**Документация**

**Дипломная работа по теме:**

**“Анализ розничных продаж (поиск инсайтов и составление рекомендаций стейкхолдерам)”**

**Профессия “Аналитик данных”, DA-77**

**Мальцева Александра Николаевна**

**г. Томск, 2024**

**Оглавление**

**Введение…………………………………………………………………………….3**

**Блок 1. Описание исходного датасета и типов данных….…4**

**Блок 2. Подготовка и преобразование данных….……………...5**

**Блок 3. Аномалии в данных……..…………………………………………...5**

**Блок 4. Анализ данных для стейкхолдеров………………………..7**

**4.1 Анализ основных категорий товаров.………………..……...7**

**4.2 Анализ посетителей………………………………….……….…….…8**

**4.3 Динамика продаж по каждому году……………………….…10**

**4.4 Анализ продаж по каждому торговому центру………..12**

**Выводы и рекомендации……………………………………….…….......14**

**Введение.**

Основная цель любого бизнеса - получение прибыли. Изучение покупательского спроса позволит выявить наиболее популярные группы товаров и сосредоточить внимание на увеличении объема продаж этих групп товаров, что способствует повышению общей прибыли.

**Круг стейкхолдеров:**

* Руководитель сети магазинов
* СЕО
* Директор(управляющий) магазином
* Руководитель отдела продаж
* Руководитель отдела маркетинга(ответственный специалист)
* Ответственный технический специалист. (Для прояснения вопроса по данным за 2023 год.)

**Ключевые метрики проекта.**

* Сумма среднего чека
* Объем продаж (год/месяц) общий и по группам товаров
* Выручка (год/месяц)
* Средняя цена на товара по категориям и по месяцам
* Спрос по категориям товаров в течение года
* Наиболее/наименее прибыльные группы товаров по выручке

**Блок 1. Описание исходного датасета и типов данных**

Определите шаги преобразования и очистки данных.

Источник дата-сета: [www.kaggle.com](https://www.kaggle.com/datasets/mehmettahiraslan/customer-shopping-dataset)

Дата-сет представляет собой данные о продажах в Торговых центрах Турции(если быть более конкретной - Стамбула) за 2021, 2022, 2023 года

Дата-сет содержит 99457 уникальных записей о розничных продажах.

Дата-сет содержит поля:

| **№** | **Имя Столбца** | **Описание** | **Тип данных** |
| --- | --- | --- | --- |
| ***1*** | **invoice\_no** | номер покупки | object |
| ***2*** | **customer\_id** | id покупателя | object |
| ***3*** | **gender** | пол покупателя | object |
| ***4*** | **age** | возраст покупателя | int64 |
| ***5*** | **category** | категория товаров | object |
| ***6*** | **quantity** | количество товара | int64 |
| ***7*** | **price** | цена за единицу товара | float64 |
| ***8*** | **payment\_method** | метод оплаты (наличные дебетовая карта, кредитная карта) | object |
| ***9*** | **invoice\_date** | дата покупки | object |
| ***10*** | **shopping\_mall** | название ТЦ | object |

Пропуски в данных отсутствуют.

**Блок 2. Подготовка и преобразование данных**

Столбец 'invoice\_date' приведен к типу данных date-time.

Добавлены столбцы: Для оценки выручки мной был добавлен столбец **‘sum\_of\_purchace’** - сформирован путём перемножения данных из столбцов ‘**quantity’ и ‘price’. Необходим для оценки выручки**

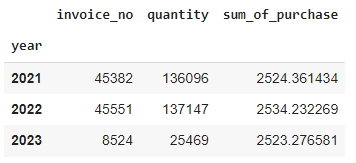
**‘month-year’** - выделен из столбца invoice\_date, содержит только месяц и год, когда совершена покупка.

**‘year’** - выделен из столбца invoice\_date, содержит только год для удобства группировки данных по году.

Сформированы отдельно датафреймы с данными о продажах за каждый год(2021,2022,2023).

**Блок 3. Аномалии в данных**

В ходе подготовки данных выявлено значительное снижение показателей за 2023 год относительно предыдущих лет:



Количество покупателей, объем продаж значительно снизились.

Предполагаемые причины снижения оборотов в 2023 году:

* Показаны только продажи в физических магазинах, не показаны онлайн продажи.
* Снижение количества туристов.
* Проблемы с поставками товаров.
* Магазин был закрыт по техническим и т.д. причинам.

На этапе проверки гипотезы о закрытии магазинов выявлено, что начиная с марта данные о продажах предоставлены за 1-3 число каждого месяца.



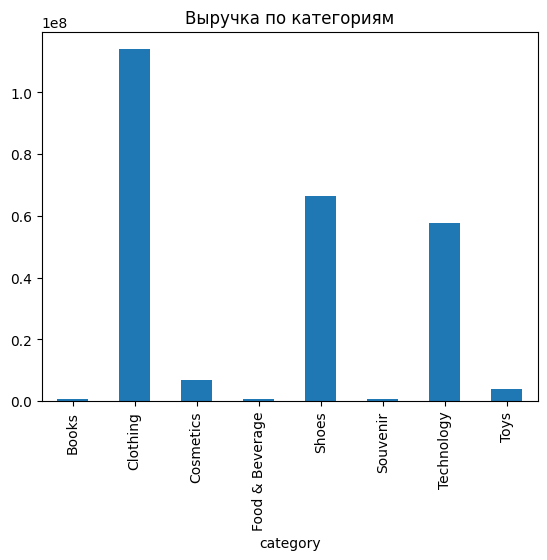
Поскольку вероятность того, что 10 крупных торговых центров 10 месяцев в году работали по 2-3 дня в месяц стремится к 0, вероятнее всего данные были утеряны в результате технического сбоя.

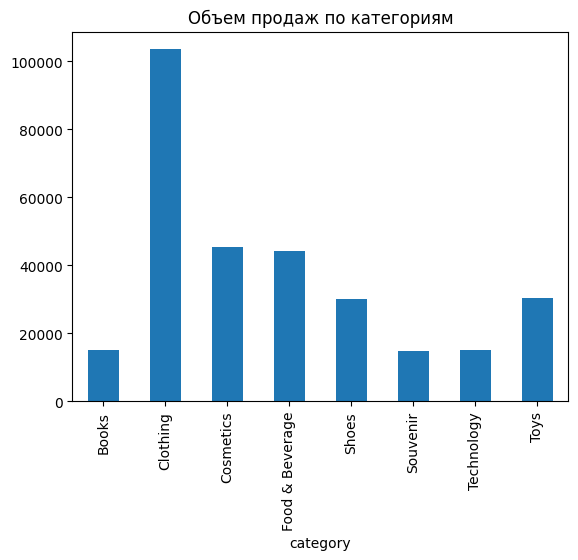
Я приняла решение оставить данные за полные месяцы - январь и февраль, а оставшиеся данные считать данными за март.

**Блок 4. Анализ данных для стейкхолдеров**

**4.1. Анализ основных категорий товаров.**

По объемам продаж и по принесенной прибыли:





А так же в сравнении по годам:



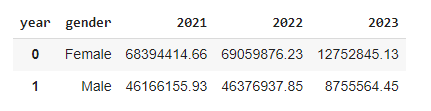
Основную прибыль заказчик получает от продажи одежды, обуви, техники. Также хочу отметить высокий объем продаж косметики, пищевой продукции и напитков, а также игрушек.

**4.2 Анализ посетителей**

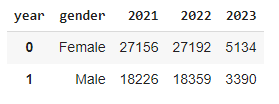
Был проведен анализ посетителей магазинов по полу и возрасту. Женщины больше ходят в Торговые центры заказчика и больше тратят.

Соотношение мужчины/женщины:

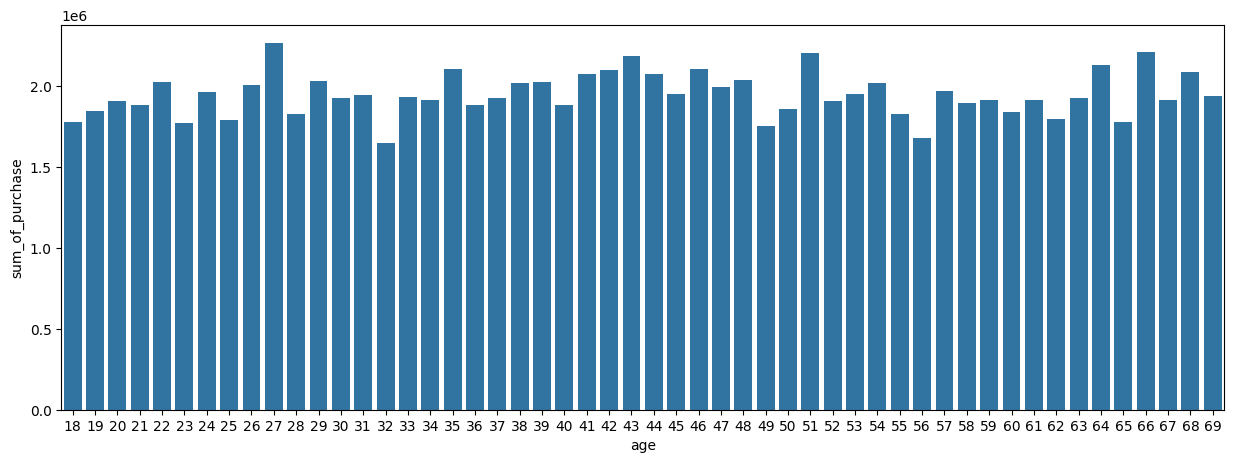
* по прибыли:



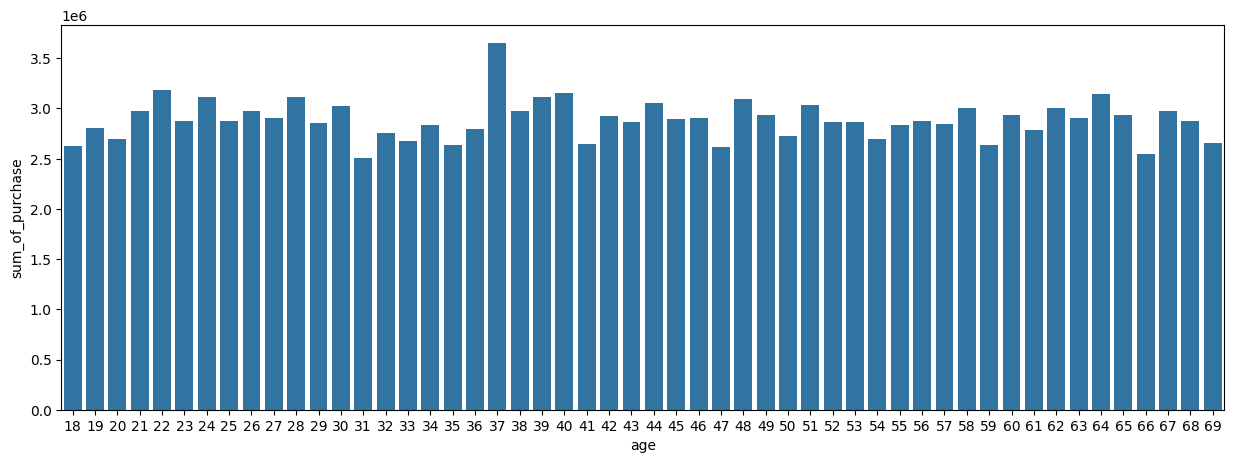
* по количеству посетителей:

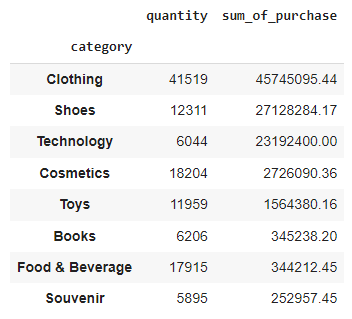


* возрастной состав мужчины:



* возрастной состав женщины:



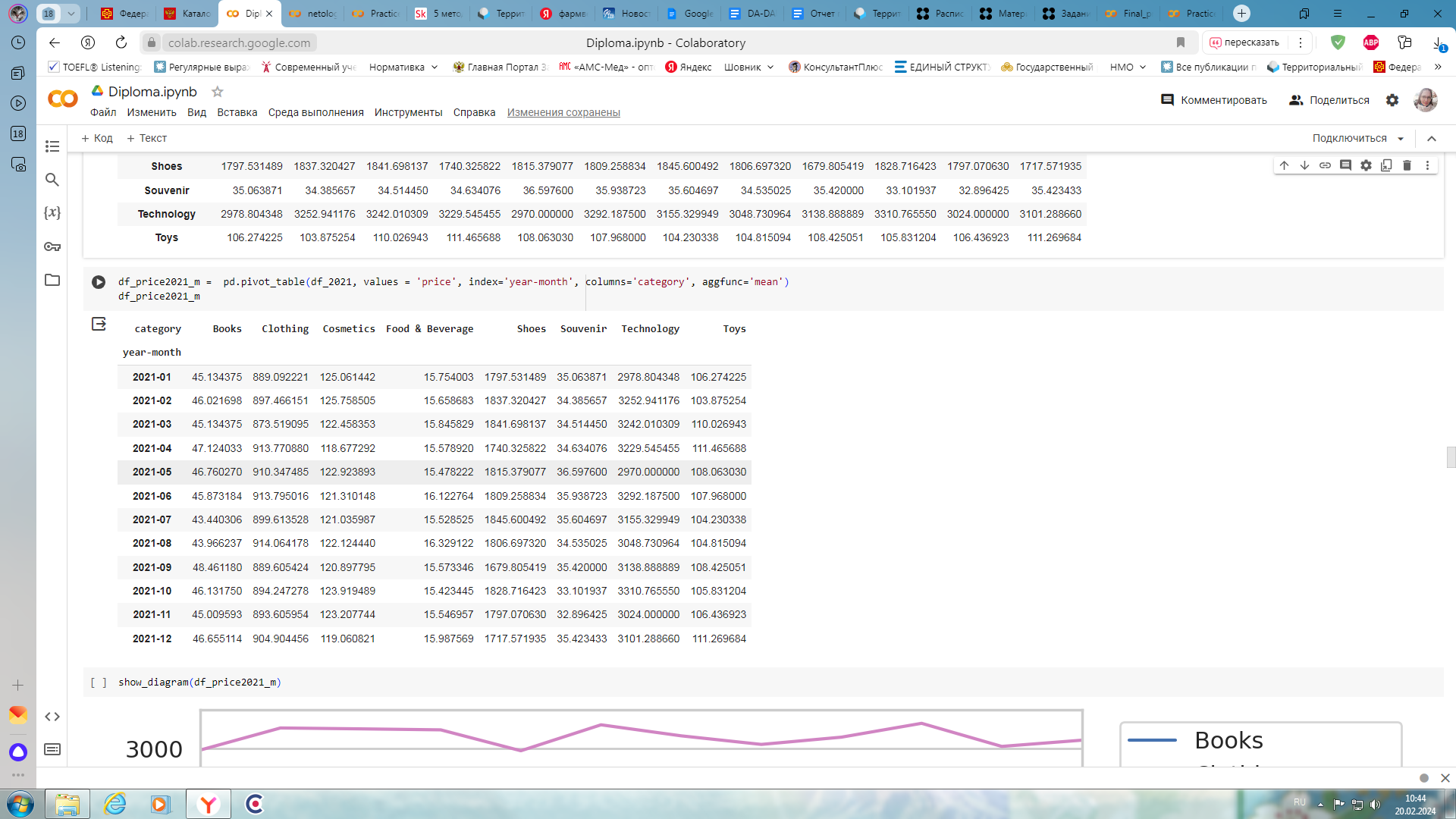
Наиболее прибыльными группами товаров так же остаются одежда, обувь, техника как среди мужчин, так и среди женщин: 



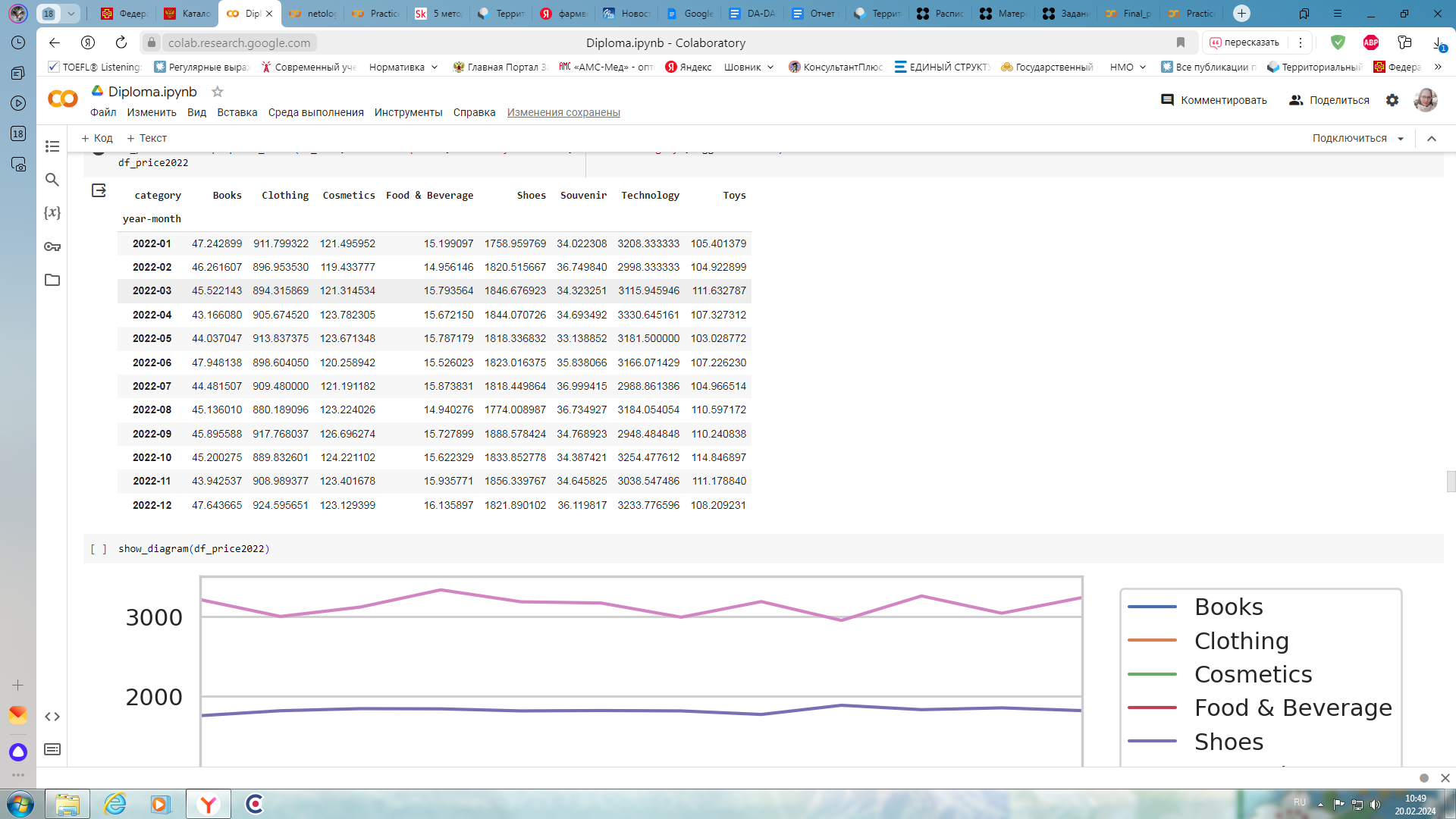
Вывод. Женщины среднего возраста приносят больше всего прибыли, меньше всего прибыли приносят молодые мужчины.

**4.3. Динамика показателей по каждому году.**

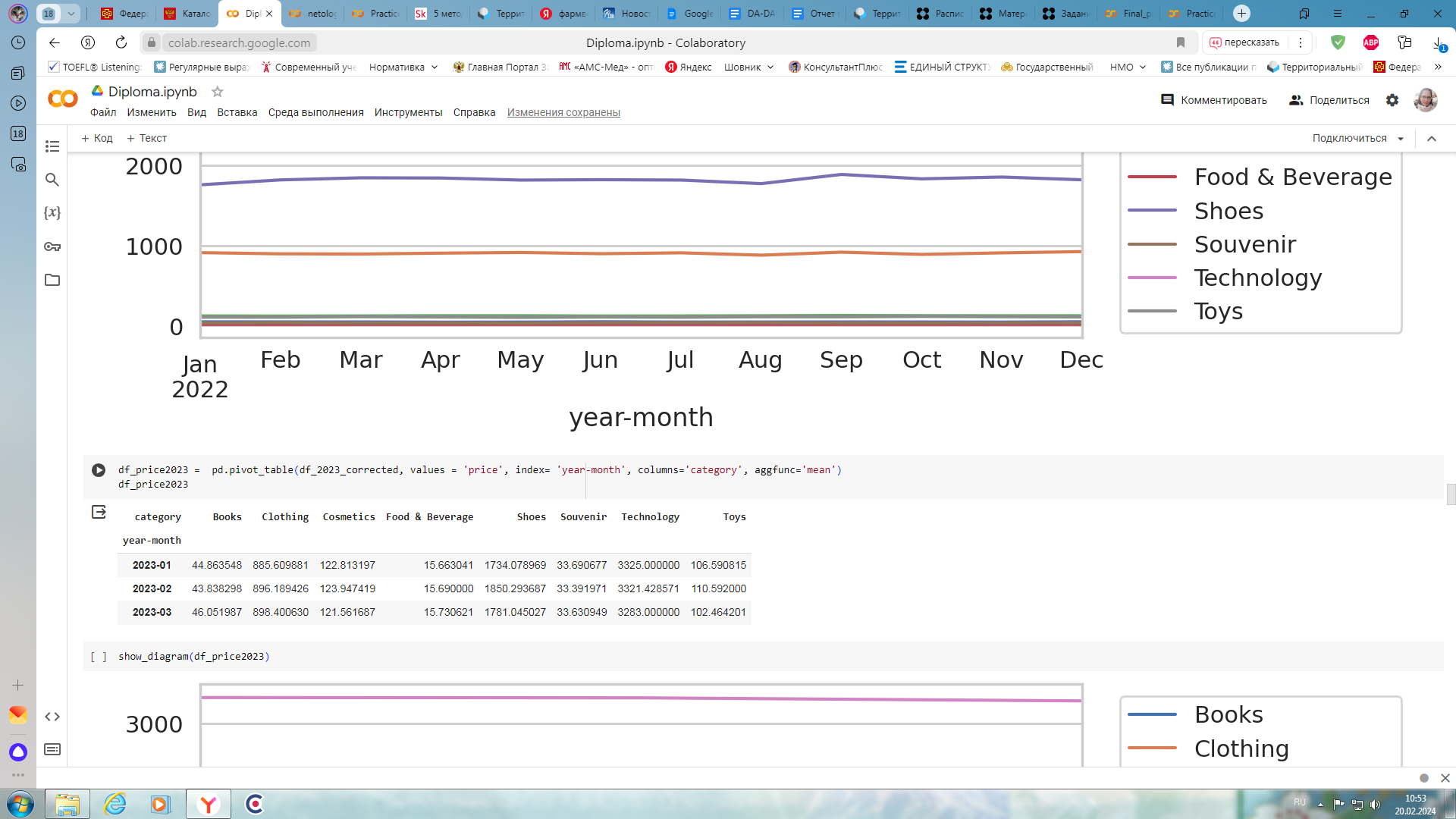
Далее был проведен анализ отдельно по годам по колебанию средней цены на товар:

2021 год:

2022:



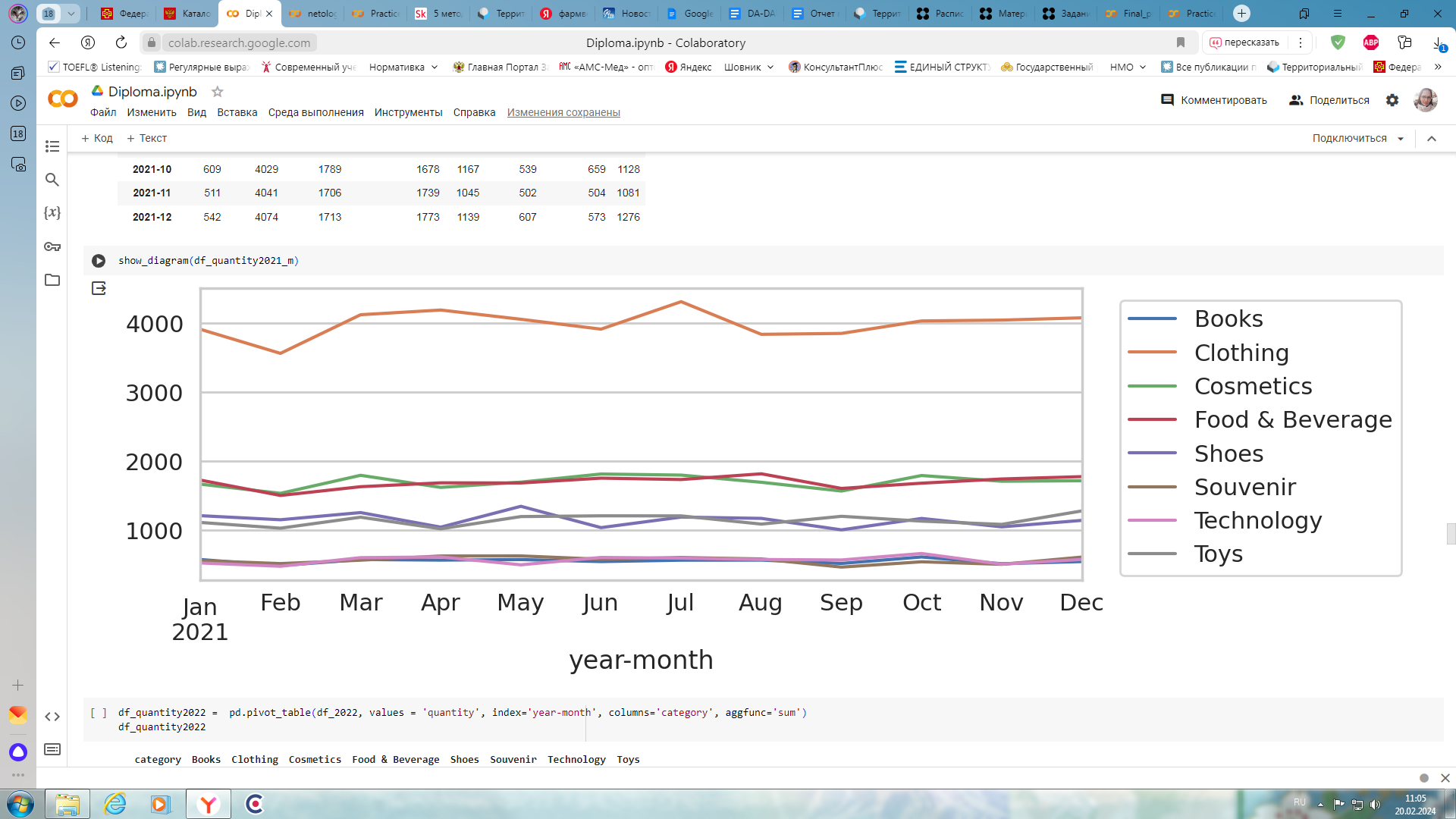
2023:



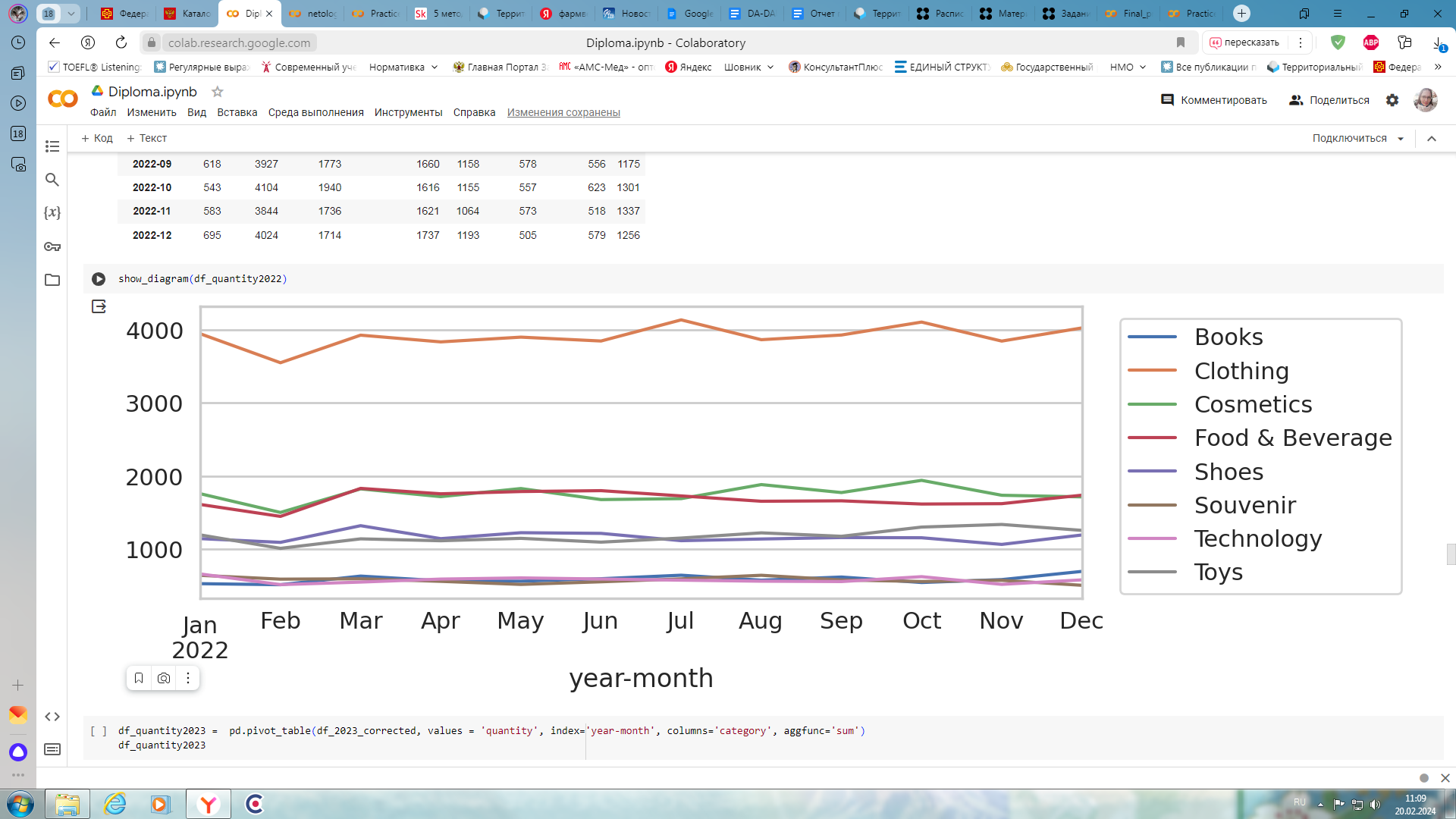
Вывод. Существенные колебания средней цены покупки наблюдаются по категории товаров одежда. Для более подробного анализа необходимо дополнить датасет конкретными наименованиями купленного товара.

Динамика по объемам продаж:

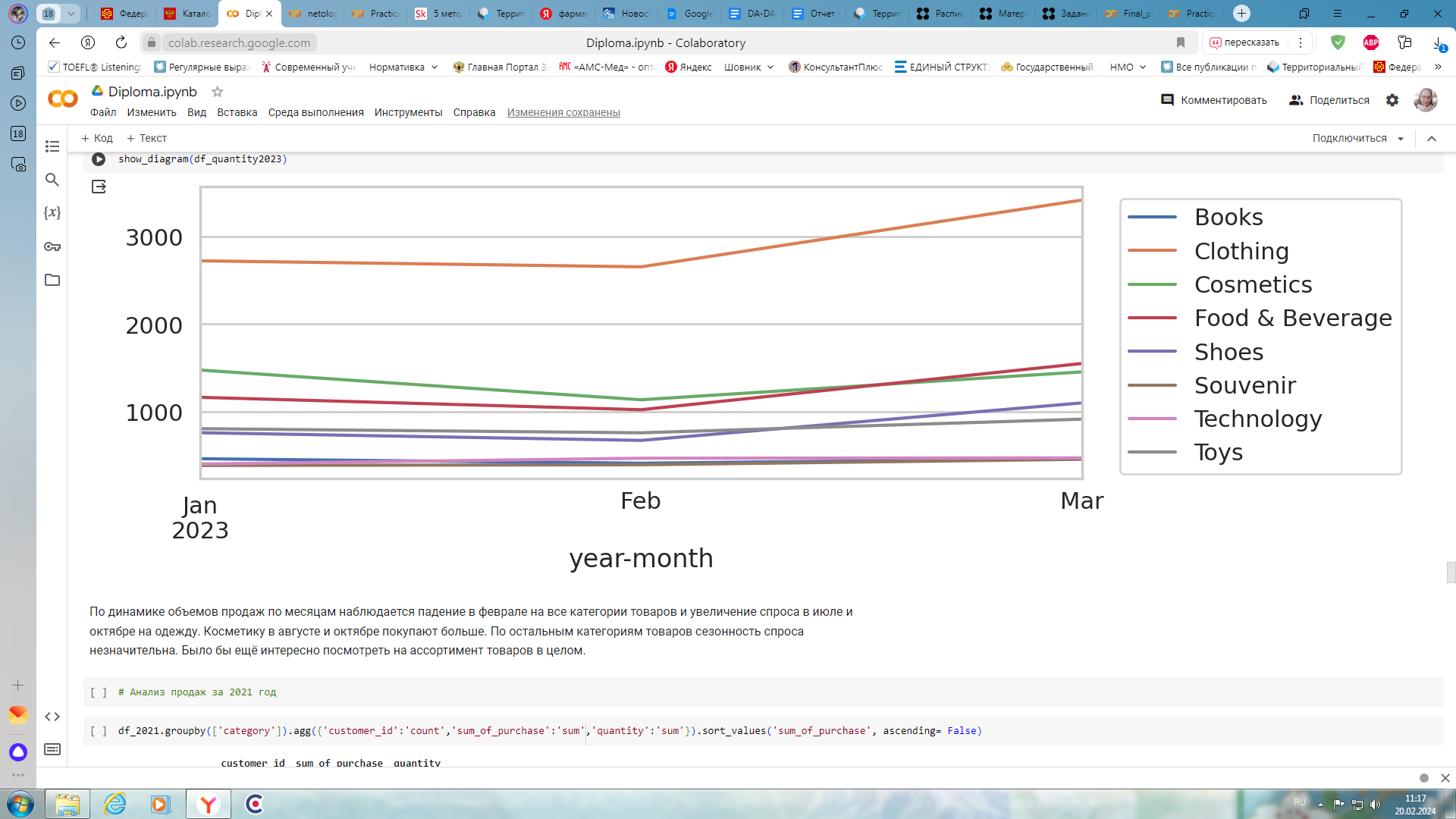
2021:



2022:



2023:

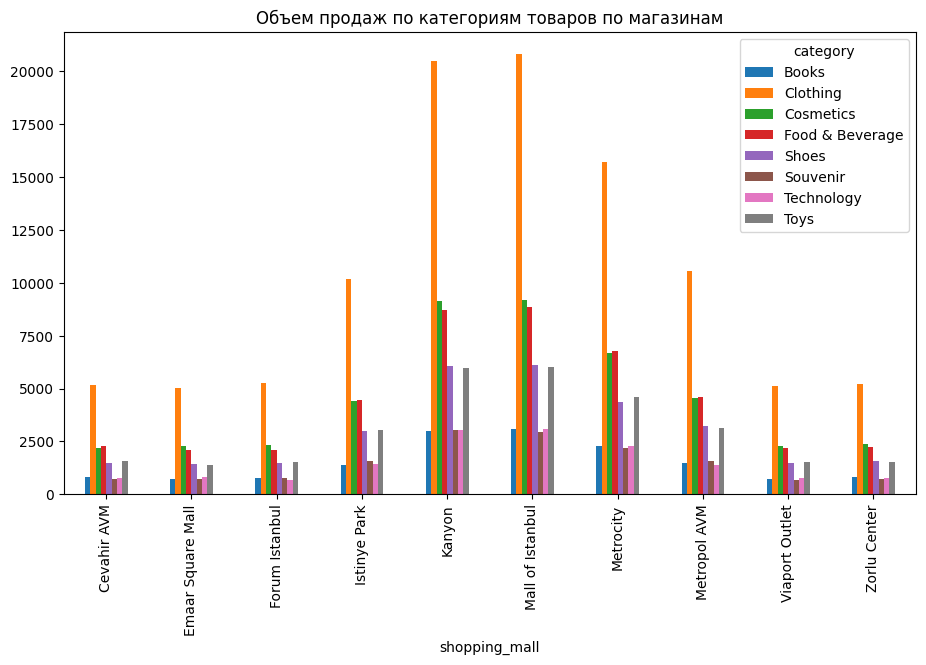


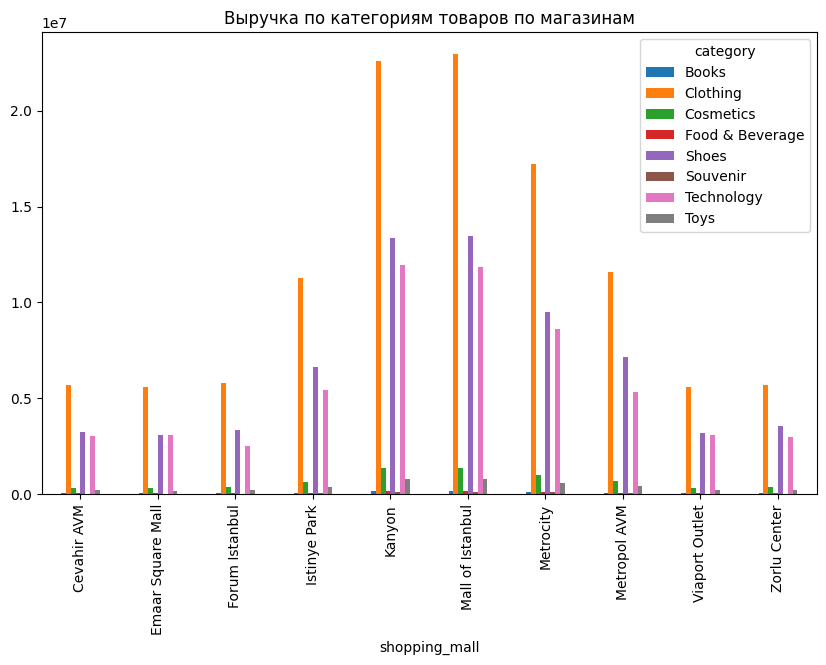
По динамике объемов продаж по месяцам наблюдается падение в феврале на все категории товаров и увеличение спроса в июле и октябре на одежду. Косметику в августе и октябре покупают больше. По остальным категориям товаров сезонность спроса незначительна. Было бы ещё интересно посмотреть на ассортимент товаров в целом.

Падение спроса в феврале может быть связано со спадом туристического сезона. Более подробное изучение сезонности товаров позволит увеличить продажи и прибыль.

**4.4. Анализ продаж по каждому ТЦ.**

Кроме того, был проведен анализ продаж в торговых центрах

по объемам продаж по категориям товаров по прибыли по категориям товаров:



Из графиков видно, что наиболее популярными торговыми центрами являются: **Kanyon и Mall of Istanbul, на 2м месте Metrocity на 3м MetropolAVM и Istinye Park.** По объему продаж тенденция сохраняется и по каждому магазину в отдельности. Лидируют одежда, косметика, продукты.

**Выводы и рекомендации**

Большую прибыль заказчику приносят женщин среднего возраста

Меньшую - молодые мужчины. В планировании маркетинговой стратегии, рекламных акций это стоит учитывать.

Поскольку объемы продаж групп товаров косметика и игрушки довольно высокие, я бы предложила более подробно проанализировать спрос по ним. Например по косметике средний ценник за 3 года почти не менялся.(в ТЦ нет товаров другой ценовой категории или есть, но они не пользуются спросом)

По имеющимся в продаже позициям косметики я предлагаю провести сравнение цен с конкурентами, возможно стоит повышать цены. Так же, возможно, стоит пересмотреть ассортимент и закупать более дорогую косметику. Например, закупить 1-2 линейки и провести A/B тестирование по результатам продаж через месяц и через 3 месяца, если результат будет положительным, далее развиваться в этом направлении. Аналогично поступить с группой товаров игрушки.(только если бизнес-заказчик не рассматривает игрушки как сопутствующий товар)

В целом, средний чек за 3 года менялся незначительно, я бы рекомендовала провести исследование и сравнить цены с конкурентами, после чего принимать решение оставлять цены как есть или повышать. Либо действовать через расширение ассортимента, и увеличение позиций в чеке за счет допродаж или акций (товар дня, 1+1, увеличенный кэшбэк на определенную группу товаров).

Дата-сет не позволяет сравнить ассортимент более подробно не только по группам товаров, но и по-позиционно. Поэтому, предлагаю пересмотреть ассортимент на предмет актуальности и соответствия трендам.

Среди 10 магазинов по прибыли и объемам продаж выделяются 2. Для привлечения клиентов в другие ТЦ, я бы предложила сделать рекламные акции на покупку в определенном ТЦ.

А также выяснить по какой причине данные о продажах сохранялись не корректно, дабы не допустить такого в будущем.

Не удалось выявить сезонность по группам товаров. Что касается сезонности, можно сказать, что продажи снижаются в феврале(что может быть связано с поставками) и увеличиваются в июле и августе.(что скорее связано с туристическим сезоном).

**Рекомендации по дополнению дата-сета:**

1. Количество товара на складе.(остаток) Позволит оценивать популярность товара и его оборачиваемость.
2. Кроме категории товара добавить наименование товара, более мелко раздробить категории товаров. Например одежда: верхняя одежда,футболки, брюки, джинсы, нижнее белье. Это позволит проводить более глубокий анализ сезонности спроса, контролировать продажи и закупать более популярные товары своевременно.

1. Наименование поставщика. Позволит отслеживать наиболее добросовестных поставщиков(не задерживают сроки поставки, меньше процент возвратов). А также отслеживать выгодные ценовые предложения
2. Закупочная цена. Позволит отслеживать изменения в цене товара, вести переговоры с поставщиком о цене следующей поставки или о возврате товара
3. Дата поставки и/или срок годности(для товаров с ограниченным сроком использования). Позволит отслеживать партии товара и вовремя реализовывать товар с подходящим сроком, а также планировать акции и другие маркетинговые решения

Подводя итог рекомендациям, хочу отметить, что на мой взгляд наиболее приоритетным направлением для заказчика будет изучение цен конкурентов и более подробное изучение товаров. Более подробному изучению будет способствовать расширение датасета, поэтому дополнение дата-сета будет следующей приоритетной задачей.