Министерство образования и науки Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего образования

«Уральский федеральный университет имени первого Президента России Б.Н. Ельцина» (УрФУ)

Институт радиоэлектроники и информационных технологий – РтФ

Школа бакалавриата

Оценка

Дата защиты

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА к проекту по модулю «Методы анализа Big Data» по теме: ЗАВИСИМОСТЬ ПОПУЛЯРНОСТИ МУЗЫКАЛЬНОГО АЛЬБОМА ОТ ДАТЫ ЕГО ВЫПУСКА

Подпись Дата

Преподаватель С.Г. Мирвода

Студенты И.А. Зарудный

Е.И. Манин

Группа РИ-450005

**Содержание**

[Введение 3](#_Toc535169449)

[1 Описание набора данных 4](#_Toc535169450)

[2 Обработка данных 5](#_Toc535169451)

[3 Фильтрация набора данных 6](#_Toc535169452)

[4 Построение моделей 7](#_Toc535169453)

[5 Разработка приложения 8](#_Toc535169454)

[Заключение 9](#_Toc535169455)

[Библиографический список 10](#_Toc535169456)

# Введение

С развитием современных технологий появились новые возможности для анализа огромного количества данных. Огромные объёмы данных обрабатываются для того, чтобы человек мог получить конкретные и нужные ему результаты для их дальнейшего эффективного применения.

В ходе работы над нашим проектом предметом для исследования была выбрана сфера музыкальной индустрии. Наверняка, многие люди замечали, что в один и тот же день может выйти сразу несколько музыкальных альбомов различных исполнителей, жанры которых могут быть диаметрально противоположными. Данная ситуация заинтересовала нас и было принято решение проверить, существует ли зависимость между популярностью альбома, количеством его продаж и непосредственно датой его выпуска.

Поставленная цель достигается при выполнении следующих задач:

1. Выбор и обработка подходящего набора данных.
2. Анализ полученных данных
3. Фильтрация набора данных.
4. Построение моделей данных.
5. Разработка приложения для работы с созданной моделью.

# Описание набора данных

Музыкальная индустрия— часть экономики и общественной жизни, объединяющая компании и граждан, зарабатывающих деньги посредством создания и продажи музыкальных произведений. Музыкальная индустрия является составной частью более обширного сектора экономики — индустрии развлечений [1].

В музыкальном бизнесе задействовано большое число людей и организаций: музыканты, создающие и исполняющие музыкальные произведения; компании и специалисты, занимающиеся звукозаписью и продажей музык, в том числе продюсеры, студии звукозаписи, звукорежиссёры, лейблы, музыкальные магазины и коллективные организации по управлению правами; организаторы гастролей; теле- и радиосети, работающие в музыкальном формате; музыкальные журналисты и критики; производители музыкальных инструментов и многие другие.

В качестве своеобразного показателя успешности релиза можно расценивать статус, присвоенный после сертификации альбома. Одной из наиболее авторитетных признана сертификация RIAA (Recording Industry Association of America).

Сертификация RIAA — процедура сертификации объёма продаж звукозаписей на территории США, на соответствие определённому статусу. Осуществляется Американской ассоциацией звукозаписывающих компаний. Появилась в 1958 как попытка стандартизировать методики измерения объёма продаж звукозаписей. Сначала существовал только статус «Золотая награда», присуждаемый за продажу более 500 000 экземпляров записи. В 1976 появился платиновый статус (продажа более 1 000 000 экземпляров). В 1984 появился мультиплатиновый статус (более 2 000 000 экземпляров) в 1999 бриллиантовый (более 10 000 000 экземпляров).

Для нашего исследования набор данных был взят с сайта https://www.riaa.com/, нас интересовали только произведения, получившие платиновый статус. Поскольку сайт не объединяет результаты в определенные наборы данных, был написан скрипт на языке Python, позволяющий с опорой на соответствующее изображение, отличительное для платинового статуса, получить необходимую информацию для релиза. Полученная информация объединялась в Dataset формата CSV.

# Обработка данных

Для нашего исследования была взята информация о релизах с платиновым статусом с 01.01.2010 по 01.01.2019. Для наглядного представления количества таких релизов и распределения их по месяцам была составлена тепловая карта. Наиболее важными данными из всего набора данных для нас являются колонки Certified.Units, в которой отображено количество проданных копий, а также колонки Release date и Genre, в которых отражены дата выхода произведения и его жанр. Для примера, также, был построен график зависимости даты выхода от количества проданных произведений в жанре POP(Попса).

# Построение моделей

## Построение первичной модели с выбросами

Для построения моделей была выбрана модель ARIMA (англ. autoregressive integrated moving average, модель Бокса — Дженкинса) — интегрированная модель авторегрессии — скользящего среднего — модель и методология анализа временных рядов. Является расширением моделей ARMA для нестационарных временных рядов, которые можно сделать стационарными взятием разностей некоторого порядка от исходного временного ряда (так называемые интегрированные или разностно-стационарные временные ряды).

## Проецирование продаж на логарифмическую шкалу

## Фильтрация набора данных и построение модели

## Проецирование продаж на тангенс

# Разработка приложения

# Заключение

# Библиографический список

1. Дейт К. Введение в системы баз данных / К. Дейт, М.Hаука, 2005 г. 246 с.
2. Карпова Т.С. Базы данных: модели, разработка, реализация / Т.С. Карпова, СПб.: Питер, 2006 г. 304 с.
3. Create First OLAP Cube in SQL Server Analysis Services, Mubin M. Shaik. [Электронный ресурс]. Доступен по ссылке – https://www.codeproject.com/Articles/658912/Create-First-OLAP-Cube-in-SQL-Server-Analysis-Ser Последняя дата обращения: 20.11.18.
4. Введение в многомерный анализ [Электронный ресурс]. Доступен по ссылке – https://habr.com/post/126810/ Последняя дата обращения: 25.11.18.
5. Кречетов Н. Продукты для интеллектуального анализа данных / Н. Кречетов, ComputerWeek-Москва. 2003 г. № 14-15. 32-39 с.
6. Вендров А.М. Практикум по проектированию программного обеспечения ЭИС: Учебное пособие / А.М. Вендров, Финансы и статистика, 2014. 215с.