

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ
ФЕДЕРАЦИИ**

**Федеральное государственное автономное образовательное
Учреждение высшего образования
«ЮЖНЫЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Институт высоких технологий и пьезотехники**



**Кафедра прикладной информатики и
инноватики**

**Направление подготовки:
09.03.03 "Прикладная информатика"**

Отчёт

По дисциплине: «Большие данные»

**Сбор, предобработка и анализ данных о музыкальных предпочтениях и
ментальном состоянии**

Выполнили студенты 3 курса 21ВТ-09.03.03.01-о3 группы:

_____ Епифанов Е.В.
подпись

_____ Скоробогатов Д.И.
подпись

Проверил старший преподаватель

_____ Яценко Д. В
подпись

Ростов-на-Дону – 2024

СОДЕРЖАНИЕ

1. ПОСТАНОВКА ЗАДАЧИ.....	3
2. ОПИСАНИЕ ДАТАСЕТА.....	4
3. ХОД РАБОТЫ.....	5
3.1 ГИПОТЕЗА.....	5
3.2 ВИЗУАЛИЗАЦИЯ ДАННЫХ ПО МУЗЫКАЛЬНЫМ ПРЕДПОЧТЕНИЯМ И МЕНТАЛЬНОМ СОСТОЯНИИ	6
3.3 СТАТИСТИЧЕСКИЕ ДАННЫЕ	7
3.4 РЕЗУЛЬТАТЫ ПРОДЕЛАННОЙ РАБОТЫ	8
4. ВЫВОДЫ.....	10
5. СПИСОК ЛИТЕРАТУРЫ	11

1. Постановка задачи

Данный кейс предлагает задачу анализа музыкальных предпочтений и их связи с ментальным состоянием человека.

Музыка в истории человечества существует уже более 30 000 лет. Связь между ней и психическим здоровьем в последние 25 лет широко обсуждается из-за ее глубокого влияния на самочувствие. Музыкальная терапия использует музыку для снижения стресса, улучшения настроения и психического благополучия. Она основана на фактических данных и способствует выработке гормонов счастья, таких как окситоцин. Исследования показывают преимущества музыкальной терапии при психических расстройствах, таких как депрессия, травмы, шизофрения. Музыка помогает справиться с эмоциями, действуя как успокаивающее средство при тревоге и нарушении регуляции.

Именно поэтому анализ данных этого датасета актуален, так как способствует развитию музыкальной терапии и улучшению психологического состояния населения.

Исходя из этого, можно провести следующий анализ:

- Ознакомиться с датасетом, изучить все его поля и визуализировать;
- Предобработать данные: убрать Null-значения, отсеять выбросы, русифицировать данные для последующего анализа и т.д.;
- Визуализировать упомянутый датасет с помощью PowerBi;
- Провести комплексный анализ итогов и результатов работы.

2. Описание датасета

Датасет Music & Mental Health Data Visualization представляет собой набор метрических данных о крабах, а также информацию о поле и возрасте. В нем содержатся следующие поля:

1. Timestamp : Дата и время отправки формы
2. Age : возраст респондента
3. Primary streaming service : основной потоковый сервис респондента
4. Hours per day : Количество часов, которые респондент слушает музыку в день во время работы
5. While working : Слушает ли респондент музыку во время учебы/работы?
6. Instrumentalist: Регулярно ли респондент играет на каком-либо инструменте?
7. Composer : Сочиняет ли респондент музыку?
8. Fav genre : любимый или наиболее популярный жанр респондента
9. Exploratory : Активно ли респондент изучает новых исполнителей/жанры?
10. Foreign languages : Регулярно ли респондент слушает музыку с текстами на языке, которым он не владеет в совершенстве?
11. BPM : : Количество ударов в минуту в любимом жанре
12. Frequency [Classical] : Как часто респондент слушает классическую музыку
13. Frequency [Country] : Как часто респондент слушает музыку в стиле кантри
14. Frequency [EDM] : Как часто респондент слушает музыку в стиле EDM
15. Frequency [Folk] : Как часто респондент слушает народную музыку
16. Frequency [Gospel] : Как часто респондент слушает хоровую музыку
17. Frequency [Hip hop] : Как часто респондент слушает хип-хоп музыку
18. Frequency [Jazz] : Как часто респондент слушает джазовую музыку
19. Frequency [K pop] : Как часто респондент слушает Кей-поп музыку
20. Frequency [Latin] : Как часто респондент слушает латиноамериканскую музыку
21. Frequency [Lofi] : Как часто респондент слушает музыку lofi
22. Frequency [Metal] : Как часто респондент слушает металл

- 23.Frequency [Pop] : Как часто респондент слушает поп-музыку
- 24.Frequency [R&B] : Как часто респондент слушает R&B музыку
- 25.Frequency [Rap] : Как часто респондент слушает рэп-музыку
- 26.Frequency [Rock] : Как часто респондент слушает рок-музыку
- 27.Frequency [Video game music] : Как часто респондент слушает музыку для видеоигр
- 28.Anxiety : Самооценка тревожности по шкале от 0 до 10 баллов
- 29.Depression : Самооценка депрессии по шкале от 0 до 10 баллов
- 30.Insomnia: Самооценка бессонницы по шкале от 0 до 10 баллов
- 31.OCD : Самооценка ОКР по шкале от 0 до 10 баллов
- 32.Music effects : Улучшает/ухудшает ли музыка психическое здоровье респондента?
- 33.Permissions : Разрешения на публикацию данных

Количество записей – 736

Данные собраны за период с 07/27/2022 по 11/08/2022 по всему миру

Автор датасета – Catherine Rasgaitis, Computer Science @ University of Washington

3. Ход работы

3.1 Гипотеза

Исходя из описания датасета, была определена гипотеза – оказывает ли влияние определенные музыкальные предпочтения респондента на его психическое состояние? Если оказывает, то как именно?

В случае успешного подтверждения гипотезы определить жанры, улучшающие психологическое состояние людей при музыкальной терапии.

Первым этапом работы предобработка данных. На этом этапе были обнаружены и заменены Null-значения в записях, обнаружены аномалии (выбросы) и отсеяны, русифицированы значения некоторых атрибутов и т.д..

Затем на основе обновленных данных была построена модель визуализации для ее дальнейшего анализа при помощи PowerBi.

Финальным этапом является комплексный анализ и обобщение полученной информации в рекомендационную модель для дальнейшего применения в сфере здравоохранения и психологии.

3.2 Визуализация данных по музыкальным предпочтениям и ментальном состоянии.

Построив диаграмму рассеяния по датасету мы можем наблюдать следующие аномалии:

На рис.1 область значений (12; 25) по оси *HoursPerDay* и на рис.2 область значений (0: 50) по оси *BPM* являются аномальными, так как их значения не коррелируют с соответствующими значениями оси *Age* и имеют достаточно большое отклонение от среднего значения по данным. В следствие чего было принято решение об отсечении этих выбросов во избежание некачественного анализа.

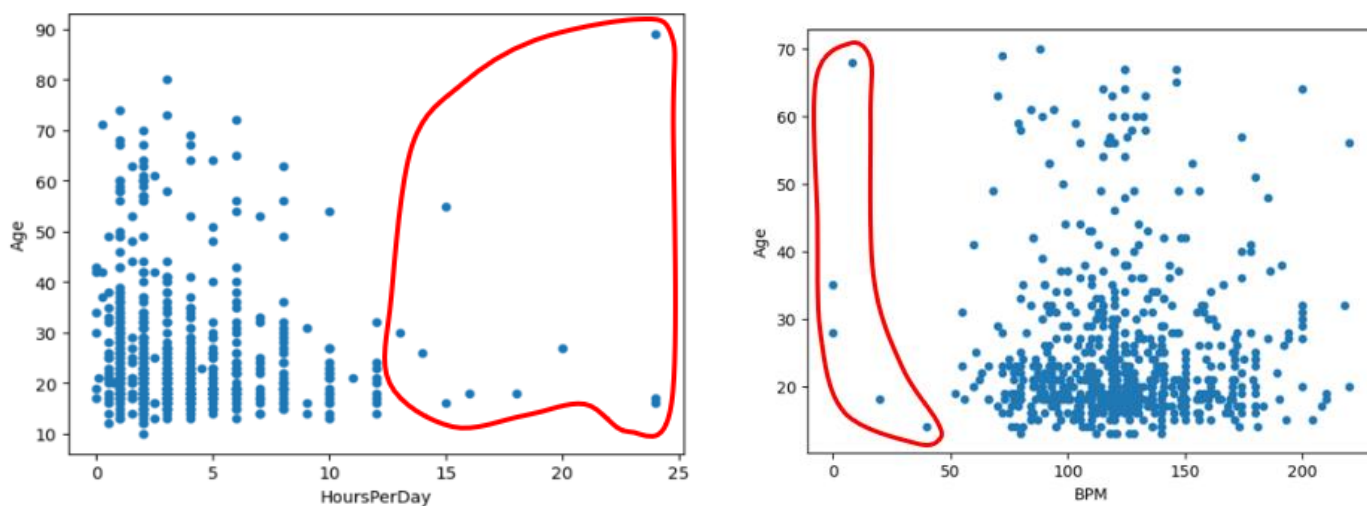


Рисунок 1-2. Визуализация датасета для дальнейшего анализа

3.3 Статистические данные

Анализируя представленные ниже диаграммы, мы выявили сделали следующие выводы:

- Депрессия: Рок может иметь зависимость от BPM, рекомендуемые жанры для людей с депрессией - Lofi и Hip-hop.
- Тревожность: Кантри или классическая музыка помогают справиться с тревожностью, респонденты, часто играющие в видеоигры, могут испытывать повышенную тревожность.

- ОКР: Госпел сильно положительно влияет на ОКР, исследований для других жанров недостаточно для вывода.
- Бессонница: Не рекомендуется музыка из видеоигр, предпочтения и образ жизни, основанный на видеоиграх, могут влиять на качество сна. Рекомендованные жанры для прослушивания - кантри, реп и R&B.

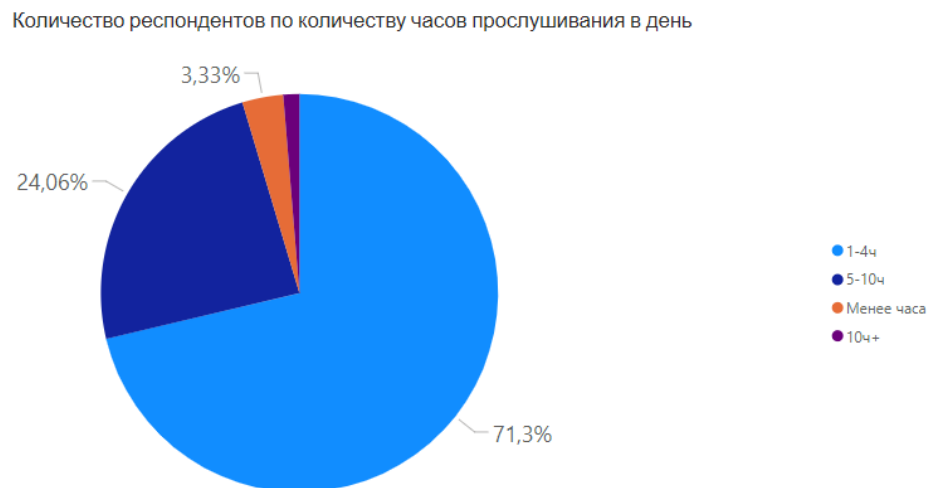


Рисунок 3. Количество респондентов по количеству часов прослушивания музыки в день

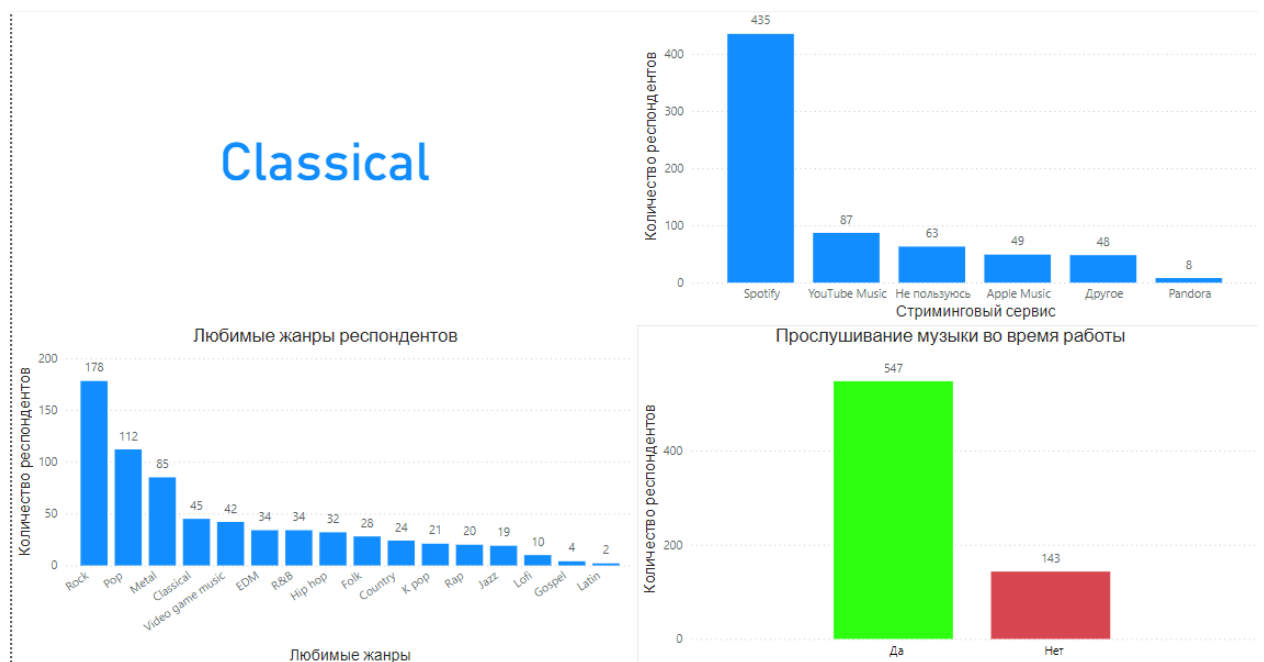


Рисунок 4. Общая информация о респондентах

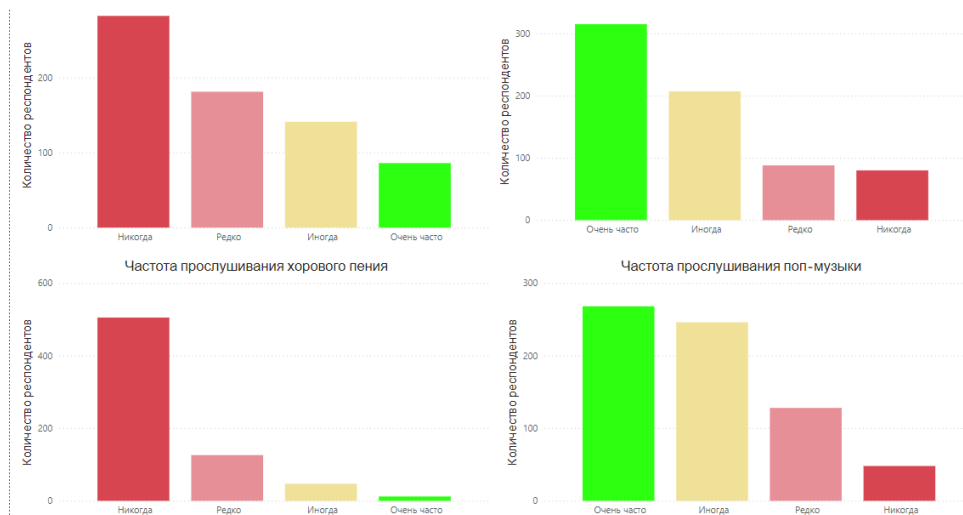


Рисунок 5. Частота прослушивания определенных жанров

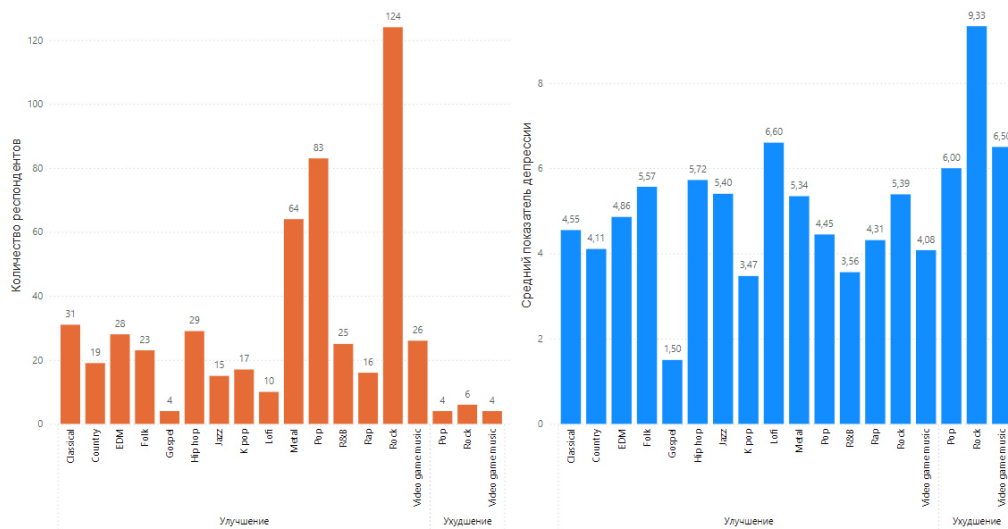


Рисунок 6. Анализ депрессии респондентов

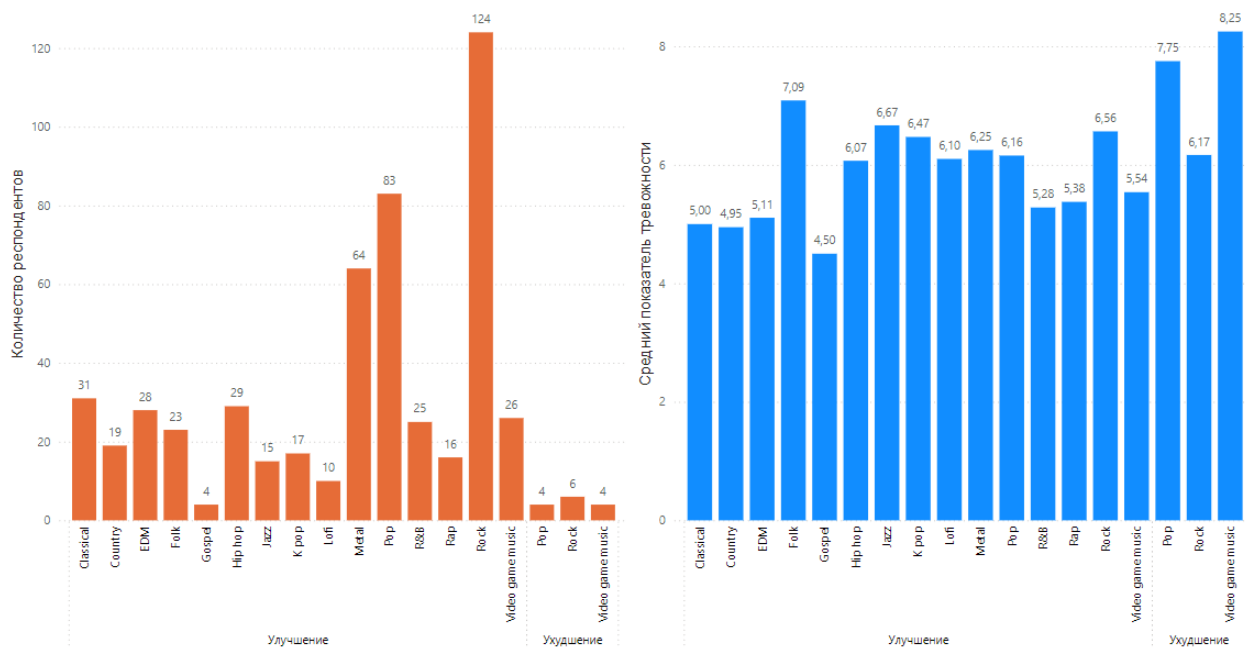


Рисунок 7. Анализ тревожности респондентов

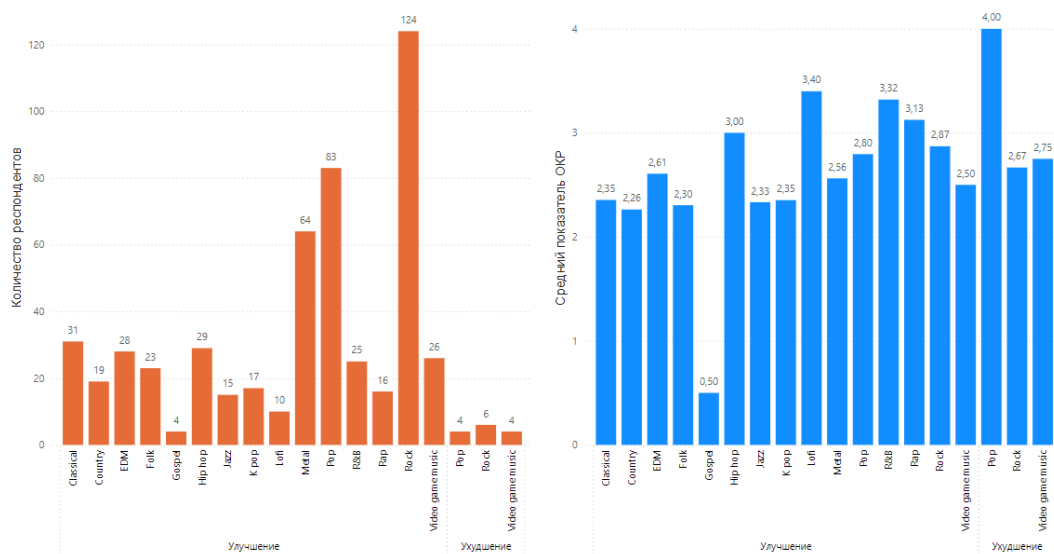


Рисунок 8. Анализ ОКР респондентов

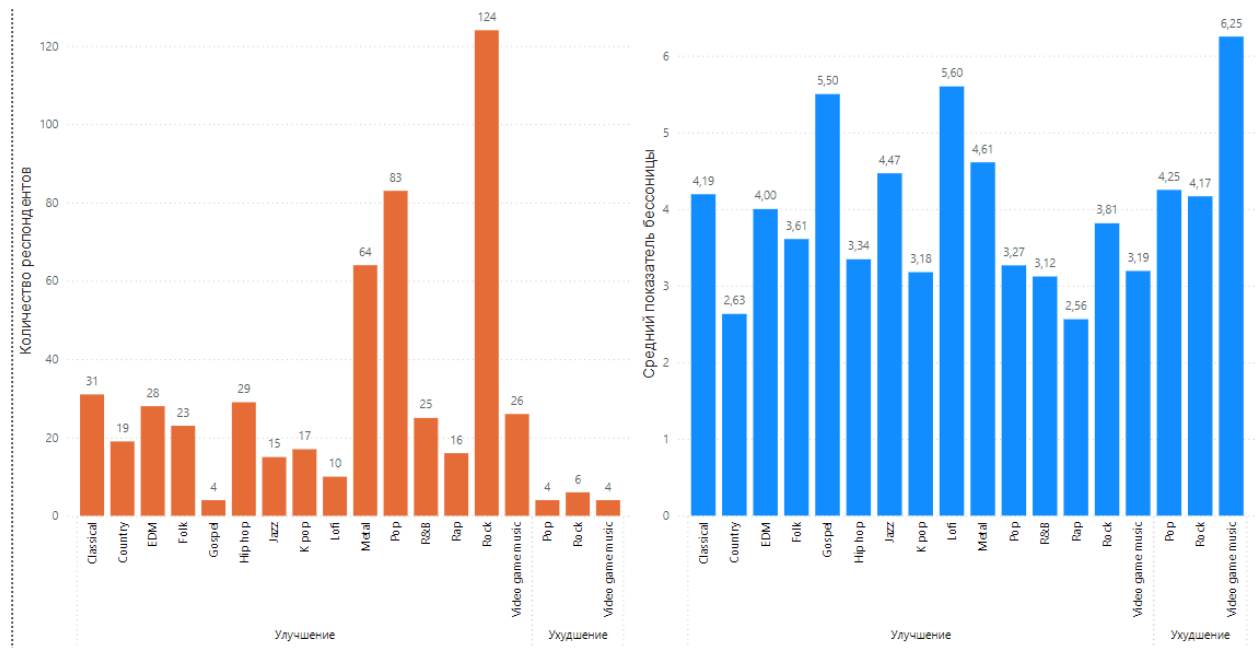


Рисунок 7. Анализ бессонницы респондентов



Classical

Рисунок 7. Анализ по BPM

3.4 Результаты проделанной работы

Гипотеза о взаимосвязи между музыкальными предпочтениями человека и его ментальным состоянием на основе нашего датасета, посредством описательного анализа, была опровергнута, так как взаимосвязь неявная и зависит от многих человеческих факторов, по сравнению с которыми музыкальные предпочтения являются не решающими.

4. Выводы

В ходе решения задачи, были выполнены следующие этапы:

- Предобработан и проанализирован датасет Music & Mental Health Data Visualization;
- Выявлена слабая взаимосвязь между музыкальными предпочтениями респондентов и их психологическим состоянием;
- Визуализированы результаты работы.

5. Список литературы

1. Music & Mental Health Data Visualization [Электронный ресурс] – URL:
<https://www.kaggle.com/code/shubhammeshram579/music-mental-health-data-visualization/notebook>
2. Пример работы других пользователей с датасетом Music & Mental Health Data Visualization [Электронный ресурс] – URL:
<https://www.kaggle.com/code/catherinerasgaitis/music-mental-health-eda#Conclusions>