Spotify Data Analysis Project

Objective: The objective of this project is to analyze Spotify song data and identify factors that influence song popularity using data visualization techniques.

Dataset: SpotifyFeatures.csv

Tools Used: Python, Pandas, Matplotlib, Seaborn, Jupyter Notebook

Research Question:

What factors influence the popularity of songs on Spotify?

This analysis will examine relationships between popularity and audio features such as energy, danceability, valence, tempo, and genre.

In [16]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

sns.set_style("whitegrid")
```

In [17]:
```
df = pd.read_csv("SpotifyFeatures.csv")

df.head()
```

Out[17]:

	genre	artist_name	track_name	track_id	popularity	acousticness
0	Movie	Henri Salvador	C'est beau de faire un Show	0BRjO6ga9RKCKjfDqeFgWV	0	0.611
1	Movie	Martin & les fées	Perdu d'avance (par Gad Elmaleh)	0BjC1NfoEOOusryehmNudP	1	0.246
2	Movie	Joseph Williams	Don't Let Me Be Lonely Tonight	0CoSDzoNIKCRs124s9uTVy	3	0.952
3	Movie	Henri Salvador	Dis-moi Monsieur Gordon Cooper	0Gc6TVm52BwZD07Ki6tIvf	0	0.703
4	Movie	Fabien Nataf	Ouverture	0IuslXpMROHdEPvSl1fTQK	4	0.950

Dataset Overview:

This dataset contains Spotify song features including genre, artist name, track name, popularity score, and various audio features such as energy, danceability, valence, tempo, and acousticness.

The dataset is used to analyze factors that influence song popularity.

In [18]:
```
df.shape
df.columns
df.info()
df.describe()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 232725 entries, 0 to 232724
Data columns (total 18 columns):
 #   Column            Non-Null Count   Dtype  
---  ------            --------------   -----  
 0   genre             232725 non-null  object 
 1   artist_name       232725 non-null  object 
 2   track_name        232724 non-null  object 
 3   track_id          232725 non-null  object 
 4   popularity        232725 non-null  int64  
 5   acousticness      232725 non-null  float64
 6   danceability      232725 non-null  float64
 7   duration_ms       232725 non-null  int64  
 8   energy            232725 non-null  float64
 9   instrumentalness  232725 non-null  float64
 10  key               232725 non-null  object 
 11  liveness          232725 non-null  float64
 12  loudness          232725 non-null  float64
 13  mode              232725 non-null  object 
 14  speechiness       232725 non-null  float64
 15  tempo             232725 non-null  float64
 16  time_signature    232725 non-null  object 
 17  valence           232725 non-null  float64
dtypes: float64(9), int64(2), object(7)
memory usage: 32.0+ MB
```

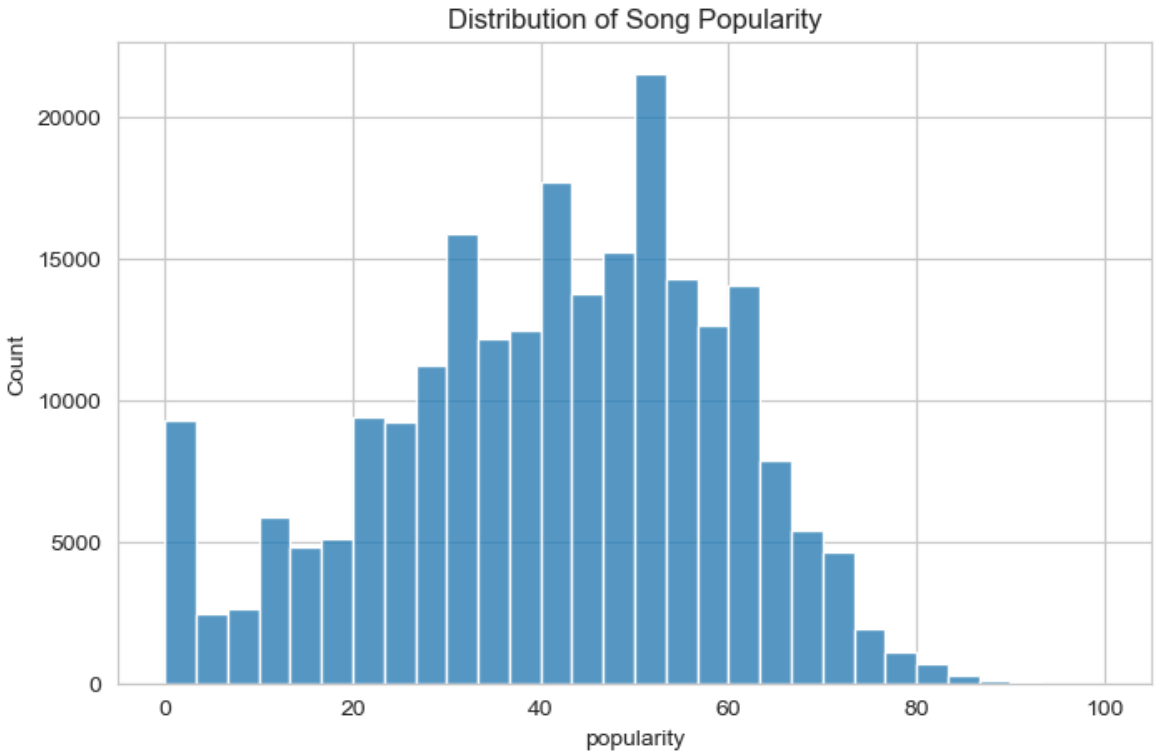
Out[18]:

	popularity	acousticness	danceability	duration_ms	energy	instr
count	232725.000000	232725.000000	232725.000000	2.327250e+05	232725.000000	23
mean	41.127502	0.368560	0.554364	2.351223e+05	0.570958	
std	18.189948	0.354768	0.185608	1.189359e+05	0.263456	
min	0.000000	0.000000	0.056900	1.538700e+04	0.000020	
25%	29.000000	0.037600	0.435000	1.828570e+05	0.385000	
50%	43.000000	0.232000	0.571000	2.204270e+05	0.605000	
75%	55.000000	0.722000	0.692000	2.657680e+05	0.787000	
max	100.000000	0.996000	0.989000	5.552917e+06	0.999000	

In [19]:
```
df = df.dropna()
print(df.isnull().sum())
print("Dataset is clean")
```

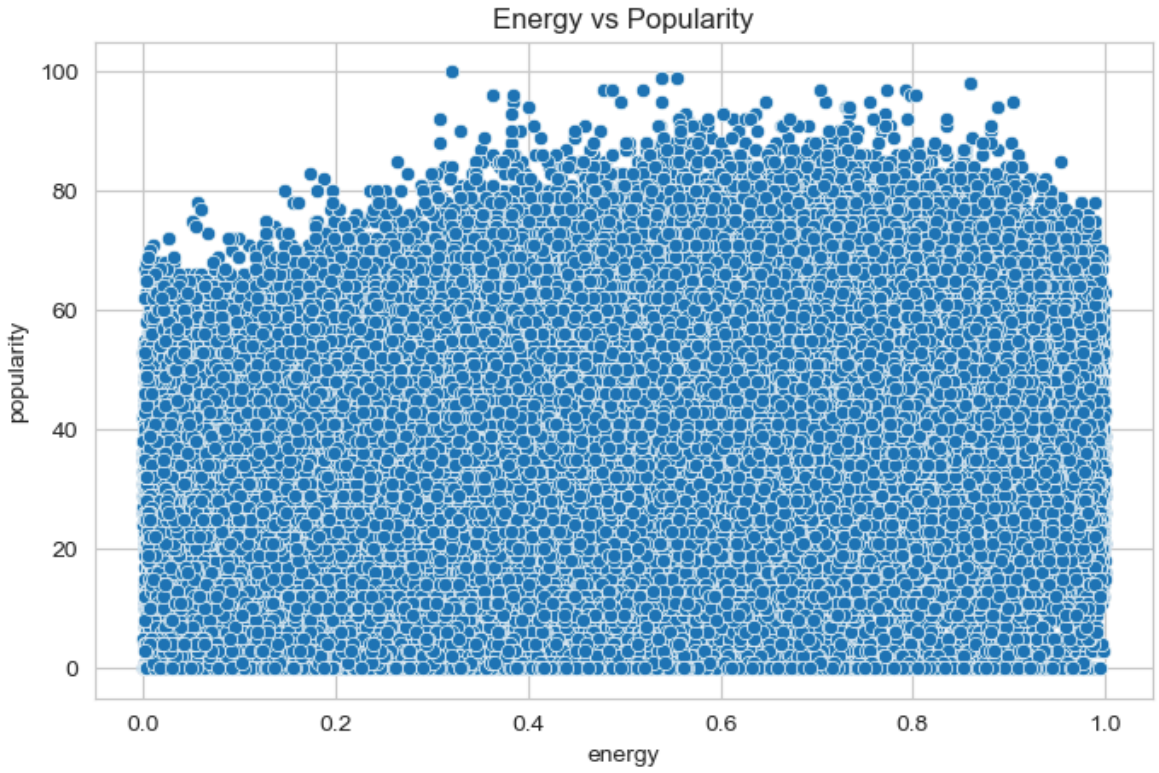
```
genre              0
artist_name        0
track_name         0
track_id           0
popularity         0
acousticness       0
danceability       0
duration_ms        0
energy             0
instrumentalness   0
key                0
liveness           0
loudness           0
mode               0
speechiness        0
tempo              0
time_signature     0
valence            0
dtype: int64
Dataset is clean
```

In [20]:
```
plt.figure(figsize=(8,5))
sns.histplot(df['popularity'], bins=30)
plt.title("Distribution of Song Popularity")
plt.show()
```



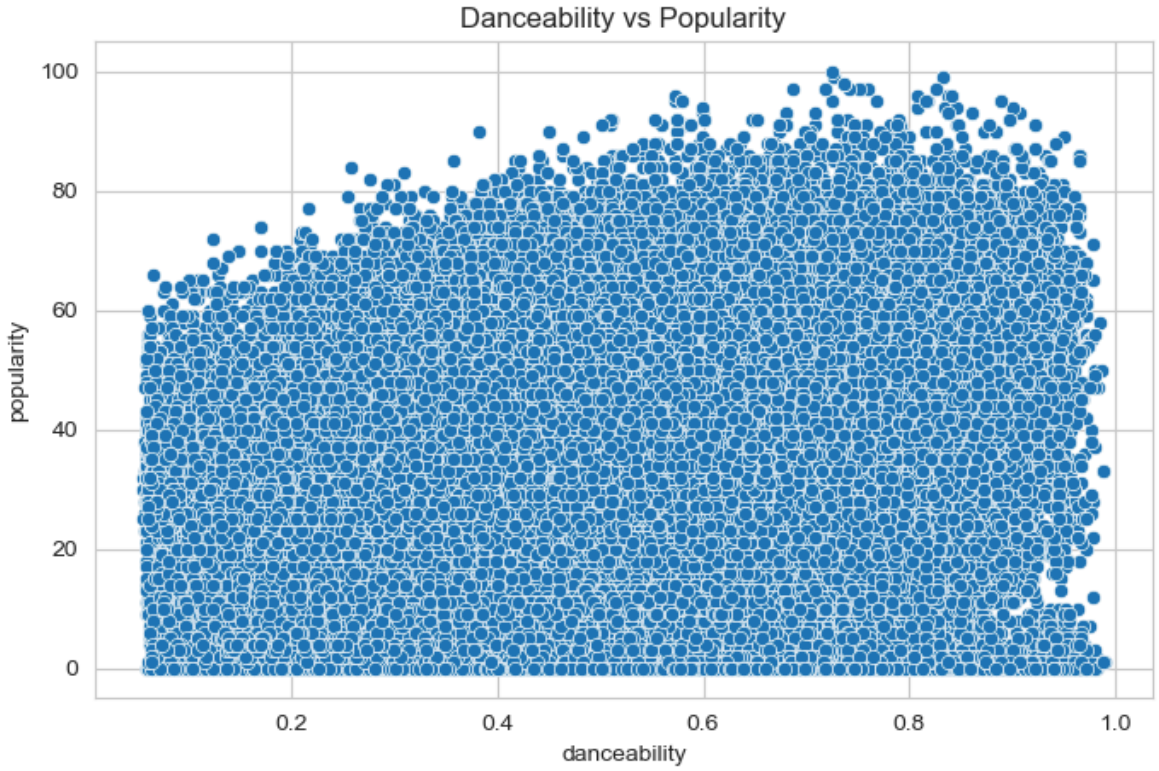
Most songs have low to medium popularity.

In [21]:
```
plt.figure(figsize=(8,5))
sns.scatterplot(x='energy', y='popularity', data=df)
plt.title("Energy vs Popularity")
plt.show()
```



Shows relationship between energy and popularity.

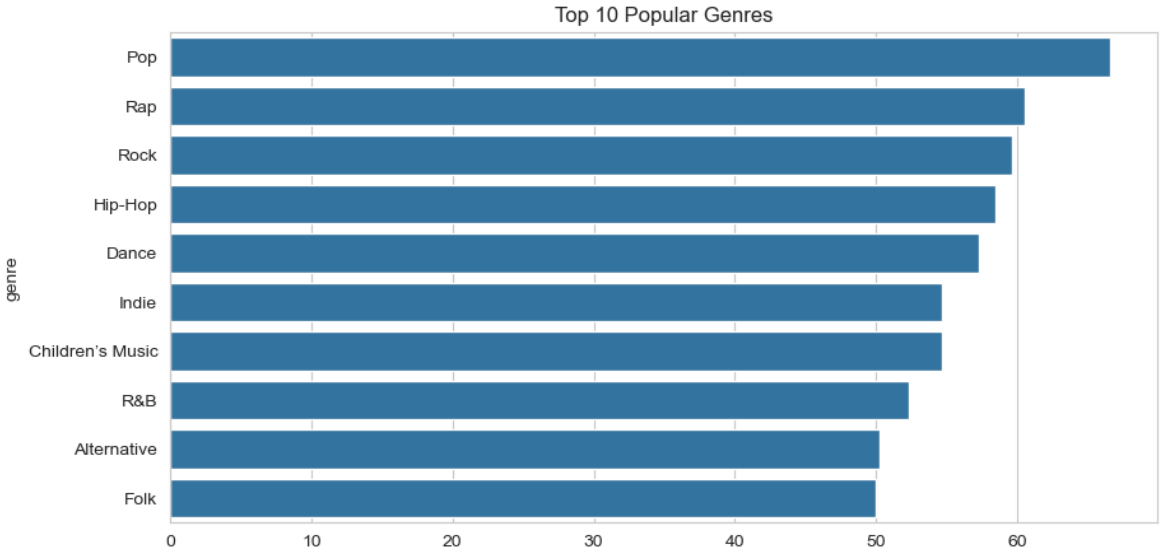
In [22]:
```
plt.figure(figsize=(8,5))
sns.scatterplot(x='danceability', y='popularity', data=df)
plt.title("Danceability vs Popularity")
plt.show()
```



Danceable songs tend to be more popular.

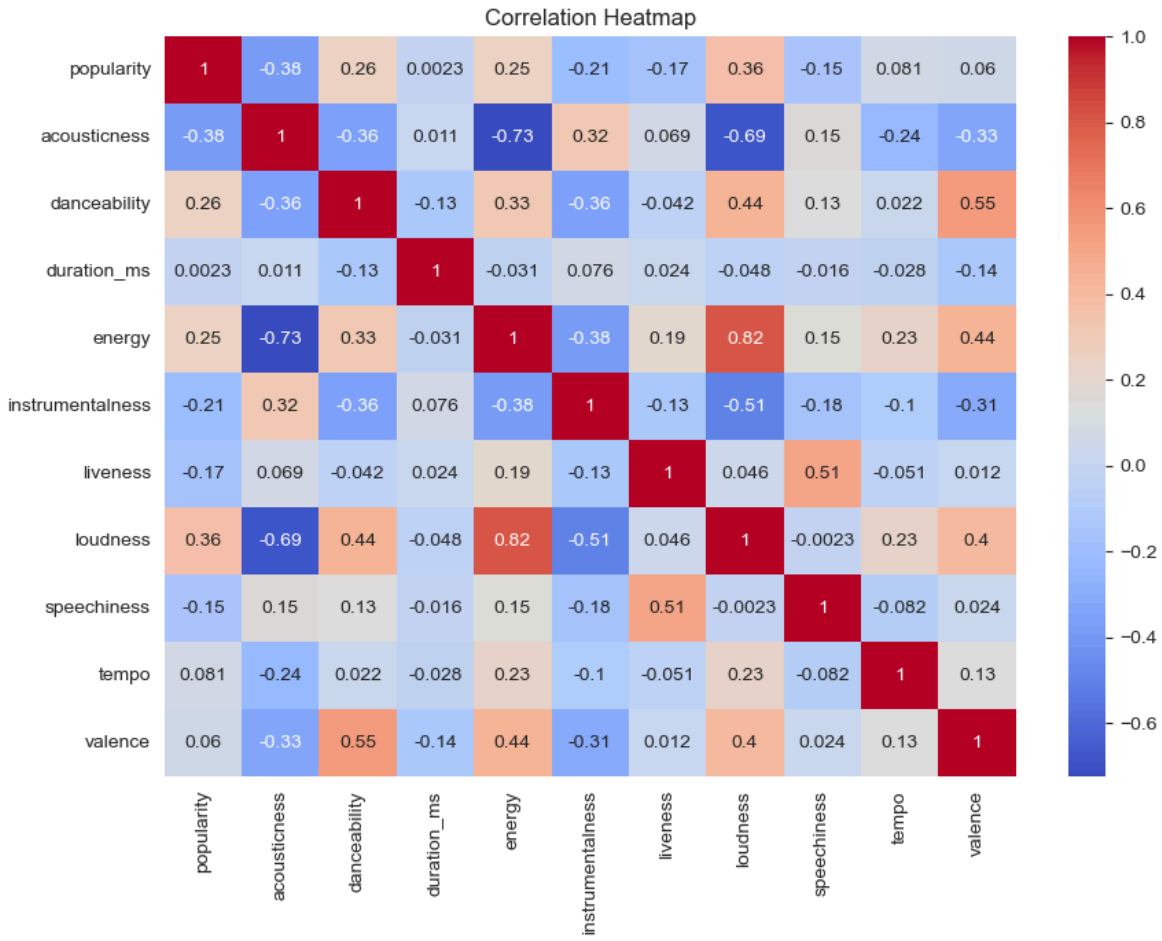
In [23]:
```
top_genres = df.groupby('genre')['popularity'].mean().sort_values(ascending=Fals
```

```
plt.figure(figsize=(10,5))
sns.barplot(x=top_genres.values, y=top_genres.index)
plt.title("Top 10 Popular Genres")
plt.show()
```



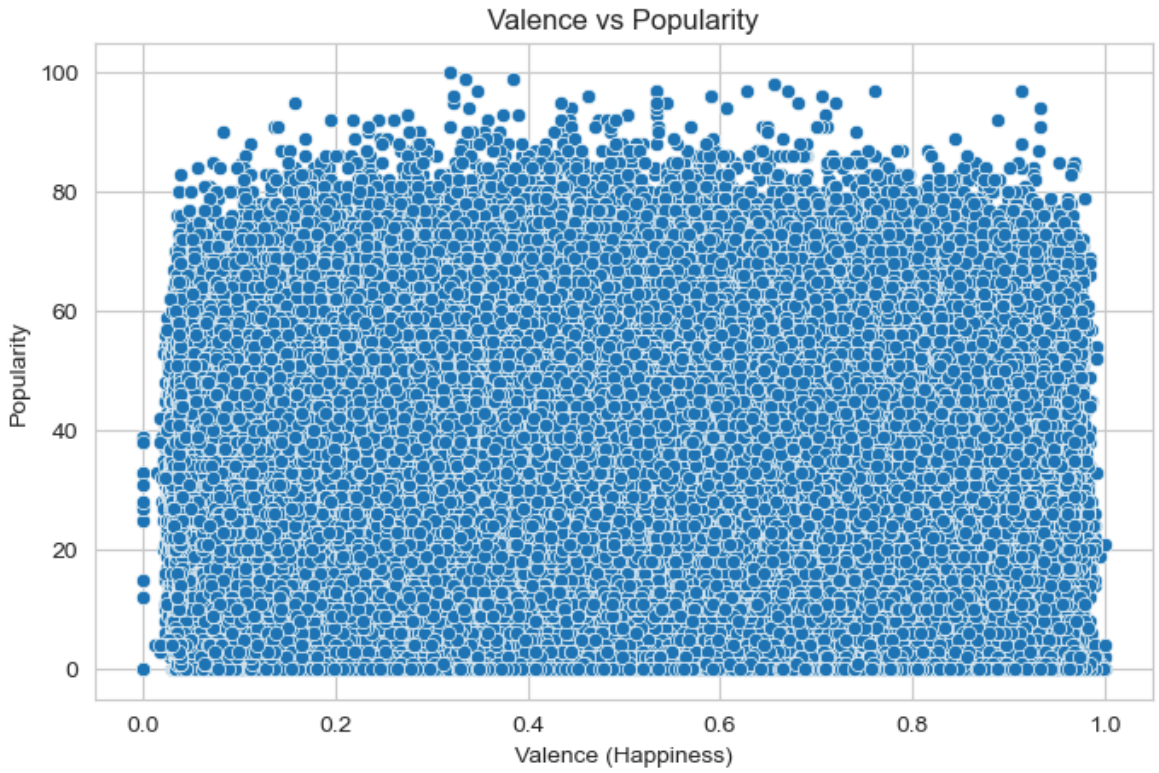
Some genres consistently show higher average popularity than others.

In [24]:
```
plt.figure(figsize=(10,7))
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap="coolwarm")
plt.title("Correlation Heatmap")
plt.show()
```



This heatmap shows relationships between all features and popularity.

In [25]:
```
plt.figure(figsize=(8,5))
sns.scatterplot(x='valence', y='popularity', data=df)
plt.title("Valence vs Popularity")
plt.xlabel("Valence (Happiness)")
plt.ylabel("Popularity")
plt.show()
```



Interpretation:

Valence represents how happy a song sounds. Some happy songs tend to be popular, but the relationship is moderate.

In [26]:
```
top_songs = df[['track_name','artist_name','genre','popularity']].sort_values(by
top_songs
```

Out[26]:

	track_name	artist_name	genre	popularity
107804	7 rings	Ariana Grande	Pop	100
9027	7 rings	Ariana Grande	Dance	100
107803	Wow.	Post Malone	Pop	99
9026	break up with your girlfriend, i'm bored	Ariana Grande	Dance	99
86951	Wow.	Post Malone	Rap	99
107802	break up with your girlfriend, i'm bored	Ariana Grande	Pop	99
138918	Con Calma	Daddy Yankee	Reggaeton	98
107909	Con Calma	Daddy Yankee	Pop	98
66643	Con Calma	Daddy Yankee	Hip-Hop	98
107810	Dancing With A Stranger (with Normani)	Sam Smith	Pop	97

Interpretation:

These are the most popular songs in the dataset. This helps identify patterns in highly successful tracks.

Key Findings:

• Most songs have medium popularity. • Energy and danceability have positive influence on popularity. • Pop, Rap, and Rock genres are most popular. • Valence has moderate impact on popularity. • Audio features play important role in song success.

Conclusion:

The analysis shows that song popularity is influenced by multiple factors such as energy, danceability, and genre. Songs with higher energy and danceability tend to have higher popularity. Some genres consistently show higher average popularity than others. Correlation analysis shows moderate relationships between popularity and audio features.

This demonstrates that audio characteristics play an important role in determining the popularity of songs on Spotify.