#### Министерство науки и высшего образования Российской Федерации

#### Федеральное государственное бюджетное образовательное учреждение

#### высшего образования

#### «Владимирский государственный университет

#### имени Александра Григорьевича и Николая Григорьевича Столетовых»

**(ВлГУ)**

**Кафедра информационных систем и программной инженерии**

Лабораторная работа №1

по дисциплине

"Бизнес анализ данных"

Выполнил:

ст. гр. ПРИм-124

Парахин К.В.

Приняла:

доц. Кафедры ИСПИ

Озерова М.И.

Владимир, 2025 г.

Часть 2. Анализ ограниченного массива данных с помощью инструмента "Сводная таблица (Pivot table)"

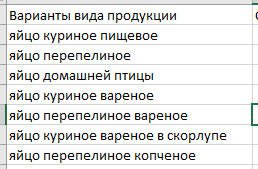
Предлагаемая выборка содержит реальные данные системы электронной сертификации продукции животного происхождения (мясо, птица, рыба, молоко, мёд и производные продукты). В сутки по стране оформляется около 13 млн. электронных сертификатов, выборка для анализа содержит 500 тыс. эВСД. эВСД – электронный ветеринарный сопроводительный документ (сертификат), который оформляется на каждую партию продукции при производстве, перемещении по территории РФ и/или переходе права собственности (продажа).

В результате анализа этой выборки необходимо извлечь и описать полученные знания об особенностях товарооборота в РФ. Количество и перечень факторов для анализа не ограничен, выбирайте сами, какие и сколько факторов разобрать, что интересно и показательно. Результаты анализа с вашими выводами и комментариями нужно отразить в виде аналитического отчета в формате PDF. Отчет должен содержать не менее 2 элементов визуализации данных каждого вида: таблица, гистограмма, круговая диаграмма, график.

Выполнение работы

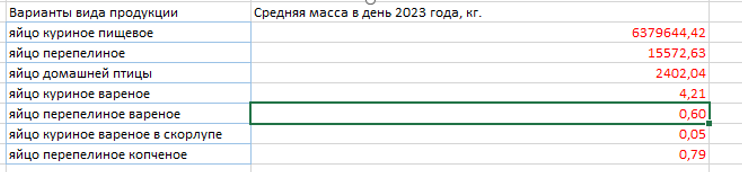
В рамках первого документа имеется выборка данных по различным видам яиц за 11 месяцев 2023 года

Для начала нужно получить уникальный список видов яичной продукции:



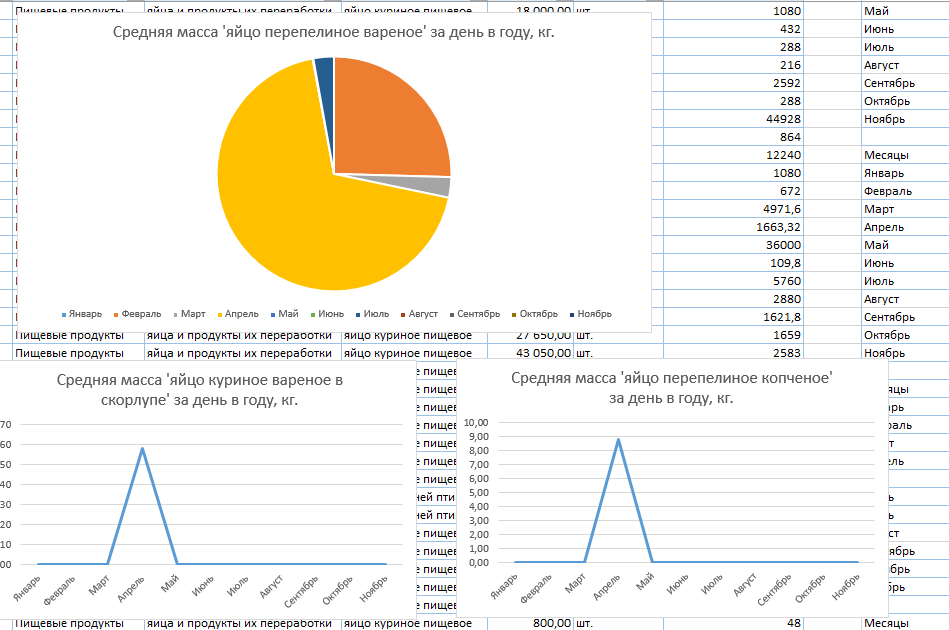
Далее как оказалось в некоторых строках для яиц указаны единицы измерения в кг, а в других – в штуках. Для единообразия привел все к массе в кг, для этого создал отдельный столбец и просчитал массу – для куриных яиц использовал 0,06 кг, для перепелиных – 0,011 кг.

Далее используя этот столбец просчитал общую массу для каждого вида продукции за 11 месяцев, поделив на 334 дня получил среднюю массу за день 2023 года



Далее раскидал выборки по видам яиц дискретно по каждому виду яйца и месяца – и учел разное кол-во дней в каждом месяце – получил распределение средней массы каждого вида продукции в день определенного месяца, с января по ноябрь 2023 года, оформил визуальное представление с помощью разных диаграмм: 2 гистограмм, 2 линейчатых, круговой и точечной диаграммы

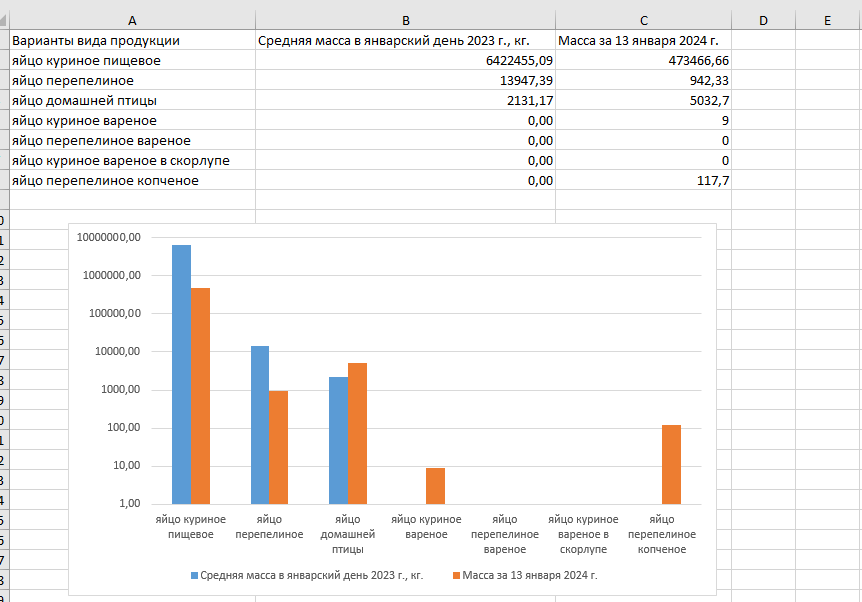




После этого использовал второй документ (.csv) – в котором находились выборки по всем единицам продукции, не только яичной, за конкретный день – 13 января 2024 года.

Аналогично прошлому файлу – привел все строки с яичной продукцией к одной единице измерения – масса, кг. Далее аналогично просчитал для каждого вида продукции массу в этот день (скриншот утерян, так как при переоткрытии сохраненного .csv файла вся разметка ячеек сломалась)

Но данные сами я вынес в отдельный .excel файл



Я специально соотнес данные для средней массы в кг. каждого вида продукции – слева за средний январский день 2023 г., а справа – за конкретный день 13 января 2024 г.

Результаты сравнения получились достаточно странные – основных видов продукции: яйцо куриное и яйцо перепелиное было произведено (перемещено) в 10 с лишним раз больше в январский день 2023 года – чем в день экстренной выборки – 13 января 2024 г.  
На гистограмме с использованием логарифмической шкалы я изобразил для каждого вида продукции сравнение между 2 случаями – вывод какой отсюда делать, точно сказать не могу, думаю – либо не все данные были учтены в выборке за этот день, либо произошел какой то резкий провал перемещения яичной продукции.

Потому что если посмотреть диаграммы за 2023 год – видно, что уровень перемещения куриных и перепелиных яиц за каждый месяц года (точнее его средний дневной показатель) держится примерно на одной отметке.

Вывод

В результате выполнения работы было проведено изучение входного набора данных в виде большой сводной csv и excel таблицы, проведена визуализация и интерпретация полученных результатов.