spotify2024-2025

December 28, 2025

[ ]:
```
# Spotify API
import spotipy
from spotipy.oauth2 import SpotifyClientCredentials
import pandas as pd
import numpy as np
import requests
import pprint
from spotipy.oauth2 import SpotifyOAuth
#----------------------#
#
import matplotlib.pyplot as plt
import seaborn as sns

session = requests.Session()
session.trust_env = False

URI = 'http://127.0.0.1:9090/callback'
CLIENT_ID = '317f5cdd80504f37b00bd8a02de80065'
CLIENT_SECRET = 'dd650241d8b743bd8c1d6f148965c6f0'

sp = spotipy.Spotify(auth_manager=SpotifyOAuth(
        client_id=CLIENT_ID,
        client_secret=CLIENT_SECRET,
        redirect_uri=URI,
        scope="user-read-private",
        open_browser=True
    ))

print("     ")
```

[ ]:
```
artist_name =[]
track_name = []
track_popularity =[]
artist_id =[]
track_id =[]
release_date = []
```

```
for i in range(0,1000,50):
    track_results = sp.search(q=f'year:{2025}', type='track', limit=50,␣
 ↪offset=i)
    for i, t in enumerate(track_results['tracks']['items']):
        artist_name.append(t['artists'][0]['name'])
        artist_id.append(t['artists'][0]['id'])
        track_name.append(t['name'])
        track_id.append(t['id'])
        track_popularity.append(t['popularity'])
        release_date.append(t['album']['release_date'])
```

[ ]:
```
track_df = pd.DataFrame({'artist_name' : artist_name, 'track_name' :␣
 ↪track_name, 'track_id' : track_id, 'release_date' : release_date,␣
 ↪'track_popularity' : track_popularity, 'artist_id' : artist_id})
print(track_df.shape)
track_df.head()
```

[19]:
```
track_df
```

[19]:

	artist_name	track_name
0	LE SSERAFIM	SPAGHETTI
1	ILLIT	NOT CUTE ANYMORE
2	SYSTEM SEOUL	
3	ALLDAY PROJECT	FAMOUS
4	ZICO	DUET
…	…	…
1995	Miranda!	Por Amar al Amor (En Vivo Estadio Ferro)
1996	Miranda!	Prisionero (En Vivo Estadio Ferro) (feat. Cris…
1997	Christine D'Clario	Para Mí Es Solo Cristo - Live
1998	Elevation Worship	When Wind Meets Fire (feat. Chris Brown & Tiff…
1999	Miranda!	Perfecta (En Vivo Estadio Ferro) (feat. FMK)

	track_id	release_date	track_popularity
0	2N9miXnewVmUrgl6JSK1FI	2025-10-24	88
1	1k0JAiH11gHL9dc5dfQjQr	2025-11-24	87
2	3b4YlIUpzYs3CZ6s9GVmzT	2025-05-16	59
3	598bDTl82CSTSDnZG0XhGK	2025-06-23	64
4	0V4J2Ass6cNSt8tCYta2Ww	2025-12-18	17
…	…	…	…
1995	15g56hOVAZfuZf9Gu7llk9	2024-12-07	15
1996	5uL9z1sVcr5T9R7zrXnIGW	2024-12-07	16
1997	0U8Ra8Aru8QqszZLIB8CLa	2024-09-28	16
1998	592no3Q0Levk7PrT39XhBs	2024-07-12	17
1999	5Z4SQVyxpjJvSr4o8qlHFi	2024-12-10	15

	artist_id
0	4SpbR6yFEvexJuaBpgAU5p

```
1      36cgvBn0aadzOijnjjwqMN
2      1YBOO3E40cq9VVNUHeQGDm
3      4gEMn0YPOdotLOygnk0Ng2
4      4XpUIb8uuNlIWVKmgKZXC0
…                          …
1995   2eEmsgWmUFMbtU7agJpnjY
1996   2eEmsgWmUFMbtU7agJpnjY
1997   6JaSyvyg28SHC0Of8YE6M9
1998   3YCKuqpv9nCsIhJ2v8SMix
1999   2eEmsgWmUFMbtU7agJpnjY

[2000 rows x 6 columns]
```

[ ]:
```python
import time

artist_popularity = {}
artist_genres = {}
artist_followers = {}

unique_artist_ids = list(set(track_df['artist_id']))

for i in range(0, len(unique_artist_ids), 50):
    batch_ids = unique_artist_ids[i:i+50]

    artists = sp.artists(batch_ids)['artists']

    for artist in artists:
        if artist is None:
            continue

        a_id = artist['id']
        artist_popularity[a_id] = artist['popularity']
        artist_genres[a_id] = artist['genres']
        artist_followers[a_id] = artist['followers']['total']

    time.sleep(0.3)
```

[45]:
```python
track_df['artist_popularity'] = track_df['artist_id'].map(artist_popularity)
track_df['artist_genres'] = track_df['artist_id'].map(artist_genres)
track_df['artist_followers'] = track_df['artist_id'].map(artist_followers)

track_df.to_csv("./data/spotify2024-2025.csv", index=False,␣
 ↪encoding="utf-8-sig")
track_df.head()
```

[45]:
```
     artist_name                                  track_name  \
0  Ufuk Beydemir                                   Rüyalarda
```

```
1    Olivia Dean  Dive - Live From Jimmy Kimmel Live! / 2024
2    Olivia Dean                              Time - Acoustic
3    Olivia Dean                              Time - Acoustic
4    Olivia Dean                    Touching Toes - Acoustic

                track_id release_date  track_popularity  \
0  5NWLfIeTRlm2K1SJulhAbW   2025-11-06                 4
1  1SecgSDJUZ31ovRC57jGyK   2024-04-23                28
2  7alWklN79r0st7vmjuw5Wb   2024-07-18                12
3  7alWklN79r0st7vmjuw5Wb   2024-07-18                12
4  0mNt6f6YeDfDDmytUzCX1n   2024-11-22                13

                artist_id  artist_popularity  artist_genres  artist_followers
0  00C6n2psbFm6XWqyPCXJ5v                 54  [turkish pop]            752656
1  00x1fYSGhdqScXBRpSj3DW                 92     [pop soul]           2361968
2  00x1fYSGhdqScXBRpSj3DW                 92     [pop soul]           2361968
3  00x1fYSGhdqScXBRpSj3DW                 92     [pop soul]           2361968
4  00x1fYSGhdqScXBRpSj3DW                 92     [pop soul]           2361968
```

[46]:
```
track_df.isna().sum()
```

[46]:
```
artist_name          0
track_name           0
track_id             0
release_date         5
track_popularity     0
artist_id            0
artist_popularity    0
artist_genres        0
artist_followers     0
dtype: int64
```

[47]:
```
track_df['release_date'] = pd.to_datetime(track_df['release_date'],␣
 ↪errors='coerce')
track_df = track_df.sort_values(by=['artist_id', 'release_date'])
track_df
```

[47]:
```
        artist_name                                        track_name  \
0     Ufuk Beydemir                                         Rüyalarda
1       Olivia Dean        Dive - Live From Jimmy Kimmel Live! / 2024
2       Olivia Dean                                  Time - Acoustic
3       Olivia Dean                                  Time - Acoustic
4       Olivia Dean                        Touching Toes - Acoustic
…             …                                                  …
1995     Tom Walker                         Lifeline - Live Session
1996          en
1997          en
```

```
1998          Jazzy                                          No Bad Vibes
1999            AVA  MOON (feat. Capo Plaza & Tony Boy) - Acoustic …

                     track_id release_date  track_popularity  \
0     5NWLfIeTRlm2K1SJulhAbW   2025-11-06                 4
1     1SecgSDJUZ31ovRC57jGyK   2024-04-23                28
2     7alWklN79r0st7vmjuw5Wb   2024-07-18                12
3     7alWklN79r0st7vmjuw5Wb   2024-07-18                12
4     0mNt6f6YeDfDDmytUzCX1n   2024-11-22                13
…                        …          …                …
1995  5u2Y9Nd0nbtIqt2BtbOkb6   2024-05-15                 3
1996  1feJ27qymxz673zQ5ZsC8b   2024-11-04                61
1997  1B0fMGDb0AT5gajmcyxG3Z   2024-11-19                59
1998  0dgdDB39uUoC8WICmTE2U8   2024-10-18                78
1999  57AUcDfPLI1H1sKpg9S1qi   2024-02-14                22

                    artist_id  artist_popularity  \
0     00C6n2psbFm6XWqyPCXJ5v                 54
1     00x1fYSGhdqScXBRpSj3DW                 92
2     00x1fYSGhdqScXBRpSj3DW                 92
3     00x1fYSGhdqScXBRpSj3DW                 92
4     00x1fYSGhdqScXBRpSj3DW                 92
…                        …                …
1995  7z2avKuuiMAT4XZJFv8Rvh                 64
1996  7z7tLLiBfmH0kZ2lNVs8LW                 63
1997  7z7tLLiBfmH0kZ2lNVs8LW                 63
1998  7zAAwgV5Wqmvpb4GzvlRkP                 70
1999  7zPS3i8YJBNeDcqXUHfCMr                 58

                                artist_genres  artist_followers
0                                [turkish pop]            752656
1                                   [pop soul]           2361968
2                                   [pop soul]           2361968
3                                   [pop soul]           2361968
4                                   [pop soul]           2361968
…                                          …                …
1995                                       []           1360411
1996  [mandopop, gufeng, c-pop, chinese r&b]            193165
1997  [mandopop, gufeng, c-pop, chinese r&b]            193165
1998                                       []            149051
1999                           [italian trap]            164721

[2000 rows x 9 columns]
```

[49]:
```
track_df['prev_track_popularity'] = (
    track_df
    .groupby('artist_id')['track_popularity']
```

```
    .shift(1)
)
track_df
```

[49]:

```
        artist_name                                          track_name  \
0     Ufuk Beydemir                                           Rüyalarda
1       Olivia Dean        Dive - Live From Jimmy Kimmel Live! / 2024
2       Olivia Dean                                  Time - Acoustic
3       Olivia Dean                                  Time - Acoustic
4       Olivia Dean                       Touching Toes - Acoustic
…             …                                                   …
1995     Tom Walker                        Lifeline - Live Session
1996          en
1997          en
1998          Jazzy                                      No Bad Vibes
1999            AVA  MOON (feat. Capo Plaza & Tony Boy) - Acoustic …

                    track_id release_date  track_popularity  \
0     5NWLfIeTRlm2K1SJulhAbW   2025-11-06                 4
1     1SecgSDJUZ31ovRC57jGyK   2024-04-23                28
2     7alWklN79r0st7vmjuw5Wb   2024-07-18                12
3     7alWklN79r0st7vmjuw5Wb   2024-07-18                12
4     0mNt6f6YeDfDDmytUzCX1n   2024-11-22                13
…                        …          …                 …
1995  5u2Y9Nd0nbtIqt2BtbOkb6   2024-05-15                 3
1996  1feJ27qymxz673zQ5ZsC8b   2024-11-04                61
1997  1B0fMGDb0AT5gajmcyxG3Z   2024-11-19                59
1998  0dgdDB39uUoC8WICmTE2U8   2024-10-18                78
1999  57AUcDfPLI1H1sKpg9S1qi   2024-02-14                22

                   artist_id  artist_popularity  \
0     00C6n2psbFm6XWqyPCXJ5v                 54
1     00x1fYSGhdqScXBRpSj3DW                 92
2     00x1fYSGhdqScXBRpSj3DW                 92
3     00x1fYSGhdqScXBRpSj3DW                 92
4     00x1fYSGhdqScXBRpSj3DW                 92
…                        …                …
1995  7z2avKuuiMAT4XZJFv8Rvh                 64
1996  7z7tLLiBfmH0kZ2lNVs8LW                 63
1997  7z7tLLiBfmH0kZ2lNVs8LW                 63
1998  7zAAwgV5Wqmvpb4GzvlRkP                 70
1999  7zPS3i8YJBNeDcqXUHfCMr                 58

                              artist_genres  artist_followers  \
0                              [turkish pop]            752656
1                                 [pop soul]           2361968
2                                 [pop soul]           2361968
```

```
3                                  [pop soul]           2361968
4                                  [pop soul]           2361968
…                                         …                  …
1995                                     []           1360411
1996  [mandopop, gufeng, c-pop, chinese r&b]            193165
1997  [mandopop, gufeng, c-pop, chinese r&b]            193165
1998                                     []            149051
1999                           [italian trap]            164721

      prev_track_popularity
0                       NaN
1                       NaN
2                      28.0
3                      12.0
4                      12.0
…                         …
1995                    3.0
1996                    NaN
1997                   61.0
1998                    NaN
1999                    NaN

[2000 rows x 10 columns]
```

[34]:
```
track_df.reset_index(drop=True)
```

[34]:
```
        artist_name                                          track_name  \
0     Ufuk Beydemir                                           Rüyalarda
1       Olivia Dean         Dive - Live From Jimmy Kimmel Live! / 2024
2       Olivia Dean                                   Time - Acoustic
3       Olivia Dean                                   Time - Acoustic
4       Olivia Dean                            Touching Toes - Acoustic
…             …                                                    …
1995     Tom Walker                             Lifeline - Live Session
1996          en
1997          en
1998          Jazzy                                        No Bad Vibes
1999            AVA  MOON (feat. Capo Plaza & Tony Boy) - Acoustic …

                    track_id release_date  track_popularity  \
0     5NWLfIeTRlm2K1SJulhAbW   2025-11-06                 4
1     1SecgSDJUZ31ovRC57jGyK   2024-04-23                28
2     7alWklN79r0st7vmjuw5Wb   2024-07-18                12
3     7alWklN79r0st7vmjuw5Wb   2024-07-18                12
4     0mNt6f6YeDfDDmytUzCX1n   2024-11-22                13
…                        …          …                 …
1995  5u2Y9Nd0nbtIqt2Btb0kb6   2024-05-15                 3
```

```
1996  1feJ27qymxz673zQ5ZsC8b   2024-11-04                61
1997  1B0fMGDb0AT5gajmcyxG3Z   2024-11-19                59
1998  0dgdDB39uUoC8WICmTE2U8   2024-10-18                78
1999  57AUcDfPLI1H1sKpg9S1qi   2024-02-14                22

                    artist_id  artist_popularity  \
0     00C6n2psbFm6XWqyPCXJ5v                 54
1     00x1fYSGhdqScXBRpSj3DW                 92
2     00x1fYSGhdqScXBRpSj3DW                 92
3     00x1fYSGhdqScXBRpSj3DW                 92
4     00x1fYSGhdqScXBRpSj3DW                 92
…                        …                 …
1995  7z2avKuuiMAT4XZJFv8Rvh                 64
1996  7z7tLLiBfmH0kZ2lNVs8LW                 63
1997  7z7tLLiBfmH0kZ2lNVs8LW                 63
1998  7zAAwgV5Wqmvpb4GzvlRkP                 70
1999  7zPS3i8YJBNeDcqXUHfCMr                 58

                                artist_genres  artist_followers  \
0                                [turkish pop]            752656
1                                   [pop soul]           2361968
2                                   [pop soul]           2361968
3                                   [pop soul]           2361968
4                                   [pop soul]           2361968
…                                          …                 …
1995                                       []           1360411
1996  [mandopop, gufeng, c-pop, chinese r&b]            193165
1997  [mandopop, gufeng, c-pop, chinese r&b]            193165
1998                                       []            149051
1999                            [italian trap]            164721

      prev_track_popularity
0                       NaN
1                       NaN
2                      28.0
3                      12.0
4                      12.0
…                         …
1995                    3.0
1996                    NaN
1997                   61.0
1998                    NaN
1999                    NaN

[2000 rows x 10 columns]
```

[50]:
```
popularity_df = track_df.dropna()
```

[51]:
```
popularity_df
```

[51]:

	artist_name	track_name	track_id	release_date	track_popularity	artist_id	artist_popularity	artist_genres	artist_followers
2	Olivia Dean	Time - Acoustic	7alWklN79r0st7vmjuw5Wb	2024-07-18	12	00x1fYSGhdqScXBRpSj3DW	92	[pop soul]	2361968
3	Olivia Dean	Time - Acoustic	7alWklN79r0st7vmjuw5Wb	2024-07-18	12	00x1fYSGhdqScXBRpSj3DW	92	[pop soul]	2361968
4	Olivia Dean	Touching Toes - Acoustic	0mNt6f6YeDfDDmytUzCX1n	2024-11-22	13	00x1fYSGhdqScXBRpSj3DW	92	[pop soul]	2361968
5	Olivia Dean	Touching Toes - Acoustic	0mNt6f6YeDfDDmytUzCX1n	2024-11-22	13	00x1fYSGhdqScXBRpSj3DW	92	[pop soul]	2361968
6	Olivia Dean	Touching Toes - Acoustic	0mNt6f6YeDfDDmytUzCX1n	2024-11-22	13	00x1fYSGhdqScXBRpSj3DW	92	[pop soul]	2361968
…	…	…	…	…	…	…	…	…	…
1989	Frédéric Chopin	Minute Waltz (Op. 64 No. 1) - Live On The Ed S…	6UFVtlQH4GCohk2OZyNRO7	2025-11-05	16	7y97mc3bZRFXzT2szRM4L4	75		
1993	Tom Walker	Head Underwater - Live Session	7FC5OU9uuQcr9t1vM75Jq6	2024-05-09	3	7z2avKuuiMAT4XZJFv8Rvh	64		
1994	Tom Walker	Lifeline - Live Session	5u2Y9Nd0nbtIqt2BtbOkb6	2024-05-15	3	7z2avKuuiMAT4XZJFv8Rvh	64		
1995	Tom Walker	Lifeline - Live Session	5u2Y9Nd0nbtIqt2BtbOkb6	2024-05-15	3	7z2avKuuiMAT4XZJFv8Rvh	64		
1997	en		1B0fMGDb0AT5gajmcyxG3Z	2024-11-19	59	7z7tLLiBfmH0kZ2lNVs8LW	63		

```
…                                        …              …
1989            [classical piano, classical]           3595526
1993                                      []           1360411
1994                                      []           1360411
1995                                      []           1360411
1997  [mandopop, gufeng, c-pop, chinese r&b]            193165

      prev_track_popularity
2                      28.0
3                      12.0
4                      12.0
5                      13.0
6                      13.0
…                      …
1989                    7.0
1993                    3.0
1994                    3.0
1995                    3.0
1997                   61.0

[934 rows x 10 columns]
```

[61]:
```
from scipy.stats import pearsonr, spearmanr

pearson_r, pearson_p = pearsonr(
    popularity_df['prev_track_popularity'],
    popularity_df['track_popularity']
)

spearman_r, spearman_p = spearmanr(
    popularity_df['prev_track_popularity'],
    popularity_df['track_popularity']
)

print(f"pearson : r = {pearson_r:.3f}, p = {pearson_p:.3f}")
print(f"spearman : r = {spearman_r:.3f}, p = {spearman_p:.3f}")
```

```
pearson : r = 0.843, p = 0.000
spearman : r = 0.835, p = 0.000
```

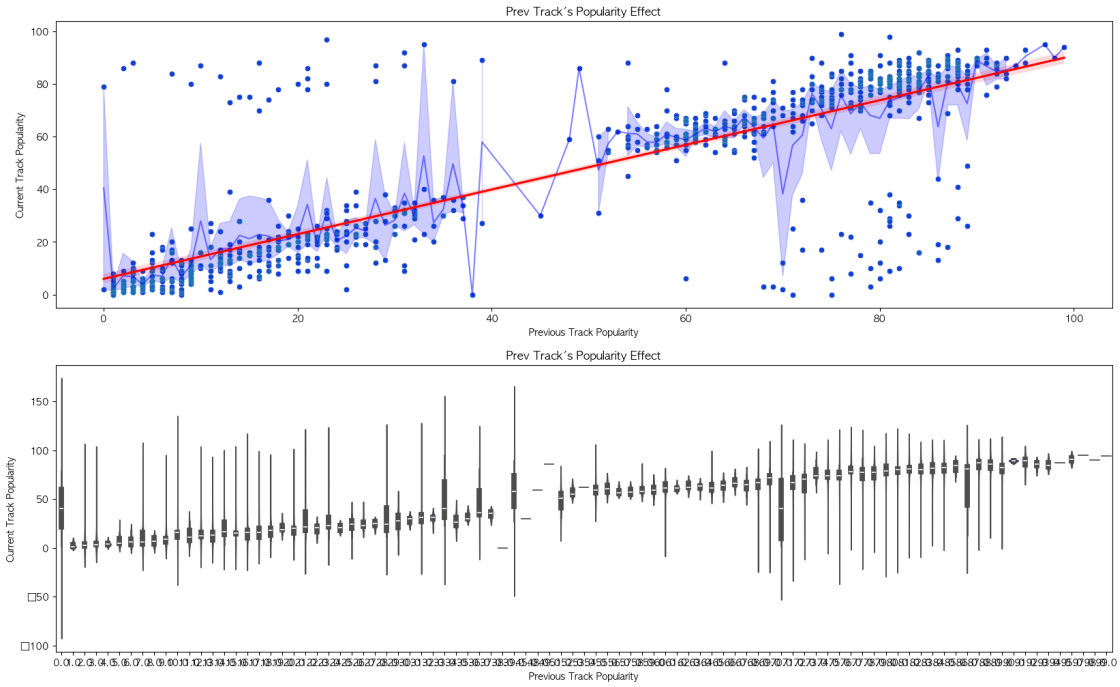
[63]:
```
import statsmodels.api as sm

X = popularity_df[['prev_track_popularity']]
X = sm.add_constant(X)
y = popularity_df['track_popularity']

model = sm.OLS(y, X).fit()
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:       track_popularity   R-squared:                       0.711
Model:                           OLS   Adj. R-squared:                  0.711
Method:                Least Squares   F-statistic:                     2291.
Date:               Fri, 26 Dec 2025   Prob (F-statistic):          2.41e-253
Time:                       14:46:19   Log-Likelihood:                -3982.9
No. Observations:                934   AIC:                             7970.
Df Residuals:                    932   BIC:                             7980.
Df Model:                          1
Covariance Type:            nonrobust
==============================================================================
=========
                             coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
---------
const                      5.9970      1.005      5.969      0.000       4.025
7.969
prev_track_popularity      0.8483      0.018     47.866      0.000       0.814
0.883
==============================================================================
Omnibus:                      150.635   Durbin-Watson:                   2.273
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2229.047
Skew:                          -0.138   Prob(JB):                         0.00
Kurtosis:                      10.563   Cond. No.                         101.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[85]:
```
plt.figure(figsize=(20,12))
plt.rc('font', family='AppleGothic')
plt.subplot(2,1,1)
sns.scatterplot(x="prev_track_popularity", y="track_popularity",␣
 ↪data=popularity_df, color='blue')
sns.lineplot(x="prev_track_popularity", y="track_popularity",␣
 ↪data=popularity_df, color='blue', alpha = 0.5)
sns.regplot(x='prev_track_popularity', y='track_popularity', data=popularity_df,
            scatter_kws={'alpha':0.5, 's':20}, line_kws={'color':'red'})
plt.xlabel('Previous Track Popularity')
plt.ylabel('Current Track Popularity')
plt.title('Prev Track\'s Popularity Effect')

plt.subplot(2,1,2)
```

```
sns.violinplot(x="prev_track_popularity", y="track_popularity",␣
 ↪data=popularity_df, color='blue')
plt.xlabel('Previous Track Popularity')
plt.ylabel('Current Track Popularity')
plt.title('Prev Track\'s Popularity Effect')
plt.show()
```

```
/Users/lucyroh/Desktop/STUDY/spotify/.venv/lib/python3.14/site-
packages/IPython/core/pylabtools.py:170: UserWarning: Glyph 8722 (\N{MINUS
SIGN}) missing from font(s) AppleGothic.
  fig.canvas.print_figure(bytes_io, **kw)
```



0.0.1

[76]:
```
popularity_df['popularity_centered'] = (
    popularity_df['track_popularity']
    - popularity_df.groupby('artist_id')['track_popularity'].transform('mean')
)
```

```
/var/folders/fv/hnh2s54s4h9_vv9q4x1rbdj40000gn/T/ipykernel_27892/764928336.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

```
  popularity_df['popularity_centered'] = (
```

[77]:
```
popularity_df['prev_popularity_centered'] = (
    popularity_df['prev_track_popularity']
    - popularity_df
      .groupby('artist_id')['prev_track_popularity']
      .transform('mean')
)
```

```
/var/folders/fv/hnh2s54s4h9_vv9q4x1rbdj40000gn/T/ipykernel_27892/3387166168.py:1
: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  popularity_df['prev_popularity_centered'] = (
```

[79]:
```
from scipy.stats import pearsonr, spearmanr

pearson_r, pearson_p = pearsonr(
    popularity_df['prev_popularity_centered'],
    popularity_df['popularity_centered']
)

spearman_r, spearman_p = spearmanr(
    popularity_df['prev_popularity_centered'],
    popularity_df['popularity_centered']
)

print(f"pearson : r = {pearson_r:.3f}, p = {pearson_p:.3f}")
print(f"spearman : r = {spearman_r:.3f}, p = {spearman_p:.3f}")
```

```
pearson : r = 0.320, p = 0.000
spearman : r = 0.201, p = 0.000
```

[82]:
```
X = popularity_df[['prev_popularity_centered']]
X = sm.add_constant(X)
y = popularity_df['popularity_centered']

model = sm.OLS(y, X).fit()
print(model.summary())
```

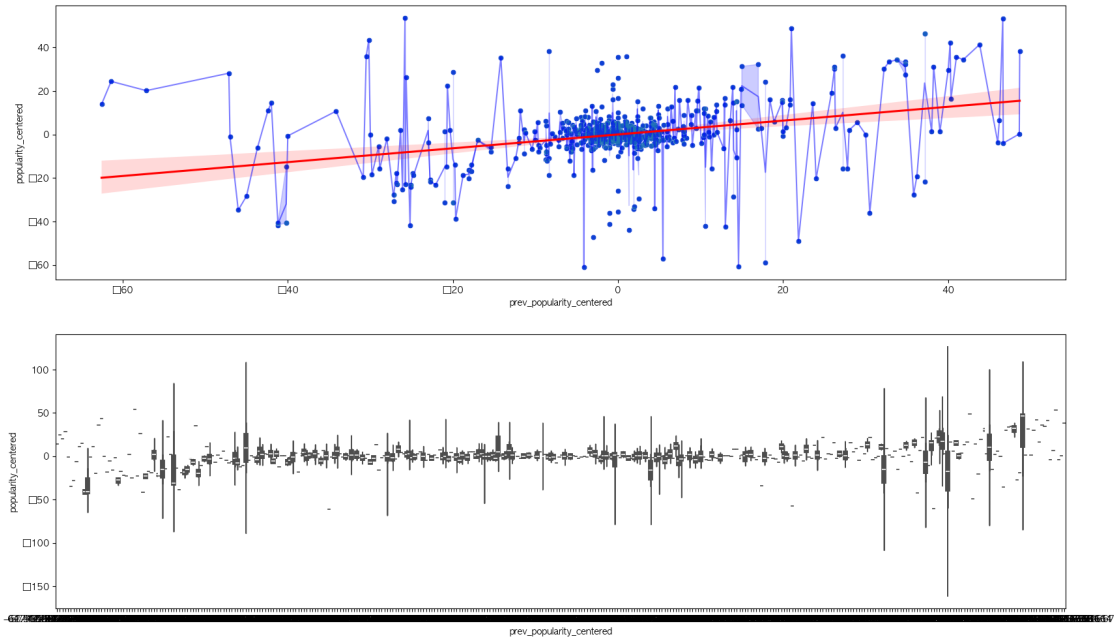
```
                            OLS Regression Results
==============================================================================
Dep. Variable:     popularity_centered   R-squared:                       0.103
Model:                             OLS   Adj. R-squared:                  0.102
Method:                  Least Squares   F-statistic:                     106.6
Date:                 Fri, 26 Dec 2025   Prob (F-statistic):           9.59e-24
```

```
Time:                           15:20:15   Log-Likelihood:                -3596.9
No. Observations:                    934   AIC:                             7198.
Df Residuals:                        932   BIC:                             7208.
Df Model:                              1
Covariance Type:               nonrobust
================================================================================
============
                                 coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
------------
const                      -2.359e-16      0.373  -6.33e-16      1.000      -0.732
0.732
prev_popularity_centered       0.3181      0.031     10.327      0.000       0.258
0.378
==============================================================================
Omnibus:                      203.774   Durbin-Watson:                   2.185
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3191.038
Skew:                          -0.530   Prob(JB):                         0.00
Kurtosis:                      11.993   Cond. No.                         12.1
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

- 1 , 0.32 .

[83]:
```
plt.figure(figsize=(20,12))
plt.rc('font', family='AppleGothic')
plt.subplot(2,1,1)
sns.scatterplot(x="prev_popularity_centered", y="popularity_centered",␣
 ↪data=popularity_df, color='blue')
sns.lineplot(x="prev_popularity_centered", y="popularity_centered",␣
 ↪data=popularity_df, color='blue', alpha = 0.5)
sns.regplot(x='prev_popularity_centered', y='popularity_centered',␣
 ↪data=popularity_df,
            scatter_kws={'alpha':0.5, 's':20}, line_kws={'color':'red'})
plt.subplot(2,1,2)
sns.violinplot(x="prev_popularity_centered", y="popularity_centered",␣
 ↪data=popularity_df, color='blue')
plt.show()
```

```
/Users/lucyroh/Desktop/STUDY/spotify/.venv/lib/python3.14/site-
packages/IPython/core/pylabtools.py:170: UserWarning: Glyph 8722 (\N{MINUS
SIGN}) missing from font(s) AppleGothic.
  fig.canvas.print_figure(bytes_io, **kw)
```



[ ]:

1 2023-2025 Analysis

1.0.1 . ?

- (popularity_centered) ?
- : popularity_centered
- : (X)
- : artist-level time series

[101]:
```
years = [2023]

artist_name =[]
track_name = []
track_popularity =[]
artist_id =[]
track_id =[]
release_date = []
for i in range(0,1000,50):
    track_results = sp.search(q=f'year:{2023}', type='track', limit=50,␣
 ↪offset=i)
    for i, t in enumerate(track_results['tracks']['items']):
        artist_name.append(t['artists'][0]['name'])
        artist_id.append(t['artists'][0]['id'])
        track_name.append(t['name'])
```

```
        track_id.append(t['id'])
        track_popularity.append(t['popularity'])
        release_date.append(t['album']['release_date'])
```

[102]:
```
df_2023 = pd.DataFrame({'artist_name' : artist_name, 'track_name' : track_name,␣
 ↪'track_id' : track_id, 'release_date' : release_date, 'track_popularity' :␣
 ↪track_popularity, 'artist_id' : artist_id})
print(df_2023.shape)
df_2023.to_csv("spotify2023.csv", encoding="utf-8-sig", index=False)
display(df_2023.head())
track_df = pd.concat([track_df, df_2023], ignore_index=True)
track_df.to_csv("spotify2023-2025.csv", encoding="utf-8-sig", index=False)
del df_2023
track_df
```

```
(1000, 6)
```

	artist_name	track_name	track_id	release_date	track_popularity	artist_id
0	TOMIOKA AI	Good bye-bye	6yrSg5ltrGVOoha4TtUBZ3	2023-09-20	55	28uwiPI48qGWTiPz6Zgnwi
1	keshi	UNDERSTAND - Acoustic	1IkbcPeZ14Ed9ZCLwU7bOV	2023-05-16	29	3pc0bOVB5whxmD50W79wwO
2	tuki.	- Bansanka	2cxcqKwQtfKq7giphxjzOu	2023-09-29	67	1Y5vJqABeI6QI6R95EDV6o
3	Way Ched	FLIRT (Feat. Leellamarz)	6oBBMles8dHGlGRT4sZvhJ	2023-11-12	56	75OcDAFGCzj0qehe1mADeM
4	BIBI	Amigos	1fzJtRgnplgXnBnr6DLpTB	2023-11-17	19	6UbmqUEgjLA6jAcXwbM1Z9

[102]:

	artist_name	track_name	track_id	release_date	track_popularity
0	Giuseppe Verdi	Cortigiani, vil razza dannata - Live On The Ed…	4xTQrn63kUDmKXBIJupxVb	2024-01-01 00:00:00	3
1	Noah Kahan	Your Needs, My Needs - Live from Red Rocks ‘23	51BSR67RxiZySQI61sFXQa	2024-01-04 00:00:00	9
2	Noah Kahan	Your Needs, My Needs - Live from Red Rocks ‘23	51BSR67RxiZySQI61sFXQa	2024-01-04 00:00:00	9
3					
4	Andree Right Hand	Dân Chơi Sao Phải Khóc			
…	…	…			
5995	Isabel LaRosa	i'm yours - Live from Irving Plaza			
5996	Pedro Capó	5 y 3 - Live Performance			
5997	AnnenMayKantereit	Lass es kreisen - Live in Köln			
5998	Little Baby Bum	Hush Little Baby			
5999	Costa Gold	Lembra Dessa			

```
3     5eY7692tmgHB9dbmq6wa2M  2024-01-05 00:00:00                73
4     0dBKcPEAsdxWJsqNDNHcPz  2024-01-05 00:00:00                56
…                        …                   …                 …
5995  3NpVUgii1cGAl3pze7CFzb           2023-05-11                13
5996  0IdDYDdJdG83RkjlROhDHv           2023-06-23                12
5997  2CavjEfW3xt5enUOUJNys8           2023-03-10                13
5998  32w2lmKGmMtSOgKVRQBvLj           2023-09-01                12
5999  3NB5u20YOFB5ZTOWp0D9w2           2023-01-20                13

                  artist_id  artist_popularity  \
0     1JOQXgYdQV2yfrhewqx96o               63.0
1     2RQXRUsr4IW1f3mKyKsy4B               84.0
2     2RQXRUsr4IW1f3mKyKsy4B               84.0
3     4UK2Lzi6fBfUi9rpDt6cik               73.0
4     4grjJqg7iwQ8RKHs8d9Snh               56.0
…                        …                 …
5995  5arKwJZEvT5uKq4o0JfqR4                NaN
5996  4QVBYiagIaa6ZGSPMbybpy                NaN
5997  23xqmJEN3oVxwzqtNIyR5m                NaN
5998  6b4gOldpp7HOBjFOlTfleW                NaN
5999  7q1aEytv83jXNECmyaMhgn                NaN

                                 artist_genres  artist_followers  \
0                          [opera, classical]          723072.0
1                                          []         4582855.0
2                                          []         4582855.0
3     [j-pop, j-rock, japanese indie, anime]         3358594.0
4     [vietnamese hip hop, v-pop, vinahouse]          448951.0
…                                        …                 …
5995                                   NaN               NaN
5996                                   NaN               NaN
5997                                   NaN               NaN
5998                                   NaN               NaN
5999                                   NaN               NaN

      prev_track_popularity
0                       NaN
1                       NaN
2                       9.0
3                       NaN
4                       NaN
…                       …
5995                    NaN
5996                    NaN
5997                    NaN
5998                    NaN
5999                    NaN
```

```
[6000 rows x 10 columns]
```

[103]:
```
track_df['release_date'] = pd.to_datetime(track_df['release_date'],␣
 ↪errors='coerce')
```

[104]:
```
artist_popularity = {}
artist_genres = {}
artist_followers = {}

unique_artist_ids = list(set(track_df['artist_id']))

for i in range(0, len(unique_artist_ids), 50):
    batch_ids = unique_artist_ids[i:i+50]

    artists = sp.artists(batch_ids)['artists']

    for artist in artists:
        if artist is None:
            continue

        a_id = artist['id']
        artist_popularity[a_id] = artist['popularity']
        artist_genres[a_id] = artist['genres']
        artist_followers[a_id] = artist['followers']['total']

    time.sleep(0.3)
```

[105]:
```
track_df['artist_popularity'] = track_df['artist_id'].map(artist_popularity)
track_df['artist_genres'] = track_df['artist_id'].map(artist_genres)
track_df['artist_followers'] = track_df['artist_id'].map(artist_followers)

track_df.to_csv("./data/spotify2023-2025.csv", index=False,␣
 ↪encoding="utf-8-sig")
track_df
```

[105]:

	artist_name	track_name	\
0	Giuseppe Verdi	Cortigiani, vil razza dannata - Live On The Ed…	
1	Noah Kahan	Your Needs, My Needs - Live from Red Rocks '23	
2	Noah Kahan	Your Needs, My Needs - Live from Red Rocks '23	
3			
4	Andree Right Hand	Dân Chơi Sao Phải Khóc	
…	…	…	
5995	Isabel LaRosa	i'm yours - Live from Irving Plaza	
5996	Pedro Capó	5 y 3 - Live Performance	
5997	AnnenMayKantereit	Lass es kreisen - Live in Köln	
5998	Little Baby Bum	Hush Little Baby	

```
5999         Costa Gold                                          Lembra Dessa

                    track_id release_date  track_popularity  \
0      4xTQrn63kUDmKXBIJupxVb   2024-01-01                 3
1      51BSR67RxiZySQI61sFXQa   2024-01-04                 9
2      51BSR67RxiZySQI61sFXQa   2024-01-04                 9
3      5eY7692tmgHB9dbmq6wa2M   2024-01-05                73
4      0dBKcPEAsdxWJsqNDNHcPz   2024-01-05                56
…                         …          …                 …
5995   3NpVUgii1cGAl3pze7CFzb   2023-05-11                13
5996   0IdDYDdJdG83RkjlROhDHv   2023-06-23                12
5997   2CavjEfW3xt5enUOUJNys8   2023-03-10                13
5998   32w2lmKGmMtSOgKVRQBvLj   2023-09-01                12
5999   3NB5u20YOFB5ZTOWp0D9w2   2023-01-20                13

                   artist_id  artist_popularity  \
0      1JOQXgYdQV2yfrhewqx96o                 63
1      2RQXRUsr4IW1f3mKyKsy4B                 84
2      2RQXRUsr4IW1f3mKyKsy4B                 84
3      4UK2Lzi6fBfUi9rpDt6cik                 73
4      4grjJqg7iwQ8RKHs8d9Snh                 56
…                          …                …
5995   5arKwJZEvT5uKq4o0JfqR4                 73
5996   4QVBYiagIaa6ZGSPMbybpy                 67
5997   23xqmJEN3oVxwzqtNIyR5m                 68
5998   6b4gOldpp7HOBjFOlTfleW                  3
5999   7q1aEytv83jXNECmyaMhgn                 58

                                          artist_genres  artist_followers  \
0                                     [opera, classical]            723072
1                                                     []           4582855
2                                                     []           4582855
3                  [j-pop, j-rock, japanese indie, anime]           3358594
4                  [vietnamese hip hop, v-pop, vinahouse]            448951
…                                                     …                …
5995                                                  []           2646491
5996                                  [latin pop, latin]           1753779
5997       [german indie, german pop, german indie pop]           2206451
5998                                                  []                12
5999  [brazilian hip hop, brazilian trap, boom bap, …           2410505

      prev_track_popularity
0                       NaN
1                       NaN
2                       9.0
3                       NaN
4                       NaN
```

```
…                       …
5995                    NaN
5996                    NaN
5997                    NaN
5998                    NaN
5999                    NaN

[6000 rows x 10 columns]
```

[109]:
```
track_df.sort_values("release_date", inplace=True)
track_df['release_date'] = pd.to_datetime(track_df['release_date'],␣
 ↪errors='coerce')
track_df.reset_index(drop=True)
track_df
```

[109]:
```
          artist_name                                  track_name  \
5935          Quevedo                                      Buenas
5172     Jackson Wang                                        Slow
5612  ShowMinorSavage  Thinkin' bout you - from BMSG TYO SESSION
5837     Buba Espinho                  É Tão Grande O Alentejo
5016         NewJeans                                         OMG
…                …                                           …
3995        ANDROMEDA                              MONTAGEM COMA
3996          KNEECAP                                Love Making
3997            bbno$                                  bing bong
3998   Nouvelle Vague                                      Shout
3999        Dayseeker                                 My Immortal

                     track_id release_date  track_popularity  \
5935  0urCgPtrxPYqbTVxzgwlqW   2023-01-01                15
5172  0zI9O5dQ614A9mplzriDT0   2023-01-01                22
5612  6jX9r1YTZWE4jicyZ7rwEE   2023-01-01                29
5837  7zaAd2XAGjiU7McAGZaiZ1   2023-01-01                24
5016  65FftemJ1DbbZ45DUfHJXE   2023-01-02                79
…                        …          …                 …
3995  1aPVNfAor4sRSzqT1QCi43          NaT                30
3996  1mEvR3z3WiOWa0KQZuWYT0          NaT                13
3997  6vS8BPkw7gHzP9hYvRVRBP          NaT                32
3998  7tbfNcsWDvL559Bfls7Cpu          NaT                12
3999  5M1wH7BBNhwZFDoaC9J1BL          NaT                17

                   artist_id  artist_popularity              artist_genres  \
5935  52iwsT98xCoGgiGntTiR7K                 84                         []
5172  1kfWoWgCugPkyxQP8lkRlY                 63                    [k-pop]
5612  0x7bC0gbku4Mu85Pv6wYva                 39                    [j-r&b]
5837  3B6lsop4CFEdGzlCksiL6R                 52                     [fado]
5016  6HvZYsbFfjnjFrWF950C9d                 79                    [k-pop]
```

```
…                       …                  …                          …
3995  1RklNDIiYVZ3dYdEUnB0cS                 68    [phonk, brazilian phonk]
3996  1ZVACPeq7ccGCoUXwtafUU                 64                          []
3997  41X1TR6hrK8Q2ZCpp2EqCz                 79                          []
3998  4h7NLIlg1oYdEtfQJfyto0                 57                    [lounge]
3999  5FjQVp1Lb0kltmwIuu5kfj                 68  [metalcore, post-hardcore]

      artist_followers  prev_track_popularity
5935           6578632                    NaN
5172           4298601                    NaN
5612             31832                    NaN
5837             42519                    NaN
5016          12013240                    NaN
…                  …                      …
3995             80399                    NaN
3996            476814                    NaN
3997           2925608                    NaN
3998            423437                    NaN
3999            526332                    NaN

[6000 rows x 10 columns]
```

[111]:
```
track_df['prev_track_popularity'] = (
    track_df
    .groupby('artist_id')['track_popularity']
    .shift(1)
)
track_df
```

[111]:
```
          artist_name                                  track_name  \
5935          Quevedo                                     Buenas
5172     Jackson Wang                                       Slow
5612  ShowMinorSavage  Thinkin' bout you - from BMSG TYO SESSION
5837     Buba Espinho                  É Tão Grande O Alentejo
5016         NewJeans                                         OMG
…                …                                           …
3995        ANDROMEDA                              MONTAGEM COMA
3996          KNEECAP                                Love Making
3997            bbno$                                  bing bong
3998   Nouvelle Vague                                      Shout
3999        Dayseeker                                 My Immortal

                    track_id release_date  track_popularity  \
5935  0urCgPtrxPYqbTVxzgwlqW   2023-01-01                15
5172  0zI9O5dQ614A9mplzriDT0   2023-01-01                22
5612  6jX9r1YTZWE4jicyZ7rwEE   2023-01-01                29
5837  7zaAd2XAGjiU7McAGZaiZ1   2023-01-01                24
```

```
5016  65FftemJ1DbbZ45DUfHJXE   2023-01-02               79
…                        …          …                …
3995  1aPVNfAor4sRSzqT1QCi43          NaT               30
3996  1mEvR3z3WiOWa0KQZuWYT0          NaT               13
3997  6vS8BPkw7gHzP9hYvRVRBP          NaT               32
3998  7tbfNcsWDvL559Bfls7Cpu          NaT               12
3999  5M1wH7BBNhwZFDoaC9J1BL          NaT               17

                     artist_id  artist_popularity               artist_genres  \
5935  52iwsT98xCoGgiGntTiR7K                 84                          []
5172  1kfWoWgCugPkyxQP8lkRlY                 63                      [k-pop]
5612  0x7bC0gbku4Mu85Pv6wYva                 39                      [j-r&b]
5837  3B6lsop4CFEdGzlCksiL6R                 52                       [fado]
5016  6HvZYsbFfjnjFrWF950C9d                 79                      [k-pop]
…                        …                …                          …
3995  1RklNDIiYVZ3dYdEUnB0cS                 68    [phonk, brazilian phonk]
3996  1ZVACPeq7ccGCoUXwtafUU                 64                          []
3997  41X1TR6hrK8Q2ZCpp2EqCz                 79                          []
3998  4h7NLIlg1oYdEtfQJfyto0                 57                    [lounge]
3999  5FjQVp1Lb0kltmwIuu5kfj                 68  [metalcore, post-hardcore]

      artist_followers  prev_track_popularity
5935           6578632                    NaN
5172           4298601                    NaN
5612             31832                    NaN
5837             42519                    NaN
5016          12013240                    NaN
…                  …                      …
3995             80399                   30.0
3996            476814                    NaN
3997           2925608                    NaN
3998            423437                    NaN
3999            526332                    NaN

[6000 rows x 10 columns]
```

[132]:
```
track_df.isna().sum()
```

[132]:
```
artist_name           0
track_name            0
track_id              0
release_date          5
track_popularity      0
artist_id             0
artist_popularity     0
artist_genres         0
artist_followers      0
```

```
prev_track_popularity       1496
popularity_centered            0
prev_popularity_centered    1496
dtype: int64
```

[133]:
```
track_df['popularity_centered'] = (
    track_df['track_popularity']
    - track_df.groupby('artist_id')['track_popularity'].transform('mean')
)
track_df['prev_popularity_centered'] = (
    track_df['prev_track_popularity']
    - track_df
      .groupby('artist_id')['prev_track_popularity']
      .transform('mean')
)
display(track_df.head(5))
display(track_df.tail(5))
```

```
         artist_name                                  track_name  \
5935          Quevedo                                      Buenas
5172     Jackson Wang                                        Slow
5612  ShowMinorSavage  Thinkin' bout you - from BMSG TYO SESSION
5837     Buba Espinho                   É Tão Grande O Alentejo
5016         NewJeans                                         OMG

                  track_id release_date  track_popularity  \
5935  0urCgPtrxPYqbTVxzgwlqW   2023-01-01                15
5172  0zI9O5dQ614A9mplzriDT0   2023-01-01                22
5612  6jX9r1YTZWE4jicyZ7rwEE   2023-01-01                29
5837  7zaAd2XAGjiU7McAGZaiZ1   2023-01-01                24
5016  65FftemJ1DbbZ45DUfHJXE   2023-01-02                79

                 artist_id  artist_popularity artist_genres  \
5935  52iwsT98xCoGgiGntTiR7K                 84            []
5172  1kfWoWgCugPkyxQP8lkRlY                 63       [k-pop]
5612  0x7bC0gbku4Mu85Pv6wYva                 39       [j-r&b]
5837  3B6lsop4CFEdGzlCksiL6R                 52        [fado]
5016  6HvZYsbFfjnjFrWF950C9d                 79       [k-pop]

      artist_followers  prev_track_popularity  popularity_centered  \
5935           6578632                    NaN           -32.500000
5172           4298601                    NaN             0.000000
5612             31832                    NaN            -2.400000
5837             42519                    NaN             0.000000
5016          12013240                    NaN            11.428571

      prev_popularity_centered
5935                       NaN
```

```
5172                      NaN
5612                      NaN
5837                      NaN
5016                      NaN
```

```
        artist_name     track_name                track_id release_date  \
3995       ANDROMEDA  MONTAGEM COMA  1aPVNfAor4sRSzqT1QCi43          NaT
3996         KNEECAP    Love Making  1mEvR3z3WiOWaOKQZuWYTO          NaT
3997           bbno$      bing bong  6vS8BPkw7gHzP9hYvRVRBP          NaT
3998  Nouvelle Vague          Shout  7tbfNcsWDvL559Bfls7Cpu          NaT
3999       Dayseeker    My Immortal  5M1wH7BBNhwZFDoaC9J1BL          NaT

      track_popularity               artist_id  artist_popularity  \
3995                30  1RklNDIiYVZ3dYdEUnB0cS                 68
3996                13  1ZVACPeq7ccGCoUXwtafUU                 64
3997                32  41X1TR6hrK8Q2ZCpp2EqCz                 79
3998                12  4h7NLIlg1oYdEtfQJfyto0                 57
3999                17  5FjQVp1Lb0kltmwIuu5kfj                 68

                  artist_genres  artist_followers  prev_track_popularity  \
3995    [phonk, brazilian phonk]             80399                   30.0
3996                          []            476814                    NaN
3997                          []           2925608                    NaN
3998                    [lounge]            423437                    NaN
3999  [metalcore, post-hardcore]            526332                    NaN

      popularity_centered  prev_popularity_centered
3995                  0.0                       0.0
3996                  0.0                       NaN
3997                  0.0                       NaN
3998                  0.0                       NaN
3999                  0.0                       NaN
```

[134]:
```
import statsmodels.api as sm
from scipy.stats import pearsonr, spearmanr

from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.model_selection import GroupKFold

from datetime import datetime
```

[135]:
```
baseline_df = track_df[
    ['popularity_centered', 'prev_popularity_centered']
].dropna()
```

[139]:
```
y_true = baseline_df['popularity_centered']
y_pred_prev = baseline_df['prev_popularity_centered']
```

```
rmse_prev = np.sqrt(
    mean_squared_error(y_true, y_pred_prev)
)
rmse_prev
```

[139]:
```
np.float64(11.69888479065329)
```

[140]:
```
y_pred_zero = np.zeros(len(y_true))

rmse_zero = np.sqrt(
    mean_squared_error(y_true, y_pred_zero)
)

rmse_zero, rmse_prev
```

[140]:
```
(np.float64(11.937183881611698), np.float64(11.69888479065329))
```

[142]:
```
eval_df = track_df[
    ['artist_id', 'popularity_centered', 'prev_popularity_centered']
].dropna()
```

[144]:
```
X = eval_df[['prev_popularity_centered']]
y = eval_df['popularity_centered']
groups = eval_df['artist_id']
```

[145]:
```
gkf = GroupKFold(n_splits=5)
```

[146]:
```
rmse_zero = []
rmse_prev = []
rmse_lm = []
r2_lm = []
```

[150]:
```
for train_idx, test_idx in gkf.split(X, y, groups):
    X_train, X_test = X.iloc[train_idx], X.iloc[test_idx]
    y_train, y_test = y.iloc[train_idx], y.iloc[test_idx]

    # Baseline 0:   X
    y_pred_zero = np.zeros(len(y_test))
    rmse_zero.append(
        np.sqrt(mean_squared_error(y_test, y_pred_zero))
    )

    # Baseline 1:       +
    y_pred_prev = X_test['prev_popularity_centered']
    rmse_prev.append(
        np.sqrt(mean_squared_error(y_test, y_pred_prev))
```

```
    )

    # Regression model
    model = LinearRegression()
    model.fit(X_train, y_train)

    y_pred_lm = model.predict(X_test)
    rmse_lm.append(
        np.sqrt(mean_squared_error(y_test, y_pred_lm))
    )
    r2_lm.append(
        r2_score(y_test, y_pred_lm)
    )
```

[151]:
```
print(f"Baseline (0) RMSE: {np.mean(rmse_zero):.3f}")
print(f"Baseline (prev) RMSE: {np.mean(rmse_prev):.3f}")
print(f"Lag Regression RMSE: {np.mean(rmse_lm):.3f}")
print(f"Lag Regression R²: {np.mean(r2_lm):.3f}")
```

```
Baseline (0) RMSE: 11.846
Baseline (prev) RMSE: 11.537
Lag Regression RMSE: 10.083
Lag Regression R²: 0.273
```

- , .
- baseline 15% .

1.1 , feature ?

[ ]:
```
track_df['artist_followers_log'] = np.log1p(track_df['artist_followers'])
```

[ ]:
```
track_df['main_genre'] = track_df['artist_genres'].apply(
    lambda x: x[0] if isinstance(x, list) and len(x) > 0 else 'unknown'
)
```

[176]:
```
#
genre_dummies = pd.get_dummies(track_df['main_genre'], prefix='genre')
track_df = pd.concat([track_df, genre_dummies], axis=1)
```

[177]:
```
feature_cols = [
    'prev_popularity_centered',
    'artist_followers_log',
    'artist_popularity'
] + list(genre_dummies.columns)
```

[178]:
```
eval_df = track_df[
    ['artist_id', 'popularity_centered'] + feature_cols
].dropna()
```

[179]:
```
X = eval_df[feature_cols]
y = eval_df['popularity_centered']
groups = eval_df['artist_id']
```

[180]:
```
gkf = GroupKFold(n_splits=5)

rmse_list = []
r2_list = []
```

[181]:
```
for train_idx, test_idx in gkf.split(X, y, groups):
    X_train, X_test = X.iloc[train_idx], X.iloc[test_idx]
    y_train, y_test = y.iloc[train_idx], y.iloc[test_idx]

    model = LinearRegression()
    model.fit(X_train, y_train)

    y_pred = model.predict(X_test)

    rmse_list.append(
        np.sqrt(mean_squared_error(y_test, y_pred))
    )
    r2_list.append(
        r2_score(y_test, y_pred)
    )
```

[182]:
```
print(f"Extended model RMSE: {np.mean(rmse_list):.3f}")
print(f"Extended model R²: {np.mean(r2_list):.3f}")
```

```
Extended model RMSE: 10.120
Extended model R²: 0.267
```

[ ]:
```

```