|  |  |
| --- | --- |
|  | **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:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

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

«Best Universities in the United Kingdom»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Студент \_\_ИУ5И-51Б\_\_\_ **\_\_\_\_\_\_\_\_\_**Кашима А.**\_\_\_\_\_\_\_**

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

Руководитель **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** Маслеников К. Ю.

(Подпись, дата) (И.О. Фамилия)

*Москва, 2024 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)

«\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_ 2024 г.

**ЗАДАНИЕ**

**на выполнение домашнего задания**

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

по теме \_\_\_\_\_\_\_ «Best Universities in the United Kingdom»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

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

Студентка группы \_\_\_\_\_\_ ИУ5И-51Б \_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Кашима Ахмед \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(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 «Best Universities in the United Kingdom»\_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 « 5 » сентября 2024 г.

**Руководитель домашнего задания**  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Маслеников К. Ю.

(Подпись, дата) (И.О.Фамилия)

**Студент \_\_\_\_\_**Кашима А.**\_\_\_\_\_\_\_\_\_\_\_\_** .

(Подпись, дата) (И.О.Фамилия)

**СОДЕРЖАНИЕ**

[ВВЕДЕНИЕ 4](#_Toc120161006)

[1 Описание предметной области 5](#_Toc120161007)

[2 Описание бизнес-процессов в нотации BPMN 2.0 6](#_Toc120161008)

[3 Хранилище данных 10](#_Toc120161009)

[4 Визуализация данных 11](#_Toc120161010)

[5 План архивации и восстановления данных 14](#_Toc120161011)

[ЗАКЛЮЧЕНИЕ 16](#_Toc120161012)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 17](#_Toc120161013)

## **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 выбран рейтинг университетов Соединенного Королевства.

В наборе данных содержатся:

1. Список университетов.
2. Мировой рейтинг.
3. Национальный рейтинг.
4. Оценка мировых рейтингов.
5. Минимальный балл для поступления.
6. Распределение иностранных студентов.
7. Оценка студентов.
8. Поступившие студенты.

Указанный набор можно проанализировать, сделать по нему выводы.

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

**2.1 Бизнес-процесс «Написание диплома»**

На рисунке 2.1 представлена схема написания студентом курсовой работы. Студент выбирает тему диплома, затем утверждает ее со своим научным руководителем и при необходимости дорабатывает, а также вместе с преподавателем устанавливают сроки выполнения и сдачи работы. Далее студент составляет план работы и приступает к написанию диплома. Разделив работу по частям, выполняет их, обосновывая и подкрепляя практически. Обязательно найти и доказать новизну и сделать выводы по проделанной работе. Закончив писать дипломную работу, студент отправляет ее на проверку научруку, который указывает на ошибки, если таковые имеются, и принимает работу. Студент далее ее оформляет в соответствии со стандартами, относит на кафедру все необходимые материалы для защиты и в итоге защищает диплом. Научрук после защиты диплома своего студента пишет отзыв по его работе. Комиссия задает вопросы выпускнику и затем выносит общую оценку за работу.

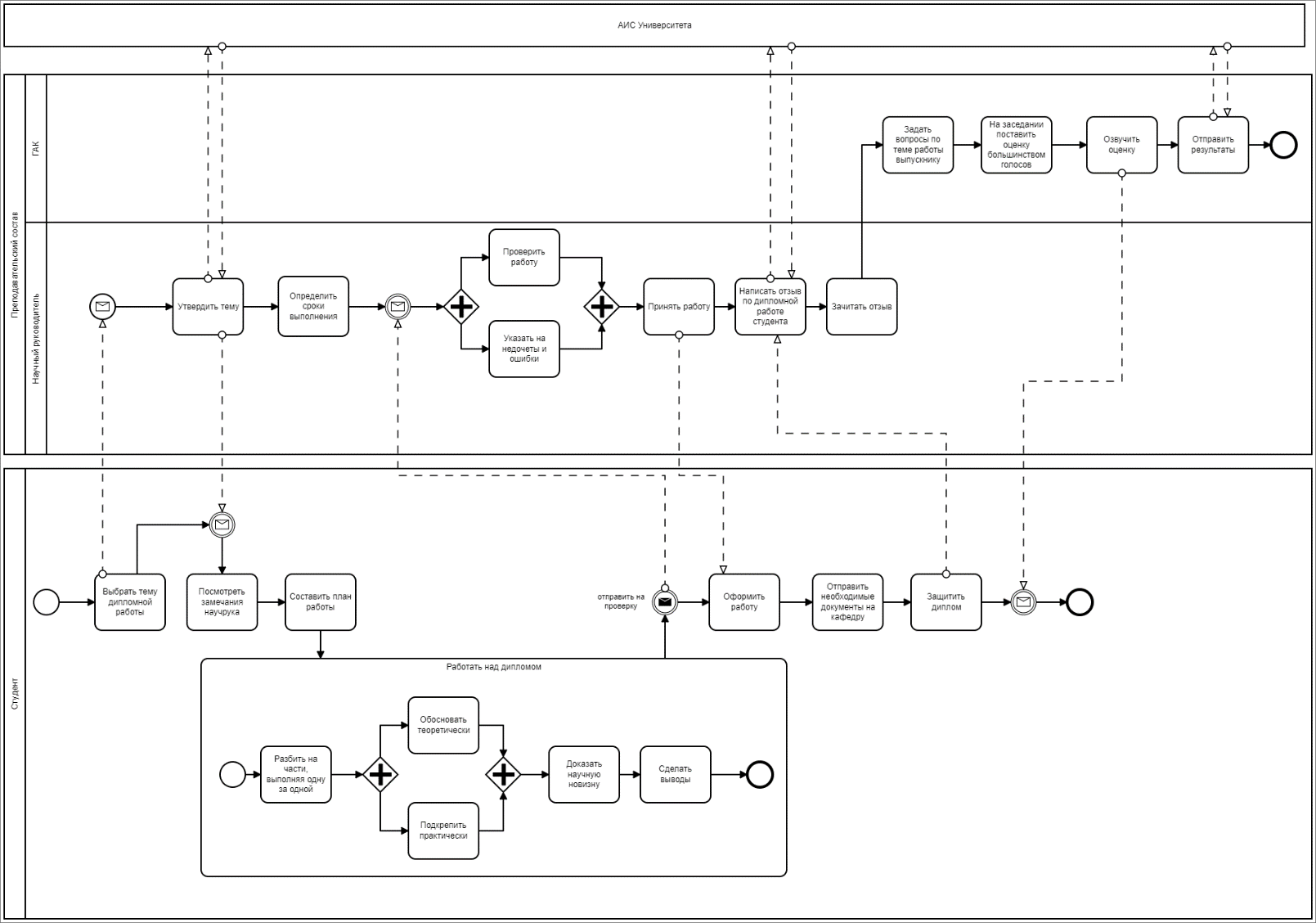


Рисунок 2.1 – Схема написания дипломной работы в нотации BPMN 2.0

**2.2 Бизнес-процесс «Поступление в ВУЗ»**

На рисунке 2.2 представлена примерная схема алгоритма поступления абитуриента в университет. После успешной сдачи ЕГЭ абитуриент выбирает ВУЗ и подает в него документы. Работники приемной комиссии проверяют все документы, в случае если каких-то документов не хватает или в случае возникших проблем, абитуриенту отказывают в приеме документов, вследствие чего абитуриент исправляет все и подает документы заново или выбирает другой ВУЗ. В случае удачной проверки работники ПК направляют личное дело абитуриента на выбранный факультет. Студенту, в свою очередь, приходит уведомление о том, что его документы приняты. Далее, в зависимости от требований ВУЗа, абитуриент либо пишет вступительные экзамены, либо проходит/не проходит по ЕГЭ. Факультет проводит вступительные испытания, проверяет работы, выставляет баллы, оформляет и отправляет результаты работникам приемной комиссии, которые составляют списки на зачисление с учетом сданных испытаний и результатов ЕГЭ. Студент получает уведомлении о пришедших результатах и ищет себя в списках на зачисление и подает согласие на зачисление/выбирает другой университет для поступления.

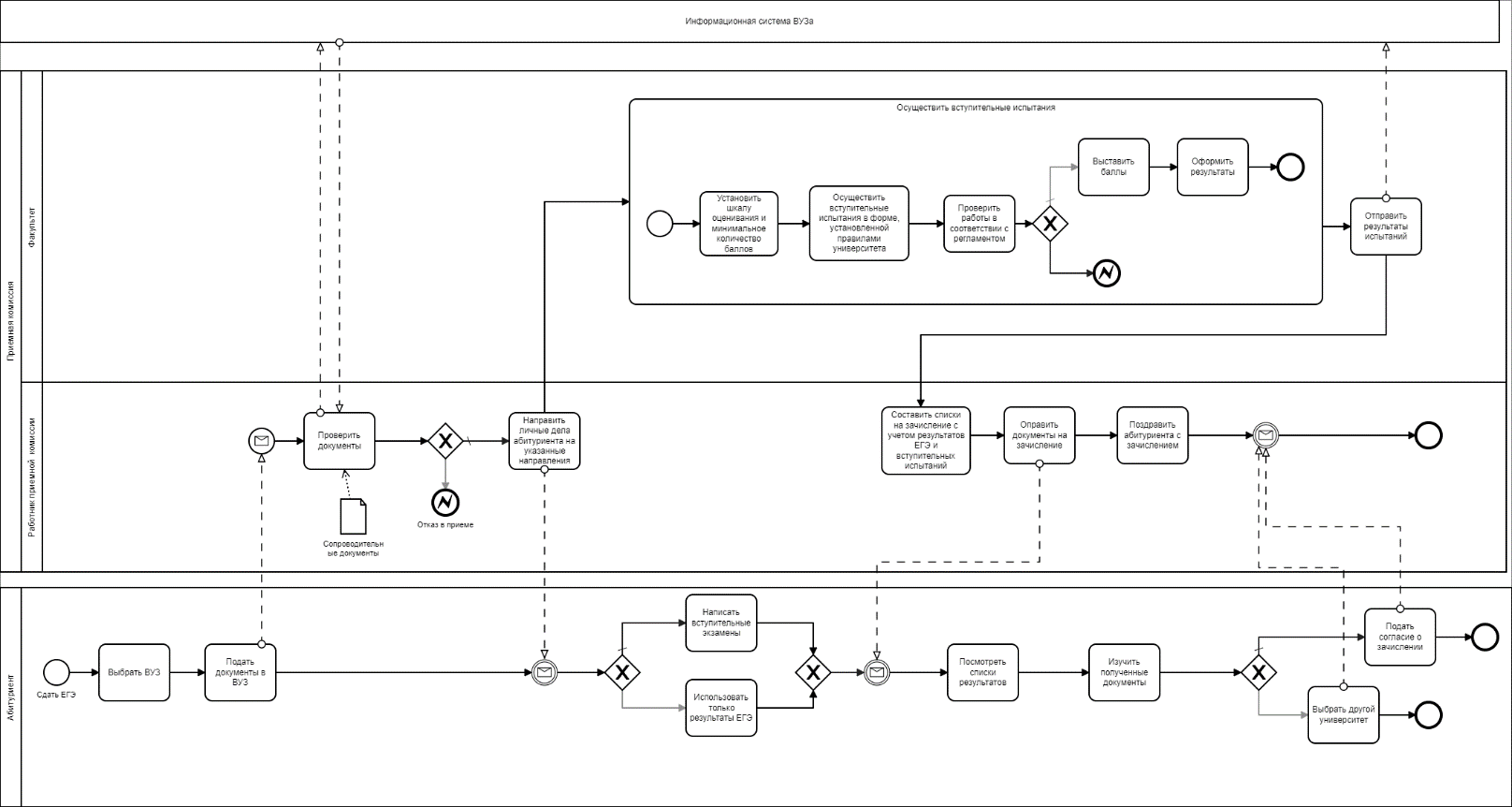


Рисунок 2.2 – Схема поступления в ВУЗ в нотации BPMN 2.0

## **Хранилище данных**

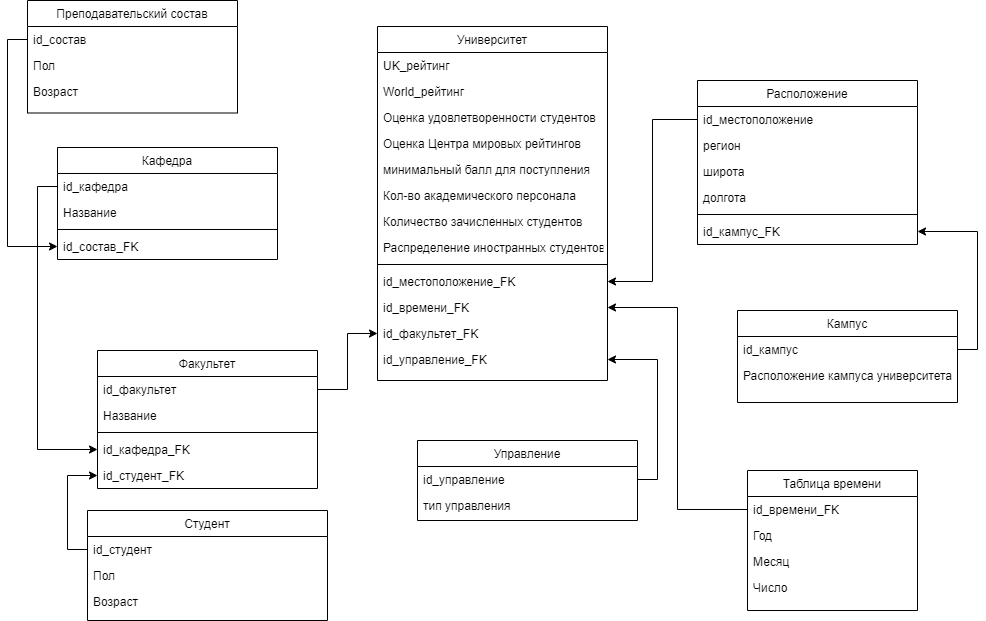


Рисунок 3 – Схема хранилища данных

Общий объем хранилища данных: 0,9 Мб.

0,9 Мб \* 2 = 1,8 Мб – объем полных копий.

100% - размер количества изменений, записанных с последней точки восстановления.

0,9 Мб \* 100% \*12 = 10,8 Мб – вес инкрементов.

0,9 Мб + 10,8 Мб = 11,7 Мб – итоговый объем хранения.

## **Визуализация данных**

**4.1. Количество университетов в каждом регионе Соединенного Королевства**

Из данных, представленных на диаграмме, можно увидеть, что самое большое количество университетов находится в Лондоне.

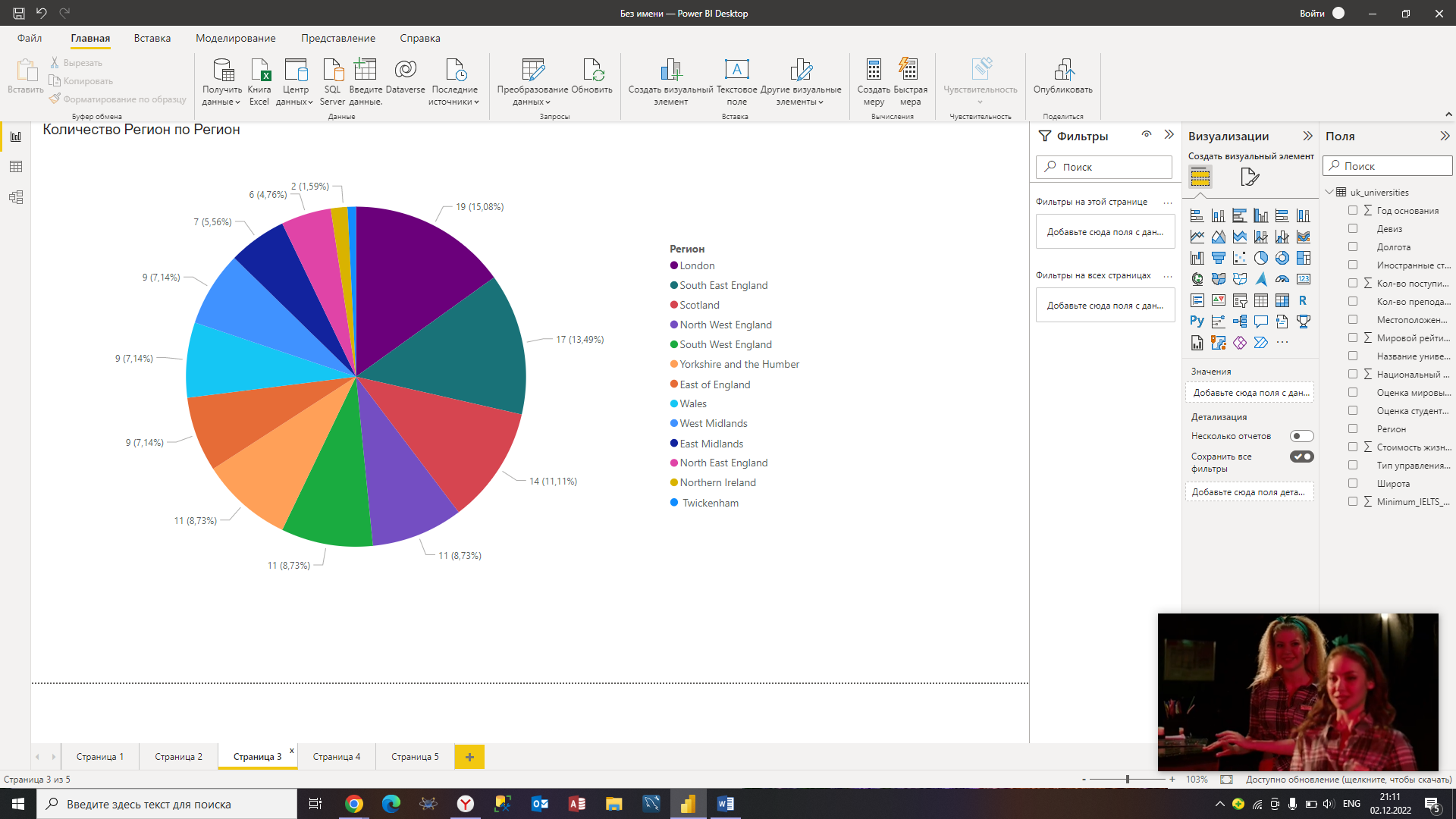
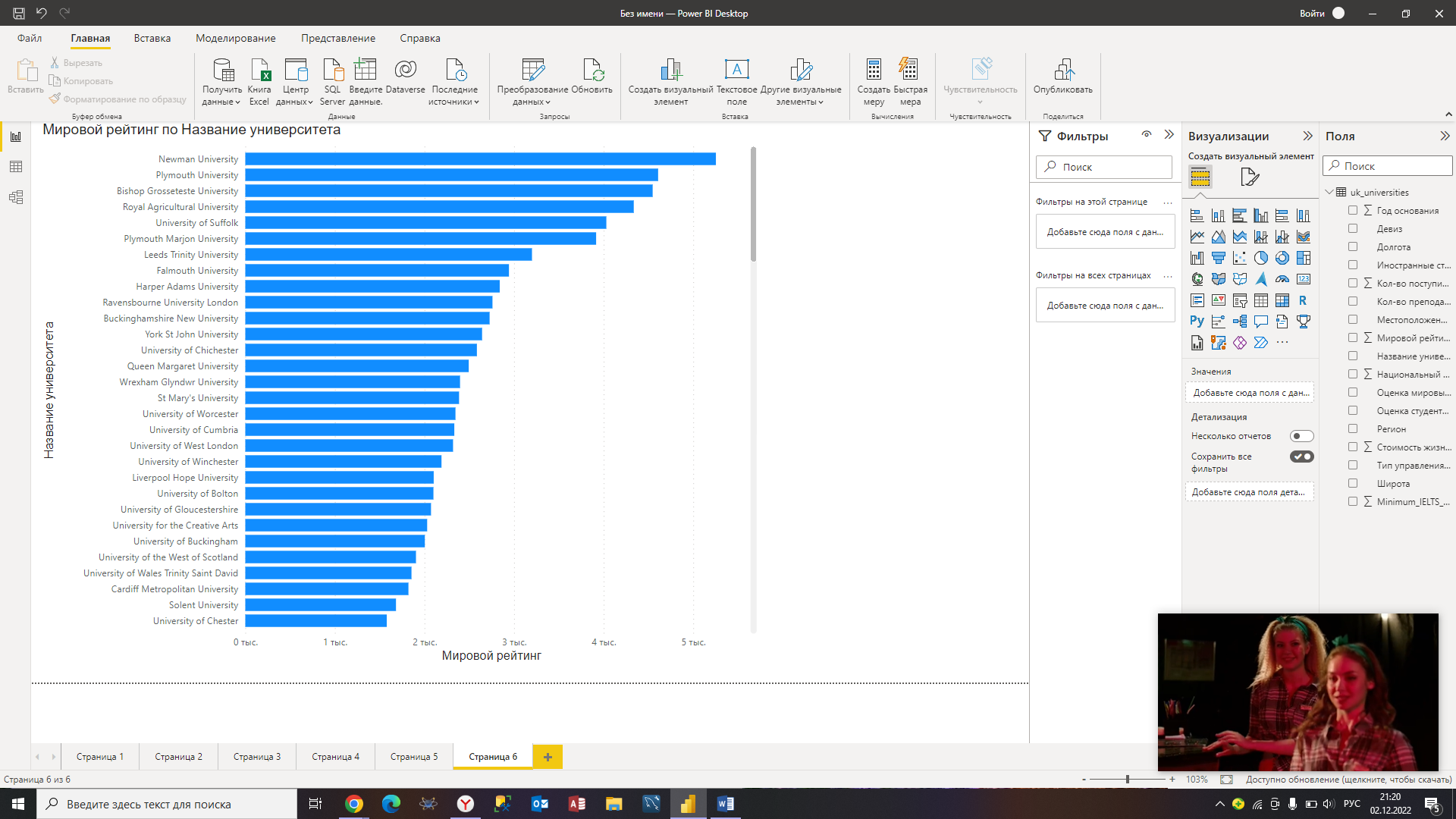


Рисунок 4.1 – Количество университетов в каждом регионе

* 1. **Мировой рейтинг университета**

Найдем самый лучший университет Соединенного Королевства с помощью гистограммы, представленной на рис.4.2, где видно, что Оксфордский Университет занимает самое высокое место по мировому рейтингу.



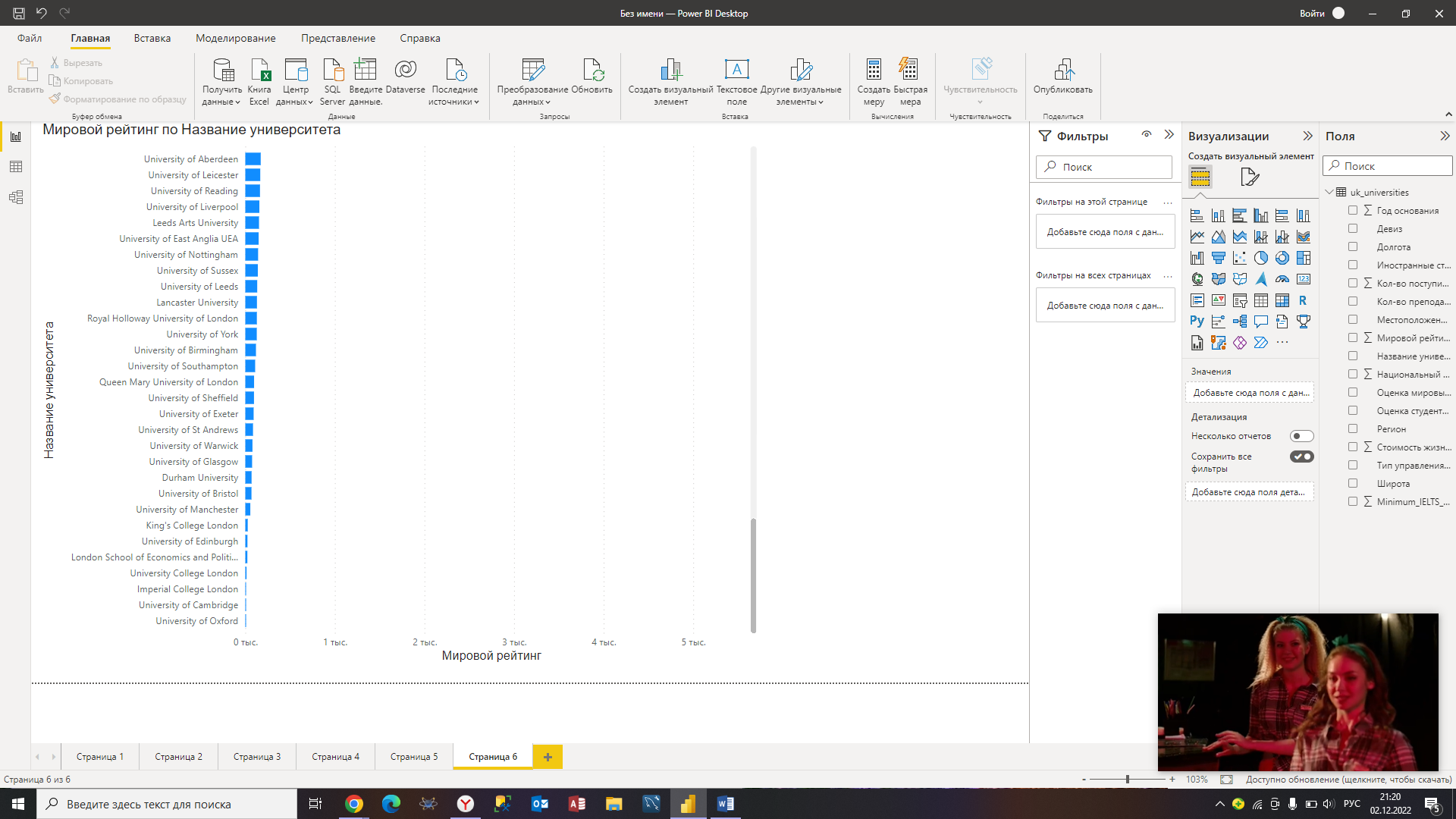
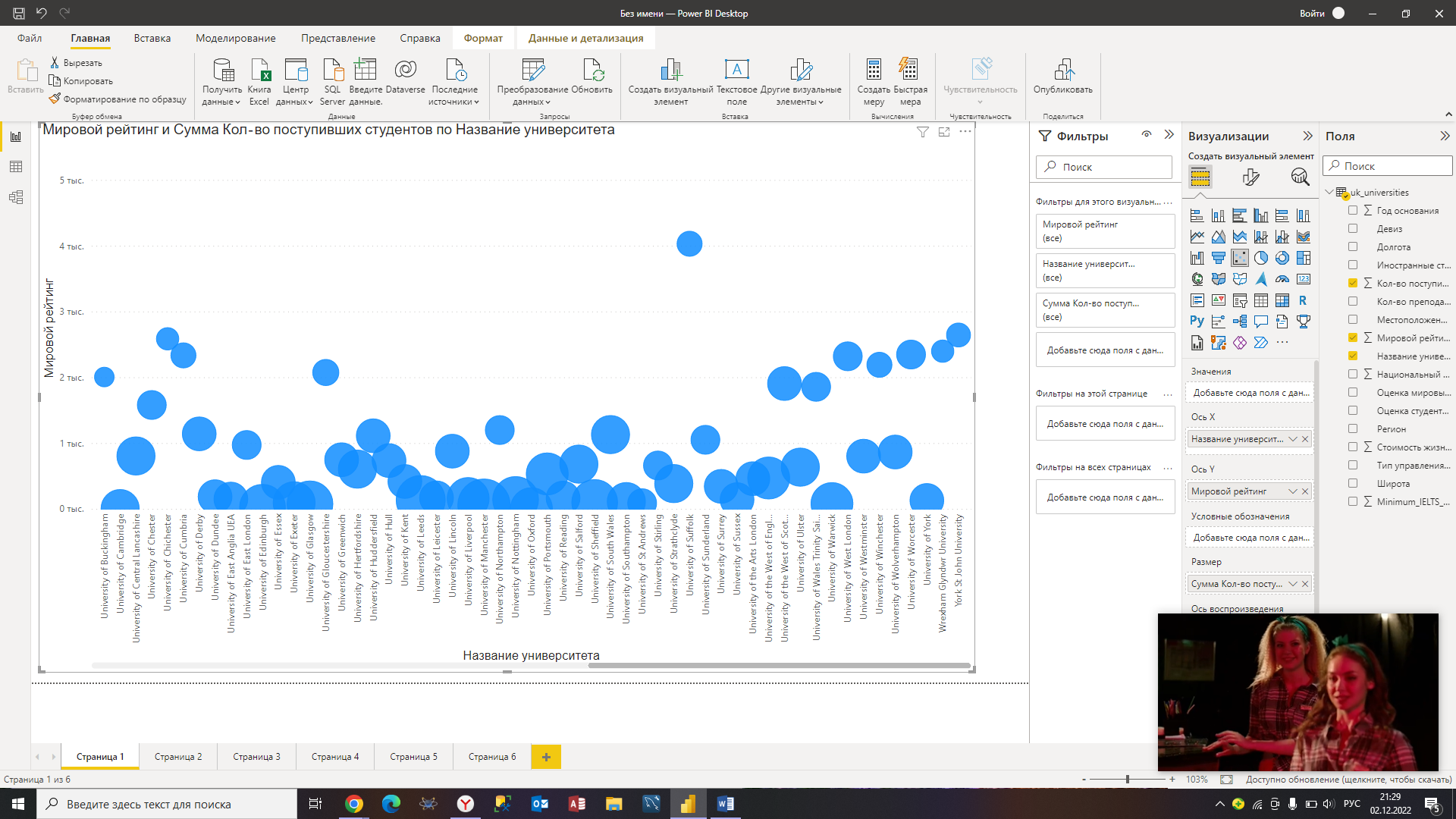


Рисунок 4.2 – Мировой рейтинга университета

* 1. **Топ университетов для поступления**

А теперь посмотрим, в какие университеты больше всего хотят поступать абитуриенты. На данной диаграмме можем увидеть такую зависимость: чем выше рейтинги университета, тем больше студентов стремятся туда поступить и наоборот, чем ниже рейтинг, тем меньше желающих там учиться.



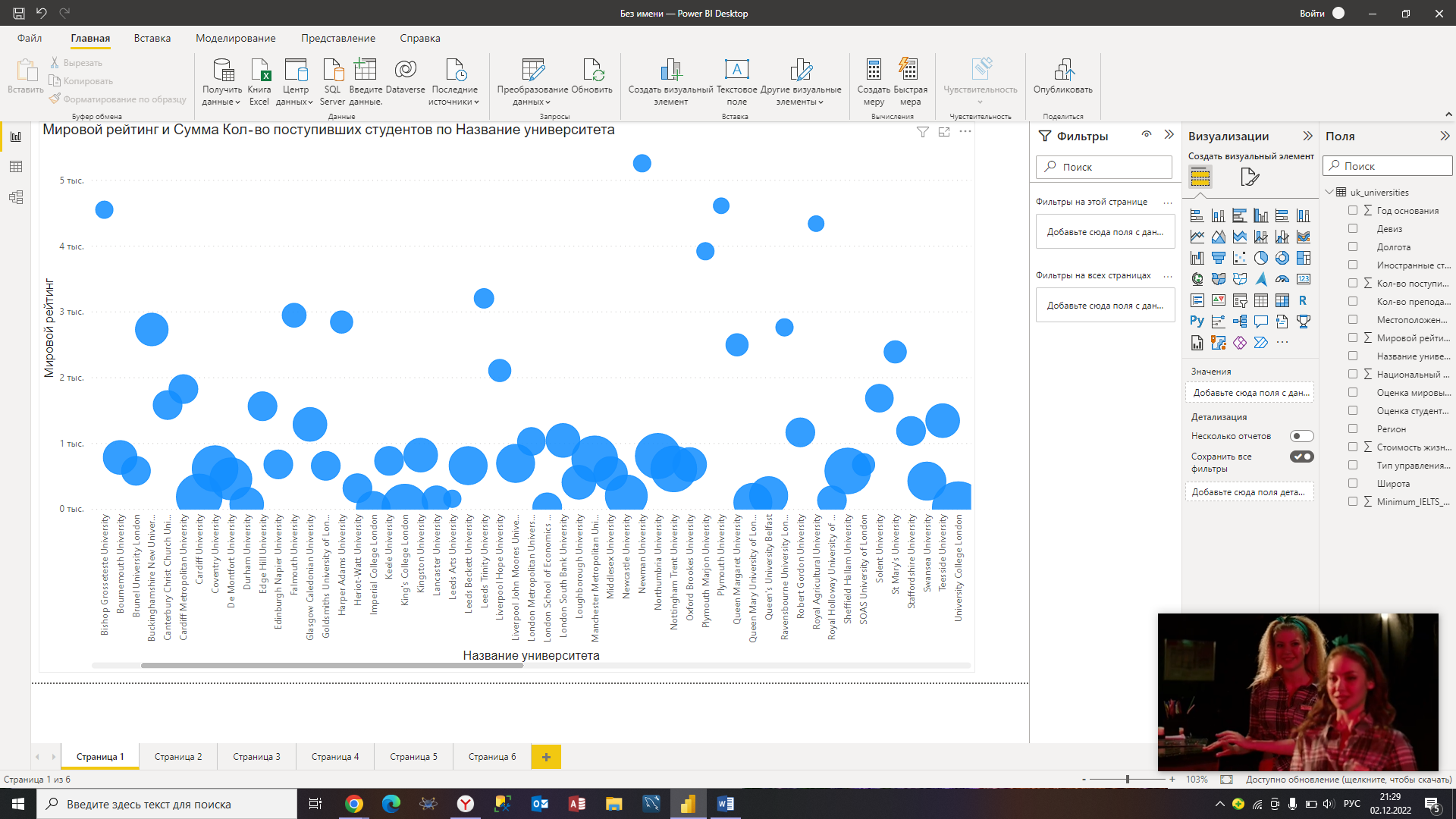


Рисунок 4.3 – Топ университетов для поступления

* 1. **Оценки студентов**

И проверим, действительно ли обучение в тех университетах, куда хотят поступать студенты, соответствует их ожиданиям.

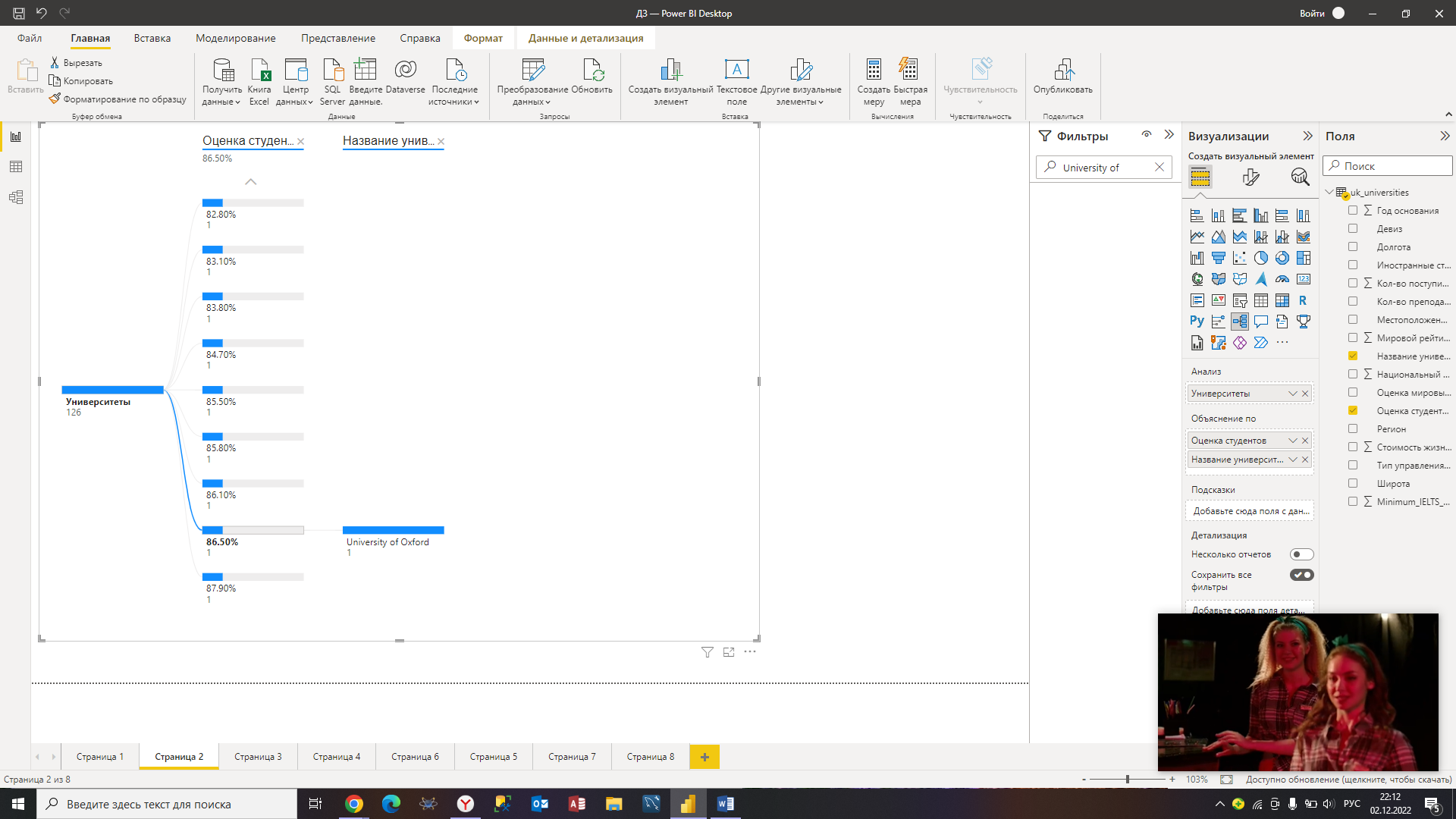
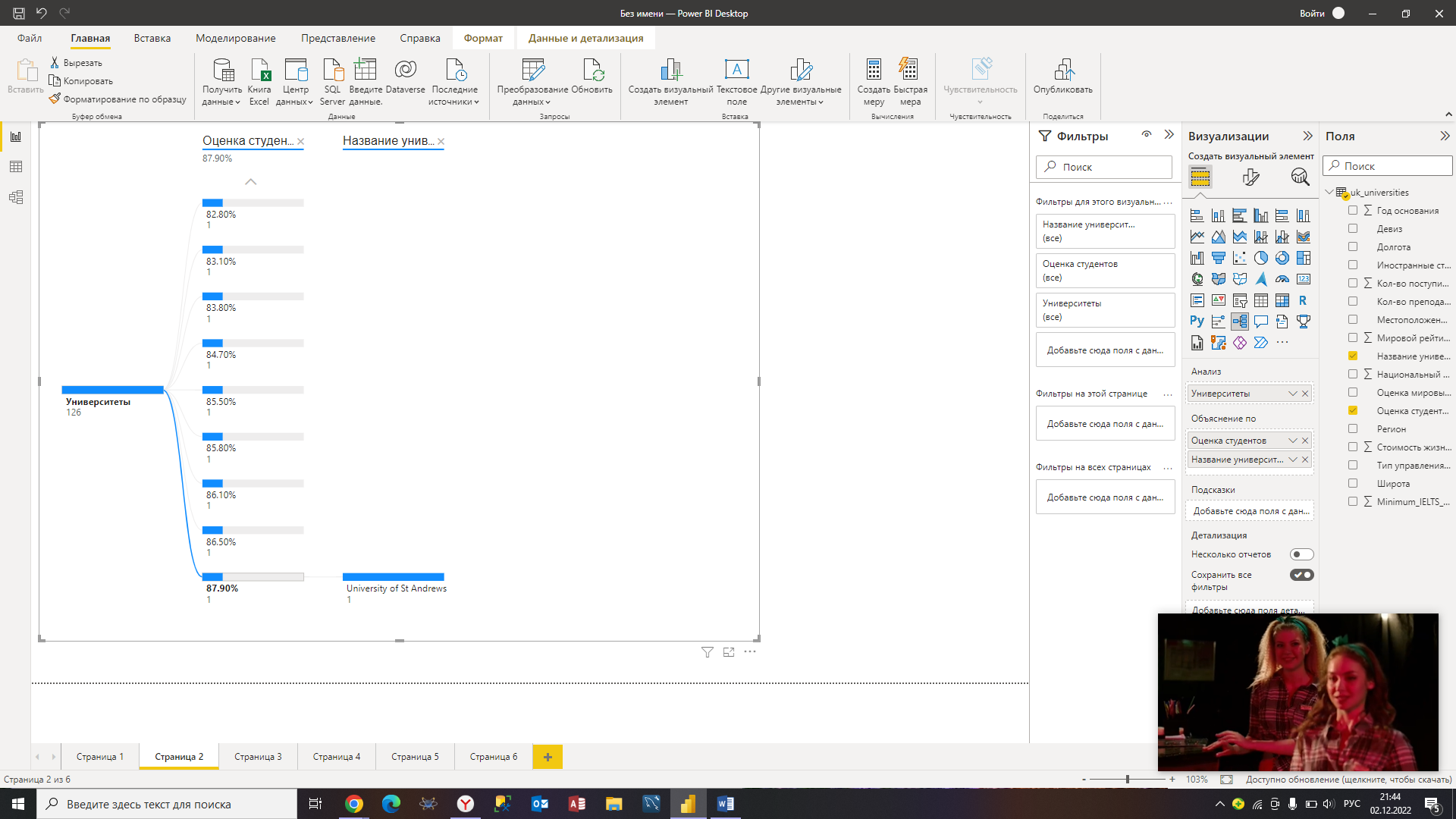


Рисунок 4.4 – Оценки студентов

Диаграмма подтверждает, что в Оксфорд не только хотят поступать ради престижа, а и то, что обучение и жизнь в этом университете устраивает студентов.



Однако самую высокую оценку удовлетворенности студентов имеет Университет св. Андрея, хотя поступают туда намного меньше, чем в университет Манчестера или в Колледж Лондона.

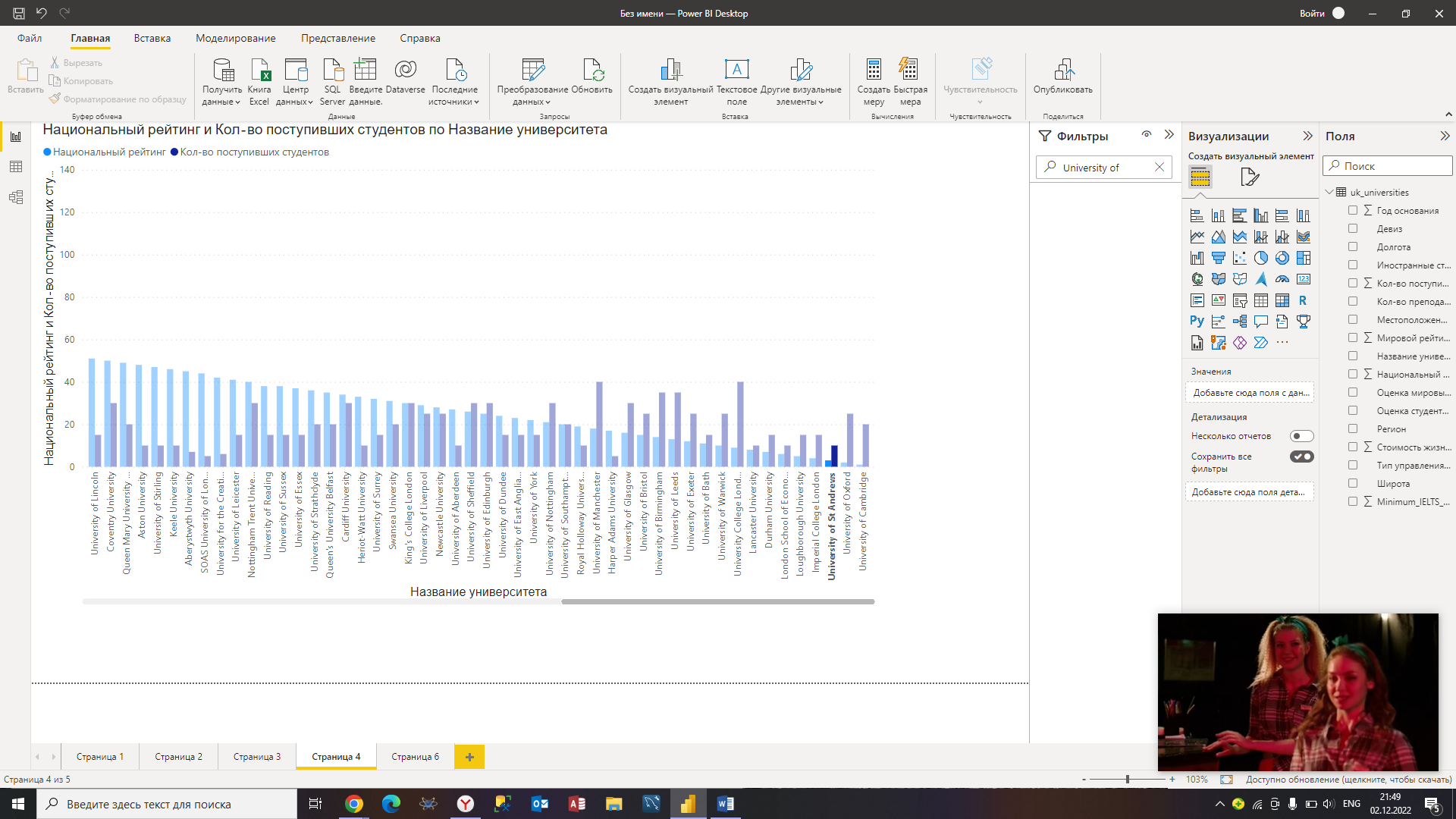
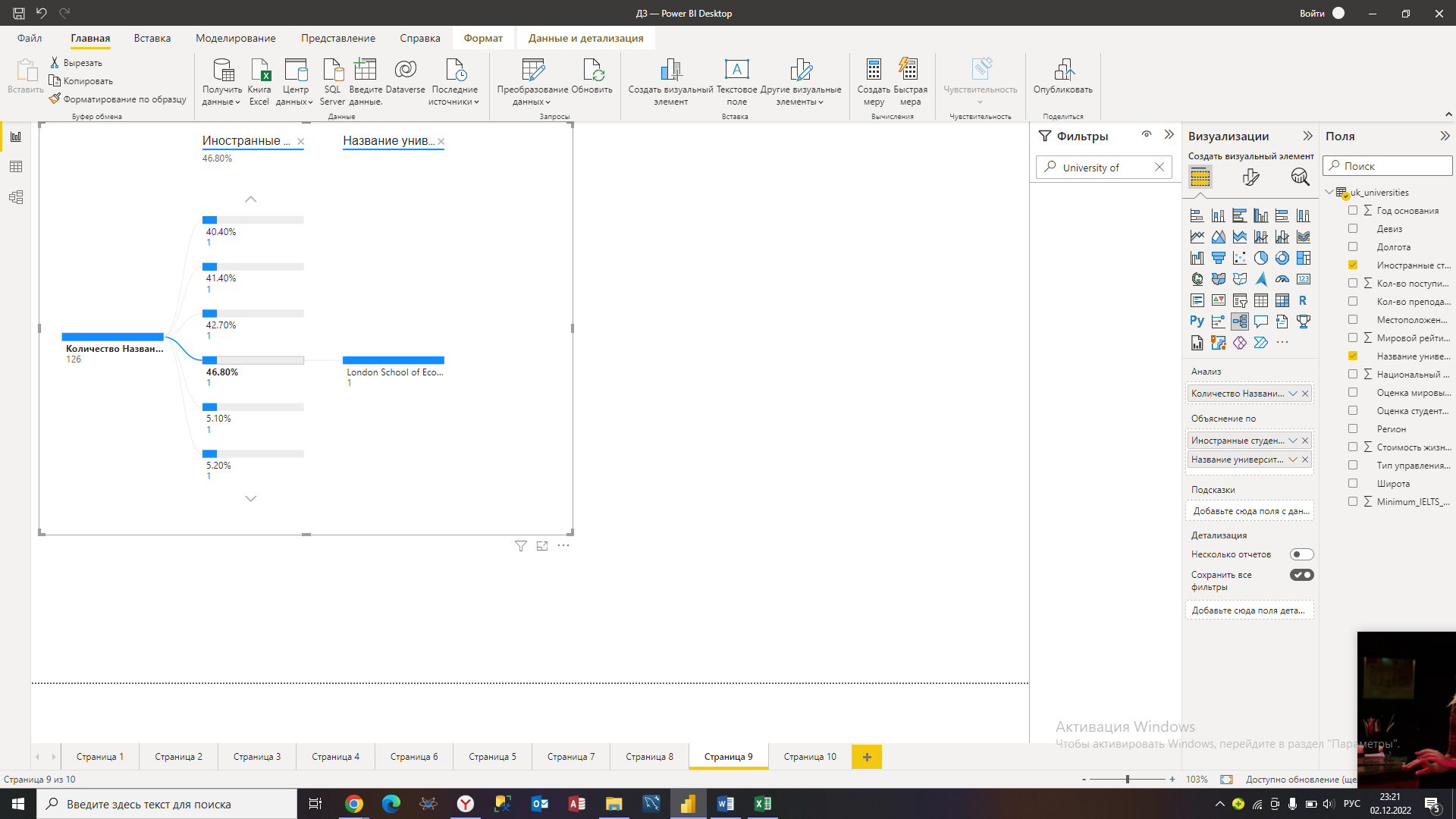
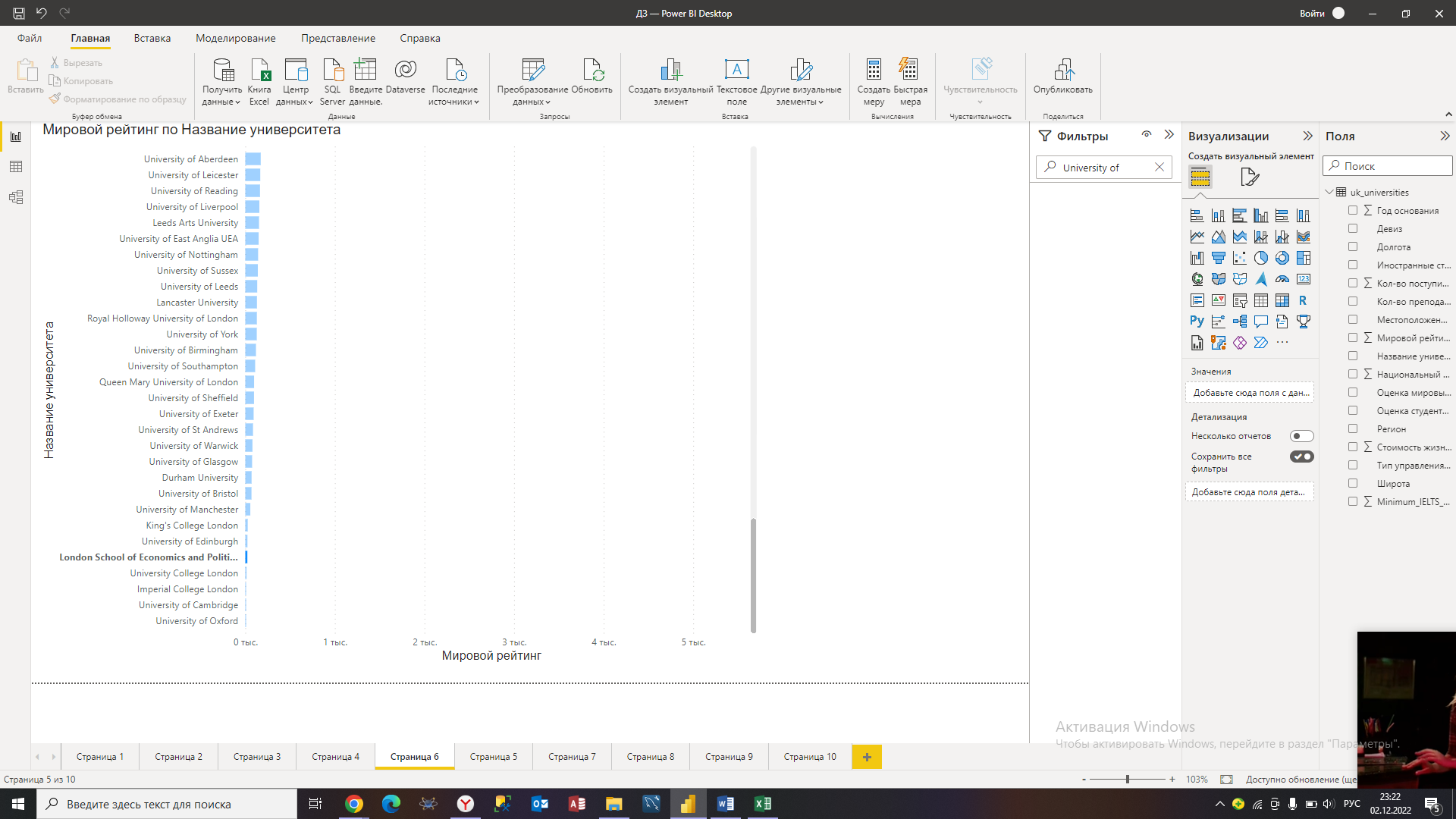


Рисунок 4.5 – График зависимости рейтинга университета от кол-ва поступивших студентов

Также о статусе университета говорит кол-во иностранных студентов, поступающих в университеты. На данном рисунке видно, что самое большое число иностранных студентов в London School of Economic and Politic.



На графике ниже можем убедиться в наших предположениях. Действительно, рейтинг данного университета достаточно высок.



## **План архивации и восстановления данных**

Целью архивации является подготовка данных для быстрого восстановления в случае необходимости. Во избежание потери важных данных необходимо разработать план их архивации и восстановления. Для создания плана архивации и восстановления данных нужно учитывать, насколько важны данные, как часто данные нуждаются в изменениях, как часто данные необходимо восстанавливать.

В первую очередь определяются критически важные данные, которые должны архивироваться ежедневно, и, наоборот, формируется список исключений для данных, не нуждающихся в архивации. В данной предметной области важными данными являются названия университетов, данные об их рейтингах, год основания, а также оценки студентов и регион, где находится университет. Самыми малозначимыми в рамках анализа являются количество преподавательского состава, кампус университета и тип управления.

Далее выбирается устройство для хранения архивных файлов и тип архивации. Для своей ПО: комбинированный тип из обычной и разностной архивации, т.е. в понедельник в 09:00 UTC выполняется обычная архивация, а в остальные дни в 09:00 UTC разностная. Такой тип требует много времени на архивацию, но обеспечивает быстрое восстановление данных.

Восстановление данных буден происходить в режиме «active-active», более известный, как синхронная репликация. В случае сбоя он способен минимизировать время простоя и потерю данных, а само восстановление проходит в автоматическом режиме. Недостатки выбранного режима заключаются в высокой стоимости оборудования и повышенной сложности поддержки и реализации.

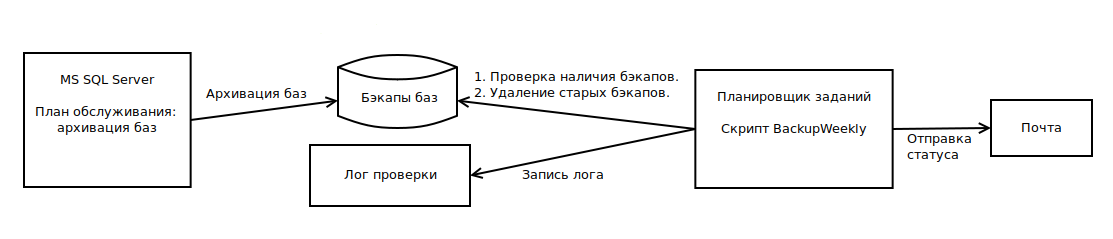


Рисунок 5.1 – Схема архивации данных

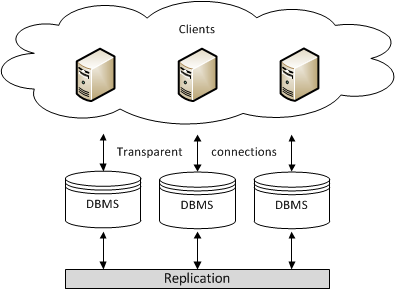


Рисунок 5.2 – Схема синхронной репликации

## **ЗАКЛЮЧЕНИ**

В результате выполнения домашнего задания было спроектировано хранилище данных по базе данных «Рейтинги университетов Соединенного Королевства». В рамках выполнения работы были описаны бизнес-процессы в нотации BPMN 2.0, проанализированы и визуализированы основные данные. Также был разработан план архивации и восстановления 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. Официальная документация Qlik Sense: [Электронный ресурс]. // [URL:https://help.qlik.com/ru-RU/cloud-services/Content/Sense\_Helpsites/Home.htm](url:https://help.qlik.com/ru-RU/cloud-services/Content/Sense_Helpsites/Home.htm) (Дата обращения: 11.11.2022)

МГПУ Архивация и восстановление данных: [Электронный ресурс].