

ХОД РАБОТ ПО ПРОЕКТУ

Автоматически сформированный отчет с презентационной версткой

Ход работ

Было получено 7 Excel таблиц с данными, спаршенными другим членом команды.

При малом числе товаров в таблице (20-40 примеров) некоторые ячейки в таблице корректировал вручную.

Далее все таблицы переносились в pandas.

Данные по возможности из текстового вида переводились в числовые и иные структуры (список записанный как строка).

Все таблицы приводились к единому виду, а значения к единым единицам измерения, были определены следующие колонки: 'Компания', 'Категория', 'Название товара', 'Цена', 'Скидочная Цена', 'Объем', 'Единицы Измерения'.

Ввиду слишком широкого списка категорий и их сходства, они были объединены в 6 категорий: Газы и газовые смеси, Промышленные технологии, Косметика, Герметики, Клеи и строительные смеси, Сварочные материалы

Далее заполнялись пропуски в данных

Колонки: 'Категория', 'Цена', 'Объем', Единицы Измерения' заполнялись на основе рыночных и торговых реалий вручную.

Колонка: 'Скидочная Цена' заполнялась в зависимости от категории товара

Для каждой категории был определён диапазон целых чисел из которого затем равномерно сэмплировались скидочные проценты.

Диапазоны и их обоснование

Газы и газовые смеси (3-10%)

Высокотехнологичное производство, стабильный спрос Низкая эластичность спроса по цене Консервативная ценовая политика

Промышленные технологии (5-15%)

B2B сегмент, долгосрочные контракты Умеренные скидки для стимулирования крупных заказов Баланс между рентабельностью и конкурентоспособностью

Косметика (5-25%)

Высокая конкуренция, импульсные покупки Частые акции и распродажи Широкий диапазон для разных ценовых сегментов

Герметики (10-20%)

Стандартизированная продукция, высокая конкуренция Необходимость стимулировать повторные покупки Умеренно-агрессивные скидки

Клеи и строительные смеси (10-30%)

Сезонный спрос (строительный сезон) Высокая конкуренция среди производителей Крупные оптовые скидки

Сварочные материалы (5-20%)

Специализированное оборудование + расходники Разные стратегии для разных подкатегорий Умеренные скидки на премиум-оборудование

При большом числе пропусков у товара (2-3, например объем, цена, единицы измерения), такие данные удалялись, поскольку чрезмерное злоупотребление искусственным заполнением исказило бы общую статистику.

На основе имеющихся данных были созданы новые колонки-признаки

Цена за единицу объема

Процент Скидки

Поквартальная динамика цен на товары с января 2021 по октябрь 2025

Для этого были взяты с сайта Росстата (<https://rosstat.gov.ru/statistics/price>) индексы цен производителей по видам экономической деятельности (с 1998 г.), а конкретнее данные с января 2021 по октябрь 2025 для категории:

Производство химических веществ и химических продуктов, на основе которых синтезировались цены сначала ежемесячные, а затем и ежеквартальные (что было необходимо для сокращения числа признаков)

Как синтезировалась история цен

Брался индекс Росстата за текущий месяц, делился на 100, (тк индексы отражали изменение относительно прошлого месяца умноженное на 100: база 100 -> выросло на 8% - индекс = 108) после чего цена товара делилась на индекс, тем самым получалась цена за прошлый месяц, затем так рекурсивно повторялось до января 2021

Ввиду большого числа признаков, ежемесячные цены были сведены к ежеквартальным

Затем к полученным ценам добавлялся шум

Для каждой категории товаров брался свой шумовой диапазон, со своим реалистичным/рыночным обоснованием:

Высокая волатильность

Газы и газовые смеси - сильно зависят от цен на энергоносители, сырьевые рынки, логистики

Клеи и строительные смеси - тесно связаны со строительным циклом, сезонностью спроса

Средняя волатильность

Герметики - зависимость от строительного рынка, но более стабильный спрос

Промышленные технологии - B2B сегмент с долгосрочными контрактами

Низкая волатильность

Косметика - потребительские товары с устойчивым спросом, брендовая лояльность

Сварочные материалы - промышленное применение, но специализированный спрос

'Газы и газовые смеси'

(0.95, 1.10), # -5% до +10% - высокая волатильность

'Промышленные технологии'

(0.98, 1.07), # -2% до +7% - средняя волатильность

'Косметика'

(0.99, 1.05), # -1% до +5% - низкая волатильность

'Герметики'

(0.97, 1.08), # -3% до +8% - средняя волатильность

'Клеи и строительные смеси'

(0.96, 1.09), # -4% до +9% - высокая волатильность

'Сварочные материалы и оборудование'

(0.98, 1.06), # -2% до +6% - низкая волатильность

Сдвиг диапазонов в положительную сторону обусловлен фактором инфляции

Затем в зависимости от категории товара равномерно сэмплировалось значение из категориального диапазона, которое умножалось на его ценовое значение за соответствующий квартал, тем самым для каждого товара получалась своя история цен с различным шумом из категориального диапазона в кажд.

За счёт такого подхода динамика цен на товары согласуется с движением индекса Росстата, но ввиду добавления уникального категориального шума, получается имитировать различные сезонности, свойственные отдельным товарам/категориям.

Также была создана таблица с информацией о компаниях, путём агрегирования исходной таблицы и создания сводных таблиц.

Таблица с информацией о компаниях включала в себя

Модальную категорию (категорию товаров, которых большинство в ассортименте компании)

Количество товаров модальной категории

Медианный процент скидки по всем товарам компании всех категорий

Количество товаров каждой категории в ассортименте компании

Медианный процент скидки по каждой категории товаров

Итого датасет для анализа принял следующий вид: 2061 товар, 6 категорий 28 столбцов признаков, 9 из которых различны, а остальные - ежеквартальная история цен. Данным свойственен сильный дисбаланс во многих аспектах: число товаров в категориях, число товаров у компаний и др.

При анализе данных были построены

Ежемесячные и ежеквартальные графики ранее упомянутых индексов цен Росстата с января 2021 по октябрь 2025 для категории: Производство химических веществ и химических продуктов

На фоне квартального графика Росстата были построены графики средней ежеквартальной динамики цен по каждой категории. А также для любого отдельно взятого товара, чтобы сравнить историческую динамику цен на товар и общий индекс Росстата (это может быть полезно для выявления недооцененных и переоцененных товаров на текущий момент, что позволит выгоднее совершать закупки)

Были построены Ящики с усами (Boxplots) по каждой категории для общего анализа статистик цен в двух вариантах:

В первом для лучшей визуальной интерпретации графиков была взята статистика без экстремальных выбросов (5% данных) или же 95-й перцентиль. (Все графики на одной картинке + каждый график детальнее на отдельной)

Во втором варианте использовались все данные, но была применена логарифмическая шкала цен, опять же для лучшей визуальной интерпретации ящиков с усами, численная же статистика также была указана в аннотации к каждому графику (все графики на одной картинке)

Построены Bar графики количества товаров по категориям и количество товаров по компаниям, последний в различных вариациях (топ-5, топ-10 и тд компаний, топ-30 компаний, не считая первых 5 самых крупных и тд) Итог: Категории по кол-ву товаров: Газы и газовые смеси - 809, Сварочные материалы и оборудование - 659, Косметика - 442, Клеи и строительные смеси - 75, Промышленные технологии - 71, Герметики - 5. Компании по кол-ву товаров: АО_АПК_БЕЛИФФ_ГАЗ - 1061, КРЫМСКАЯ-РОЗА - 191 КЕСТО - 153, Novogas - 6, ЯРПОЖИНВЕСТ - 44

Была построена Тепловая карта медианных цен (цена, цена за единицу объема, скидочная цена) по категориям. Итог: Газы и газовые смеси обладают наибольшей медианной ценой(5800) и медианной скидочной ценой(5568), Косметика имеет наибольшую цену на единицу объема (8990 за литр). Самыми дешевыми на единицу объема являются Клеи и Строительные смеси (123) и Промышленные технологии (123), Герметики же имеют самую низкую медианную цену (650) и медианную скидочную цену(590).

Были построены графики зависимости цены от объема в 4x различных вариантах шкал (обычные, log-log, обычная-log, log-обычная), это стало возможным за счёт единообразных единиц измерения (в подавляющем большинстве литры и килограммы в соотношении 2/1) на этих графиках можно проследить определенную кластерную структуру данных, и вполне возможную линейную разделимость с такими признаками (например на log-log можно легко отделить Косметику от Газов и Газовых смесей) а также Сварочные материалы материалы и оборудование, но немного в меньшей степени). Товары же мало представленных категорий визуально отличить сложно.

Были построены Карты плотности (density map) всё в тех же шкалах цена-объем в двух вариантах (обычные и log-log)

Построены Bar графики рейтингов компаний по категориям, за счёт созданной таблицы с информацией о компаниях, в трёх вариантах:

Рейтинг на основе медианной скидки компании на товары категории

Рейтинг на основе количества товаров категории в ассортименте компании

Рейтинг на основе метрики «Индекс привлекательности», равной произведению размера медианной скидки из варианта 1, и натурального логарифма от количества товаров из варианта 2. Эта метрика учитывает как и величину скидки на товары, так и их количество, что позволяет определить компанию, где с большей вероятностью можно будет найти желаемый товар по хорошей скидке.

Более детальный разбор метрики

Привлекательность = Скидка $\times \log(\text{Товары} + 1)$

Что измеряет

1. Широта выбора (чем больше товаров, тем выше вероятность найти нужное)
2. Глубина скидок (чем больше скидка, тем выгоднее покупка)

Комплексный показатель - учитывает оба фактора одновременно

Интерпретация: высокий показатель - компания с широким ассортиментом и большими скидками, низкий показатель - либо мало товаров, либо маленькие скидки.

Логарифмирование количества товаров применяется для отражения убывающей полезности расширения ассортимента и снижения влияния статистических выбросов, те на случай, когда компании с крупным ассортиментом и низкими скидками могут заполнить весь рейтинг, оставив далеко позади компании с небольшим ассортиментом, но приятными скидками.

Топ 3 компаний по скидке в каждой Категории

Газы и газовые смеси

КРИО СЕРВИС - 9% 18шт.

АлтайСтройДеталь - 7.5% 20шт.

АО АПК БЕЛИФФ ГАЗ - 6.99% 448шт.

Сварочные материалы и оборудование

АО АПК БЕЛИФФ ГАЗ - 12.99% 613шт.

КРИО СЕРВИС - 12% 12шт.

ЯРПОЖИНВЕСТ - 2.01% 24шт.

Косметика

Family Cosmetics - 25.12% 1шт.

Мануфактура натуральной косметики и мыла Живица - 25.05% 1шт.

Ievrana - 25.02% 1шт.

Клей и строительные смеси

KESTO - 7.43% 75шт.

Промышленные технологии

КЕСТО - 7.27% 71шт.

Герметики

КЕСТО - 8.11% 5шт.

Топ 3 компаний по количеству товаров в каждой Категории

Газы и газовые смеси

АО АПК БЕЛИФФ ГАЗ - 6.99% 448шт.

Novogas - 2.01% 76шт.

HPC Research - 2% 33шт.

Сварочные материалы и оборудование

АО АПК БЕЛИФФ ГАЗ - 12.99% 613шт.

ЯРПОЖИНВЕСТ - 2.01% 24шт.

КРИО СЕРВИС - 12% 12шт.

Косметика

КРЫМСКАЯ-РОЗА - 14.01% 191шт.

Compliment - 15.02% 25шт.

mirrolla - 20% 9шт.

Клеи и строительные смеси

КЕСТО - 7.43% 75шт.

Промышленные технологии

КЕСТО - 7.27% 71шт.

Герметики

КЕСТО - 8.11% 5шт.

Топ 3 компаний по метрике «Индекс привлекательности» в каждой Категории

Газы и газовые смеси

АО АПК БЕЛИФФ ГАЗ - 6.99% 448шт. Index = 42.67

КРИО СЕРВИС - 9% 18шт. Index = 26.01

АлтайСтройДеталь - 7.5% 20шт. Index = 22.47

Сварочные материалы и оборудование

АО АПК БЕЛИФФ ГАЗ - 12.99% 613шт. Index = 83.37

КРИО СЕРВИС - 12% 12шт. Index = 29.82

ЯРПОЖИНВЕСТ - 2.01% 24шт. Index = 6.39

Косметика

КРЫМСКАЯ-РОЗА - 14.01% 191шт. Index = 73.58

Compliment - 15.02% 25шт. Index = 48.35

mirrolla - 20% 9шт. Index = 43.94

Клей и строительные смеси

КЕСТО - 7.43% 75шт. Index = 32.08

Промышленные технологии

КЕСТО - 7.27% 71шт. Index = 30.99

Герметики

КЕСТО - 8.11% 5шт. Index = 13.05

Соответственно лидерами в своих категориях являются компании: АО АПК
БЕЛИФФ ГАЗ, КРИО СЕРВИС, КРЫМСКАЯ-РОЗА, КЕСТО