Проект

По

Вероятности и статистика

Изготвила:

Ирина Атанасова, 81658

КН, 3ти курс, 7та група

**Проект за музикална класация**

# 1. Разглеждаме първите 10 песни от „Top 50 Spotify Songs” – 2019 спрямо :

1) Име на песента

2) Име на изпълнител

3) Жанр

4) Темпо на песента – бийтове за минута

5) Енергия – колкото по-висока е стойността, толкова по-енергична е песента

# 2.Въвеждане на данните

# name\_song = c("Senorita","China","boyfriend (with Social House)","Beautiful People (feat. Khalid)","Goodbyes (Feat. Young Thug)","I Don't Care (with Justin Bieber)","Ransom","How Do You Sleep?","Old Town Road - Remix","bad guy")

# name\_artist = c("Shawn Mendes","Anuel AA","Ariana Grande","Ed Sheeran","Post Malone","Ed Sheeran","Lil Tecca","Sam Smith","Lil Nas X","Billie Eilish")

# genre = c("canadian pop","reggaeton flow","dance pop","pop","dfw rap","pop","trap music","pop","country rap","electropop")

# tempo = c(117,105,190,93,150,102,180,111,136,135)

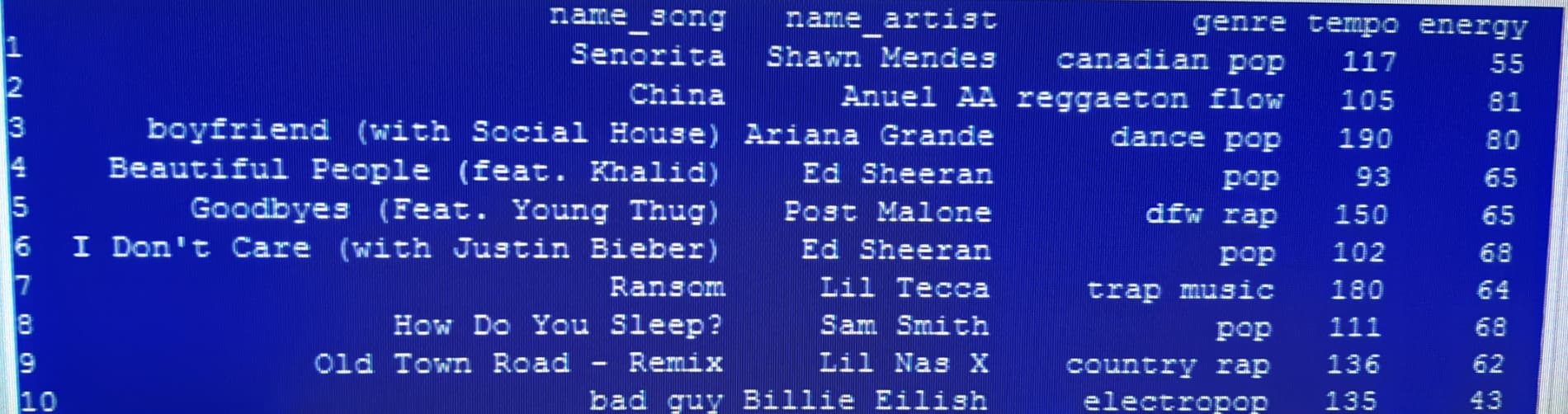
# energy = c(55,81,80,65,65,68,64,68,62,43)

Зареждаме данните в data.frame

survey = data.frame(name\_song, name\_artist, genre, tempo,energy)

survey

и получаваме:



# 3.Анализиране на данните

**3.1 Анализ на съотношението между жанровете и песните**.

Искаме да проверим дали има жанрове , които да са доминиращи спрямо броя песни.

Функцията barplot() показва графичното разпределение.

barplot(table(name\_song,genre), ylab = "songs", xlab = "genre", col = rainbow(8))

meta genre.emf

Виждаме, че най-много са песните, на които жанра е pop – 3 на брой, а останалите са по 1 песен на жанр, тоест pop-а доминира, а другите са равномерно разпределени.

**3.2 Анализ на съотношението между песните и изпълнителите.**

Искаме да проверим дали имаме изпълнител, който да има повече песни в класацията ,спрямо останалите.

Функцията barplot() показва графичното разпределение.

barplot(table(name\_song,name\_artist), ylab = "count of songs", xlab = "artist", col = rainbow(8))

meta artists.emf

# Виждаме, че най-много са песните, чийто изпълнител е Ed Sheeran – 2 на брой , а другите изпълнители имат по 1 песен .

**3.3 Анализ на темпото на песните.**

summary(tempo) не показва кое темпо в коя категория попада :

Min. 1st Qu. Median Mean 3rd Qu. Max.

93.0 106.5 126.0 131.9 146.5 190.0

Искаме да анализираме в какъв диапазон като темпо имаме най-много на брой песни.

Функцията hist() показва графичното разпределение на темпото на песните.

hist(tempo, ylab = "count", col = rainbow(6))

meta tempo.emf

# Тази хистограма показва ,че най-много песни имат темпо между 100 и 120 бийта в минута – 4 на брой, след тях са тези между 120 и 140 – 2 на брой, а останалите са по 1 на брой за другите стойности на темпо.

# 3.4 Анализ на енергията на песните.

# summary(energy) показва ясно как се разпределят енергиите:

# Min. 1st Qu. Median Mean 3rd Qu. Max.

# 43.0 62.5 65.0 65.1 68.0 81.0

# Искаме да проверим дали имаме диапазон спрямо енергиите, където да попадат повече на брой песни, спрямо останалите.

Функцията hist() показва графичното разпределение на енергиите на песните.

hist(energy, ylab = "count",col = c("red","blue","green","yellow","pink"))

legend("topleft", legend = c("Very Low","Low","Medium","High","Very High"), fill = c("red","blue","green","yellow","pink"))

legend() е използвано, за да класифицира различните енергии, спрямо интервала от стойности, в който попадат.

meta energy new.emf

Хистограмата показва, че най-много са песните, чиято енергия е между 60 и 70 (средната)– 6 на брой, а останалите са по 1 на брой за другите стойности на енергията.Тоест, най-много песни попадат в средните стойности спрямо енергията.

По-подробно можем да видим разпределението, спрямо енергията с кръгова диаграма.

pie(table(energy), col = rainbow(8))

meta pie energy.emf

Тук се вижда за всяка една стойност на енергията нейния дял спрямо цялото и се вижда, че доминиращи са 65 и 68, тоест там попадат най-много песни.

**3.5 Анализ на съотношението между жанра и енергията на песните.**

Искаме да видим дали има някаква зависимост между това песента в кой жанр е и това, каква е нейната енергия.

Функцията barplot() показва графичното разпределение.

barplot(table(genre,energy), ylab = "count", xlab = "energy" , col= rainbow(8))

meta neshto si.emf

Най много песни са с енергия 65 и 68 като 3 от тях са жанр pop и една е dfw rap.

**3.6 Анализ на съотношението между темпото и енергията на песните.**

Искаме да направим проверка за линейна зависимост между темпото и енергията.

plot(energy~tempo)

abline(lm(energy~tempo), col="red")

meta line.emf

Не се наблюдава струпване на точки около линията, което означава, че

нямаме някакво линейно отношение между двете числови променливи - tempo и energy .

**4.Заключение.**

Забелязваме, че определено pop –а е доминиращ жанр, но няма твърде голяма разлика с останалите, което показва, че и другите жанрове са харесвани от публиката. Най-категорични резултати се получават при анализа на музикалното темпо и енергия – тук с много голяма преднина са представени песни със средни (medium) стойности за енергия, докато при темпото доминират мелодии в диапазона 100-140 бийта в секунда.

Като цяло изследването показва практически равномерно разпределение на предпочитанията на публиката относно жанрове и изпълнители, а отчетените предпочитания при темпо и енергия и не се отнасят до конкретен музикален стил. Така получените резултати не позволяват да се дефинират достатъчно ясни закономерности, тоест освен изискването за средна енергия нямаме допълнителни други характеристики, които статистически значимо да определят шанса на дадена песен за попадане в класацията.