

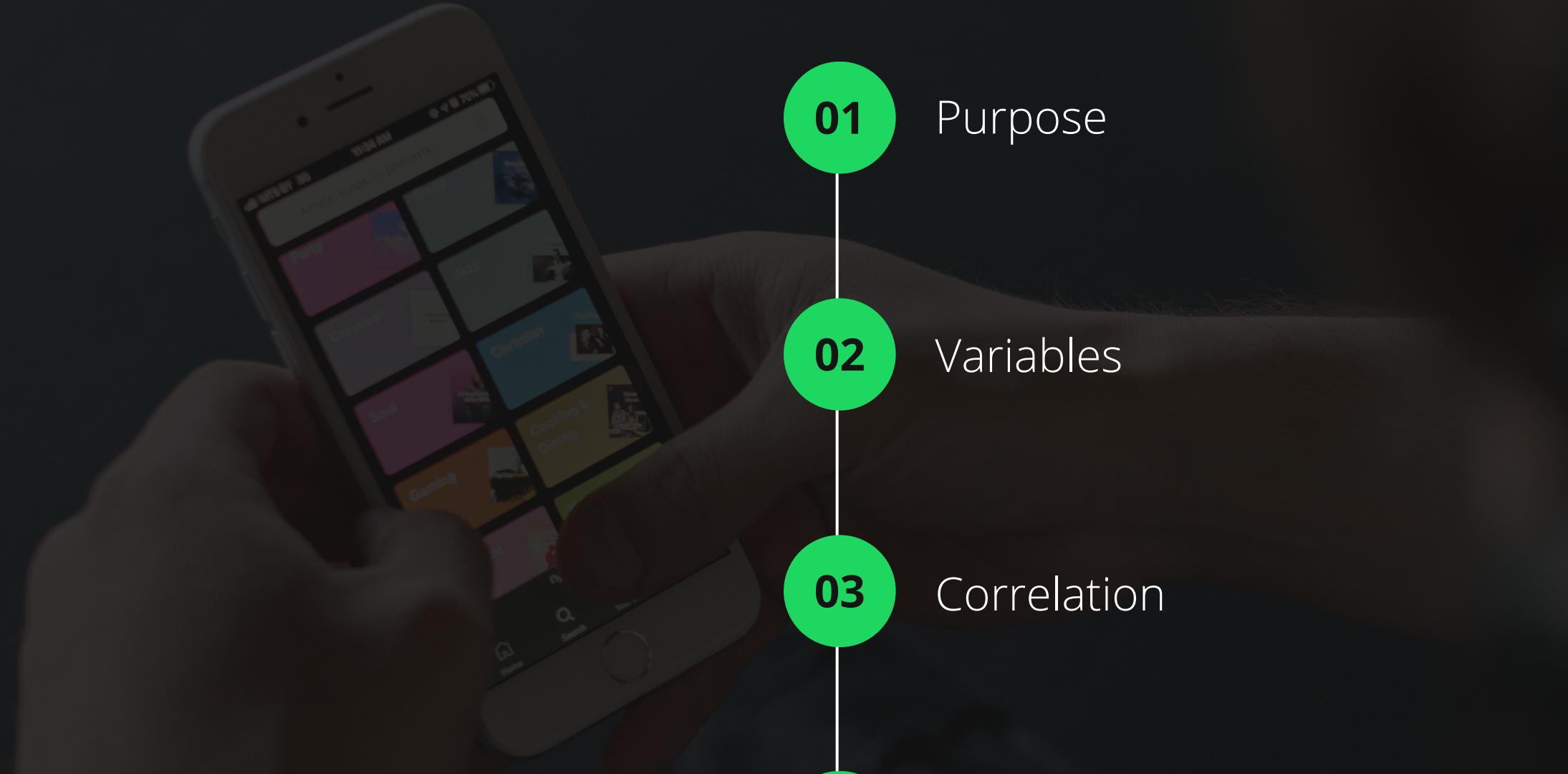


SPOTIFY

In Depth User Analysis
Prof. Thomas Kurnicki

Forecasting & Predicting the Future Using Data
Team 4 - Alyssa Brunen, Catarina Jesus, Albin Möller, Jinning Yan

AGENDA

- 
- 01 Purpose
 - 02 Variables
 - 03 Correlation
 - 04 Regression & GINI Tree
 - 05 Forecasting with ARIMA & GARCH



Purpose

WHY ARE USER ANALYTICS IMPORTANT?

Spotify prides itself on its users, artists and variety. The business is successful, but we need to analyze current data and forecast future possibilities to ensure future success.

This report analyzes the future path for Spotify and provides music creators insights on consumer preference.

MAIN VARIABLES TO CONSIDER

What makes a song popular?



Loudness



Danceability



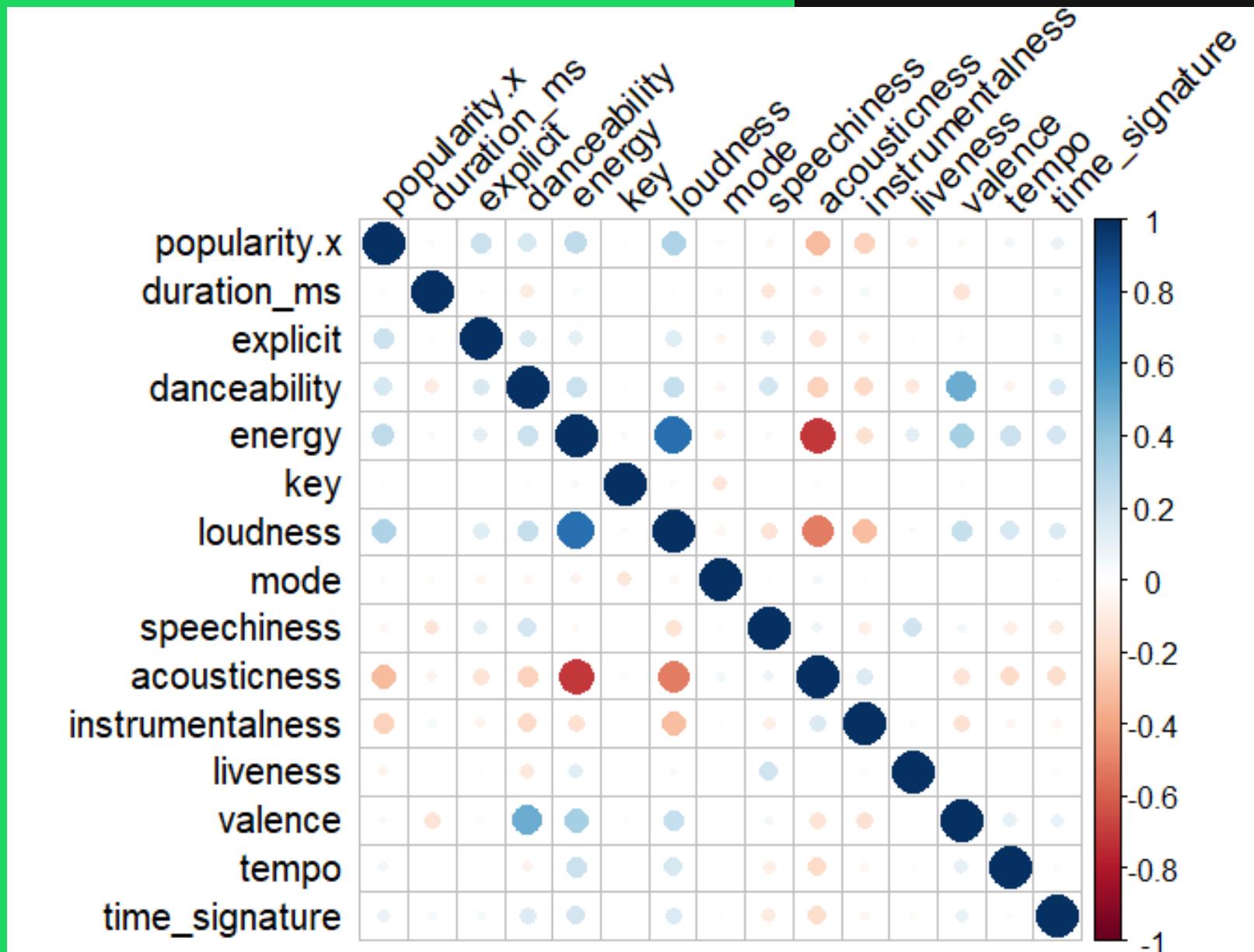
Tempo



COMING UP:

BUSINESS INSIGHT

CORRELATION MATRIX



Evaluation of the strength and direction of relationships between variables, aiding in identifying potential multicollinearity and selecting appropriate predictors.



"Loudness" and "Energy" share a strong positive correlation with coefficient of 0.756.



Consequently, variable "Energy" will be excluded in further analysis to avoid multicollinearity.

<i>Variable (normalized)</i>	<i>Exp(Estimate)-1</i>
loudness_norm	4540.034
danceability_norm	9.596778
explicit_norm	2.542597
tempo_norm	0.5617859
acousticness_norm	-0.3546678
instrumentalness_norm	-0.5145748
liveness_norm	-0.5381892
speechiness_norm	-0.6997711
valence_norm	-0.8298202
duration_ms_norm	-0.9929523

REGRESSION MODEL

(normalised data)

Normalization of data has been employed in order to scale the features to a consistent range, preventing certain variables from dominating the model and ensuring fair comparisons between different predictors.

REGRESSION MODEL

(non-normalised data)

*For every additional unit of **danceability**, the odds of business success increase by 1000%*

*For every additional unit of **followers**, the odds of business success increase by 900%*

*For every additional unit of **explicit**, the odds of business success increase by 231%*

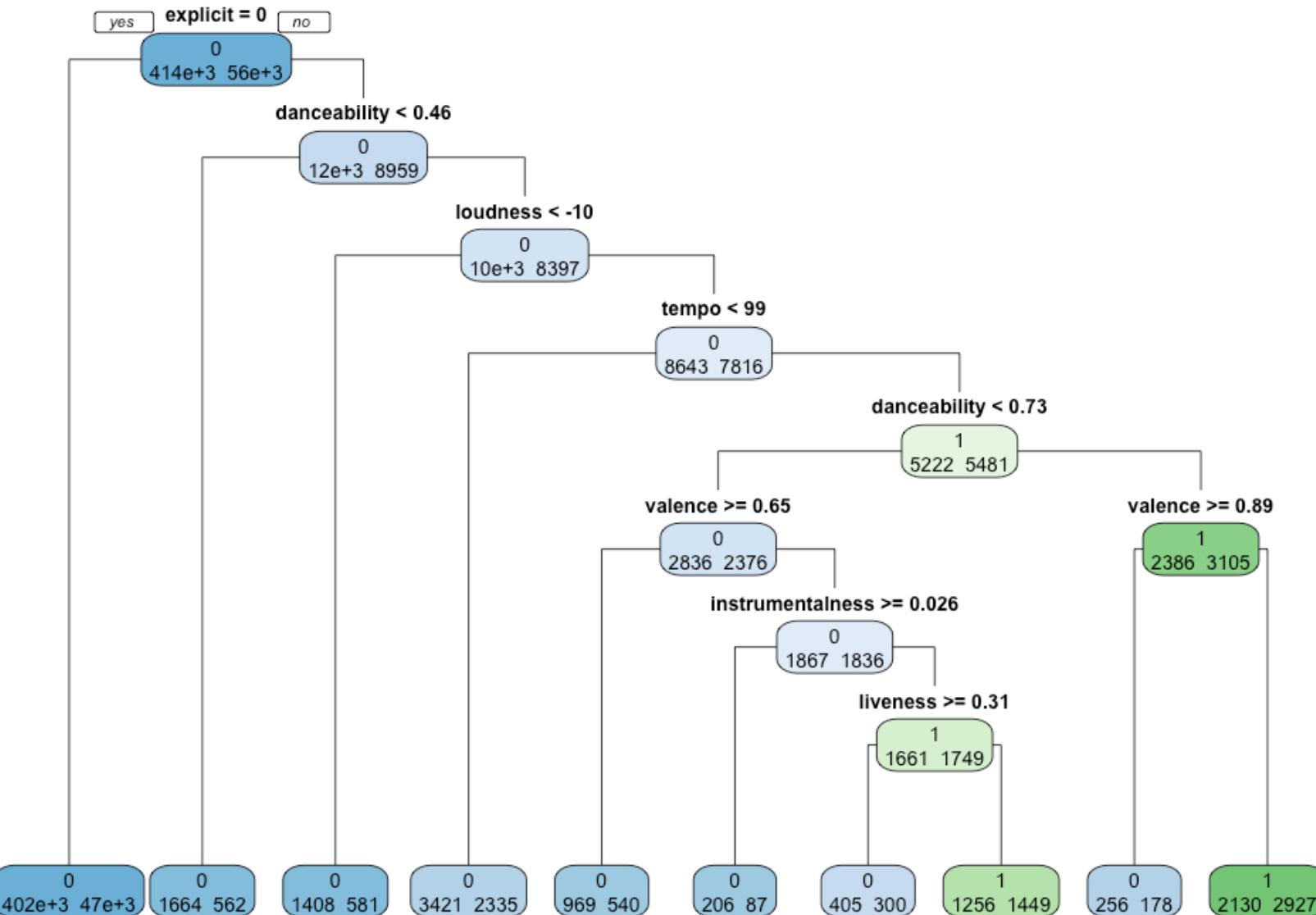
*For every additional unit of **loudness**, the odds of business success increase by 12%*

*For every additional unit of **duration**, the odds of business success decrease by 815%*

*For every additional unit of **valence**, the odds of business success decrease by 82%*

*For every additional unit of **speechiness**, the odds of business success decrease by 79%*

GINI DECISION TREE



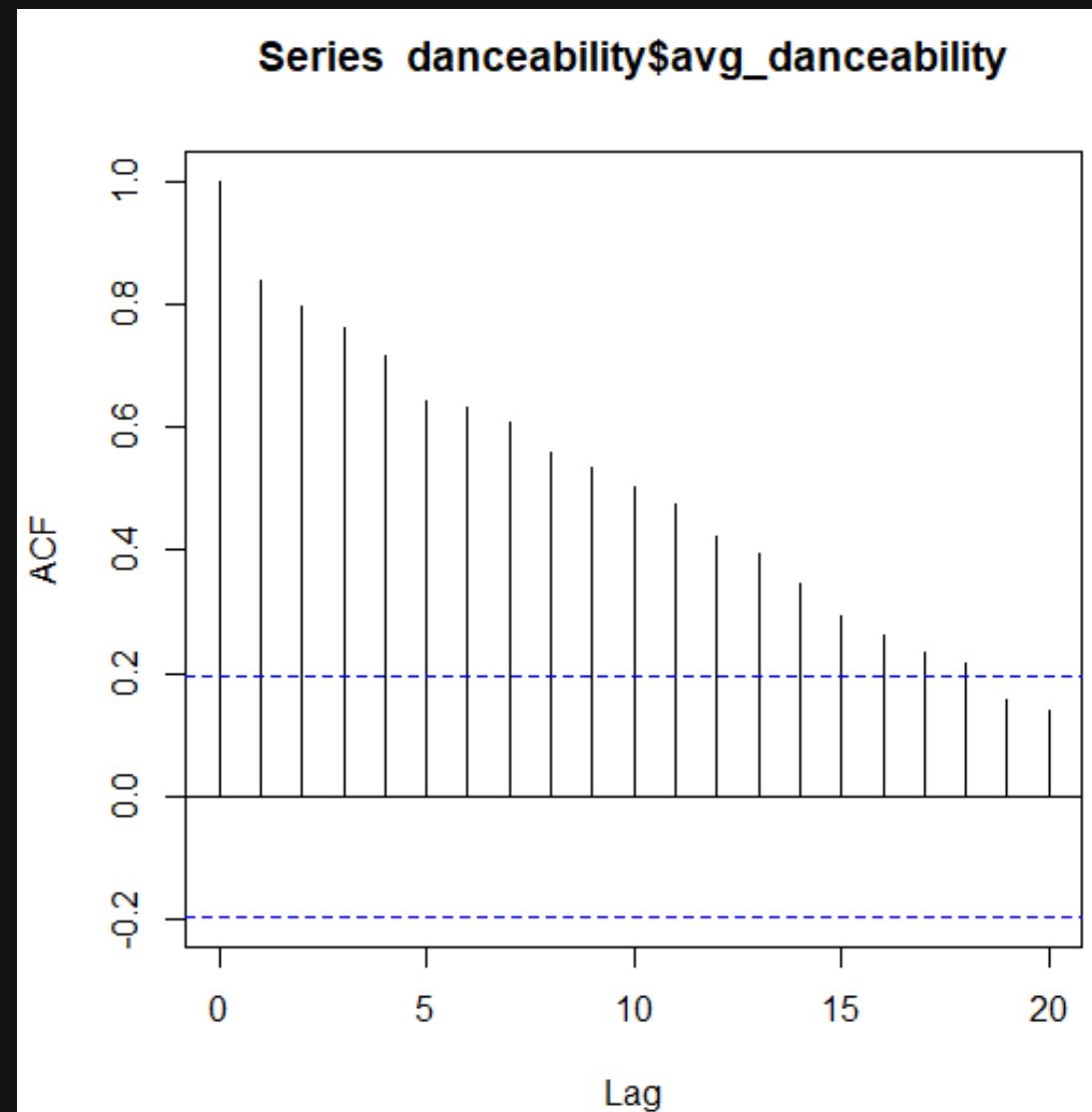
Accuracy: 86.08%

Feature Importance: Explicit > Danceability > Loudness > Valence > Tempo > Duration

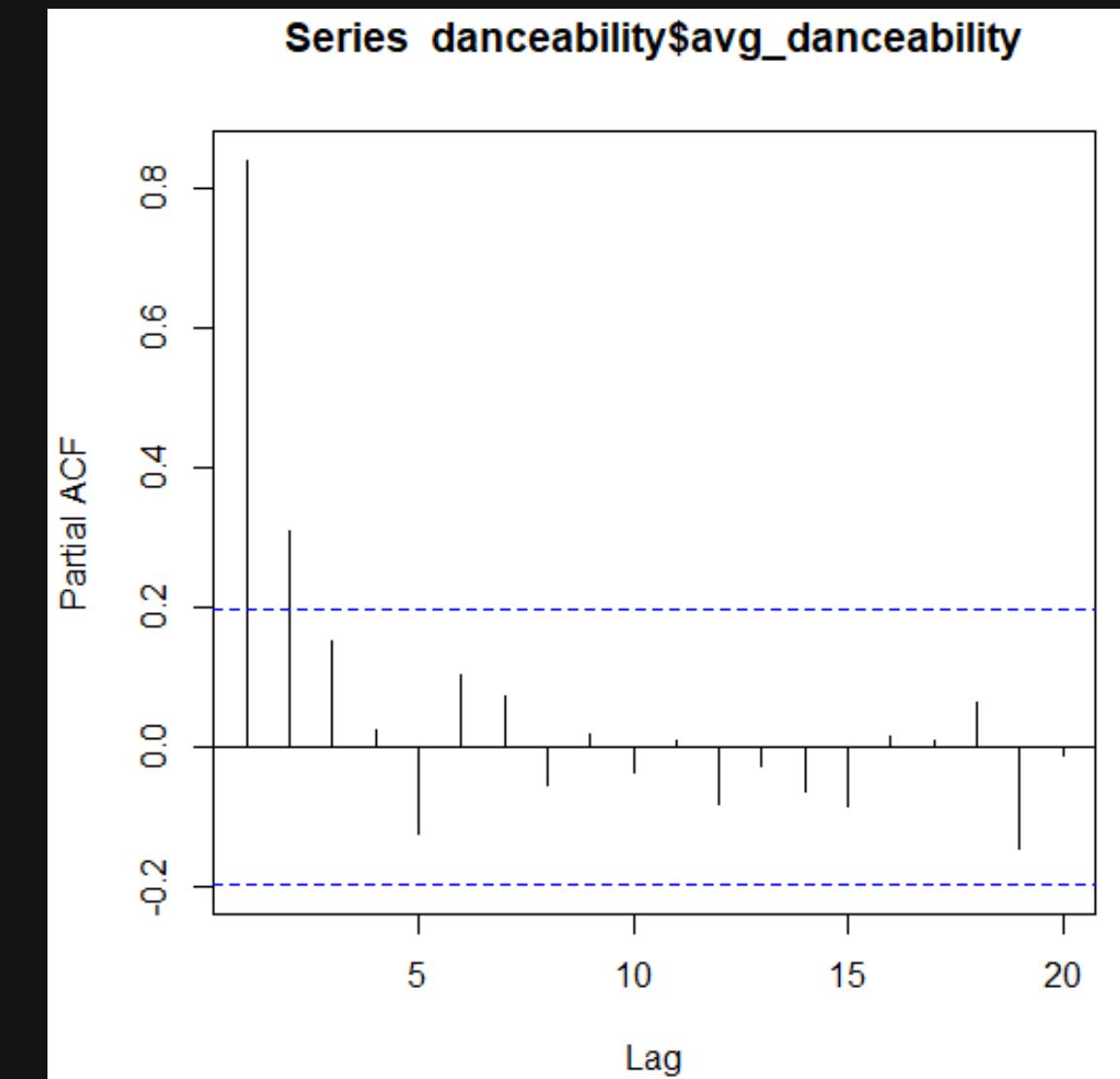
Example: Will this song be popular?

- Explicit: 1
- Danceability: 0.5
- Loudness: -8
- Valence: 0.7
- Tempo: 110
- Duration_ms: 200,000

INSIGHT 2: DANCEABILITY ARIMA



Danceability ACF



Danceability pACF

ARIMA Forecasting for the next 5 years: 0.6689181, 0.6761957, 0.6721995, 0.6941535, 0.6851626

Level of uncertainty: 2.16%, 2.26%, 2.53%, 2.59%, 3.28%

HOW DO WE USE THESE PREDICTIONS?

1

What does it tell us? **Trend!**

It tells us the predicted danceability values for each respective year in the future for 5 years.

2

Business Decisions:

01

Product Development:

Predicted values can guide product development strategies.

02

Marketing:

Targeted campaigns and advertisements that highlight danceability.

03

Business Differentiation:

It informs companies and creators on what type of music will be more popular, and pivot early.

3

Errors:

Inaccuracy of the forecast
Possible reasons:

01

Possible seasonal or periodic patterns.

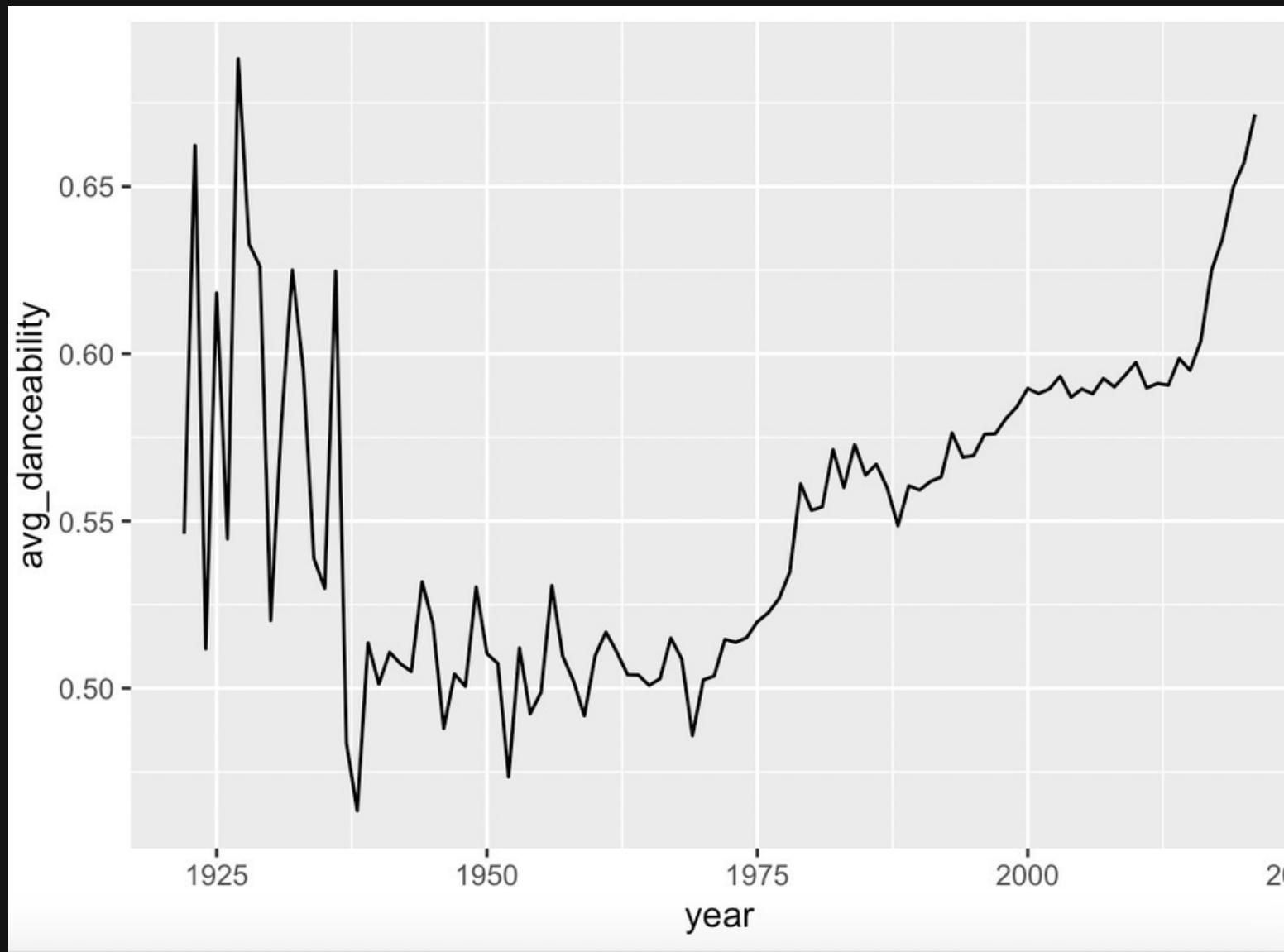
02

Variability of danceability factor
Related businesses should be adaptive and adjust when time arrive.

03

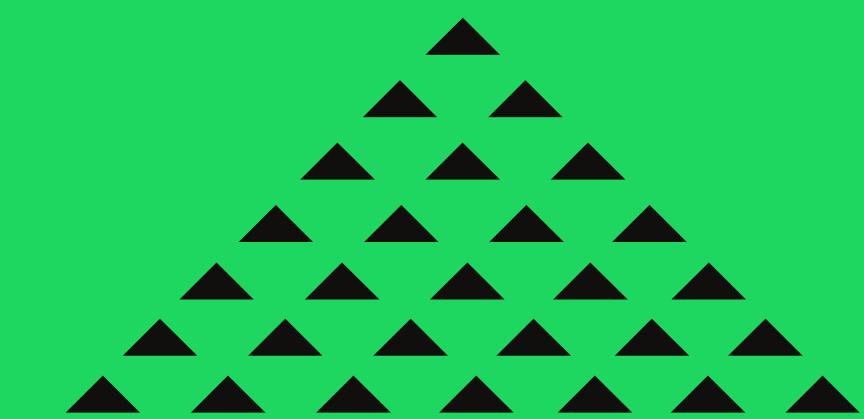
Data quality issue.

INSIGHT 3: DANCEABILITY GARCH



GARCH Model:

- $t+1$: 0.077818
- $t+2$: 0.078014
- $t+3$: 0.078210
- $t+4$: 0.078405
- $t+5$: 0.078600



Volatility & Stability

According to the GARCH model, the future volatility of danceability in music is quite low. Between 7.78% to 7.86%.

01

Risk:

After the GARCH analysis, businesses can understand uncertainty and variability in future music success and popularity.

02

Decision Making:

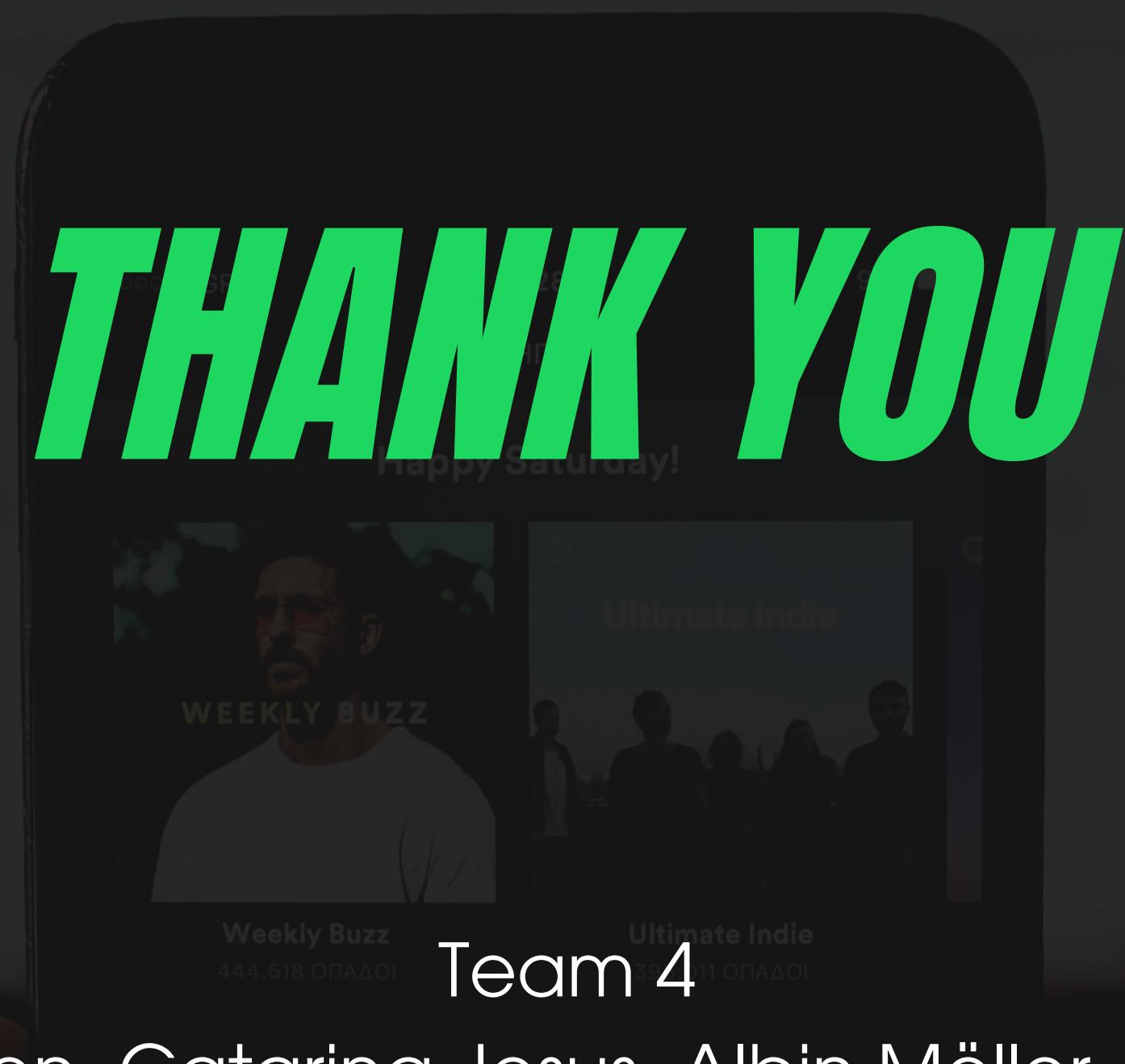
Informed choices for investments and marketing strategies.

03

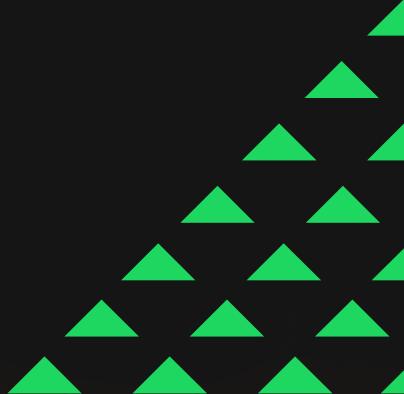
OTHER ANALYSIS

We also created ARIMA forecasting and GARCH models for other variables, such as loudness and tempo.

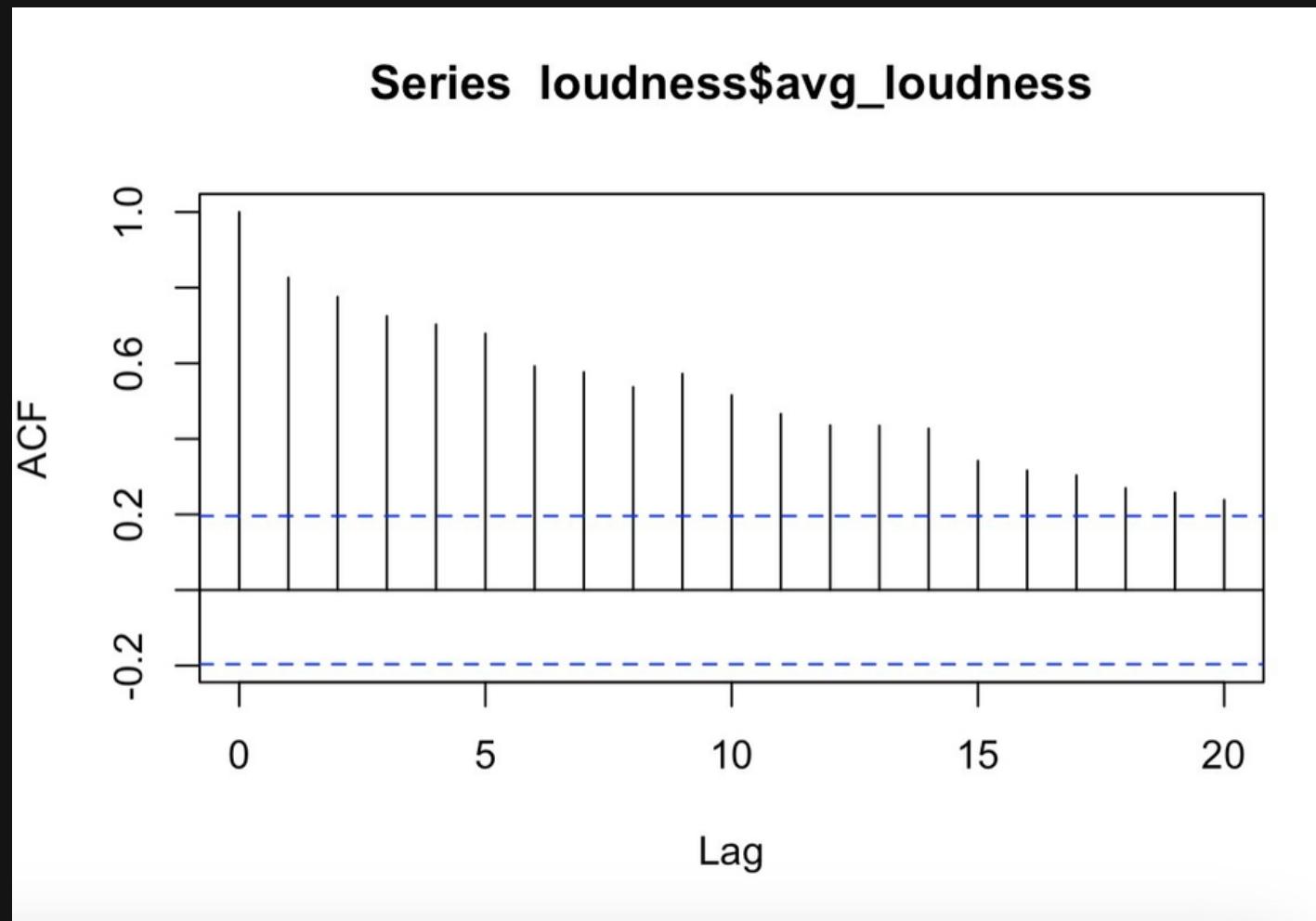
Please refer to appendix and R script.



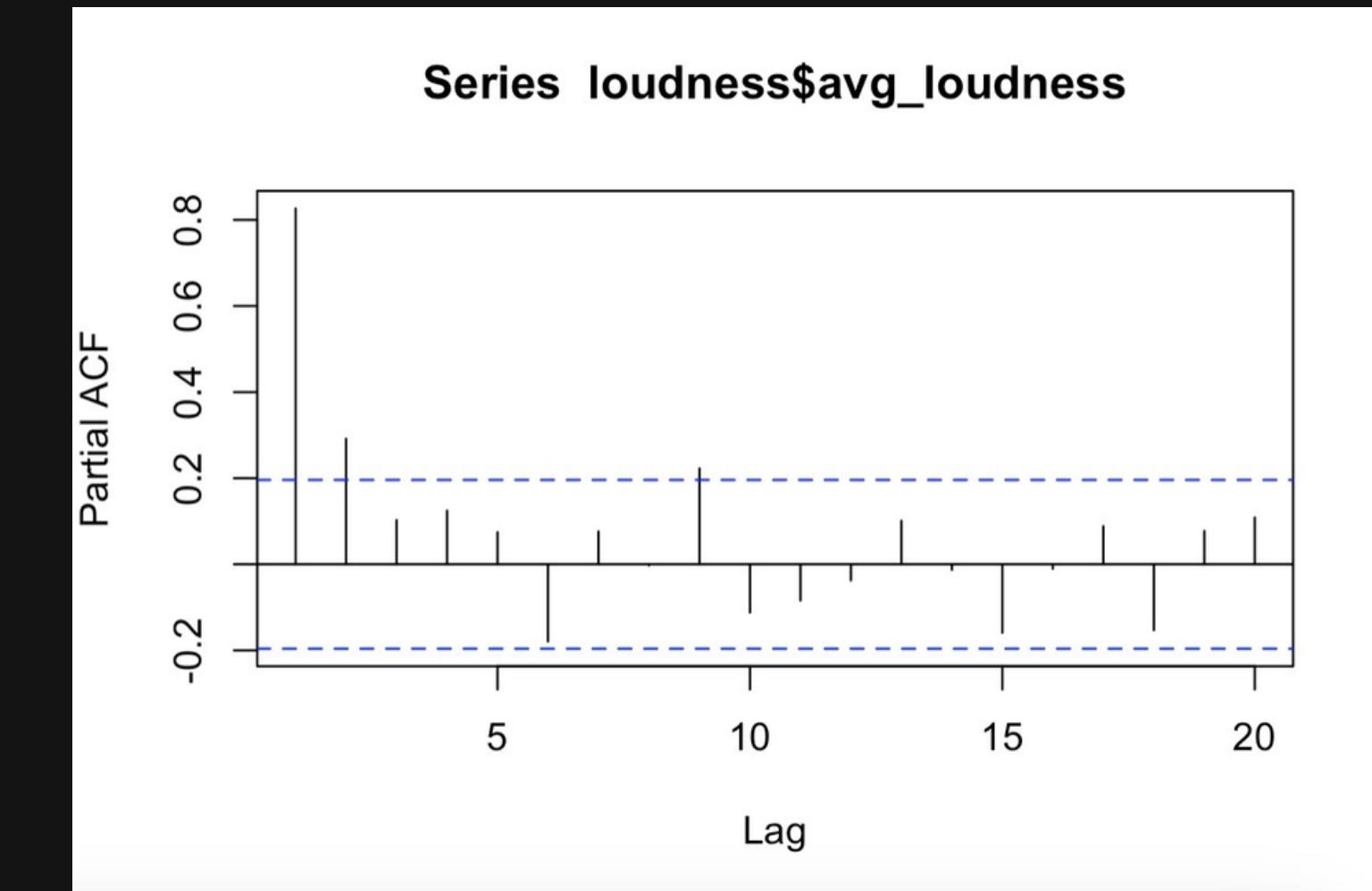
Alyssa Brunen, Catarina Jesus, Albin Möller, Jinning Yan



APPENDIX 1: LOUDNESS ARIMA



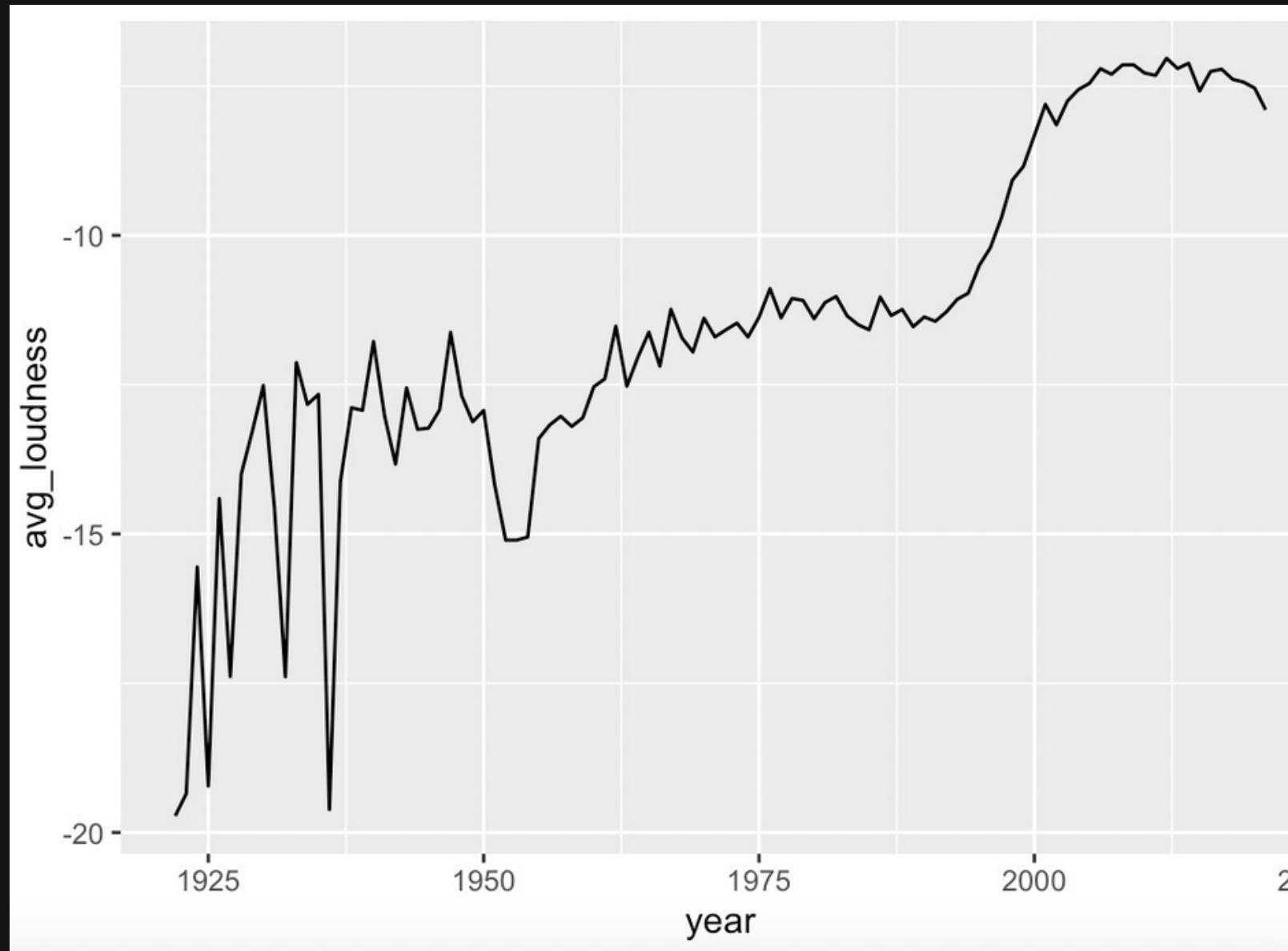
Loudness ACF



Loudness pACF

*ARIMA Forecasting for the next 5 years: -8.574335 -8.155909 -8.814771 -9.463423 -9.092720
Level of uncertainty: 0.9224355 1.0329742 1.1331575 1.1865815 1.4306653*

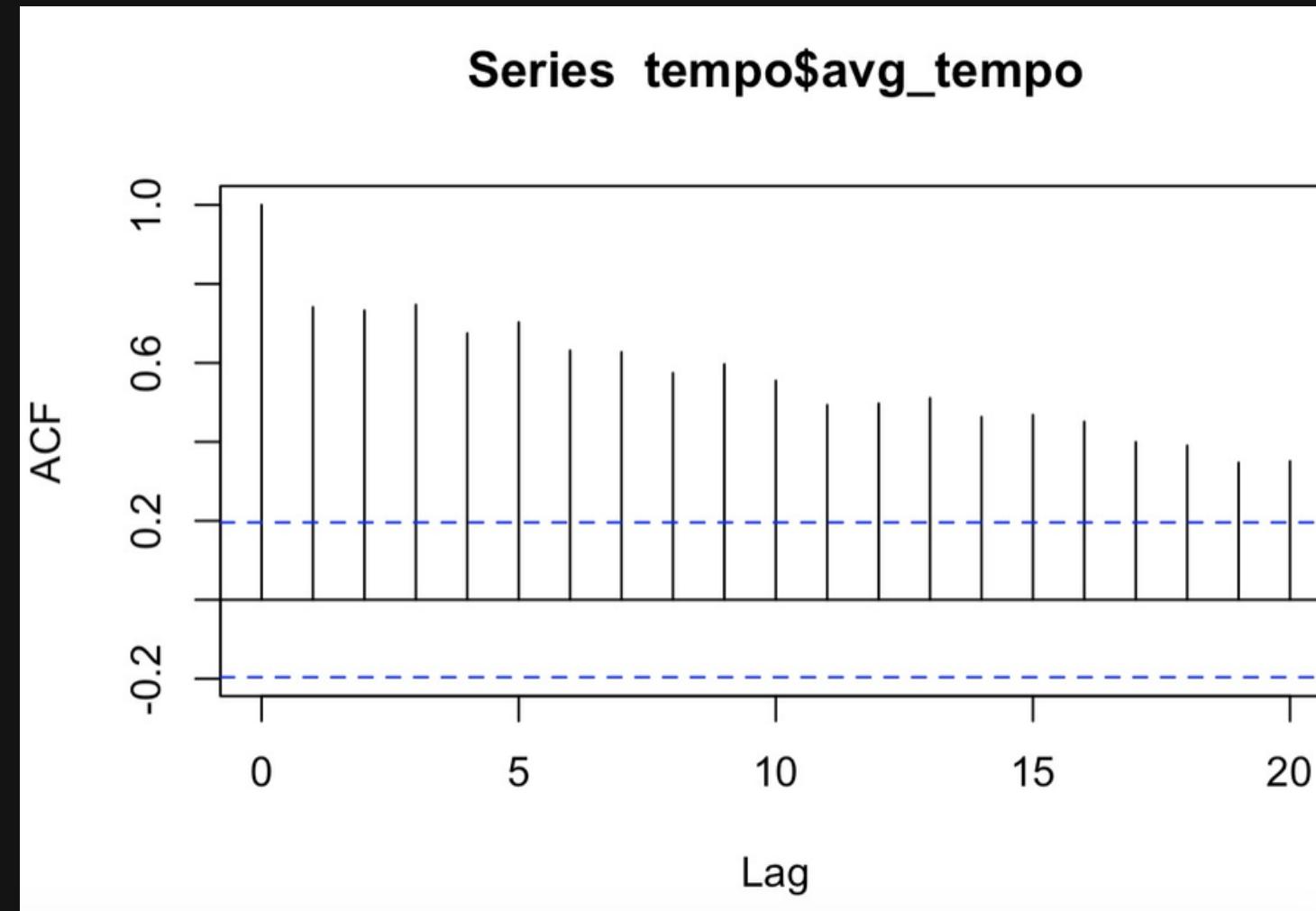
APPENDIX 2: LOUDNESS GARCH



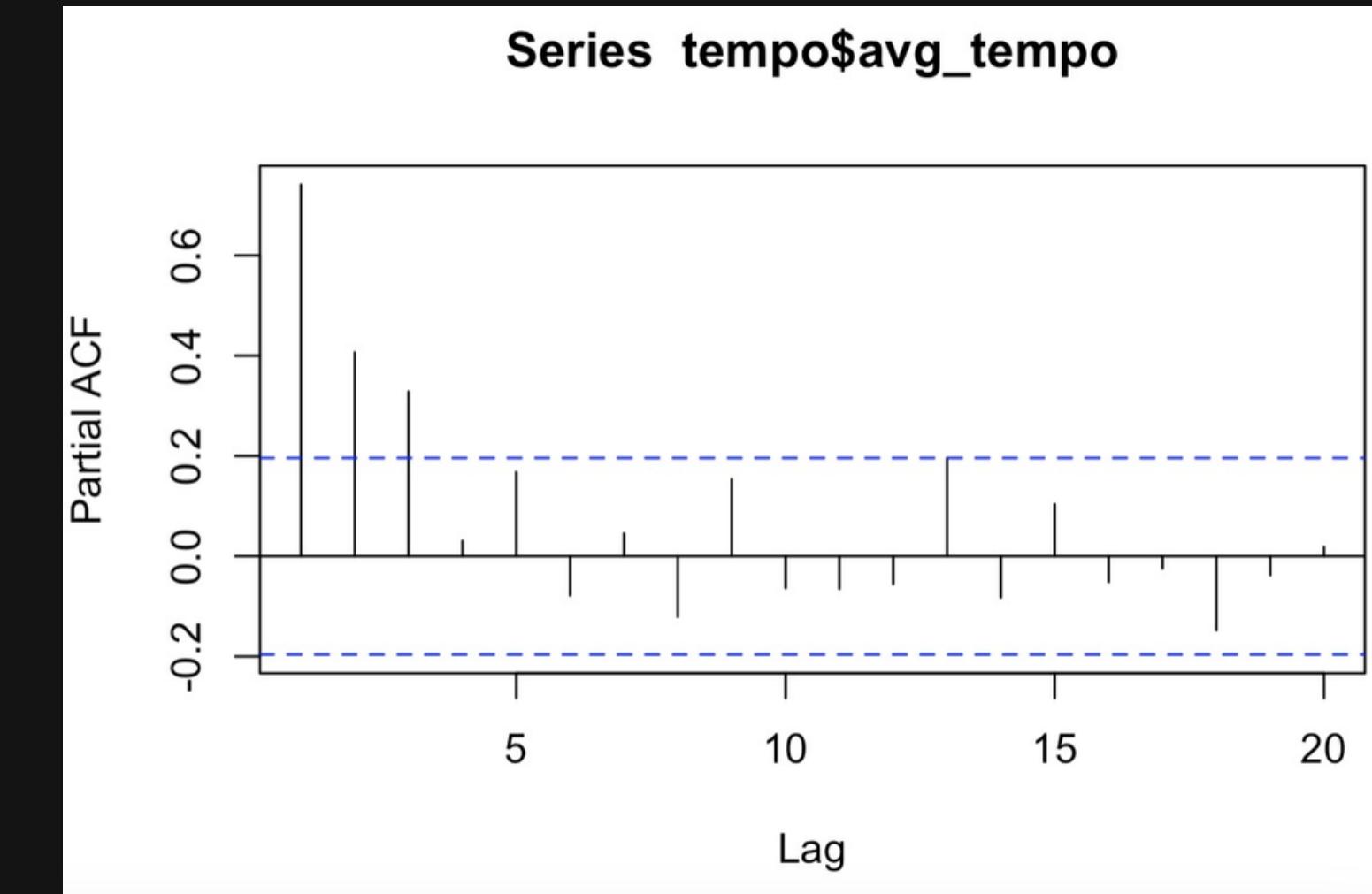
GARCH Model loudness:

$t+1$: 3.0506
 $t+2$: 3.0508
 $t+3$: 3.0511
 $t+4$: 3.0514
 $t+5$: 3.0517

APPENDIX 3: TEMPO ARIMA



Tempo ACF

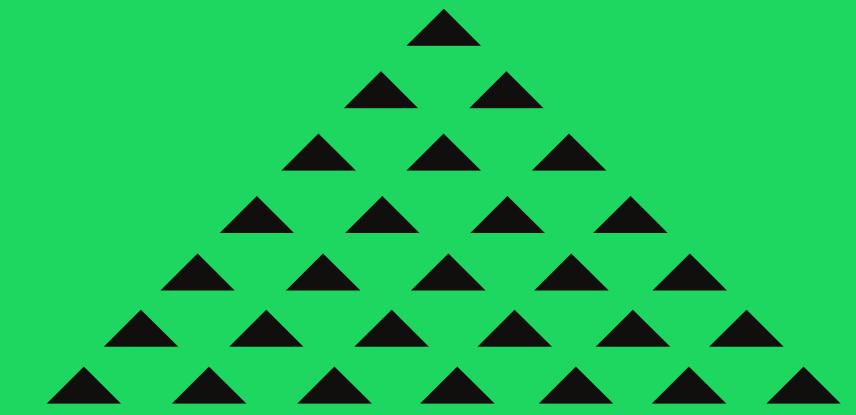
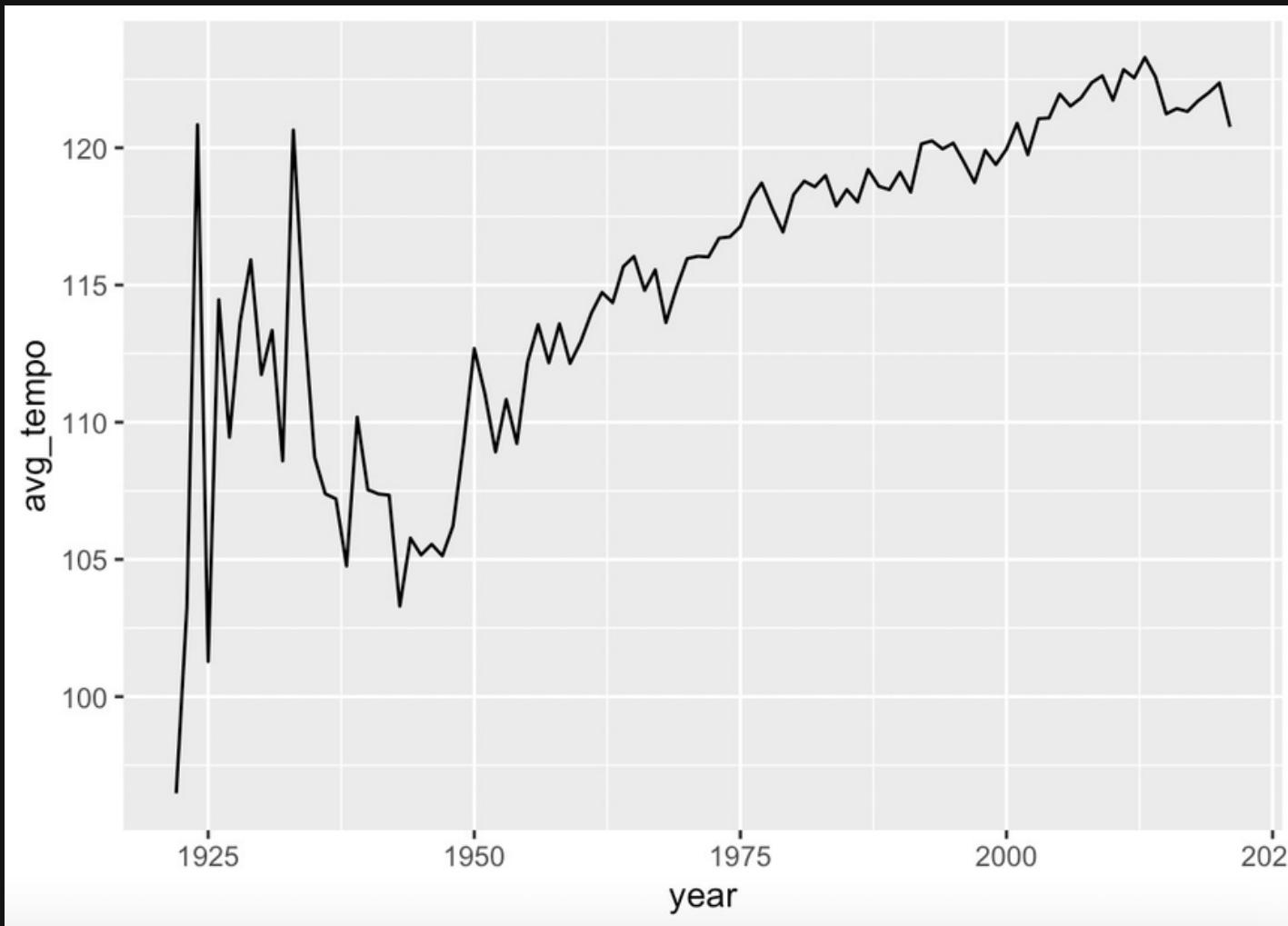


Tempo pACF

ARIMA Forecasting for the next 5 years: 121.5818 119.0877 121.1205 119.7098 118.4814

Level of uncertainty: 1.951986 2.072900 2.409080 2.826541 2.886078

APPENDIX 4: TEMPO GARCH



GARCH Model tempo:

$t+1$: 2.0140
 $t+2$: 2.0090
 $t+3$: 2.0042
 $t+4$: 1.9997
 $t+5$: 1.9955