

Проект Музыкальная терапия



Киреев Максим
Панина Екатерина

Задачи

- **Регрессия:** Предсказать уровни тревожности/депрессии/тревожности/ ОКР на основе музыкальных предпочтений человека.
- **Кластеризация:** С помощью метода кластеризации (KMeans) сегментировать пользователей по музыкальным привычкам и психоэмоциональному фону для нахождения взаимосвязей.

Anxiety	Depression	Insomnia	OCD	Music effects	Permissions
3.0	0.0	1.0	0.0	NaN	I understand.
7.0	2.0	2.0	1.0	NaN	I understand.
7.0	7.0	10.0	2.0	No effect	I understand.
9.0	7.0	3.0	3.0	Improve	I understand.
7.0	2.0	5.0	9.0	Improve	I understand.
NaN	NaN	NaN	NaN	NaN	NaN

[illegible]

Исследование данных

Timestamp	735
Age	61
Primary streaming service	6
Hours per day	27
While working	2
Instrumentalist	2
Composer	2
Fav genre	16
Exploratory	2
Foreign languages	2
BPM	135
Frequency [Classical]	4
Frequency [Country]	4
Frequency [EDM]	4
Frequency [Folk]	4
Frequency [Gospel]	4
Frequency [Hip hop]	4
Frequency [Jazz]	4
Frequency [K pop]	4
Frequency [Latin]	4
Frequency [Lofi]	4
Frequency [Metal]	4
Frequency [Pop]	4
Frequency [R&B]	4
Frequency [Rap]	4
Frequency [Rock]	4
Frequency [Video game music]	4
Anxiety	12
Depression	12
Insomnia	12
OCD	13
Music effects	3
Permissions	1
dtype: int64	

	count	mean	std	min	25%	50%	75%	max
Age	735.0	2.520680e+01	1.205497e+01	10.0	18.0	21.0	28.0	89.0
Hours per day	736.0	3.572758e+00	3.028199e+00	0.0	2.0	3.0	5.0	24.0
BPM	629.0	1.589948e+06	3.987261e+07	0.0	100.0	120.0	144.0	999999999.0
Anxiety	736.0	5.837636e+00	2.793054e+00	0.0	4.0	6.0	8.0	10.0
Depression	736.0	4.796196e+00	3.028870e+00	0.0	2.0	5.0	7.0	10.0
Insomnia	736.0	3.738451e+00	3.088689e+00	0.0	1.0	3.0	6.0	10.0
OCD	736.0	2.637228e+00	2.842017e+00	0.0	0.0	2.0	5.0	10.0

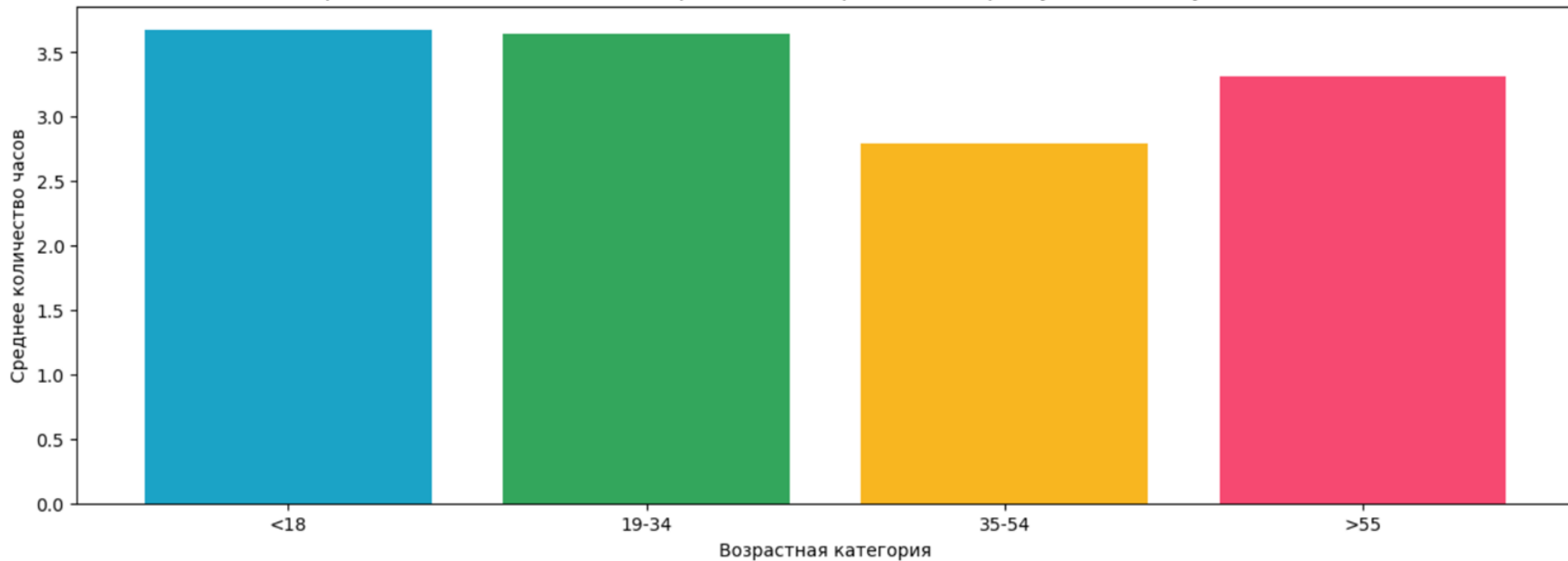
BPM

568	999999999.0
644	624.0
610	220.0
248	220.0
662	218.0
...	
688	NaN
700	NaN
706	NaN
712	NaN
717	NaN

639	0.0
524	0.0
339	0.0
429	4.0
567	8.0
372	20.0
421	40.0
214	52.0
427	55.0
58	55.0

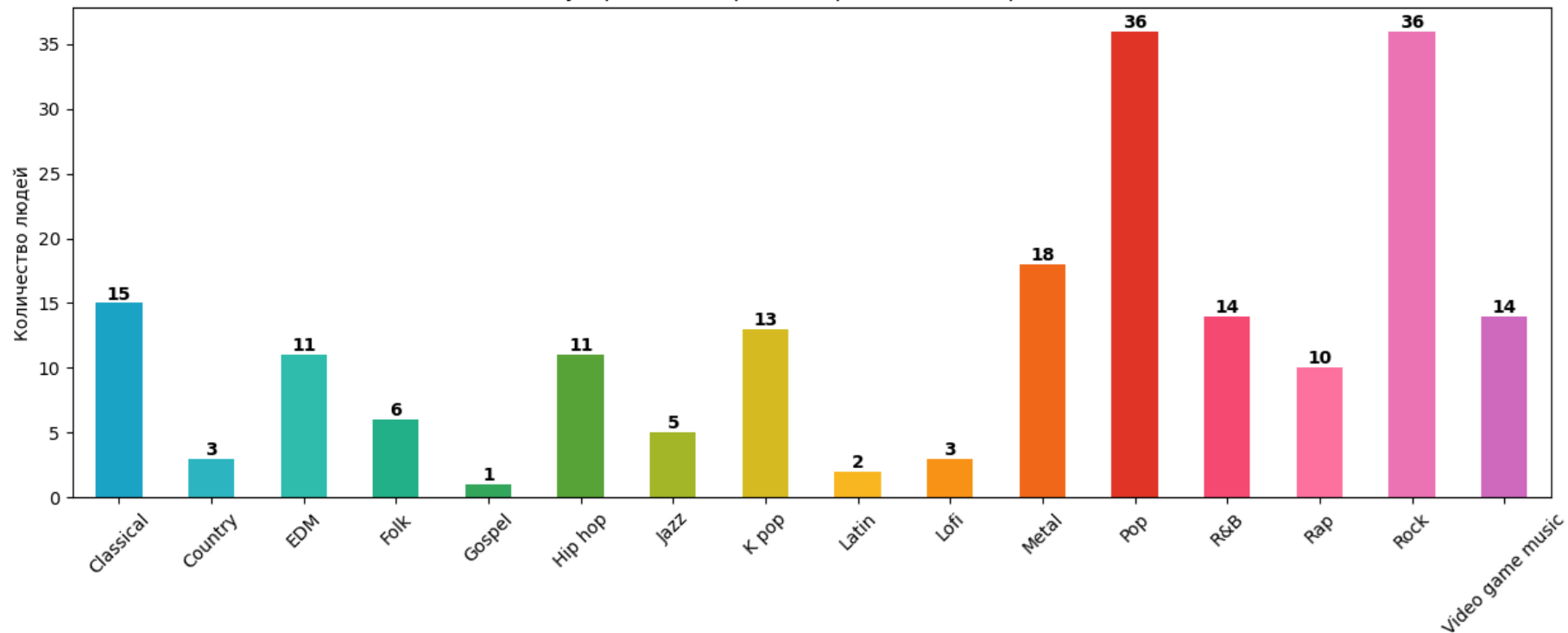
Возрастные категории

Среднее количество часов, которые человек проводит за прослушиванием музыки в день

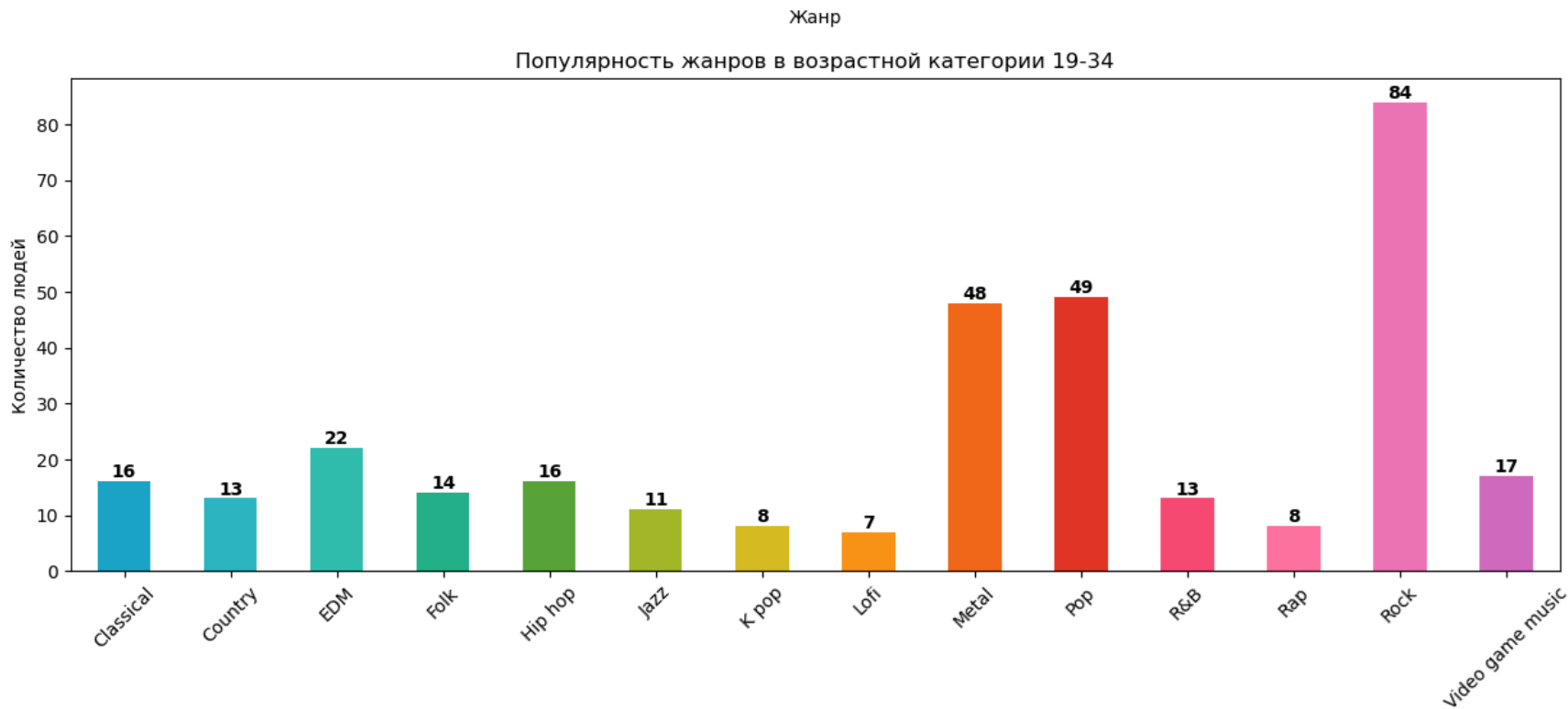


Жанры до 18 лет

Популярность жанров в возрастной категории <18



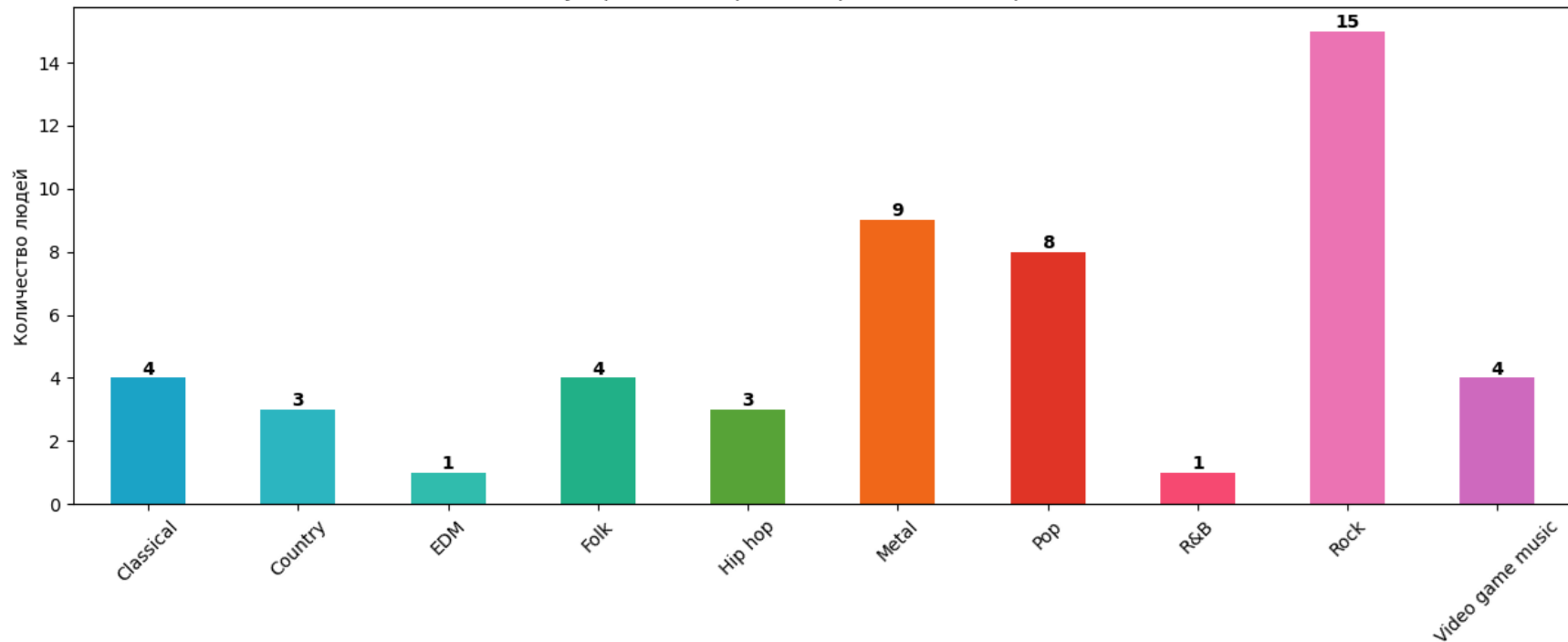
Жанры с 19 до 34 лет



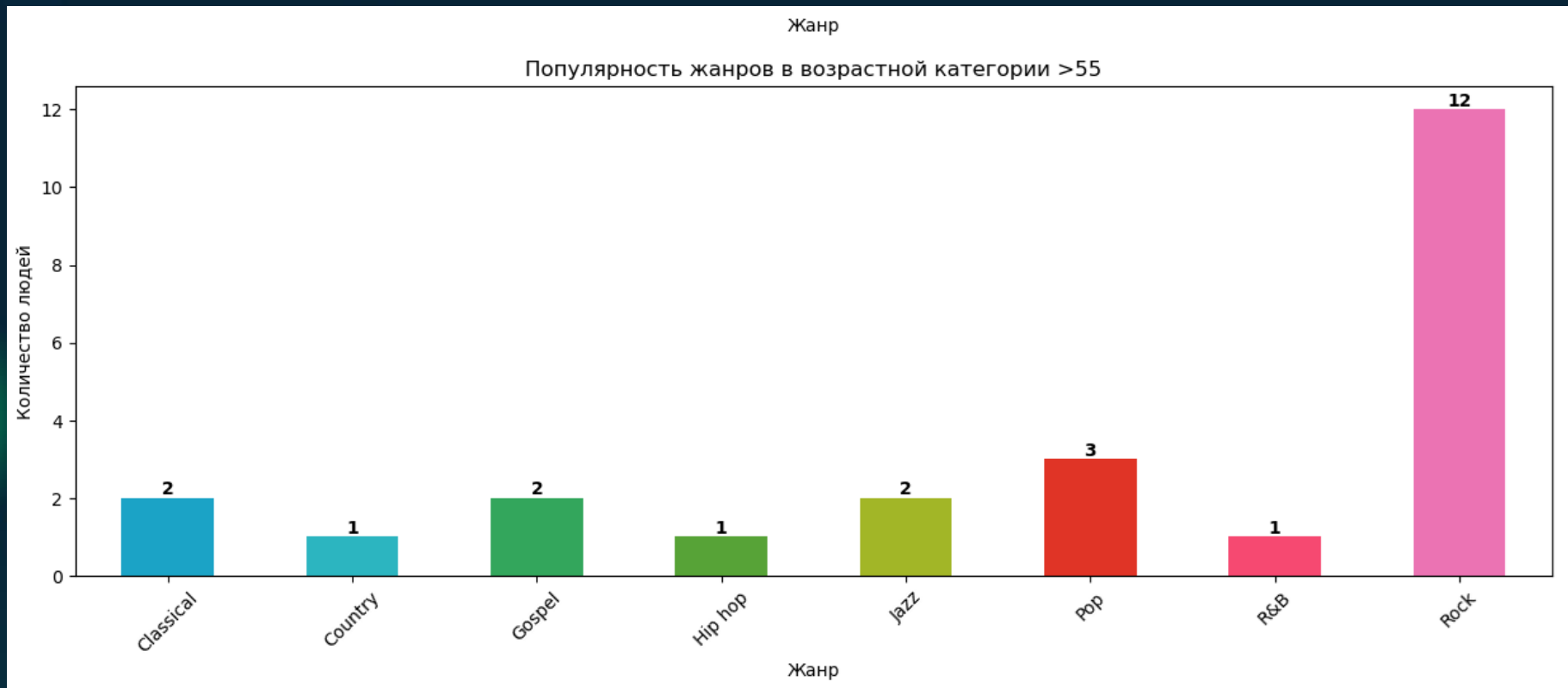
Жанры с 35 до 54 лет

Жанр

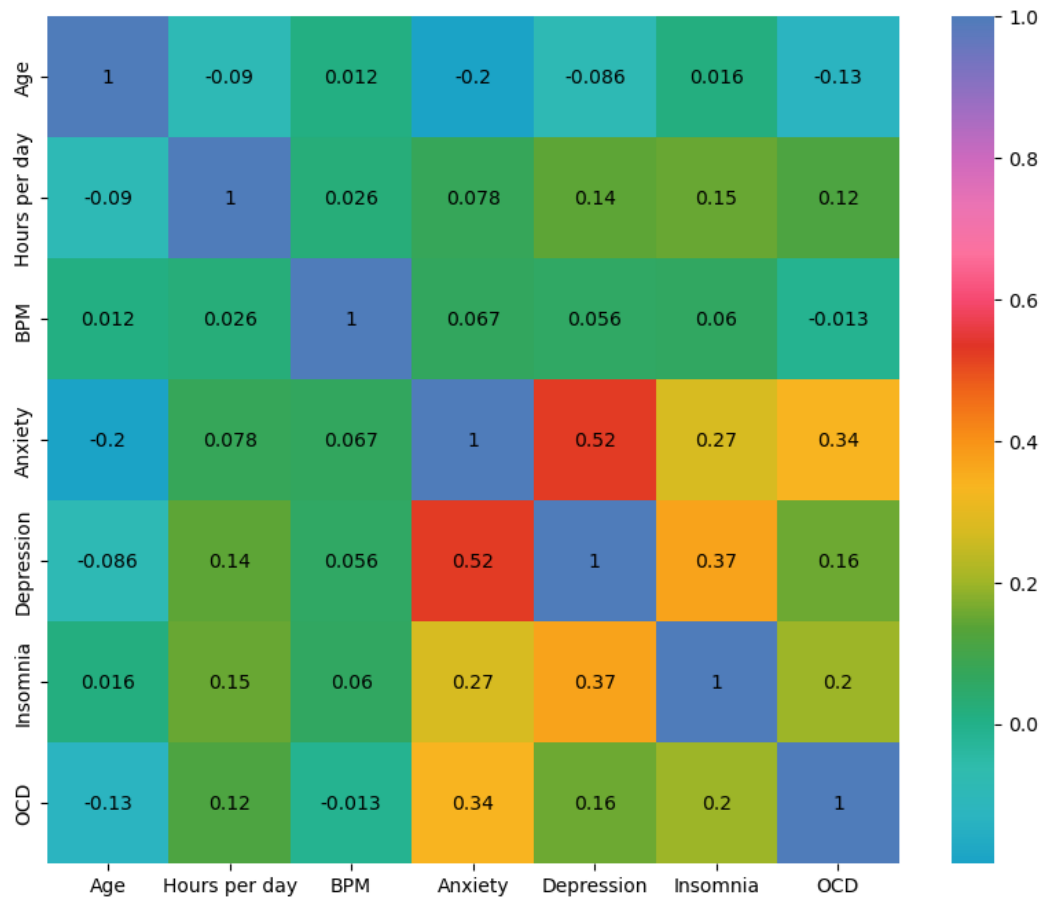
Популярность жанров в возрастной категории 35-54

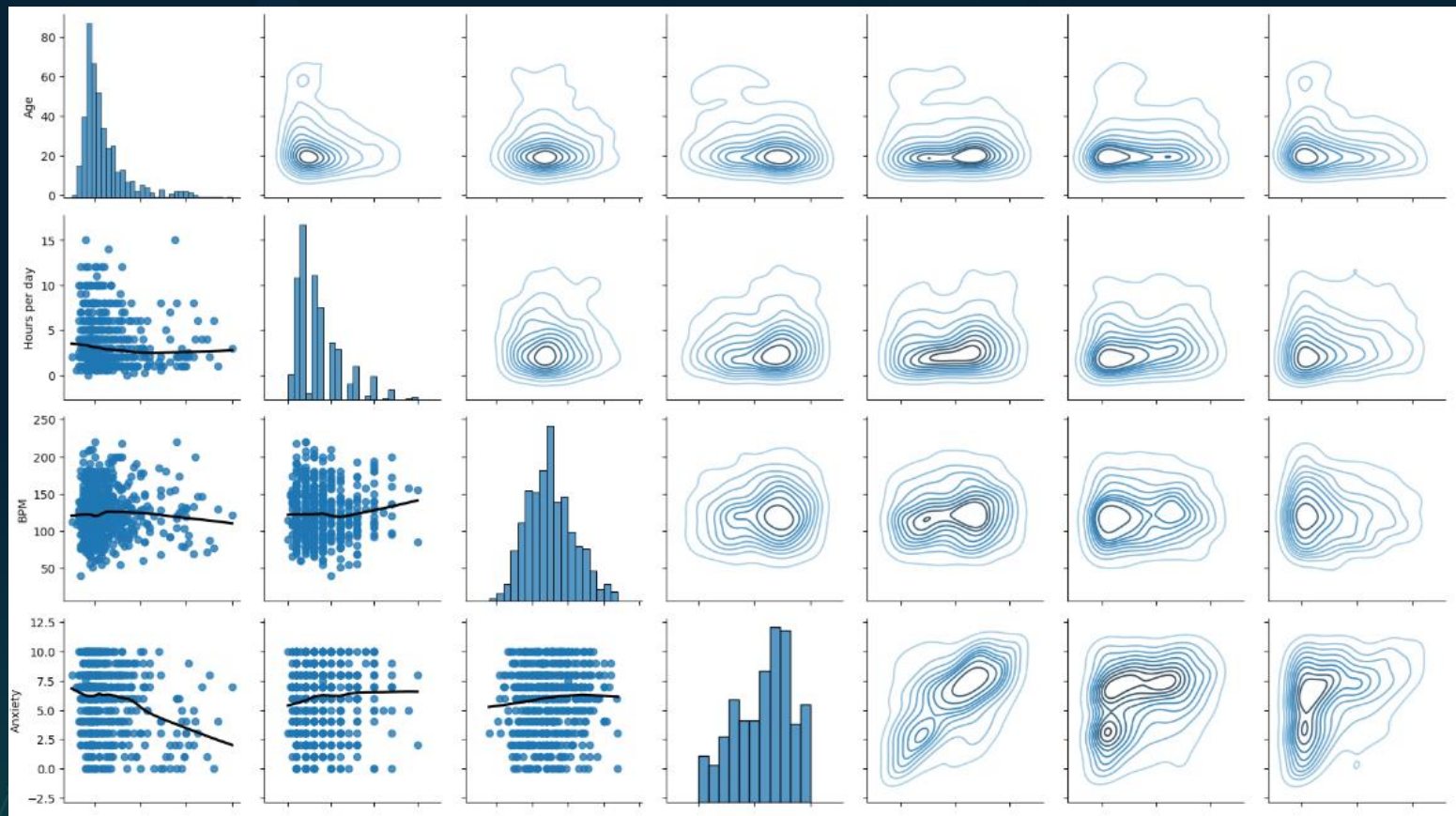


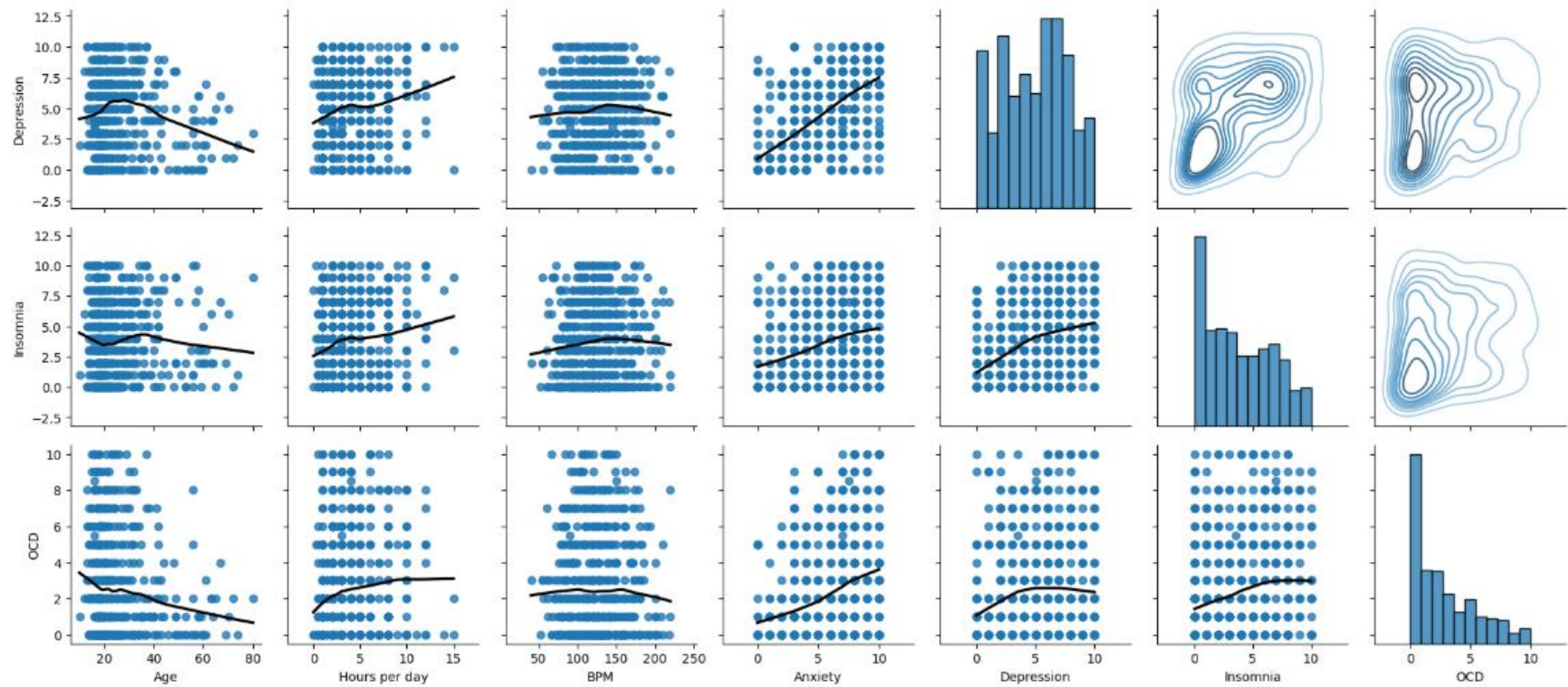
Жанры после 55 лет



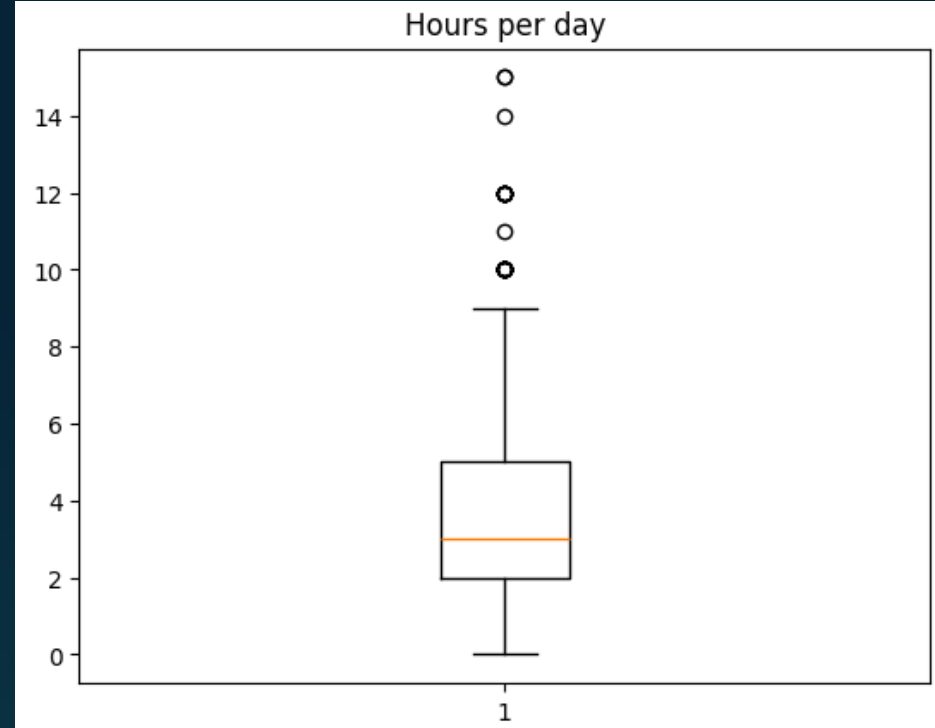
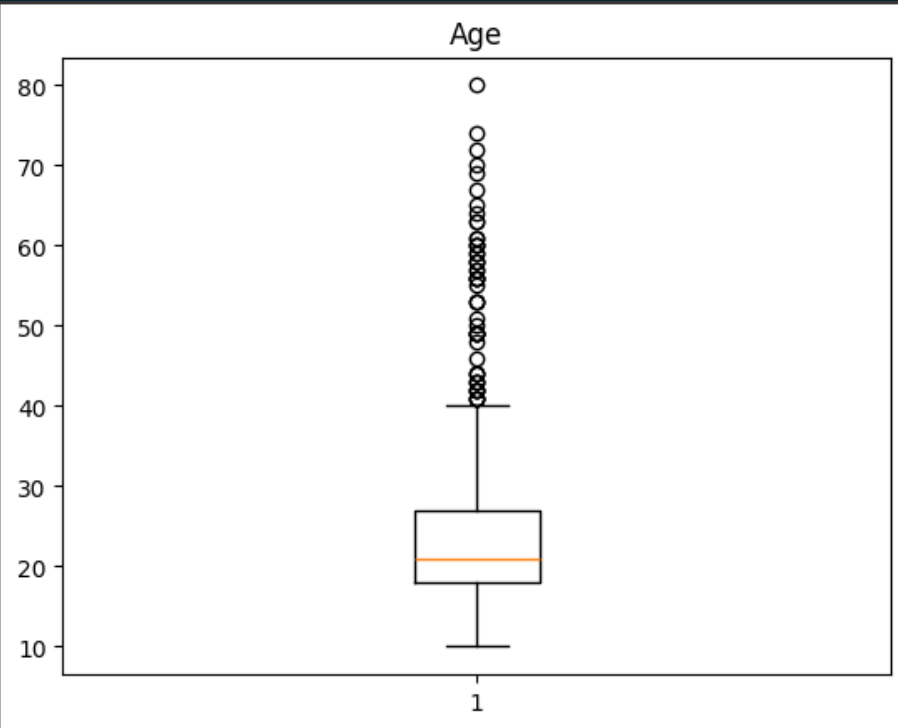
Корреляционная матрица



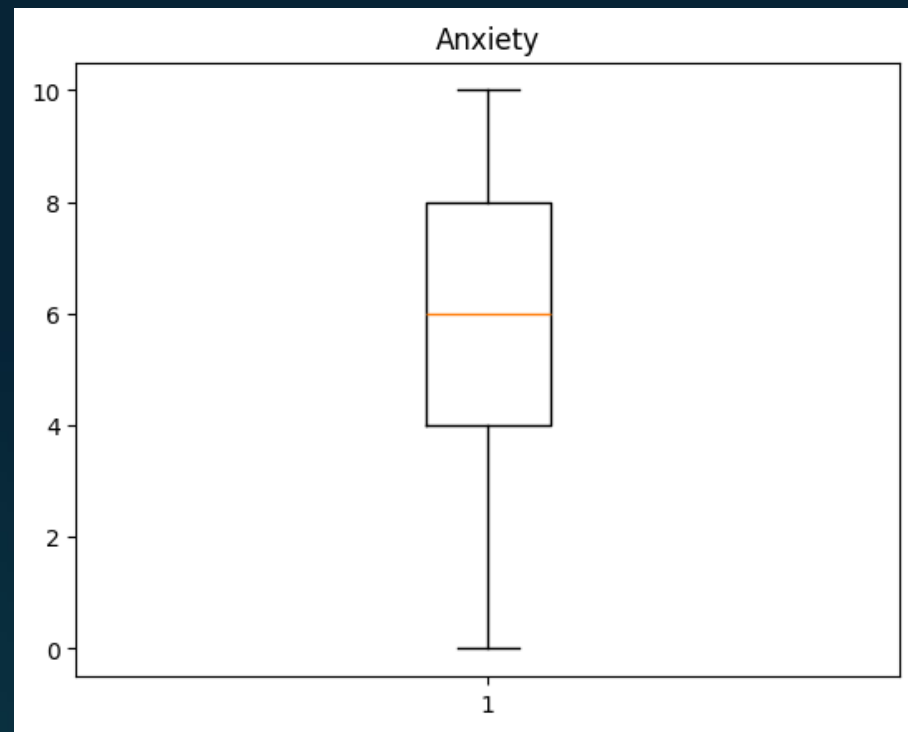
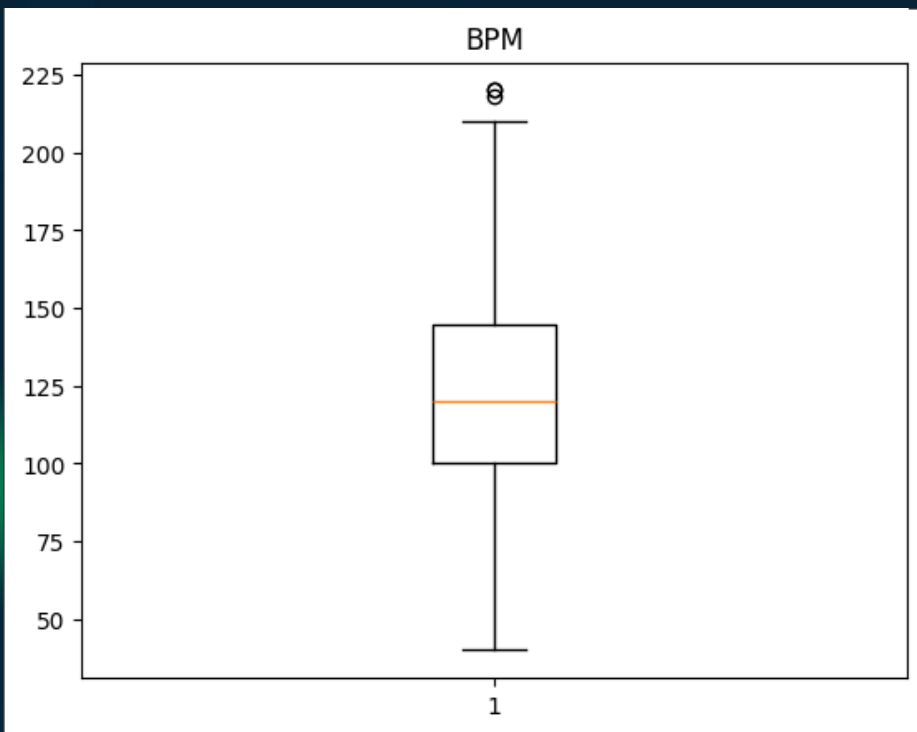




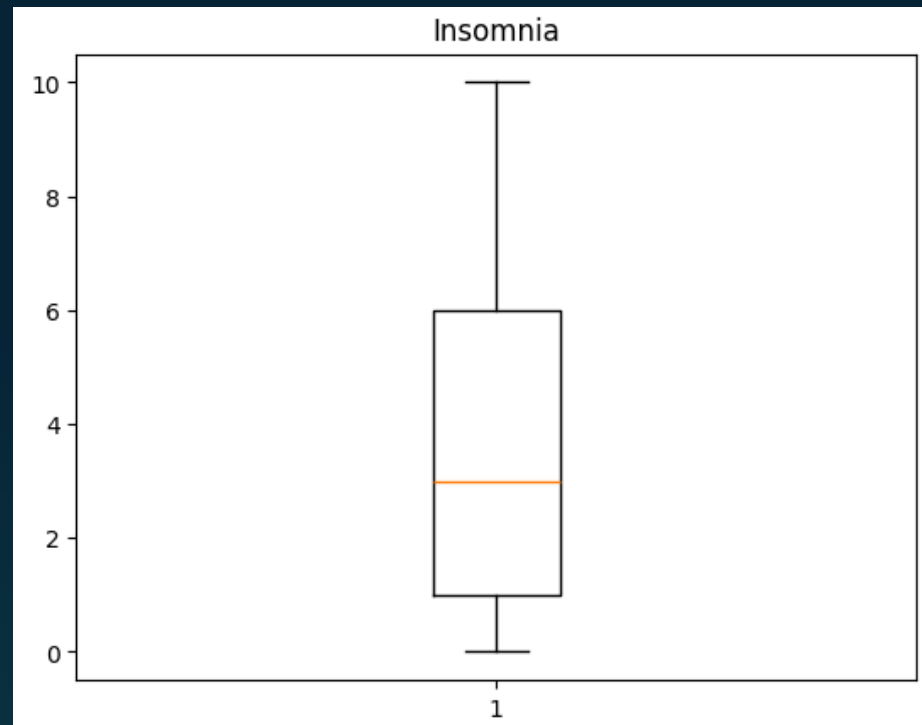
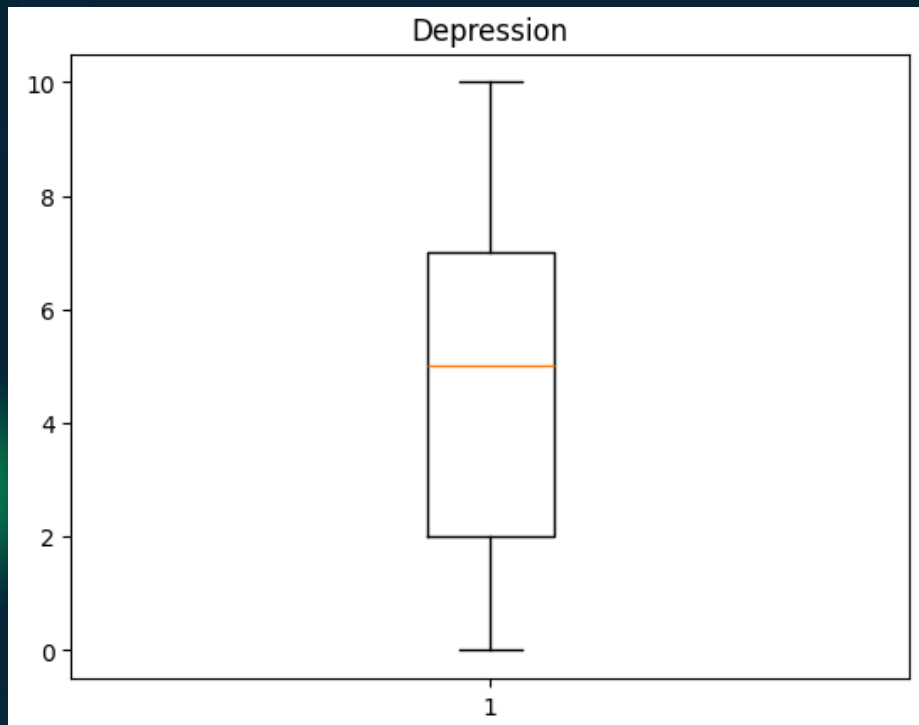
Распределения



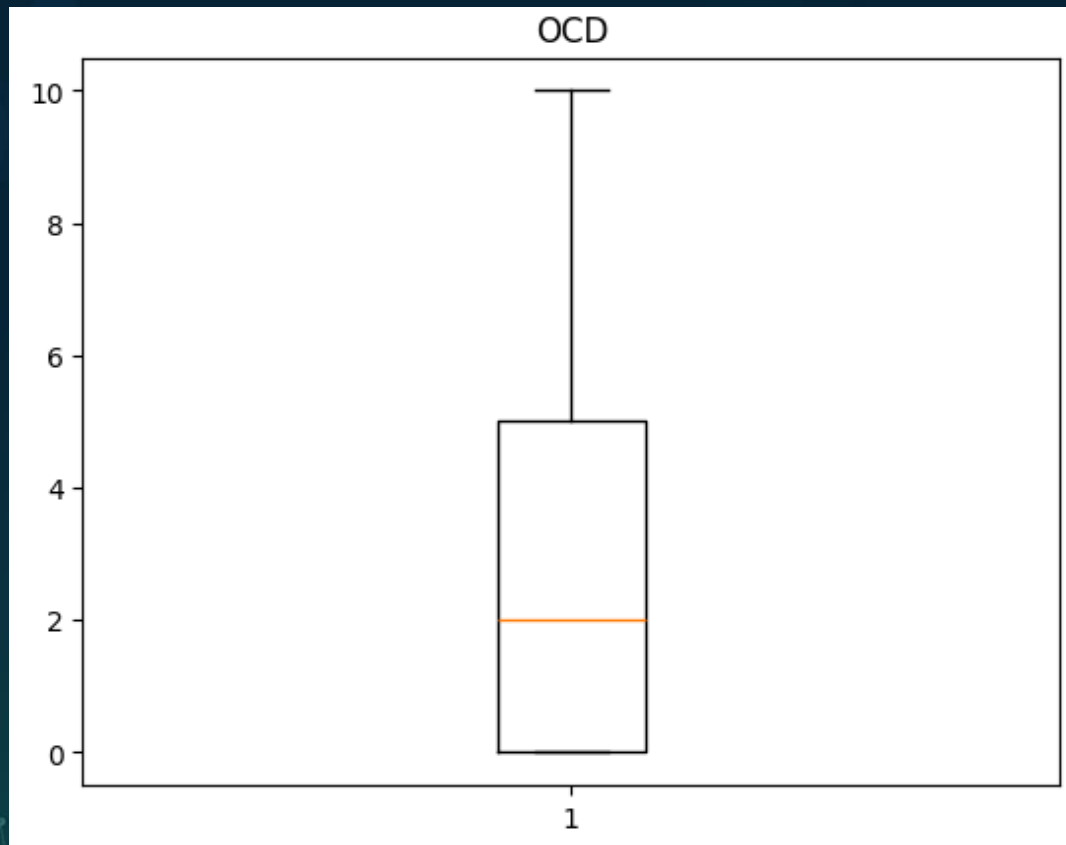
Распределения



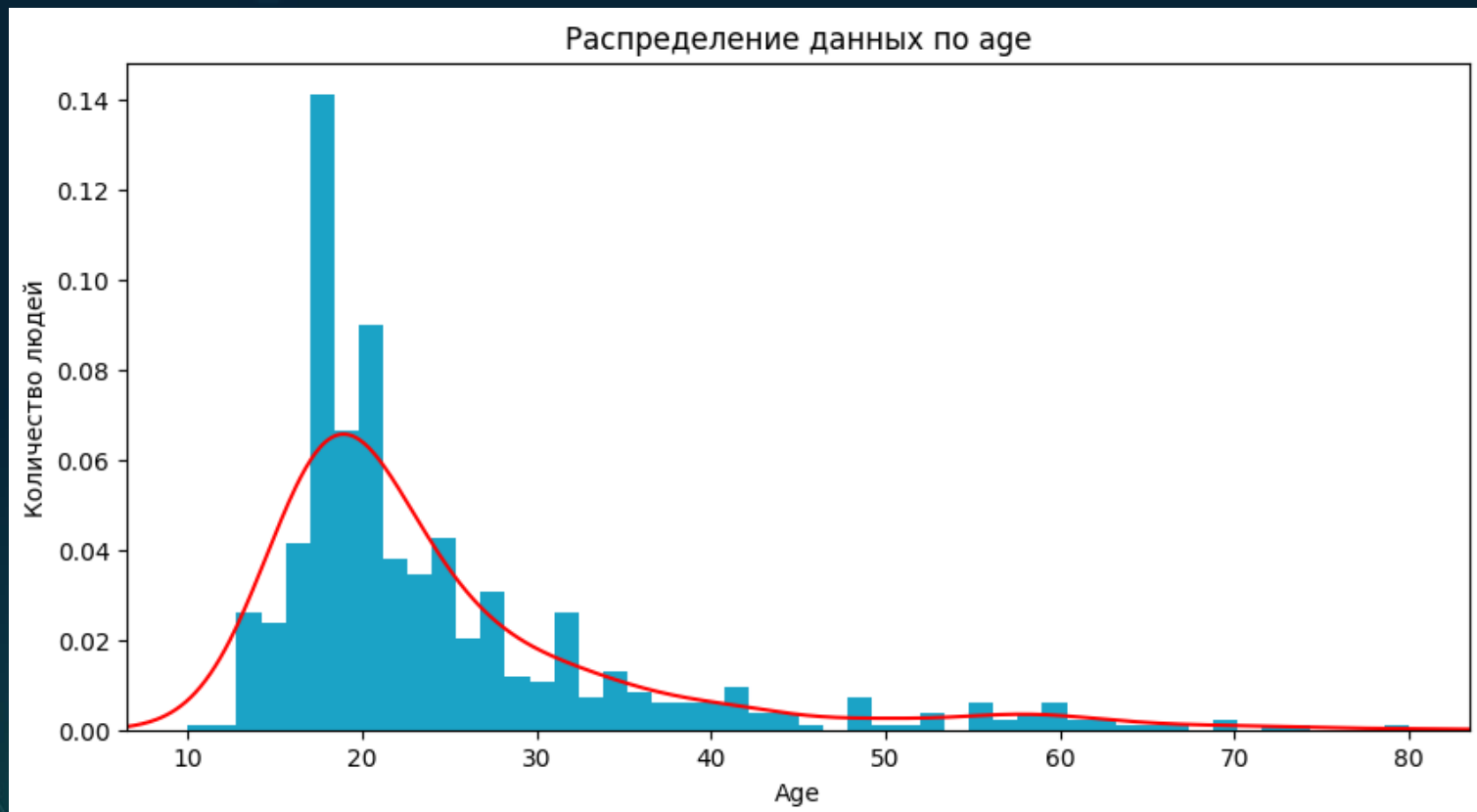
Распределения



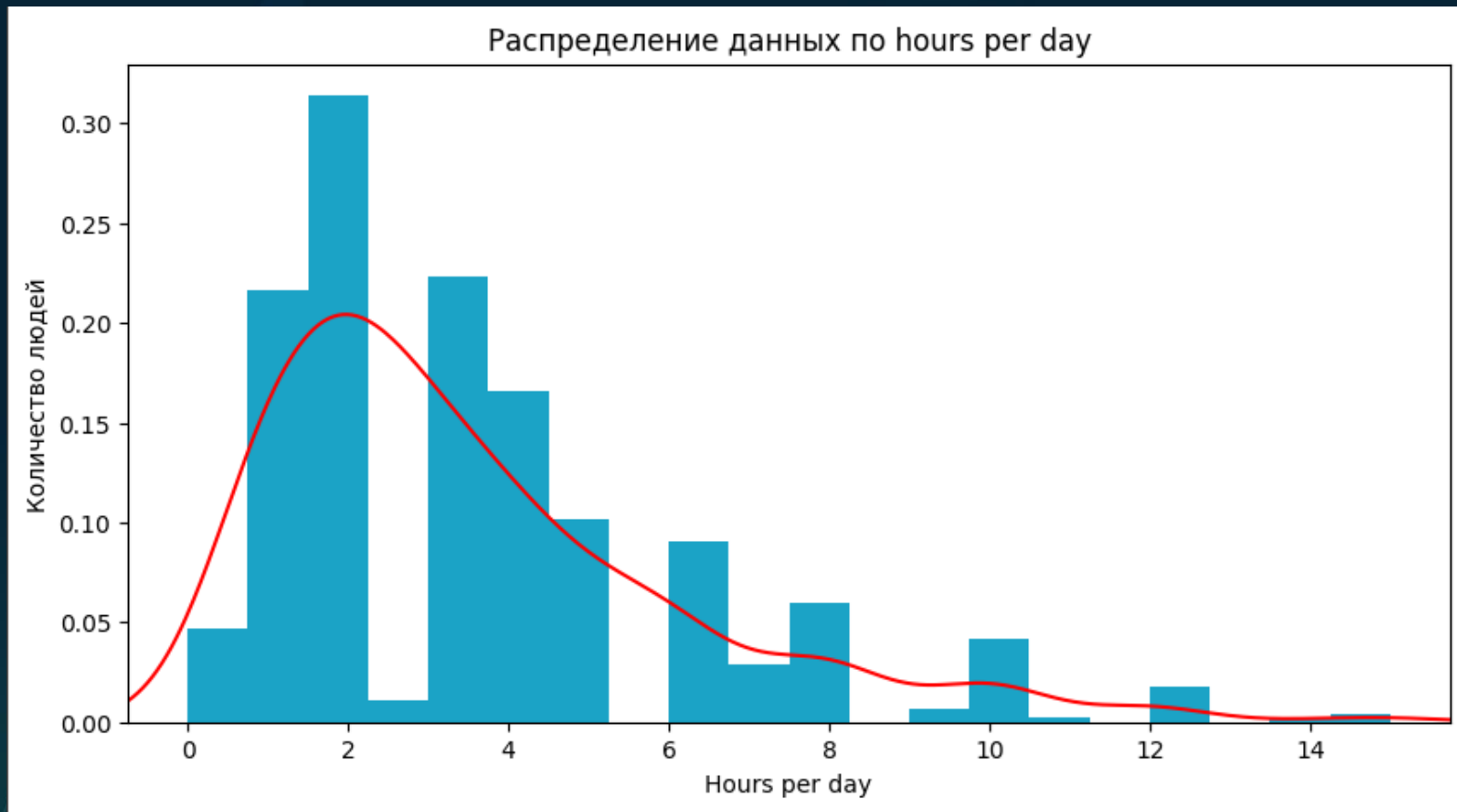
Распределения



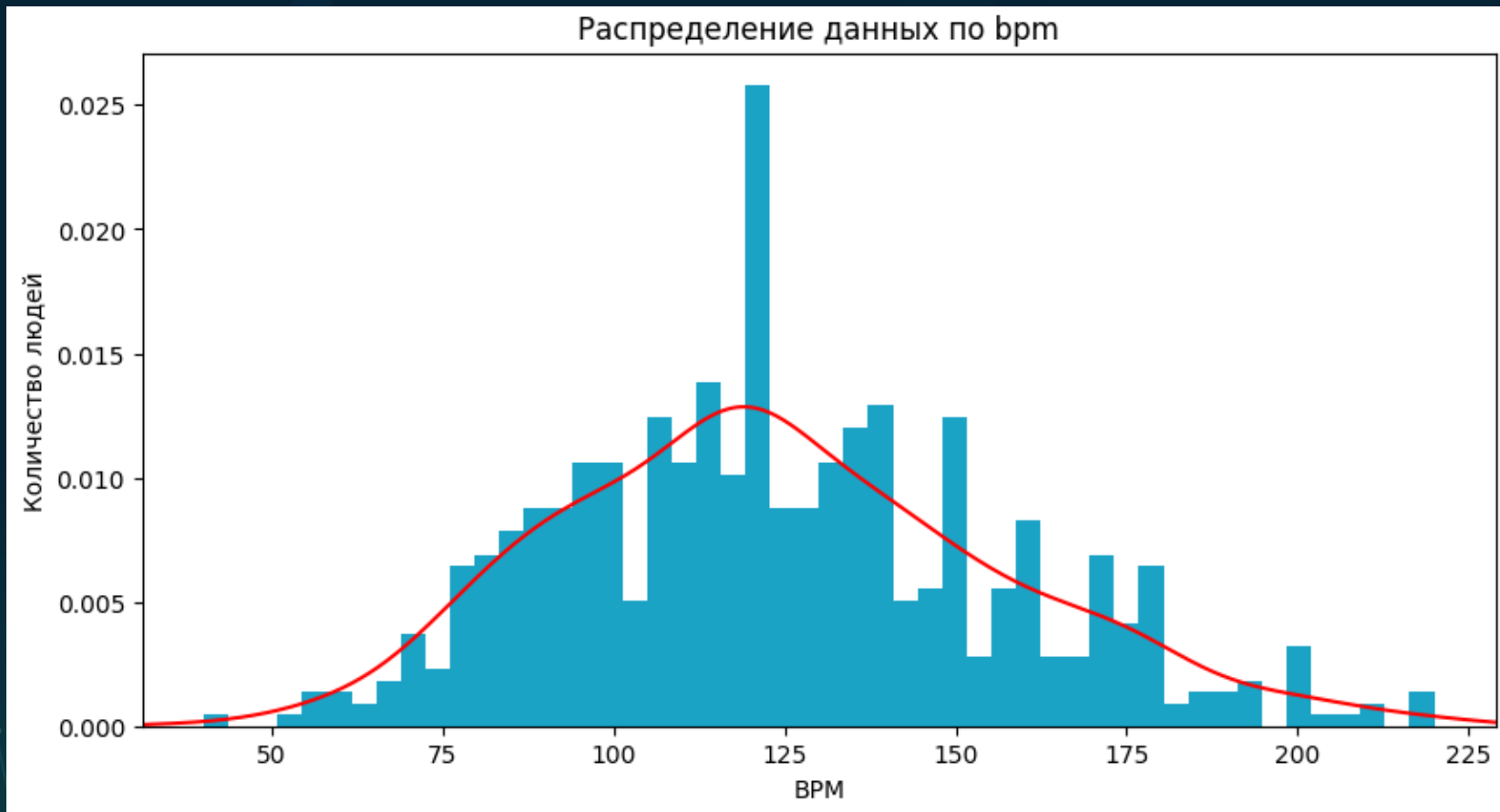
Распределение по возрасту



Распределение по времени

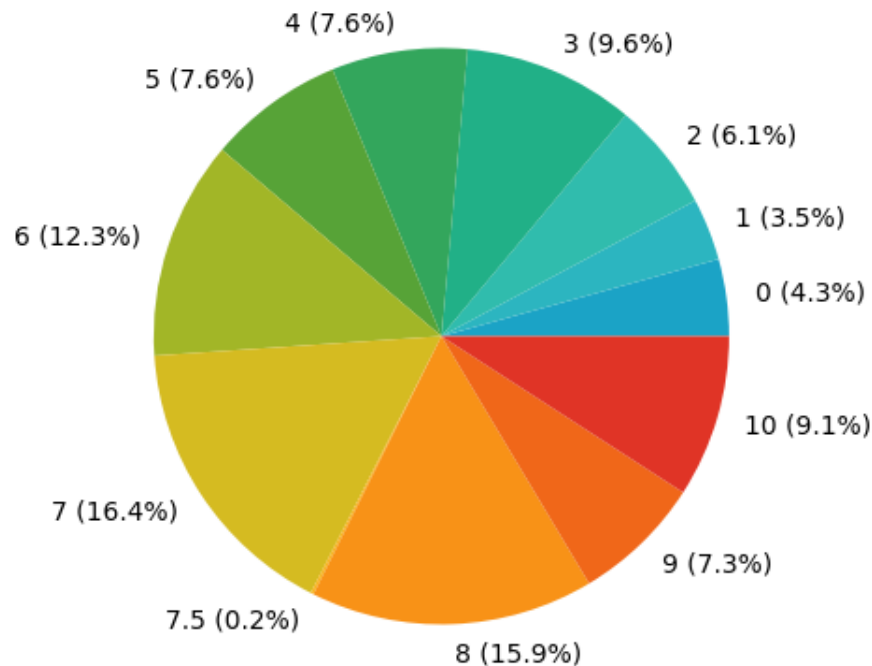


Распределение по BPM

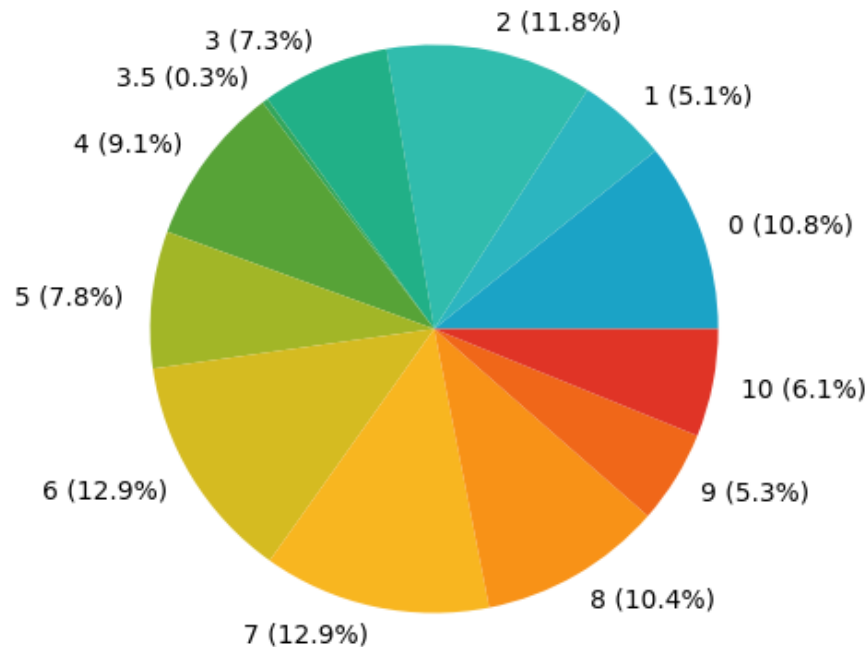


Распределение по болезням

Распределение данных по anxiety

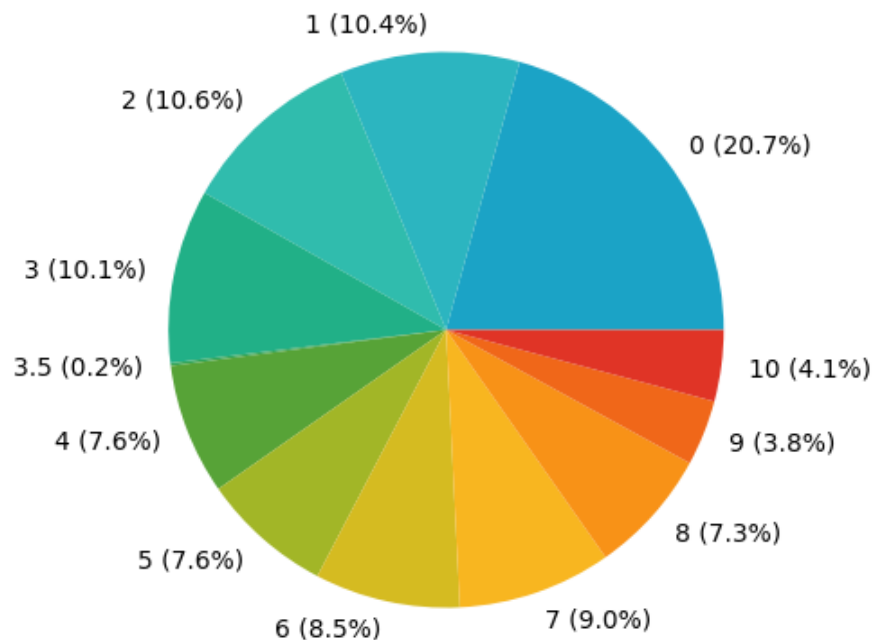


Распределение данных по depression

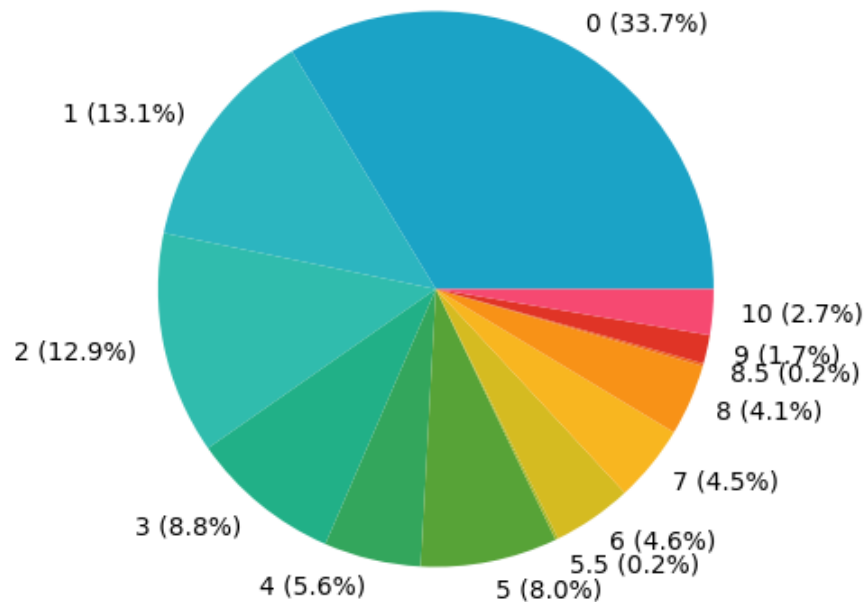


Распределение по болезням

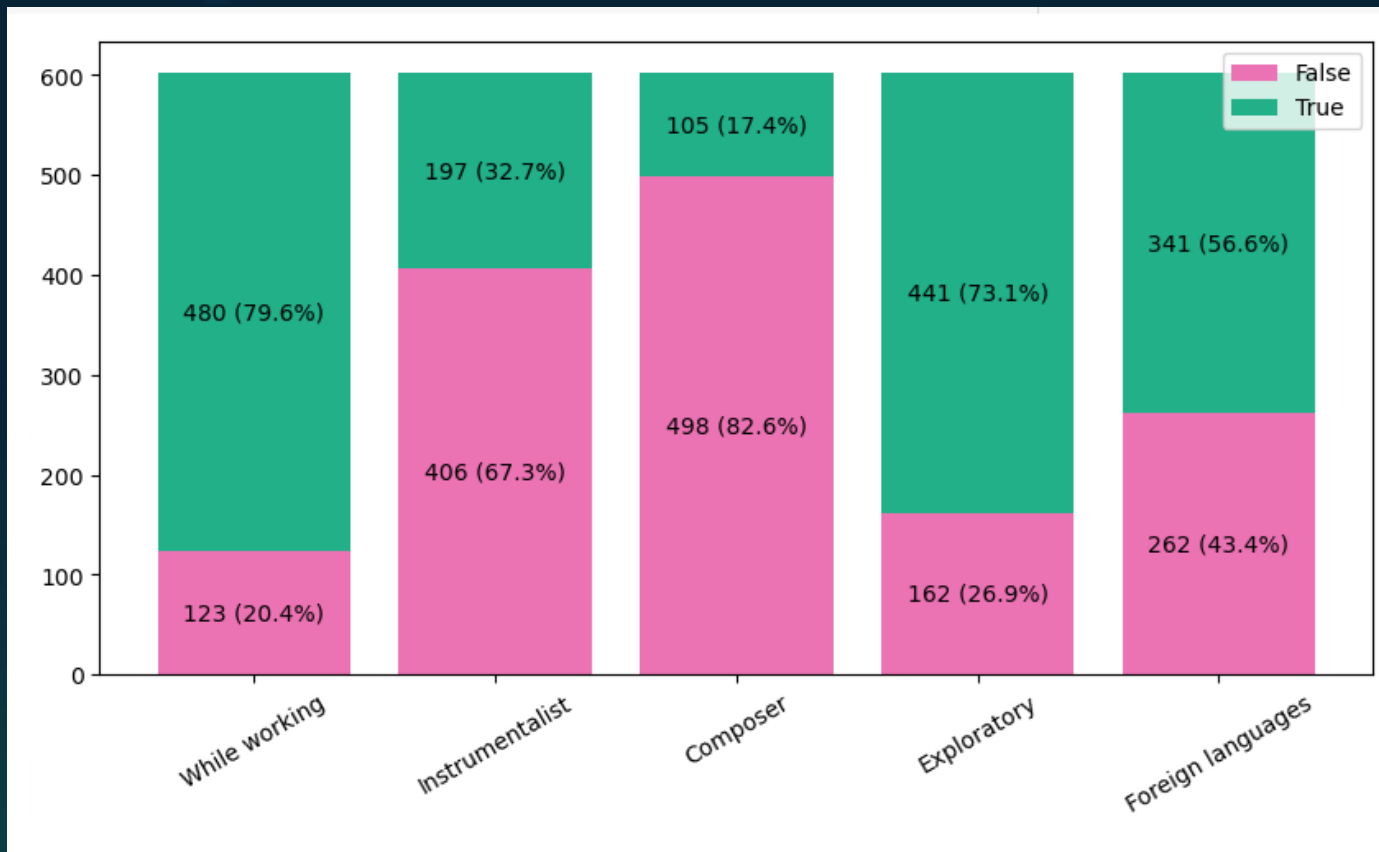
Распределение данных по insomnia



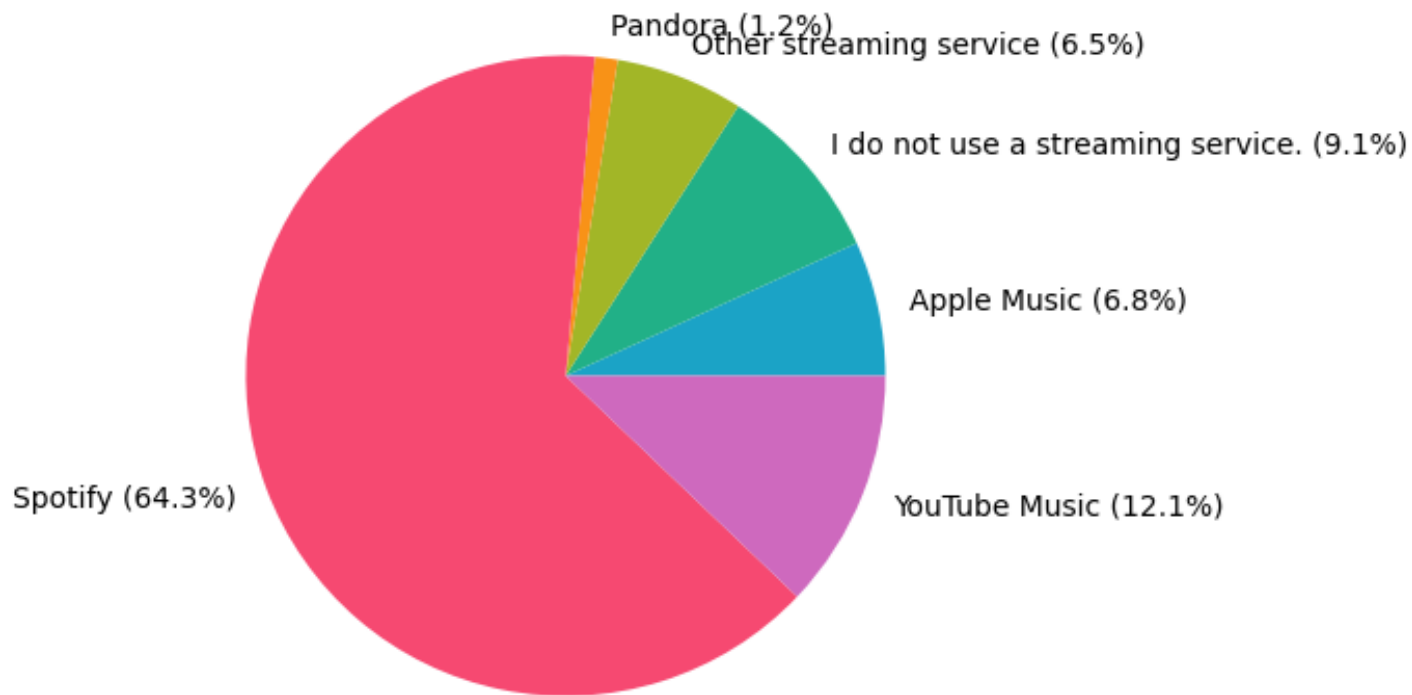
Распределение данных по ocd



Да или Нет

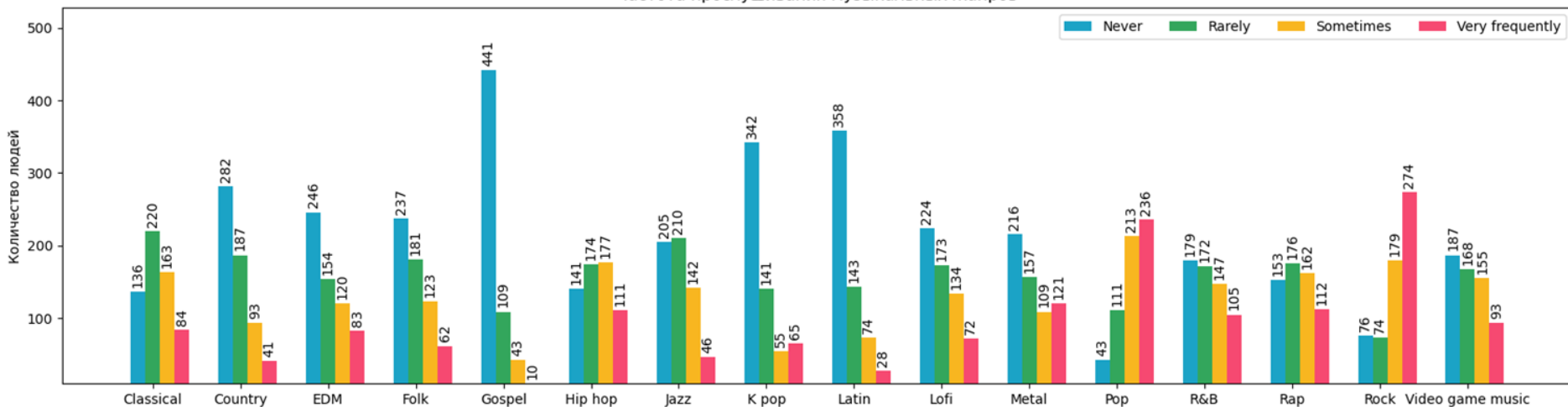


Стриминговый сервис

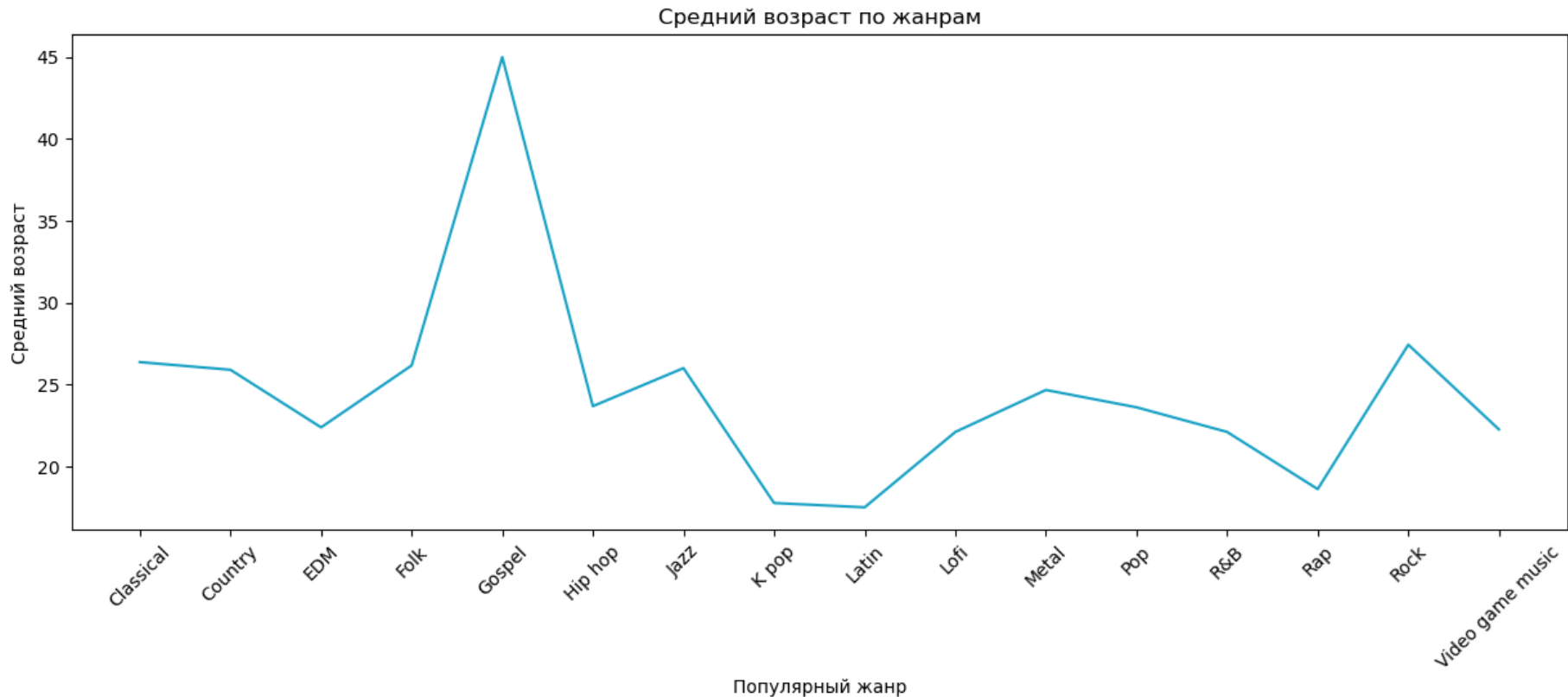


Любимый жанр

Частота прослушивания музыкальных жанров



Средний возраст по жанрам



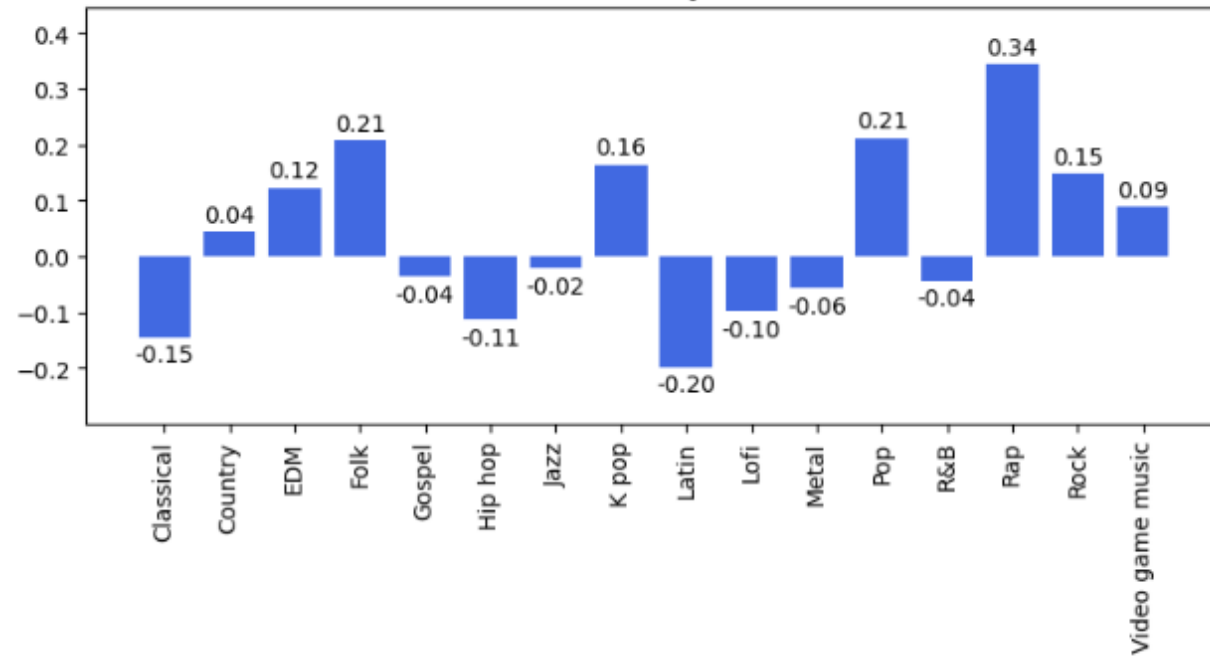
Регрессия

Ordinary Least Squares (OLS) — метод наименьших квадратов.

- Категориальные переменные (частотность прослушивания жанров) преобразуются в числовые значения
- Целевые переменные (Anxiety, Depression, Insomnia, OCD) сохраняются в матрицу Y
- В матрицу X добавляется столбец из единиц для учета свободного члена в регрессии

$$\beta = (X^T X)^{-1} X^T Y$$

Anxiety

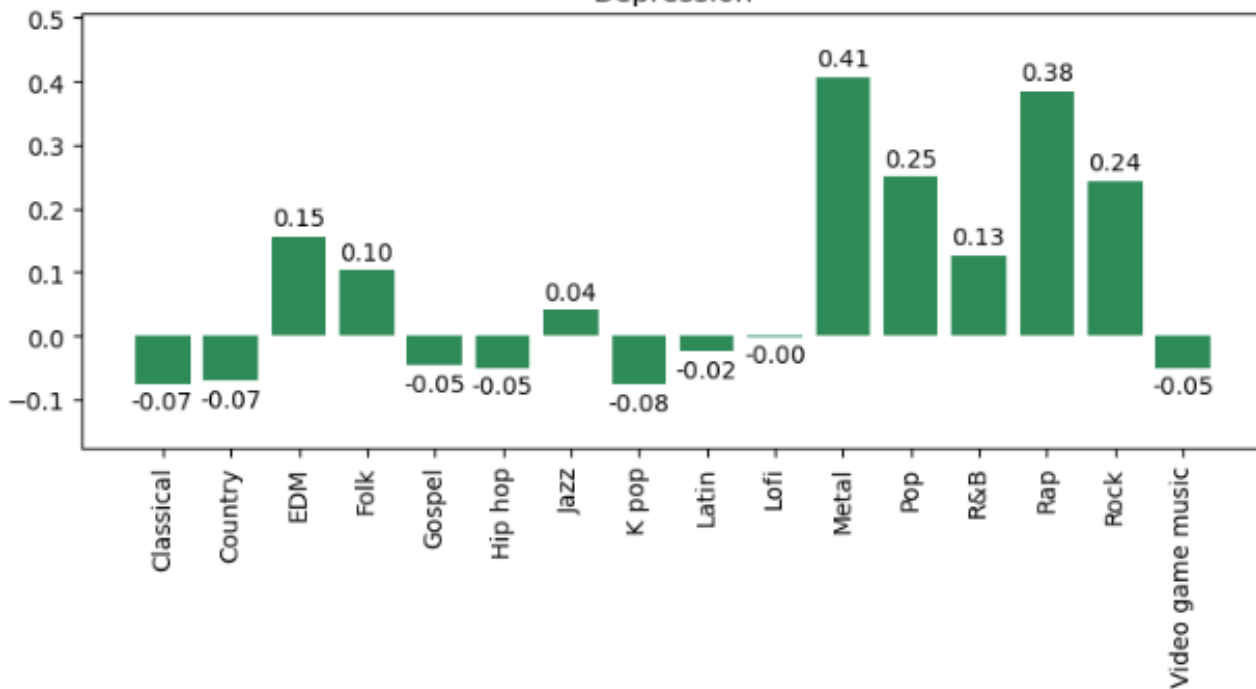


Тревожность

Тревожность (Anxiety)

- **Лучшие жанры:** Classical, Hip hop, Latin
- **Худшие жанры:** Rap, Pop, Folk
- **Музыкальная рекомендация:**
 - **Утро:** Лёгкая классика (например, Бах или Шопен) для расслабления.
 - **Вечер:** Hip hop – оптимистичные мелодии для снятия напряжения.
 - **Исключить:** Агрессивный рэп и интенсивный поп.

Depression

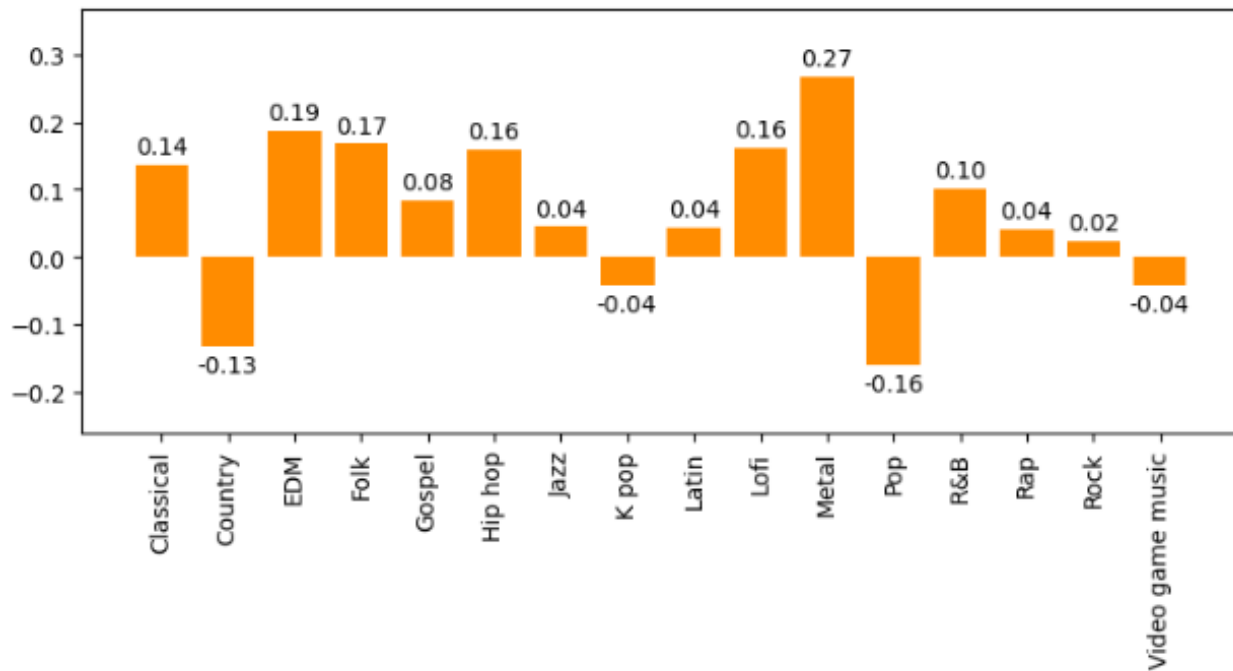


Депрессия

☁ Депрессия (Depression)

- **Лучшие жанры:** Classical, Country, Video game music
- **Худшие жанры:** Metal, Rap, Rock
- 🎧 **Музыкальная рекомендация:**
 - **Утро:** Кантри и классика – мягкое начало дня для эмоционального подъёма.
 - **День:** Саундтреки из игр (например, Stardew Valley OST) – помогают сохранить позитивный настрой.
 - **Исключить:** Тяжёлый метал, интенсивный рок и депрессивный рэп.

Insomnia



Бессоница

☾ Бессонница (Insomnia)

- **Лучшие жанры:** Country, Pop, K-pop
- **Худшие жанры:** Metal, EDM, Folk
- 🎧 **Музыкальная рекомендация:**
 - **Вечер перед сном:** Мягкий кантри и спокойный поп (например, песни Taylor Swift) для успокоения.
 - **Ночная рутина:** Медленные композиции К-поп для расслабления нервной системы.
 - **Исключить:** Метал и быстрые треки EDM перед сном.

OCD



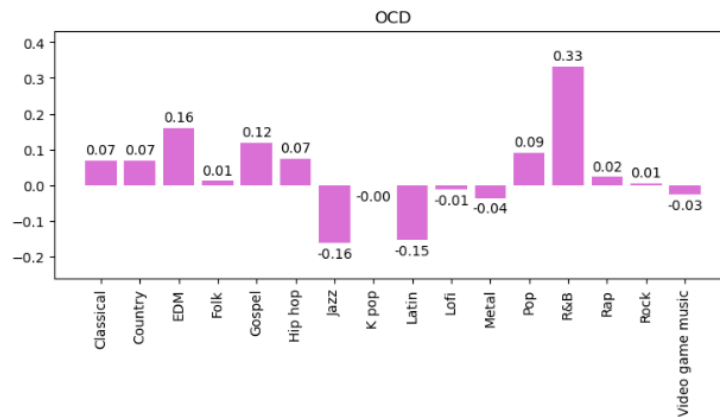
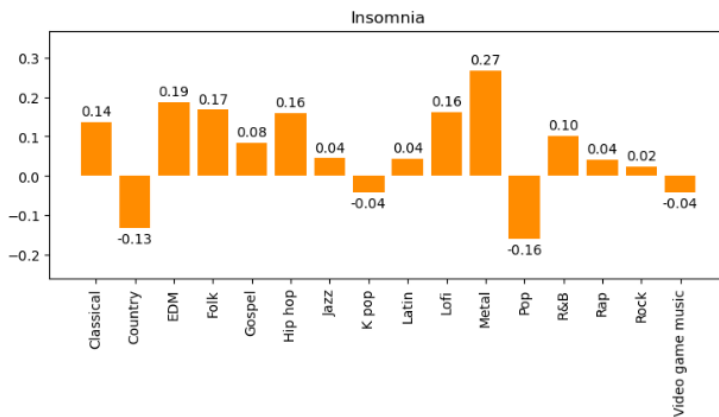
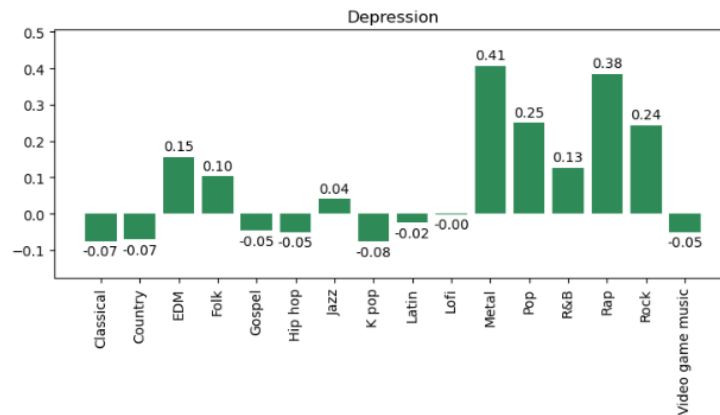
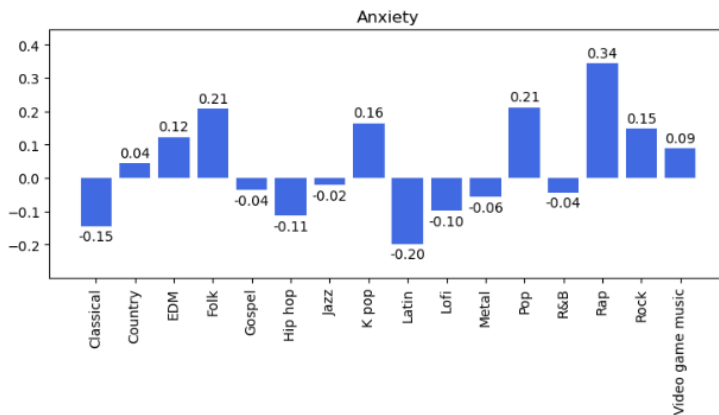
Обсессивно-компульсивное расстройство (OCD)

- **Лучшие жанры:** Jazz, K-pop, Latin
- **Худшие жанры:** R&B, EDM, Folk
- **Музыкальная рекомендация:**
 - **День:** Джазовая музыка для концентрации и уменьшения навязчивых мыслей.
 - **Перерывы в работе:** Весёлые латинские ритмы и K-поп – для переключения внимания и снятия навязчивости.
 - **Исключить:** Эмоционально нагруженные R&B-композиции и монотонный EDM.

ОКР

Полученные коэффициенты модели

Коэффициенты линейной регрессии по жанрам



Пример работы

Формула прогноза в линейной регрессии

Линейная регрессия предсказывает целевую переменную по формуле:

$$\hat{Y} = X \cdot \beta$$

Прогноз для человека 16:

Anxiety: 6.27
Depression: 5.65
Insomnia: 4.84
OCD: 3.72

Реальные ответы человека 16:

Anxiety: 6.0
Depression: 7.0
Insomnia: 5.0
OCD: 4.0

Прогноз для человека 19:

Anxiety: 6.73
Depression: 4.44
Insomnia: 3.69
OCD: 3.12

Реальные ответы человека 19:

Anxiety: 7.0
Depression: 4.0
Insomnia: 2.0
OCD: 5.0

Прогноз для человека 24:

Anxiety: 5.01
Depression: 3.32
Insomnia: 3.05
OCD: 2.11

Реальные ответы человека 24:

Anxiety: 4.0
Depression: 4.0
Insomnia: 4.0
OCD: 3.0

Кластеризация

Подготовка данных для кластеризации:

Числовые признаки: мы используем возраст, количество часов прослушивания музыки в день, переменную While working (переводим ответы "Yes" и "No" в нули и единицы).

Кодирование частот жанров: преобразуем текстовые ответы ('Never', 'Rarely', ...) в числовые значения от 0 до 3.

Формирование итоговой таблицы признаков (x_full): объединяем числовые переменные и жанры.

Масштабирование: применяем StandardScaler, чтобы все признаки имели среднее значение 0 и стандартное отклонение 1.

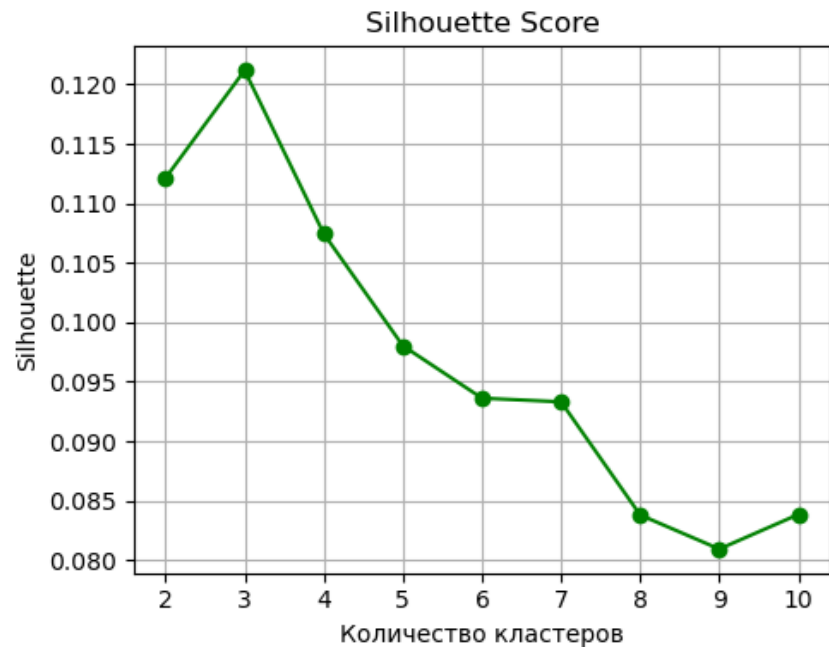
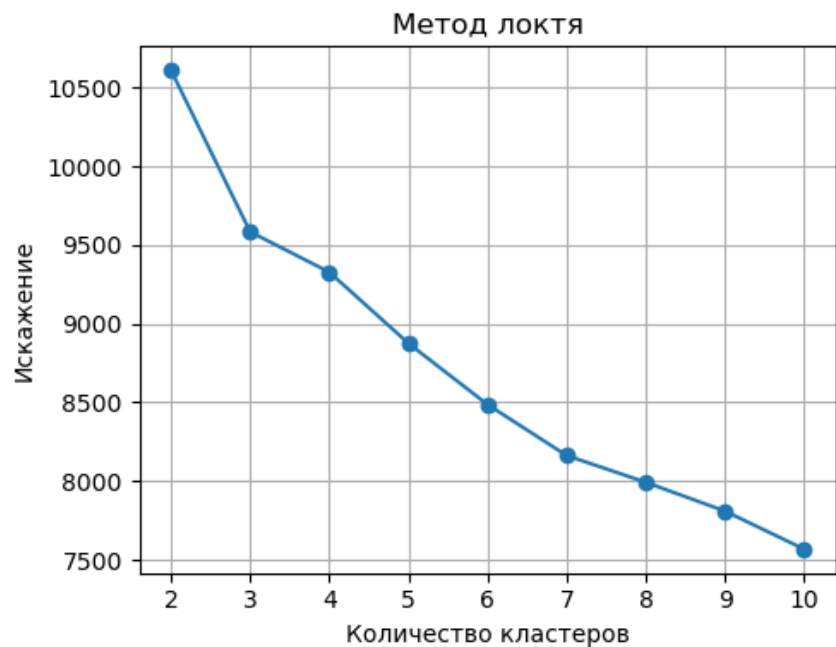
Подбор оптимального количества кластеров

На этом этапе мы определяем, сколько кластеров лучше всего использовать для KMeans, используя два метода:

- 1. Elbow Method**

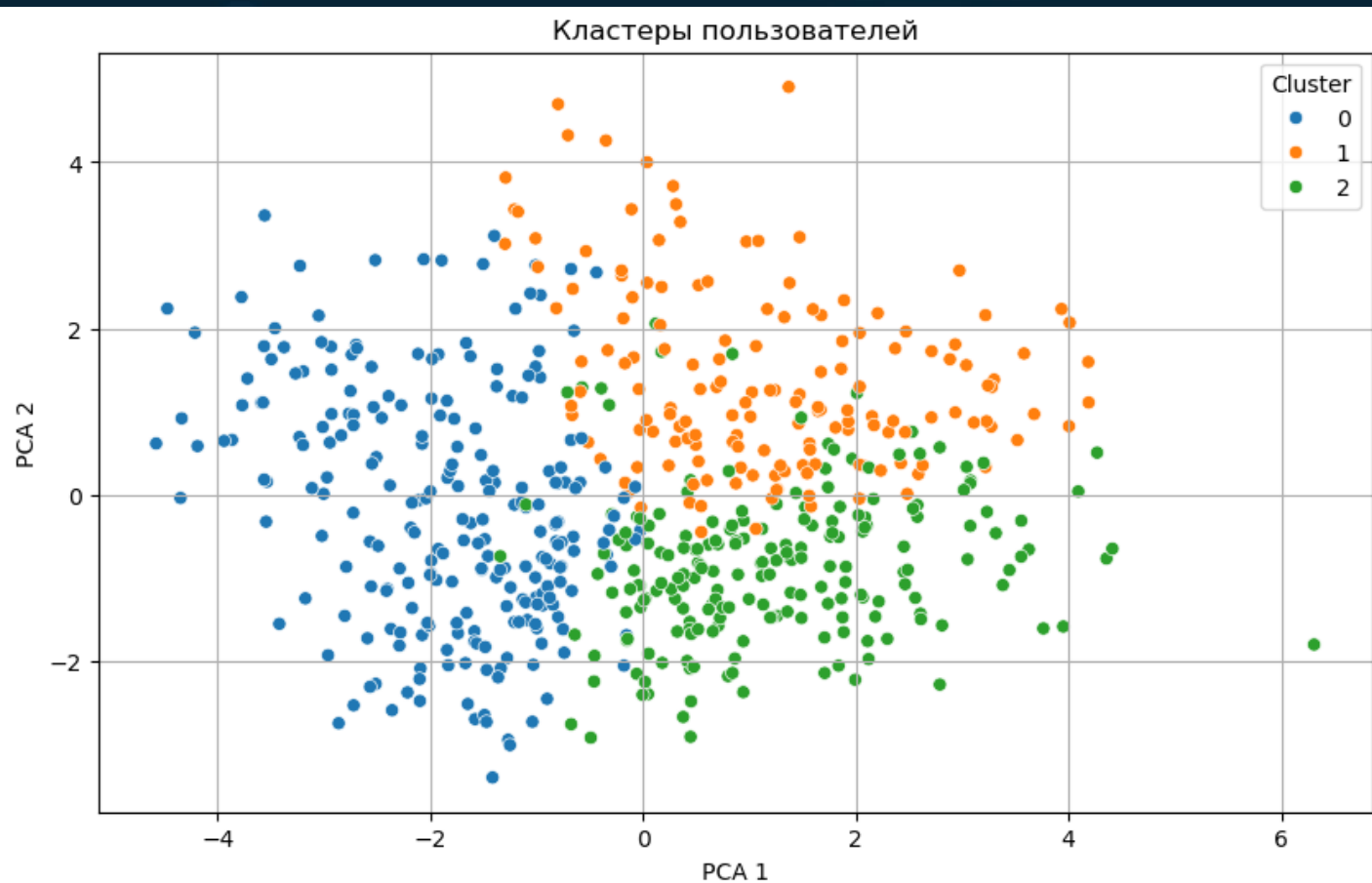
- 2. Silhouette Score**

Результат



Лучший Silhouette Score при $n_clusters = 3$, значение: 0.1212

Кластеризация KMeans



Данные про кластеры

Общие данные про кластеры:

Кластер 0:

- Среднее время прослушивания: 2.8 ч
- Средний возраст: 27.5 лет
- Любимые жанры: Rock, Pop, Classical

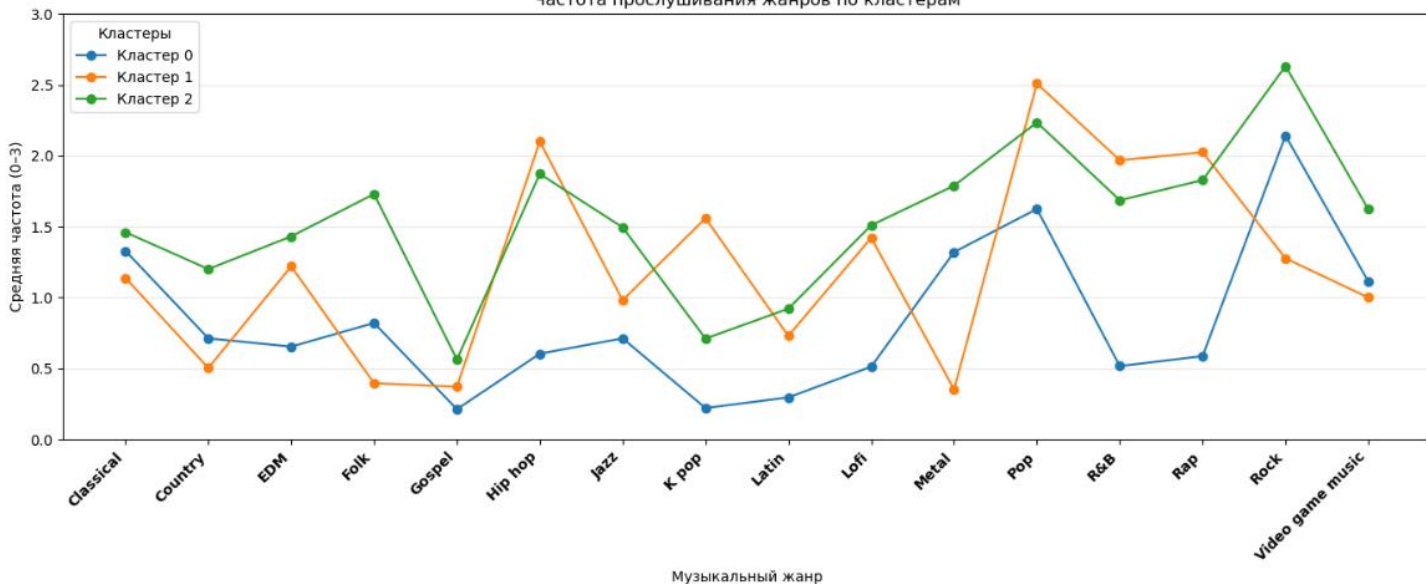
Кластер 1:

- Среднее время прослушивания: 3.6 ч
- Средний возраст: 20.1 лет
- Любимые жанры: Pop, Hip hop, Rap

Кластер 2:

- Среднее время прослушивания: 4.4 ч
- Средний возраст: 24.6 лет
- Любимые жанры: Rock, Pop, Hip hop

Частота прослушивания жанров по кластерам



Выводы



Выводы по кластерам

На основе кластеризации пользователей по музыкальным предпочтениям и возрасту и времени прослушивания музыки можно выделить следующие моменты:



Кластер 0:

- **Наименее вовлечённая группа:** слушают музыку в среднем 2.8 часа в день.
- **Самые старшие:** средний возраст — 27.5 лет.
- **Предпочтения:** Rock, Pop, Classical.
- **Особенности:**
 - На фоне остальных — меньше интереса к Hip hop, Rap, K-pop.
 - Выше интерес к более «традиционным» жанрам (Classic, Rock).



Кластер 1:

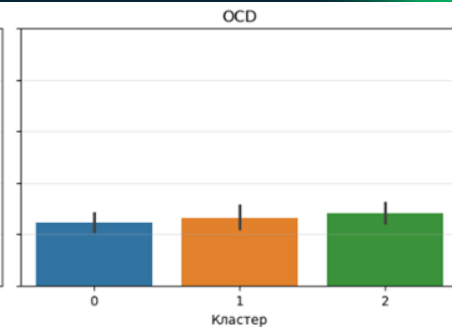
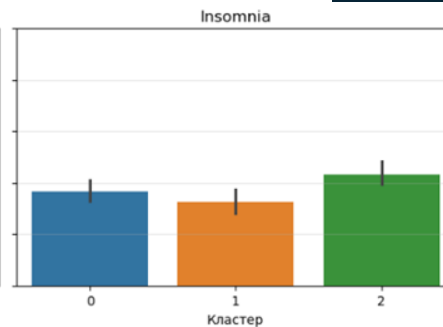
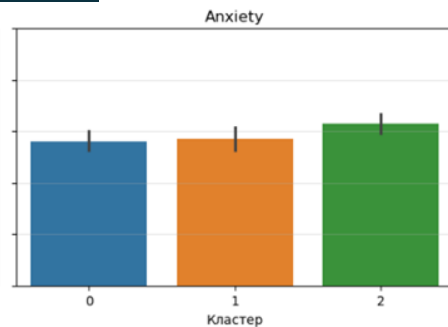
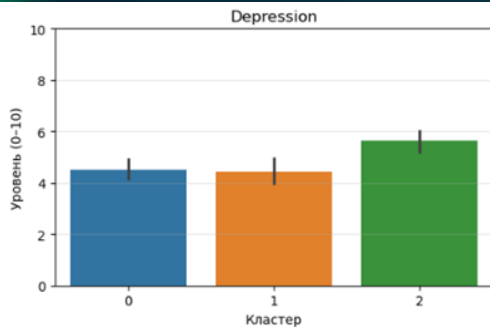
- **Среднеактивные слушатели:** 3.6 ч/день.
- **Самые молодые:** 20.1 года.
- **Предпочтения:** Pop, Hip hop, Rap.
- **Особенности:**
 - Ярко выраженный интерес к современной популярной и танцевальной музыке.
 - Максимальные значения по жанрам: Pop, R&B, K-pop.



Кластер 2:

- **Самые активные:** в среднем 4.4 ч/день.
- **Средний возраст** — 24.6 лет.
- **Предпочтения:** Rock, Pop, Hip hop.
- **Особенности:**
 - Высокие значения почти по всем жанрам — это самая вовлечённая и музыкально разнообразная группа.
 - Особенно выделяются: Rock, Metal, Video game music.

Статистика заболеваний по кластерам



Связь заболеваний и кластеров



Связь музыкальных предпочтений с психоэмоциональным состоянием

На основе анализа психоэмоциональных показателей можно сделать следующие выводы:

■ Кластер 2 — эмоционально перегруженные меломаны

- Пользователи из этого кластера проявляют **наиболее высокие уровни депрессии, тревожности и бессонницы**.
- Их музыкальный профиль охватывает **много разнообразных и насыщенных жанров** — от металла и рока до lo-fi и фолка.
- Это может указывать на **попытку справиться с внутренним напряжением через музыку**, либо на то, что такие жанры выбираются людьми с уже высоким уровнем эмоционального напряжения.
- Высокая музыкальная вовлечённость — **возможно, форма самотерапии**, но не всегда эффективно снижающая стресс.

■ Кластер 0 — умеренно уравновешенные слушатели

- Состояние участников сбалансировано, но не идеально: показатели тревожности и бессонницы не самые низкие.
- Эти люди **слушают музыку умеренно и довольно разносторонне**, но не в эмоционально напряжённых жанрах.
- Такое распределение может говорить о том, что **музыка не является для них главным инструментом эмоциональной регуляции**, а скорее привычной фоново-социальной активностью.

■ Кластер 1 — молодые и устойчивые

- Самые **низкие уровни психоэмоционального стресса**, особенно по депрессии и бессоннице.
- Ядро их музыкального вкуса — **современная популярная музыка с позитивной и энергичной атмосферой** (pop, k-pop, R&B, hip-hop).
- Это может указывать на то, что **простая, ритмичная и социально вовлечённая музыка** способствует психологической устойчивости или, наоборот, её предпочитают устойчивые люди.



Обобщения и гипотезы

- Не сами жанры вызывают стресс, но **выбор жанров может быть связан с внутренним состоянием**.
- Чем **богаче и мрачнее музыкальный профиль**, тем **выше вероятность эмоционального напряжения**.
- Жанры с акцентом на **агрессию, драму или меланхолию** (metal, rock, lofi, folk) — чаще выбираются кластерами с повышенной тревожностью.
- Более **лёгкие и социальные жанры** (pop, R&B, hip-hop) — связаны с **низкими уровнями стресса**, что потенциально может использоваться в музыкальной терапии.

Спасибо за внимание!