**DS 529 FINAL PROJECT REPORT**

Arda Karabiber - S033175

Gökhan Paslı - S027932

Esra Koyuncu - S030088

Sevde Bozdoğan - S026167

1. **Introduction**

We are a team hired by a Turkish digital media agency to analyze the trends of the Turkish song listeners. Our task is to analyze the previous weeks and predict the weekly streams that will occur on 6-12 Jan 2023.

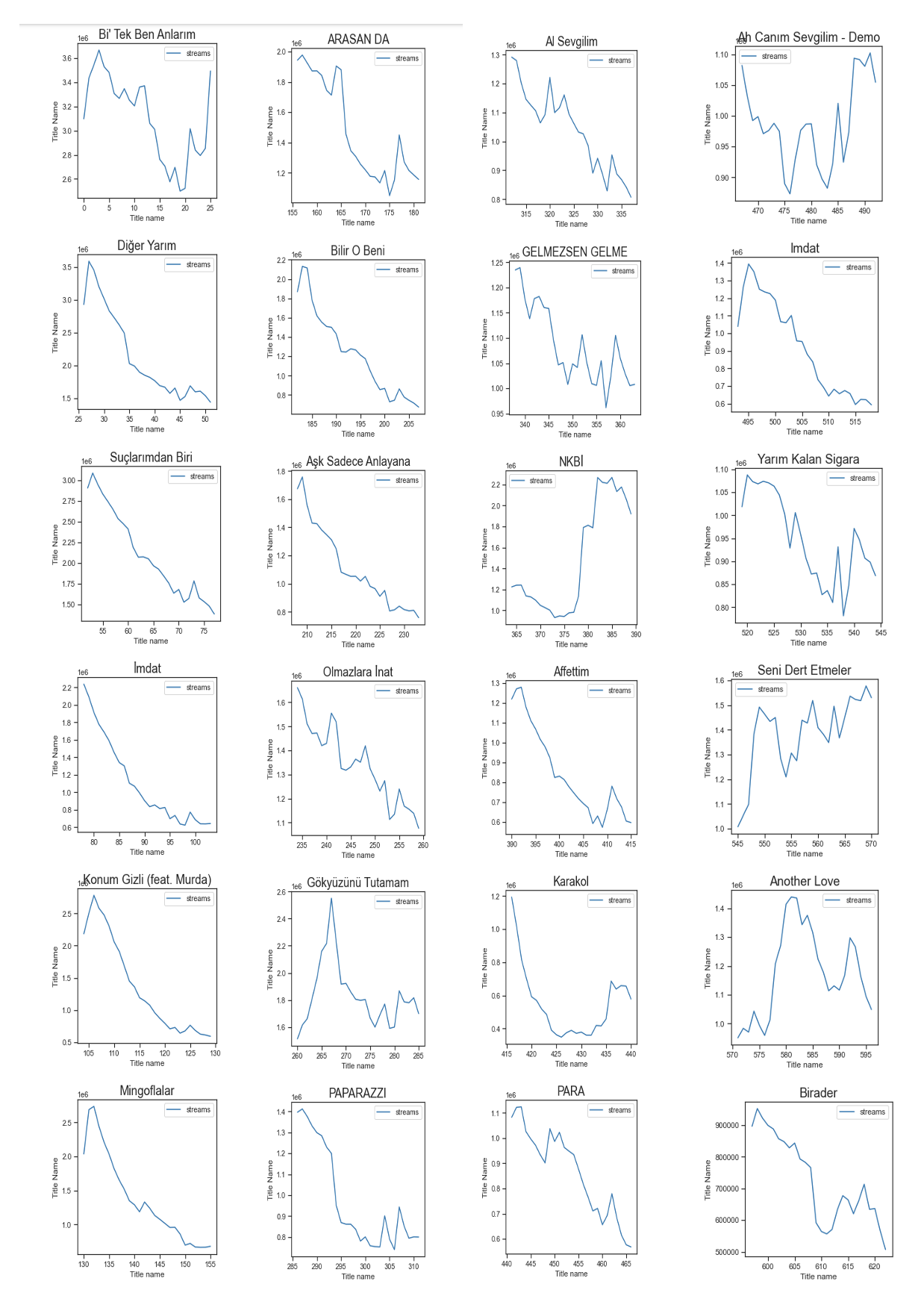
1. **Methodology**

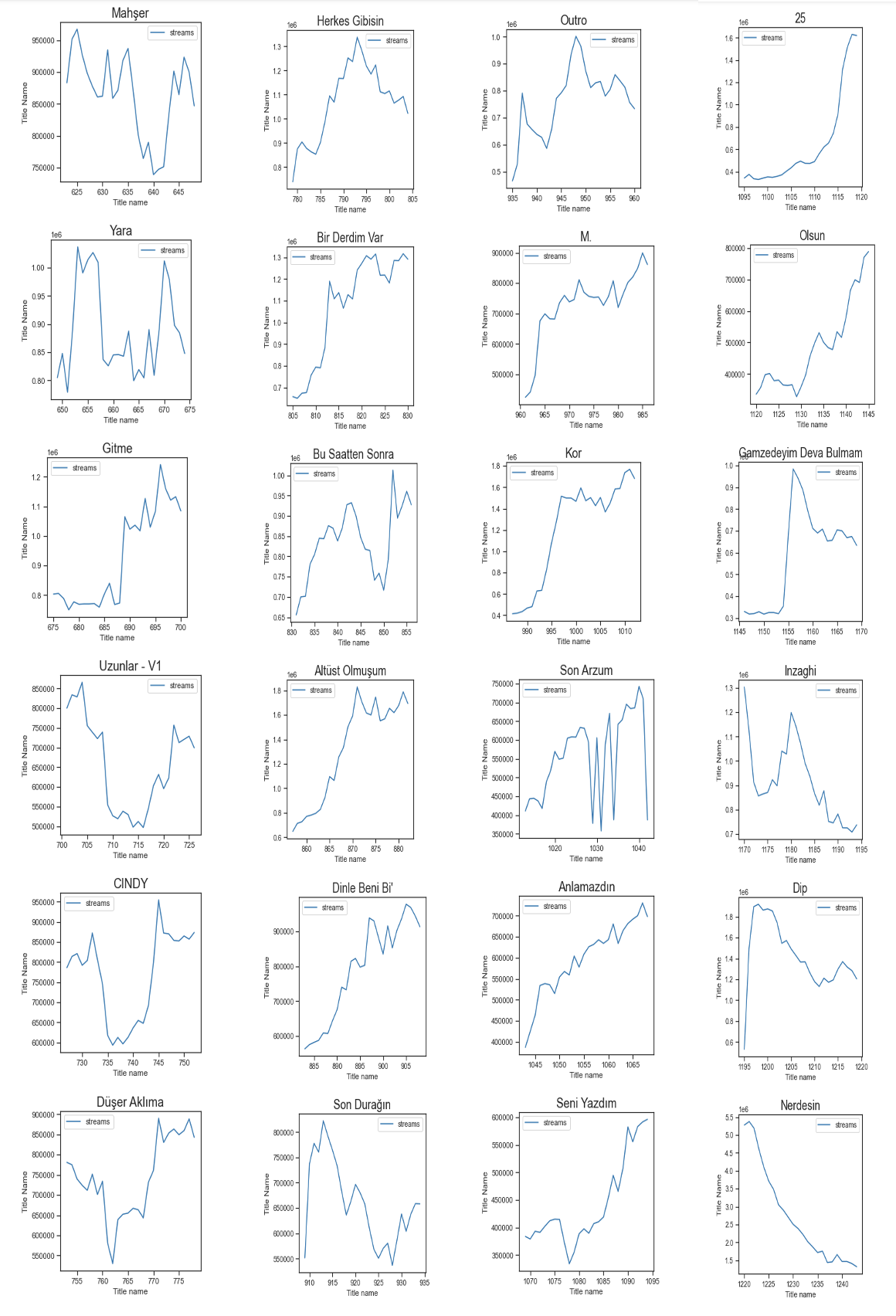
To complete this task, we used historical data generated from Spotify database. We used the data generated from: <https://charts.spotify.com/charts/overview/tr>. We used the last 26 weeks of observations to analyze the data. We tested 5 different prediction models on 10 random songs selected from the 100 songs list. Models we have used are: Logistic Harmonic Regression, ARIMA, Exponential Smoothing, Two Weeks Moving Average Method (with Box Cox transformation), and STL+ARIMA. The first 80% of the available weeks were used as training data and the remaining 20% as test data for each model. Then we compared our results based on the RMSE outcomes. And our next step was to select the model with the lowest RMSE to continue with. The 10 songs in the table below were taken into account in the model selection. When choosing songs, the criteria of songs with different genres and different trends should be taken into consideration.

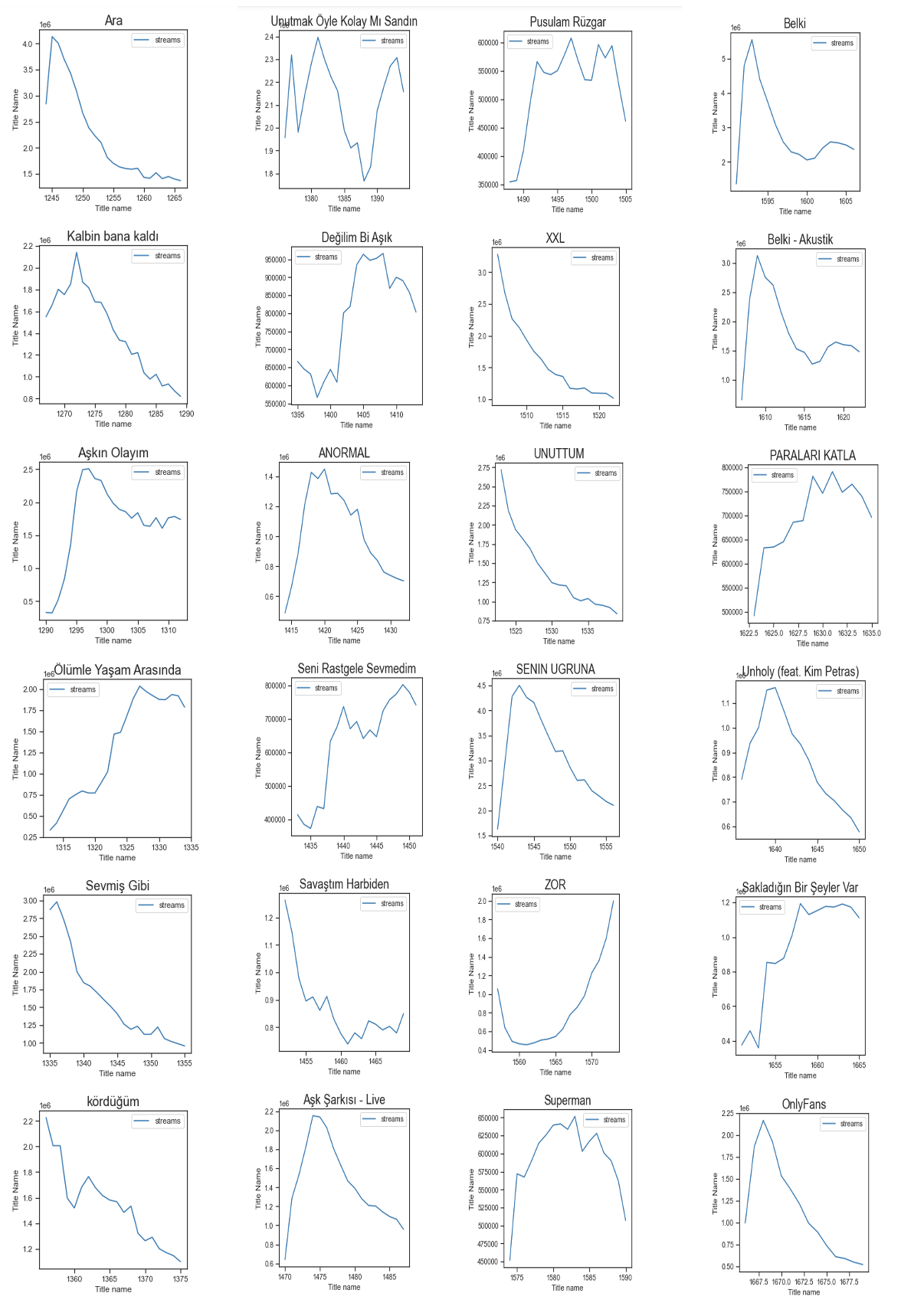
**Table 1:** Metric Comparison of Selected Songs

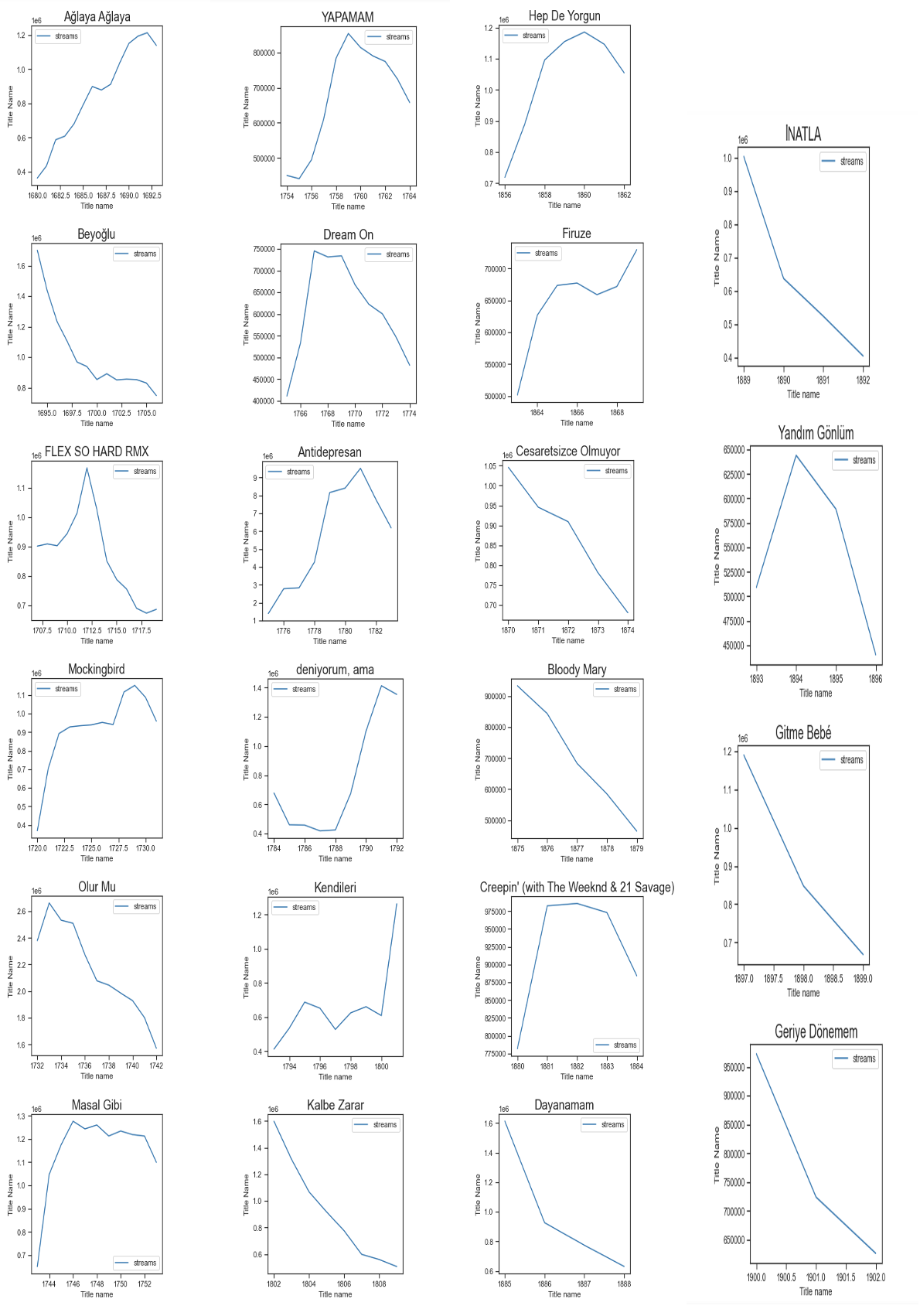
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **RMSE** | | | | |
| **Song Name** | **Logistic Harmonic Regression** | **ARIMA** | **Exponential Smoothing** | **MA2 (Boxcox)** | **STL+ARIMA** |
| **Ölümle Yaşam Arasında** | 10.629 | 150.698 | 69.284 | 34.563 | 1.339.757 |
| **Seni Dert Etmeler** | 259.646 | 152.295 | 57.592 | 27.635 | 326.146 |
| **Ağlaya Ağlaya** | 706 | 217.590 | 50.686 | 25.316 | 731.290 |
| **XXL** | 4.669 | 20.697 | 54.165 | 25.993 | 1.570.228 |
| **Kalbin bana kaldı** | 25.823 | 914.834 | 63.426 | 31.854 | 827.497 |
| **Düşer Aklıma** | 172.012 | 849.078 | 23.745 | 12.739 | 117.598 |
| **YAPAMAM** | 3 | 21.822 | 49.390 | 18.591 | 266.439 |
| **Bloody Mary** | 1,68 | 135.745 | 120.247 | 58.613 | 468.110 |
| **Son Durağın** | 260.354 | 23.574 | 35.222 | 18.591 | 129.901 |
| **Superman** | 1.711 | 42.250 | 36.933 | 19.353 | 85.764 |

It is observed that Dynamic Harmonic Regression gave the best results for data with trend and Moving Average gave the best results for the data with no trend. After selecting our models, we decided to use these models on our data which is categorized based on trend occurrence.

**Figure 1:** Time Series Plots for All 100 Songs







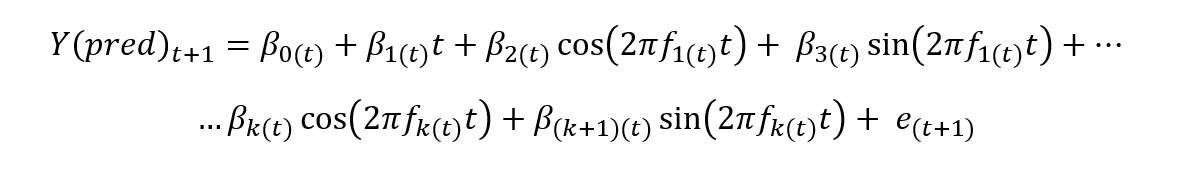
1. **Results**

**Dynamic Harmonic Regression**:

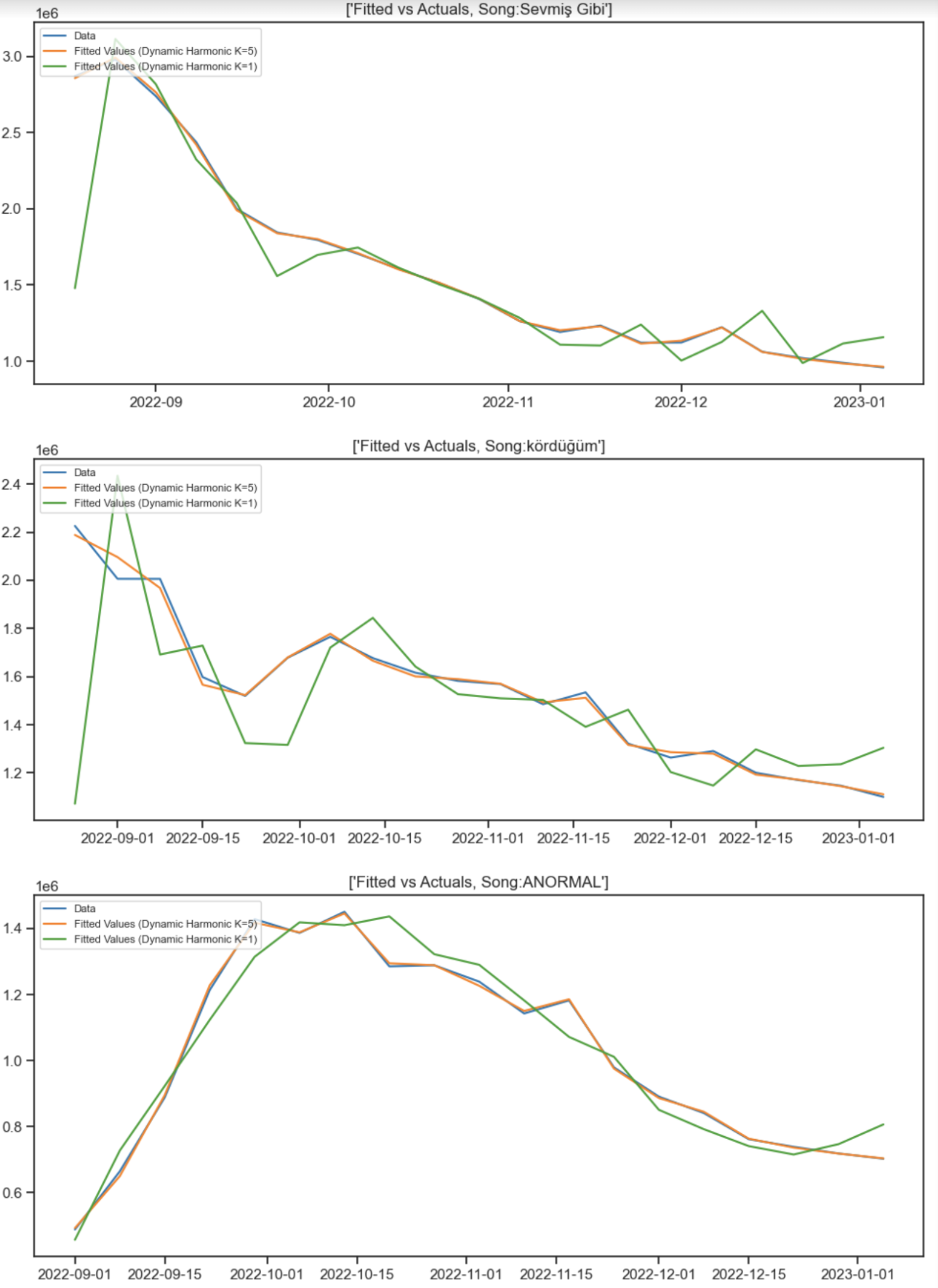
**Songs:** GELMEZSEN GELME, Son Arzum, Anlamazdın, Seni Yazdım, 25, Olsun, Nerdesin, Ara, Kalbin bana kaldı, Aşkın Olayım, Ölümle Yaşam Arasında, Sevmiş Gibi, kördüğüm, ANORMAL, Seni Rastgele Sevmedim, Savaştım Harbiden, Aşk Şarkısı - Live, XXL, UNUTTUM, SENIN UGRUNA, ZOR, Belki, Belki - Akustik, PARALARI KATLA, Unholy (feat. Kim Petras), Sakladığın Bir Şeyler Var, OnlyFans, Ağlaya Ağlaya, Beyoğlu, FLEX SO HARD RMX, Mockingbird, Olur Mu, Masal Gibi, YAPAMAM, Dream On, Antidepresan, deniyorum, ama, Kendileri, Kalbe Zarar, Fırtınadayım, Yalanı Bırak, VUR, 10MG, Hep De Yorgun, Firuze, Cesaretsizce Olmuyor, Bloody Mary, Dayanamam, İNATLA, Yandım Gönlüm, Gitme Bebé, Geriye Dönemem

**Data:** We have collected weekly data for the last 26 weeks.

**Prediction Equation:** We have used dynamic harmonic regression method for prediction with the following parameters:

  
**Reason:** Dynamic harmonic regression was chosen because this method gives the lowest RMSE value for the songs that show a clear trend. It is a regression model based on ARIMA errors. Two different dynamic harmonic regression model are tried, one is with K=5 and the other is with K=1. K=5 gives better results, because AIC decreases as K increases from 1 to 5 and yields lower RMSE. It is clear in the actual vs. fitted plots for three example songs (All plots are located in the python notebook file):

**Figure 2:** Actual vs. Fitted Plots for Three Example



The table with the RMSE values is given below.

**Table 2:** RMSE Results of Dynamic Harmonic Regression

|  |  |
| --- | --- |
| **Song** | **RMSE** |
| GELMEZSEN GELME | 170123,9 |
| Son Arzum | 149819,4 |
| Anlamazdın | 71976,4 |
| Seni Yazdım | 44478,5 |
| 25 | 8205,3 |
| Olsun | 38620,9 |
| Nerdesin | 133412,3 |
| Ara | 42111,6 |
| Kalbin bana kaldı | 25823,2 |
| Aşkın Olayım | 20981,2 |
| Ölümle Yaşam Arasında | 10628,8 |
| Sevmiş Gibi | 9477,2 |
| kördüğüm | 26469,5 |
| ANORMAL | 7268,4 |
| Seni Rastgele Sevmedim | 6592,0 |
| Savaştım Harbiden | 3077,3 |
| Aşk Şarkısı - Live | 5663,3 |
| XXL | 4661,7 |
| UNUTTUM | 4584,4 |
| SENIN UGRUNA | 45715,6 |
| ZOR | 2788,0 |
| Belki | 28214,2 |
| Belki - Akustik | 22889,1 |
| PARALARI KATLA | 1179,7 |
| Unholy (feat. Kim Petras) | 920,1 |
| Sakladığın Bir Şeyler Var | 11986,3 |
| OnlyFans | 3812,4 |
| Ağlaya Ağlaya | 706,3 |
| Beyoğlu | 1230,3 |
| FLEX SO HARD RMX | 2376,5 |
| Mockingbird | 790,6 |
| Olur Mu | 0,000003 |
| Masal Gibi | 0,000006 |
| YAPAMAM | 0,000002 |
| Dream On | 0 |
| Antidepresan | 0,000001 |
| deniyorum, ama | 0,000001 |
| Kendileri | 0,000001 |
| Kalbe Zarar | 0,000001 |
| Fırtınadayım | 0,000001 |
| Yalanı Bırak | 0,000001 |
| VUR | 0,000001 |
| 10MG | 0,000001 |
| Hep De Yorgun | 0,000001 |
| Firuze | 0,000001 |
| Cesaretsizce Olmuyor | 0,000002 |
| Bloody Mary | 0,000002 |
| Dayanamam | 0,000002 |
| İNATLA | 0,000002 |
| Yandım Gönlüm | 0,000002 |
| Gitme Bebé | 0,000003 |
| Geriye Dönemem | 0,000003 |

**2 Weeks Moving Average:**

**Songs:** Diğer Yarım, Suçlarımdan Biri, İmdat, Konum Gizli (feat. Murda), Mingoflalar, Bilir O Beni, Aşk Sadece Anlayana, Olmazlara İnat, Al Sevgilim, NKBİ, Affettim, PARA, Imdat, Birader, Gitme, Herkes Gibisin, Bir Derdim Var, Altüst Olmuşum, Dinle Beni Bi', M., Kor, Dip, Bi' Tek Ben Anlarım, Unutmak Öyle Kolay Mı Sandın, Gökyüzünü Tutamam, Seni Dert Etmeler, ARASAN DA, Another Love, Ah Canım Sevgilim - Demo, Creepin' (with The Weeknd & 21 Savage), Bu Saatten Sonra, Mahşer, Yarım Kalan Sigara, Yara, Değilim Bi Aşık, CINDY, Düşer Aklıma, Keskin, Outro, Derinden/bazen hayat çok kötü gidiyor ve ben onu nasıl anlatacağımı bilemiyorum, PAPARAZZI, Inzaghi,Uzunlar-V1, Son Arzum, Gamzedeyim Deva Bulmam, Karakol, Son Durağın, Pusulam Rüzgar, Superman

**Data:** We have collected weekly data for the last 26 weeks. The first 80% of the available weeks were used as training data and the remaining 20% as test data for each

**Prediction Equation:** We have used 2 weeks moving average method for prediction with the following parameters:

**Reason:** 2 weeks moving average method was chosen because this method gives the lowest RMSE value for test data. The table with the RMSE values is given below.

**Table 3:** RMSE Results of 2 Weeks Moving Average Method

|  |  |
| --- | --- |
| **Song** | **RMSE** |
| Diğer Yarım | 47498,3 |
| Suçlarımdan Biri | 65580,0 |
| İmdat | 38633,1 |
| Konum Gizli (feat. Murda) | 28149,8 |
| Mingoflalar | 33355,3 |
| Bilir O Beni | 32759,7 |
| Aşk Sadece Anlayana | 12840,7 |
| Olmazlara İnat | 29680,7 |
| Al Sevgilim | 32727,6 |
| NKBİ | 47257,4 |
| Affettim | 38486,7 |
| PARA | 31955,5 |
| Imdat | 15821,7 |
| Birader | 27739,0 |
| Gitme | 42125,1 |
| Herkes Gibisin | 18463,8 |
| Bir Derdim Var | 23291,4 |
| Altüst Olmuşum | 37380,5 |
| Dinle Beni Bi' | 16982,1 |
| M. | 18522,9 |
| Kor | 47335,2 |
| Dip | 35831,0 |
| Bi' Tek Ben Anlarım | 161417,4 |
| ARASAN DA | 80403,0 |
| PAPARAZZI | 85536,0 |
| Karakol | 65428,5 |
| Ah Canım Sevgilim - Demo | 29598,9 |
| Yarım Kalan Sigara | 31042,8 |
| Seni Dert Etmeler | 27635,0 |
| Another Love | 39937,2 |
| Mahşer | 27239,4 |
| Uzunlar - V1 | 31007,0 |
| CINDY | 5550,6 |
| Düşer Aklıma | 12738,8 |
| Bu Saatten Sonra | 53139,4 |
| Son Durağın | 18590,7 |
| Outro | 18895,5 |
| Gamzedeyim Deva Bulmam | 15659,9 |
| Inzaghi | 16624,3 |
| Unutmak Öyle Kolay Mı Sandın | 52414,8 |
| Değilim Bi Aşık | 18001,1 |
| Pusulam Rüzgar | 27604,8 |
| Superman | 19353,1 |
| Derinden / bazen hayat çok kötü gidiyor ve ben onu nasıl anlatacağımı bilemiyorum | 37877,6 |
| Keskin | 66339,6 |
| Creepin' (with The Weeknd & 21 Savage) | 53032,2 |
| Gökyüzünü Tutamam | 62266,8 |
| Yara | 35722,0 |