МИНИСТЕРСТВО ЦИФРОВОГО РАЗВИТИЯ,  
СВЯЗИ И МАССОВЫХ КОММУНИКАЦИЙ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное бюджетное образовательное учреждение высшего образования  
«ПОВОЛЖСКИЙ ГОСДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ТЕЛЕКОММУНИКАЦИЙ И ИНФОРМАТИКИ»

Факультет «Дистанционного обучения»

**ЛАБОРАТОРНАЯ РАБОТА №1**

Тема: «Предобработка данных в текстовом редакторе и редакторе таблиц»

|  |  |  |
| --- | --- | --- |
| Проверил: |  |  |
|  |  |  |
| Выполнил: |  | студент 3-го курса группы ИСТ-90Д Соловьёва М.И. |

Самара, 2022 г.

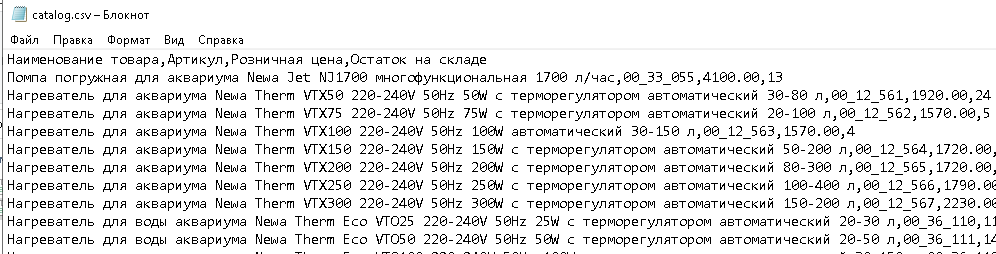
**Лабораторная работа №1**

**Цель работы:** получить базовые навыки предобработки данных в текстовом редакторе и редакторе таблиц.

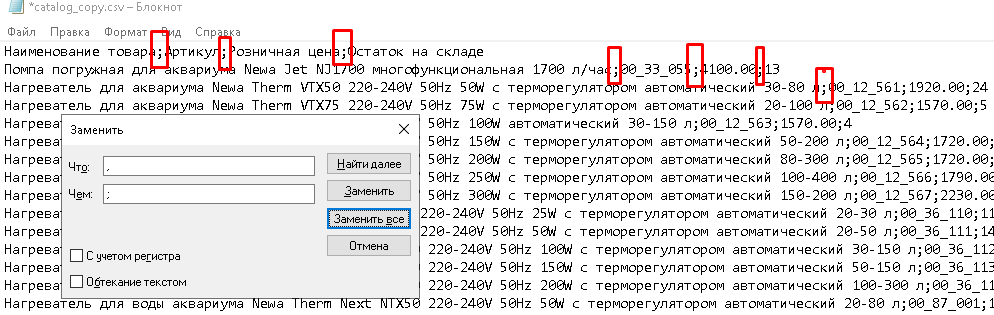
Ход выполнения работы:

1. Скачиваем с сайта <https://global-torg.com/services/csv/> файл catalog.csv.

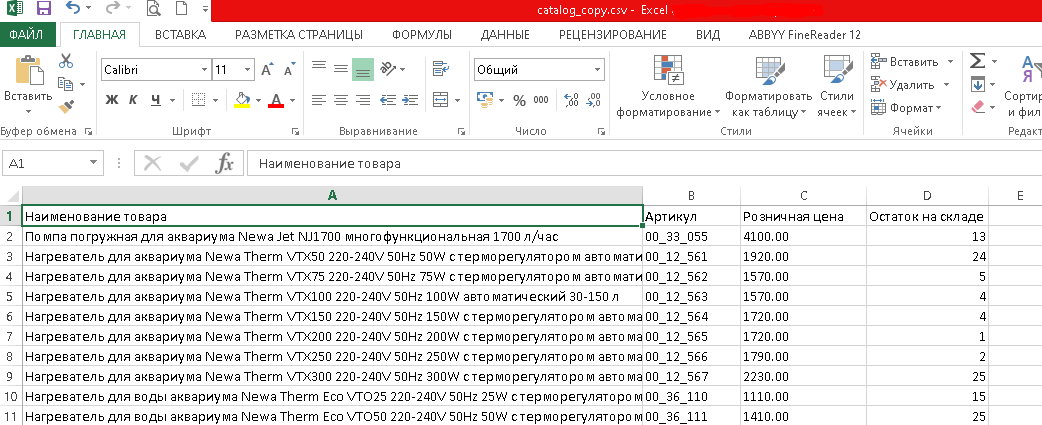
2. Делаем его копию: catalog\_copy.csv. Открываем в текстовом редакторе (Блокнот):



В качестве символа-разделителя выступает « , ». Производим ее замену на точку с запятой « ; »:



3. Открываем измененный файл в MS Excel:

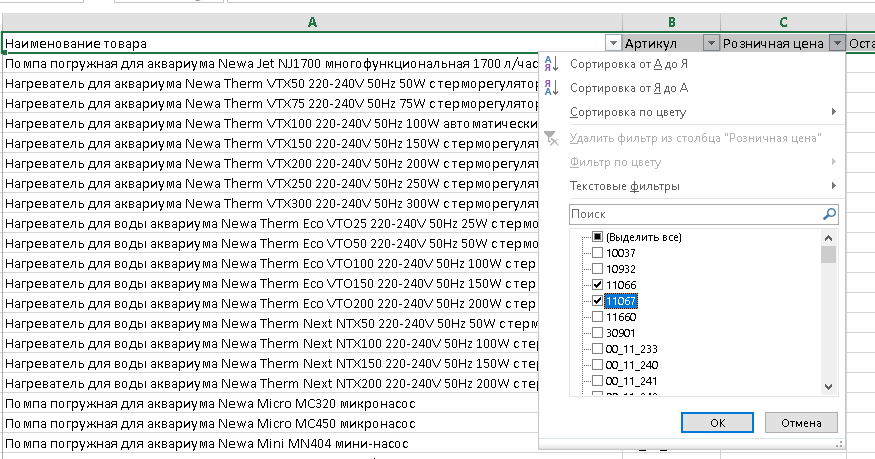


4. Разбиение по столбцам выполнено.

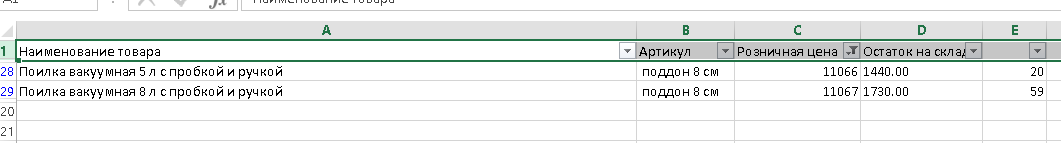
5. Названия столбцов в наличии.

6. Скриншот приложен в пункте 3.

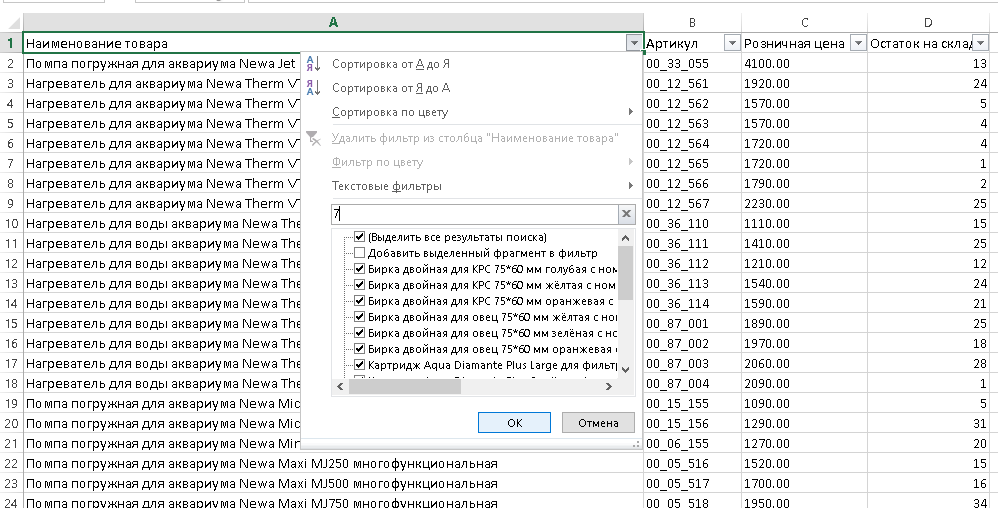
7. Включаем фильтр столбцов и производим фильтрацию по двум параметрам:



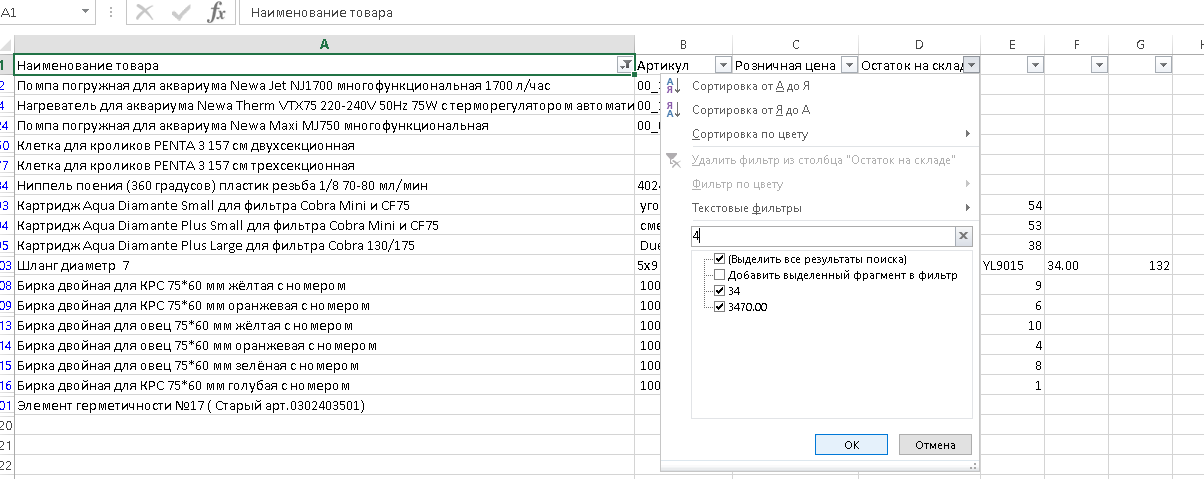
Результат фильтрации



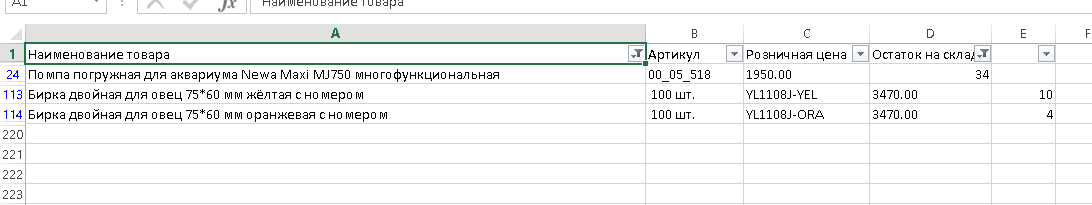
Вводим значение «7» в строку поиска первого столбца:



Здесь же в 4 столбце вводим в строку поиска «4» и производим фильтрацию:

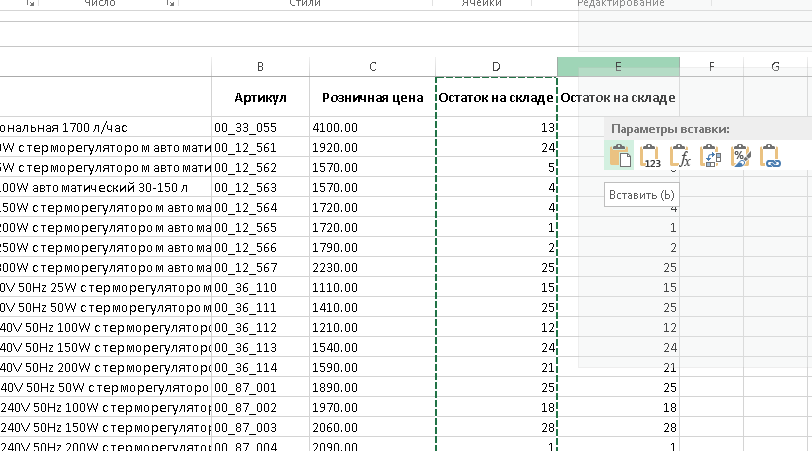


Результаты двойной фильтрации:

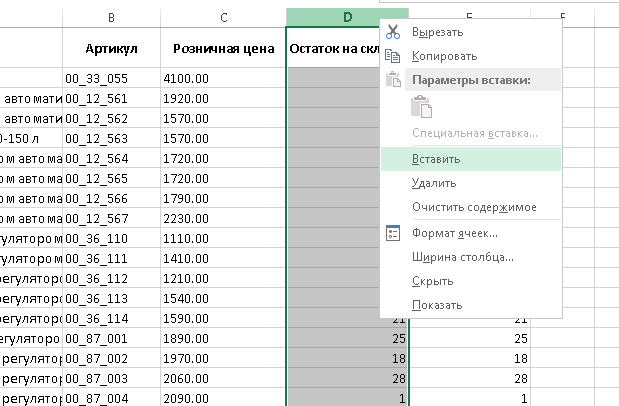


8. Добавляем дополнительный столбец:

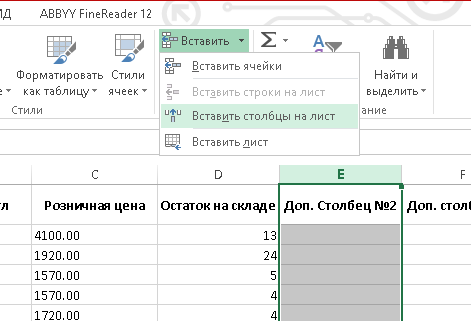
- способ 1 – копируем столбец и, выделив соседний, вставляем. Получаем точную копию:



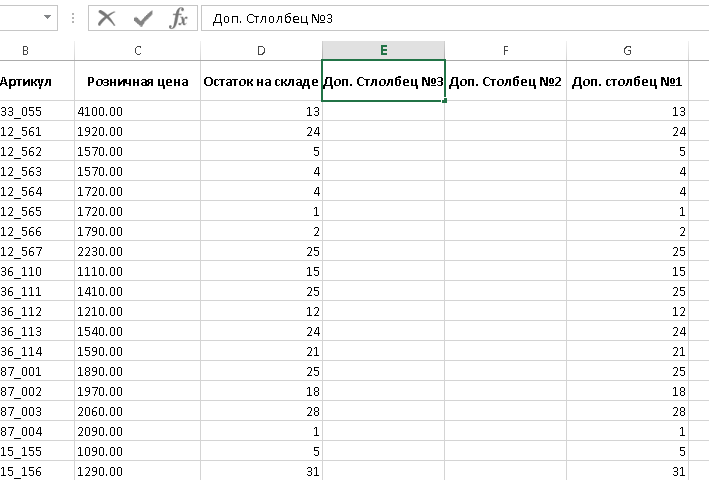
- способ 2 – вставка через контекстное меню



- способ 3 – вставка через основное меню, выделив предварительно столбец, левее которого будет помещено дополнение

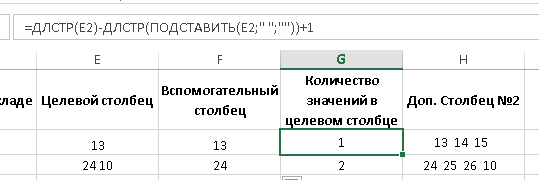


В итоге получилось:

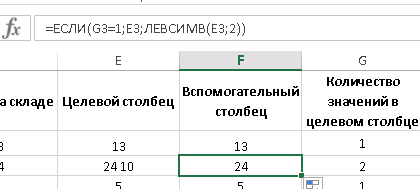


9. Определяем Доп. столбец №1 целевым столбцом

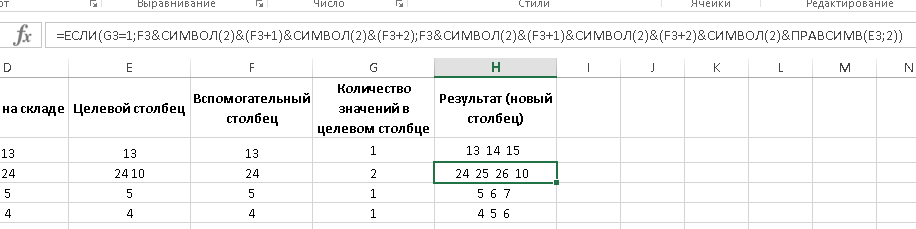
Для упрощения формул столбец со значением количества значений в целевом столбце,



и вводим вспомогательный столбец, в котором будем отделять первое значение из целевого столбца, а если оно одно, то его и прописывать.

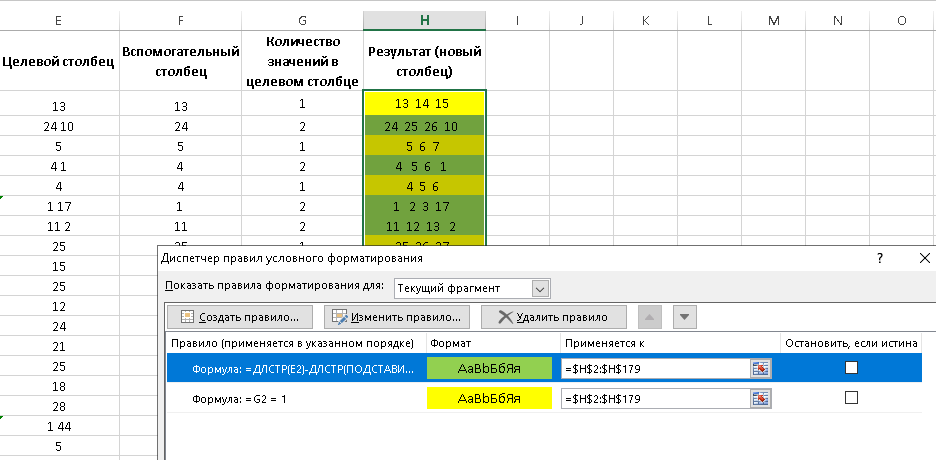


В результате (Новый столбец) выводим окончательное количество значений через пробел



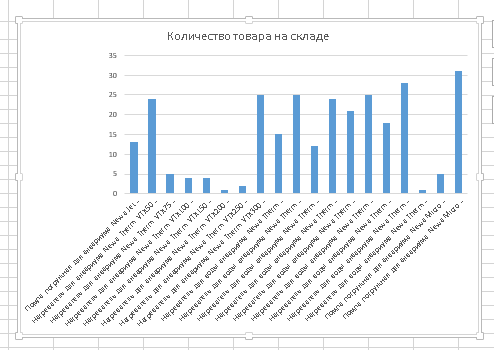
Здесь проверяется количеств значений в целевом столбце и в зависимости от этого первое значение неизменно, ко второму плюсуется 1, а к 3 – число 2. Если же изначально было 2 значения, второе записываем неизменным.

10. Раскрашиваем ячейки двумя способами в зависимости от количества значений в целевом столбце. Способы отражены в диспетчере правил условного форматирования.

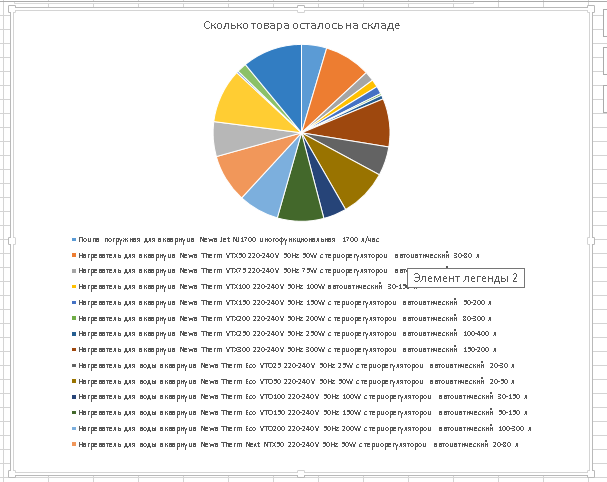


11. Строим гистограммы.

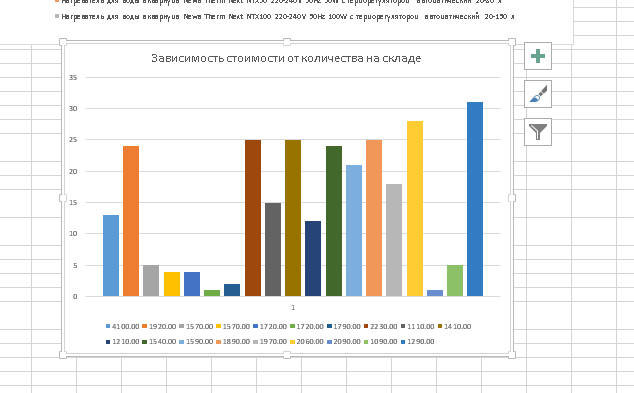
11.1 Гистограмма количества товара на складе:



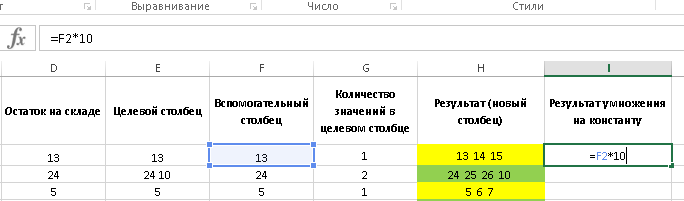
11.2 Гистограмма круговая количества товара на складе



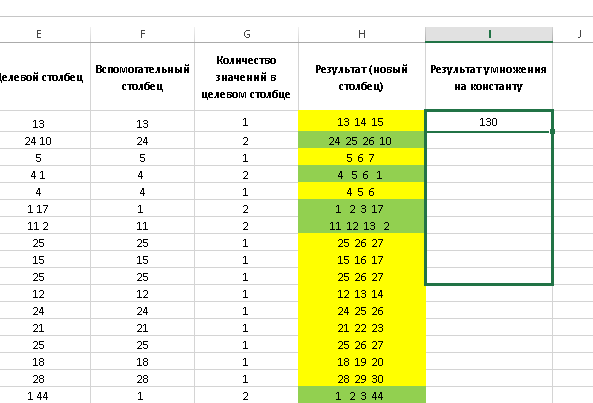
11.3 Гистограмма зависимости стоимости от количества товара на складе



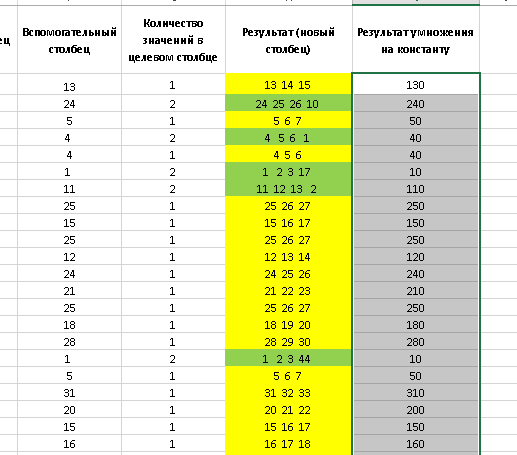
12. Выполняем умножение на константу для произвольного столбца из файла. Для этого значения из столбца «Вспомогательный столбец»(F) умножим на число 10 (чтобы было проще проверить правильность). Результат разместим в столбце «Результат умножения на константу»



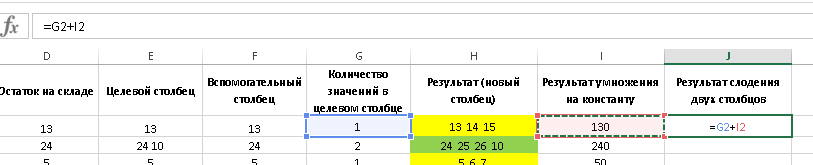
Цепляем результат за точку в правом нижнем углу и тянем вниз по столбцу:



Получаем результат:



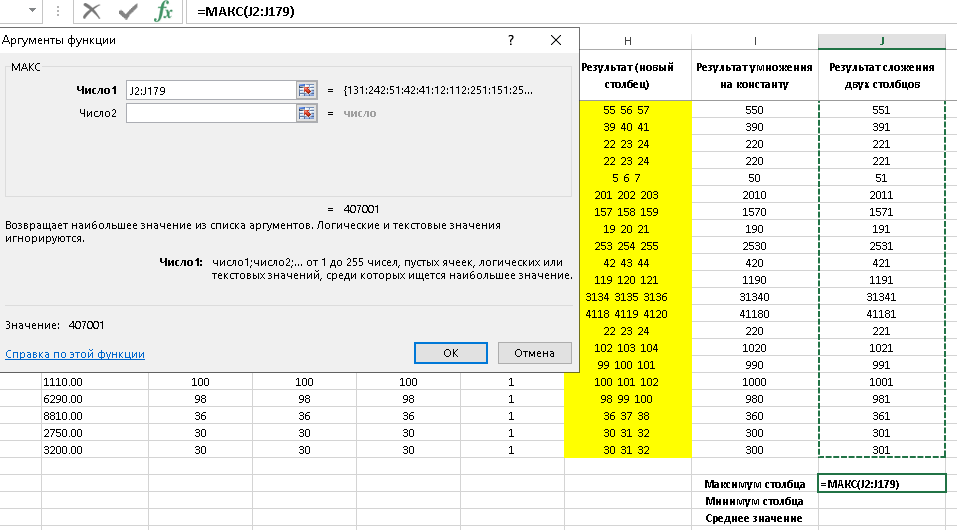
13. Выполняем сложение двух столбцов. Для этого создаем новый столбец «Результат сложения двух столбцов». Для ячеек этого столбца прописываем формулу.



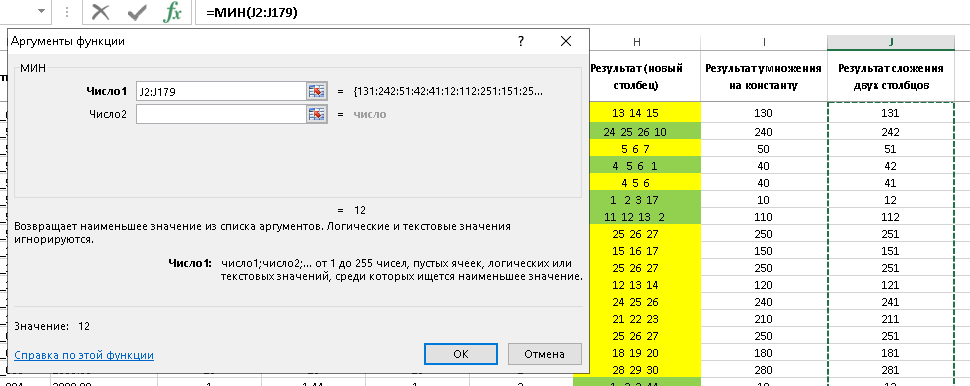
По вышеуказанному принципу растягиваем формулу на все ячейки столбца:

14. Вычислим максимальное, минимальное и среднее значение для столбца «Результат сложения двух столбцов». Для этого снизу, после завершения всех данных сделаем соответствующие надписи и пропишем формулу для вычисления каждой величины:

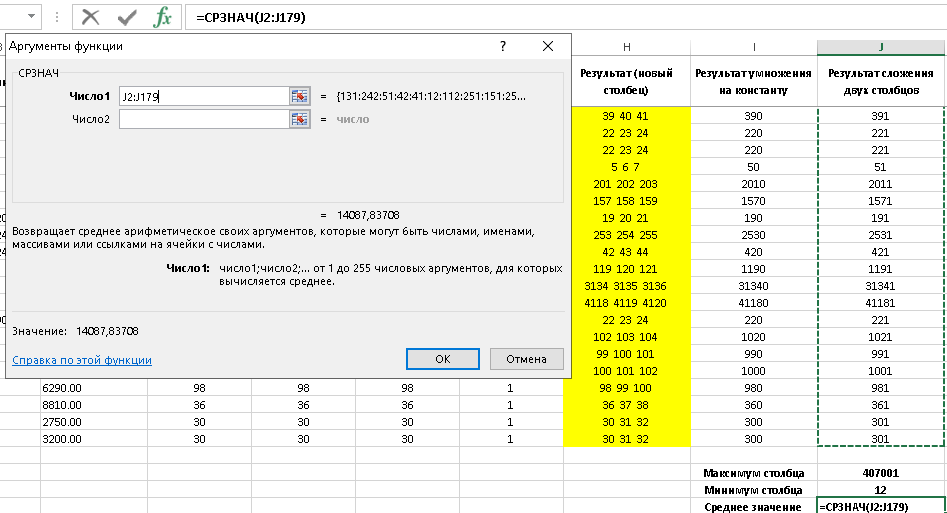
- максимальное:



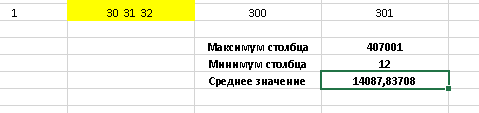
- минимальное:



- среднее значение:

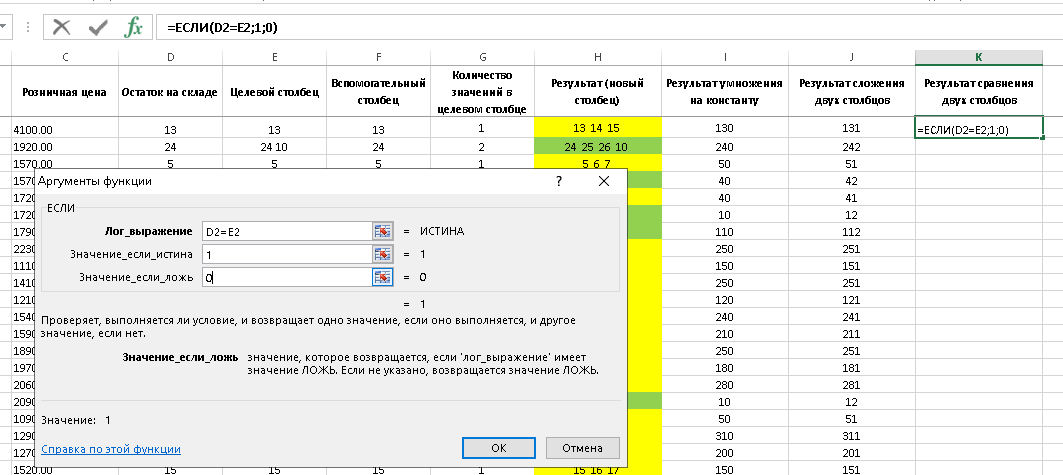


Итоговый результат:



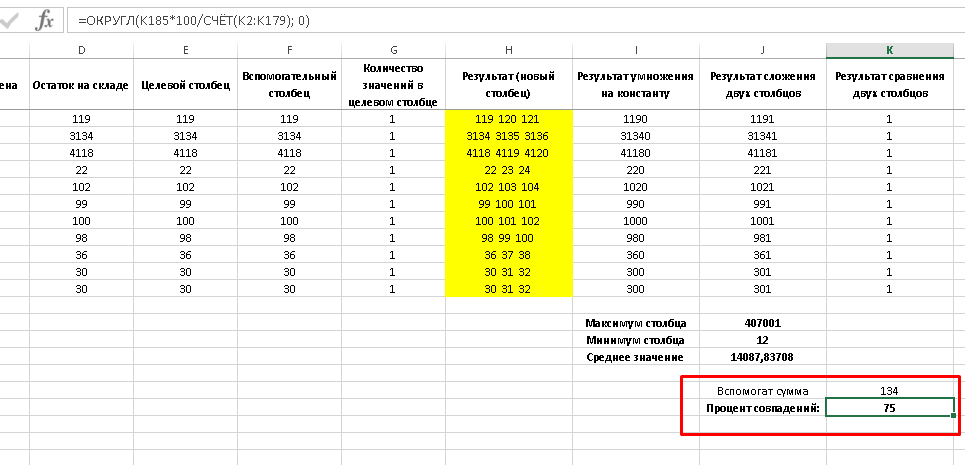
15. Сравниваем значения двух столбцов и вычисляем процент совпадений. Для этого создаем столбец «Результат сравнения двух столбцов». Сравнивать будем «Вспомогательный столбец» и «Целевой столбец». Причина такого выбора в большом количестве совпадений данных.

В результирующем столбце пропишем условие, при соблюдении которого в ячейке будет выводиться 1, если же условие false – 0.

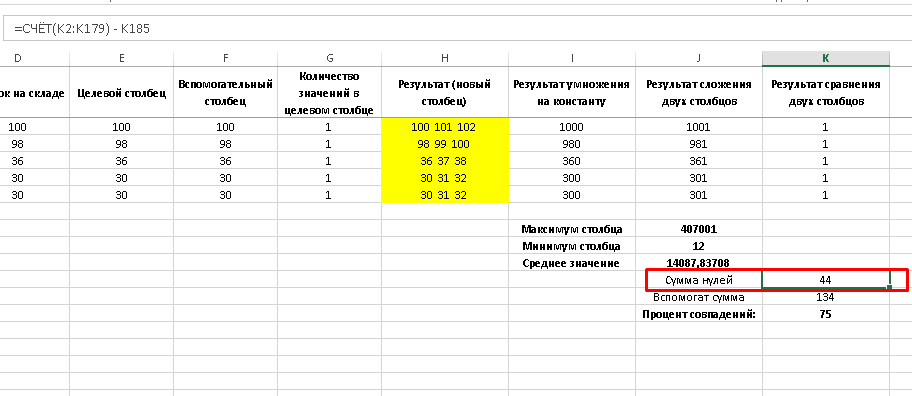




Вычисляем процент совпадений. Для этого вводим ячейку «Вспомогательная сумма», которая считает количество единичек в столбце, то есть – количество совпадений. Далее в строке ниже производим основной расчет. Для этого изначально результат округляем до ближайшего целого числа, затем прописываем формулу расчета обычной пропорции, в знаменателе которой прописан расчет количества непустых ячеек из рассматриваемого диапазона.



16. Считаем процент несовпадений. Для этого вводим дополнительную ячейку «Сумма нулей», в которой из количества непустых ячеек вычитаем сумму всех единичек.

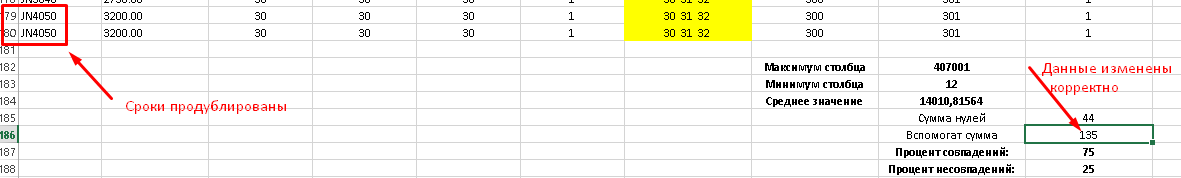


Далее производим расчет аналогично пункту 15

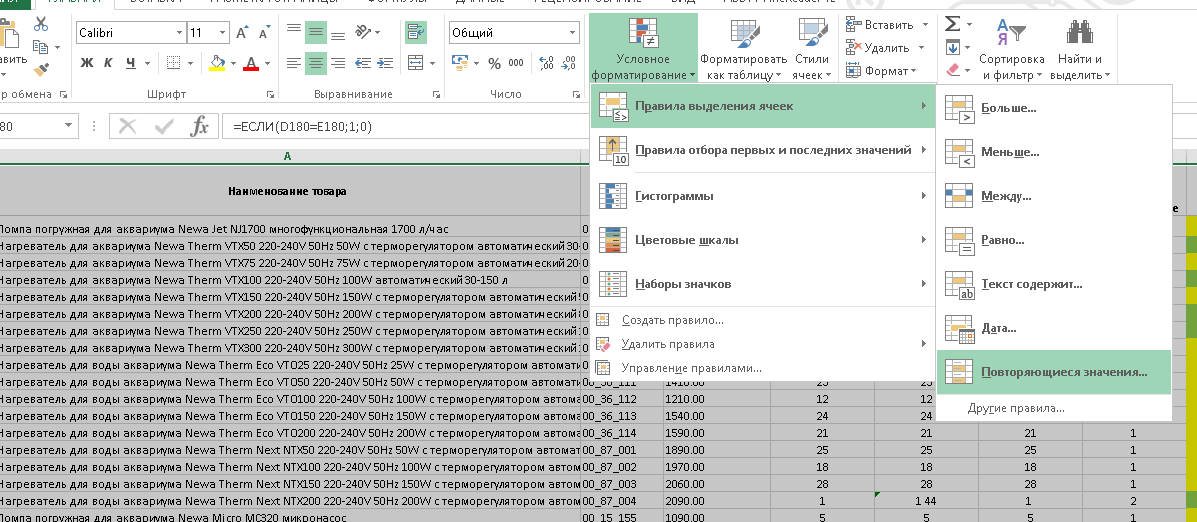


Действительно, сложив проценты совпадений и несовпадений, получаем 100%, что подтверждает правильность расчетов.

17. Производим поиск дубликатов средствами текстового редактора. Для этого создадим дубликат последней строки, чтобы было удобнее контролировать неизменность рассчитанных выше данных.



Выделяем таблицу, далее кликаем на «Условное форматирование»



Выявляем наши дубликаты:



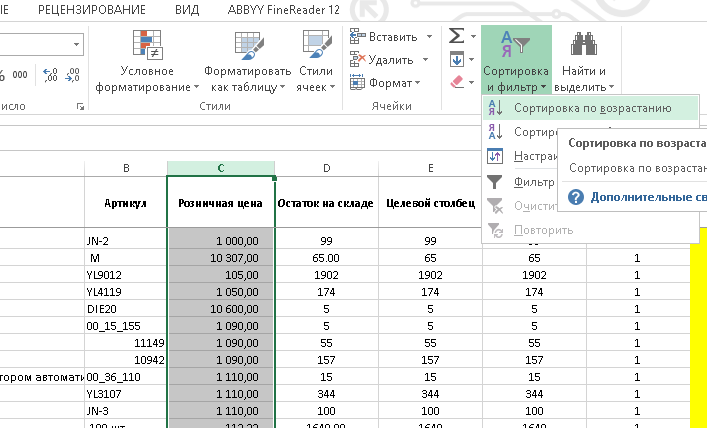
Теперь, когда видны дубликаты, можно принять решение на удаление или внесение изменений.

18. Выполняем поиск выбросов данных.

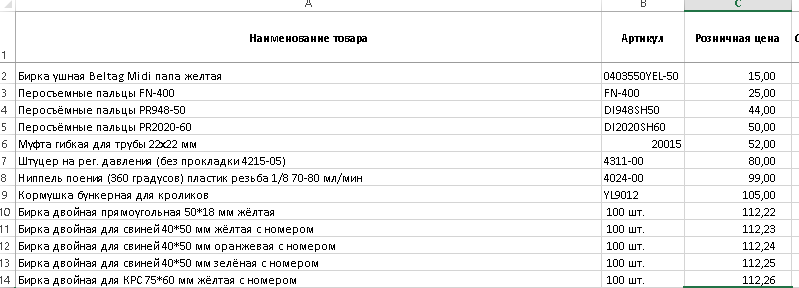
Выброс – это значение, которое значительно выше или ниже, чем большинство значений в ваших данных. При использовании Excel для анализа данных выбросы могут искажать результаты. Например, среднее значение набора данных может действительно отражать ваши значения.

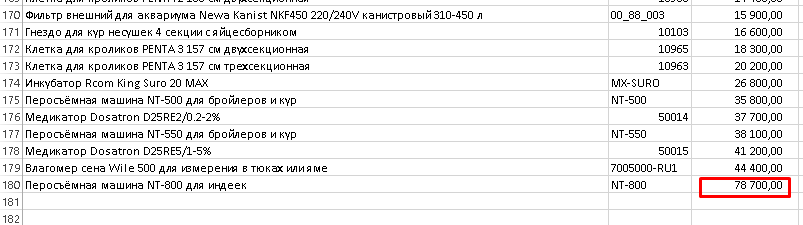
Для выявления выбросов можно использовать сортировку данных в необходимом столбце. И проанализировать самые большие и самые маленькие значения, сравнив их со следующими.

Я возьму столбец «Розничная цена»



Производим сортировку по возрастанию и анализируем показатели в начале и конце таблицы.





В конце таблицы видим значительное превышение стоимости товара (почти в 2 раза). Данный показатель может повлиять на нашу статистику, поэтому при расчете определенных показателей (максимум, среднее значение) есть смысл его исключить из расчетов.

**Вывод по работе.**

Проведенная работа показала множество возможностей, которые предоставляет пользователем MS Excel. Программа действительно позволяет решать массу задач от простейших расчетов до сложнейших вычислений и анализа данных с построением всевозможных графиков и диаграмм, полностью автоматизируя вычислительные процессы и выдавая пользователю результат в любой удобной для него форме или импортируя этот результат в любые офисные программы.

Рассмотренные примеры – это лишь крохотный кусочек этой программы, профессиональное освоение которой откроет перед пользователем массу возможностей, сделав его весьма конкурентным претендентом на рынке труда практически во всех отраслях.

Приложения:

1. Файл catalog.csv – исходный файл.

2. Файл catalog\_copy.csv – сохраненный файл согласно п.2 задания лабораторной работы.

3. Файл catalog\_copy.xlsx – открытая в Excel копия файла из п.2, в котором проводилась вся работа.