# Executive Summary

## Objective

Our goal is to enhance the user experience on our platform by developing an advanced music recommendation system. This system will leverage our dataset to analyze user preferences and deliver personalized song recommendations. This can be achieved by comparing the effectiveness of two models : the BaselineOnly() method from the Surprise library, and a fine-tuned Singular Value Decomposition (SVD) model. The final model’s outcome should deliver recommendations that match user tastes and facilitate the discovery of new and appealing music.

## Methodology

1. Exploratory Data Analysis

*Popular Artists (measured by the number of unique listeners) and Top Hits (See Figure 4 & 5)*

* Having a large listener base doesn't guarantee an artist's songs will be hits. For instance, out of the top 10 artists with the most unique listeners, only Kings of Leon and Florence + The Machine, and OneRepublic have songs in the top 10 most popular list. Interestingly, the artist with the highest total play counts among the top 5 has the smallest number of unique listeners. This indicates a smaller, dedicated fanbase can contribute significantly to an artist's play counts.

*Top listeners and the number of songs they listen to (See Figure 6)*

* Our data shows that the top 5 most active users have each played songs at least 300 times, with the top two exceeding 400 plays. Contrary to the expectation that users with high play counts would listen to a broad range of songs, we observe a different pattern: these top 5 users account for about 1770 plays, but surprisingly, these plays are concentrated on just 12 unique songs. Notably, the most active user alone has played a single song over 400 times, highlighting a very selective listening behavior among our most engaged users.

1. Modeling

Our methodology kicks off with the BaselineOnly() method to set a performance benchmark. This provides a clear point of comparison for more advanced models. We then move on to training and fine-tuning the SVD model, aiming to not only surpass our initial benchmark but also to develop a dynamic system for song recommendations. Our optimization process is made efficient through a custom function, while the accuracy of our model is evaluated using RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) metrics. These metrics help us gauge the precision of our play count predictions to ensure our recommendations are as accurate as possible.

1. Evaluations

Increasing the learning rate, number of factors, and epochs in our model improved its performance, as indicated by reduced RMSE and MAE values, signaling better accuracy. However, we noticed diminishing returns in improvement upon further increases in these hyperparameters (see Figure 2).

Our fine tuned model has an RMSE of 2.49, it shows our predictions are, on average, within 2.49 counts of actual user play counts.

Although the benchmark model had the lowest RMSE, we opted for the SVD model because of its edge in personalization and detecting hidden patterns:

* Latent factors in SVD can help unearth hidden patterns in user-item interactions, allowing for highly personalized recommendations.
* Also, through latent factors, it can detect complex and subtle patterns in the data, such as similarities among users or items that are not immediately apparent. This can lead to more accurate predictions of play counts, as it considers deeper nuances in user behavior and music attributes.
* The difference in RMSE is extremely small.

1. Recommendations

Our model's recommendations have been fascinating, as it introduces me to both familiar tunes and genres I rarely explore, which has been a great way to discover new music (See Figure 3).

A common hurdle for recommendation systems is figuring out what to suggest to new users. Our data analysis shows that just because a song has a lot of unique listeners doesn't mean it'll have high play counts, and vice versa. But there's a reason certain songs and artists stand out: they either have a universal appeal or they're particularly beloved by certain groups.

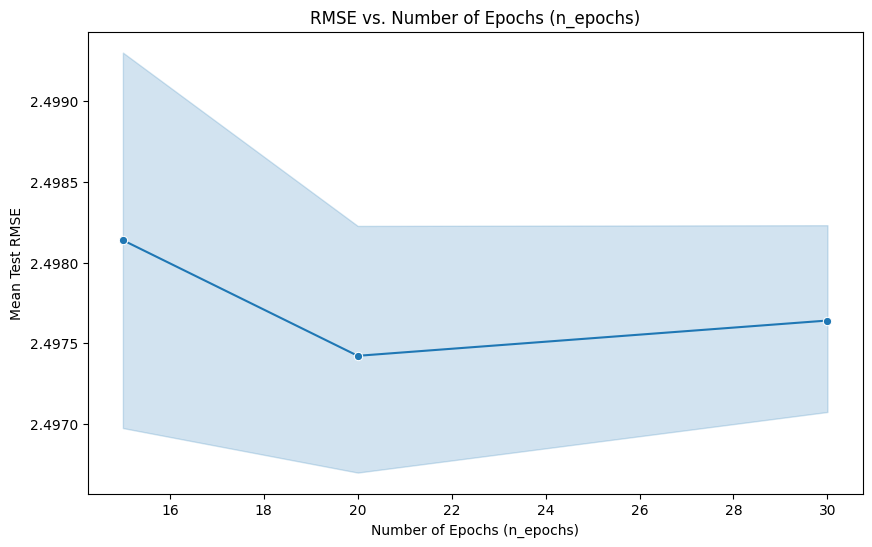
With this in mind, our approach for welcoming new users involves a mix. We take the top 5 songs with the most plays for their broad appeal and combine them with tracks from the artists who attract the most unique listeners. Then, we shuffle and serve up these 10 picks. This way, we're not just playing the hits, we're also introducing new users to a range of popular artists, offering a snapshot of diverse musical tastes.

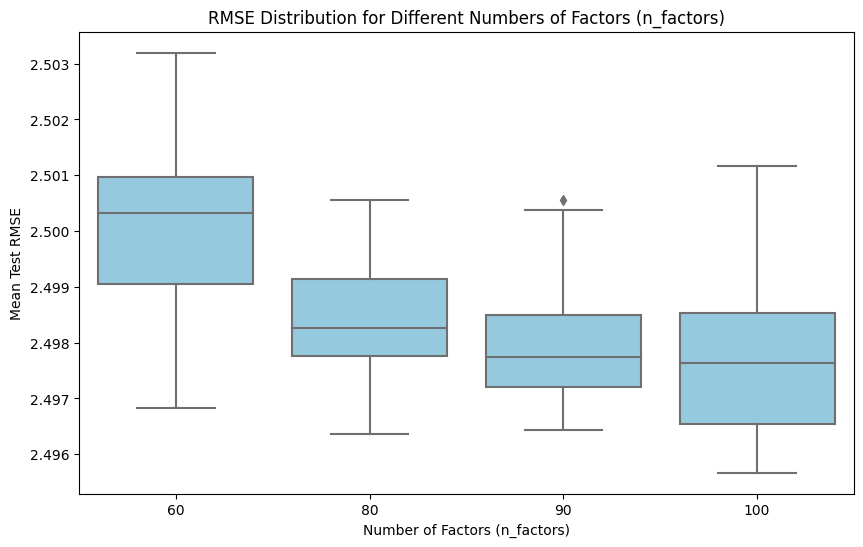
# Appendix

|  | Root Mean Squared Errror (RMSE) | Mean Absolute Error (MAE) |
| --- | --- | --- |
| Benchmark Model | 2.47 | 1.82 |
| Pre-Tuned SVD | 2.50 | 1.85 |
| Fine-Tuned SVD | 2.49 | 1.82 |

Figure 1: Evaluation Metrics

Figure 2: Relationship between hyperparameters and Root Mean Squared Error





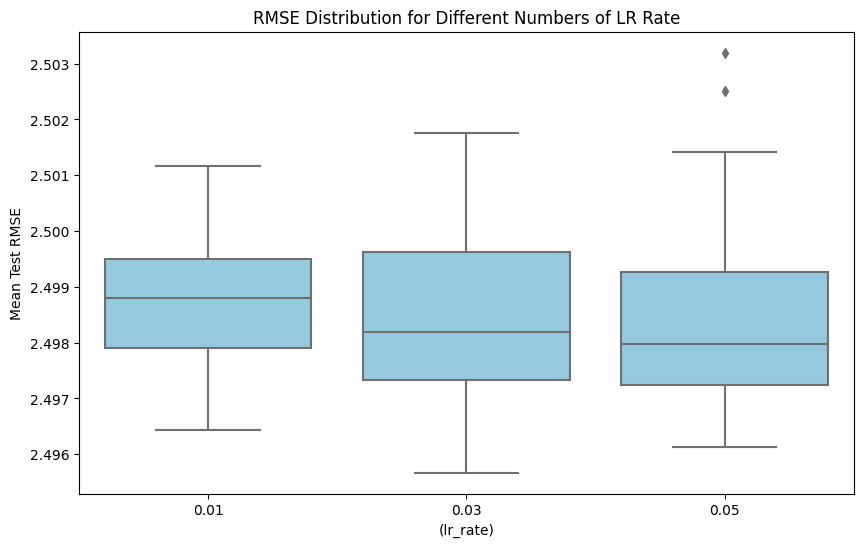


Figure 3: Recommendations for me, my friend, and cold start users (subset of the list)

*Top 10 Song Recommendations for : namtr*

Roly Poly by Bob Wills And His Texas Playboys: 7.8138

Stranger by Thousand Foot Krutch: 7.5061

A Boy Without A Girl by Frankie Avalon: 7.4376

Whiskey\_ You're the Devil by The Clancy Brothers: 7.3826

August Moon by Ottmar Liebert: 7.3671

Smoking Gun by The Robert Cray Band: 7.3220

Fingers Become Thumbs by Future Of The Left: 7.2997

The Time Is Now by WWE John Cena and Tha Trademarc: 7.2527

Enter The Dragon by Charlie Hunter: 7.2341

Un Dia Gris by Paulina Rubio: 7.2271

*Top 10 Song Recommendations for : boshen*

A Deeper Shade Of Soul by Ray Barretto: 8.1062

Un Dia Gris by Paulina Rubio: 7.9366

Come Thou Fount Of Every Blessing by Sufjan Stevens: 7.7399

Cold Blooded (Acid Cleanse) by The fFormula: 7.6264

Two Steps Down by Damnation A.d.: 7.4790

Entre Tus Alas by Camila: 7.4741

Imaginary Love by Rufus Wainwright: 7.4351

Love And Live by Ektomorf: 7.4313

Living In The White by Meneguar: 7.4111

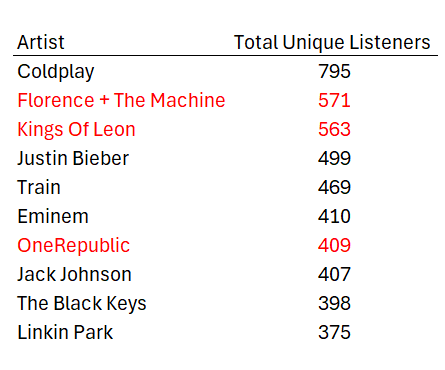
The Quest by HYPOCRISY: 7.3959

*Recommendation for cold start users*

Welcome to our platform, here are 10 songs you might like:

1. You're The One by Dwight Yoakam
2. Sehr kosmisch by Harmonia
3. Pocket Of A Clown (2006 Remastered LP Version) by Dwight Yoakam
4. The Scientist by Coldplay
5. Cosmic Love by Florence + The Machine
6. Dog Days Are Over (Radio Edit) by Florence + The Machine
7. Army of Me by Björk
8. Horn Concerto No. 4 in E flat K495: II. Romance (Andante cantabile) by Barry Tuckwell/Academy of St Martin-in-the-Fields/Sir Neville Marriner
9. Revelry by Kings Of Leon
10. Clocks by Coldplay

Figure 4: Popular Artist vs Top Hits



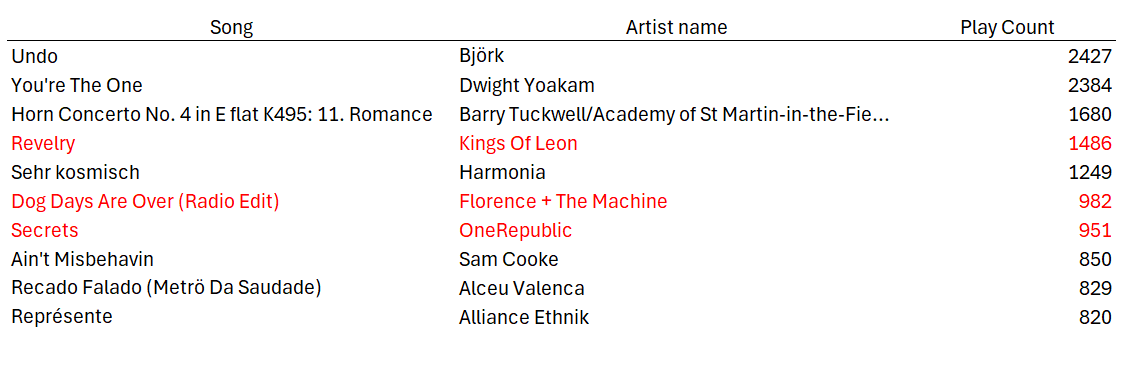


Figure 5: The artist with the most cumulative play counts actually has the lowest number of unique listeners among the top 5.

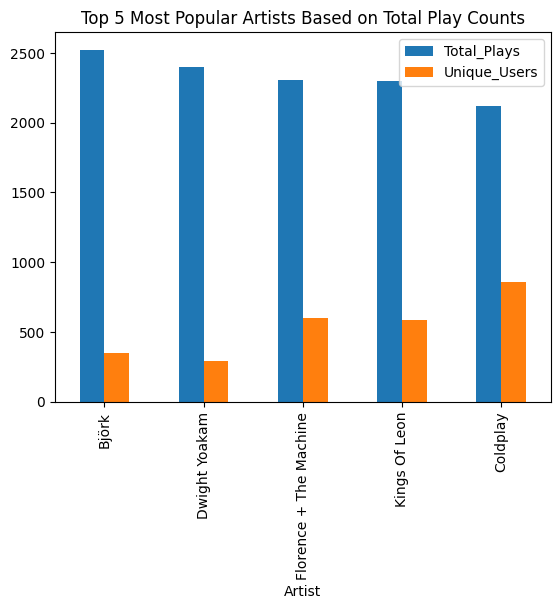


Figure 6: Top listeners and the number of songs they listen to

