

Введение

Retail Dataset представляет собой базу данных 34 крупных супермаркетов США за период с 2017 по 2020 годы. Каждый супермаркет имеет свою разновидность товаров, количество проданных и оставшихся товаров, цену, выручку.

Цель данного проекта – провести анализ супермаркетов с точки зрения прибыльности, определить ходовые товары, понять динамику продаж с целью облегчения принятия решений на основе данных.

Метаданные

Набор данных представлен 3 файлами в формате CSV: Inventory (Остатки товара), Products (Товары), Sales (Продажи). Каждый файл был стандартизирован и представлен следующими столбцами и форматом данных:

Файл Inventory:

- Product ID (Целое число) - идентификационный номер товара
- Store ID (Целое число) - идентификационный номер супермаркета
- Store Name (Общий) – название супермаркета
- Address (Общий) – адрес супермаркета
- Neighborhood (Район) – район, в котором находится супермаркет
- Quantity Available (Целое число) – количество оставшегося товара

Файл Products:

- Product ID (Целое число) - идентификационный номер товара
- Product Name (Общий) – наименование товара
- Supplier (Общий) – поставщик
- Product Cost (Десятичное число) – стоимость товара в долларах США

Файл Sales:

- Sales ID (Целое число) - идентификационный номер продажи
- Store ID (Целое число) - идентификационный номер супермаркета
- Product ID (Целое число) - идентификационный номер товара
- Date (Дата) – дата продажи
- Unit Price (Десятичное число) – цена за единицу
- Quantity (Целое число) – количество проданного товара

Файл Inventory содержит 1000 строк и 6 столбцов. Файл Products содержит 1000 строк и 4 столбца. Файл Sales содержит 200000 строк и 6 столбцов.

Инструменты:

1. Очистка данных (Data Cleaning) – Microsoft Excel
2. Визуализация данных (Data Visualization) - Tableau

Все визуализации были созданы в Tableau и могут быть найдены [здесь](#). Все файлы доступны по ссылке: [enupilov/Retail \(github.com\)](https://github.com/enupilov/Retail).

Очистка данных (Data Cleaning)

Файл Inventory:

1. Дубликаты не найдены.
2. Столбцы ProductId, StoreId, StoreName, QuantityAvailable переименованы в Product ID, Store ID, Store Name, Quantity Available.
3. Формат столбцов Product ID, Store ID, Store Name, Quantity Available изменен на числовой.
4. Пустая ячейка столбца Storename была заполнена значением Dollar Tree согласно связанному значению Store ID (21791).
5. Удален ноль в начале некоторых значений столбца Address.
6. Название столбца neighborhood было исправлен на заглавную букву.

Файл Products:

1. Дубликаты не найдены.
2. Столбцы ProductId, ProductName, ProductCost переименованы в Product ID, Product Name, Product Cost.
3. Формат столбцов Product ID и Product Cost изменен на числовой и валюту в долларовых единицах.
4. Удалена одна пустая строка.

Файл Sales:

1. Дубликаты не найдены.
2. Столбцы SalesId, StoreId, ProductId, UnitPrice переименованы в Sales ID, Store ID, Product ID, Unit Price.
5. Формат столбцов Sales ID, Store ID, Product ID, Quantity изменен на числовой. Формат столбца Unit Price изменен на валюту в долларовых единицах.
3. Создан столбец Revenue (Выручка).

Анализ данных (Data Analysis)

Оборачиваемость запасов

Для определения выручки было решено создать в Microsoft Excel столбец Revenue, произведение цены за единицу (Unit Price) и количества проданного товара (Quantity).

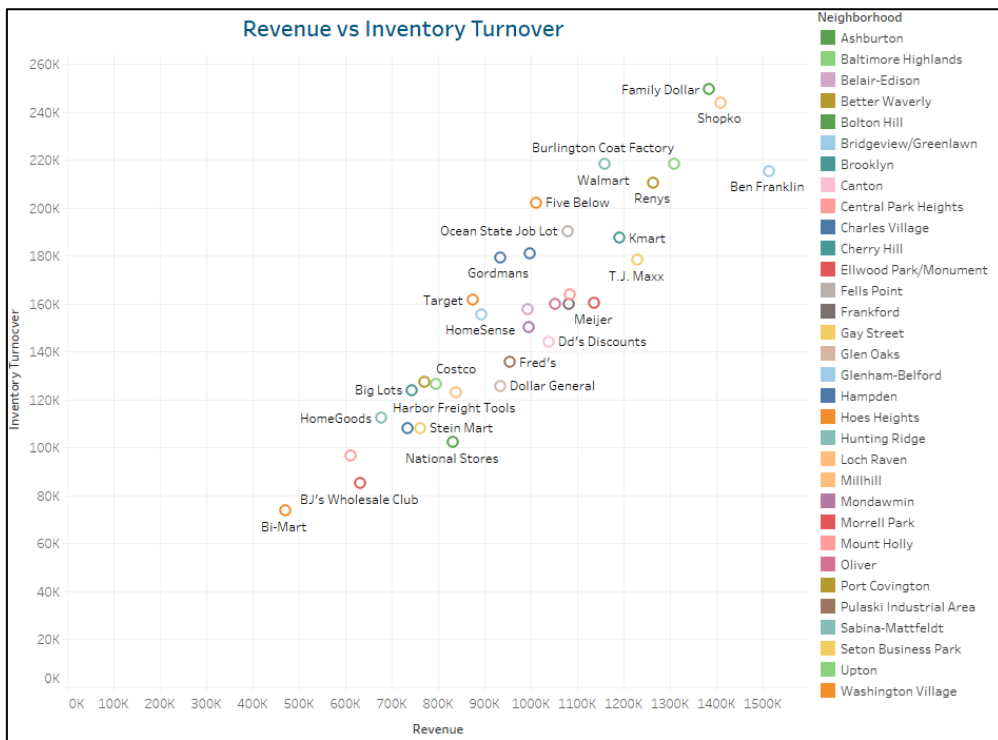
$$\text{Revenue} = [\text{Unit Price}] * [\text{Quantity}]$$

Для анализа товарного склада было создано в Tableau вычисляемое поле Inventory Turnover (Оборачиваемость запасов):

$$\text{Inventory Turnover} = \text{SUM}([\text{Revenue}]) / \text{AVG}([\text{Quantity Available}])$$

Данное поле позволит нам понять как часто супермаркеты продают и заменяют свои запасы. Чем выше оборачиваемость запасов, тем быстрее супермаркет продает свой товар и обновляет запасы. Низкая оборачиваемость запасов будет свидетельствовать о проблемах с продажами, устареванием товара или неэффективным управлением запасами.

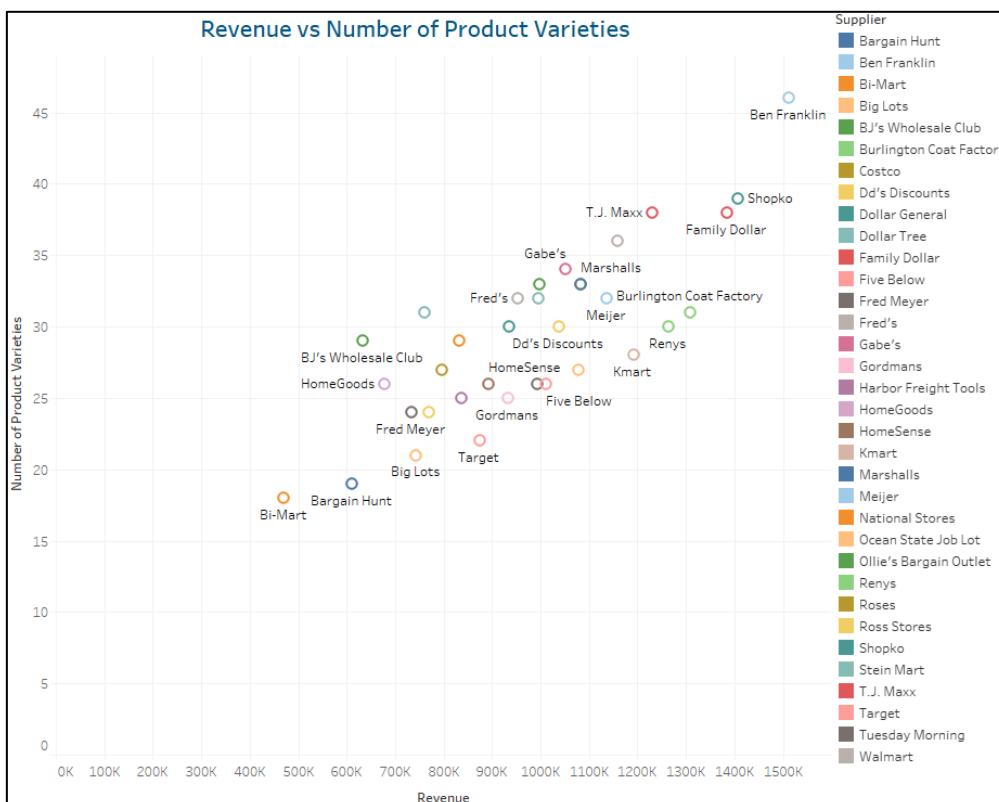
В целом, наблюдается линейная зависимость между выручкой и оборачиваемостью запасов. График показывает, что имеется и высокая выручка, и высокая оборачиваемость, что говорит о здоровом и успешном бизнесе.



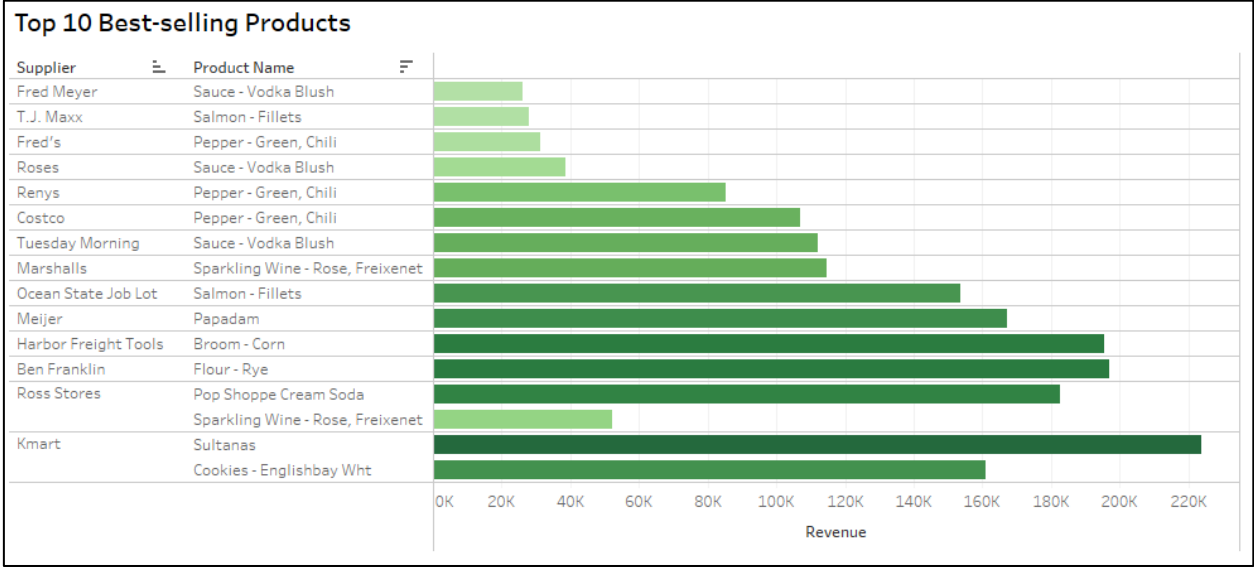
Корреляция товара и выручки

В рамках анализа товаров также наблюдается линейная зависимость между выручкой и числом разновидностей товаров.

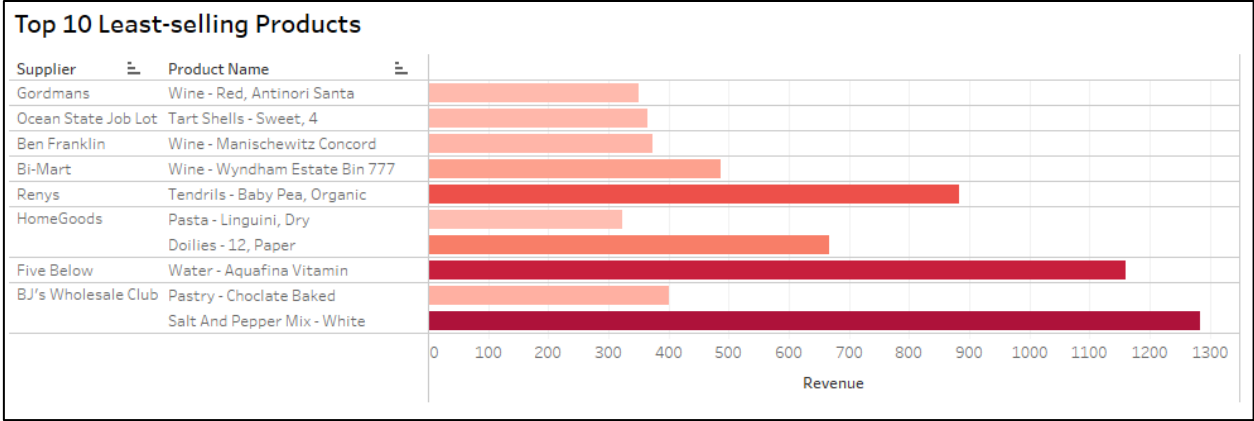
Примечательно, что супермаркеты, имеющие большую выручку и оборачиваемость запасов, имеют и большую разновидность товаров. Такая зависимость может быть характерна для компаний, которые успешно используют стратегию диверсификации продуктов, чтобы привлечь больше клиентов или удовлетворить разнообразные потребности рынка.



Касательно продаваемых товаров хотелось бы отметить Зеленый перец чили (Pepper – Green, Chili). Данный продукт является ходовым в трех супермаркетах. Но, как ни странно, самым продаваемым продуктом является изюм (Sultanas.)



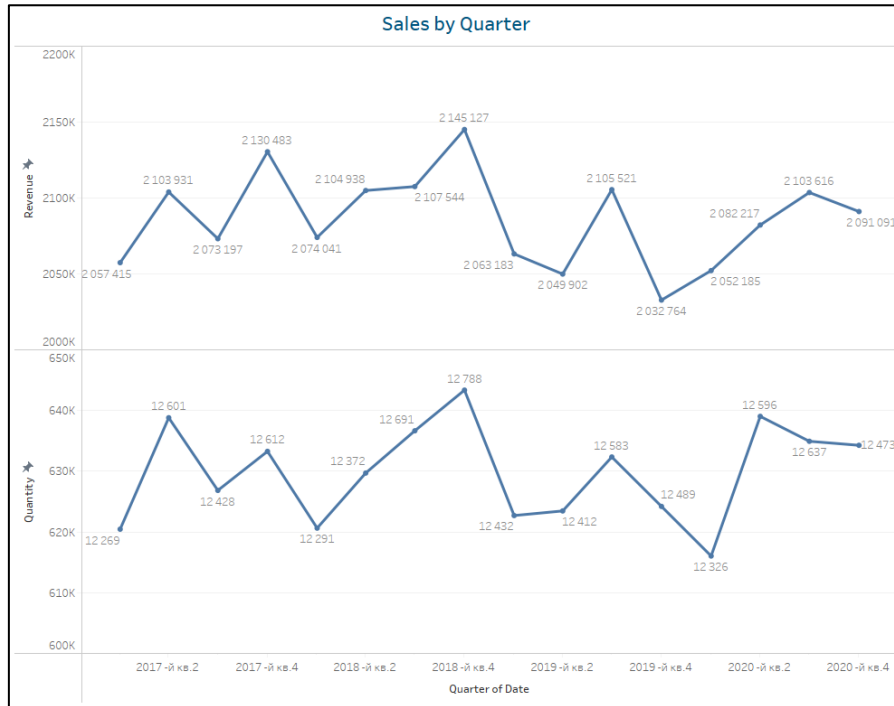
В категории слабо продаваемых отмечается вино. Возможно супермаркету Ben-Franklin, который имеет и так не плохие продажи, стоит отказаться от продажи вина Manischewitz Concord и заменить его на игристое вино Rose, Freixenet, так как данный вид продукта входит в Топ-10 самых продаваемых товаров.



Динамика продаж

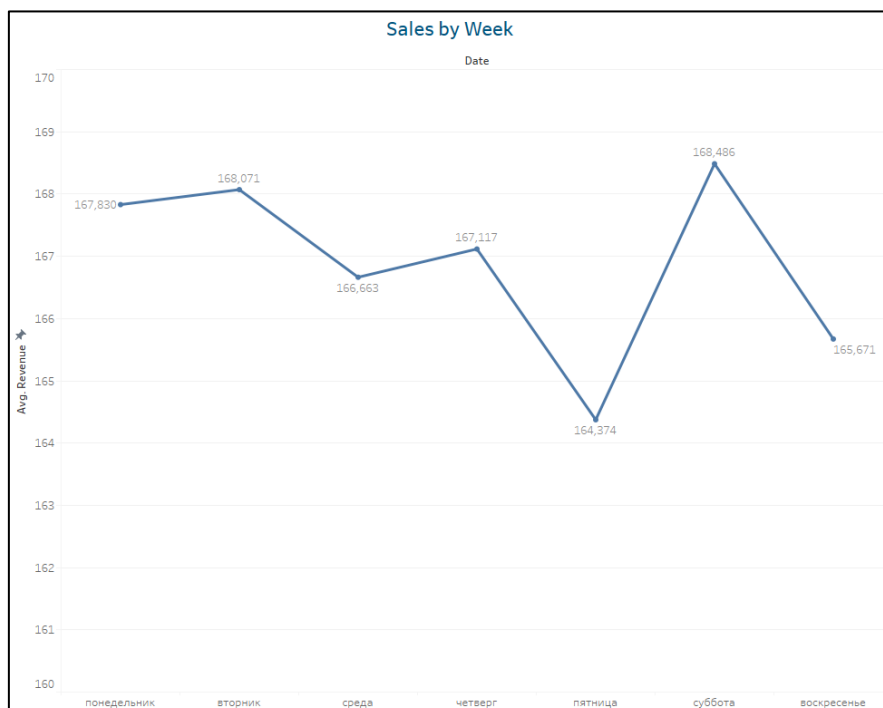
Касательно продаж были выполнены визуализации продаж поквартально, за неделю и с прогнозом на следующий год.

Анализ поквартально весь период показал, что у продаж не имеется четкой тенденции на рост. Также при сравнении с количеством проданного товара можно отметить корреляцию.



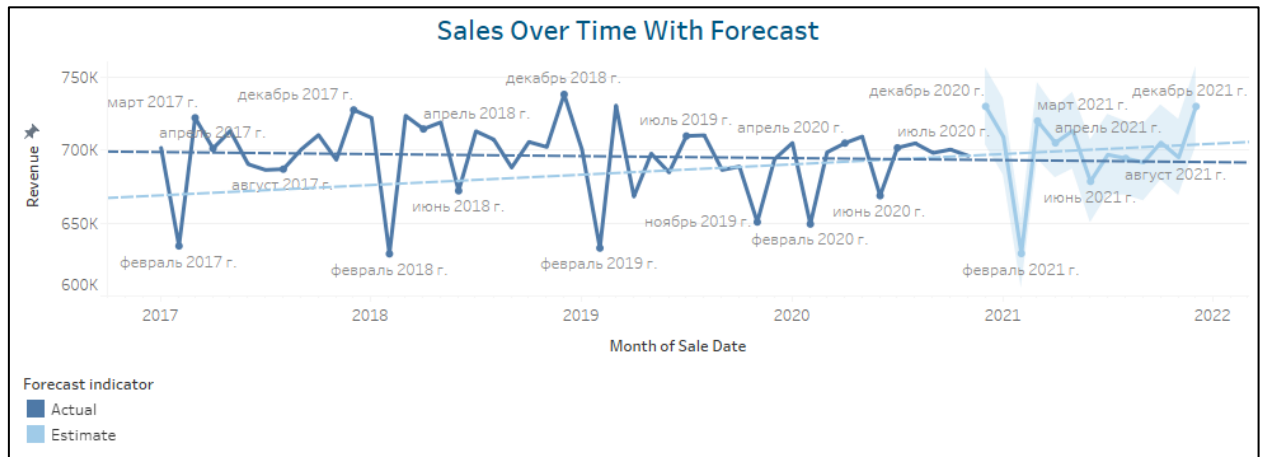
В 2018 году можно отметить, что количество продаж увеличиваются, а выручка остается стабильной. Похожая ситуация отмечается и в 2020 году. Это может указывать на увеличение объема продаж товаров с низкой ценой.

В разрезе недели отмечается, что по пятницам продажи самые слабые. Но пик продаж приходится на субботу. Возможно, стоит пересмотреть скидочную стратегию для увеличения пятничных продаж.



Анализируя весь период продаж, можно отметить, что для зимнего периода характерно ослабевание в продажах. В декабре начинается упадок продаж, которые длится аж до февраля. В феврале же начинается рост продаж.

Прогноз указывает на продолжение восходящей тенденции продаж, характеризующейся паттернами, подобными тем, что наблюдались в исторических данных.



Заключение

Данный анализ выявил ценные идеи о работе супермаркетов США. Изучая тенденции продаж, корреляции выручки мы получили понимание динамики продаж супермаркетов. Обнаруженная годовая сезонность предоставляет стратегические направления для оптимизации бизнес-стратегий. По мере движения вперед, эти результаты, без сомнения, будут информировать принятие решений на основе данных, способствуя улучшению прибыльности и удовлетворенности клиентов в операциях супермаркетов.