Ссылка на блокнот Google Colab: <https://colab.research.google.com/drive/1t6N01llDAqKfV18T_1Dx64MaHnQgcUvD?usp=sharing>

Ссылка на отчёт в Power BI: https://app.powerbi.com/view?r=eyJrIjoiMGIwYzg0MTEtOWQwNS00MjA2LTgzZDUtYzJlZTRhMzFmNTdkIiwidCI6ImI2MjkxZjY5LWNiNGYtNDU0ZS1hMTEwLWM0YjAxODI4MTAyMyIsImMiOjEwfQ%3D%3D

**Описание данных**

**Таблица customers**: Содержит данные о клиентах, включая контактную информацию, адреса и уникальные идентификаторы.

Таблица **customerProfile**: Описывает профили клиентов с информацией о семейном положении, доходе, образовании, возрасте и других характеристиках.

Таблица **products**: Содержит сведения о продуктах (велосипедах и велосипедной экипировки), их категории, характеристиках и стандартной стоимости.

Таблица **promotions**: Информация о маркетинговых акциях, включая тип и категорию.

Таблица **manufacturers**: Перечень производителей с их уникальными идентификаторами.

Таблица **manufacturer details**: Дополнительные детали о производителях, такие как группировка и тип рынка.

Таблица **data**: Основная таблица транзакций с финансовыми показателями, связями с клиентами, продуктами, акциями и другими параметрами.

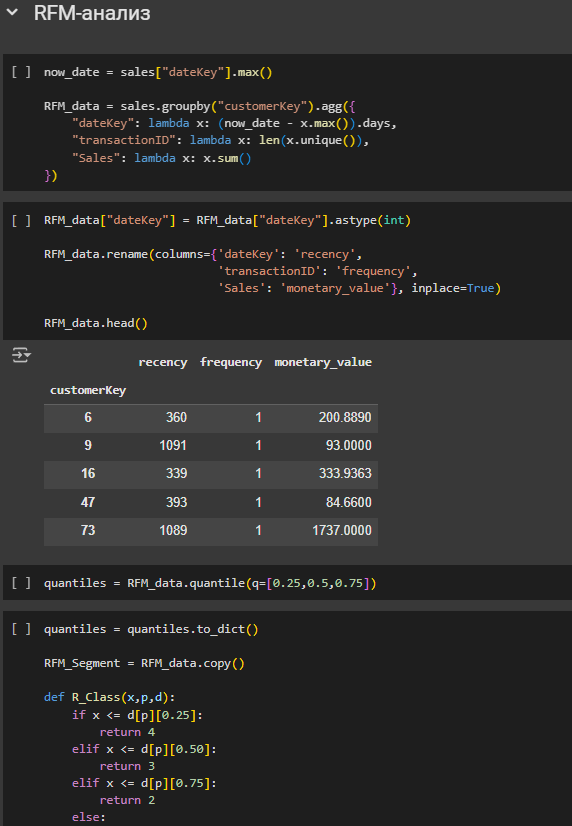
Задача: Провести предварительную обработку данных, их анализ с помощью инструментов Data Science, сравнение полученных результатов и визуализация основных выводов для дальнейшего принятия решения специалистами в данной области.

Цель: Показать практическое применение своих навыков

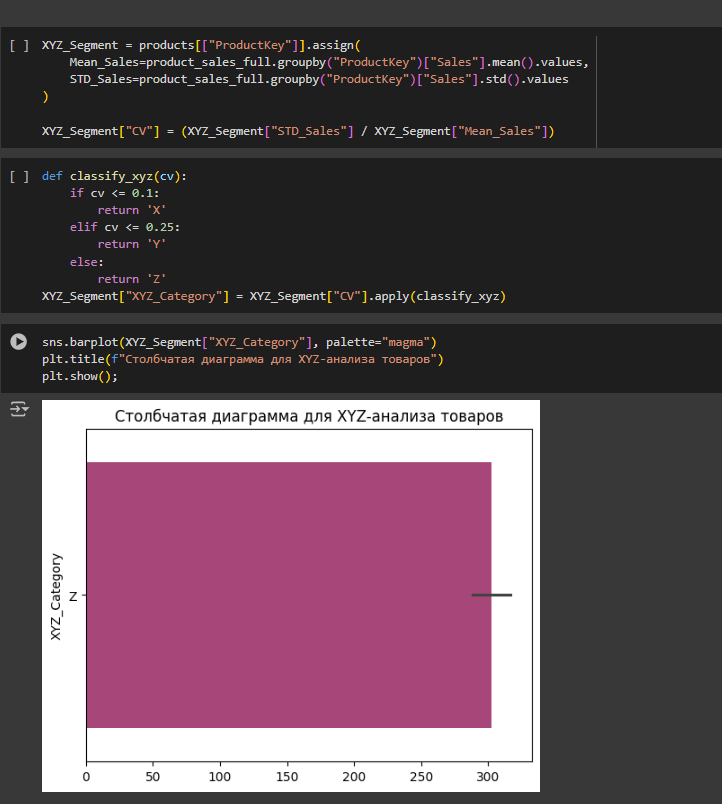
**Анализ в Google Colab**

На данном этапе я провожу первоначальное изучение данных, включая проверку на наличие пропущенных значений, дубликатов и соответствие данных их типам. Выполняю консолидацию таблиц, исследую данные с помощью группировок, построения визуализаций и расчёта описательных статистик. Также проверяю данные на наличие выбросов и анализирую их влияние. Формулирую задачу по обогащению данных за счёт проведения ABC-XYZ и RFM анализов.

**Насыщение данных**

При помощи библиотек Python (Pandas и Numpy):





Добавляю эти данные в качестве дополнительных столбцов к существующей

таблице и продолжаю работу с ней уже в блокноте Google Colab. Удаляю лишние

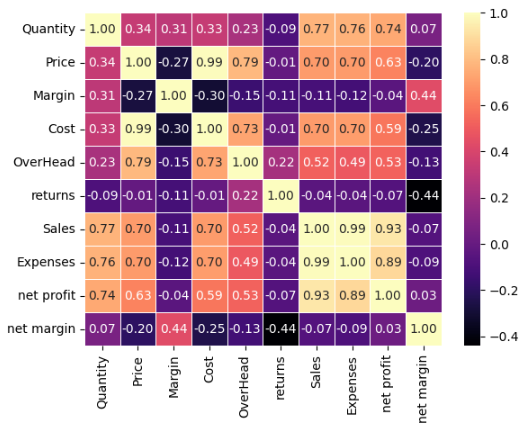
столбцы.

По результатам анализа ABC‑XYZ выявлено 3 категорий: AZ, BZ, CZ.

Данные категории вынесу в отдельный признак для дальнейшего анализа в Power BI.

По данным о продажах можно выделить низкую дисперсию у числовых показателей (продажи, прибыль, маржа, расходы и т. д.)

Корреляция (Пирсона) данных о продажах:

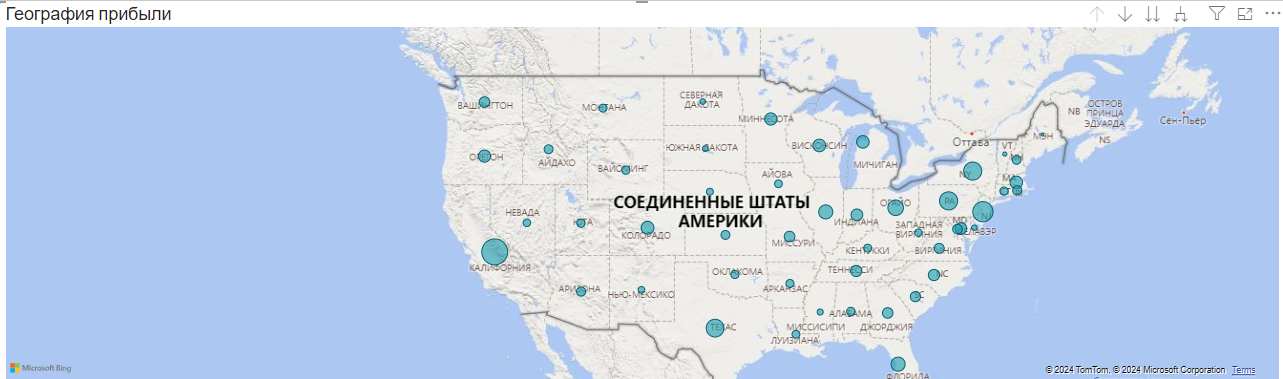


Выделить можно следующие пары с высокой корреляцией: Цена – Себестоимость, Продажи – Расходы, Продажи – Чистая прибыль, Расходы – Чистая прибыль

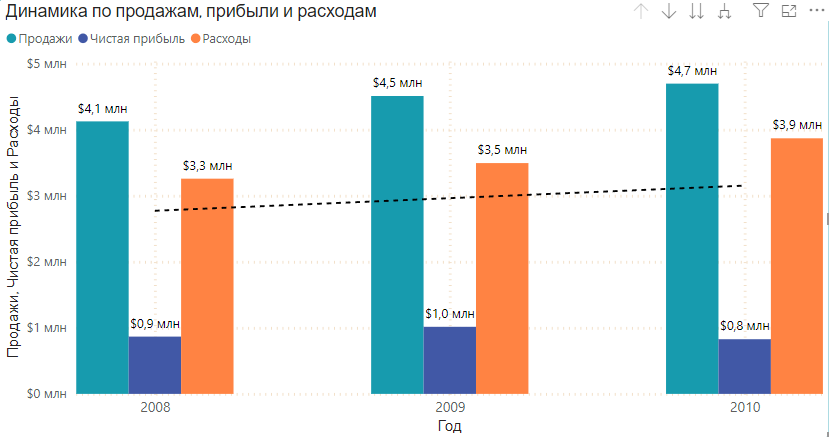
**Анализ в Power BI**

Добавляю новые меры:

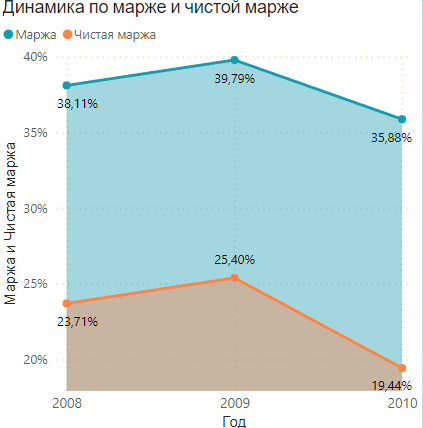
1. Расходы Всего = Расходы + Накладные расходы
2. Средний чек = Продажи / Кол-во проданных товаров
3. Процент возвратов = Возвраты / (Купленные товары + Возвраты)

****

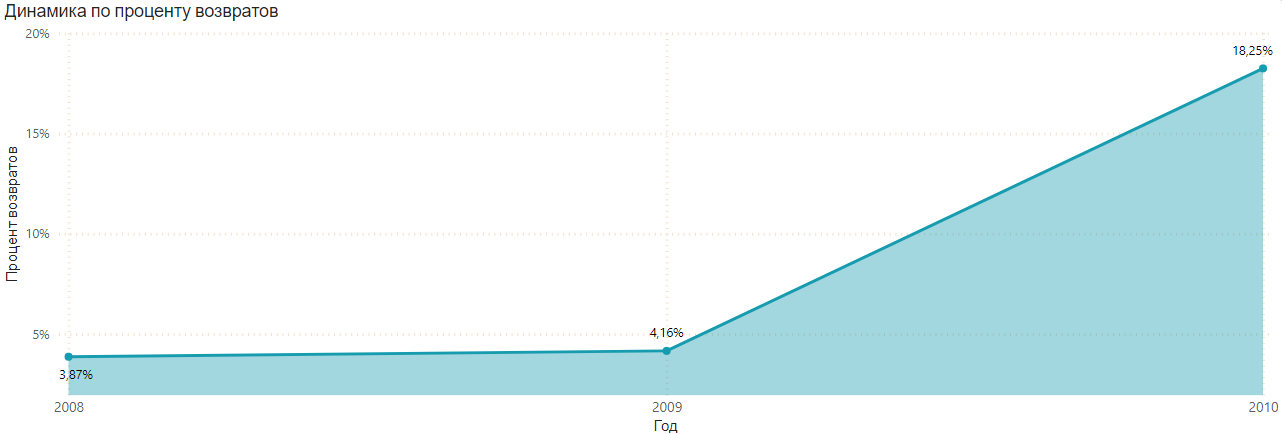
Больше всего прибыли принесли следующие штаты: Калифорния (12,4%), Нью Йорк (7%), Нью Джерси (5,4%), Пенсильвания (5,5%)



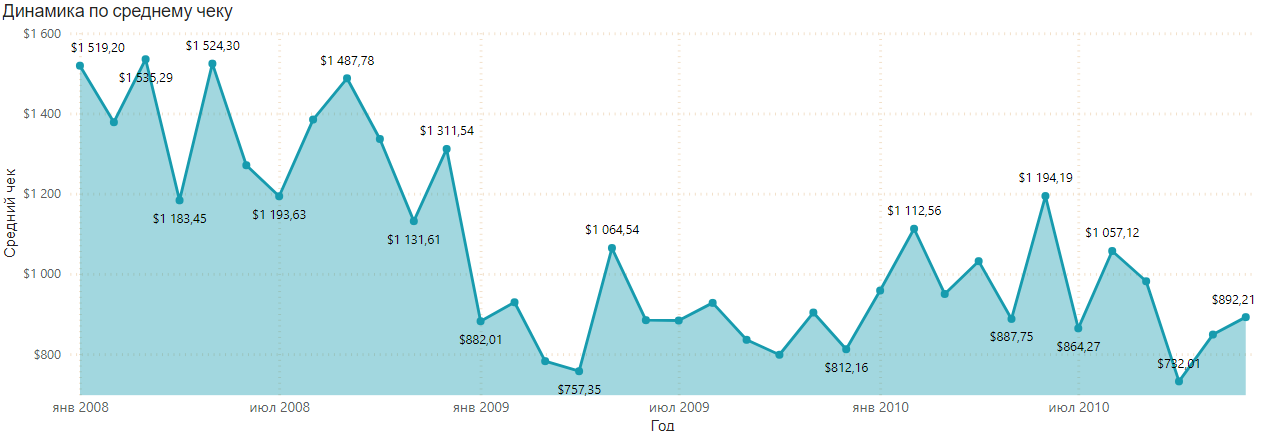
Динамика продаж за 2008-2010 года почти не менялась; линия тренда почти перпендикулярна оси X. Прибыль, несмотря на поднятие продаж (на 0,2 млн $), уменьшилась в 2010 по сравнению с 2009 (на 0,2 млн $).

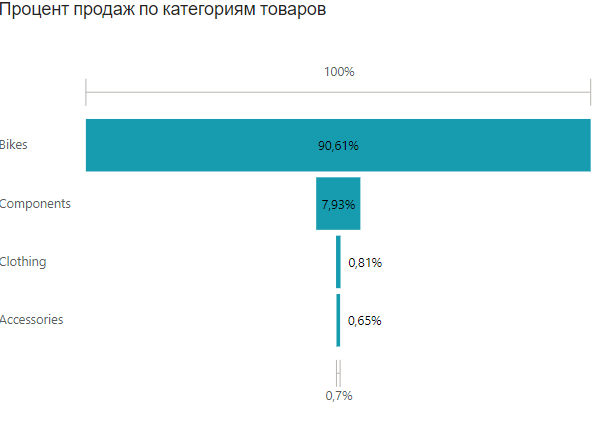


Динамика маржинальности имеет переменный характер. В 2010 году большое снижение показателя чистой маржи почти на 6% по сравнению с 2009 годом.

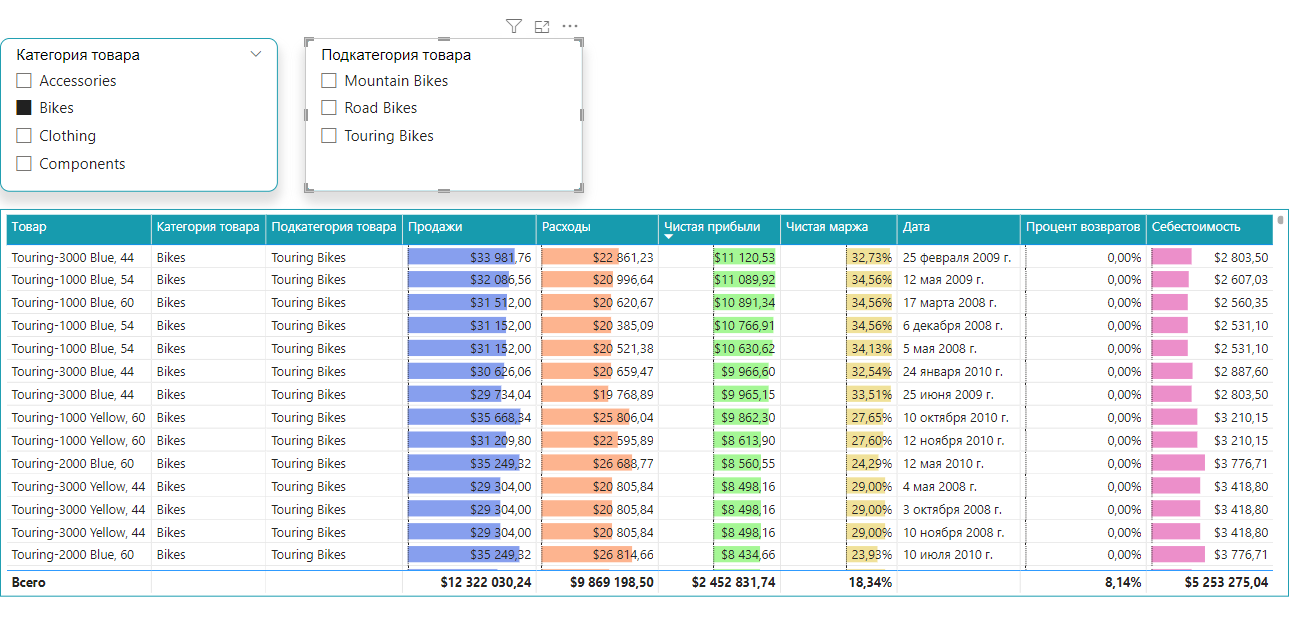


Динамика процента возвратов имеет значительный рост за 2010 год в сравнение с 2009 годом (на 14%).





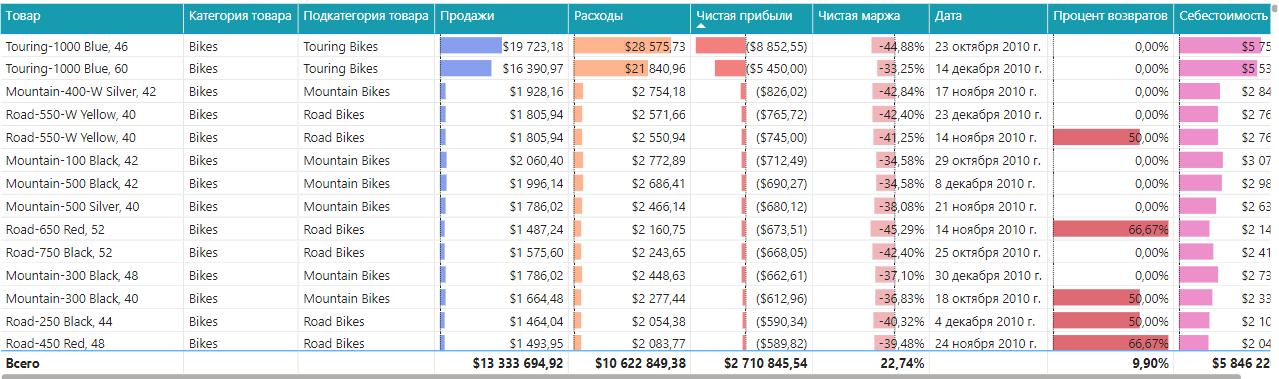
2008 год – лучший год по среднему чеку, 2009 год – худший. Причиной является доля продаж, приходящаяся на велосипеды (самая дорогая категория товаров продавалась чаще в 2008 году). Это же и описывает положение показателя среднего чека в 2010 году.



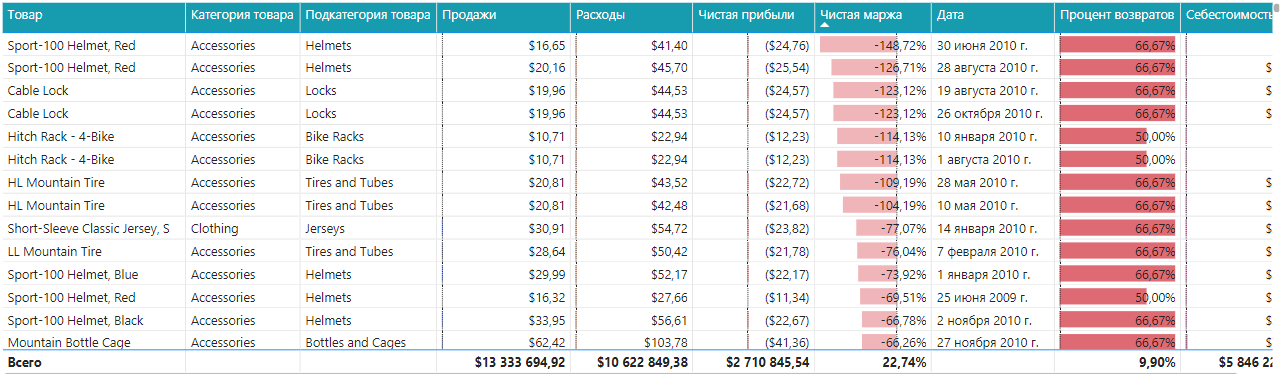
Всего велосипеды принесли почти 2,5 млн $ чистой прибыли.

Есть три категории велосипедов:

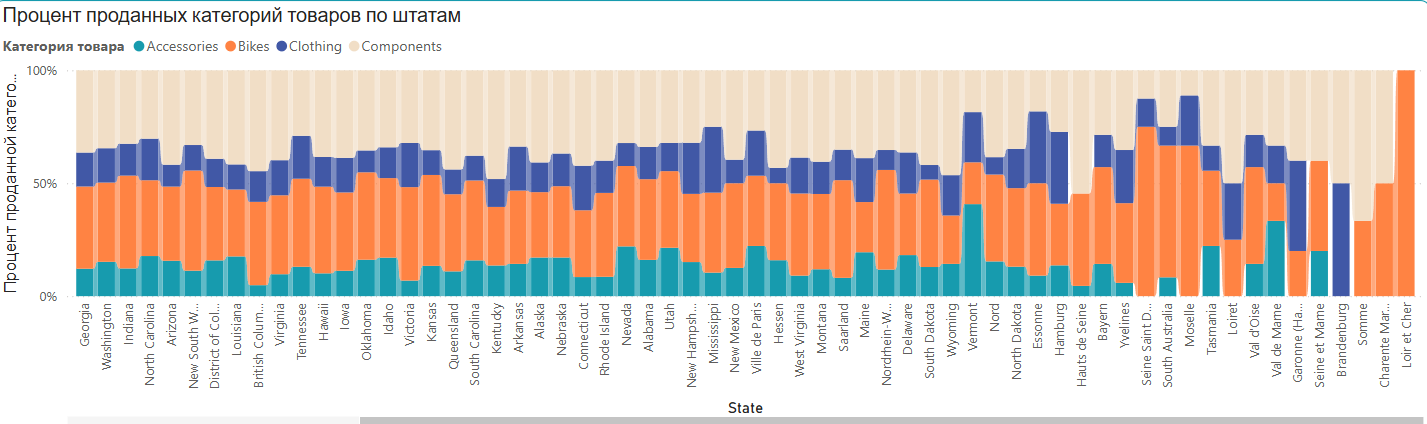
1. Mountain Bikes (16,6 % прибыли)
2. Road Bikes (22,3 % прибыли)
3. Touring Bikes (61,1 % прибыли, также имеет самую высокую маржу в 20%)



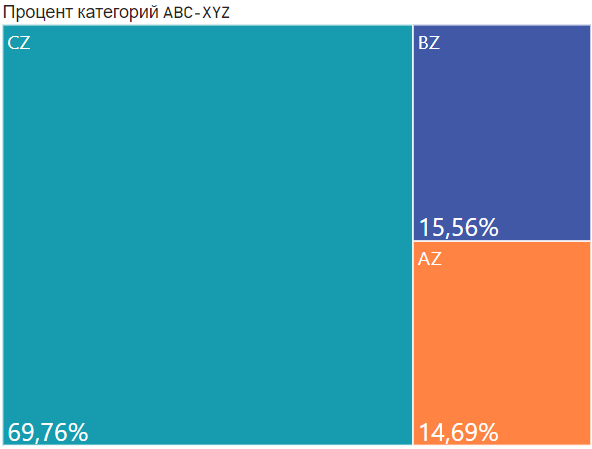
Относительно чистых убытков “анти-лидерами” списка являются товары Bike Touring-1000 Blue, 46 и Bike Touring 1000 Blue, 60



Относительно чистой маржинальности “анти-лидерами” списка являются товары категории Аксессуары: Sport-100 Helmet Red, Cable Lock, Hitch Rack – 4-Bike, HL Mountain Tire.

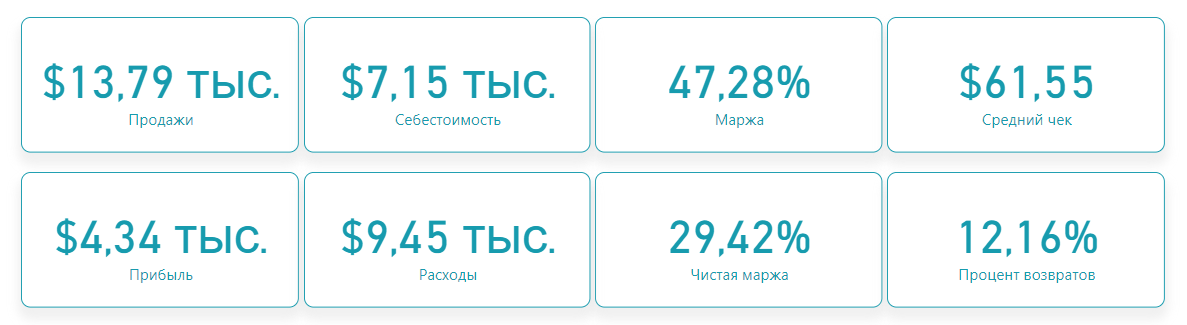


Видим большое преобладание продаж, приходящихся на велосипеды, в некоторых регионах. Данные регионы находятся не в США, а во Франции, и тем самым в данных регионах было всего несколько продаж, поэтому график имеет такие аномалии в правой стороне.

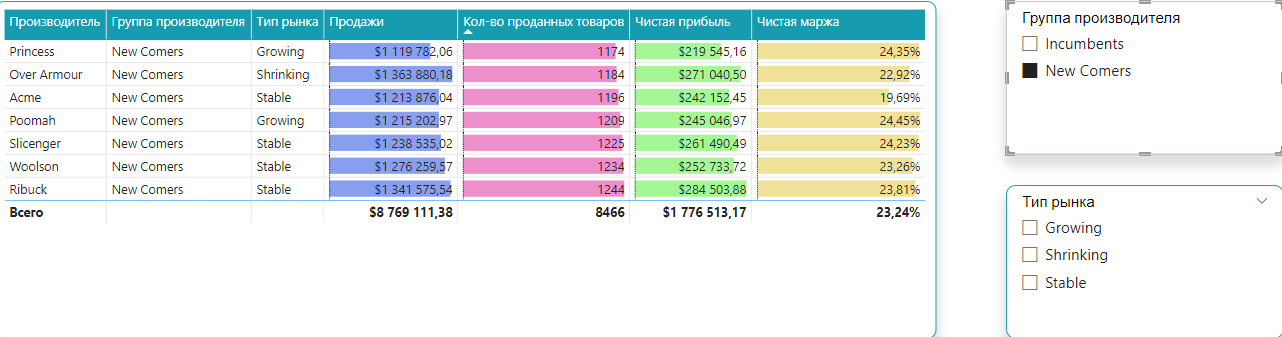
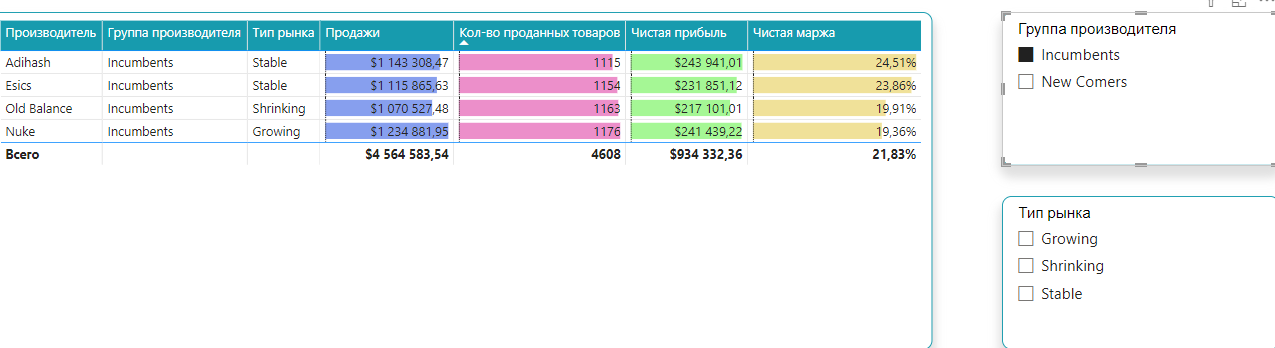


Как говорилось ранее, все товары разбиваются на 3 категории:

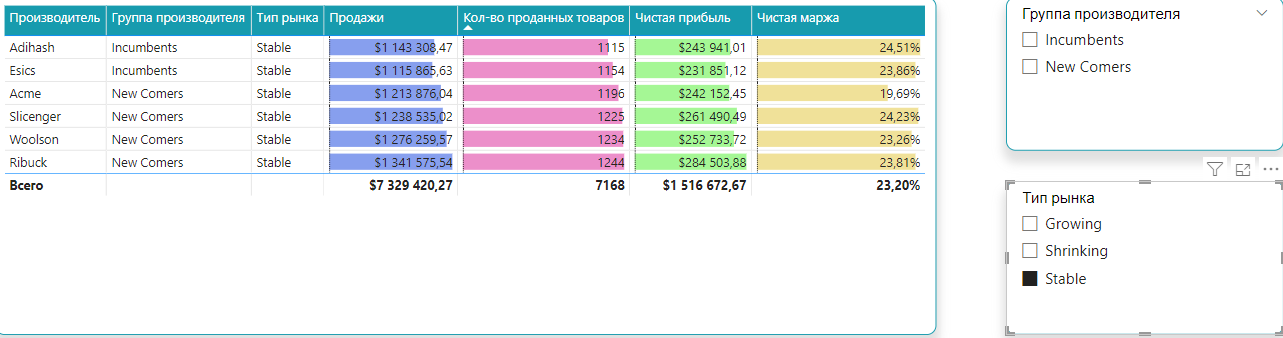
1. CZ – Основная часть товаров, приносящая лишь 6% прибыли (в эту категорию не попадают велосипеды – основной источник дохода)
2. BZ – 15,5% товаров, приносащие 14% прибыли (в эту категорию попадает большенство велосипедов и некоторая часть компонентов для велосипедов)
3. AZ – 14,7% товаров, приносящие 80% прибыли (в эту категорию попадает оставшаяся часть велосипедов)



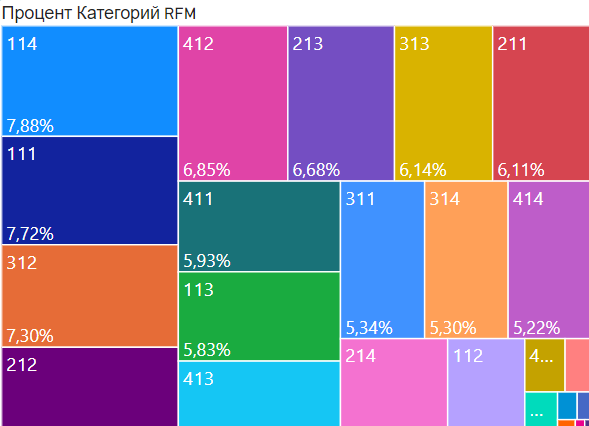
У товаров есть стиль: для мужчин, для женщин, унисекс или не указано. Случайно найдена зависимость стиля товара и чистой маржинальности. У товаров для мужчин она выше почти на 7% чем общая. Но это связанно с тем, что все “мужские товары” – это одежда, которая сама по себе имеет большую маржинальность по сравнению с велосипедами (на 9% больше).



Новички на рынке показывают более хорошие показатели по сравнению с действующими компаниями. Отталкиваемся от маржинальности, которая на 1,4% выше у новичков (не берём в расчёт продажи и прибыль, так как новичков попросту больше и они имеют больше продаж).



Компании со стабильным типом рынка также показывают более хорошие показатели маржинальности (0,5% выше по сравнению с общей средней маржинальностью)



Выделим 4 самых больших категории клиентов:

1. **114 (7,88%):** Это клиенты, которые недавно сделали покупку, но не очень лояльны и не тратят много, поэтому они могут потребовать специальных предложений или стимулов для увеличения лояльности.
2. **111 (7,72%):** Эти клиенты активно покупают, но их средний чек невысок. Возможно, они заинтересованы в недорогих товарах, и важно предложить им более выгодные предложения для увеличения средней суммы покупки.
3. **312 (7,3%):** Эта группа клиентов может быть платежеспособной, но их активность упала. Здесь могут быть потенциальные возможности для повторной активации с помощью маркетинговых кампаний, которые акцентируют внимание на стиле жизни и потребностях.
4. **212 (7,12%):** Это клиенты, которые когда-то тратили много, но не были активными в последнее время. Необходимо проанализировать причины их отсутствия и разрабатывать стратегические предложения, чтобы вернуть их в активных покупателей.

**Вывод по работе**

В ходе анализа данных выполнено их объединение, очистка и обогащение новыми признаками. С помощью **Google Colab** проведены ABC-XYZ и RFM-анализы, что позволило сегментировать товары и клиентов для дальнейшей оптимизации бизнеса. Использование **Power BI** обеспечило глубокий анализ финансовых и операционных данных с добавлением новых показателей.

**Основные выводы:**

1. **ABC-XYZ анализ**:
   * Категория AZ приносит 80% прибыли и включает ключевые товары (велосипеды).
   * Категории BZ и CZ в основном состоят из менее доходных товаров, таких как аксессуары.
2. **RFM-анализ клиентов**:
   * Выделены группы клиентов, требующие индивидуальных подходов:
     + Лояльные с высокой активностью.
     + Клиенты с низкой лояльностью, которым нужны дополнительные стимулы.
3. **Анализ продаж и прибыли**:
   * Лидеры по доходу: Калифорния, Нью-Йорк, Нью-Джерси.
   * Туристические велосипеды (Touring Bikes) занимают 61,1% прибыли, показывая наивысшую маржинальность.
   * Убыточные товары — из категории аксессуаров.
4. **Динамика данных**:
   * Значительных сезонных колебаний в продажах за 2008–2010 годы не выявлено.
   * Рост возвратов в 2010 году требует пересмотра качества продукции и работы с клиентами.
5. **Инсайты по типам компаний**:
   * Новички на рынке показывают более высокую маржинальность по сравнению с устоявшимися производителями (+1,4%).
   * Стабильные рынки также демонстрируют преимущество в маржинальности (+0,5%).

**Значение работы**

* Сегментация товаров и клиентов позволяет точнее ориентироваться в стратегиях продаж и маркетинга.
* Выявленные ключевые продукты и регионы помогают сосредоточить усилия на наиболее прибыльных сегментах.
* Анализ поведения клиентов и факторов оттока предоставляет основу для улучшения клиентской лояльности.