

Проект

Продуктовая аналитика

Подготовил Кузьмин А.

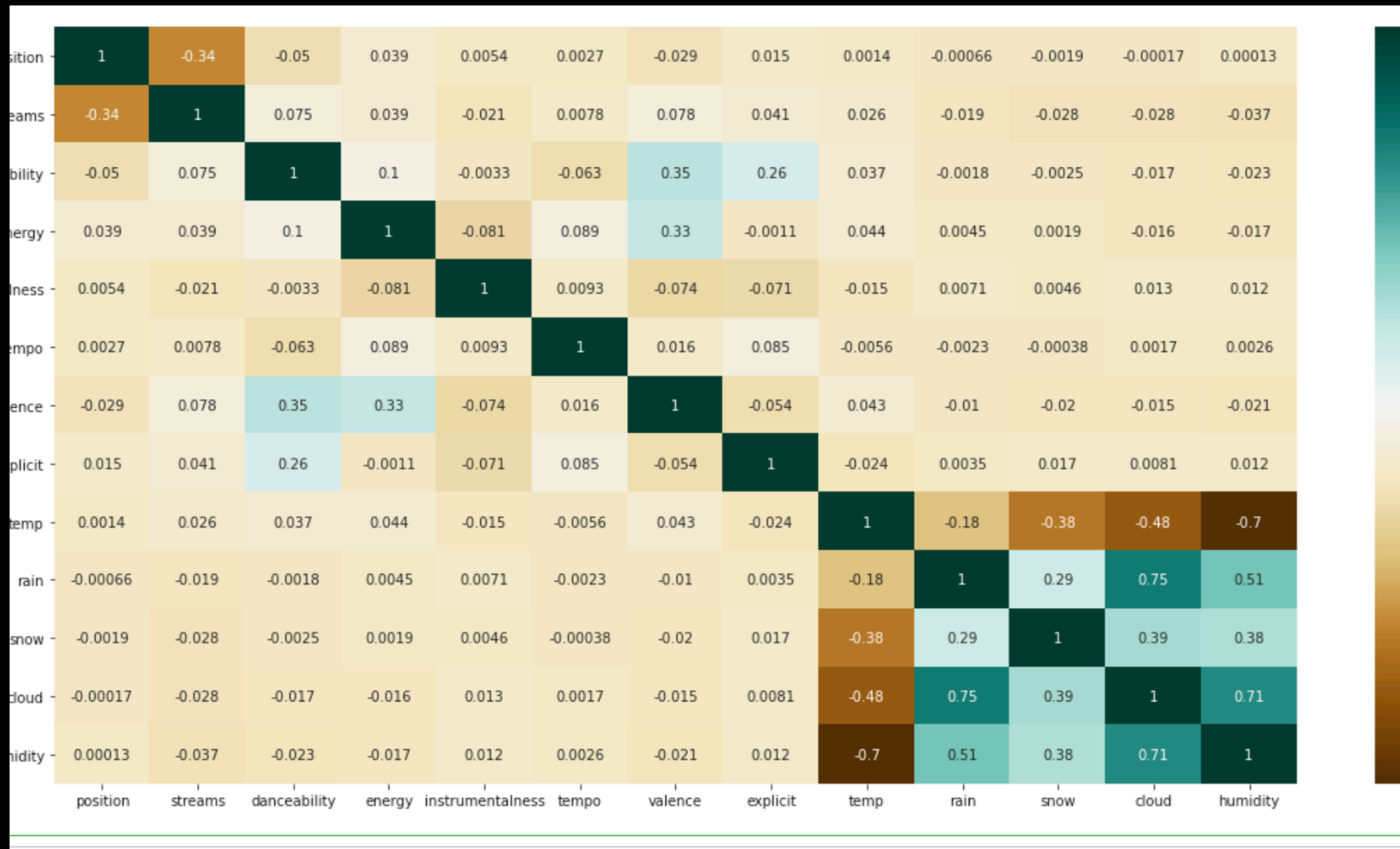
Основные сведения о данных

- Данные связанные с Spotify , World-weather.
- 3 группы данных : музыка, погодные условия, дата и время
- Приведено множество различных параметров музыки и погоды

Постановка ГИПОТЕЗЫ

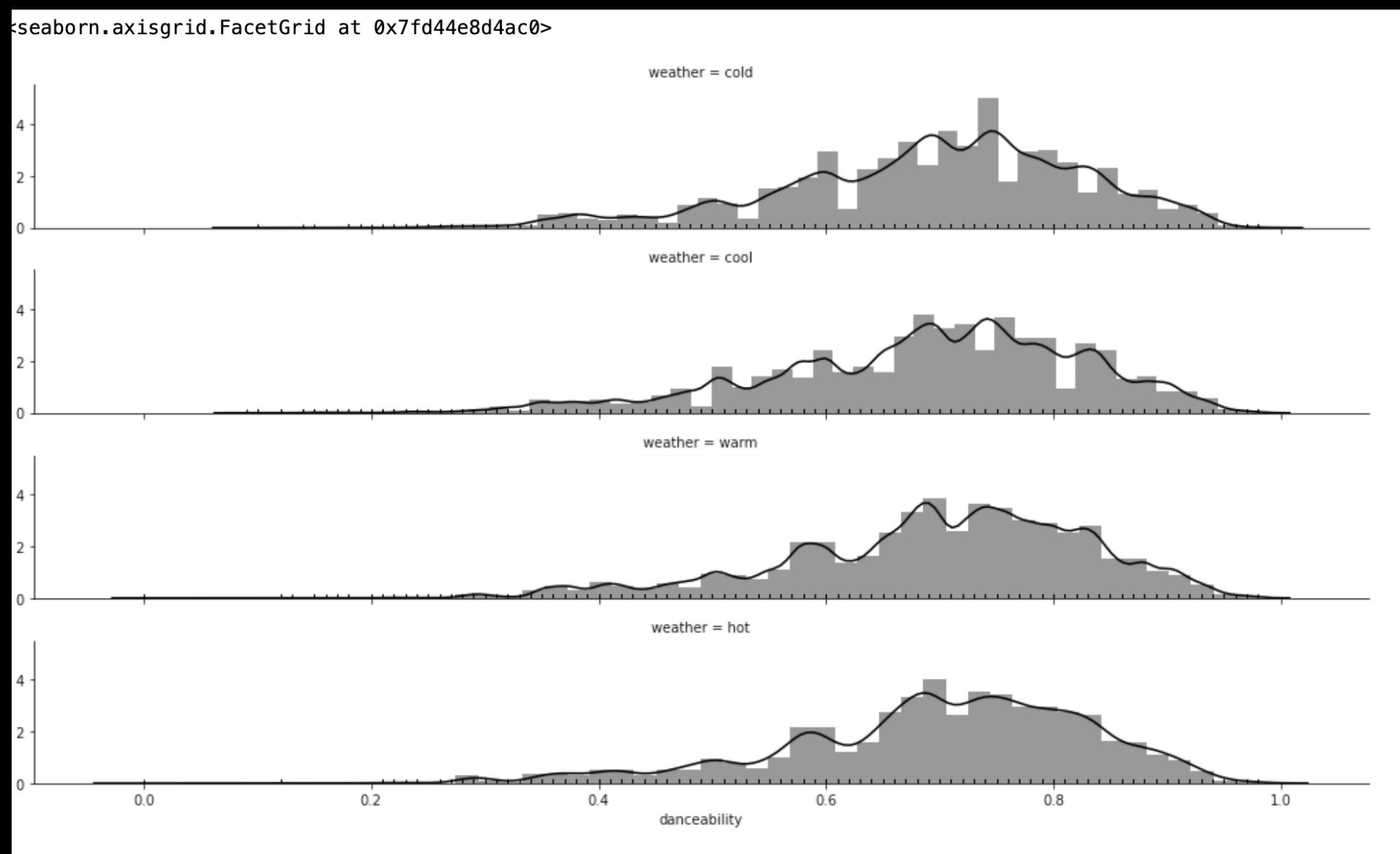
Зависимость параметров музыки и погоды. В качестве параметра был выбран 'danceability'. Погода разделена по категориям.

Корреляционная матрица



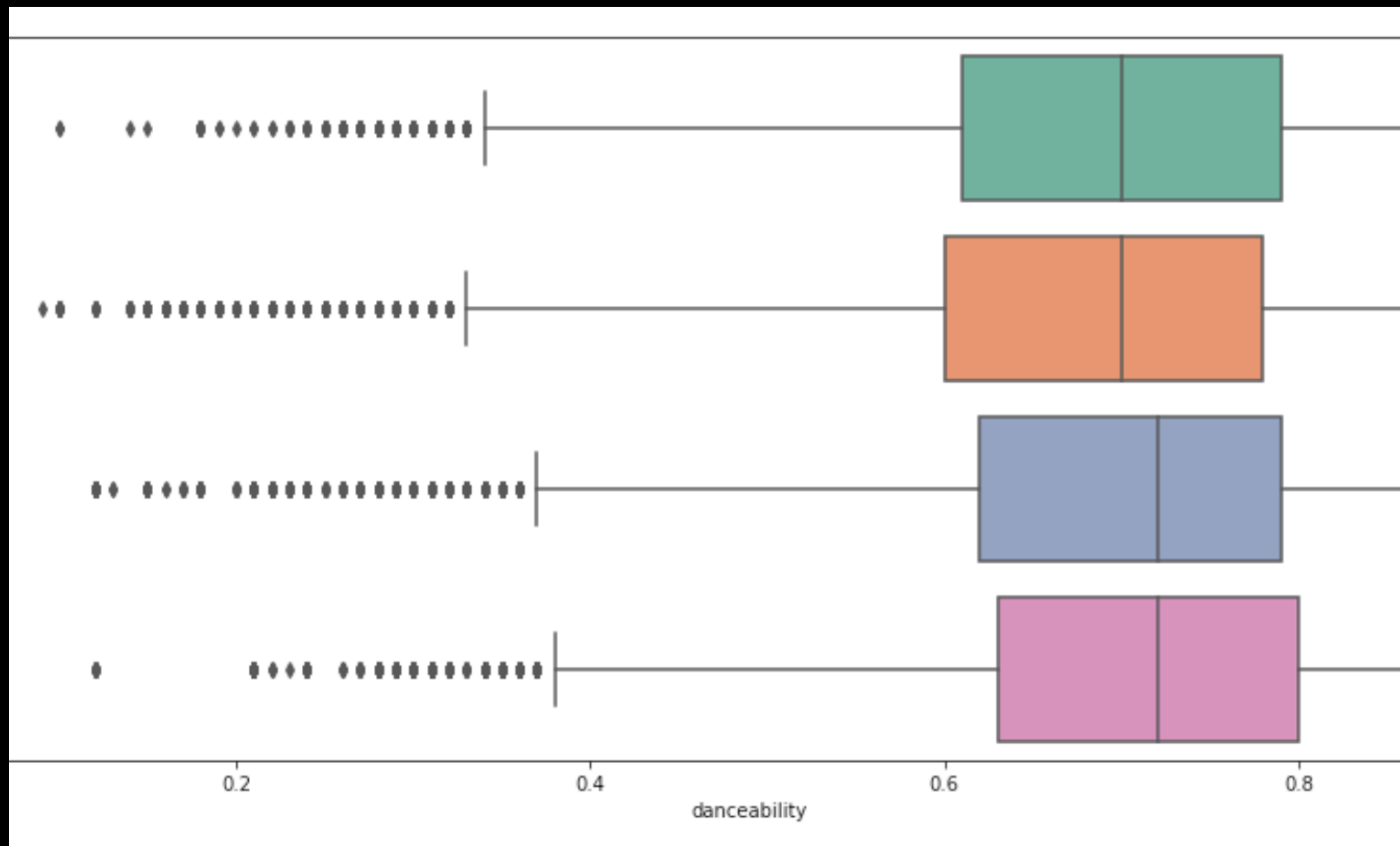
Вывод : Слабая зависимость между погодой и музыкой.

Парное распределение по параметру музыки и погоде



Вывод : Т.к. Во всех 4х графиках значения распределены в одном диапазоне , то это говорит нам о независимости и отвердении гипотезы.

Box plot



Вывод : аналогичный предыдущему

Распределения в группах по погоде

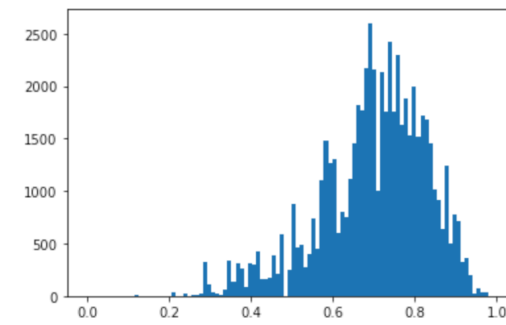
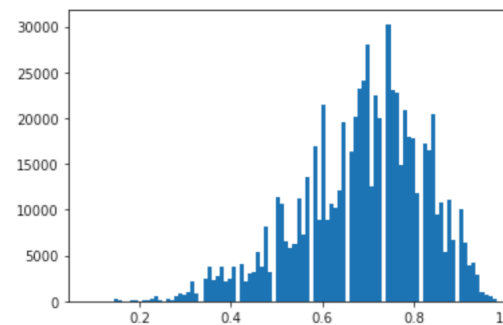
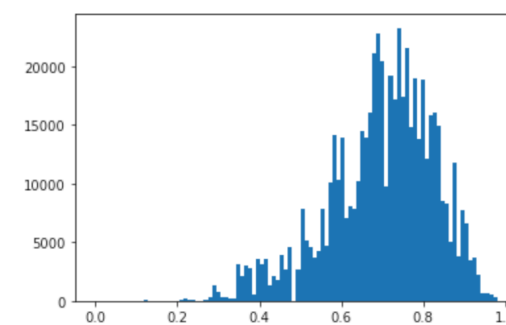
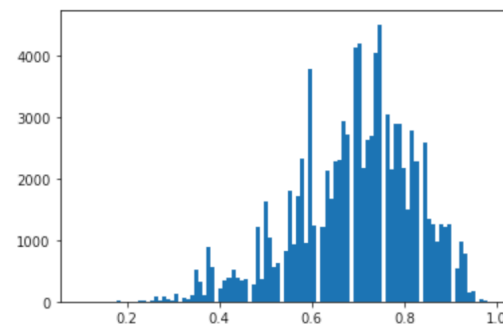
```
: alpha = 0.05
st = shapiro(data[data['weather'] == 'cold']['danceability'])
print('Distributions is {}normal\n'.format( {True:'not ',False:''}[st[1] < alpha]))
st = shapiro(data[data['weather'] == 'cool']['danceability'])
print('Distributions is {}normal\n'.format( {True:'not ',False:''}[st[1] < alpha]))
st = shapiro(data[data['weather'] == 'warm']['danceability'])
print('Distributions is {}normal\n'.format( {True:'not ',False:''}[st[1] < alpha]))
st = shapiro(data[data['weather'] == 'hot']['danceability'])
print('Distributions is {}normal\n'.format( {True:'not ',False:''}[st[1] < alpha]))
```

Distributions is not normal

Distributions is normal

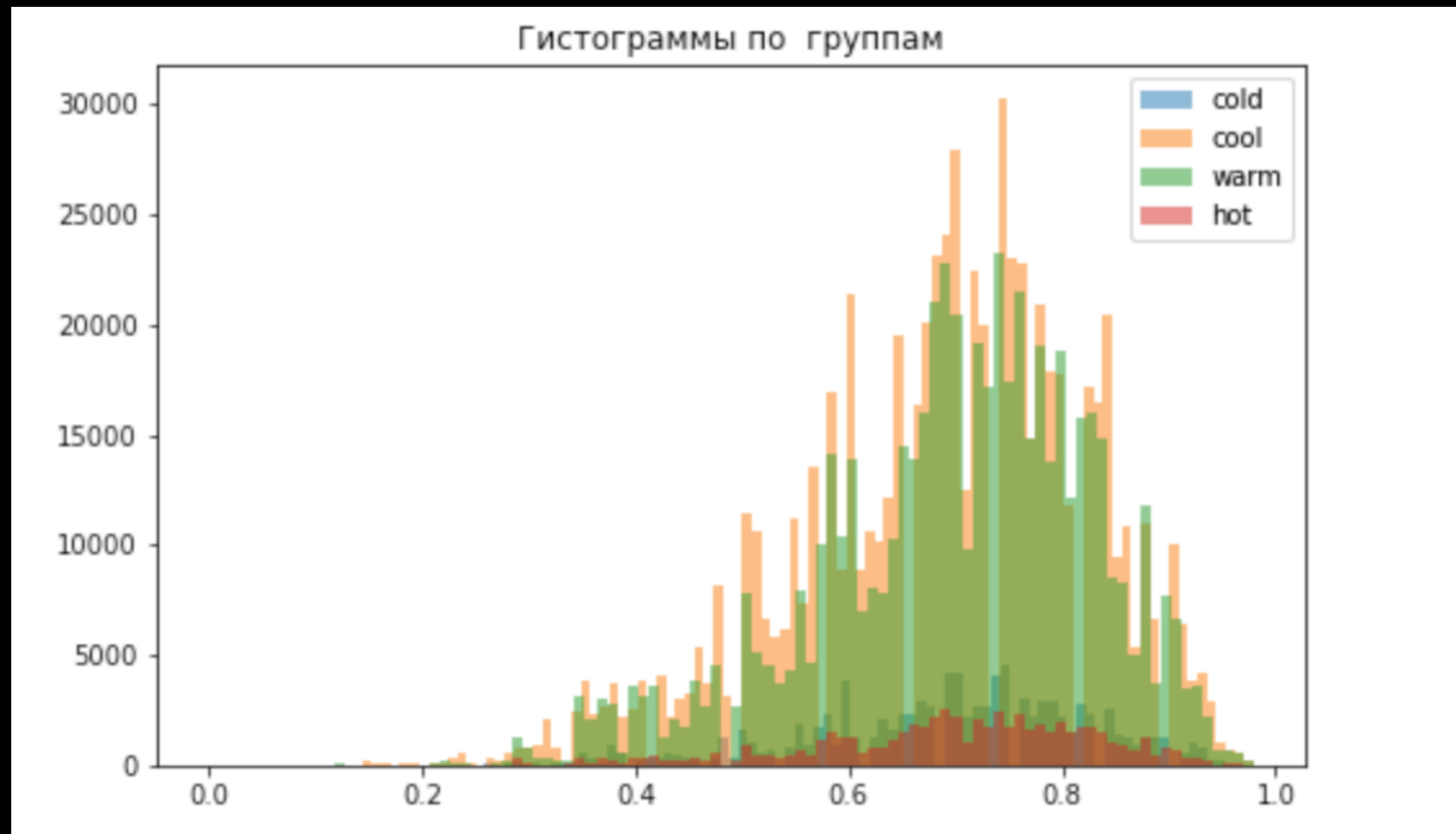
Distributions is normal

Distributions is not normal



Имеют вид схожий с нормальным. Однако 2 группы нормальные, 2 - нет.

Мат. Ожидания Дисперсии



Вывод : мат. Ожидания и дисперсии в группах равны. Данное заключение проиллюстрировано.

Примечание : Выбор критериев связан с : нечувствительностью к нормальным ,
Учетом равенства сразу в четырех группах, неравенством количества в группах значений.

Вывод

Т.к. мат. ожидания и дисперсии равны, то это означает, что независимо от температуры, параметр 'danceability' распределен одинаково. Таким образом, можно заключить, что температура не влияет на данный параметр, следовательно отвергаем гипотезу.

**Спасибо за
внимание!**