|  |  |
| --- | --- |
| Gerb-BMSTU_01 | **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\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

на тему: \_\_\_\_\_\_«Рейтинг лучших университетов с 2017 по 2022 год»\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Студент \_\_ИУ5-53Б\_\_\_ **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_**Пономарева М.О.

(Группа) (Подпись, дата) (И.О.Фамилия)

Руководитель \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_Маслеников К. Ю.\_

(Подпись, дата) (И.О.Фамилия)

*Москва, 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)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_

(И.О.Фамилия)

« \_\_\_\_ » \_\_\_\_\_\_\_\_\_\_\_\_ 2022 г.

**ЗАДАНИЕ**

**на выполнение домашнего задания**

по дисциплине \_\_\_\_\_\_\_\_\_\_\_\_\_ Оперативный анализ данных \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

по теме \_\_\_\_\_\_\_ «Рейтинг лучших университетов с 2017 по 2022 год» \_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Студент группы \_\_\_\_\_\_ ИУ5-53Б \_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Пономарева Марина Олеговна\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Фамилия, имя, отчество)

График выполнения домашнего задания: 25% к 3 нед., 50% к 6 нед., 75% к 9 нед., 100% к 12 нед.

Техническое задание \_Разработать интерактивную систему оперативного анализа данных для некоторой информационной системы по выбранной предметной области «Рейтинг лучших университетов с 2017 по 2022 год», 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 \_\_16\_\_ 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** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_ Пономарева М.О

(Подпись, дата) (И.О.Фамилия)

**СОДЕРЖАНИЕ**

[ВВЕДЕНИЕ 4](#_Toc120953609)

[1 Описание предметной области 5](#_Toc120953610)

[2 Описание бизнес-процессов в нотации BPMN 2.0 6](#_Toc120953611)

[2.1 Бизнес-процесс «Материальные выплаты студентам» 6](#_Toc120953612)

[2.2 Бизнес-процесс «Работа приемной комиссии» 7](#_Toc120953613)

[3 Хранилище данных 8](#_Toc120953614)

[4 Визуализация данных 9](#_Toc120953615)

[4.1Карта количества высших уч. заведений попавших в рейтинг 9](#_Toc120953616)

[4.2.Зависимость рейтинга от размеров и типа университета 10](#_Toc120953617)

[4.3.Изменение рейтинга с 2017 по 2022 год 11](#_Toc120953618)

[4.4. Зависимость баллов рейтинга от соотношения количества студентов и преподавателей. 12](#_Toc120953619)

[5. План архивации и восстановления данных 13](#_Toc120953620)

[ЗАКЛЮЧЕНИЕ 15](#_Toc120953621)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 16](#_Toc120953622)

## **ВВЕДЕНИЕ**

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.

## **Описание предметной области**

В качестве предметной области была выбрана тема – рейтинг лучших мировых университетов в 2017 – 2022 годах.

В наборе данных содержатся:

1. Страна, в которой находится университетов.
2. Количество баллов в рейтинге.
3. Год составления рейтинга.
4. Тип университета.
5. Размер университета.
6. Количество студентов на одного преподавателя.
7. Количество факультетов.
8. Количество иностранных студентов.
9. Место в рейтинге.
10. Регион.

Такую сводку можно проанализировать, сделать по ней выводы и использовать в следующих целях:

1. Выявление изменений в уровне университетов за несколько лет.
2. Выявление стран с наибольшим количеством университетов, попавших в рейтинг.
3. Выявление зависимости рейтинга от размера и типа университета.
4. Выявление зависимости рейтинга от соотношения числа студентов и преподавателей.

## **Описание бизнес-процессов в нотации BPMN 2.0**

**2.1 Бизнес-процесс «Материальные выплаты студентам»**

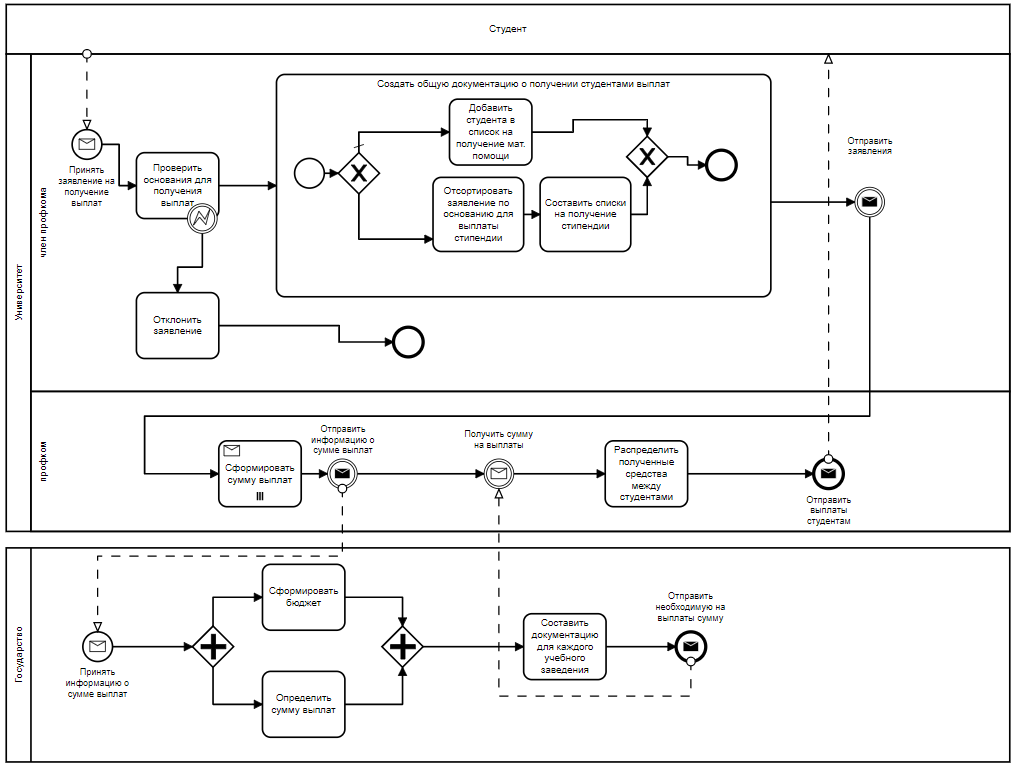
Процесс «Материальные выплаты студентам» представлен на рисунке 1. После того как член профкома получает от студента заявление на получение материальных выплат он проверяет основания для получения выплат. В случае если оснований нет заявление отклоняется, в ином случае на основе данного и всех остальных заявлений создается общая документация о получении студентами выплат и отправляется в профком, который отправляет информацию об общей сумме выплат государству. После получения информации о необходимой сумме выплат и формирования бюджета, университету отправляется необходимая для выплат сумма. Она же распределяется между студентами. ****

Рисунок 2.1 – Схема материальные выплаты студентам

в нотации BPMN 2.0

**2.2 Бизнес-процесс «Работа приемной комиссии»**

Процесс «Работа приемной комиссии» представлен на рисунке 2. После того как управление ВУЗом составляет учебный план, приемная комиссия может начать получение заявлений на подачу документов от абитуриентов. В случае, если заявление верное, абитуриент добавляется в список поступающих, иначе заявление отклоняется. После этого список поступающих обновляется, и после окончания приемной комиссии, абитуриента либо зачисляют, либо ему отказывают в зачислении. После чего списки зачисленных отправляются на кафедры, где формируют группы студентов и составляют расписание. После чего расписание публикуется на сайте кафедры.

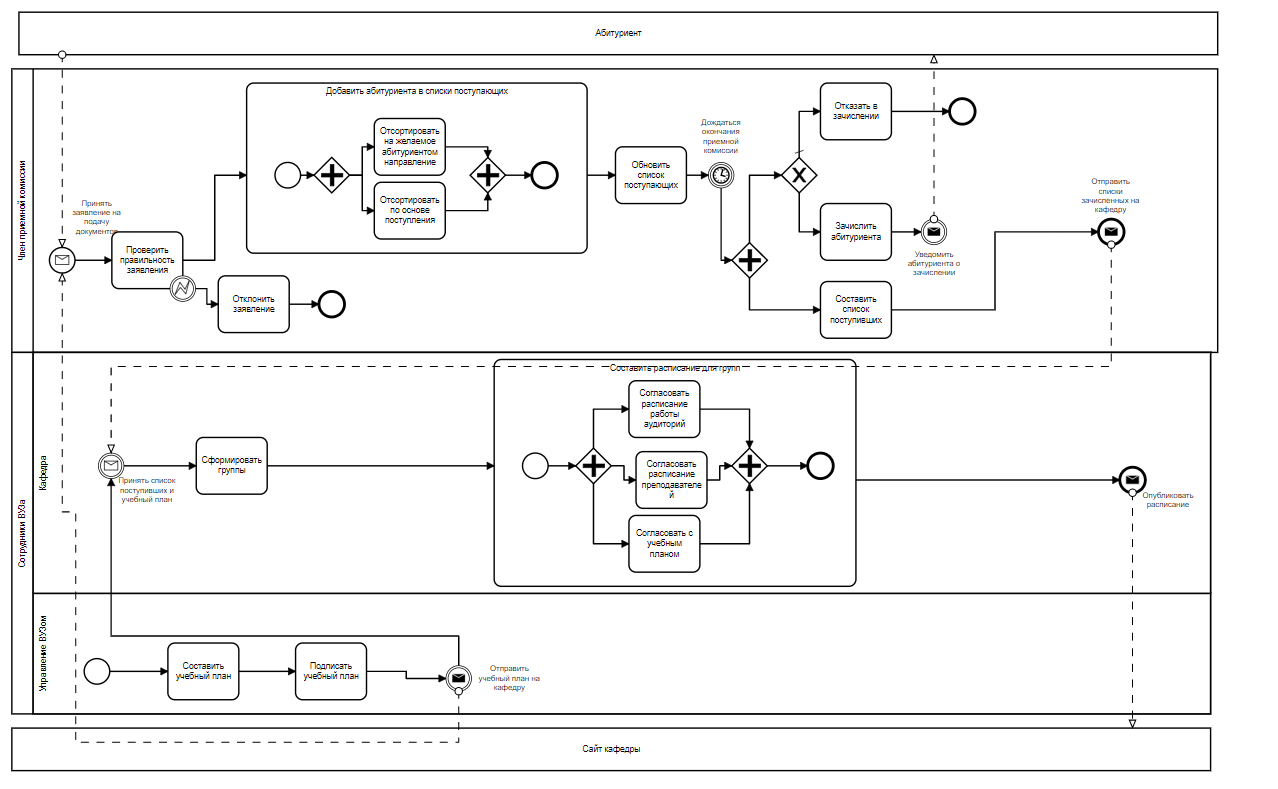


Рисунок 2.2 – Схема работа приемной комиссии в нотации BPMN 2.0

## **Хранилище данных**

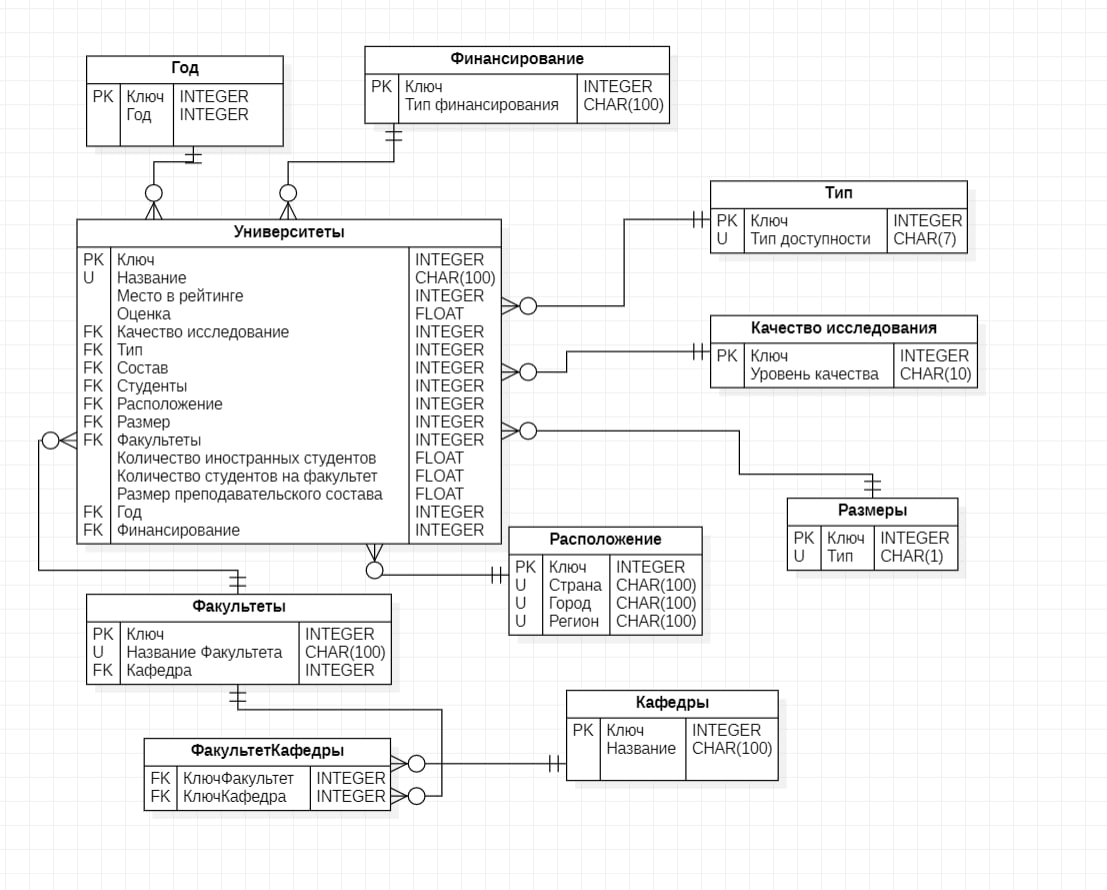


Рисунок 3 – Схема хранилища данных

На рисунке 3 представлена схема хранилища данных. Расчет объема хранилища данных: INT – 23 поле, CHAR – 9 полей, в 6 - по 100 символов, в 1 – 10 символов, в 1 – 1 символ, в 1 – 7 символов, FLOAT – 4 поля.

Полная копия создается из инкрементов один раз в неделю, необходимо 14 точек восстановления. Значит, постоянно будем хранить 2 полные копии и 12 инкрементов. Общий вес ВМ = 6483 \* (23 \* 4 + 6 \* 100 \* 1 + 10 \* 1 + 1 \* 1 + 10 \* 7 + 4 \* 4) = 0,63 Мб.

Размер инкрементальной копии зависит от количества изменений, записанных с последней точки восстановления. Для удобства допустим, что изменений происходит немного – 10%. Тогда вес инкрементов – 0,63 \* 10% \* 12 = 0,76 Мб. Итого 0,63 + 0,76 = 1,39 Мб. – то, что будет храниться постоянно, это нижняя граница графика.

## **Визуализация данных**

* 1. **Карта количества высших уч. заведений попавших в рейтинг**

На рисунке 4.1 изображена карта количества высших уч. заведений, попавших в рейтинг на одну страну. Данная карта показывает, что наибольшее количества таких университетов находится в США.

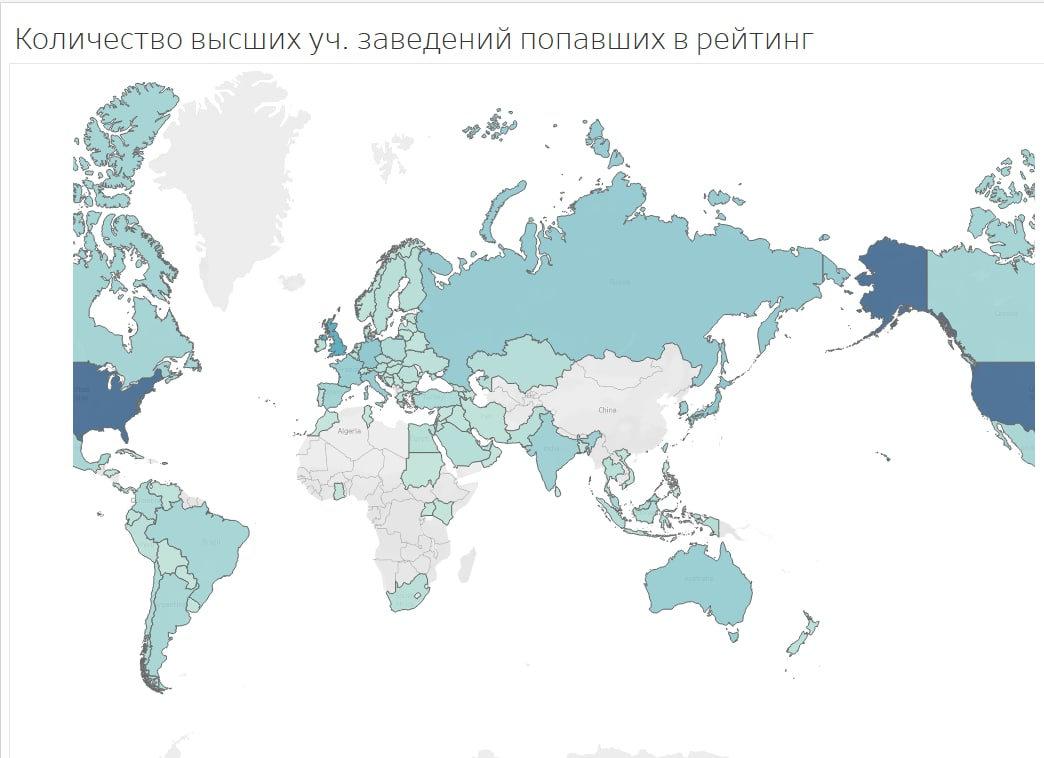


Рисунок 4.1 – Карта количества высших уч. заведений попавших в рейтинг

* 1. **Зависимость рейтинга от размеров и типа университета**

На рисунке 4.2 изображена диаграмма зависимости рейтинга от размеров и типа университета. На данной диаграмме видно, что рейтинг университета практически не зависит от его размера, однако частные университеты в среднем имеют больший бал в рейтинге.

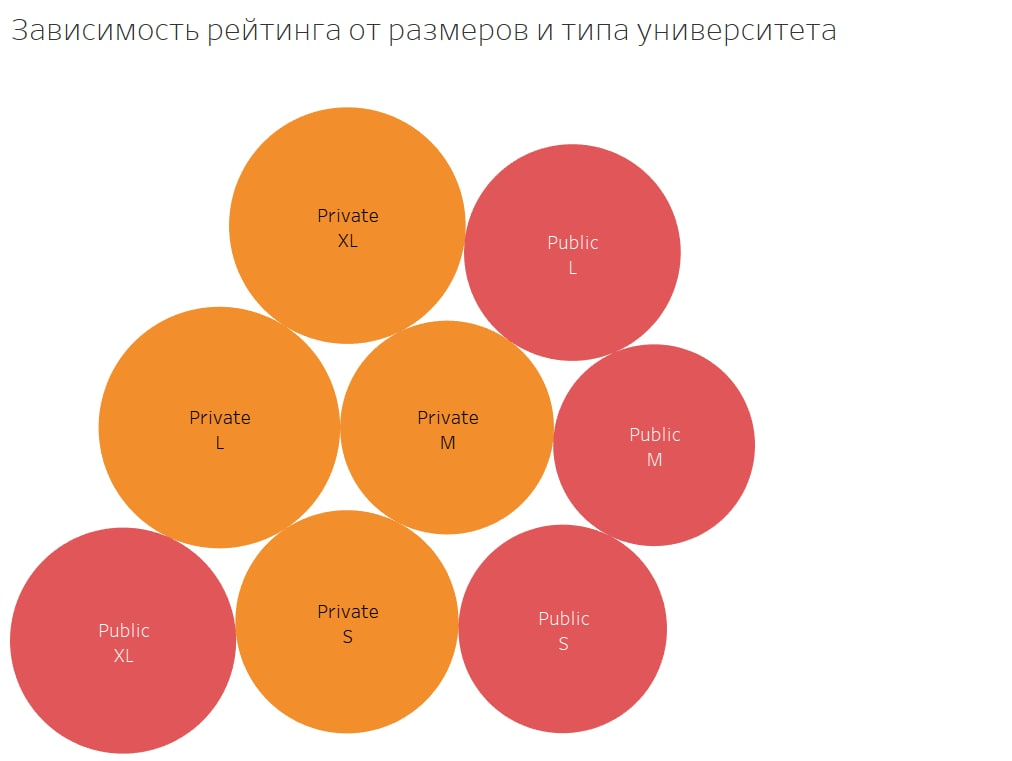


Рисунок 4.2 – Диаграмма зависимости рейтинга от размеров и типа университета

* 1. **Изменение рейтинга с 2017 по 2022 год**

На рисунке 4.3 изображена карта изменения рейтинга с 2017 по 2022 год. Несмотря на то, что изменения в оценках не так велики, можно проследить, что с годами в большинстве университетов кол-во баллов рейтинга растет.

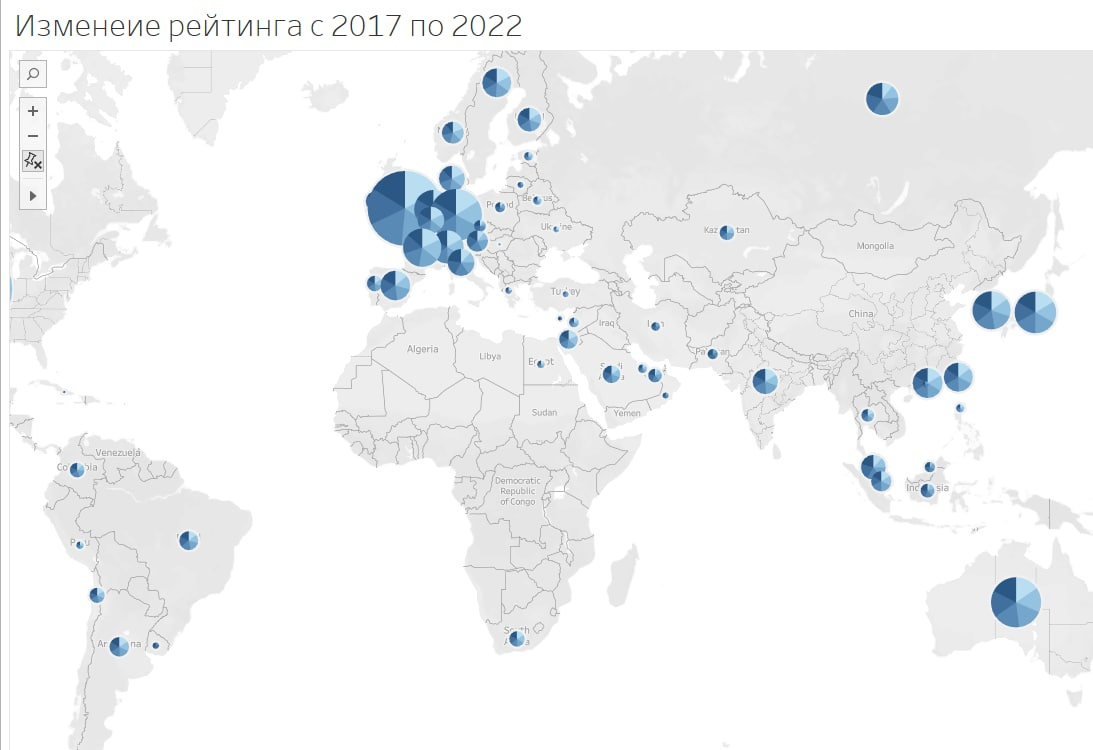


Рисунок 4.3 – Карта изменения рейтинга с 2017 по 2022 год

* 1. **Зависимость баллов рейтинга от соотношения количества студентов и преподавателей.**

На рисунке 4.4 изображена диаграмма зависимости баллов рейтинга от соотношения количества студентов и преподавателей. На диаграмме четко видно, что среднее количество баллов уменьшается с увеличением соотношения количества студентов и преподавателей.

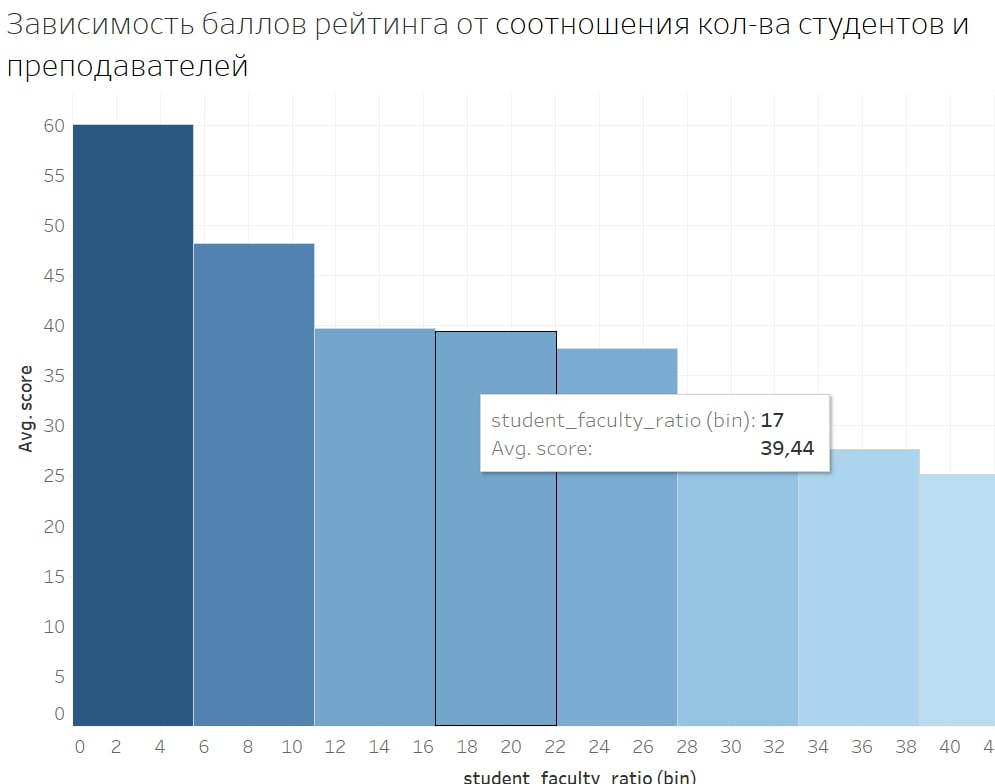
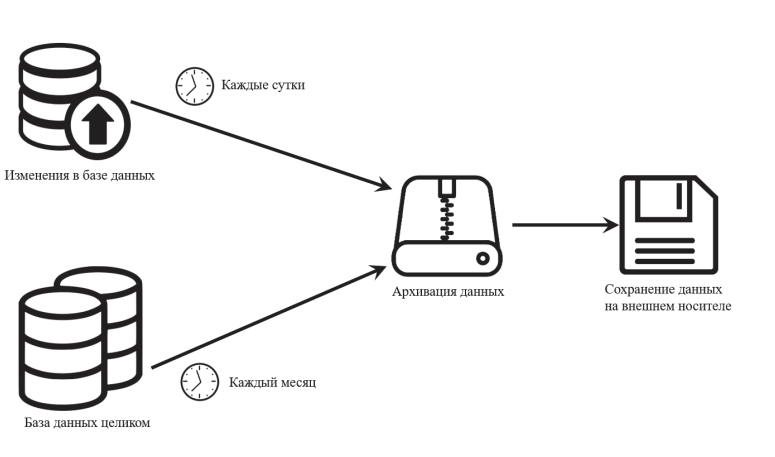


Рисунок 4.4 – График зависимости возрастного рейтинга проекта от его оценки критиками

## **План архивации и восстановления данных**

На рисунке 5.1 представлена схема архивации данных. Раз в месяц предусматривается полное резервное копирование. Делать это чаще не имеет смысла, т.к. вероятность аварийной ситуации и утери всех данных невелика. Кроме того, раз в сутки в 4 часа утра (UTL) предусматривается дифференциальное резервное копирование. Этот период времени оптимален для того, чтобы защитить данные от вирусных программ или незначительных сбоев, не требующих восстановления БД с нуля. При дифференциальном резервном копировании происходит сохранение только тех файлов, которые были подвергнуты изменению с момента последнего полного копирования. Благодаря этому данный вид копирования можно применять относительно часто.



в 4 часа утра

Рисунок 5.1 – Схема архивации данных

На рисунке 5.2 представлена схема восстановления данных. При незначительных сбоях или заражении вирусом может быть достаточно восстановить данные с момента последнего дифференциального резервного копирования. Но при утрате всех данных потребуется полное восстановление базы данных с нуля.

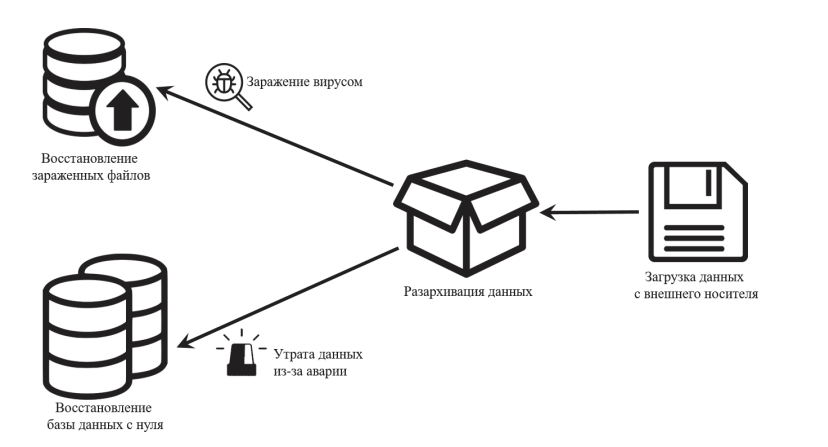


Рисунок 5.2 – Схема восстановления данных

## **ЗАКЛЮЧЕНИЕ**

В результате выполнения домашнего задания была спроектирована интерактивная система оперативного анализа данных для некоторой информационной системы по выбранной предметной области. В рамках выполнения работы были построены и проанализированы схемы бизнес-процессов в нотации BPMN 2.0, спроектировано хранилище данных на основе выбранного набора данных, а также произведена визуализация основных данных. В дополнение был продуман план архивации и восстановления данных. Как итог, были получены и укреплены на практике знания в сфере анализа и визуализации данных.

## **СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

1. Фёдоров И. Г. Моделирование бизнес-процессов в нотации BPMN2.0: Монография, Москва 2013 г. МЭСИ. – 255 стр.
2. 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.
3. Как рассчитать пространство для хранения бэкапов: [Электронный ресурс]. // URL: https://vc.ru/services/331674-kak-rasschitat-prostranstvo-dlya-hraneniya-bekapov (Дата обращения: 01.12.2022)