```
# Шаг 0. Импорт библиотек Python, необходимых для работы.

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Шаг 1. Выбор данных.

music = pd.read_csv('/Users/vladislavsolovev/Desktop/IT/Data Science/Практика/Наборы Данных (Датасеты)/Kaggle/Spotify_1Million_Tracks.csv', delimiter = ',')
print(music)
musicDATA = music[['artist_name', 'track_name', 'year', 'popularity', 'danceability', 'tempo', 'duration_ms']]
print(musicDATA)
musicJM = musicDATA.loc[musicDATA['artist_name'] == "Jason Mraz"]
print(musicJM)
musicJH = musicDATA.loc[musicDATA['artist_name'] == "Joshua Hyslop"]
print(musicJH)
musicAB = musicDATA.loc[musicDATA['artist_name'] == "Andrew Belle"]
print(musicAB)
musicNC = musicDATA.loc[musicDATA['artist_name'] == "Nicola Conte"]
print(musicNC)
musicAT = musicDATA.loc[musicDATA['artist_name'] == "Amon Tobin"]
print(musicAT)
musicDG = musicDATA.loc[musicDATA['artist_name'] == "David Gray"]
print(musicDG)
musicHP = musicDATA.loc[musicDATA['artist_name'] == "Harley Poe"]
print(musicHP)
```

```
# Шаг 2. Создание случайных выборок.

# Музыкант(1) - Jason Mraz. Общее количество треков - 193.
m_y_JM_2004 = musicJM.loc[musicJM['year'] == 2004]
print(m_y_JM_2004)
m_y_JM_2005 = musicJM.loc[musicJM['year'] == 2005]
print(m_y_JM_2005)
m_y_JM_2006 = musicJM.loc[musicJM['year'] == 2006]
print(m_y_JM_2006)
m_y_JM_2007 = musicJM.loc[musicJM['year'] == 2007]
print(m_y_JM_2007)
m_y_JM_2008 = musicJM.loc[musicJM['year'] == 2008]
print(m_y_JM_2008)
m_y_JM_2009 = musicJM.loc[musicJM['year'] == 2009]
print(m_y_JM_2009)
m_y_JM_2010 = musicJM.loc[musicJM['year'] == 2010]
print(m_y_JM_2010)
m_y_JM_2011 = musicJM.loc[musicJM['year'] == 2011]
print(m_y_JM_2011)
m_y_JM_2012 = musicJM.loc[musicJM['year'] == 2012]
print(m_y_JM_2012)
m_y_JM_2014 = musicJM.loc[musicJM['year'] == 2014]
print(m_y_JM_2014)
m_y_JM_2017 = musicJM.loc[musicJM['year'] == 2017]
print(m_y_JM_2017)
m_y_JM_2018 = musicJM.loc[musicJM['year'] == 2018]
print(m_y_JM_2018)
m_y_JM_2020 = musicJM.loc[musicJM['year'] == 2020]
print(m_y_JM_2020)
m_y_JM_2021 = musicJM.loc[musicJM['year'] == 2021]
```

```
print(m_y_JM_2021)
m_y_JM_2022 = musicJM.loc[musicJM['year'] == 2022]
print(m_y_JM_2022)
m_y_JM_2023 = musicJM.loc[musicJM['year'] == 2023]
print(m_y_JM_2023)

# Музыкант(2) - Joshua Hyslop. Общее количество треков - 61.
m_y_JH_2011 = musicJH.loc[musicJH['year'] == 2011]
print(m_y_JH_2011)
m_y_JH_2012 = musicJH.loc[musicJH['year'] == 2012]
print(m_y_JH_2012)
m_y_JH_2015 = musicJH.loc[musicJH['year'] == 2015]
print(m_y_JH_2015)
m_y_JH_2016 = musicJH.loc[musicJH['year'] == 2016]
print(m_y_JH_2016)
m_y_JH_2017 = musicJH.loc[musicJH['year'] == 2017]
print(m_y_JH_2017)
m_y_JH_2018 = musicJH.loc[musicJH['year'] == 2018]
print(m_y_JH_2018)
m_y_JH_2020 = musicJH.loc[musicJH['year'] == 2020]
print(m_y_JH_2020)
m_y_JH_2022 = musicJH.loc[musicJH['year'] == 2022]
print(m_y_JH_2022)

# Музыкант(3) - Andrew Belle. Общее количество треков - 66.
m_y_AB_2008 = musicAB.loc[musicAB['year'] == 2008]
print(m_y_AB_2008)
m_y_AB_2010 = musicAB.loc[musicAB['year'] == 2010]
print(m_y_AB_2010)
m_y_AB_2011 = musicAB.loc[musicAB['year'] == 2011]
```

```
print(m_y_AB_2011)
m_y_AB_2012 = musicAB.loc[musicAB['year'] == 2012]
print(m_y_AB_2012)
m_y_AB_2013 = musicAB.loc[musicAB['year'] == 2013]
print(m_y_AB_2013)
m_y_AB_2014 = musicAB.loc[musicAB['year'] == 2014]
print(m_y_AB_2014)
m_y_AB_2016 = musicAB.loc[musicAB['year'] == 2016]
print(m_y_AB_2016)
m_y_AB_2017 = musicAB.loc[musicAB['year'] == 2017]
print(m_y_AB_2017)
m_y_AB_2018 = musicAB.loc[musicAB['year'] == 2018]
print(m_y_AB_2018)
m_y_AB_2020 = musicAB.loc[musicAB['year'] == 2020]
print(m_y_AB_2020)
m_y_AB_2021 = musicAB.loc[musicAB['year'] == 2021]
print(m_y_AB_2021)
m_y_AB_2023 = musicAB.loc[musicAB['year'] == 2023]
print(m_y_AB_2023)

# Музыкант(4) - Nicola Conte. Общее количество треков - 126.
m_y_NC_2000 = musicNC.loc[musicNC['year'] == 2000]
print(m_y_NC_2000)
m_y_NC_2004 = musicNC.loc[musicNC['year'] == 2004]
print(m_y_NC_2004)
m_y_NC_2007 = musicNC.loc[musicNC['year'] == 2007]
print(m_y_NC_2007)
m_y_NC_2008 = musicNC.loc[musicNC['year'] == 2008]
print(m_y_NC_2008)
m_y_NC_2009 = musicNC.loc[musicNC['year'] == 2009]
```

```
print(m_y_NC_2009)
m_y_NC_2011 = musicNC.loc[musicNC['year'] == 2011]
print(m_y_NC_2011)
m_y_NC_2014 = musicNC.loc[musicNC['year'] == 2014]
print(m_y_NC_2014)
m_y_NC_2016 = musicNC.loc[musicNC['year'] == 2016]
print(m_y_NC_2016)
m_y_NC_2017 = musicNC.loc[musicNC['year'] == 2017]
print(m_y_NC_2017)
m_y_NC_2018 = musicNC.loc[musicNC['year'] == 2018]
print(m_y_NC_2018)
m_y_NC_2019 = musicNC.loc[musicNC['year'] == 2019]
print(m_y_NC_2019)
m_y_NC_2020 = musicNC.loc[musicNC['year'] == 2020]
print(m_y_NC_2020)
m_y_NC_2021 = musicNC.loc[musicNC['year'] == 2021]
print(m_y_NC_2021)
m_y_NC_2022 = musicNC.loc[musicNC['year'] == 2022]
print(m_y_NC_2022)
m_y_NC_2023 = musicNC.loc[musicNC['year'] == 2023]
print(m_y_NC_2023)

# Музыкант(5) - Amon Tobin. Общее количество треков - 115.
m_y_AT_2000 = musicAT.loc[musicAT['year'] == 2000]
print(m_y_AT_2000)
m_y_AT_2002 = musicAT.loc[musicAT['year'] == 2002]
print(m_y_AT_2002)
m_y_AT_2003 = musicAT.loc[musicAT['year'] == 2003]
print(m_y_AT_2003)
m_y_AT_2005 = musicAT.loc[musicAT['year'] == 2005]
```

```
print(m_y_AT_2005)
m_y_AT_2007 = musicAT.loc[musicAT['year'] == 2007]
print(m_y_AT_2007)
m_y_AT_2009 = musicAT.loc[musicAT['year'] == 2009]
print(m_y_AT_2009)
m_y_AT_2010 = musicAT.loc[musicAT['year'] == 2010]
print(m_y_AT_2010)
m_y_AT_2011 = musicAT.loc[musicAT['year'] == 2011]
print(m_y_AT_2011)
m_y_AT_2012 = musicAT.loc[musicAT['year'] == 2012]
print(m_y_AT_2012)
m_y_AT_2015 = musicAT.loc[musicAT['year'] == 2015]
print(m_y_AT_2015)
m_y_AT_2018 = musicAT.loc[musicAT['year'] == 2018]
print(m_y_AT_2018)
m_y_AT_2019 = musicAT.loc[musicAT['year'] == 2019]
print(m_y_AT_2019)
m_y_AT_2021 = musicAT.loc[musicAT['year'] == 2021]
print(m_y_AT_2021)
m_y_AT_2022 = musicAT.loc[musicAT['year'] == 2022]
print(m_y_AT_2022)

# Музыкант(6) - David Gray. Общее количество треков - 174.
m_y_DG_2001 = musicDG.loc[musicDG['year'] == 2001]
print(m_y_DG_2001)
m_y_DG_2002 = musicDG.loc[musicDG['year'] == 2002]
print(m_y_DG_2002)
m_y_DG_2005 = musicDG.loc[musicDG['year'] == 2005]
print(m_y_DG_2005)
m_y_DG_2006 = musicDG.loc[musicDG['year'] == 2006]
```

```
print(m_y_DG_2006)
m_y_DG_2007 = musicDG.loc[musicDG['year'] == 2007]
print(m_y_DG_2007)
m_y_DG_2009 = musicDG.loc[musicDG['year'] == 2009]
print(m_y_DG_2009)
m_y_DG_2010 = musicDG.loc[musicDG['year'] == 2010]
print(m_y_DG_2010)
m_y_DG_2012 = musicDG.loc[musicDG['year'] == 2012]
print(m_y_DG_2012)
m_y_DG_2014 = musicDG.loc[musicDG['year'] == 2014]
print(m_y_DG_2014)
m_y_DG_2015 = musicDG.loc[musicDG['year'] == 2015]
print(m_y_DG_2015)
m_y_DG_2016 = musicDG.loc[musicDG['year'] == 2016]
print(m_y_DG_2016)
m_y_DG_2019 = musicDG.loc[musicDG['year'] == 2019]
print(m_y_DG_2019)
m_y_DG_2020 = musicDG.loc[musicDG['year'] == 2020]
print(m_y_DG_2020)
m_y_DG_2021 = musicDG.loc[musicDG['year'] == 2021]
print(m_y_DG_2021)
m_y_DG_2022 = musicDG.loc[musicDG['year'] == 2022]
print(m_y_DG_2022)

# Музыкант(7) - Harley Poe. Общее количество треков - 127.
m_y_HP_2004 = musicHP.loc[musicHP['year'] == 2004]
print(m_y_HP_2004)
m_y_HP_2006 = musicHP.loc[musicHP['year'] == 2006]
print(m_y_HP_2006)
m_y_HP_2010 = musicHP.loc[musicHP['year'] == 2010]
```

```
print(m_y_HP_2010)
m_y_HP_2012 = musicHP.loc[musicHP['year'] == 2012]
print(m_y_HP_2012)
m_y_HP_2013 = musicHP.loc[musicHP['year'] == 2013]
print(m_y_HP_2013)
m_y_HP_2015 = musicHP.loc[musicHP['year'] == 2015]
print(m_y_HP_2015)
m_y_HP_2017 = musicHP.loc[musicHP['year'] == 2017]
print(m_y_HP_2017)
m_y_HP_2018 = musicHP.loc[musicHP['year'] == 2018]
print(m_y_HP_2018)
m_y_HP_2019 = musicHP.loc[musicHP['year'] == 2019]
print(m_y_HP_2019)
m_y_HP_2020 = musicHP.loc[musicHP['year'] == 2020]
print(m_y_HP_2020)
m_y_HP_2021 = musicHP.loc[musicHP['year'] == 2021]
print(m_y_HP_2021)
m_y_HP_2022 = musicHP.loc[musicHP['year'] == 2022]
print(m_y_HP_2022)

# Шаг 3. Создание кластеров.

# Первый Кластер.
Jason_Mraz = pd.concat([m_y_JM_2008, m_y_JM_2011, m_y_JM_2014,
m_y_JM_2017, m_y_JM_2018, m_y_JM_2020, m_y_JM_2021, m_y_JM_2023])
print(Jason_Mraz)
Andrew_Belle = pd.concat([m_y_AB_2008, m_y_AB_2011, m_y_AB_2014,
m_y_AB_2017, m_y_AB_2018, m_y_AB_2020, m_y_AB_2021,
m_y_AB_2023])
print(Andrew_Belle)
```

```
Nicola_Conte = pd.concat([m_y_NC_2008, m_y_NC_2011, m_y_NC_2014,
m_y_NC_2017, m_y_NC_2018, m_y_NC_2020, m_y_NC_2021,
m_y_NC_2023])
print(Nicola_Conte)
c_1 = pd.concat([Jason_Mraz, Andrew_Belle, Nicola_Conte])

# Второй Кластер.
Joshua_Hyslop = pd.concat([m_y_JH_2012, m_y_JH_2015, m_y_JH_2022])
print(Joshua_Hyslop)
Amon_Tobin = pd.concat([m_y_AT_2012, m_y_AT_2015, m_y_AT_2022])
print(Amon_Tobin)
David_Gray = pd.concat([m_y_DG_2012, m_y_DG_2015, m_y_DG_2022])
print(David_Gray)
Harley_Poe = pd.concat([m_y_HP_2012, m_y_HP_2015, m_y_HP_2022])
print(Harley_Poe)
c_2 = pd.concat([Joshua_Hyslop, Amon_Tobin, David_Gray, Harley_Poe])

# Выведем сформированные кластеры на экран.
print(c_1)
print(c_2)

# Шаг 4. Сортировка данных в кластерах.

# Подготовка данных Первого Кластера.
cluster_1 = c_1[['artist_name', 'popularity', 'tempo']]
print(cluster_1)

# Подготовка данных Второго Кластера.
cluster_2 = c_2[['artist_name', 'danceability', 'duration_ms']]
print(cluster_2)
```

```
# Проведём сортировку данных Первого Кластера по возрастанию.
CL_1 = cluster_1.sort_values(by=["tempo"])
print(CL_1)

# Проведём сортировку данных Второго Кластера по возрастанию.
CL_2 = cluster_2.sort_values(by=["duration_ms"])
print(CL_2)

# Шаг 5. Визуализация данных в полученных кластерах и построение графиков.

# Визуализация данных Первого Кластера.
sns.set_style("darkgrid", {'axes.facecolor':'thistle', 'axes.edgecolor':'indigo'})
plt.title("Зависимость популярности трека от его темпа", fontsize=16, color='darkmagenta')
tempo = CL_1["tempo"]
popularity = CL_1["popularity"]
plt.scatter(tempo, popularity, color = "darkviolet")
plt.xlabel('tempo', fontsize=12, color='mediumorchid')
plt.ylabel('popularity', fontsize=12, color='mediumorchid')
plt.legend()
plt.show();

# Визуализация данных Второго Кластера.
sns.set_style("darkgrid", {'axes.facecolor':'paleturquoise',
'axes.edgecolor':'darkblue',  'axes.labelcolor':'darkslateblue'})
duration_ms = CL_2["duration_ms"]
danceability = CL_2["danceability"]
Dan = sns.lineplot(x = duration_ms, y = danceability, legend = "full")
```

```
Dan.set_title("Зависимость танцевальности трека от его длительности",
fontsize=16, color='navy')
plt.show();

# Шаг 6. Подтверждение гипотез на основе анализа исходного файла с
полным набором данных.

# Проверка Первой Гипотезы.
sns.set_style("darkgrid", {'axes.facecolor':'thistle', 'axes.edgecolor':'indigo'})
plt.title("Зависимость популярности трека от его темпа", fontsize=16,
color='darkmagenta')
temp = musicDATA["tempo"]
popularnost = musicDATA["popularity"]
plt.scatter(temp, popularnost, color = "darkviolet")
plt.xlabel('tempo', fontsize=12, color='mediumorchid')
plt.ylabel('popularity', fontsize=12, color='mediumorchid')
plt.legend()
plt.show();

# Проверка Второй Гипотезы.
sns.set_style("darkgrid", {'axes.facecolor':'paleturquoise',
'axes.edgecolor':'darkblue', 'axes.labelcolor':'darkslateblue'})
duration = musicDATA["duration_ms"]
dance = musicDATA["danceability"]
Dur = sns.scatterplot(x = duration, y = dance)
Dur.set_title("Зависимость танцевальности трека от его длительности",
fontsize=16, color='navy')
plt.show()
```