|  |  |
| --- | --- |
|  | **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: «Used Cars Dataset» – «Продажа поддержанных автомобилей»\_\_\_. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

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

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

Студент \_\_ИУ5Ц-74Б\_\_\_ **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_Папин А. В.\_**\_\_\_

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

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

(Подпись, дата) (И.О.Фамилия)

*Москва, 2023 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)

« \_\_\_\_ » \_\_\_\_\_\_\_\_\_\_\_\_ 2023 г.

**ЗАДАНИЕ**

**на выполнение домашнего задания**

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

по теме \_\_\_\_\_\_\_ «Used Cars Dataset» – «Продажа поддержанных автомобилей»

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

Студент группы \_\_\_\_\_\_ ИУ5Ц-74Б \_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Папин Алексей Владимирович .

(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 по топу продажей поддержанных автомобилей, которые были собраны и получены с сайта cars.com. Произвести визуализацию данных, построить хранилище данных, составить и обосновать план архивации и восстановления хранилища данных.

***Оформление научно-исследовательской работы:***

Расчетно-пояснительная записка на \_17\_ 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 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_ Папин А. В. \_\_\_\_\_**

(Подпись, дата) (И.О.Фамилия)

**СОДЕРЖАНИЕ**

[1 ВВЕДЕНИЕ 4](#_Toc152365386)

[2 Описание предметной области 5](#_Toc152365387)

[3 Описание бизнес-процессов в нотации BPMN 2.0 6](#_Toc152365388)

[4 Хранилище данных 9](#_Toc152365389)

[5 Визуализация данных 10](#_Toc152365390)

[6 План архивации и восстановления данных 14](#_Toc152365391)

[7 ЗАКЛЮЧЕНИЕ 16](#_Toc152365392)

[8 СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 17](#_Toc152365393)

## **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 информационной системы поддержанных автомобилей, собранном из cars.com. Данные были собраны в апреле 2023 года;
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 и нахождение подходящего способа архивации и восстановления хранилища данных.

## **Описание предметной области**

В качестве предметной области был выбран «Used Cars Dataset». Данные были собраны в апреле 2023 года. В данных содержат следующие информации, которые позволяют нам анализировать текущий рынок поддержанных автомобилей, чтобы ориентировать спрос и предложений для увеличения успеха продажи. В наборе данных содержатся:

1. manufacturer – название производителя автомобиля
2. model – название модели автомобиля
3. year – год, когда был выпущен автомобиль
4. mileage – миль, пройденных автомобилем с момента выпуска
5. engine – автомобильный двигатель
6. transmission – тип трансмиссии автомобиля
7. drivetrain – тип трансмиссии автомобиля
8. fuel\_type – тип топлива, которое потребляет автомобиль
9. mpg – количество миль, которое автомобиль может проехать, используя один галлон топлива (мили на галлон)
10. exterior\_color – цвет кузова автомобиля
11. interior\_color – цвет салона автомобиля
12. accidents\_or\_damage – попадал ли автомобиль в АВАРИИ
13. one\_owner – принадлежал ли автомобиль одному лицу
14. personal\_use\_only – использовался ли автомобиль только в личных целях
15. seller\_name – имя продавца
16. seller\_rating – рейтинг продавца
17. driver\_rating – рейтинг автомобиля, данный водителями
18. driver\_reviews\_num – количество отзывов об автомобилях, оставленных водителями
19. price\_drop – снижение цены по сравнению с начальной ценой
20. price – цена автомобиля

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

1. Автомобили с более низким пробегом имеют более высокую цену.
2. Автомобили с более новыми годами выпуска стоят дороже
3. Рейтинг продавца и рейтинг водителей связаны с ценой автомобиля.
4. Автомобили с определенным типом топлива имеют более высокую эффективность топливопотребления.
5. Автомобили с определенным типом трансмиссии имеют более высокую эффективность топливопотребления.
6. Автомобили с определенными цветами кузова стоят дороже.
7. Автомобили с определенными цветами салона стоят дороже.
8. Рейтинг продавца коррелирует с рейтингом автомобиля, данным водителями.
9. Средняя цена на автомобили меняется с течением времени.

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

**2.1 Бизнес-процесс «Приобрести поддержанного автомобиля в онлайн-сервисе покупателем»**

В бизнес-процессе на Рисунок 1 представлена схема желание продажи автомобилей. Для этого пользователя необходимо открыть онлайн-сервис, который предоставляет возможность продавать поддержанных и новых автомобилей. После открытия онлайн-сервиса и регистрации по нему, пользователя следует выбрать объявление, которое ему понравилось, и сообщить продавцу. Продавцу приходит уведомление от клиента, и он начинает брать данные с БД этого автомобиля, чтобы проверить наличие на складе и оценить уровень повреждения, а также учитывать желания клиента: цвет и характеристики. Клиент, как правило, будет начать торг, когда у автомобиля присутствует повреждения. Если желание клиента не достигнуто, то диалог закончен и машина не передается клиенту. В противном случае, начинается оформление на заказ машины, и передается данные, и если детальные информации человека соответствуют корректности, то документ будет подписан и передается оформленные документы продавцам и клиентам, после чего клиент может получить машину. В противном случае документы отправляются на повторную проверку и исправление.

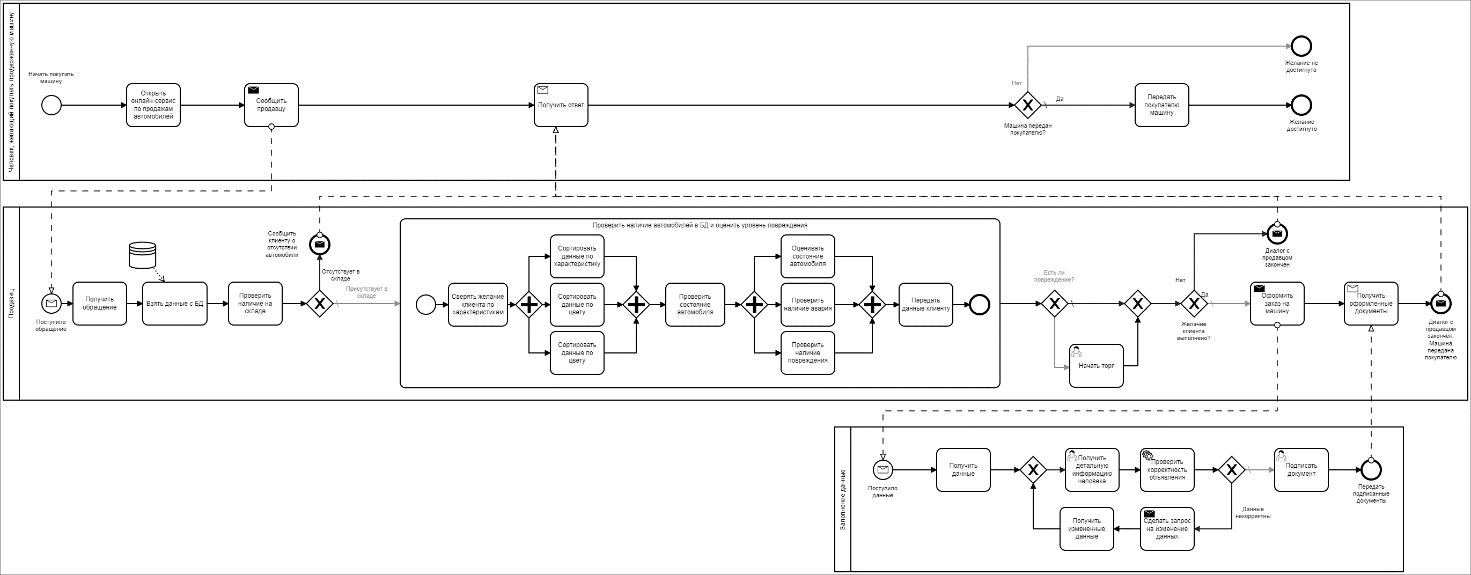


Рисунок 1 – Схема покупки поддержанного автомобиля пользователем в нотации BPMN 2.0

**2.2 Бизнес-процесс «Выставление и продажи поддержанных автомобилей в онлайн-сервис продавцом»**

В бизнес-процессе на Рисунок 2 представлена схема выставление и продажи поддержанных автомобилей в онлайн-сервис продавцом. Пользователь, который решил выставить свою автомобиль в продажах, может открыть онлайн-сервис, отвечающий за купли-продажи автомобилей. Для этого ему необходимо сначала открыть объявление о продажах и передать данные о автомобиле. Как только данные поступили в онлайн-сервисе, они проходят по нескольких этапов: получение детальной информации человека, проверки корректности объявления, в котором содержат ряд проверок: заполнение данных по характеристику, цвету, оценки текущей состоянии автомобиля, наличия авария и повреждения. После вышеперечисленных этапов данные фиксируются, и если они корректны, то сохраняется в БД и оформляется документы, которые подпишет продавец. В противном случае продавцу придется заново подать запрос на изменение данных. После успешного этапа, передает уведомление продавцу о успешного выставления автомобиля в продажах. Также продавцу предоставляется возможность воспользоваться услугу эвакуатора для перевозки автомобиля. Как только автомобиль будет доставлен в пункт назначении, то ему сообщит о успешном задании после чего объявление станет актуальным.

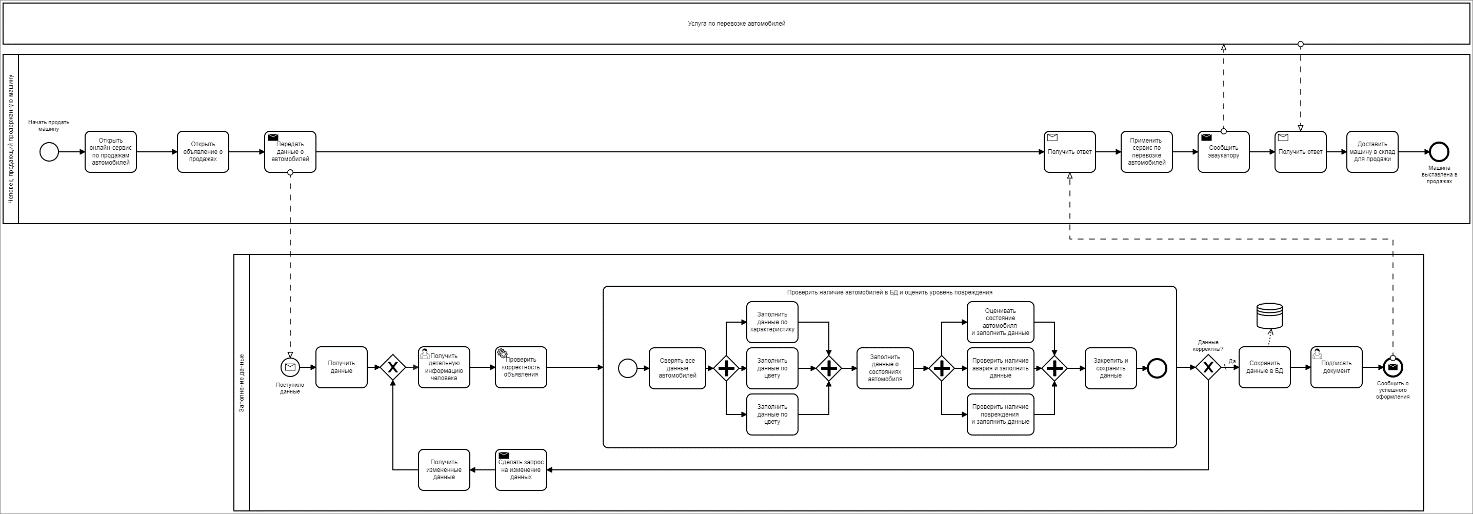


Рисунок 2 – Выставление и продажи поддержанных автомобилей в онлайн-сервис продавцом

## **Хранилище данных**

Рисунок 3 – Схема хранилища данных

## **Визуализация данных**

* 1. **Количество производимых автомобилей по производителям**

Данный график на Рисунок 4 иллюстрирует список моделей, который по порядку убывания лидирует по количеству производимых и в объявлении о продажах. Сразу заметно, что в США преобладает свои марки: Ford и Chervolet, но и японские – Toyota.

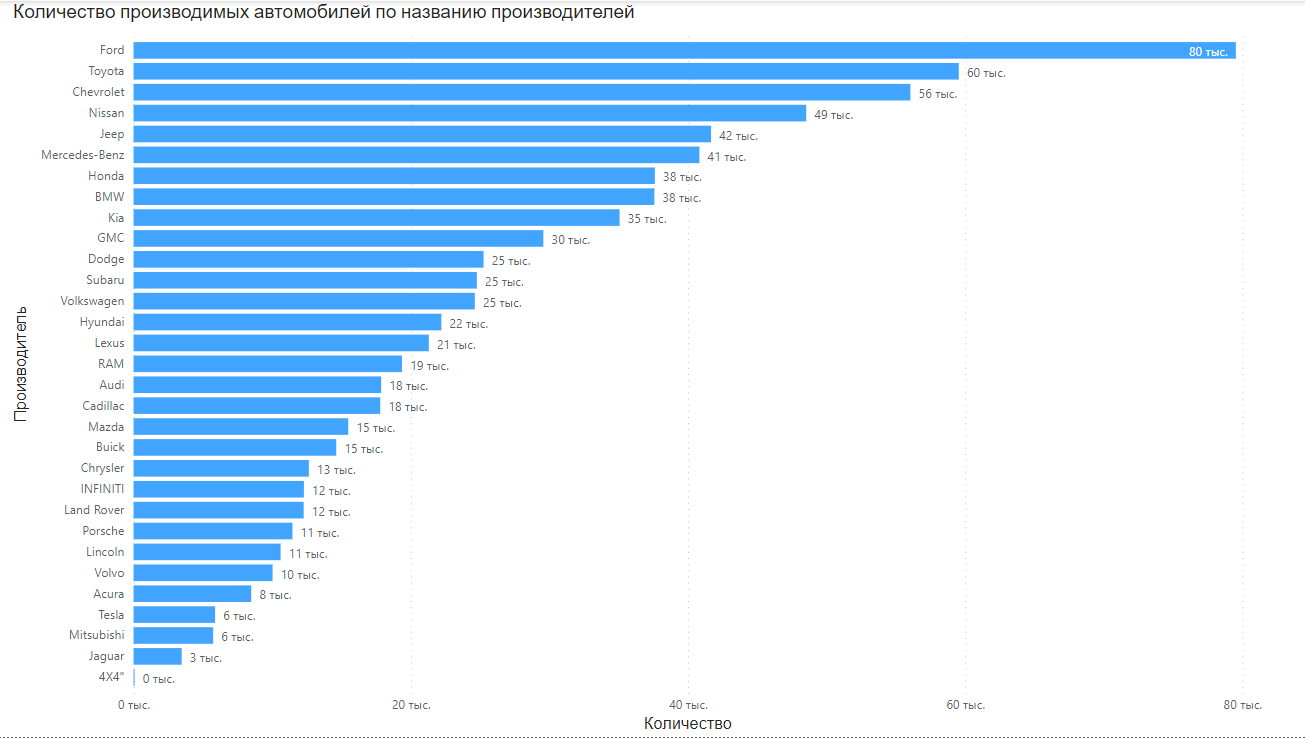


Рисунок 4 – Количество производимых автомобилей по производителям

* 1. **Распространенный тип топлива**

Данный график на Рисунок 5 иллюстрирует распространенный тип топлива, которые пользуются в США. В первое место попал – бензиновый тип, а второе – гибридный, третье – дизельный. Очевидно, что бензин лидирует на рынке.

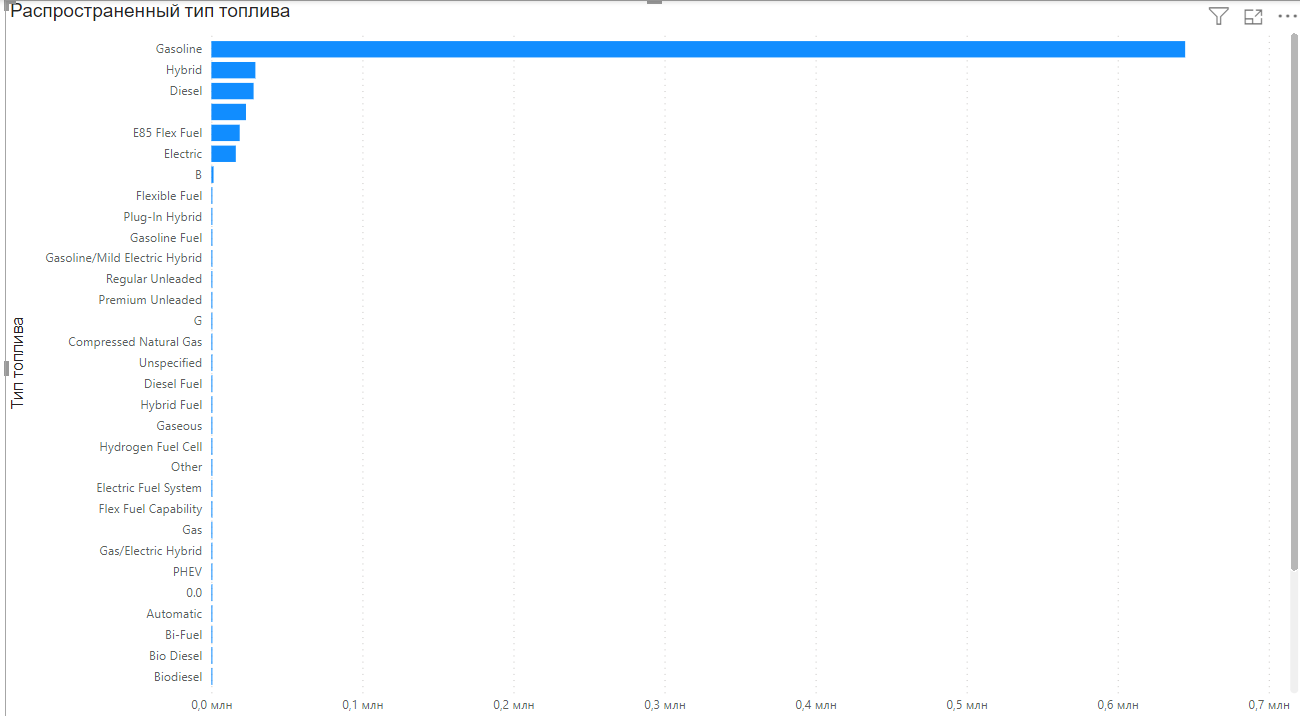


Рисунок 5 – Распространенный тип топлива

* 1. **Количество автомобилей с типом трансмиссии**

Данный график на Рисунок 6 иллюстрирует типов трансмиссии, которые часто можем встретиться в объявлении продажи. По этому графику можно сделать вывод, что в США предпочитают АКПП.

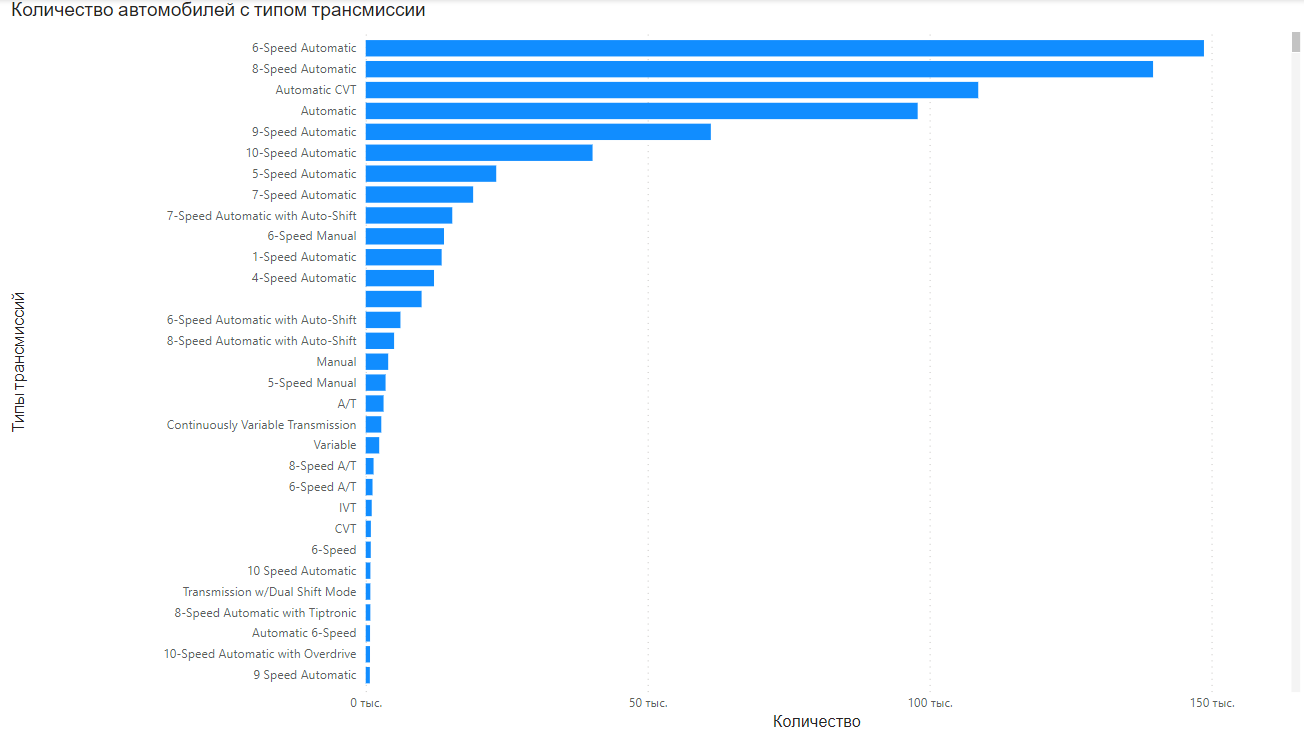


Рисунок 6 – Количество автомобилей с типом трансмиссии

* 1. **Количество новых и старых автомобилей**

Данный график на Рисунок 7 иллюстрирует график, который отображается количество новых и старых автомобилей за протяженный год с 1920 по 2023. По графику можно сделать вывод, что автомобили, выпускаемые недавно, чаще выставляют в объявлениях нежели раритетных.

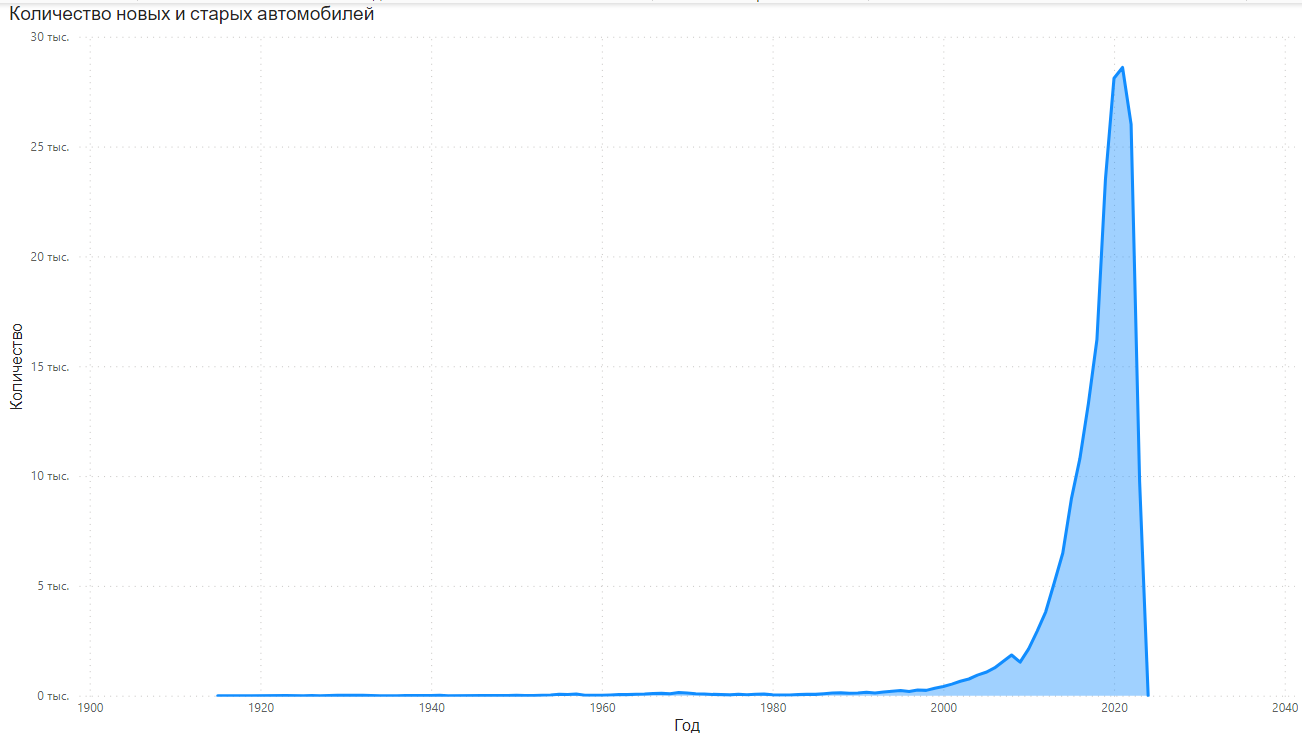


Рисунок 7 – Количество новых и старых автомобилей

* 1. **Зависимость цены от рейтинга продавца**

Данный график на Рисунок 8 показывает зависимость цены от рейтинга продавца. Чем выше оценку ставит продавец, тем высока вероятность того, что автомобиль дорогим получится.

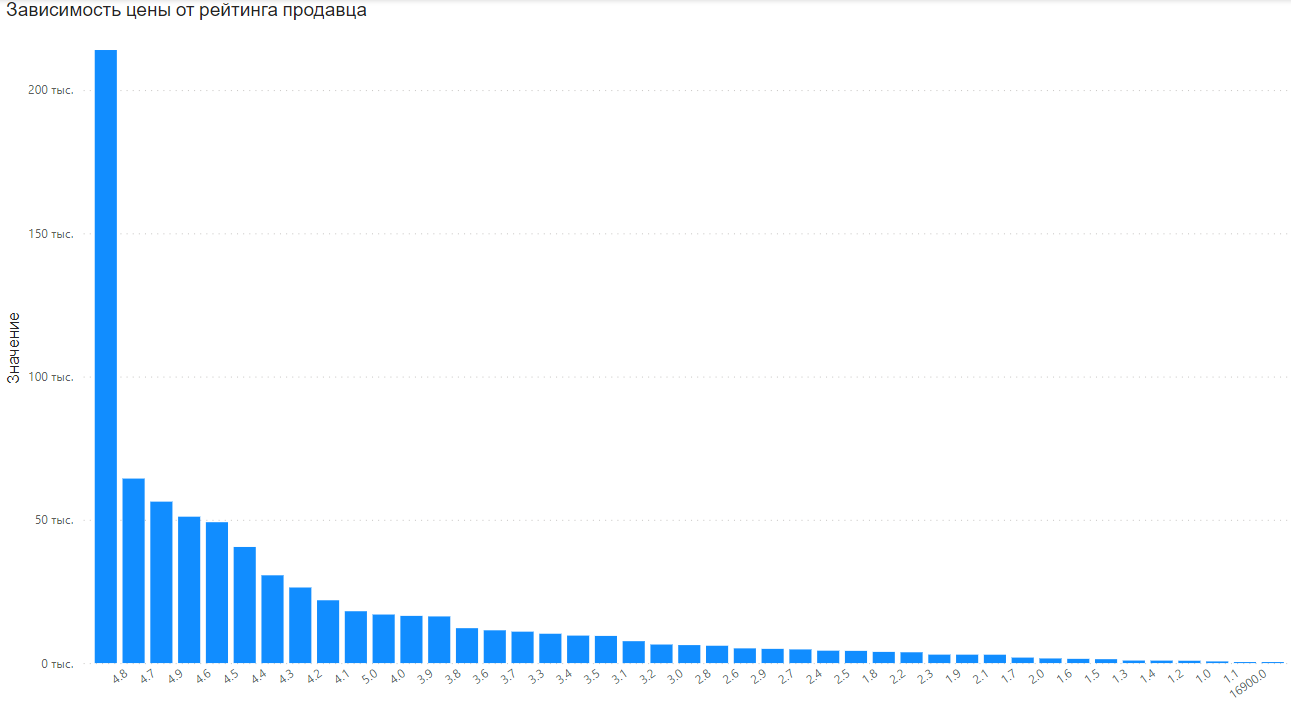


Рисунок 8 – Зависимость цены от рейтинга продавца

* 1. **Зависимость цены от рейтинга водителей**

Данный график на Рисунок 9 показывает зависимость цены от рейтинга водителей. Чем выше оценку ставит продавец, тем высока вероятность того, что автомобиль дорогим получится.

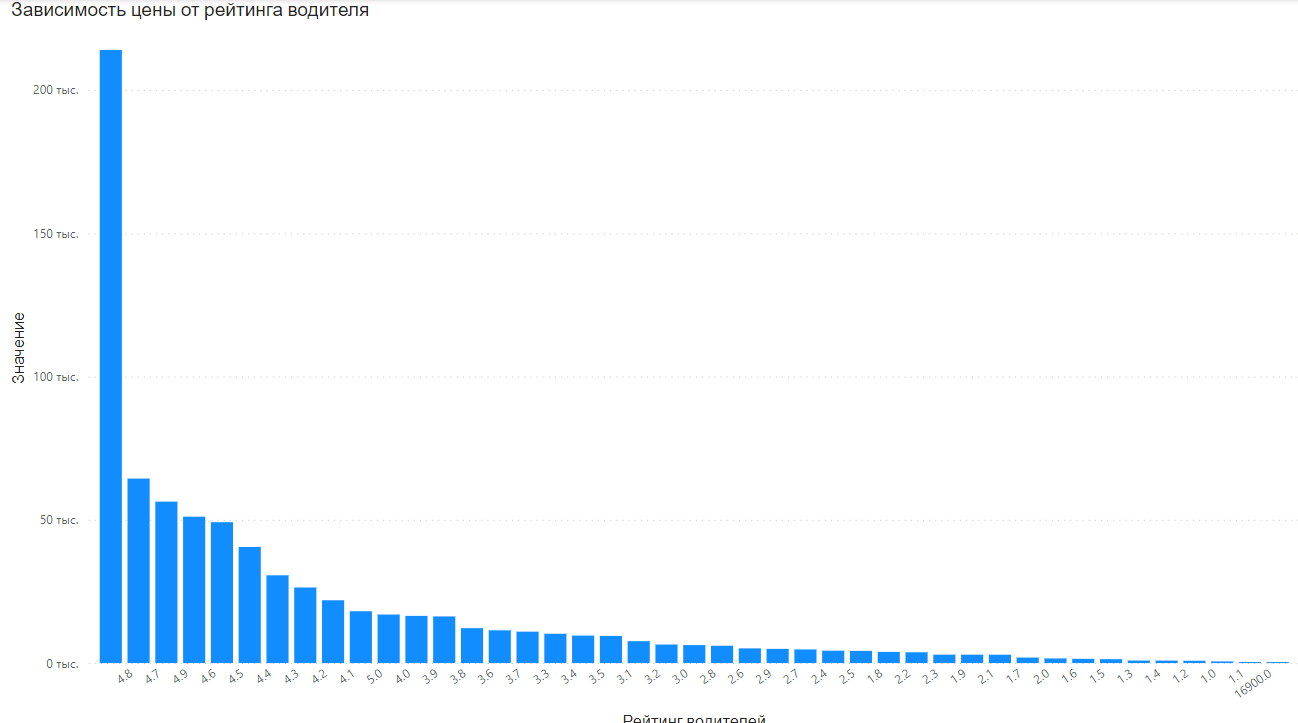


Рисунок 9 – Зависимость цены от рейтинга водителей

## **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 названия треков, количество прослушиваний, артисты, исполнившие треки, наивысшая позиция в чарте, количество недель в чарте, популярность, жанр треков, а также их длительность. Самыми малозначимыми в рамках анализа являются дата релиза, темп и громкость треков.

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.

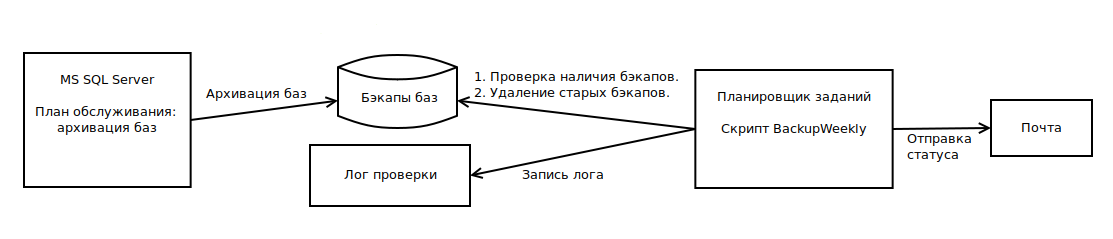


Рисунок 10 – Схема архивации данных

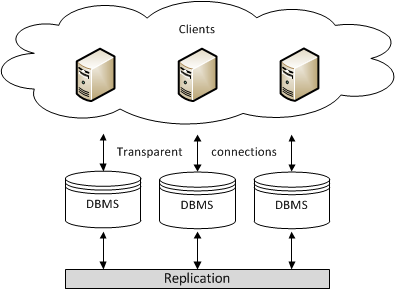


Рисунок 11 – Схема синхронной репликации

## **ЗАКЛЮЧЕНИЕ**

В результате выполнения домашнего задания была спроектирована интерактивная система оперативного анализа данных для информационной системы «Used Cars Dataset». 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 Qlik Sense: [Электронный ресурс]. // <URL:https://help.qlik.com/ru-RU/cloud-services/Content/Sense_Helpsites/Home.htm> (Дата обращения: 11.11.2022)
4. МГПУ Архивация и восстановление данных: [Электронный ресурс]. // URL: <https://studfile.net/preview/7834724/> (Дата обращения: 11.11.2022)