

Investigating Trend Longevity of Popular Music

Ezra Passalacqua

12/13/2024

1 Overview

The Billboard Hot 100 is a weekly ranking of popular music in the United States, combining record sales, radio play, and, more recently, streaming, to create a general metric of the relative popularity of music [2]. Billboard claims that the data represents over 90% of sales (tracked by the third-party tool Luminate), all streams on major streaming services, and radio play on the majority of major radio stations (categorized by the Nielsen Audio Ratings system, which weights airplay differently based on time of day). All of these statistics are combined using a proprietary system to obtain a ranking for each song. The charts are released every Tuesday online and Saturday in print (our data reflects the print publication dates, which have changed over time, but are consistent to one per week). Each chart covers the Friday two weeks prior through the previous Thursday [2].

As the landscape of popular music continues to shift, claims about popular music trends continue to surface. The purpose of this analysis is to provide some clarity on two main questions. Primarily, we will analyze the “speed” of the Hot 100’s progression. We wonder: is there an observable trend in the average amount of time songs spend on the Hot 100? By graphing and decomposing the Hot 100 time series, we are able to look for a general trend. We also are curious: is there any truth to common pop culture terms like “song of the summer”? Does music popularity shift throughout the year? We will observe any seasonal cycles to gain some clarity on this. Lastly, to answer if this behavior can be modeled, we will test multiple time series models using both held out data and cross-validation to obtain a model for predicting future Hot 100 outcomes that accounts for both general trends and any present cycles.

2 Data

2.1 Data Set Description

This data set is a time series, with one outcome, “average”.

The original data was obtained from the git repository linked below [1]. Because the machine running this analysis was insufficient to run a large scale data set, I have pulled a collection of all Hot 100 data since January 3, 1970 through the publication on December 29, 1979. This totals 522 weeks of data.

As described above, each date in the original data is associated with the top 100 songs of a given week. Though the written date of each chart can vary within the week over the data set, each date represents the data from the previous week, so we consider the data to be regularly spaced, and fit for a time series.

To analyze the general longevity of the music on any given chart, I calculated a custom metric to be the sole outcome variable. Each song on the Hot 100 has a measure of how many weeks it has spent on the chart, including the current week. The “average” metric I have calculated averages the number of weeks spent on the chart between all 100 songs for each given week. Though it is not perfect, this gives a single number that represents how long the average song has spent on the chart.

There are no missing values in the series so nothing need be imputed.

2.2 Summary Statistics for Average

Table 1: Summary Statistics for Average Length of Time on Hot 100

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
5.09	6.29	6.98	7.027	7.628	9.86

2.3 Plotting Data

2.3.1 Box Plot

We can see by looking at this box plot that there is a singular outlier above, and the data is otherwise slightly right skewed. I chose to ignore this outlier, as it is unlikely to incentivise large changes in the model predictions due to the size of the data set.

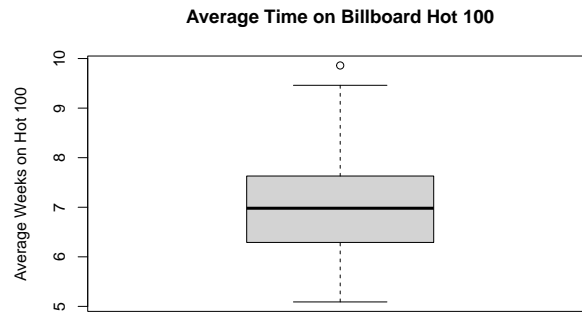


Figure 1: Box Plot of Average Times on Chart Across Billboard Hot 100

2.3.2 Histogram

Viewing the histogram, this right skewed distribution becomes even more apparent.

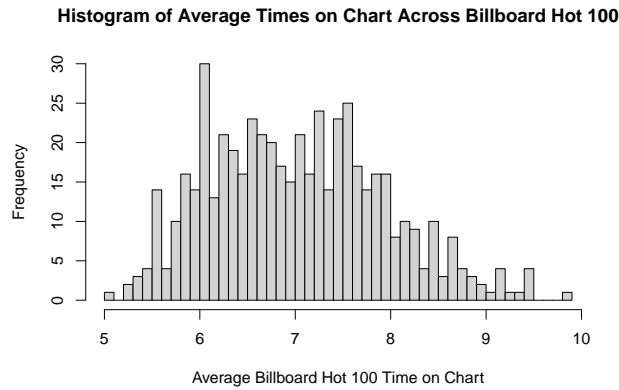


Figure 2: Histogram of Average Times on Chart Across Billboard Hot 100

2.3.3 Time Series Graph

Observing this time series shows a general increase as time goes on, as well as possibly seasonal cycles, but no trend cycles. However, there is also a lot of noise that obscures these observations.

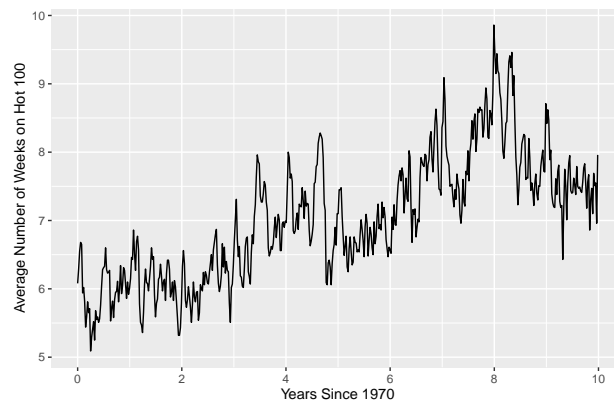


Figure 3: Average Number of Weeks on Hot 100 by Week

2.3.4 Time Series Decomposition

The decomposition of the time series clearly shows the upward trend (which is additive) and a yearly seasonal cycle, peaking around the new year, around early January. Notably, there is also an observable trend cycle with an approximate 3 year period, beginning in 1962. However, this cycle is only observed twice, so the meaning is debatable.

To better observe the seasonal cycle, a plot of the seasonal cycle from the first 3 years of the data set is shown below:

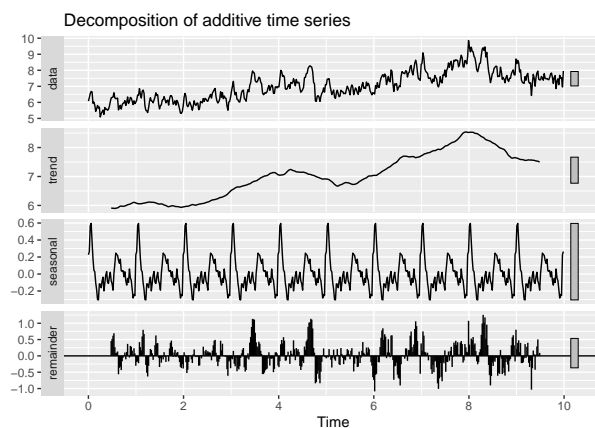


Figure 4: Components of Average Billboard Ranking

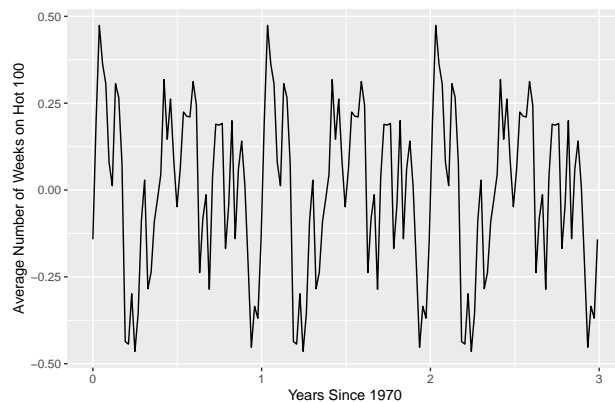


Figure 5: Seasonal Trend of Average Time on Hot 100

2.4 Stationarity

The execution of an Augmented Dickey-Fuller Test results in a rejection of the null hypothesis due to a p-value of 0.01 which concludes the series is stationary. The KPSS test results in a p-value of 0.01 which is significant causing is to reject the null hypothesis and conclude the series is non-stationary. Due to the outcomes of these tests, we conclude the series is not strict stationary.

3 Modeling This Series

3.1 Training and Test Sets

For testing, my training set consists of my entire data set, minus the last year of data (this is 470 weeks or approximately 9 years. My testing set is the last year of data (52 weeks, 1 year).

3.2 Benchmarks

3.2.1 Average Model

I started by getting a fit baseline using an average model to my data with a forecast horizon of 1 year. The residuals are shown below.

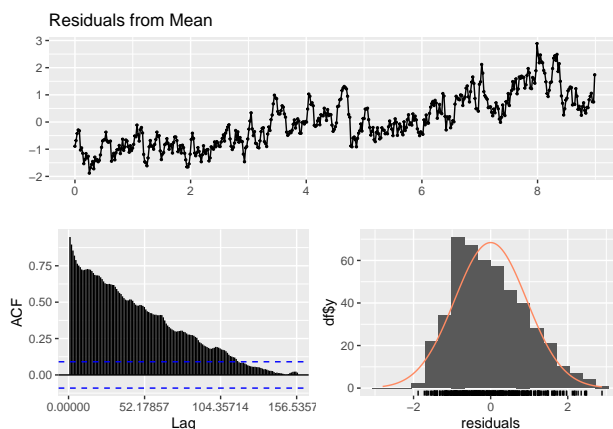


Figure 6: Average Model Residuals

The residuals can be seen to have the same trend as the original timeseries. We can still observe a general decrease followed by an increase and a seasonal cycle not picked up by the model. In addition, the correlogram shows an extremely high level of correlation in the residuals, and a histogram of the residuals follows the right skewed distribution. The Ljung-Box test shows a p-value $< 2.2e-16$, meaning extremely high correlation in residuals, which mirrors this data.

This correlation is expected for our data set. For example, if the average time on the chart in a given week is high, meaning many songs have been on the Hot 100 for an extended period, the average of the previous week is necessarily high, as every song with a period greater than 1 must have been on the chart the previous week. By definition, this time series is autocorrelated.

3.2.1.1 Evaluation Against Test Data

Table 2: Average Model RMSE on Training and Test Sets

	RMSE
Training set	0.932962
Test set	1.448575

This RMSE of about 1.449 on the test set is relatively large for the magnitude of values we are working with, and shows that the average baseline model does not fit the test set well.

3.2.1.2 Cross-Validation

Table 3: Average Model RMSE from 10-Step Cross Validation

RMSE from 10-Step Cross Validation	
	0.9432022

Using cross validation with a 10-step forecast horizon, we find that the RMSE is lower, but still relatively high compared to the magnitude of the data set, at about 0.943.

3.2.2 Seasonal Naive Model

Because this data has a seasonal component, the next step I took was running a seasonal naive (random walk) model. This is the most basic model I can use that still somewhat accounts for the trends present in my time series.

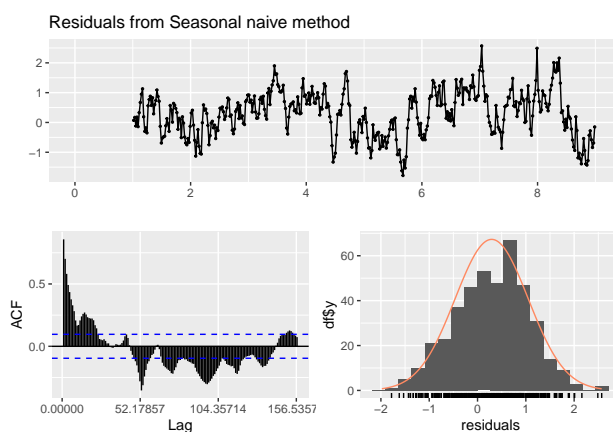


Figure 7: Seasonal Naive Model Residuals

The residuals here, at first glance, look much better. There is not an obvious visible trend in the residuals plot and the correlogram looks slightly more promising than the previous (average) model, with at least some residual values within the significance threshold. The histogram of the residuals also shows a much more normal distribution, with a mean around 0. However, even though the graphs appear to improve slightly, a Ljung-Box test, again results in a near-zero p-value (less than $2.2e-16$). This means there is still a strong correlation in the residuals that the model is not picking up.

3.2.2.1 Evaluation Against Test Data

Table 4: Seasonal Naive Model RMSE on Training and Test Sets

	RMSE
Training set	0.8112364
Test set	1.4400000

This RMSE of about 1.44 on the test set is slightly smaller than our average model. However, it is still extremely large for the magnitude of our data set.

3.2.2.2 Cross-Validation

Table 5: Seasonal Naive Model RMSE from 10-Step Cross Validation

RMSE from 10-Step Cross Validation
0.8429068

Using cross validation with a 10-step forecast horizon, we find that the RMSE of about 0.843 is noticeably lower than the CV RMSE of the average model. However, it is still higher than ideal for our magnitude of data.

3.3 ARIMA Model

To more properly model the data through more complex processes, I chose the ARIMA model. I chose this model because it accounts for autocorrelation which is very much present in the current data set. The ARIMA function I am using automatically corrects for non-stationarity, so it is not needed to adjust for our series being not strict stationary.

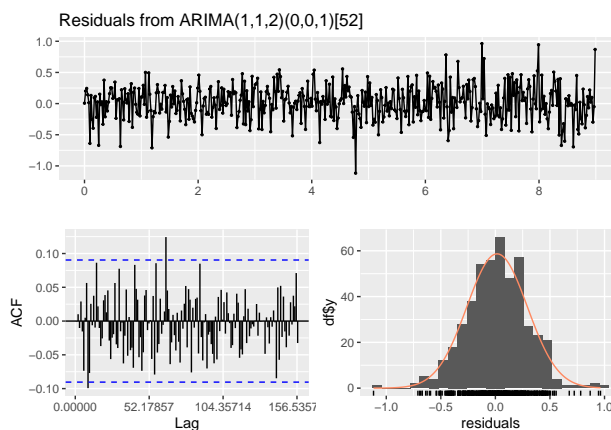


Figure 8: ARIMA Model Residuals

The residuals on this model look much more promising than the previous two. Not only does the plot look like mostly random noise, but the correlogram also seems to be completely random, with only a few lags above the significance threshold. Furthermore, the residuals histogram is approximately normal, with a mean around 0 (excluding outliers). A Ljung-Box test shows a higher, insignificant p-value of approximately 0.2, meaning, there is no observed correlation in the residuals, meaning no observable trends that are not being picked up.

3.3.1 Evaluation Against Test Data

Table 6: ARIMA Model RMSE on Test Set

ARIMA RMSE on Test Set
0.2388568

Here we can see the RMSE on the test set improving drastically, dropping to around 0.2389.

3.3.2 Cross-Validation

Table 7: ARIMA RMSE from 10-Step Cross Validation

RMSE from 10-Step Cross Validation
0.5770741

Using cross validation with a 10-step forecast horizon, we see the RMSE is much improved over the other models, at about 0.577. While this is a drastic improvement over the average and random walk models, it is important to know that the range of our data is 4.77. Therefore, an RMSE of 0.577 is still relatively large compared to the data set.

3.4 Evaluating Predictive Ability

The compiled RMSE values for all three models are shown below.

Table 8: RMSE Values for All Tested Models Against Held Out Data and With 10-Step Cross Validation

Model	Test Set RMSE	Cross-Validation RMSE
Average	1.44857446808511	0.94320220762632
Naive	1.44	0.842906849255444
ARIMA	0.23885676211354	0.577074065922024

Here, a noticeable drop in RMSE for both forms of testing can be observed from the Average to the Naive model, and again from the Naive to the ARIMA model.

Though the RMSE of the ARIMA model is relatively large compared to the size of the data set, it is noticeably smaller than that of the Average and Naive models, both against the test set, and with cross-validation. Therefore, the ARIMA model is the best model used for predicting future data.

3.5 Variable Importances

Because this is a single variable time series, time is considered to be the sole predictor. Therefore, we can not rank the importance of predictors in this model.

3.6 Modeling Future Behavior

It is important to note that the predictions from this model should not be used to determine anything more than what goes into the “average” metric. In other words, the predictions from any model on this data set should only serve to predict general trends in length of time on the Hot 100.

This plot, shown below, uses the ARIMA model to predict the next year (52 weeks) of chart progression. Below is the predicted values for the first 5 weeks of 1980.

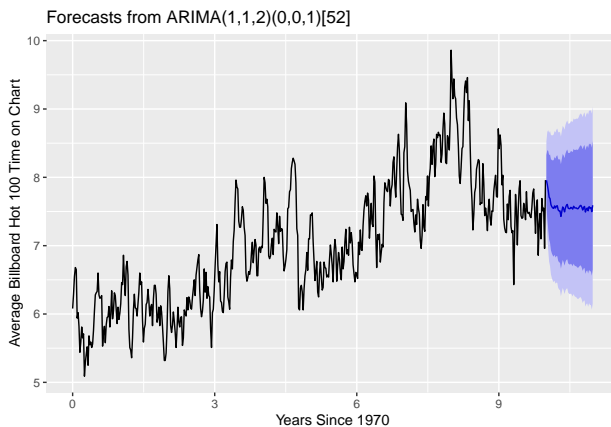


Figure 9: ARIMA Forecasts for One Year.

Table 9: Predicted Amount of Time on Hot 100 for First 5 Weeks of 1980

Week of 1980	Mean ARIMA Prediction
1	7.957146
2	7.895491
3	7.813270
4	7.708698
5	7.688908

4 Conclusion

To answer if the “speed” of pop music is changing (at least along the time period of our data set), we look back to the time series decomposition. The additive underlying trend dictates that, on average, the time a song spent on the Hot 100 was increasing from 1970-1980.

The same decomposition also showed us that there is an underlying seasonal trend where the most popular songs peak in early January.

Lastly, by fitting the Arima model to the data, we are able to predict with a reasonable accuracy the average time on the Hot 100 for the year of 1980.

5 References

[1] Billboard Hot 100 Data (mhollingshead on GitHub)

<https://github.com/mhollingshead/billboard-hot-100>

[2] Billboard Hot 100 Legend

<https://www.billboard.com/billboard-charts-legend/>