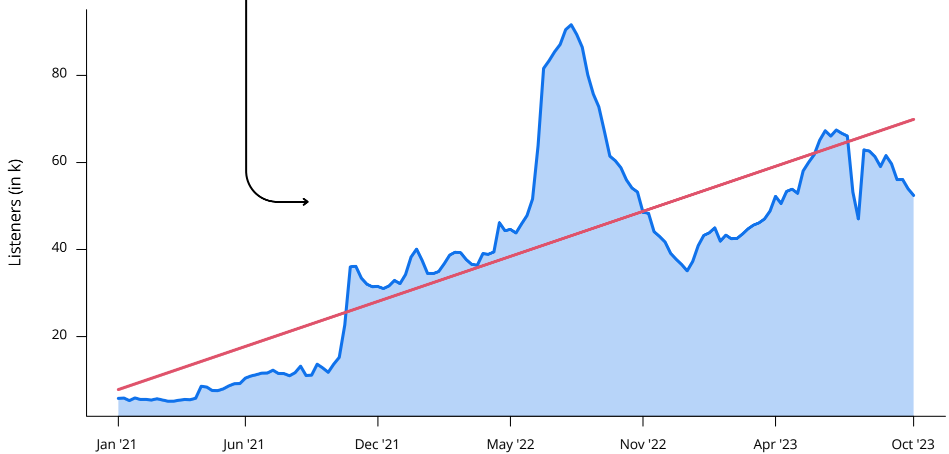
Modeling Listeners

Time Series Linear Model

tslm(listeners ~ trend)

TSLM, simple but effective, helps in capturing the **linear increasing trend**

Adj R^2 = 0.6148



Time Series Linear Model

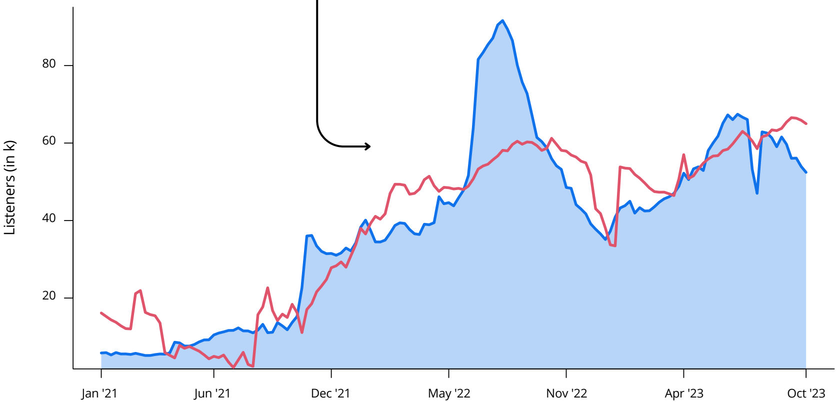
tslm(listeners ~ trend + plst.count + avg.reach)

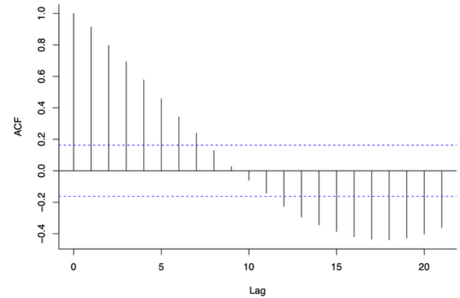
By adding two explanatory variables, the model succeeds in capturing the overall dynamic more effectively.

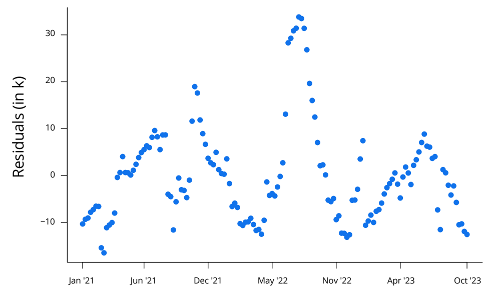
Adj R^2 = 0.7875

The residuals appear to be **positively autocorrelated**.

ARIMA may be employed to address non-stationarity and to capture other complex patterns.





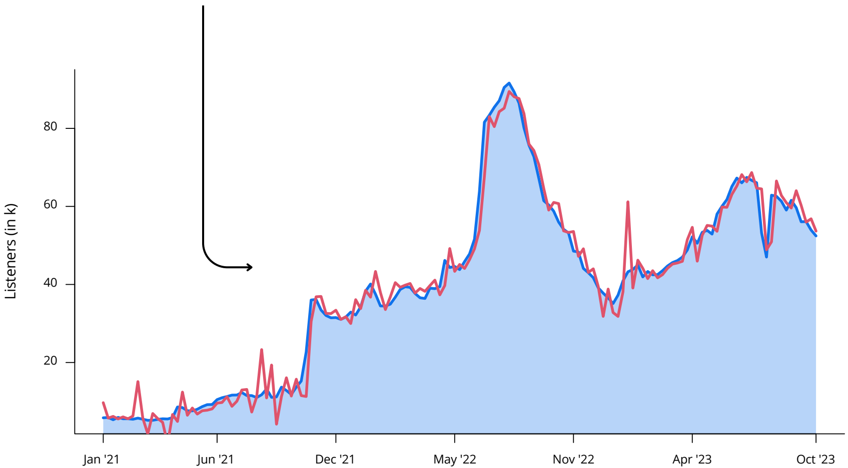


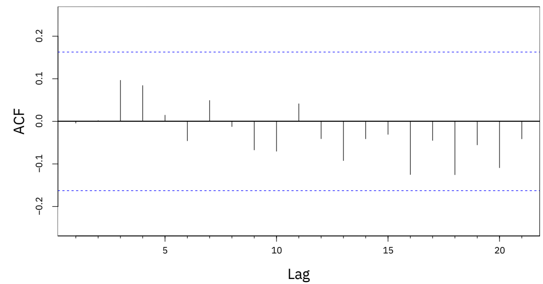
TSLM + ARIMA

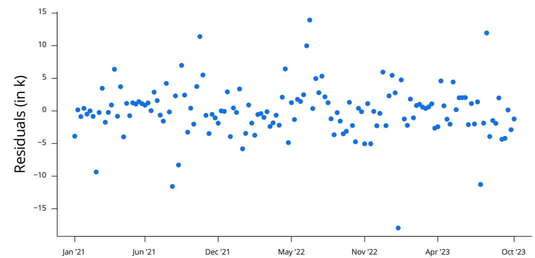
tslm(listeners ~ trend + plst.count + avg.reach) + ARIMA(1,0,1)

Now, the model fits better to the variability of the data

AIC = 2821





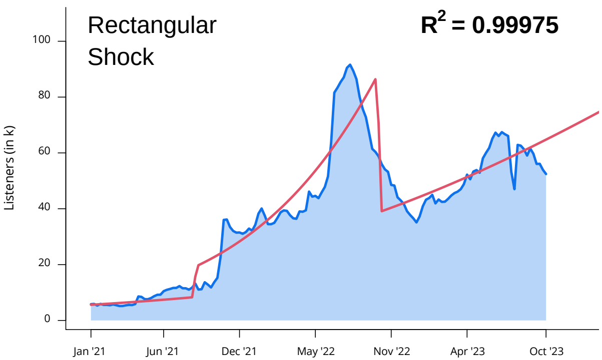


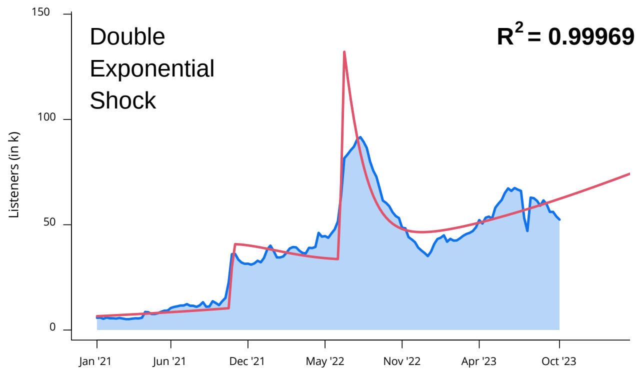
Generalized Bass Model

Rectangular shock vs double exponential shock

Both models effectively capture the data, highlighting the **growth trend** and the significant **spikes** associated with the **artist's top two singles**, causing a brief exponential surge in listeners.

To choose the most suitable model between the two, we use the **R squared** as a criterion, indicating that the model with the **rectangular shock is preferred**.

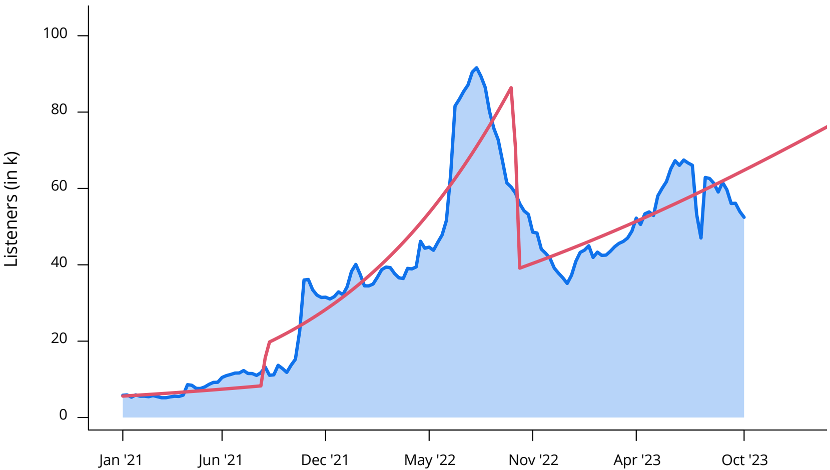


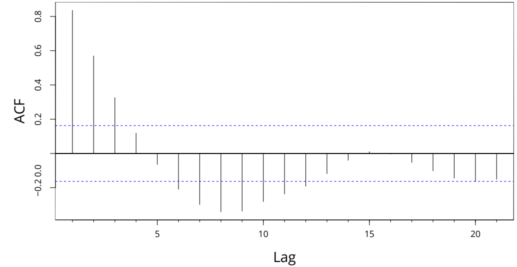


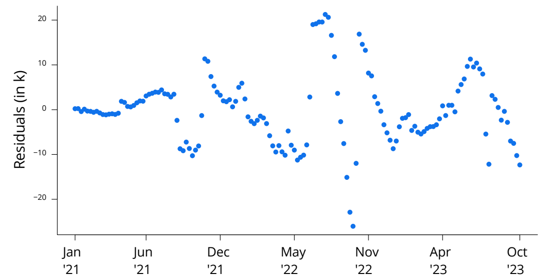
Generalized Bass Model

Rectangular shock

Observing the residuals, we notice the presence of **autocorrelation**. This situation can be addressed with a **refinement** using **ARIMA**.



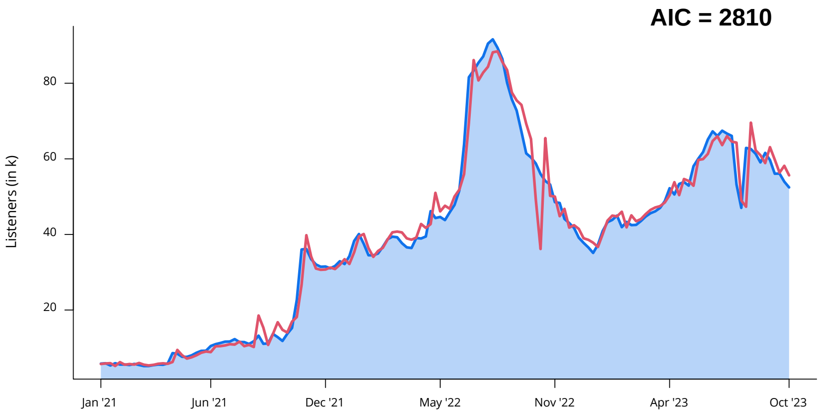


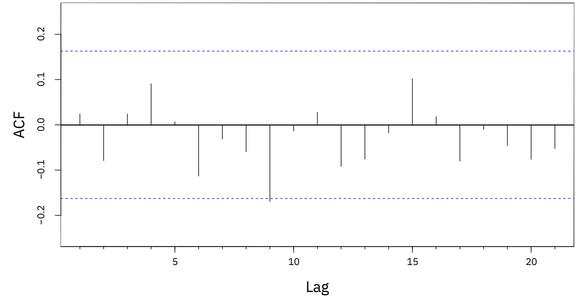


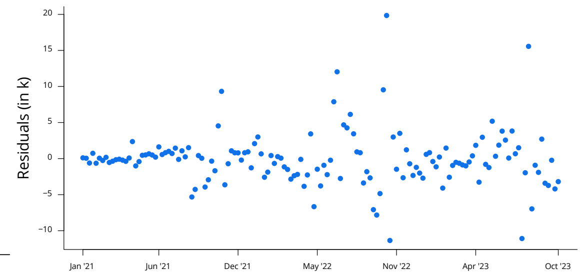
GBM + ARIMA

Rectangular shock

- The model fits well to the data and can be used for predictions.
- The residuals do not exhibit autocorrelation.



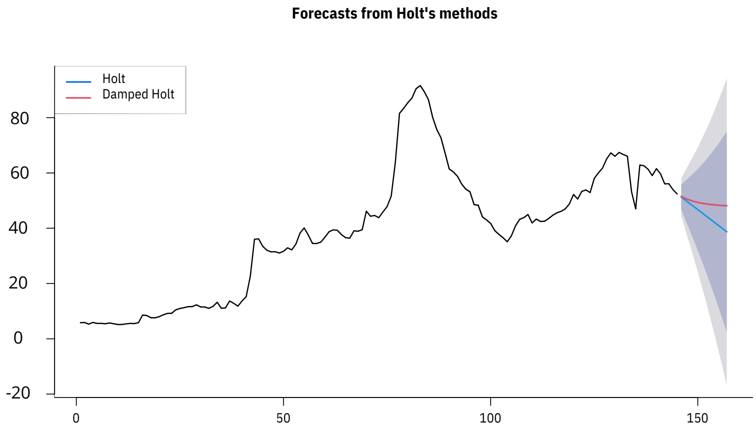


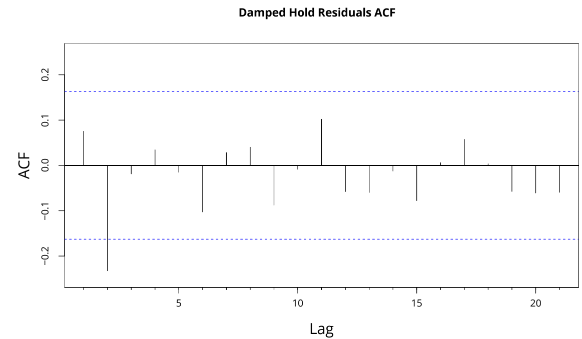


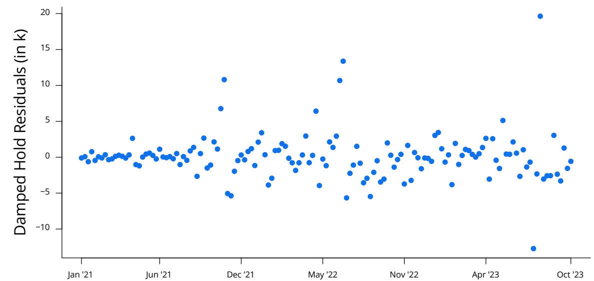
Holt's methods

Holt vs Damped Holt

In comparing the two methods, Damped Holt appears to be preferable as it has a lower AIC of **3082**, while Holt has an AIC of **3092**.







Comparing **models**

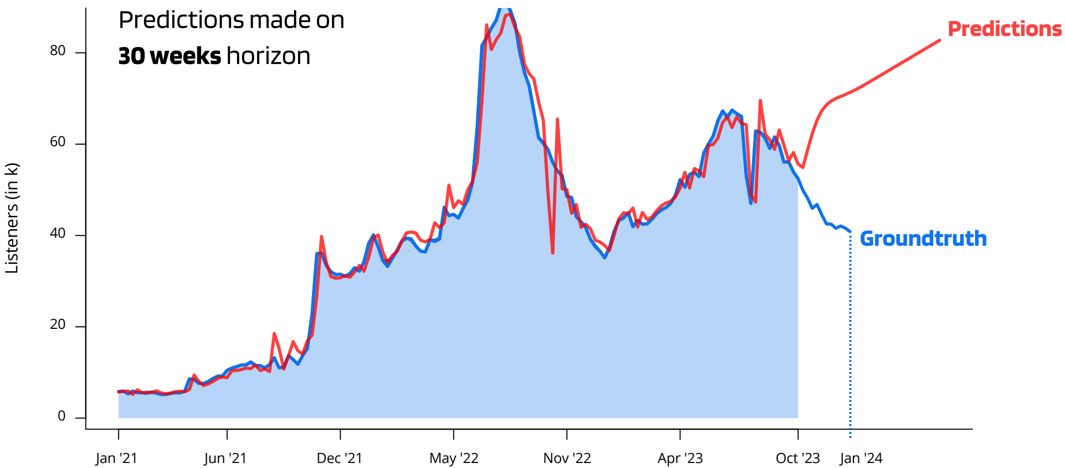
Model	AIC	RMSE	MAE	MAPE
TSLM + ARIMA	2821	3948	2675	113
GBM + ARIMA	**2810**	3807	2368	68.99
Damped Holt	3082	**3287**	**1919**	**5.52**

Based on the metrics mentioned above, the best models appear to be **GBM + ARIMA** in terms of AIC, and **Dumped Holt** in terms of error.

Both models fit well with the data and can be considered for making predictions about the artist's future.

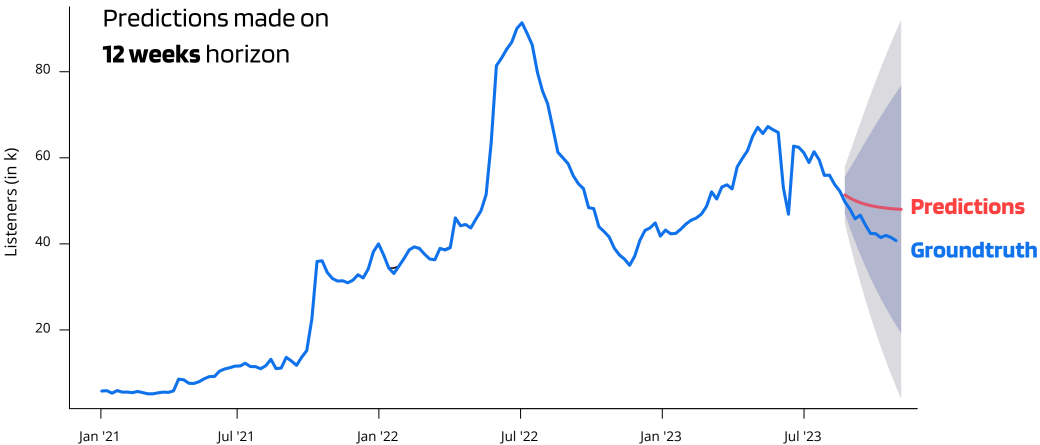
GBM + ARIMA: predictions

This model is interesting due to its **interpretability**. In particular, it is appreciated for its ability to **capture** and describe the **increasing trend** that has characterized the development of the artist's career from their early publications to the present day. For this reason, we consider it a suitable model for **medium to long-term** predictions.



Damped Holt: predictions

Exponential smoothing, by definition, is a model suitable for short-term forecasts, as it assigns increasing weights to the most recent data, placing greater **importance** on **recent observations** compared to older ones.
For this reason, it is worthwhile to assess the **short-term** predictions of this model.



Conclusions.

Final considerations

- To sum up, the analysis conducted highlights a **growth trend** up to today.
- Its dependence on private playlists makes the **streams particularly volatile**.
- The fact that in the last year he has **published less frequently and regularly** has had a negative effect on the streams. This is also evident from the short-term forecasts.
- The analysis indicates **potential for long-term growth**, heavily reliant on his release frequency, song quantity, promotion strategy, number of collaborations and effort in identity building.

Future works

- In addition to streams and presence on private playlists, it is crucial to **gain recognition**.
- It's important to carefully **promote** the tracks, be active on **social media** and **collaborate** with other artists.
- Adding such data in the future would be appropriate to further understand the dynamics related to the **growth** of an emerging artist.

Thank you for your attention