**Проект SAS. Краузе Наталья и Демченко Карина.**

**Первый Этап обработки данных в таблице**

Инструменты: notebook and libraries pandas, numpy

Ход работы:

Что мы хотим убрать из таблицы и почему:

1. Все строчки с “Новым Статусом” != “Доставлено”. Там есть “Статус” и “Новый статус”, выбрали новый статус. Потому что нам в общем интересно что в итоге с товаром, а если новый статус какой-то неопределенный то для витрины он не имеет особого смысла.

Анализ1:

df = df[(df['НовыйСтатус'] != 'Доставлен') & (df['Статус'] == 'Доставлен')]

Есть строка 'НовыйСтатус' = ‘К отгрузке’, а 'Статус' = ‘Доставлен’

Значит, ориентироваться надо на 'Статус'

df = df[((df['НовыйСтатус'] == 'Доставлен')] - получилось 359207 rows

Итог1:

df = df[((df['Статус'] == 'Доставлен')] - получилось 530029 rows

1. Удалить все строки с количеством доставленного товара == 0

df = df[((df['Количество'] > 0)] - получилось 516157 rows

1. Строки в колонке ‘Отменено’, где стоит “да ” - нам не интересные покупатели, которые отменили заказ для составления витрины.

df = df[(df['Отменено'] != 'Да')] - получилось 511529 rows

1. Строки если сумма заказа = 0 или сумма документа = 0, тк как нам не интересно это для составления витрины.

df = df[(df[СуммаЗаказаНаСайте] != 0) & (df[СуммаДокумента] != 0)] - получилось - получилось 511529 rows

Итог2: Цифра не изменилась, значит, ошибок по этому пункту в таблице не было.

1. Удалить все строки с количеством продано клиенту == 0

df = df[((df['КоличествоПроданоКлиенту'] > 0)] - получилось 508555 rows