**Car Sales Project (SQL + Power BI Project)**

**Введение**

Данный проект содержит анализ продаж автомобилей, а также ценовой политики автодилеров в 7 штатах США с 10.01.2022 по 09.09.2023. Проект создан с использованием Microsoft Excel, SQL (Microsoft SQL Server) и Microsoft Power BI.

**Описание данных**

Датасет взят с сайта “Kaggle” и содержит ряд показателей:

* Id автомобиля
* Дату приобретения
* Имя покупателя
* Пол покупателя
* Годовой доход покупателя
* Название автодилера
* Название бренда и модели автомобиля (В дальнейшем будет разделена на два новых столбца «Компания» и «Модель»)
* Двигатель, установленный в автомобиле
* Коробка передач автомобиля
* Цвет автомобиля
* Цена автомобиля
* Номер автодилера
* Кузов автомобиля
* Телефонный номер покупателя
* Регион нахождения автодилера

Формат датасета – лист Microsoft Excel.

**Подготовка к анализу**

В первую очередь, столбец «Company & Model» был разделен на отдельные столбцы «Company» и «Model» (шаг выполнен в программе Microsoft Excel). Это было сделано для того, чтобы проанализировать продажи по конкретным автомобильным брендам без уточнения модели автомобиля.

Далее, преобразованный датасет был загружен в Microsoft SQL Server Management. К базе данных был составлен ряд запросов для дальнейшего анализа и получения данных для визуализаций:

SELECT COUNT(Car\_id) AS total\_orders from Car\_Sales

Для вычисления общего числа проданных машин



SELECT SUM(Price) as Total\_Revenue from Car\_Sales

Для подсчета общей выручки



SELECT AVG(Annual\_Income) as Avg\_Annual from Car\_Sales

Для подсчета среднегодового дохода покупателей



SELECT COUNT(DISTINCT(Company)) AS Number\_Of\_Companies FROM Car\_Sales

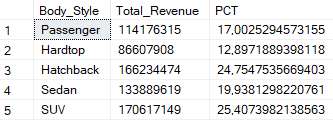
Для подсчета уникального числа автодилеров



SELECT Body\_Style, sum(Price) as Total\_Revenue, sum(Price) \* 100 / (SELECT sum(Price) from Car\_Sales) as PCT from Car\_Sales

Group by Body\_Style

Для вычисления доли прибыли по типу кузова автомобиля

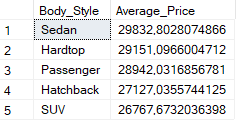


SELECT Body\_Style, AVG(Price) as Average\_Price from Car\_Sales

GROUP BY Body\_Style

Order by Average\_Price Desc

Для подсчета средней цены по кузову автомобиля



SELECT top 5 Dealer\_Name, sum(Price) as Total\_Revenue from Car\_Sales

Group by Dealer\_Name

Order by Total\_Revenue Desc

Для выборки 5 автодилеров-лидеров по объему выручки

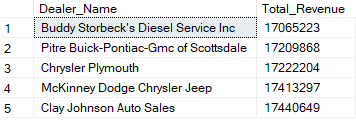


SELECT top 5 Dealer\_Name, sum(Price) as Total\_Revenue from Car\_Sales

Group by Dealer\_Name

Order by Total\_Revenue Asc

Для выборки 5 автодилеров-аутсайдеров по объему выручки

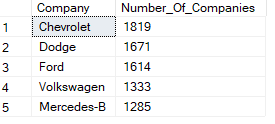


SELECT TOP 5 Company, COUNT(Company) AS Number\_Of\_Companies FROM Car\_Sales

GROUP BY Company

ORDER BY Number\_Of\_Companies DESC

Для выборки 5 наиболее популярных брендов (лидеров по числу проданных машин)



Данный шаг был необходим для того, чтобы выявить **точные** данные из датасета, потому что Power BI округляет значения и некоторые данные воспринимаются некорректно.

Далее, датасет был импортирован в Power BI. На основе данных при помощи функций DAX были рассчитаны «меры», а на основе мер – часть необходимых визуализаций:

* Общее число заказов (подсчет всех Id номеров машин в датасете при помощи функции COUNT)
* Общая выручка (сумма всех цен машин в датасете при помощи функции SUM)
* Среднегодовой доход покупателя (среднее значение годового дохода при помощи функции AVERAGE)
* Число автодилеров (подсчет **уникальных** записей в датасете при помощи функции DISTINCTCOUNT)

Показатели были представлены в виде карточек.

Затем были построены оставшиеся визуализации:

* Круговая диаграмма, отражающая объем и долю выручки от продаж машины с определенным типом кузова от общей выручки
* Столбчатая диаграмма, иллюстрирующая среднюю цену автомобиля с определенным типом кузова
* Две столбчатые диаграммы, изображающие пятерку лидеров и аутсайдеров по объему выручки
* Гистограмма, показывающая топ 5 самых популярных брендов (по числу проданных машин)
* Диаграмма рассеяния для выявления взаимосвязи между ценой автомобиля и объемом выручки

Также, наверх дашборда был добавлен фильтр по дилерам, так что пользователь может посмотреть информацию по конкретному дилеру.

**Анализ**

Итак, общая выручка составила примерно 671,53 млн $. Наиболее существенную долю занимают продажи внедорожников и хэтчбеков – 177 млн и 166 млн, или 25,41% и 24,75% соответственно. При этом автомобили в таких кузовах имеют наименьшую среднюю цену – 27,13 тыс $ для хэтчбека и 26,77 тыс $ для внедорожника. Такая низкая цена делает эти машины наиболее привлекательными для покупателей и, можно сказать, обеспечивает дилерам наилучшие продажи.

Обратимся к дилерам-лидерам и дилерам-аутсайдерам. Эти два лагеря различает совершенно разный подход к ценовой политике. Если посмотреть на дилеров-лидеров, то становится заметно, что цены на внедорожники хэтчбеки близки к средним, а цены на остальные кузова немного выше средних. Можно сделать предположение, что оптимальная цена, которая обеспечивает максимальную прибыль выше реальных средних цен на автомобили с этими типами кузова. Ситуация с дилерами-аутсайдерами противоположная – цены на внедорожники, и хэтчбеки завышена, что наоборот, делает эти предложения непривлекательными, а цены на остальные типы кузова занижены, что лишает дилера дополнительной выручки.

Если обратиться к диаграмме рассеяния, то можно заметить, что повышение цены на автомобиль ведет к пусть и небольшому, но увеличению выручки. Однако не стоит воспринимать это, как призыв к действию, поскольку, как уже говорилось ранее, повышение цены делает предложение для покупателя менее привлекательным и снижает число проданных машин. Более того, бесконтрольное повышение цены может привести к тому, что на товар будет установлена цена бойкота, когда товар не будут покупать вообще и дилер ничего не заработает.

Также интересно, что три самых популярных автомобильных бренда – американские. Машины именно этих брендов покупали больше всего. Данную информацию также необходимо учитывать при составлении модельного ряда.

Подводя итоги, стоит заметить, что грамотная ценовая политика положительно влияет на выручку, а, соответственно, и на успешность компании. Специфика автомобильной отрасли заключается в том, что оптимальная цена на автомобиль может различаться, в зависимости от типа кузова.

**Возможности для дальнейшего анализа**

Данный дашборд раскрывает лишь ценовой аспект вопроса в то время, как существуют другие факторы, которые также могут влиять на объем выручки компании. Эти факторы также требуют изучения. На основе показателей в датасете, можно провести следующий анализ:

* Проанализировать предпочтения покупателей в зависимости от их пола и адаптировать под них свою рекламную кампанию. Например, если автомобили марки «Dodge» больше популярны у мужчин, то рекламная компания автомобилей этой марки должна быть ориентирована на мужчин.
* Проанализировать предпочтения покупателей в зависимости от региона, вполне возможно, что они будут различаться, из-за специфики конкретного штата.
* Изучить, какие цвета пользуются наибольшей популярностью и изменить модельный ряд в соответствии с предпочтениями покупателей.
* Разделить покупателей на четыре равных группы (по квартилям), в зависимости от величины годового дохода, и также посмотреть на предпочтения покупателей в каждой из групп: какие бренды, какой кузов, какая коробка передач популярны в определенной группе.