import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sb

import statsmodels.formula.api as sm

path = r"C:\Users\junnk\Downloads\top2018songs (1).csv"

top2018 = pd.read\_csv(path)

top2018.head(10)

top2018['artists'].value\_counts().head(20)

# Heatmap illustrating correlation of variables

f,ax=plt.subplots(figsize=(20,10))

sb.heatmap(top2018.iloc[:,3:16].corr(),annot=True,linewidth=.5,fmt=".2f",ax=ax)

plt.title("Correlations of Variables")

plt.show()

# Top 10 songs by speechiness

top2018[['name','artists','speechiness']].sort\_values(by='speechiness',ascending=False).head(10)

# Top 10 songs by acousticness

top2018[['name','artists','acousticness']].sort\_values(by='acousticness',ascending=False).head(10)

# Top ten songs by danceability

top2018[['name','artists','danceability']].sort\_values(by='danceability',ascending=False).head(10)

# Top 10 songs by valence

top2018[['name','artists','valence']].sort\_values(by='valence',ascending=False).head(10)

# Top 10 songs with highest valence

top2018[['name','artists','valence']].sort\_values(by='valence',ascending=False).tail(10)

plt.scatter(top2018["valence"], top2018["acousticness"], marker="\*", s=10)

plt.title("Valence vs. Acousticness")

plt.ylabel("valence")

plt.xlabel("acousticness")

plt.grid(True)

plt.scatter(top2018["valence"], top2018["speechiness"], marker="\*", s=10)

# Incorporate the other graph properties

plt.title("Valence vs. Speechiness")

plt.ylabel("valence")

plt.xlabel("speechiness")

plt.grid(True)