Billboard Music Trend

Chia-Wen Chen, Jyun-Hao Chen, Shih-Man Wang

Team Members

Mickey Chen 🐭

Jyun-Hao Chen

Shih-Man Wang

📜 Agenda

01 Background

02 Research Questions

03 Derivative Data

04 Analysis Approach

05 Findings

01 Background

🎧 Motivation

- Significant role in our daily lives.

- Highly profitable industry.

- The transition of popular music and identify the types of music people enjoy more today.

02 Research Questions

🎯 Research Questions

1. What are the most frequently charted music genres in history, and how have music attributes changed in a period?

2. Is there a correlation between music attributes and popularity on Billboard? And the prediction of music trends for the 2020s.

3. Do audiences prefer fixed songs, or are new songs introduced weekly?

03 Derivative data

Derivative Data

Derivative Data

```
df_musicfactors_trends.head()
```

	week_id	danceability	energy	loudness	speechiness	acousticness	instrumentalness	liveness	valence	tempo
0	2020-05-02	0.696938	0.647797	-5.826313	0.117678	0.183440	0.005359	0.185391	0.538859	121.486109
1	2020-05-09	0.684484	0.649797	-5.826937	0.102336	0.174324	0.005352	0.187403	0.517977	119.390906
2	2020-05-16	0.682131	0.644623	-5.907557	0.117861	0.211857	0.005612	0.196775	0.511180	118.065852
3	2020-05-23	0.677696	0.639393	-5.930232	0.093173	0.194417	0.006131	0.180252	0.510527	119.614250
4	2020-05-30	0.714136	0.626780	-6.353305	0.143324	0.185313	0.005829	0.183998	0.488729	121.897237

04 Analysis Approach

Narrative Analysis

Multiple Linear Regression

Narrative Analysis

1. **Separate "Genre" & 11 Music Attributes**

   They have different data types.

   - Genre: String

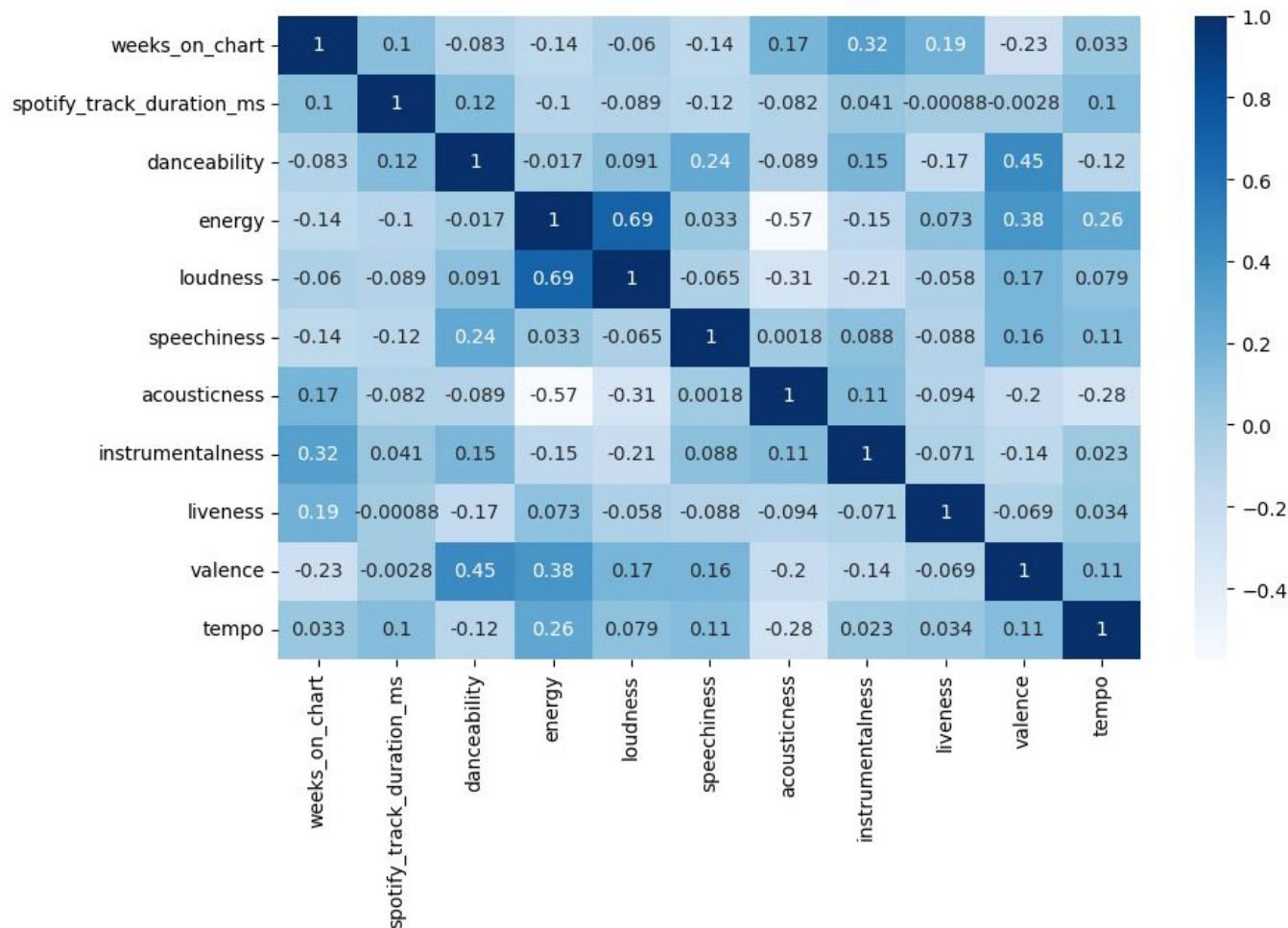   - 11 Music Attributes: Float type. Range of 0 to 1.

2. **Limit duration from 2020 to 2021.**

   Too much data and hard to see the trends by weeks.

Multiple Linear Regression Model

Goal: Determine the correlation between Attributes and weeks_on_chart.
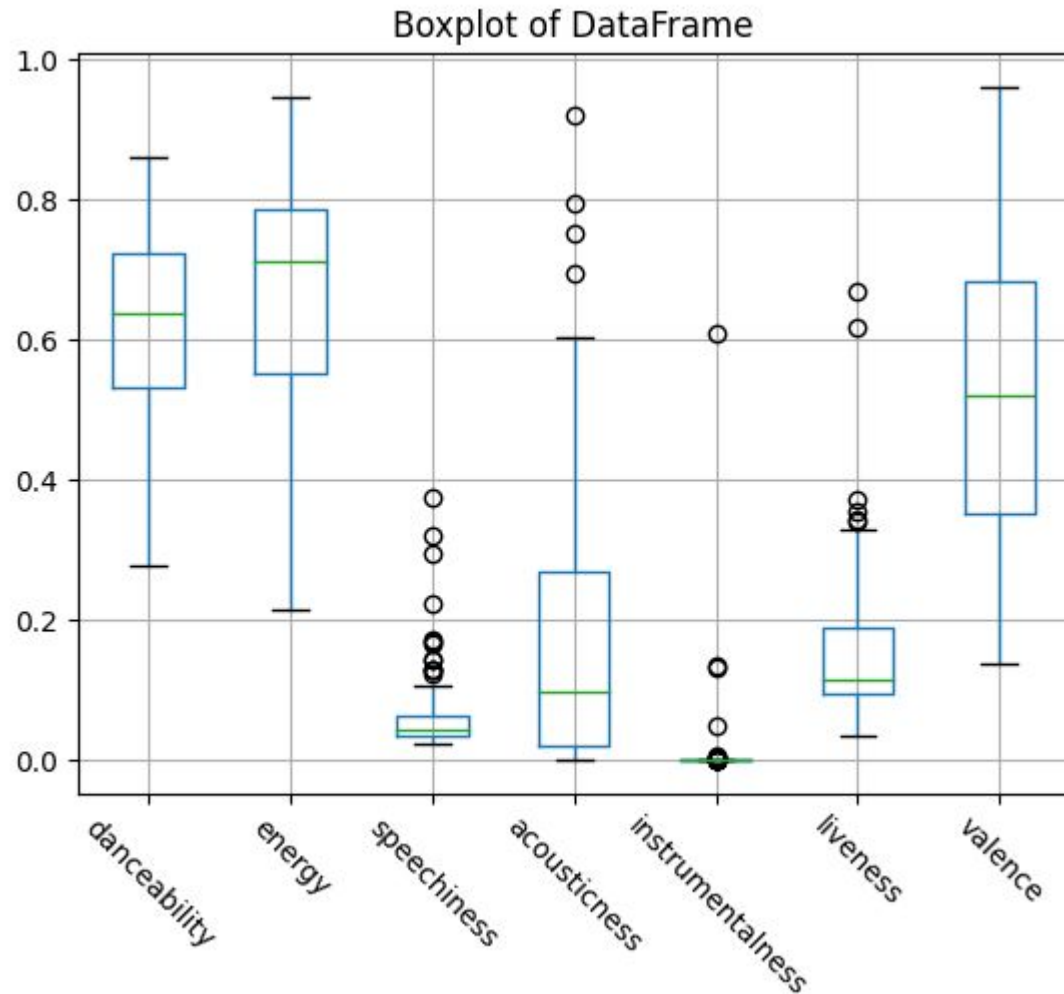
1. Check relationships between the attributes.

Correlation between independent variables is low.
>> No Multicollinearity.

Multiple Linear Regression Model

2. Identify outliers again.

Boxplot of DataFrame

Outliers in: instrumentalness, liveness, and acousticness.

To execute accurate prediction

>> Remain the outlier and not delete.

Multiple Linear Regression Model

3. Develop model for independent variables and dependent variables.

- Independent Variables : Samples = 1:10
- Apply multiple linear regression model to generate result summary.
  >>For advanced summarization & Analysis of results.

```python
aa = df_mr.sort_values(by=['weeks_on_chart'], ascending=False).dropna(subset=['liveness','spotify_track_duration_ms', 'spotify_genre']).head(100)
```

```python
aa = aa.drop(columns=['week_id','song','performer','song_id','spotify_genre','spotify_track_explicit','key','mode','time_signature'])
```

```python
x = sm.add_constant(aa.iloc[:,1:])
y = aa['weeks_on_chart']
model = sm.OLS(y, x)
result = model.fit()
result.summary()
```

05 Findings

🔍 Music attributes and Billboard rank don't have distinct relevance.

```
                          OLS Regression Results
 Dep. Variable:     weeks_on_chart      R-squared:        0.229
       Model:       OLS              Adj. R-squared:      0.142
      Method:       Least Squares        F-statistic:     2.637
        Date:       Wed, 26 Apr 2023  Prob (F-statistic): 0.00734
        Time:       20:11:23          Log-Likelihood:    -320.58
No. Observations:   100                  AIC:             663.2
 Df Residuals:      89                   BIC:             691.8
   Df Model:        10
Covariance Type:    nonrobust
```

	coef	std err	t	P>|t|	[0.025	0.975]
const	53.0083	10.372	5.111	0.000	32.399	73.618
spotify_track_duration_ms	1.748e-05	2.01e-05	0.872	0.386	-2.24e-05	5.73e-05
danceability	-0.7473	6.425	-0.116	0.908	-13.514	12.020
energy	-3.5542	7.811	-0.455	0.650	-19.074	11.966
loudness	0.4504	0.474	0.950	0.345	-0.492	1.393
speechiness	-12.7439	11.371	-1.121	0.265	-35.338	9.850
acousticness	4.9271	3.940	1.250	0.214	-2.903	12.757
instrumentalness	34.6931	10.895	3.184	0.002	13.044	56.342
liveness	13.7441	6.083	2.259	0.026	1.657	25.831
valence	-3.8104	4.030	-0.945	0.347	-11.818	4.197
tempo	0.0217	0.026	0.833	0.407	-0.030	0.074

```
   Omnibus:       48.226  Durbin-Watson:     0.475
Prob(Omnibus):    0.000   Jarque-Bera (JB):  127.572
     Skew:        1.791       Prob(JB):       1.99e-28
   Kurtosis:      7.218      Cond. No.        5.12e+06
```

The independent variables can only explain 22% relationship of the weeks_on_chart.

>> 78% of elements are not found.

"Const", "liveness", "instrumentalness" have a p-value < 0.05.

>> Significant differences from the dependent variable.

17

🔍 Music attributes and Billboard rank don't have distinct relevance.

```
                        OLS Regression Results
        Dep. Variable:   weeks_on_chart      R-squared:       0.146
              Model:     OLS             Adj. R-squared:      0.129
             Method:     Least Squares       F-statistic:     8.306
               Date:     Wed, 26 Apr 2023  Prob (F-statistic): 0.000468
               Time:     20:11:24         Log-Likelihood:     -325.65
    No. Observations:    100                      AIC:        657.3
        Df Residuals:    97                       BIC:        665.1
            Df Model:    2
       Covariance Type:  nonrobust

                           coef   std err     t      P>|t|   [0.025  0.975]
              const       52.2948 1.137    46.007   0.000   50.039  54.551
    instrumentalness      35.7923 10.117   3.538    0.001   15.713  55.872
           liveness       13.4826 5.939    2.270    0.025   1.695   25.270

           Omnibus:       50.779   Durbin-Watson:      0.326
     Prob(Omnibus):       0.000   Jarque-Bera (JB):    137.679
              Skew:       1.890        Prob(JB):        1.27e-30
          Kurtosis:       7.330        Cond. No.        16.1

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```
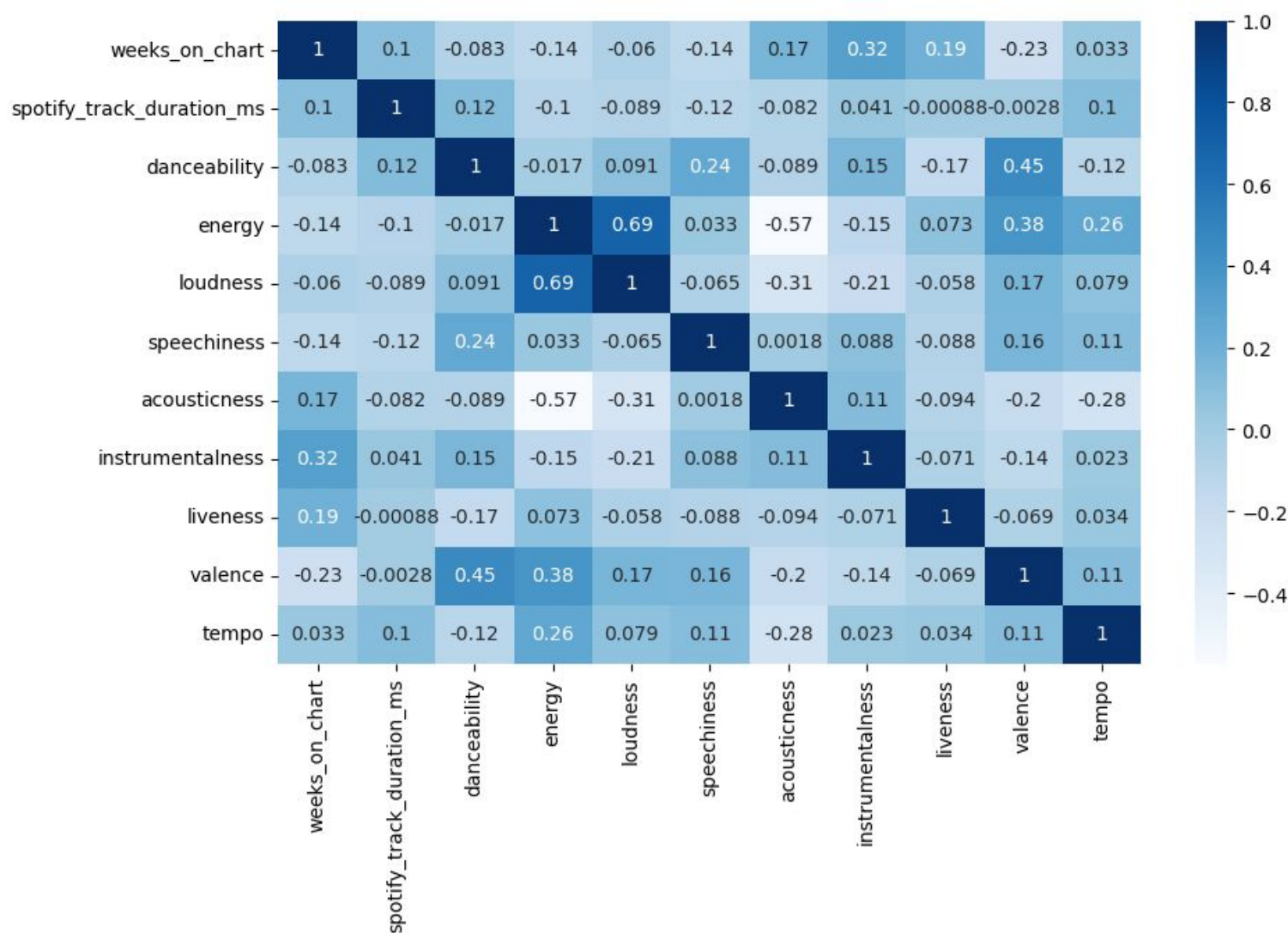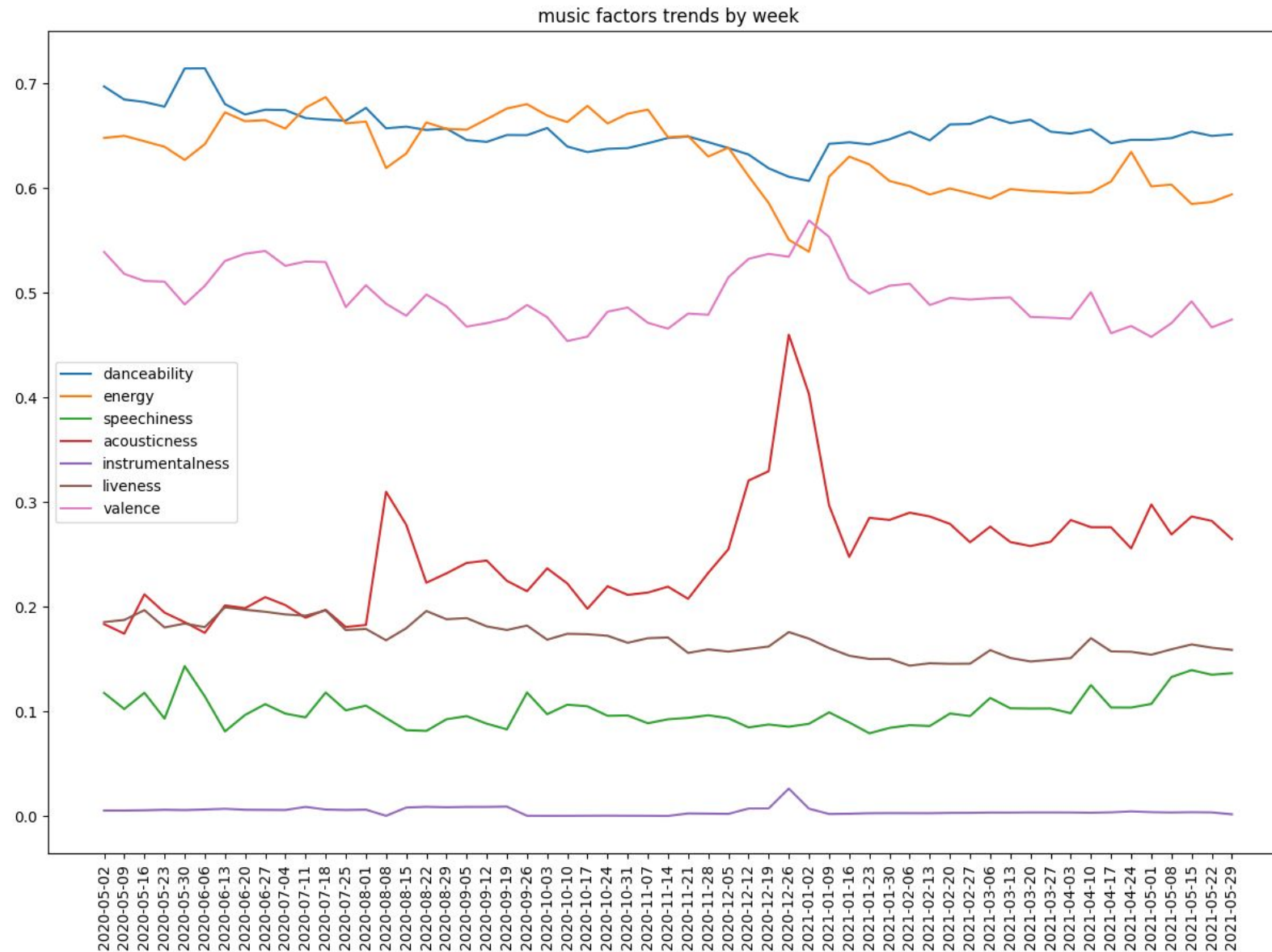
Remove the variables that do not have significant difference.

Only explain 14.6% reasons of weeks_on_chart.

>> R^ decline as lesser independent variables are in the model.

🔍 Music attributes and Billboard rank don't have distinct relevance.

🔍 **Music attributes and Billboard rank don't have distinct relevance.**
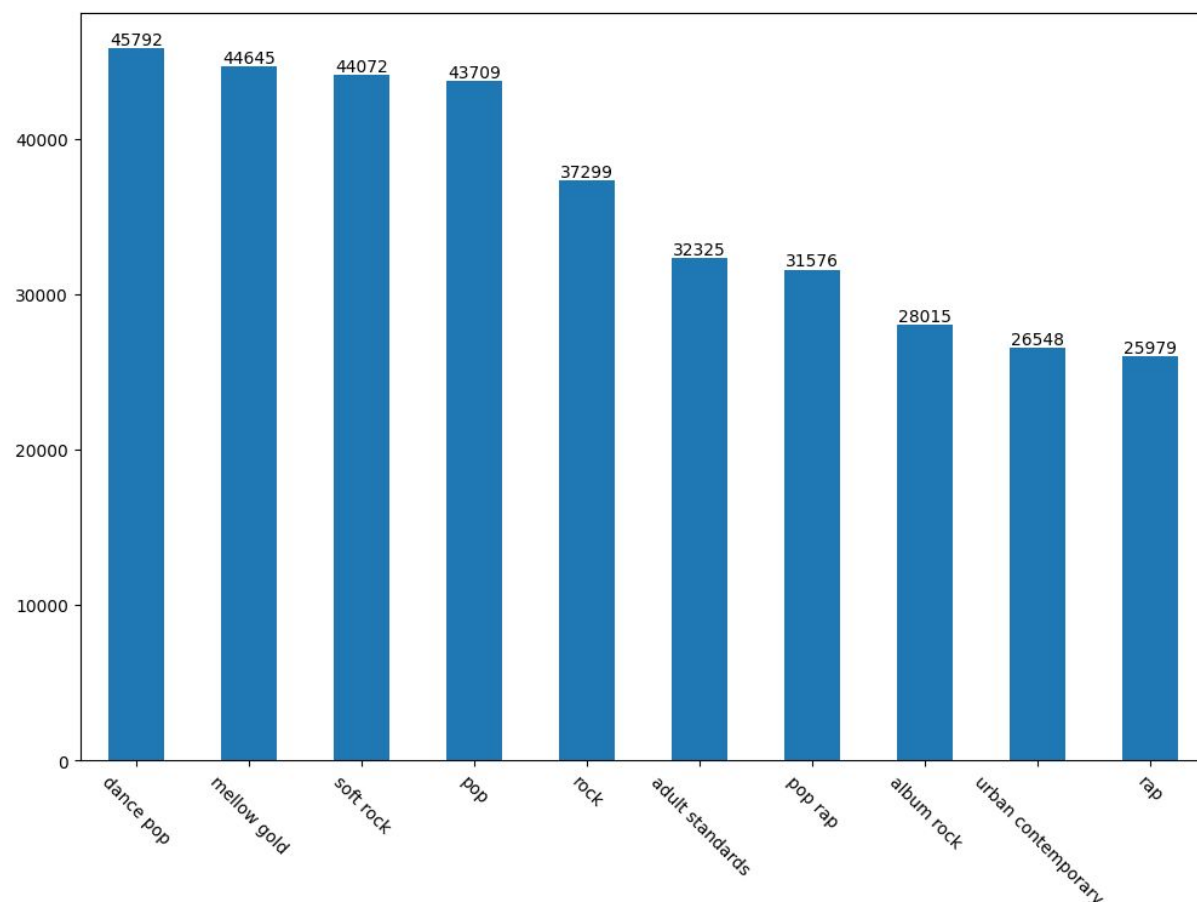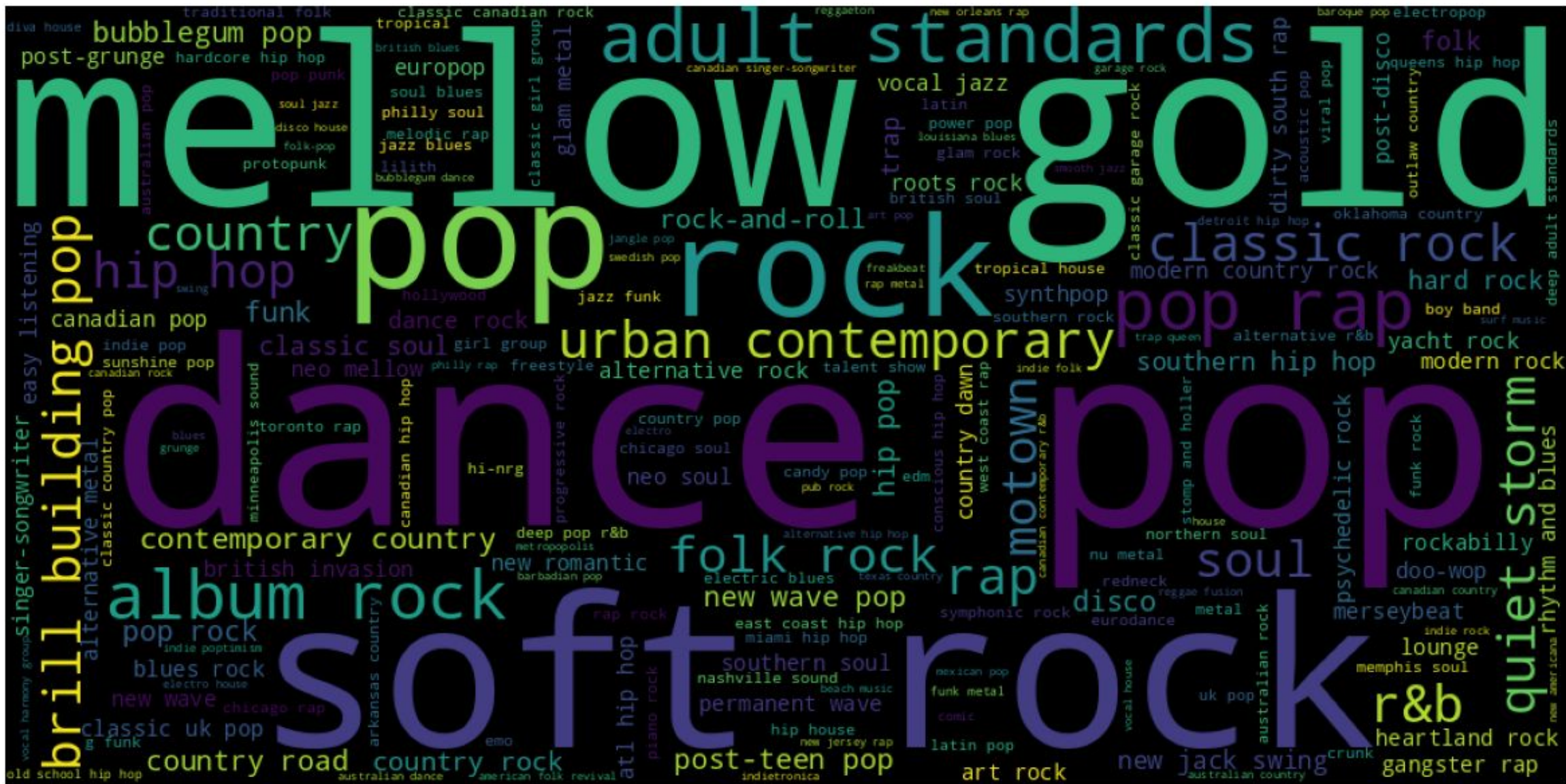

music factors trends by week

🔍 Music genre that people love the most

1. Separate music genres from the dataset [Mellow gold, pop, classic]
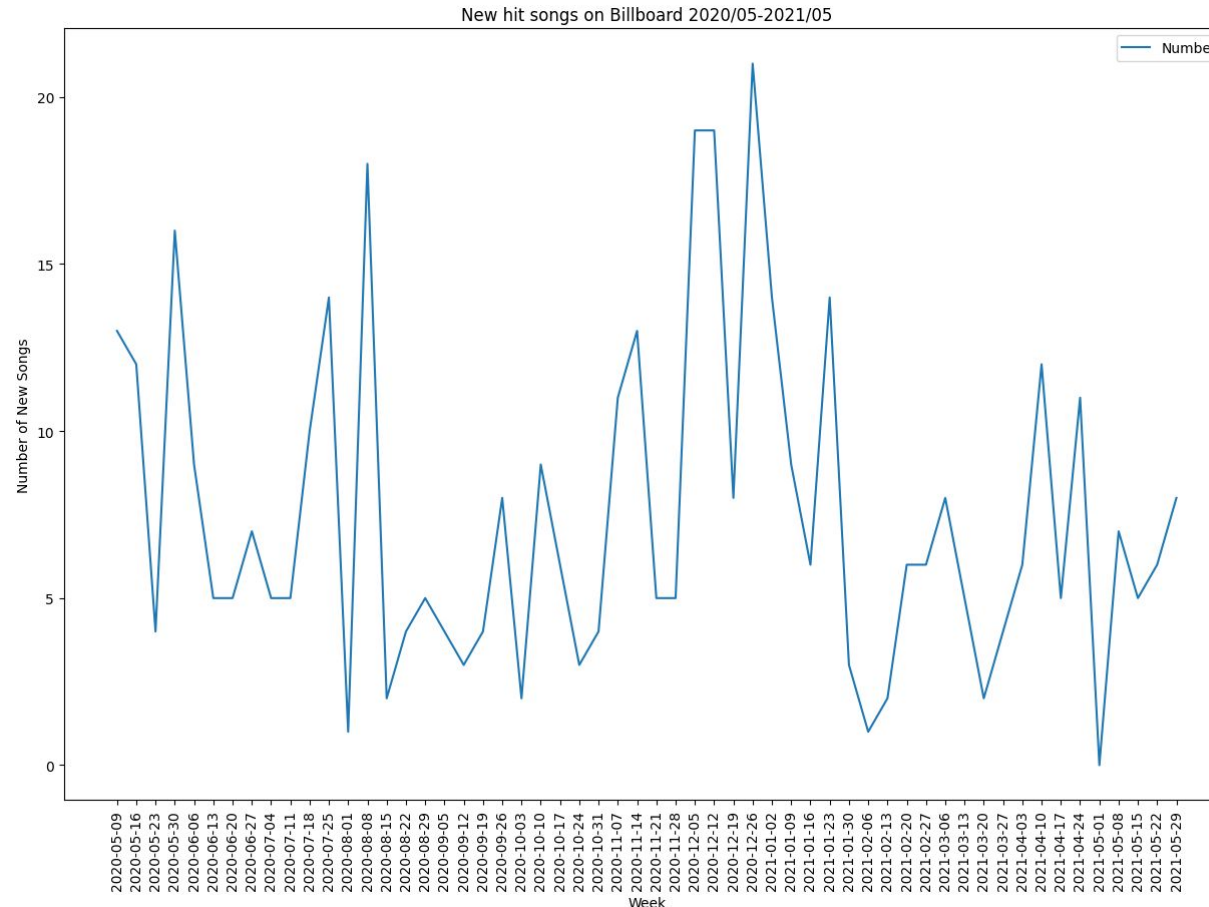
2. Calculate the number of different genres

🔍 **People love new songs but also stick to old songs**

1. Record the song_id that appears each week.

2. Compare and calculate the difference.

New hit songs on Billboard 2020/05-2021/05

Thank you for your listening!

Chia-Wen Chen, Jyun-Hao Chen, Shih-Man Wang