



Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Московский государственный технический университет  
имени Н.Э. Баумана»  
(национальный исследовательский университет)  
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ \_\_\_\_\_ ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ \_\_\_\_\_

КАФЕДРА \_\_\_\_\_ СИСТЕМЫ ОБРАБОТКИ ИНФОРМАЦИИ И УПРАВЛЕНИЯ (ИУ5) \_\_\_\_\_

## РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

### к домашнему заданию

по дисциплине: Оперативный анализ данных/Средства визуализации данных

на тему: «Korean drama list».

Студент ИУ5-54Б  
(Группа)

\_\_\_\_\_  
(Подпись, дата) Свечникова А.А.  
(И.О.Фамилия)

Руководитель

\_\_\_\_\_  
(Подпись, дата) Маслеников К.Ю.  
(И.О.Фамилия)

2022 г.

Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Московский государственный технический университет имени Н.Э. Баумана  
(национальный исследовательский университет)»  
(МГТУ им. Н.Э. Баумана)

---

УТВЕРЖДАЮ

Руководитель секции ИУ5  
(Индекс)

« \_\_\_\_\_ » \_\_\_\_\_ 20 \_\_\_\_ г.  
(И.О.Фамилия)

**З А Д А Н И Е**  
**на выполнение домашнего задания**

по дисциплине Оперативный анализ данных/Средства визуализации данных  
по теме \_\_\_\_\_

Студент группы ИУ5-54Б

Свечникова Анна Андреевна

(Фамилия, имя, отчество)

График выполнения домашнего задания: 25% к 3 нед., 50% к 6 нед., 75% к 9 нед., 100% к 15 нед.

**Техническое задание:** Разработать интерактивную систему оперативного анализа данных для информационной системы «Korean drama list» по выбранной предметной области, произвести анализ бизнес-процессов, произвести визуализацию данных, построить хранилище данных, составить и обосновать план архивации и восстановления хранилища данных

**Оформление научно-исследовательской работы:**

Расчетно-пояснительная записка на 18 листах формата А4.

Перечень графического (иллюстративного) материала (чертежи, плакаты, слайды и т.п.)

Дата выдачи задания « 2 » сентября 2022 г.

Руководитель домашнего задания

Маслеников К.Ю.  
(Подпись, дата) (И.О.Фамилия)

Студент

Свечникова А.А.  
(Подпись, дата) (И.О.Фамилия)

## Содержание

ВВЕДЕНИЕ.....	4
1 Описание предметной области.....	5
2 Описание бизнес-процессов в нотации BPMN 2.0.....	6
3 Хранилище данных.....	8
4 Визуализация данных.....	9
ЗАКЛЮЧЕНИЕ.....	14
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	18

## **ВВЕДЕНИЕ**

В работе производится анализ данных о телесериалах для выявления ранее неизвестных фактов и закономерностей.

Цели работы:

- 1) Углубление и расширение теоретических знаний в сфере анализа и визуализации набора данных.
- 2) Получение навыков разработки интерактивных систем оперативного анализа данных для некоторой информационной системы.
- 3) Освоение описания, построения и анализа бизнес-процессов в нотации BPMN 2.0.
- 4) Приобретение опыта проектирования хранилища данных.
- 5) Практическое освоение визуализации данных.
- 6) Поиск способов архивации и восстановления хранилища данных.

## **1 Описание предметной области**

В качестве предметной области выбраны сериалы, снятые телекомпаниями Южной Кореи.

В наборе данных содержатся:

- 1) Название сериала.
- 2) Число эпизодов.
- 3) Начало показа по телевидению.
- 4) Окончание показа по телевидению.
- 5) Продолжительность одного эпизода.
- 6) Оценка по данным портала Mydramalist.
- 7) Число зрителей, поставивших оценки на портале.
- 8) Телекомпания.
- 9) Исполнитель(-ница) главной роли.

Этот набор данных можно проанализировать и сделать следующие выводы:

- 1) Какая телекомпания завершила больше всего проектов.
- 2) Какие актеры имеют наиболее обширную фильмографию.
- 3) Какова динамика зрительских оценок за последние несколько лет.
- 4) Какие жанры пользуются наибольшей популярностью у зрителей и у телекомпаний.

## 2 Описание бизнес-процессов в нотации BPMN 2.0

### 2.1 Бизнес-процесс «Подбор актёра на роль»

На рисунке 2.1 представлен процесс подбора актёра на роль в телесериале. Режиссёр составляет описание желаемого типажа и отправляет его кастинг-директору. Кастинг-директор составляет условия контракта: описание проекта, график и места съёмок и гонорар, и отправляет их в актерское агентство вместе с описанием типажа. Получив от актерского агентства портфолио кандидатов, он направляет его режиссёру. Режиссер может и сам направить приглашение для желаемого кандидата в агентство. Просмотрев портфолио, режиссёр утверждает список участников проб и направляет его кастинг-директору, который организует пробы: находит помещение, необходимую аппаратуру, приглашает осветителей и операторов. Информацию о пробах кастинг-директор направляет режиссёру и в агентство. Если после проведения проб режиссёр выбрал исполнителя роли, он сообщает кастинг-директору о принятом решении и тот заключает контракт с актёром через агентство. Иначе режиссер сообщает кастинг-директору о необходимости провести повторные пробы и процесс повторяется заново начиная с утверждения списка кандидатов.

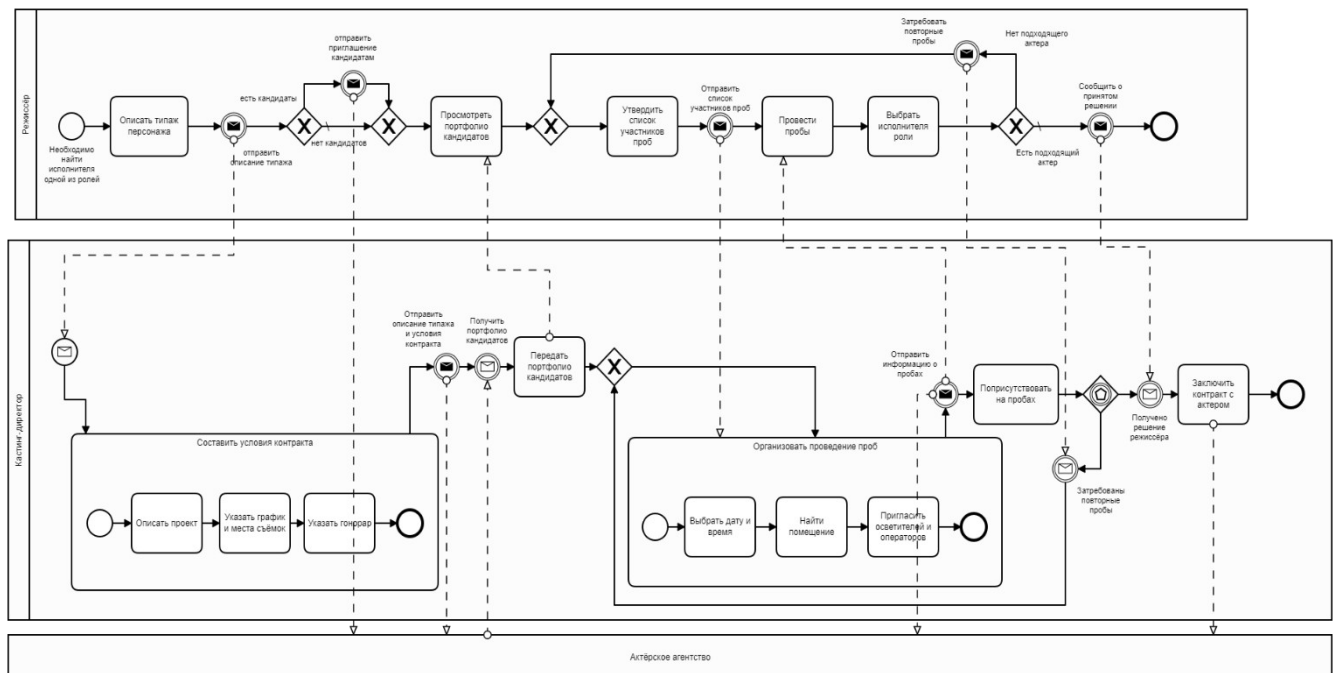


Рисунок 2.1 – Процесс подбора актёра в нотации BPMN 2.0

## 2.2 Бизнес-процесс «Поиск сервиса для показа сериала»

На рисунке показан процесс поиска стримингового сервиса для показа нового сериала. Компания предлагает сервису свой новый проект. Получив предложение, сервис оценивает его с помощью команды юристов. Сервис может принять предложение, отклонить его или выдвинуть другие условия. О своем решении он сообщает телекомпании. Если получен отказ, или выдвинутые условия не подходят, телекомпания подтверждает отказ от сделки и процесс завершается. Если получено согласие, или выдвинутые условия подходят, то телекомпания отправляет согласие и стриминговый сервис с помощью юристов оформляет сделку. Телекомпания также оформляет сделку с помощью юристов и предоставляет сервису необходимые видеоматериалы в обмен на оплату. Затем телекомпания распределяет выручку. Стриминговый сервис запускает пиар-компанию сериала: создаёт его страницу в своем каталоге, делает тематические публикации в соцсетях и закупает рекламу, после чего начинает показ сериала.

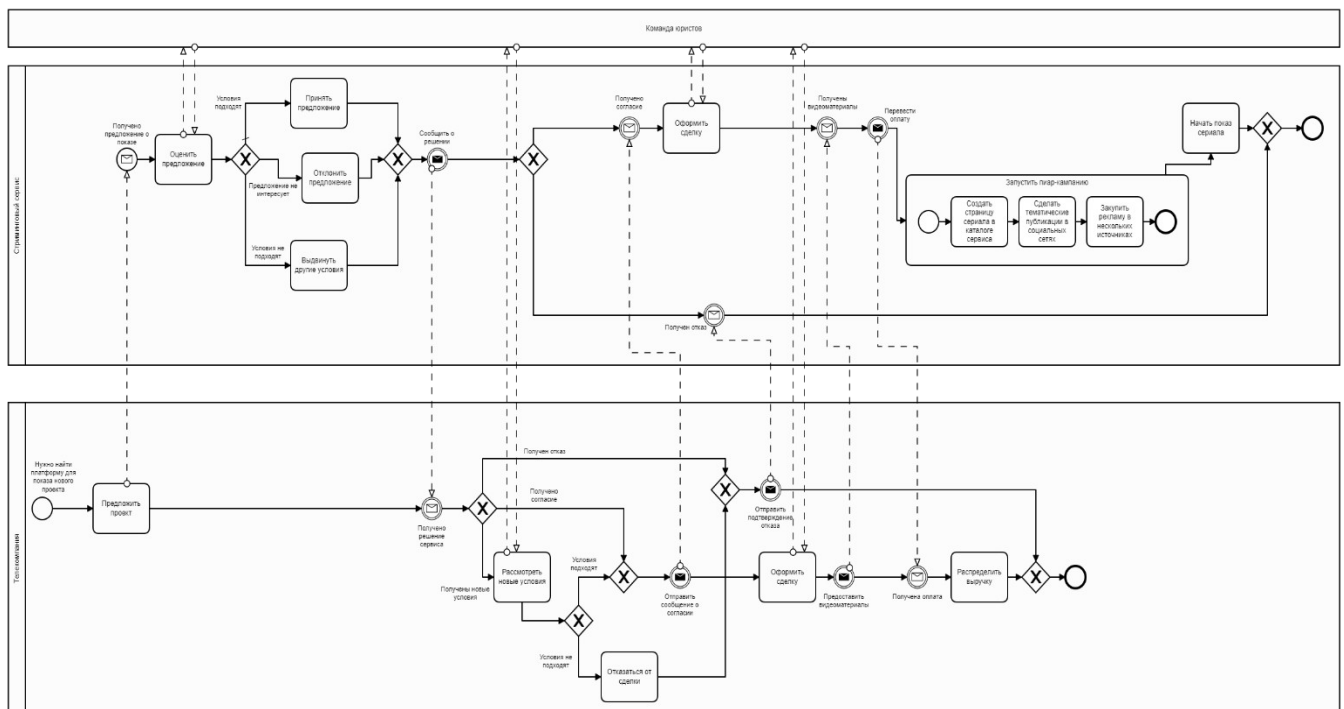


Рисунок 2.2 – Поиск сервиса для показа сериала в нотации BPMN 2.0

### 3 Хранилище данных

На рисунке 3 представлена схема хранилища данных. Таблица «Сериал» представляет собой таблицу фактов, а таблицы «Жанр», «Телекомпания», «Актер» и «Трансляция» - таблицы измерений. Также в хранилище есть таблица времени.

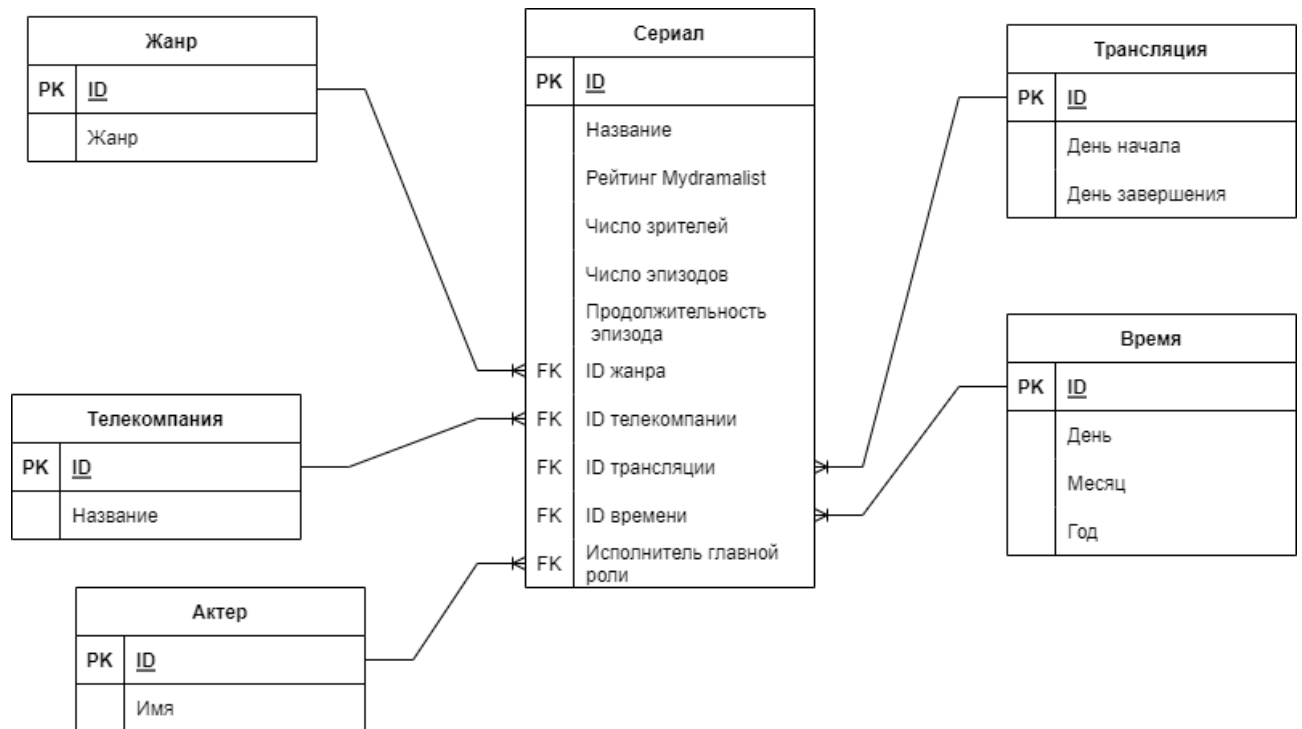


Рисунок 3 – Схема хранилища данных



## 4 Визуализация данных

### 4.1 Число сериалов у телекомпаний

На этой диаграмме видно, что первые три крупнейшие телекомпании имеют существенный отрыв от остальных. Наибольшее число сериалов отсняты компанией MBC.

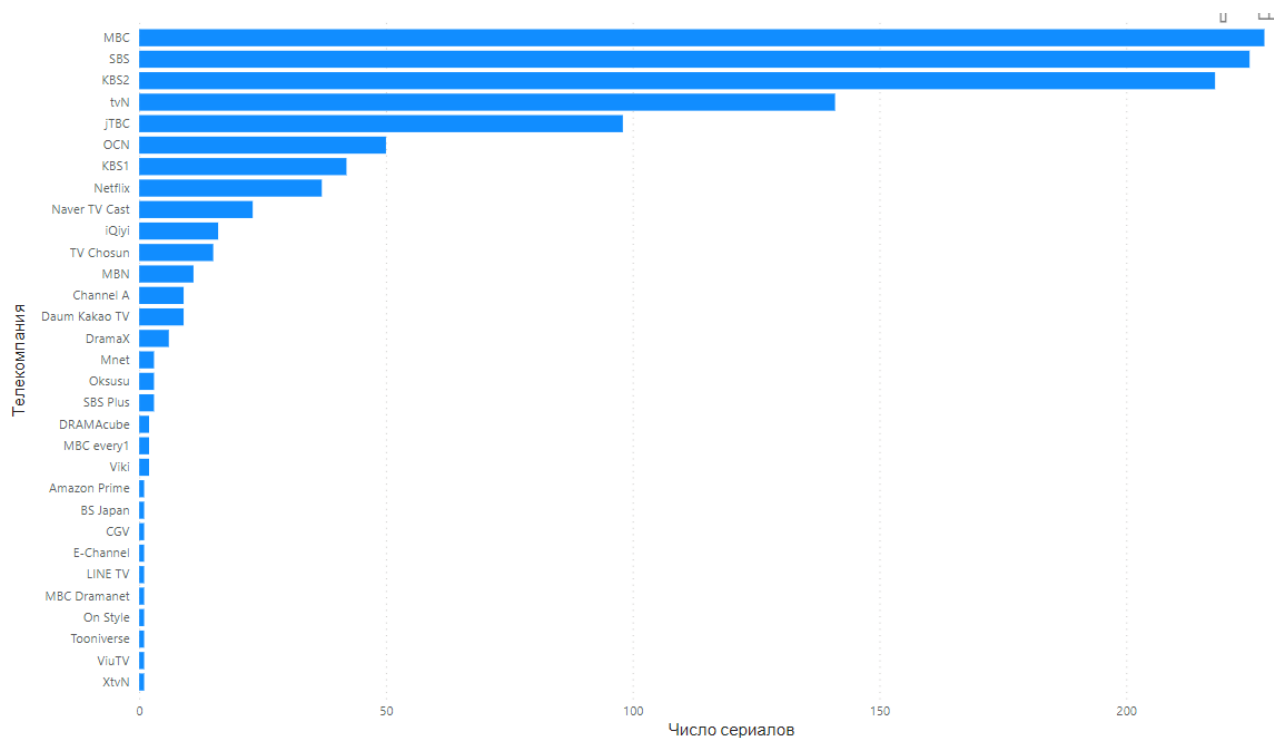


Рисунок 4.1 – Число сериалов у телекомпаний

### 4.2 Число сериалов по годам

Данная диаграмма демонстрирует, что набор данных содержит данные преимущественно о сериалах, выпущенных в период с 2010 года по настоящее время.

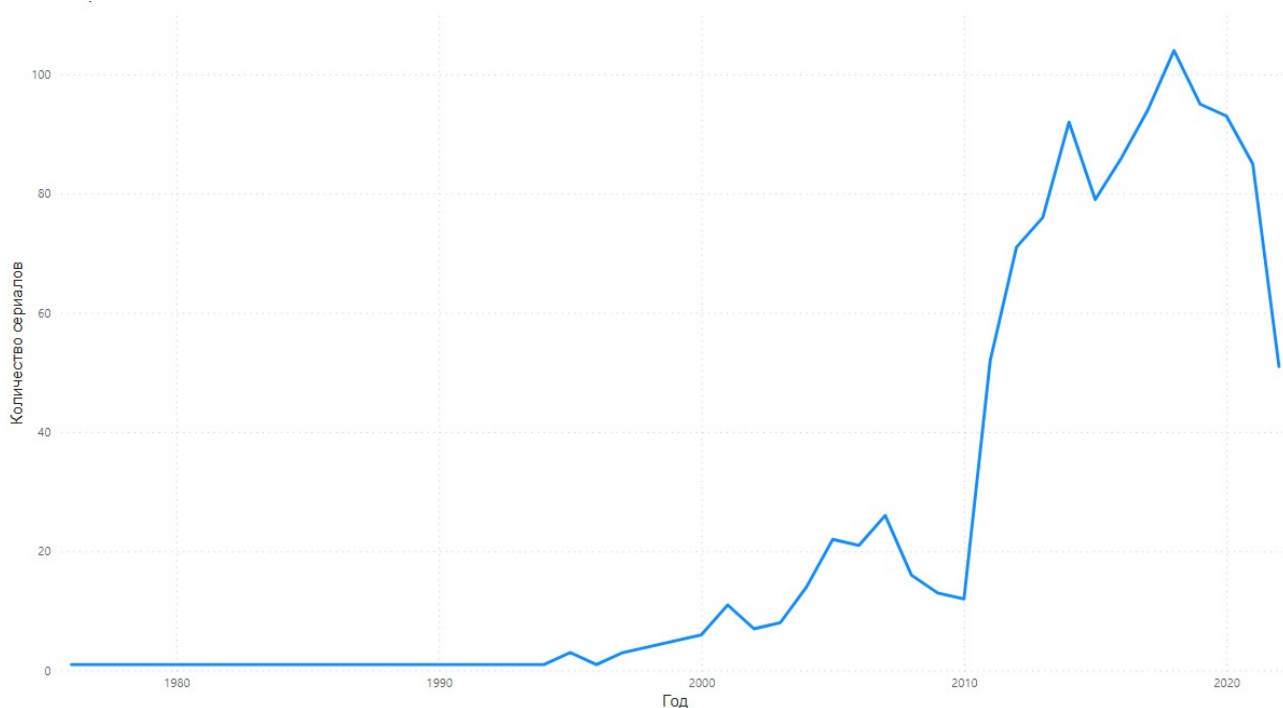


Рисунок 4.2 – Число сериалов по годам

#### 4.3 Рейтинг сериалов: средний, минимальный, максимальный

Датчик показывает, что минимальный рейтинг в наборе данных составляет 4 балла, а максимальный – 10 баллов. Средний рейтинг составляет 7,7 балла.

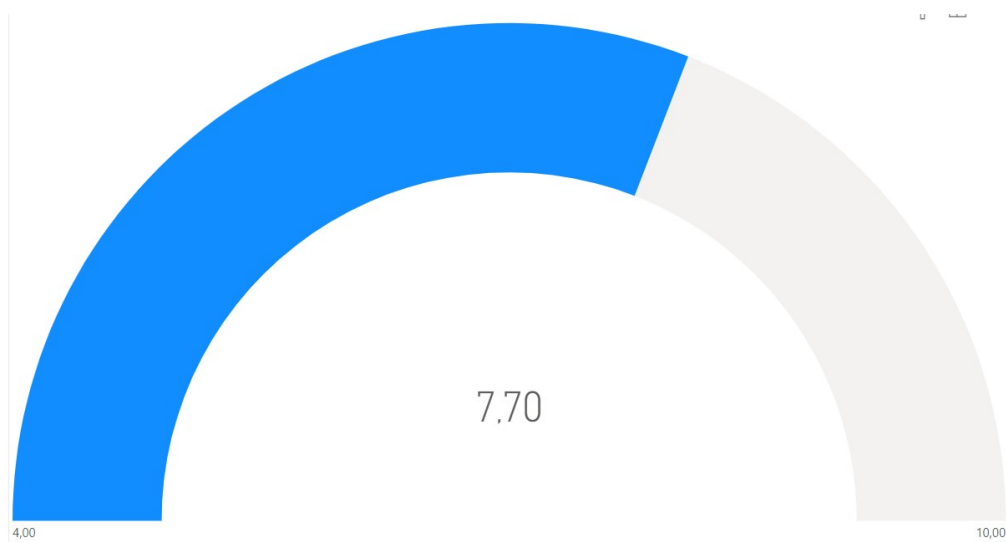


Рисунок 4.3 – Рейтинг сериалов: средний, минимальный, максимальный

#### 4.4 Средний рейтинг по годам

На данной диаграмме представлен средний рейтинг сериалов, выпущенных с 2000 года по настоящее время. Красной пунктирной линией отмечен средний

рейтинг, найденный в предыдущем пункте. Можно заметить, что в последнее время пользователи дают более высокие оценки.



Рисунок 4.4 – Средний рейтинг по годам

#### 4.5 Число сериалов различных жанров

Круговая диаграмма наглядно отражает распределение сериалов по жанрам. Наиболее часто встречаются комедия и романтика.

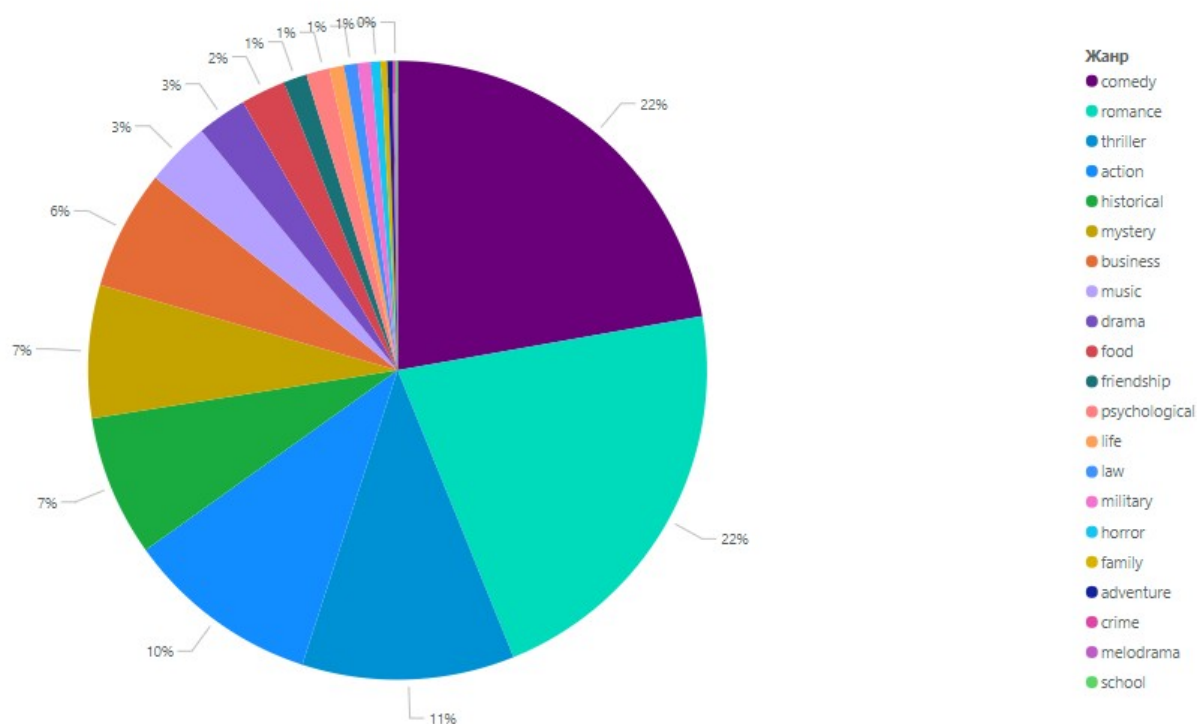


Рисунок 4.5 – Число сериалов различных жанров.

#### 4.6 Число зрителей для различных жанров

Данная диаграмма отражает число зрителей сериалов каждого жанра. Видно, что спрос здесь совпадает с предложением лишь отчасти – жанр комедия, к которому относится большая доля сериалов, имеет и наибольшее число зрителей, но второй по популярности жанр – экшен – занимает только четвертое место в предыдущем пункте.



Рисунок 4.6 – Число зрителей для различных жанров

#### 4.7 Средний рейтинг различных жанров

Средний рейтинг жанров распределен достаточно равномерно. На удивление, сериалы наиболее популярного жанра – комедийные – имеют не самые высокие рейтинги. Возможно, это вызвано высокой конкуренцией в этом жанре. Наиболее высокий рейтинг имеют сериалы в жанре повседневность.

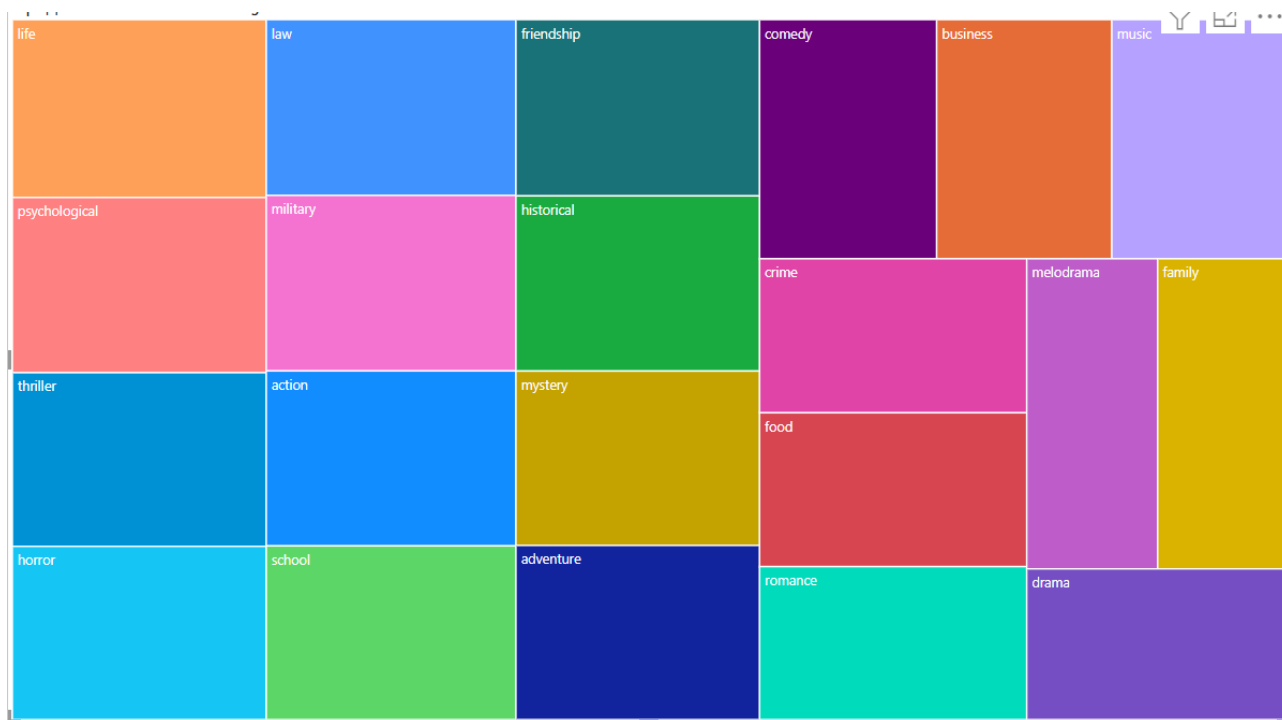


Рисунок 4.7 – Число зрителей для различных жанров

#### 4.8 Востребованность актеров

На данной диаграмме отражены пять актеров, снявшихся в наибольшем числе сериалов. Кроме того, на диаграмме отражено число различных жанров, в которых работал актер. Лидером по обоим пунктам является Jang Hyuk (Чан Хёк).

Число жанров везде меньше числа сериалов, так как у актера обычно формируется устойчивое амплуа, в котором он снимается.

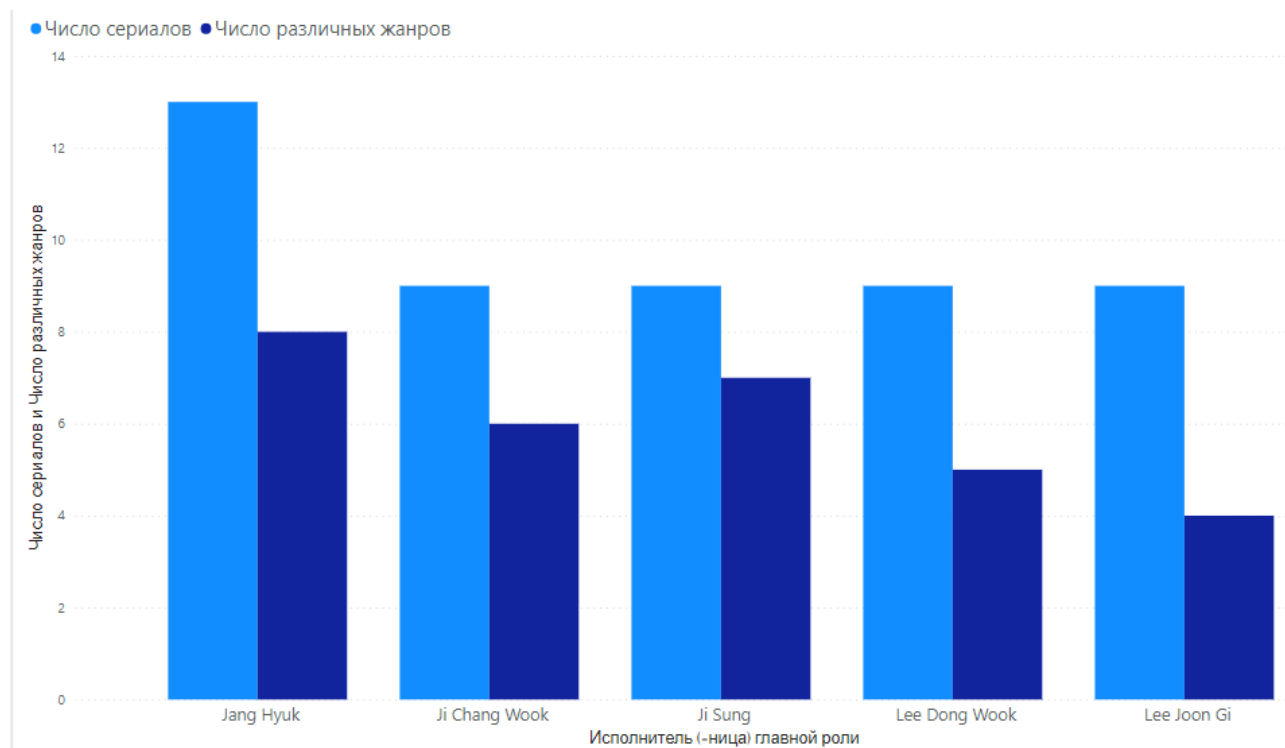


Рисунок 4.8 – Востребованность актеров

## 5 План архивации и восстановления данных

Архивация данных обеспечивает возможность быстро восстановить их, если возникнет такая необходимость. Для создания плана архивации и восстановления нужно учесть, как часто изменяются данные, какие данные являются критически важными. В данной системе критически важными являются следующие данные: названия сериалов, число серий, рейтинги, число оценивших зрителей и жанры.

Для хранения архивированных данных можно использовать SSD-диски, так как они энергонезависимы и обеспечивают быстрый доступ к данным. Для архивации будет использоваться метод инкрементального бэкапа: раз в неделю, в понедельник в 8:00 будет производиться полная архивация, а затем в течение недели архивироваться будут только те данные, которые изменились с момента последней архивации. Для восстановления системы в случае сбоя потребуется восстановить сначала полный бэкап, а затем следующую за ним цепочку изменений. Этот метод восстановления довольно длительный, но зато такой тип архивации снижает сетевую нагрузку и уменьшает время, затраченное на архивацию.

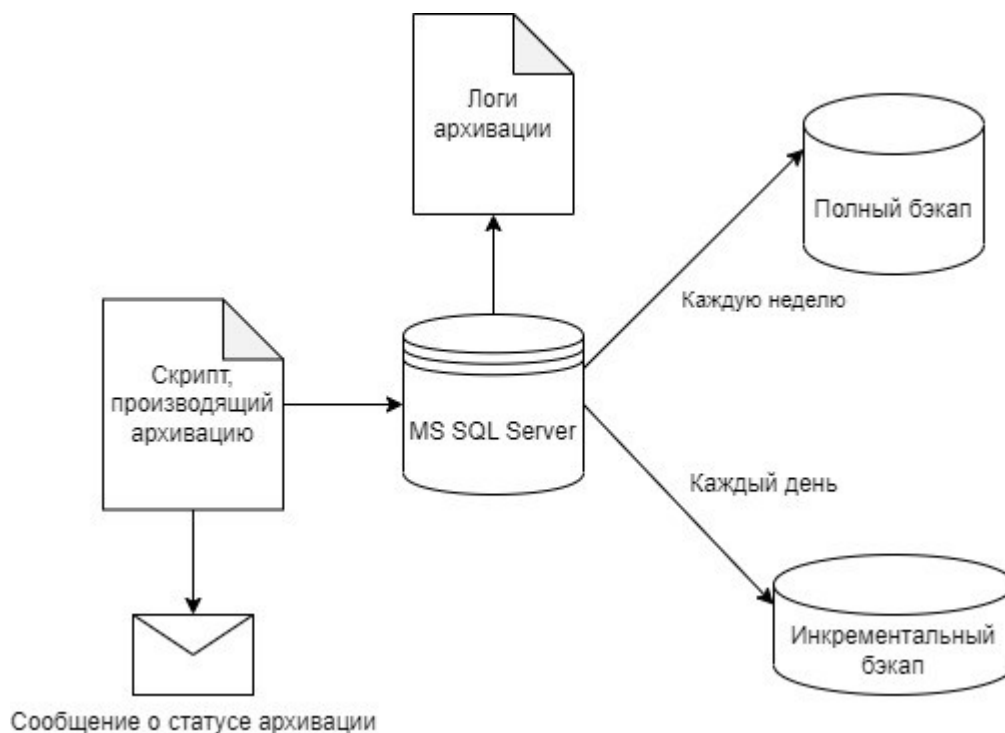


Рисунок 5 – Схема архивации данных

Необходимо рассчитать объём хранилища: 2 поля типа INT, 2 поля типа DATE, 2 поля типа FLOAT, 2 поля типа VARCHAR по 50 символов и три поля типа VARCHAR по 20 символов, число строк 1153.

$$(14 \times 4 + 2 \times 2 + 2 \times 8 + 2 \times 50 \times 1 + 3 \times 20 \times 1) \times 1153 = 266 \text{ Кбайт.}$$

Предположим, что объем изменившихся за день данных составляет 10%. Тогда размер инкремента составит 26,6 Кбайт в день.

Храним 2 полные копии и 7 инкрементов. Следовательно, общий объем хранилища равен  $266 + 266 + 26,6 \times 7 = 718,2$  Кбайта.



## **ЗАКЛЮЧЕНИЕ**

В результате выполнения домашнего задания была спроектирована интерактивная система оперативного анализа данных для информационной системы «Korean drama list». Были построены описания бизнес-процессов в нотации BPMN 2.0 и спроектировано хранилище данных, а также произведена визуализация основных данных с использованием программы Microsoft Power BI. Кроме того, была разработана стратегия архивации и восстановления данных. В результате выполнения работы были получены и закреплены на практике знания в области анализа и визуализации данных.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Анализ данных и процессов: учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. - 3-е изд., перераб. и доп. - СПб.: БХВ-Петербург, 2009. - 512 с.
2. Статья Backup Types Explained: Full, Incremental, and Differential: [Электронный ресурс]. // URL: Types of Backup: Full, Incremental, Differential, and More (nakivo.com) (Дата обращения: 22.12.2022)
3. Фёдоров И. Г. Моделирование бизнес-процессов в нотации BPMN2.0: Монография, Москва 2013 г. МЭСИ. – 255 стр.
4. Официальная документация Microsoft PowerBI: [Электронный ресурс]. // URL: <https://learn.microsoft.com/ru-ru/power-bi/> (Дата обращения 22.12.2022)
5. Статья Архивация и восстановление данных от Microsoft: [Электронный ресурс]. // URL: <https://learn.microsoft.com/ru-ru/security-updates/security/20212685> (Дата обращения: 22.12.2022)