|  | **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:\_ «Анализ результатов олимпийских игр с 1896 года» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

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

Студент \_\_ИУ5-55Б\_\_\_ **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_Киндыыкбаев Т.Н.\_\_**

(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 \_\_\_\_\_\_\_ «Анализ результатов олимпийских игр с 1896 года» \_\_\_\_\_\_\_\_\_\_\_\_

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

Студент группы \_\_\_\_\_\_ ИУ5-55Б \_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Киндыкбаев Тимур Нуржигитович\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(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 информационной системы результатов Олимпийских игр, 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 \_\_20\_\_ 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 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_ Киндыкбаев Т.Н. \_**

(Подпись, дата) (И.О.Фамилия)

СОДЕРЖАНИЕ

[ВВЕДЕНИЕ 4](#_heading=h.gjdgxs)

[1](#_heading=h.30j0zll) Описание предметной области 5

[2](#_heading=h.tyjcwt) Описание бизнес-процессов в нотации BPMN 2.0 6

[3](#_heading=h.3dy6vkm) Хранилище данных 8

[4](#_heading=h.1t3h5sf) Визуализация данных 9

[ЗАКЛЮЧЕНИЕ 17](#_heading=h.4d34og8)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 18](#_heading=h.2s8eyo1)

## **ВВЕДЕНИЕ**

Цели работы:

1. углубление и расширение теоретических знаний в сфере анализа и визуализации набора данных;
2. освоение описания, построения и анализа бизнес-процессов в нотации BPMN 2.0;
3. приобретение опыта проектирования хранилища данных;
4. практическое освоение визуализации данных;
5. поиск и нахождение подходящего способа архивации и восстановления хранилища данных.

## **Описание предметной области**

Предметной областью данного домашнего задания являются Олимпийские игры. Для анализа были собраны данные о результатах Олимпийских игр с 1986 года. В наборе данных содержатся:

1. Название страны
2. Код страны
3. Количество участников от страны на летних Олимпийских играх
4. Количество золотых медалей на летних Олимпийских играх
5. Количество серебряных медалей на летних Олимпийских играх
6. Количество бронзовых медалей на летних Олимпийских играх
7. Суммарное количество медалей на Олимпийских играх
8. Количество участников от страны на зимних Олимпийских играх
9. Количество золотых медалей на зимних Олимпийских играх
10. Количество серебряных медалей на зимних Олимпийских играх
11. Количество бронзовых медалей на зимних Олимпийских играх

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

1. Поиск стран, которые побеждают чаще всего;
2. Выявление закономерностей в результатах Олимпийских игр;
3. Предсказание интересных «матчей» на основе прошлых результатов;
4. Предсказание победителей по результатам медального зачета
5. Предложить покупателям билеты на наиболее интересные матчи

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

**2.1 Бизнес-процесс «Покупка билетов на Олимпийские игры»**

В бизнес-процессе покупки билетов на Олимпийские игры есть существенные отличия от процесса покупки в онлайн-магазине. Клиент формирует запрос на покупку билета, указывая гражданство. Страна-организатор Олимпийских игр проверяет гражданство клиента, и, если клиент имеет гражданство страны-организатора, формируется группа билетов в наличии. Если же клиент имеет другое гражданство, формируется запрос на сайт-агрегатор этой страны и билет оформляется из наличия в базе сайта-агрегатора. После этого клиент выбирает доступные билеты из списка, оплачивает покупку через банк, печатает билет и отправляется на Олимпийские игры

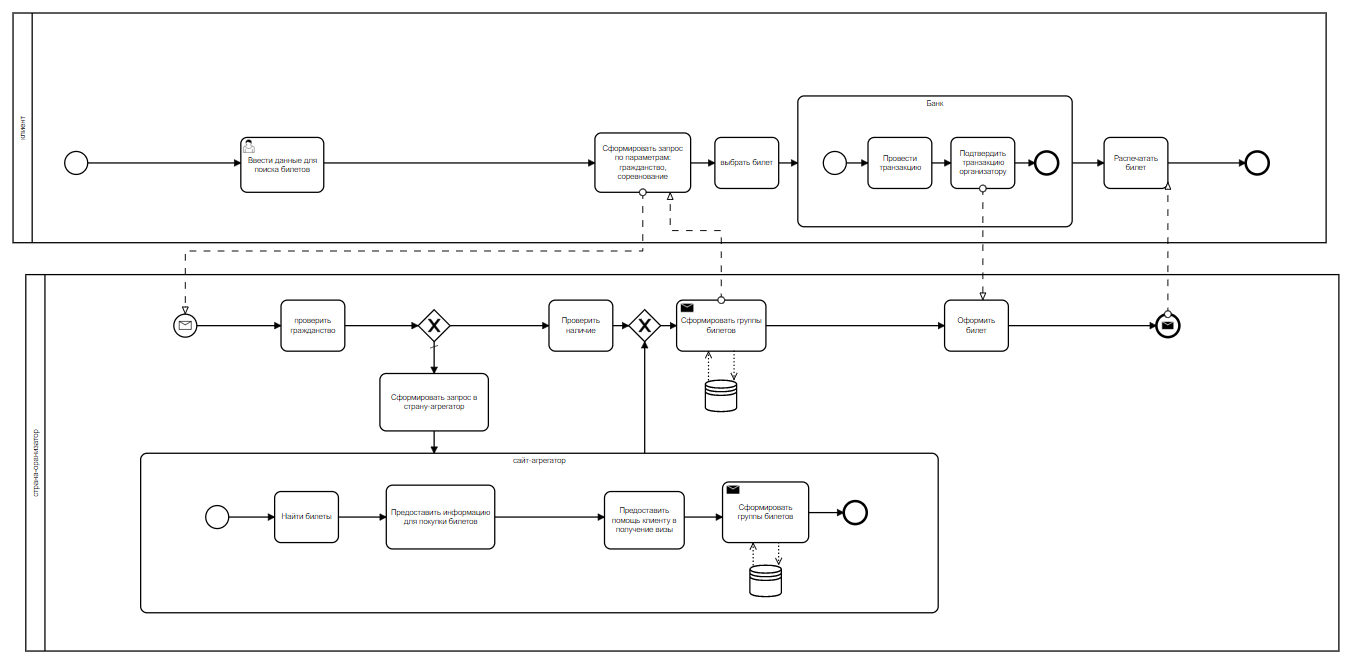


Рисунок 1.1 – Схема покупки билетов в нотации BPMN 2.0

**2.2 Бизнес-процесс «Заключение договора о рекламе»**

Рекламный баннер на олимпийских играх должен пройти много этапов перед тем, как его повесят на стадионе. В данном бизнес-процессе мы рассмотрим процесс заключения договора о рекламе.

В первую очередь выбираются условия на составление договора об оказании рекламных услуг. Далее выбирается кандидат в спонсоры и направляется запрос на заключение договора в их рекламный отдел. Параллельно агент выбирает место для рекламы.

После завершения составления условий договора его отправляют на проверку в компанию спонсора. Спонсор рассматривает договор и принимает решение об изменении условий, проводит работу с юридическим отделом.

В зависимости от принятого спонсором решения, агент составляет новые условия и заново направляет условия спонсору, или, если спонсор утвердил договор сразу, визирует документ и завершает этот бизнес-процесс

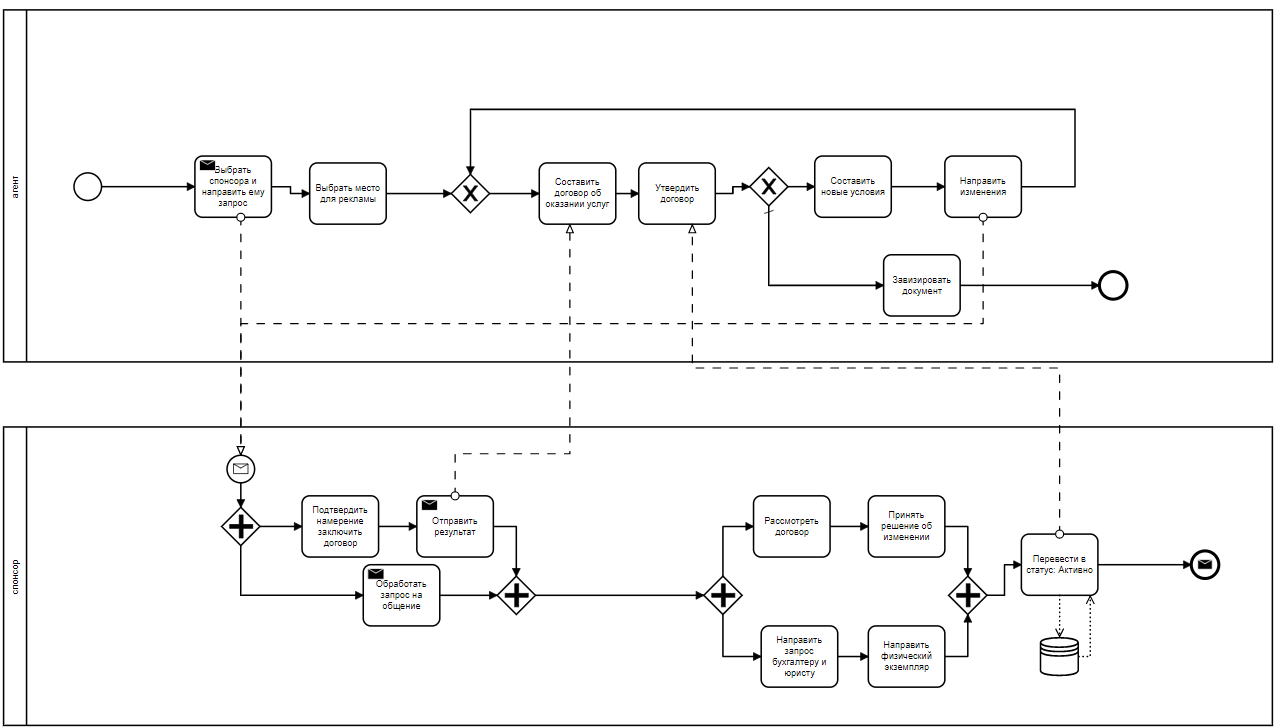


Рисунок 1.2 – Схема заключения договора на рекламу BPMN 2.0

## **Хранилище данных**

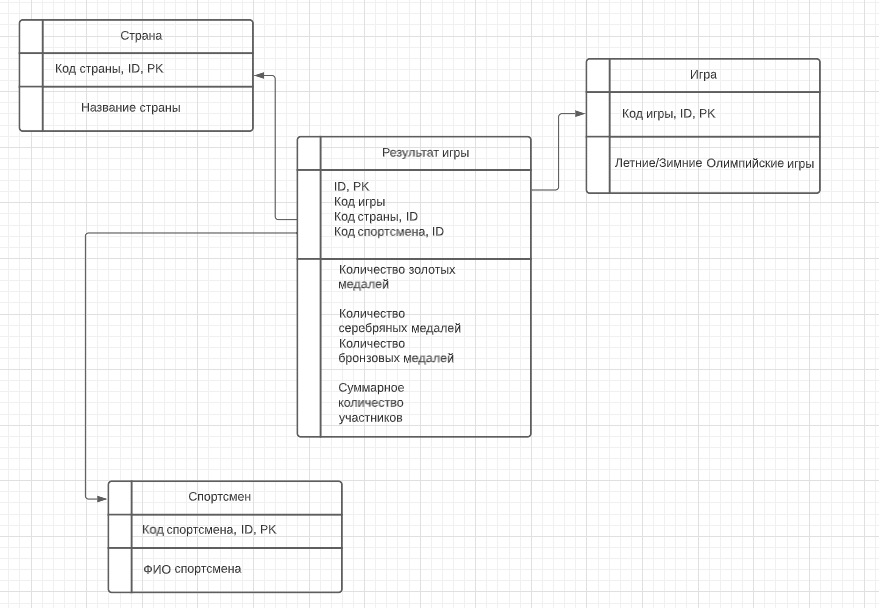


Рисунок 3 – Схема хранилища данных

Расчет объема хранилища данных: INT – 3 полей, VARCHAR – 2 поля, в каждом - по 30 символов,

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 = 56000 \* (3 \* 4 + 2 \* 30 \* 1) = 4, 03 Мб.

Размер инкрементальной копии зависит от количества изменений, записанных с последней точки восстановления. Для удобства допустим, что изменений происходит немного – 10%. Тогда вес инкрементов – 8,22 \* 10% \* 12 = 4,83 Мб. Итого 8,86 Мб. – то, что будет храниться постоянно, это нижняя граница графика.

## **Визуализация данных**

* 1. **Изображение результатов медального зачета на графике**

Данный график иллюстрирует количество золотых, серебряных и бронзовых медалей, полученных на всех Олимпийских играх для стран, сумма медалей которых имеет наибольшие значения.

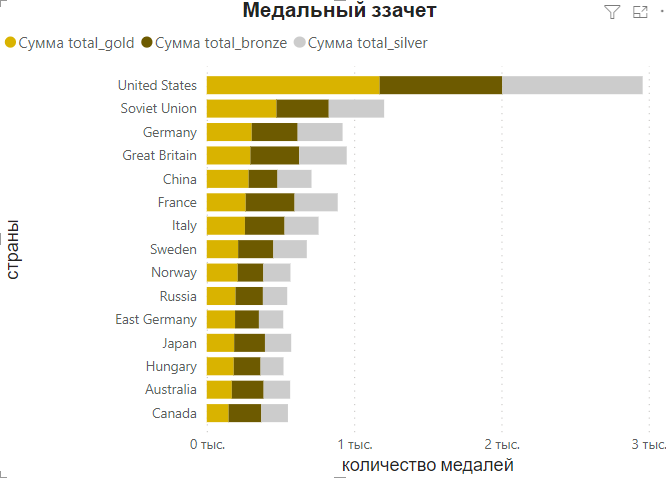


Рисунок 3.1 – Диаграмма результатов медального зачета

По графику заметно, что некоторые страны показывают значительно выделяющийся результат на фоне других. Такие значения показывают USA, а значит их вклад в историю Олимпийских игр очень значителен, однако на основе этого утверждать, что USA будут победителем медального зачета нельзя.

Победителем медального зачета за всю историю Олимпийских игр можно назвать USA.

* 1. **Распределение медалей на летних Олимпийских играх**

Данный график показывает, какие страны показывали наилучшие результаты на летних Олимпийских играх. За показатель результата я взяла суммы медалей по летним Олимпийским играм, потому что именно этот показатель наиболее описывает вклад в историю Олимпийских игр.

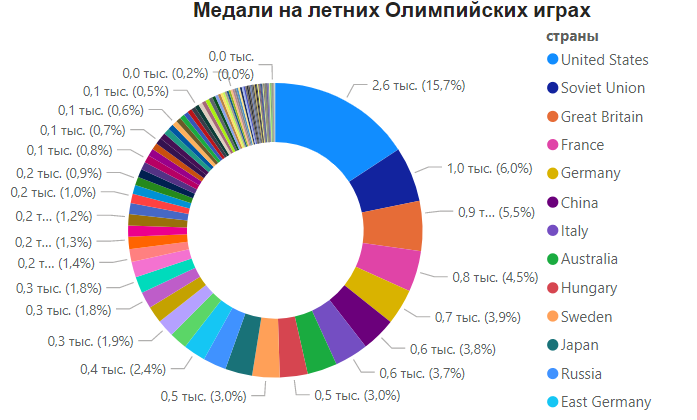


Рисунок 3.2 – Распределение медалей на летних Олимпийских играх

На графике видно, что страны, которые фигурировали в предыдущем разделе, здесь в большей степени тоже участвуют, однако порядок не совпадает. То есть какие-то страны лучше выступают на летних играх, какие-то на зимних, и эти данные нужно исследовать более подробно, чтобы изучить связь между летними и зимними Олимпийскими играми

* 1. **Распределение медалей на зимних Олимпийских играх**

Аналогичный график для зимних игр показывает распределение медалей зимних Олимпийских игр для стран.

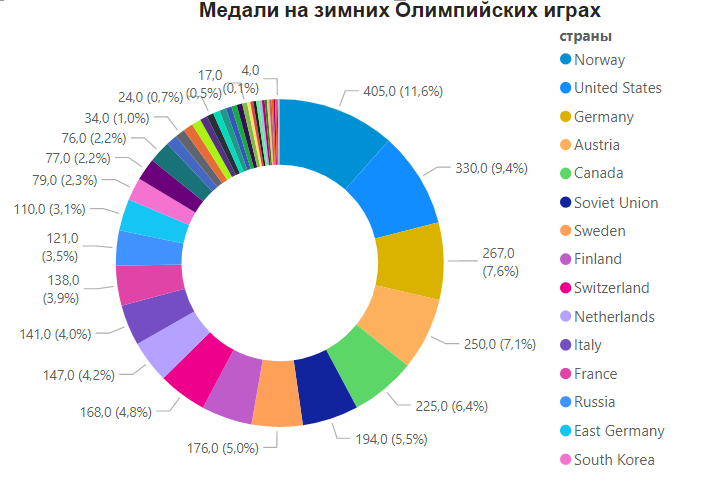


Рисунок 3.3 – Распределение медалей на зимних Олимпийских играх

На этом графике получились очень интересные результаты – они довольно сильно отличаются от результатов летних игр. Победителем медального зачета для зимних Олимпийских игр становится Норвегия, которая значительно обогнала США. Для Норвегии действительно правдиво высказывание о том, что выступление на зимних играх значительно вероятнее пройдет лучше, чем выступление на летних играх.

Также среди стран-победительниц появились Канада, Швейцария и Финляндия, то есть страны Северной Европы и Америки, значит возможно, что результаты зимних Олимпийских игр для стран северных регионов повышаются

* 1. **Диаграмма команд участников Олимпийских игр**

Эта диаграмма показывает суммарное количество участвующих команд от каждой страны на летних и зимних Олимпийских играх. Наибольшее количество участников отправили следующие сраны: Франция, Великобритания, Греция, Швейцария, Австрия, Дания, Италия, Швеция. То есть в основном масса участников приходится на страны Европы и Северной Америки, однако достаточно участников приходится на разные страны мира.

При этом, вспоминая результаты медального зачета, по-видимому результаты Олимпийских Игр не сильно зависят от количества отправленных участников

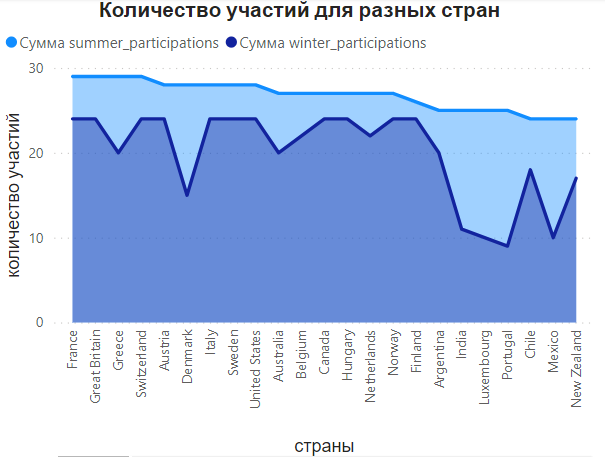


Рисунок 3.4 – Диаграмма количества участников

* 1. **Диаграмма соотношения медалей**

Мы выдвинули предположение, что количество медалей для зимних Олимпийских игр и летних Олимпийских игр отличается: хорошее выступление на летних Олимпийских играх не гарантирует хорошего выступления на зимних Олимпийских играх.

По графику видно, что количество медалей на летних Олимпийских играх распределено неравномерно, а вот результаты зимних игр довольно равномерно. Получается большой разброс между результатами, и зависимость нужно изучать более детально

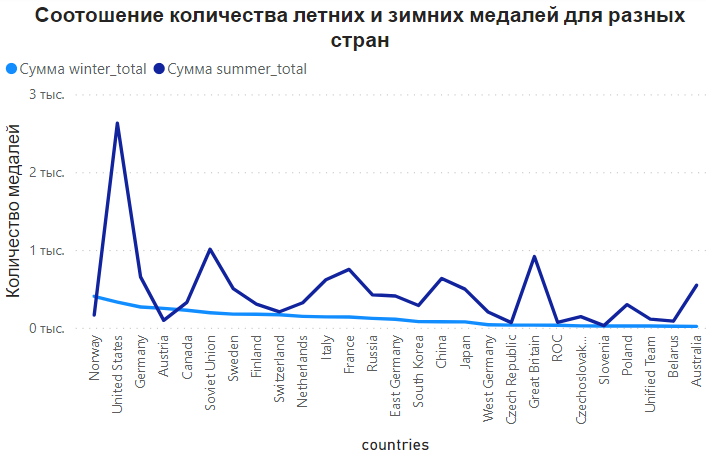


Рисунок 3.5 – Диаграмма количества автомобилей в зависимости от типа топлива

* 1. **Совокупность таблиц для получения полной информации о результатах Олимпийских игр**

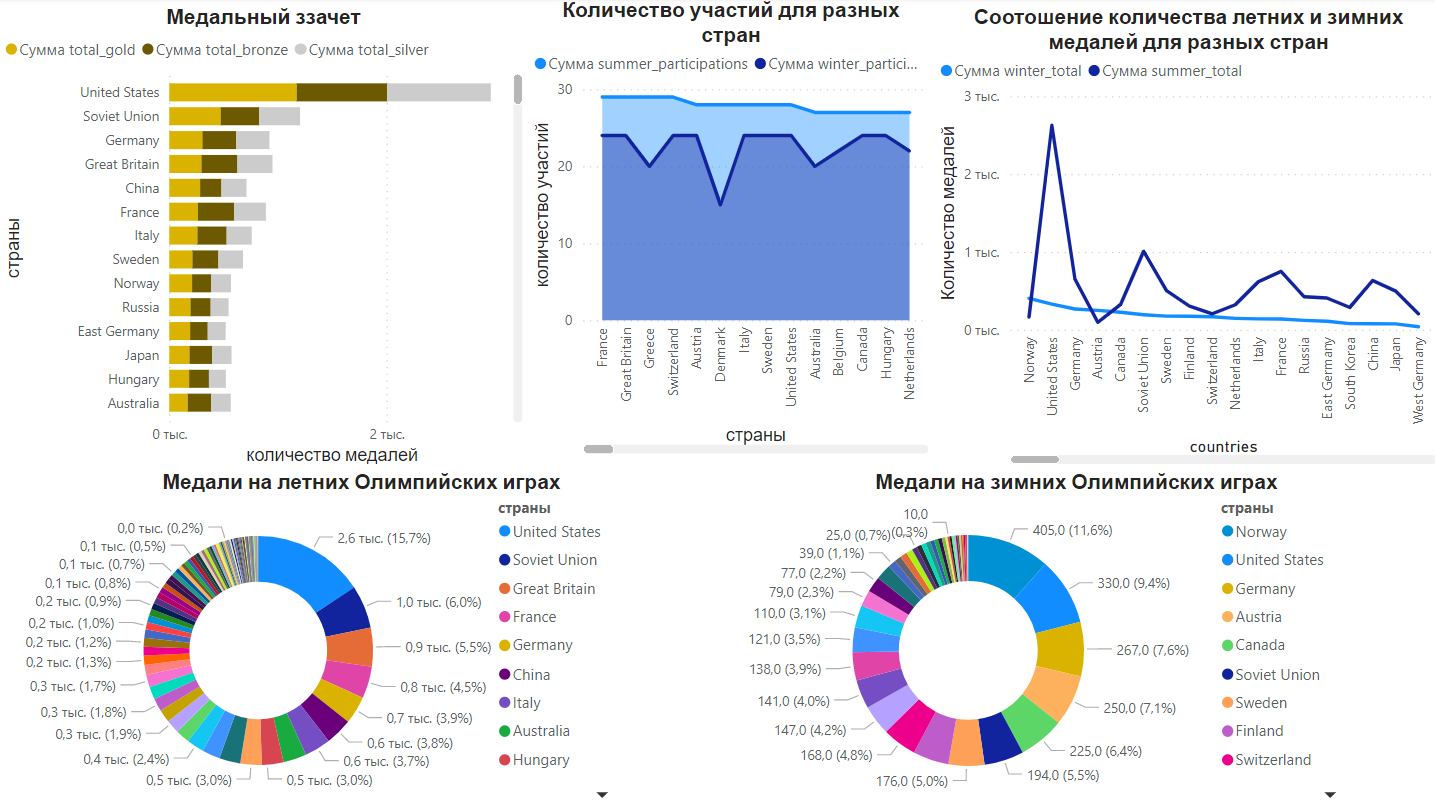


Рисунок 3.6 – Совокупность диаграмм для получения полной информации о результатах Олимпийских игр

Результатом аналитической работы по визуализации данных является возможность комплексно проанализировать результаты Олимпийских игр. Есть возможность выбрать определенную страну и сразу посмотреть всю основную информацию, выделенную на остальных графиках. Помимо этого, данная общая диаграмма дает общее представление о закономерностях, описанных в этой работе

**5 План архивации и восстановления данных**

Целью архивации является подготовка данных для быстрого восстановления в случае необходимости. 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 состояние автомобилей, данные об их пробеге, данные о стоимости машин, а также данные о типе коробки передач. Самыми малозначимыми в рамках анализа являются данные о стране производства и цвете кузова.

Далее выбирается устройство для хранения архивных файлов и тип архивации. Для своей ПО я выбрал комбинированный тип из обычной и разностной архивации, т.е. в понедельник выполняется обычная архивация, а в остальные дни 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.

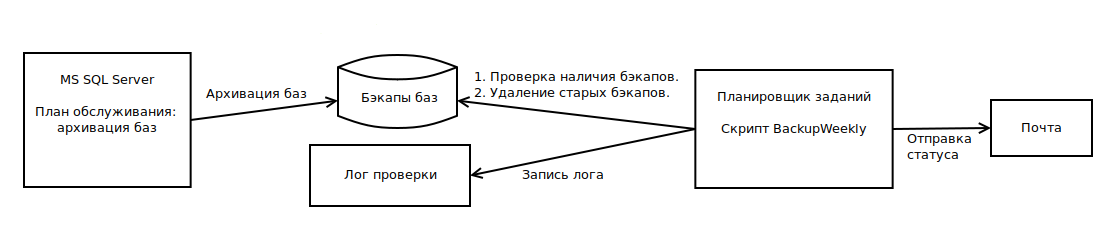


Рисунок 5.1 – Схема архивации данных

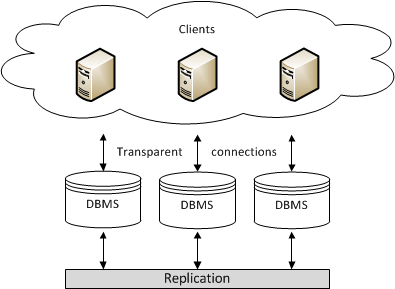


Рисунок 5.2 – Схема синхронной репликации

## **ЗАКЛЮЧЕНИЕ**

В результате выполнения домашнего задания была спроектирована интерактивная система оперативного анализа данных для некоторой информационной системы результатов Олимпийских игр с начала их истории. 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. Официальная документация Tableau: [Электронный ресурс]. // [URL:https://help.tableau.com/ru-RU/cloud-services/Content/Sense\_Helpsites/Home.htm](about:blank) (Дата обращения: 11.11.2022)
4. МГПУ Архивация и восстановление данных: [Электронный ресурс]. // URL: <https://studfile.net/preview/7834724/> (Дата обращения: 11.11.2022)