В таблице о данных товарной иерархии # pr\_df

Анализ столбцов:

Категориальные:

• pr\_sku\_id - собрано 2050 уникальный продуктов.

• pr\_group\_id - всего уникальных групп товаров – 9 шт.

• pr\_cat\_id – всего уникальных категорий товаров – 43 шт.

• pr\_subcat\_id – всего уникальных подкатегорий товар – 170 шт.

• pr\_uom\_id – 2 уникальных значения – продажа по весу или штучно.

1 – 1279

17 – 771

Видно, что товар, продающие по «1» больше

Категориальные признаки распределены неравномерно (см. Графики):

• «Распределения для pr\_group\_id»

• «Распределения для pr\_cat\_id»

• «Распределения для pr\_subcat\_id»

ТУТ НАДО ДОДЕЛАТЬ ГРУППИРОВКУ ПО ГРУППАМ ТОВАРОВ, ПОТОМ ЧТО ИХ МАЛО, ЭТО БУДЕТ ИНФОРМАТИВНО

Пропущенные значения:

Прощенные значения отсутствуют.

sales\_df\_train

Анализ столбцов:

Целочисленные:

• pr\_sales\_type\_id

1 – 40.6%

17 – 59.4%

• pr\_sales\_in\_units и pr\_promo\_sales\_in\_units

Есть значения больше 2000, вероятнее всего это граммы. Модель должна будет сама определить эту грань и найти зависимость. Возможно стоит сделать DBCSAN, для кластеризации units.

• pr\_sales\_in\_rub и pr\_promo\_sales\_in\_rub

Есть аномально большие значения продаж в рублях, возможно это просто большие заказы.

Есть отрицательные значения. Их очень мало – 78. Возможно аномалии.

Общий вывод по числовым столбцам:

Видно, как все числовые столбцы имеют очень большой коэффициент скошенности право. Товары, продающиеся по весу и по штучно не имеют принципиального отличия в данных – просто целые числа, однако есть очень большие – вероятнее всего это и есть вес.

По продажам в цене есть и отрицательные (но их крайне мало 78) и очень большие значения – но там данные чистые хоть с небольшими нюансами (есть небольшое количество продуктов, но большая стоимость больше 10000)- вероятнее всего это просто аномальный спрос или очень большие заказы и дорогие готовые продукты.

Смешанные:

• st\_id – всего 12 уникальных id магазинов.

• pr\_sku\_id -

• date (ГГ-ММ-ДД):

Всего уникальных дней – 352.

Начальная дата – 2022-08-01

Конечная дата - 2023-07-18

Из графика «График продаж ниже (красное) и выше (зеленое) медианы за все время по дням» видно, что

Среднее значение кол-ва продаж в день за весь период 2508.57

Медианной значение кол-ва продаж в день за весь период 2461.5

• Доля дней с кол-вом продаж ниже медианы 2461.5: 50.00%

• Доля дней с кол-вом продаж ниже средней 2508.57: 55.68%

Также, стоит отметить, что в декабре ближе к середине месяца продажи весьма высокие и до конца месяца не уходят ниже медианы.

Из графика «График продаж ниже (красное) и выше (зеленое) медианы за все время по месяцам» видно, что

Среднее значение кол-ва продаж в день за весь период 75760.5

Медианной значение кол-ва продаж в день за весь период 73584.58

• Доля месяцев с количеством продаж ниже медианы 75760.5: 50.00%

• Доля месяцев с количеством продаж ниже средней 73584.58: 33.33%

Доля нижне среднего кол-во не очень информативна, поскольку есть аномально маленькие продажи в Июле – разница между Июлем и Июнем -31830. Возможно связано это с сезоном отпусков.

Что можно сказать точно глядя на два графика касательно месяцев:

1) Месяцы, кол-во продаж в которые было меньше медианы (6шт.):

\* Январь - 70560

\* Февраль -69012

\* Июль -47322

\* Август -75236

\* Сентябрь -72903

\* Ноябрь -73849

Месяца, кол-во продаж в которые было больше медианы (6 шт.):

\* Март -78307

\* Апрель -76285

\* Май -79783

\* Июнь -79152

\* Октябрь -77553

\* Декабрь -83053

Янв - 8.0%

Фев - 7.8%

Мар - 8.9%

Апр - 8.6%

Май - 9.0%

Июн - 9.0%

Июл - 5.4%

Авг - 8.5%

Сен - 8.3%

Окт - 8.8%

Ноя - 8.4%

Дек - 9.4%

Динамика продаж в основном чередуется. Один месяц рост, следующий небольшое относительно роста падение. Однако есть месяц, в который падение продаж очень серьезное - \*\*40.2%\*\* - в процентном отношении к предыдущему. После, в Августе, почти все потери нивелируются и выходят на Июньский уровень.

\*\*Лучший месяц\*\* по продажам – Декабрь (\*\*83053\*\*) и \*\*худший\*\* – Июль (\*\*47322\*\*)

По дням недели:

Дни недели с продажами ниже медианы: Понедельник, Вторник, Среда

Дни недели с продажами выше медианы: Четверг, Пятница, Суббота, Воскресенье

Доля дней недели с количеством продаж ниже медианы 124632.0: 42.86%

День недели с минимальными 116321 продажами: Понедельник

День недели с максимальными 139385 продажами: Суббота

По сезонам:

Доля сезонов с количеством продаж ниже медианы 223465.0: 50.00%

Сезоны с продажами ниже медианы: Зима, Лето

Сезоны с продажами выше медианы: Весна, Осень

Минимальное кол-во продаж в месяц Лето= 201710

Минимальное кол-во продаж в месяц Весна = 234375

Доли каждого сезона по кол-ву продажам:

Весна - 26.2%

Осень - 25.4%

Зима - 25.2%

Лето - 22.8%

Рекомендации бизнесу по увеличению продаж:

Нужно сосредоточиться на слабых сезонах, месяцах и днях недели:

Увеличить промо-акции в начале недели – В ПН, ВТ и СР наблюдаются небольшой спад по продажам, его нужно нивелировать либо промо-акциями, либо более мягким ценообразованием и скидками (возможно псевдоскидками). Особенно сделать упор на Понедельник, поскольку именно в этот день недели минимальные продажи – 116321.

Также стоит уделить самому проблемному месяцу – Июлю. В Июль слишком мало продаж - 47322. Конечно, возможно это из-за сезона отпусков, однако стимулирующие акции должны быть введены в этот месяц и проверить гипотезу.

Сезонность. С ней можно бороться, продавая блюда, которые обычно популярны в это время года. Например, запустить продажу мороженного, новогодних блюд и шашлыков в разное время года, если этого еще нет.

Продаж без промо больше – 524082 (50.4), чем с промо 358933 (40.6%)

Прощенные значения:

Пропущенных значений нет