Muhammed Turhan 28840

Project Report

1. **Exporting CSV and Converting to Pandas DataFrame:**
   - The project begins with a section on exporting data and converting it to a Pandas DataFrame, which is a crucial step for data analysis.
2. **Data Import:**
   - This Project gets it data from chosic.com as csv file as it was used in our earlier assignment. There is no other tools or applications are used for data importing process.
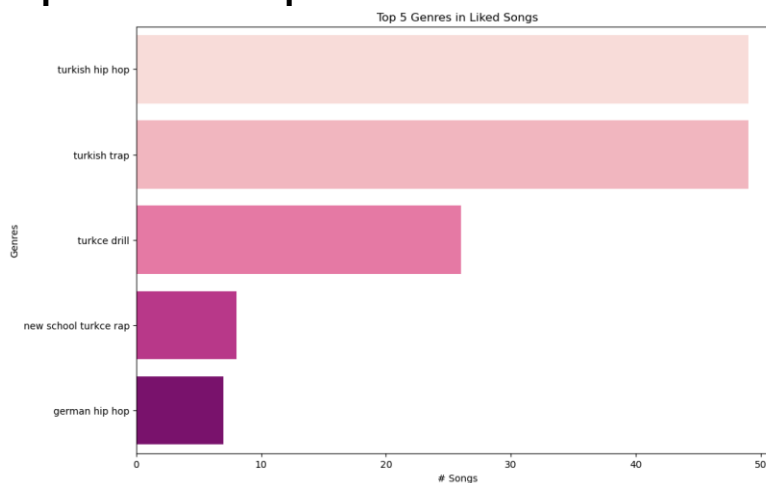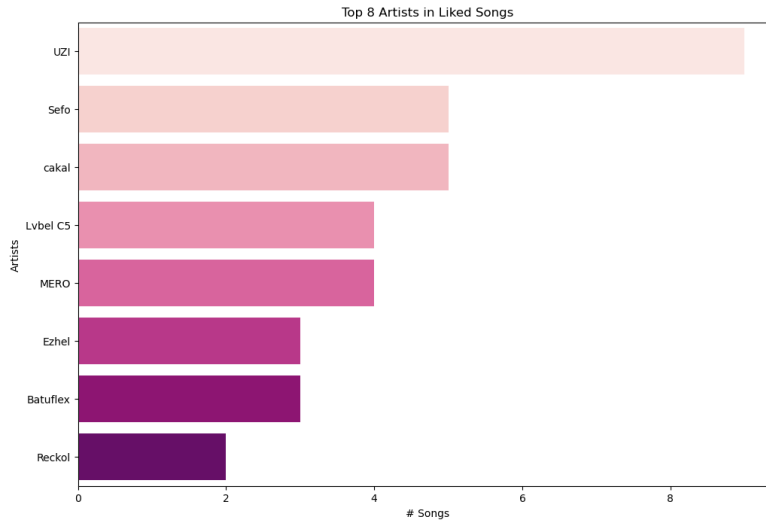3. **CSV to DataFrame:**
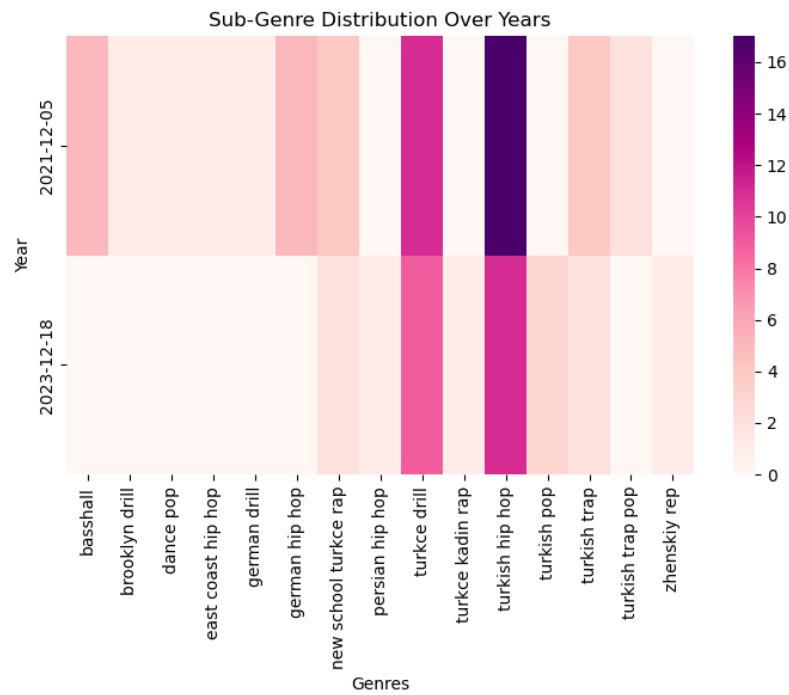   - The data is loaded from a CSV file into a Pandas DataFrame.
4. **Data Analysis:**
   - The notebook includes a section titled "Data Analysis," indicating a focus on examining the Spotify data.
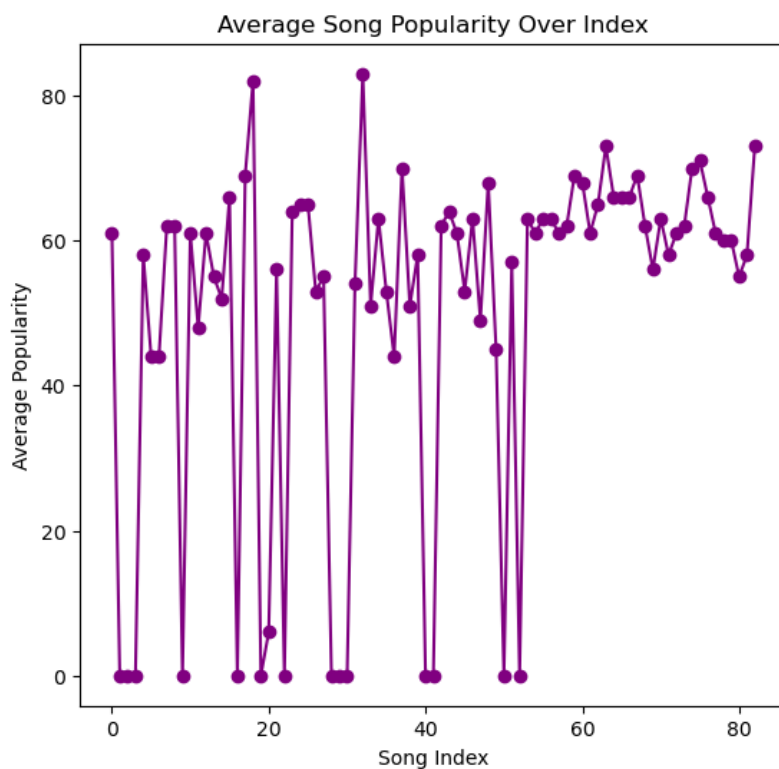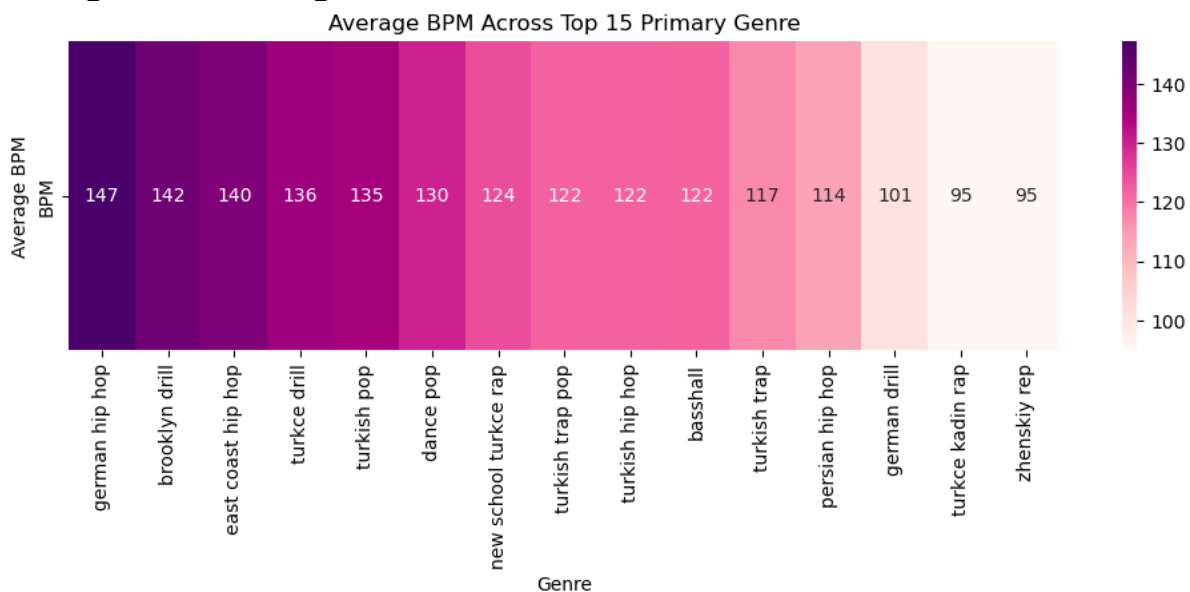5. **Top 5 Genres & Top Artists:**

Top 8 Artists in Liked Songs

6. Genre Distribution over adding time:

Sub-Genre Distribution Over Years

7. Average Popularity Over Time:

Average Song Popularity Over Index

8. Average BPM across genres:

Average BPM Across Top 15 Primary Genre

german hip hop	brooklyn drill	east coast hip hop	turkce drill	turkish pop	dance pop	new school turkce rap	turkish trap pop	turkish hip hop	basshall	turkish trap	persian hip hop	german drill	turkce kadin rap	zhenskiy rep
147	142	140	136	135	130	124	122	122	122	117	114	101	95	95

9. Looking deeper into BPM-Loudness-Popularity Correlation:

Correlation Heatmap

Average BPM of Top 5 Artists

Average Popularity of Top 5 Artists

Average Loudness of Top 5 Artists
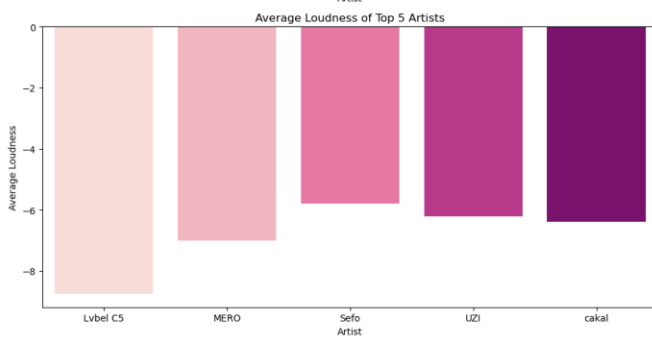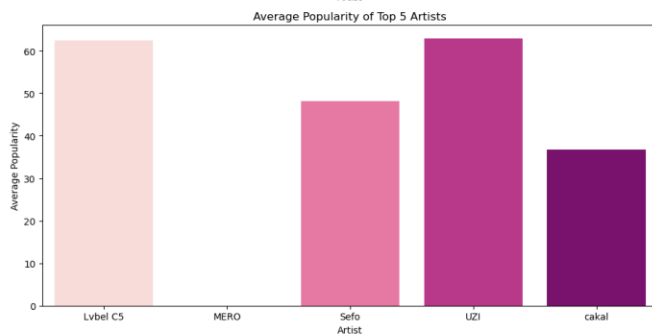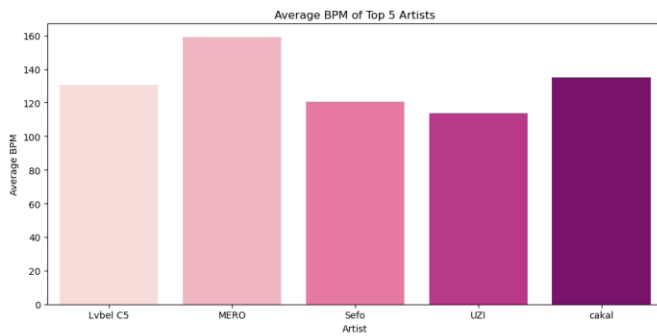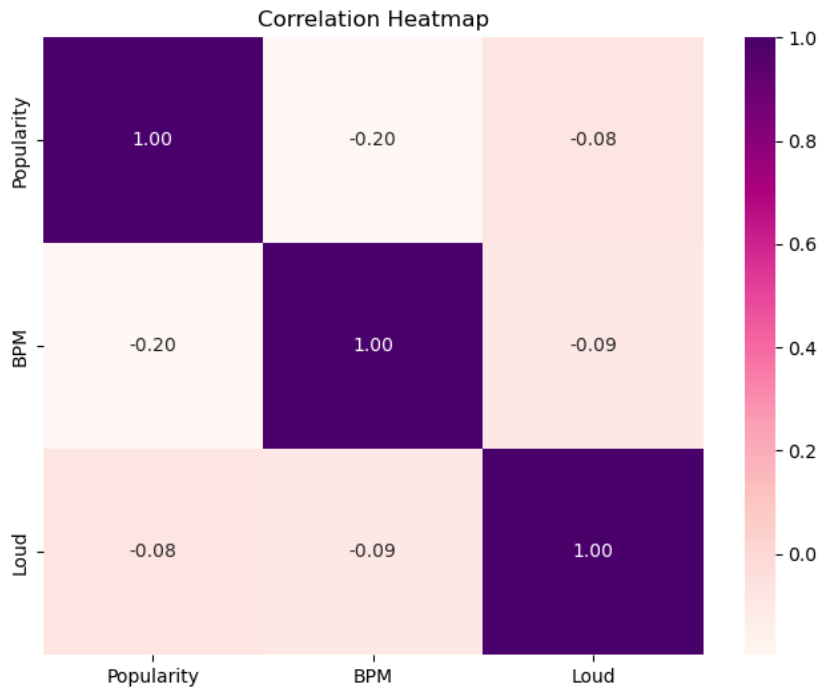
10. Implementing Machine Learning in order to create 10 song playlist matching my song taste:

- **First we labelled the data:**

```
df['Top Song'] = df.apply(lambda x: 1 if (115 < x['BPM'] < 155) and
            ('Hip Hop' in str(x['Parent Genres'])) else 0, axis=1)
```

- **Then we trained the data with a 30% test size**

```
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y,
test_size=0.3, random_state=42)

X_train.shape, X_test.shape, y_train.shape, y_test.shape
```

- **After training, achieved these scores:**

```
Accuracy:  0.8
Confusion Matrix:
[[13  2]
 [ 3  7]]
Report:
              precision    recall  f1-score   support

           0       0.81      0.87      0.84        15
           1       0.78      0.70      0.74        10

    accuracy                           0.80        25
   macro avg       0.80      0.78      0.79        25
weighted avg       0.80      0.80      0.80        25
```

- **Finally, we created our playlist with a 1 false positive result(last of the playlist)**

As a Final Task - Creating 10 Song Playlist

The playlist includes 1 false positive top song just to match its accuracy, it is the last one :)

```python
test_indices = X_test.index

true_positives = test_indices[(y_pred == 1) & (y_test == 1)]
false_positives = test_indices[(y_pred == 1) & (y_test == 0)]

selected_songs_indices = true_positives.tolist() + false_positives[:1].tolist()

playlist = df.loc[selected_songs_indices, ['Song', 'Artist', 'BPM', 'Primary Genre']]

playlist
```

	Song	Artist	BPM	Primary Genre
30	GECELERİN DERDİ	BEGE	125	turkce drill
0	Dünya Hain	Batuflex	145	turkish hip hop
12	Hayrola	Artz,Bugy,Ezhel	151	turkish hip hop
49	Hayrane	Reo Manchs	140	new school turkce rap
70	İllegale	Batuflex	138	turkish hip hop
67	TALIBANA	DEHA INC.,Reckol	123	new school turkce rap
56	Sarıl Bana	Velet,Dilan	129	turkish hip hop
69	Perros Blancos	Reckol,cakal	140	turkce drill