Тема: Аналіз впливу рейтингу товару на продажі у сегменті електроніки

Опис проєкту: Мій PET-проєкт буде спрямований на аналіз впливу рейтингу товару та залежність продажів від нього в сегменті електроніки. Планую вивчити, як високий рейтинг товару впливає на його популярність серед споживачів та як це впливає на обсяги продажів.

Моя робота включатиме аналіз доступних даних з датасетів "E-commerce", "Smartphones Sales Dataset" та "Video Game Sales with Ratings". Досліджувати буду кореляцію між рейтингом товару та обсягами продажів, а також вплив інших факторів, таких як ціна, розмір пам'яті або жанр продукту, на цю взаємодію.

Метою мого проєкту буде розробка інсайтів для невеликих виробників та роздрібних мереж електроніки щодо того, як оптимізувати стратегії маркетингу та продажів на основі рейтингів товарів.

Технології до використання для проекту:

BigQuery (SQL): для роботи з великими обсягами даних, BigQuery буде корисним при їх початковій обробці та підготовці за допомогою SQL.

Jupyter Notebook (Python): Python разом з бібліотеками pandas, Matplotlib та Seaborn дозволить вам провести більш детальний аналіз даних, використовуючи програмування для автоматизації процесу та створення візуалізацій для більш детального аналізу.

Tableau або Power BI. інструменти що відмінно підходять для візуалізації даних та створення дашбордів. Вони дозволять зробити дані більш зрозумілими, легше сприйнятними для аудиторії та інтерактивними.

Використані датасети:

* [E-commerce Customer Behavior - Sheet1](https://drive.google.com/file/d/1MPI3Yywof3rCMi7kV6I7YQBYIkh8vimF/view?usp=drive_link)
* [S.Sales](https://drive.google.com/file/d/1LHcv5Q6ZdJsSoFjbwWklBpDhjdursJin/view?usp=drive_link)
* [Video\_Games\_Sales\_as\_at\_22\_Dec\_2016](https://drive.google.com/file/d/125dylxdXw3uJ3uX7e8nTH560928pXMSc/view?usp=drive_link)

Їх ресурси:

[Kaggle](https://www.kaggle.com/datasets/uom190346a/e-commerce-customer-behavior-dataset)(E-commerce Customer Behavior)

Власні дані

Парсинг

Загальний план роботи над проєктом:

* Етап 1: Підготовка і розуміння даних (22 березня - 23 березня)

Отримання та очищення даних: Завантаження даних з відповідного джерела (база даних у BigQuery).

Видалення непотрібних або некоректних даних.

Первинний аналіз даних: Використання Excel/ BigQuery для швидкого аналізу загальних показників та виявлення основних трендів.

* Етап 2: Аналіз та підготовка даних (23 березня - 24 березня)

Більш детальний розгляд даних за допомогою Python: Використання Jupyter Notebook для обробки та аналізу даних з використанням внутрішніх бібліотек.

Вирішення конкретних завдань аналізу

Підготовка даних для візуалізації: Фільтрація та групування даних за необхідними критеріями.

* Етап 3: Візуалізація та дашборди (24 березня - 27 березня)

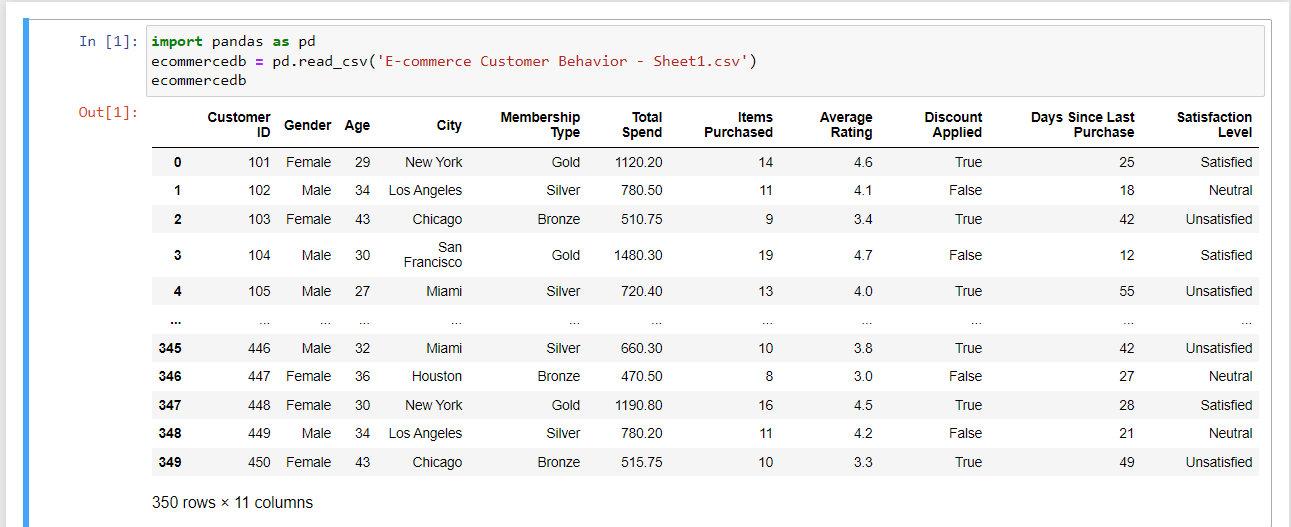
Створення візуалізацій за допомогою Tableau або Power BI: Створення графіків, діаграм та інтерактивних візуалізацій для кращого розуміння даних.

Створення дашбордів: Компонування візуалізацій у дашборди для зручного представлення результатів та аналізу.

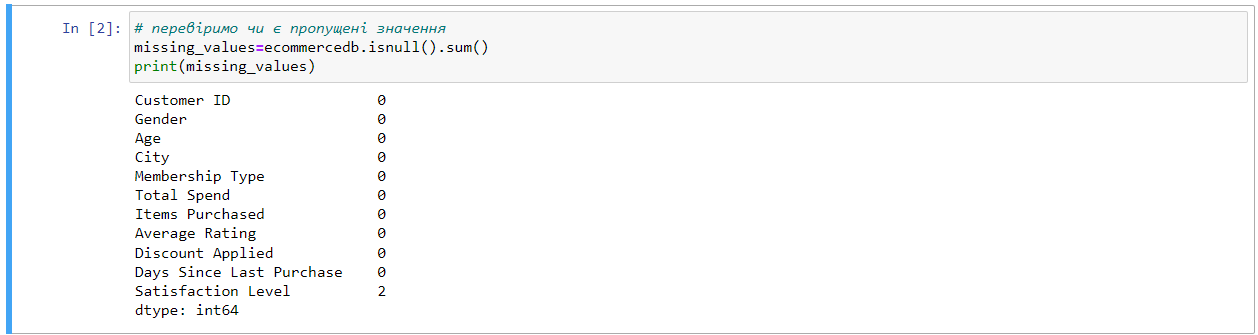
ПРОЕКТ

**Jupyter Notebook (Python)**

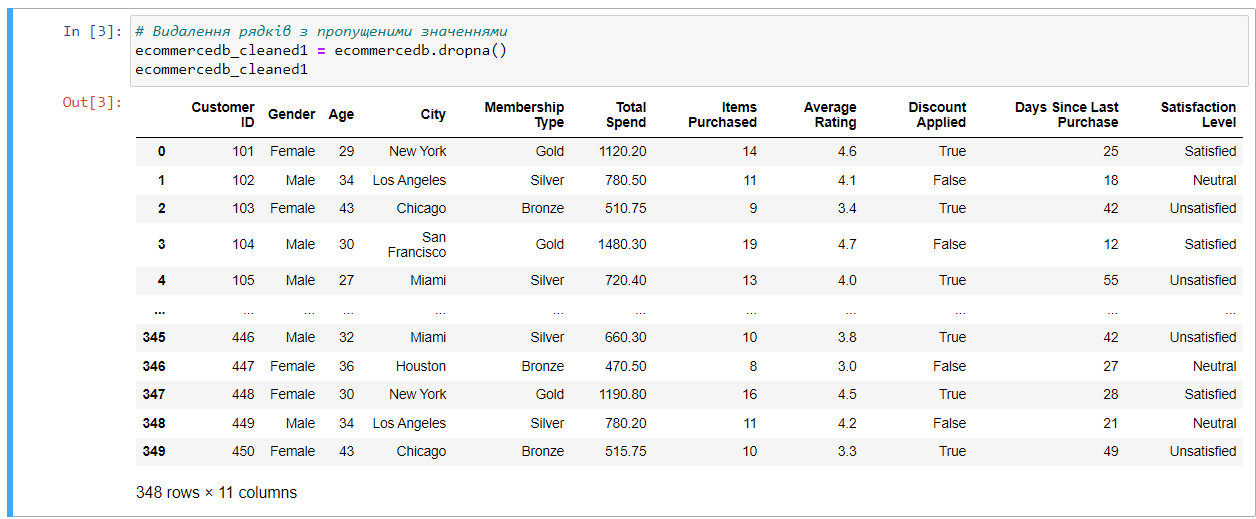
Спочатку для аналізу даних в рамках мого PET-проекту я використовуватиму Python через середовище Jupyter Notebook. Першим кроком буде завантаження та підготовка даних з дата сету, які включають інформацію про покупців, їх активність, їх рейтинги та покупки.



Після завантаження даних з дата сету, перевіримо наявність пропущених значень в різних стовпчиках. Цей етап є важливим для перевірки цілісності перенесених даних і забезпечує чистоту та точність аналізу.



Бачимо, що в стовпці "Satisfaction Level" наявні 2 пропущених значення, тому переходимо до очистки даних шляхом видалення рядків з пропущеними значеннями.

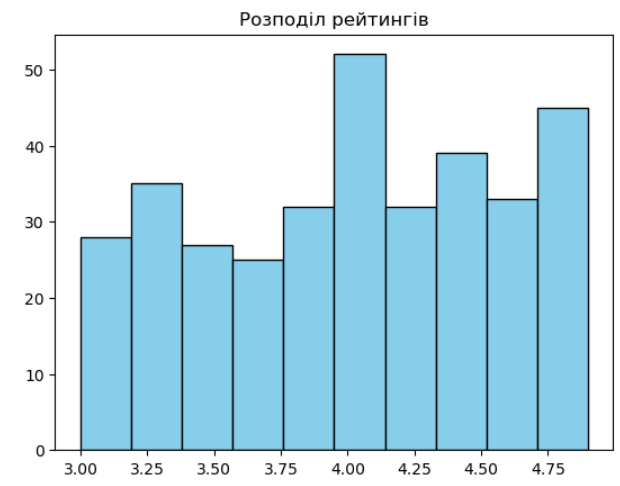


Від пропущених значень переходимо до перевірки наявності дублікатів.



Очистивши дані, ми можемо перейти до обчислення статистичних показників рейтингу та побудови графіку розподілу рейтингів з підготовленого дата сету.





Загально результати аналізу показують, що більшість товарів мають високий рейтинг близько 4.0+. Це свідчить про переважну задоволеність споживачів якістю товарів.(детальний аналіз за візуалізаціями).

Оцінимо ближче, яка кількість оцінок по кожному рейтингу. Виберемо кількість для кожного рейтингу та розсортуємо їх за рейтингом.



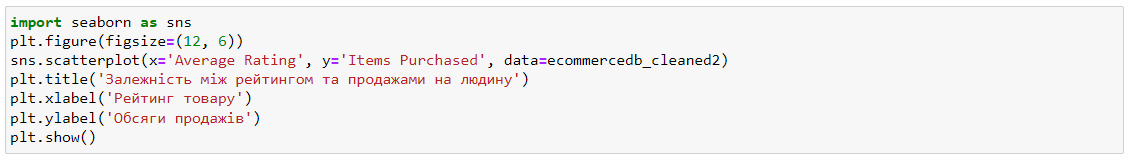
Товари з рейтингом 4.1 мають найбільшу кількість оцінок - 34, загалом маємо близько-однакові значення, проте позитивні рейтинги трішки вищі.

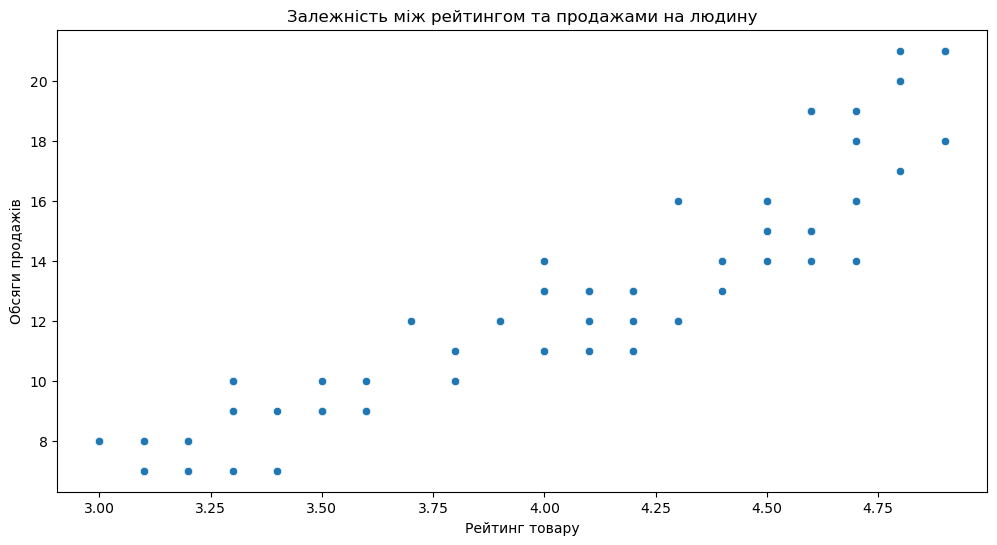
Перейдемо до кореляції між середнім рейтингом і кількістю куплених товарів. Кореляція дозволяє нам оцінити, наскільки сильно зв'язані ці дві змінні: чим ближче до 1 або -1 значення кореляції, тим сильніший зв'язок між ними.



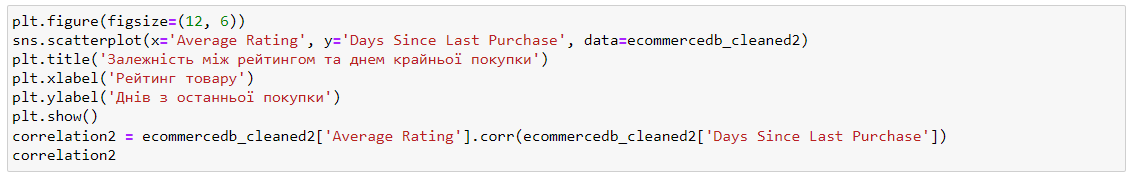
0.9214, це свідчить про сильний позитивний зв'язок між цими двома змінними. І це означає, що товари з вищим рейтингом потенційно мають тенденцію до більшого обсягу продажів.

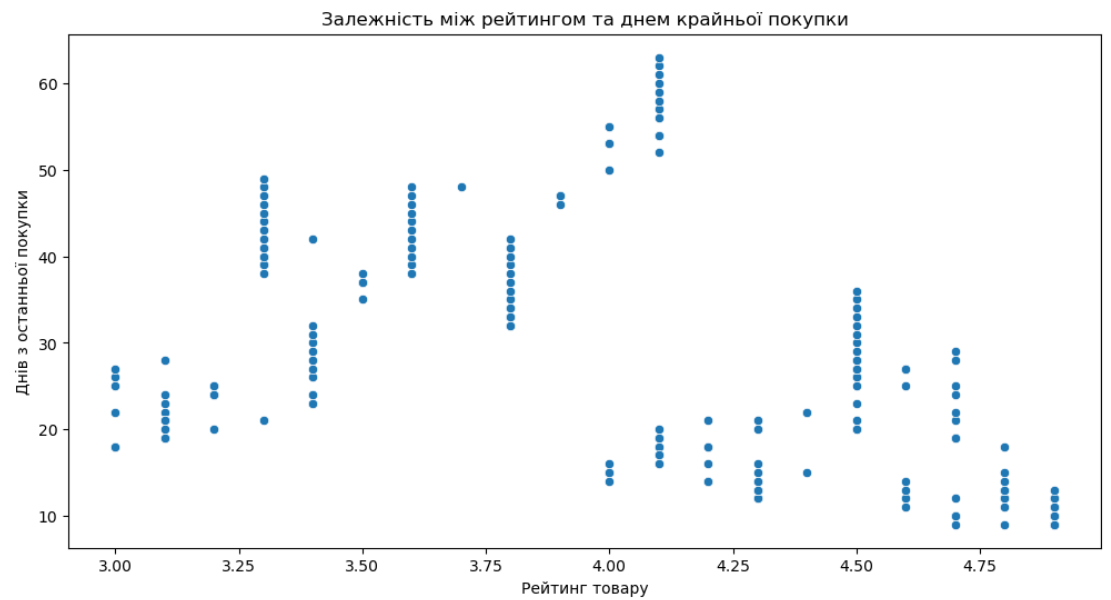
Для візуального аналізу зв'язку між середнім рейтингом товарів та їх кількістю продажів на людину ми використовуємо діаграму розсіювання (scatterplot). Ця візуалізація дозволить нам відобразити взаємозв'язок між двома змінними і побачити, як змінюється кількість продажів зі зростанням рейтингу товару.





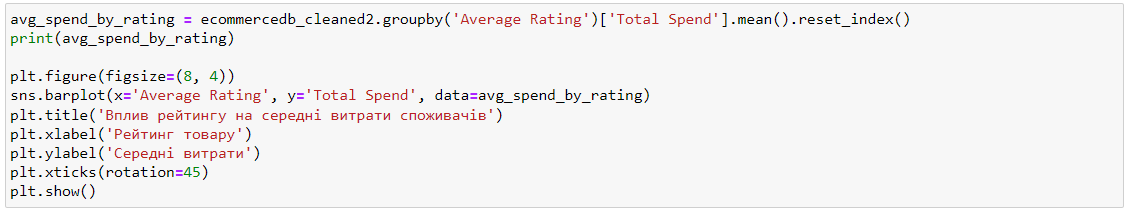
Також використовуємо діаграму розсіювання для аналізу зв'язку між середнім рейтингом товарів та кількістю днів з останньої покупки.

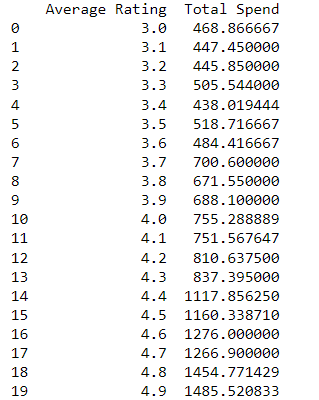


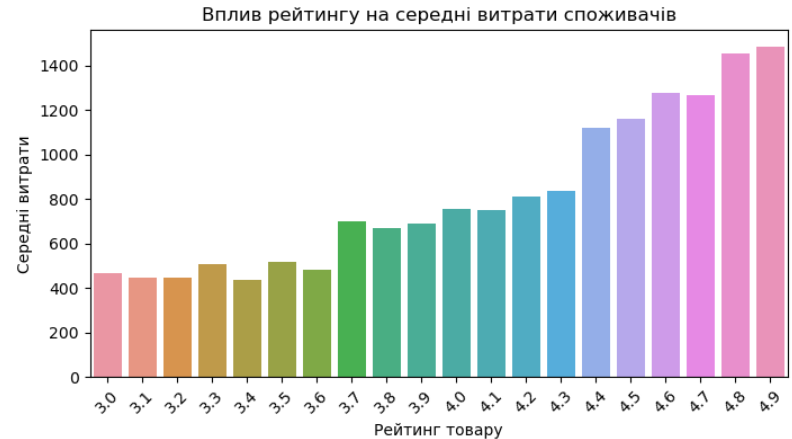


Отримане значення кореляції -0.4359 показує, що існує помірний негативний зв'язок між середнім рейтингом товарів та кількістю днів з останньої покупки. Це означає, що зі зростанням рейтингу товарів спостерігається ймовірна тенденція до зменшення кількості днів між покупками у споживачів.

Для дослідження впливу середнього рейтингу товарів на середні витрати споживачів, ми використовуємо групування за рейтингом та обчислення середніх витрат. Після цього ми візуально оцінимо результати за допомогою стовпчастого графіка.





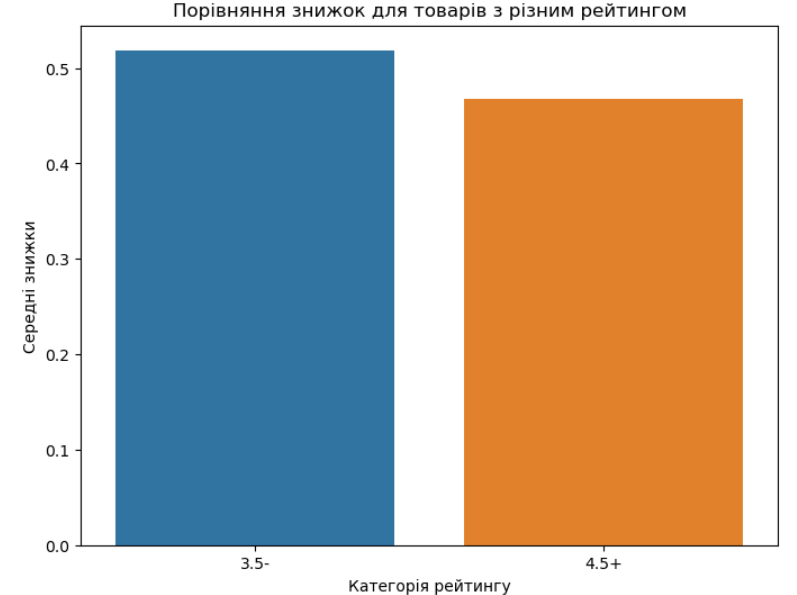


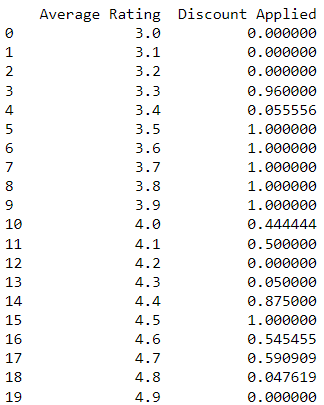
Для дослідження впливу рейтингу товарів на середні знижки, ми використовуємо категоризацію рейтингів та обчислення середніх значень знижок для кожної категорії, після ми візуалізуємо результати за допомогою стовпчастих графіків з метою порівняння знижок для товарів з різним рейтингом.

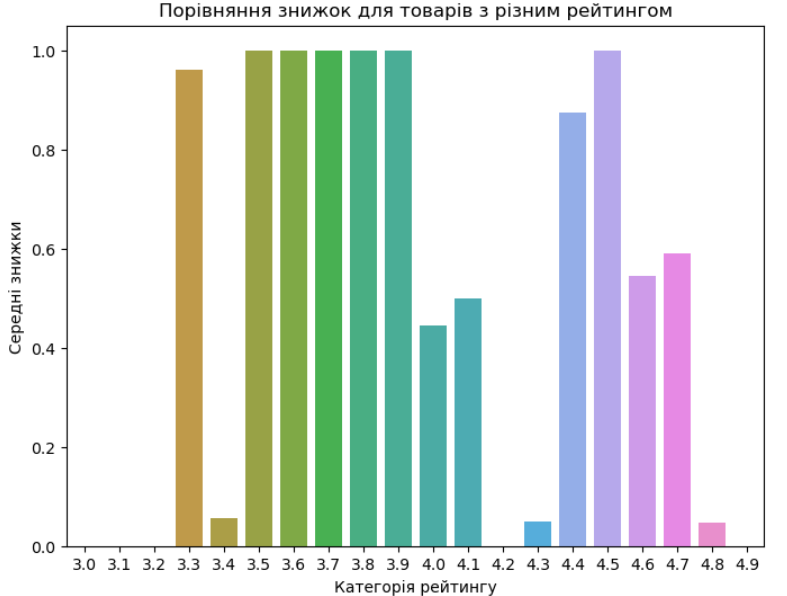
Rating Category Discount Applied

3.5- 0.518828

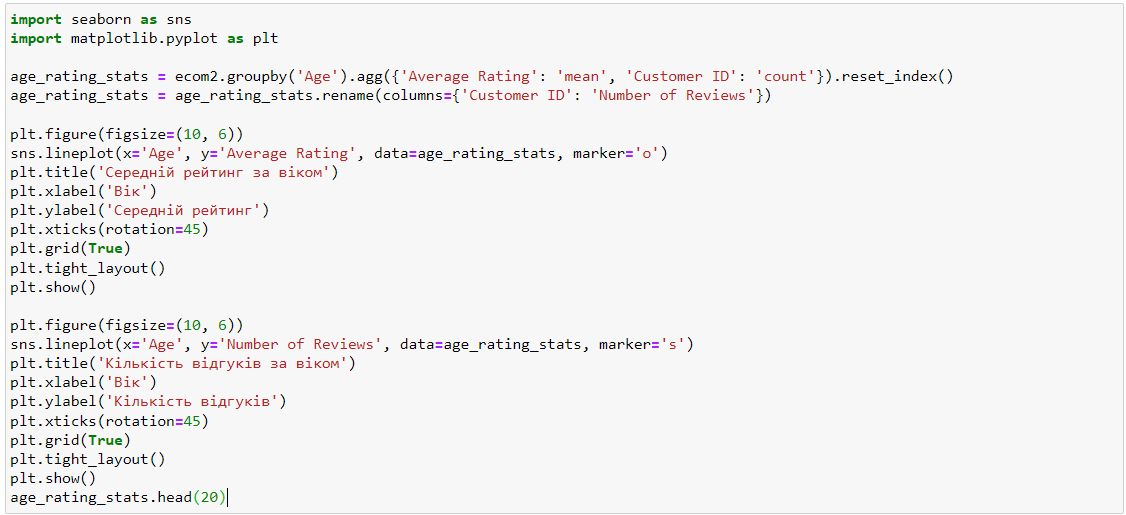
4.5+ 0.467890

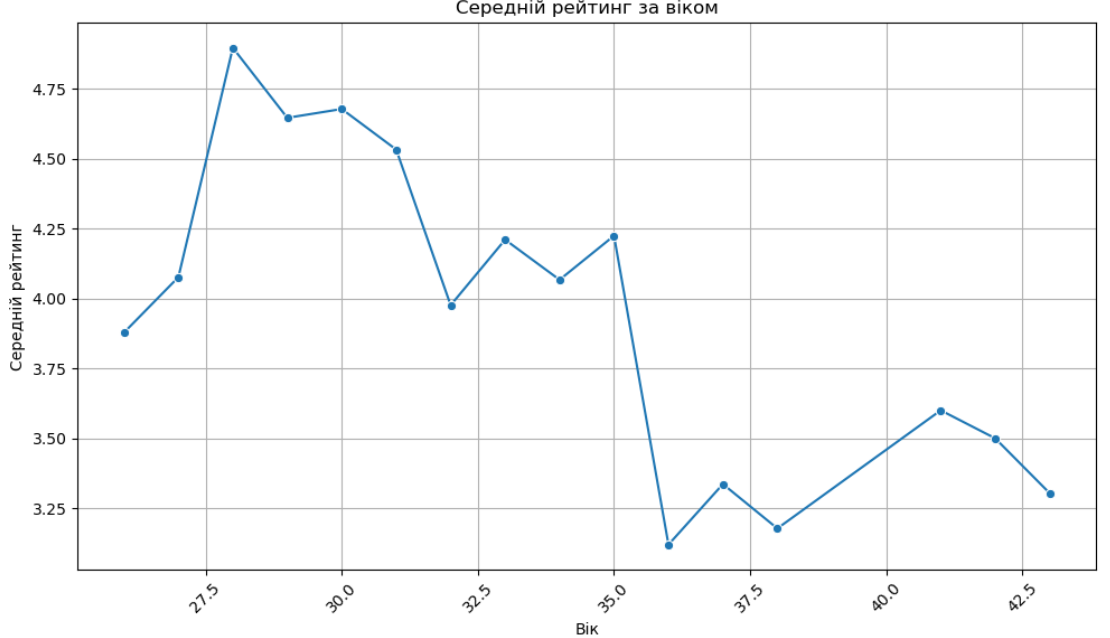




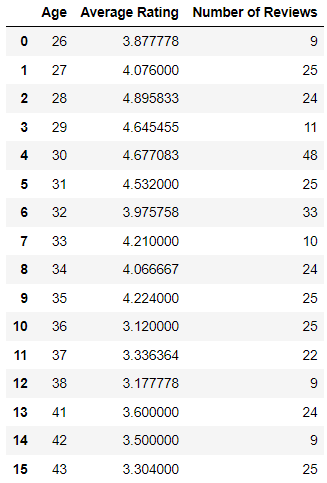


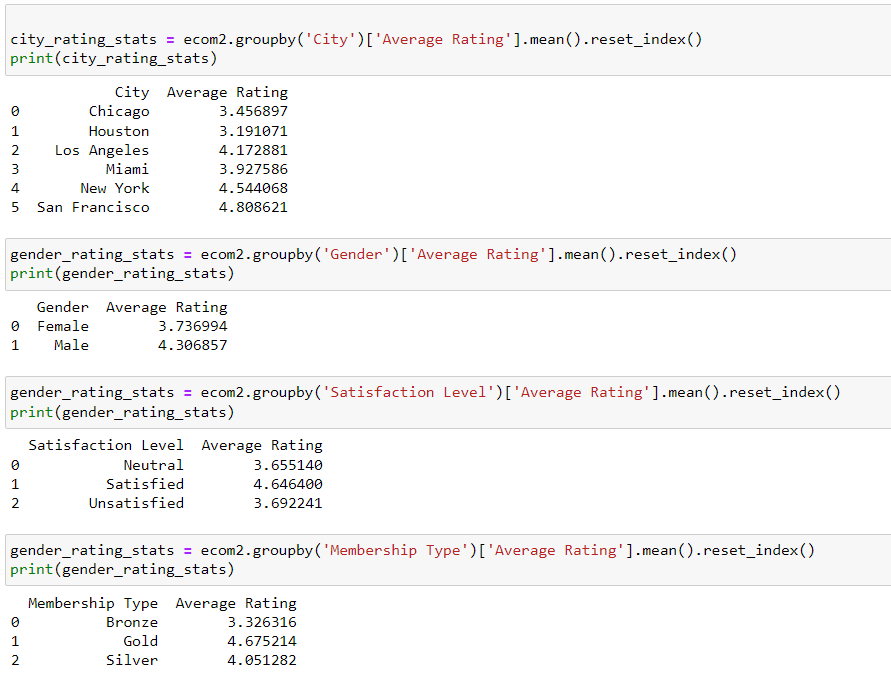
Оглянемо взаємозв'язок між віком споживачів та середнім рейтингом товарів, а також кількістю відгуків. Ми використовуємо групування за віком та обчислення середнього рейтингу і кількості відгуків для кожної вікової категорії.







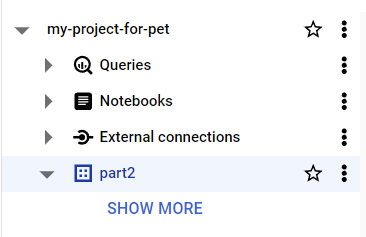




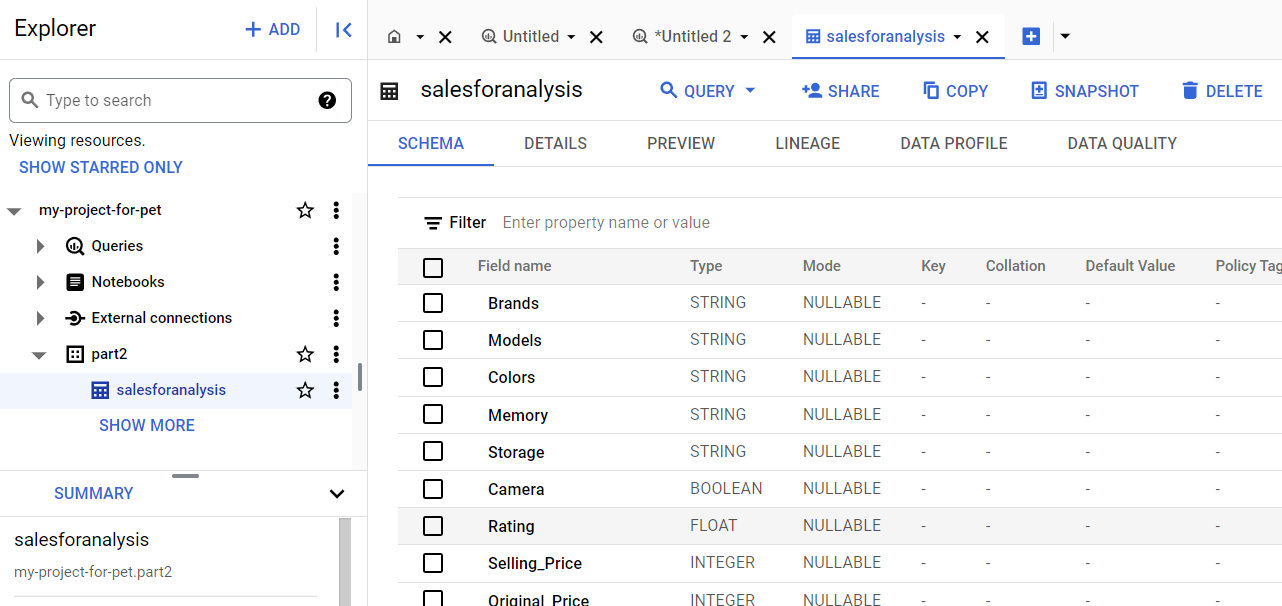
Отримавши основне розуміння даних та початковий аналіз даного дата сету, ми можемо перейти до наступного. У роботі з цим дата сетом ми перейдемо до роботи з BigQuery.

**BigQuery**

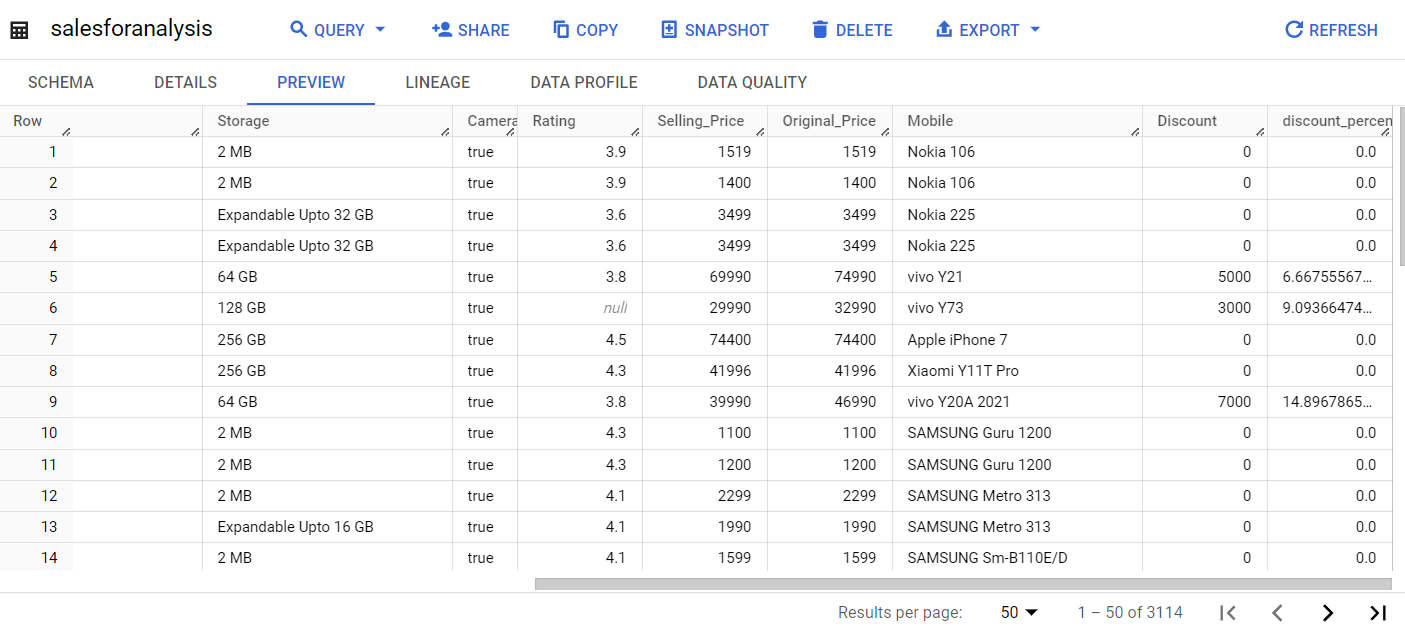
Створимо проект та завантажимо другий дата сет в таблицю.

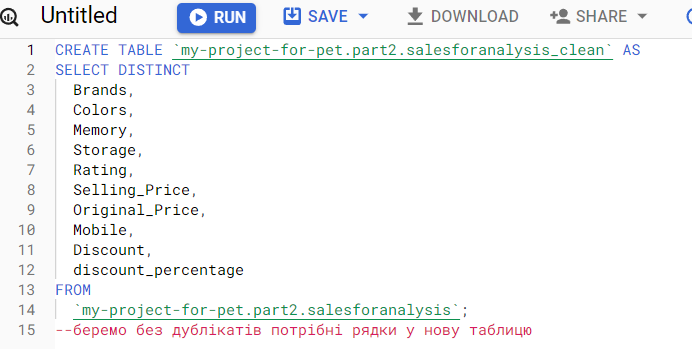


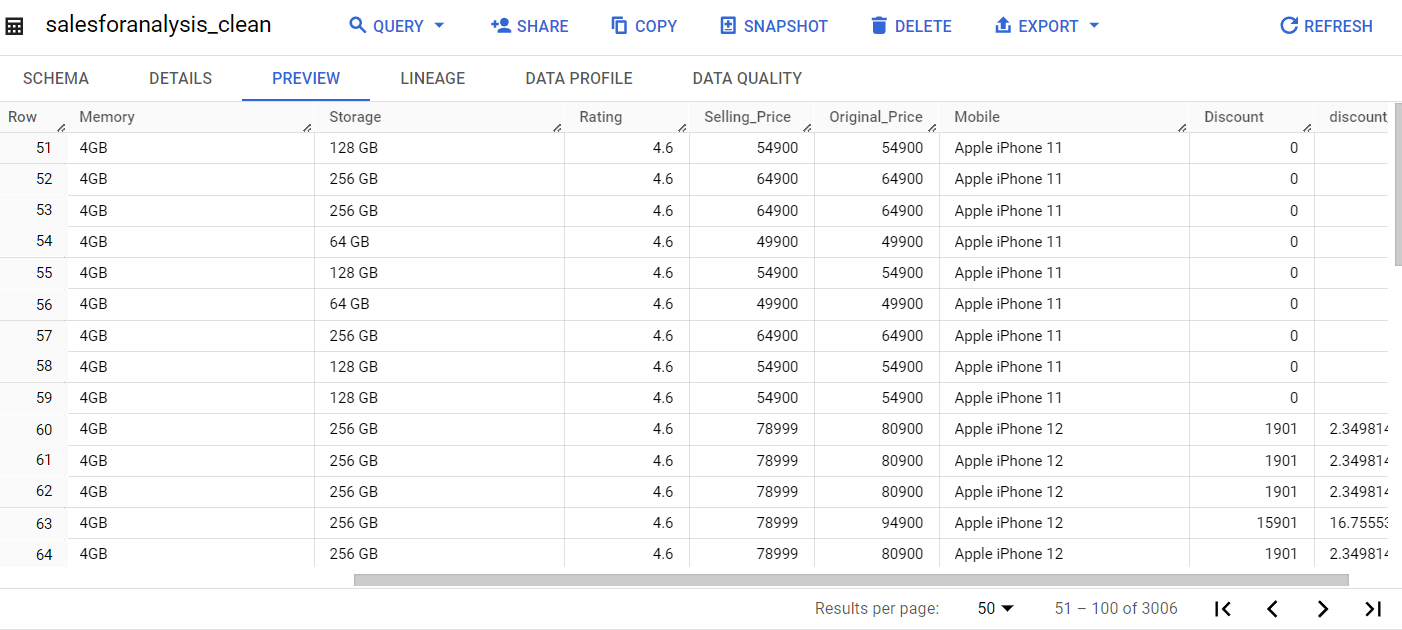




Зробимо окрему табличку під очищені дані

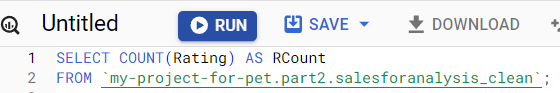




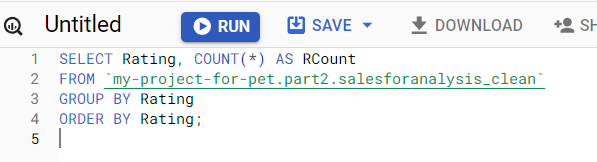


Ми відкинули 108 рядків-дублікатів

Знову почнемо з аналізу рейтингу, знайдемо кількість оцінок, загальну та роздрібну.

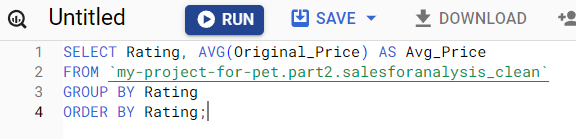


| Row | RCount |  |
| --- | --- | --- |
| 1 | 2876 |  |

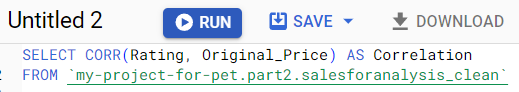


| Row | Rating | RCount |  |
| --- | --- | --- | --- |
| 1 | *null* | 130 |  |
| 2 | 2.3 | 1 |  |
| 3 | 2.4 | 1 |  |
| 4 | 2.7 | 2 |  |
| 5 | 2.8 | 3 |  |
| 6 | 3.0 | 6 |  |
| 7 | 3.1 | 2 |  |
| 8 | 3.2 | 3 |  |
| 9 | 3.3 | 8 |  |
| 10 | 3.4 | 3 |  |
| 11 | 3.5 | 17 |  |
| 12 | 3.6 | 35 |  |
| 13 | 3.7 | 65 |  |
| 14 | 3.8 | 114 |  |
| 15 | 3.9 | 134 |  |
| 16 | 4.0 | 207 |  |
| 17 | 4.1 | 249 |  |
| 18 | 4.2 | 319 |  |
| 19 | 4.3 | 683 |  |
| 20 | 4.4 | 506 |  |
| 21 | 4.5 | 301 |  |
| 22 | 4.6 | 164 |  |
| 23 | 4.7 | 44 |  |
| 24 | 4.8 | 2 |  |
| 25 | 4.9 | 1 |  |
| 26 | 5.0 | 6 |  |

Перейдемо до огляду середньої ціни по кожному рейтингу.



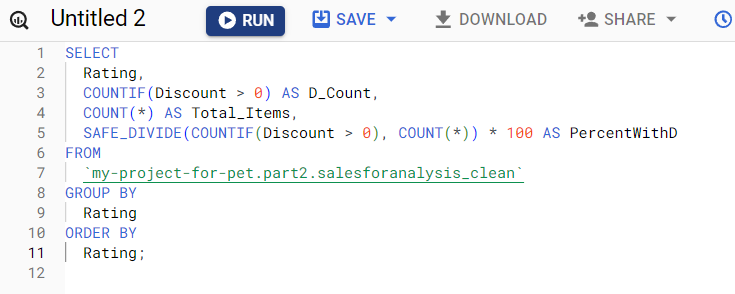
| Row | Rating | Avg\_Price |  |
| --- | --- | --- | --- |
| 1 | *null* | 64330.238461538444 |  |
| 2 | 2.3 | 11999.0 |  |
| 3 | 2.4 | 14939.0 |  |
| 4 | 2.7 | 9632.0 |  |
| 5 | 2.8 | 9309.3333333333339 |  |
| 6 | 3.0 | 74073.0 |  |
| 7 | 3.1 | 8620.0 |  |
| 8 | 3.2 | 20329.666666666668 |  |
| 9 | 3.3 | 12270.5 |  |
| 10 | 3.4 | 10196.333333333334 |  |
| 11 | 3.5 | 8360.4705882352919 |  |
| 12 | 3.6 | 9469.7428571428572 |  |
| 13 | 3.7 | 10717.830769230772 |  |
| 14 | 3.8 | 13258.5789473684 |  |
| 15 | 3.9 | 13489.947761194029 |  |
| 16 | 4.0 | 15800.560386473437 |  |
| 17 | 4.1 | 13975.253012048206 |  |
| 18 | 4.2 | 15598.335423197497 |  |
| 19 | 4.3 | 20789.888726207915 |  |
| 20 | 4.4 | 25709.057312252957 |  |
| 21 | 4.5 | 36358.7475083056 |  |
| 22 | 4.6 | 84549.8780487805 |  |
| 23 | 4.7 | 130120.22727272726 |  |
| 24 | 4.8 | 20999.0 |  |
| 25 | 4.9 | 22990.0 |  |
| 26 | 5.0 | 36849.833333333336 |  |



| Row | Correlation |  |
| --- | --- | --- |
| 1 | 0.436030372212344 |  |

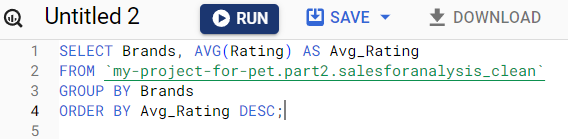
Коефіцієнт кореляції 0.436 показує помірну позитивну залежність між рейтингом товару і його початковою ціною (Original\_Price).

Також оцінимо за рейтингом: кількість взятих товарів зі знижкою, загальну кількість товарів по рейтингу і процент взятих зі знижкою товарів.



| Row | Rating | D\_Count | | | | Total\_Items | PercentWithD | |  | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | *null* | 49 | | | | 130 | 37.692307692307693 | |  | |
| 2 | 2.3 | 0 | | | | 1 | 0.0 | |  | |
| 3 | 2.4 | 0 | | | | 1 | 0.0 | |  | |
| 4 | 2.7 | 1 | | | | 2 | 50.0 | |  | |
| 5 | 2.8 | 1 | | | | 3 | 33.333333333333329 | |  | |
| 6 | 3.0 | 2 | | | | 6 | 33.333333333333329 | |  | |
| 7 | 3.1 | 1 | | | | 2 | 50.0 | |  | |
| 8 | 3.2 | 1 | | | | 3 | 33.333333333333329 | |  | |
| 9 | 3.3 | 1 | | | | 8 | 12.5 | |  | |
| 10 | 3.4 | 2 | | | | 3 | 66.666666666666657 | |  | |
| 11 | 3.5 | 0 | | | | 17 | 0.0 | |  | |
| 12 | 3.6 | 3 | | | | 35 | 8.5714285714285712 | |  | |
| 13 | 3.7 | 12 | | | | 65 | 18.461538461538463 | |  | |
| 14 | 3.8 | 21 | | | | 114 | 18.421052631578945 | |  | |
| 15 | 3.9 | 23 | | | | 134 | 17.164179104477611 | |  | |
| 16 | 4.0 | 38 | | | | 207 | 18.357487922705314 | |  | |
| 17 | 4.1 | 63 | | | | 249 | 25.301204819277107 | |  | |
| 18 | 4.2 | 110 | | | | 319 | 34.482758620689658 | |  | |
| 19 | 4.3 | 366 | | | | 683 | 53.587115666178619 | |  | |
|  |
| 20 | 4.4 | | 264 | 506 | | | | 52.173913043478258 | |  | |
| 21 | 4.5 | | 133 | | 301 | | | 44.186046511627907 | |  | |
| 22 | 4.6 | | 56 | | 164 | | | 34.146341463414636 | |  | |
| 23 | 4.7 | | 8 | | 44 | | | 18.181818181818183 | |  | |
| 24 | 4.8 | | 2 | | 2 | | | 100.0 | |  | |
| 25 | 4.9 | | 0 | | 1 | | | 0.0 | |  | |
| 26 | 5.0 | | 0 | | 6 | | | 0.0 | |  | |

Розглянемо розподіл оцінок по брендам.



| Row | Brands | Avg\_Rating |  |
| --- | --- | --- | --- |
| 1 | Apple | 4.56654411764706 |  |
| 2 | Google Pixel | 4.5068965517241368 |  |
| 3 | IQOO | 4.4 |  |
| 4 | realme | 4.3924342105263161 |  |
| 5 | POCO | 4.3876712328767109 |  |
| 6 | vivo | 4.3827586206896534 |  |
| 7 | Xiaomi | 4.3046875 |  |
| 8 | OPPO | 4.297177419354842 |  |
| 9 | Infinix | 4.2760000000000007 |  |
| 10 | SAMSUNG | 4.2085014409221957 |  |
| 11 | Motorola | 4.1421568627450975 |  |
| 12 | ASUS | 4.08090909090909 |  |
| 13 | Nokia | 4.0010752688172042 |  |
| 14 | LG | 3.9816326530612249 |  |
| 15 | Lenovo | 3.9706896551724138 |  |
| 16 | HTC | 3.9690909090909083 |  |
| 17 | GIONEE | 3.8500000000000 | |

Перевіримо ж, товари яких брендів займають найвищі та найнижчі місця серед рейтингу:

WITH MinRatings AS (

  SELECT

    Mobile,

    AVG(Rating) AS Avg\_Rating,

    ROW\_NUMBER() OVER(ORDER BY AVG(Rating) ASC) AS Min\_Rank

  FROM

    `my-project-for-pet.part2.salesforanalysis\_clean`

  WHERE

    Rating > 0  -- вибираємо записи з ненульовим рейтингом

  GROUP BY

    Mobile

),

MaxRatings AS (

  SELECT

    Mobile,

    AVG(Rating) AS Avg\_Rating,

    ROW\_NUMBER() OVER(ORDER BY AVG(Rating) DESC) AS Max\_Rank

  FROM

    `my-project-for-pet.part2.salesforanalysis\_clean`

  GROUP BY

    Mobile

)

SELECT

  'Top Min Ratings' AS Rank\_Category,

  Mobile,

  Avg\_Rating

FROM

  MinRatings

WHERE

  Min\_Rank <= 5

UNION ALL

SELECT

  'Top Max Ratings' AS Rank\_Category,

  Mobile,

  Avg\_Rating

FROM

  MaxRatings

WHERE

  Max\_Rank <= 5

ORDER BY

  Rank\_Category, Avg\_Rating;

| Row | Rank\_Category | Mobile | Avg\_Rating |  |
| --- | --- | --- | --- | --- |
| 1 | Top Max Ratings | Apple iPhone 13 | 4.7 |  |
| 2 | Top Max Ratings | Apple iPhone 13 Pro | 4.7 |  |
| 3 | Top Max Ratings | vivo Z1x | 4.75 |  |
| 4 | Top Max Ratings | vivo S2 | 5.0 |  |
| 5 | Top Max Ratings | SAMSUNG SM-B310EZDDINS | 5.0 |  |
| 6 | Top Min Ratings | Motorola Quench X T3 | 2.4 |  |
| 7 | Top Min Ratings | GIONEE M7 | 2.7 |  |
| 8 | Top Min Ratings | GIONEE G3 | 2.8 |  |
| 9 | Top Min Ratings | GIONEE S11 | 2.8 |  |
| 10 | Top Min Ratings | OPPO R1 R829 | 3.0 |  |

І на кінець, розглянемо які були середні знижки на товари по кожному рейтингу .

SELECT

  Rating,

  AVG(discount\_percentage) AS Avg\_Discount\_Percentage

FROM

  `my-project-for-pet.part2.salesforanalysis\_clean`

WHERE

  discount\_percentage > 0  --записи з ненульовим

GROUP BY

  Rating

ORDER BY

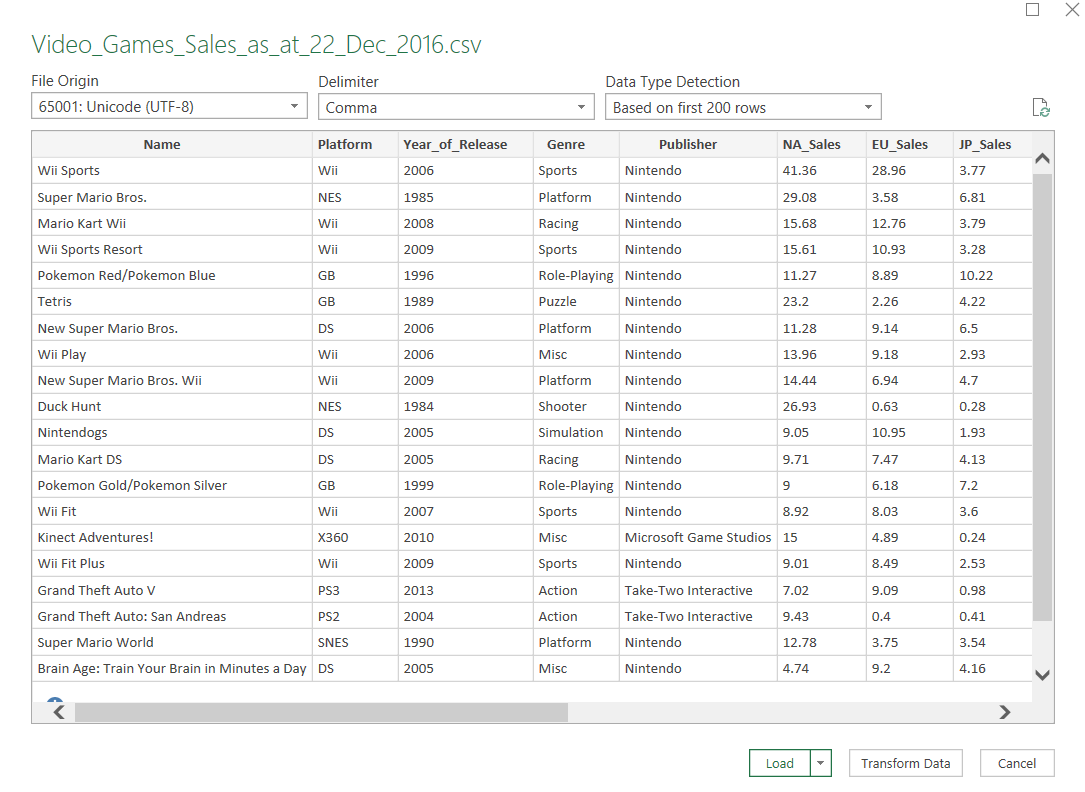
  Rating desc;

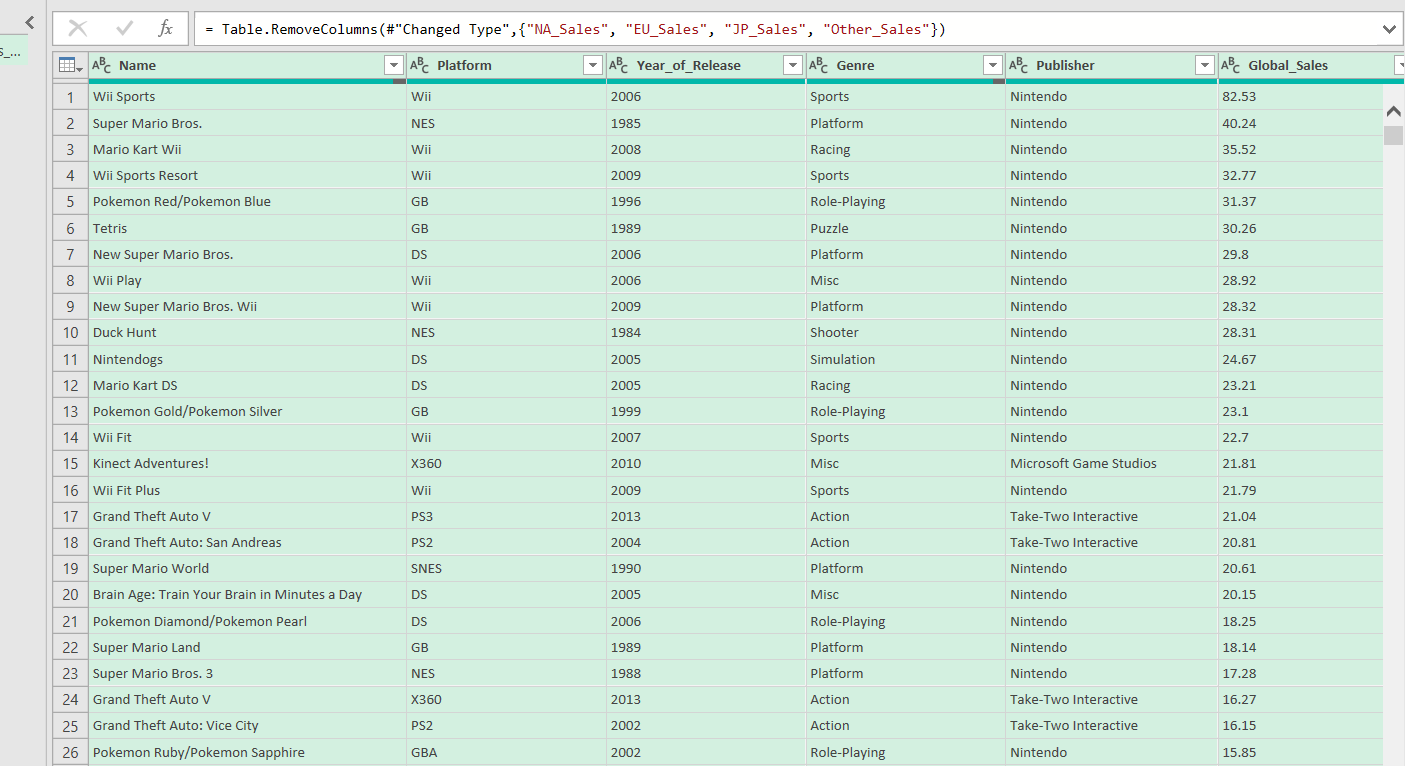
| Row | Rating | Avg\_Discount\_Percentage |  |
| --- | --- | --- | --- |
| 1 | 4.8 | 16.667460355255013 |  |
| 2 | 4.7 | 21.781400599470075 |  |
| 3 | 4.6 | 17.858541121250596 |  |
| 4 | 4.5 | 13.323872217712735 |  |
| 5 | 4.4 | 15.345897434643613 |  |
| 6 | 4.3 | 14.331136376841323 |  |
| 7 | 4.2 | 15.506731486796019 |  |
| 8 | 4.1 | 19.630996161213712 |  |
| 9 | 4.0 | 22.573083042211138 |  |
| 10 | 3.9 | 18.864211756935234 |  |
| 11 | 3.8 | 28.248366447935638 |  |
| 12 | 3.7 | 38.827619280691465 |  |
| 13 | 3.6 | 47.94413836234687 |  |
| 14 | 3.4 | 50.484737999916263 |  |
| 15 | 3.3 | 44.4493832648072 |  |
| 16 | 3.2 | 37.604002106371773 |  |
| 17 | 3.1 | 42.758620689655174 |  |
| 18 | 3.0 | 21.052742382854646 |  |
| 19 | 2.8 | 0.11697274535033338 |  |
| 20 | 2.7 | 0.10026067776218166 |  |
| 21 | null | 12.583967648268 |  |

Тепер, переглянувши більш детально другий дата сет, можемо перейти до аналізу третього. Для його аналізу, ми використаємо EXEL.

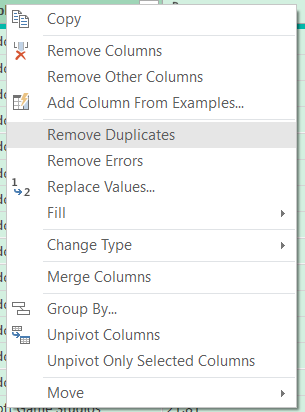
**EXEL**

Спершу перетворимо CSV-файл на .xlsx формат з яким легше працювати в EXEL. Також перенесемо дані в Power Query Editor для роботи зі стовпцями.

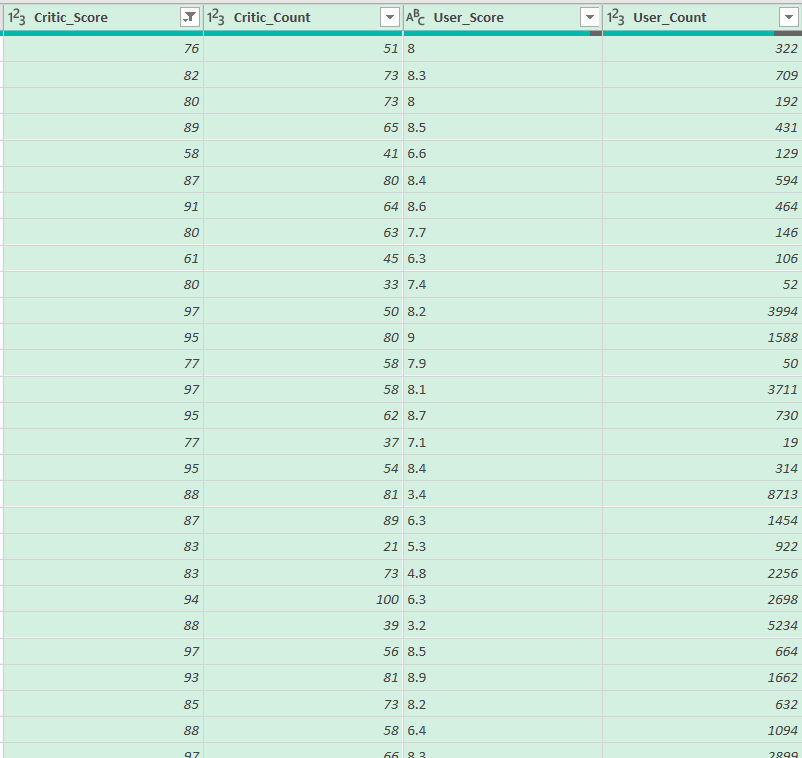




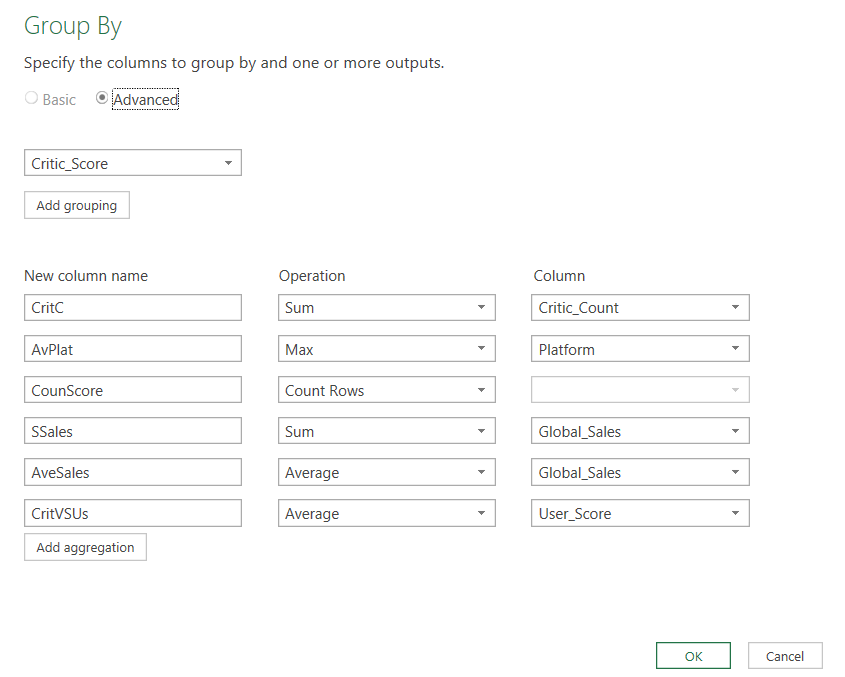
Далі очищуємо дані від помилок та дублікатів перед аналізом даних.

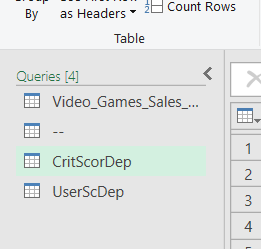


Також додаймо фільтр по відкиданню нульових значень в оцінках

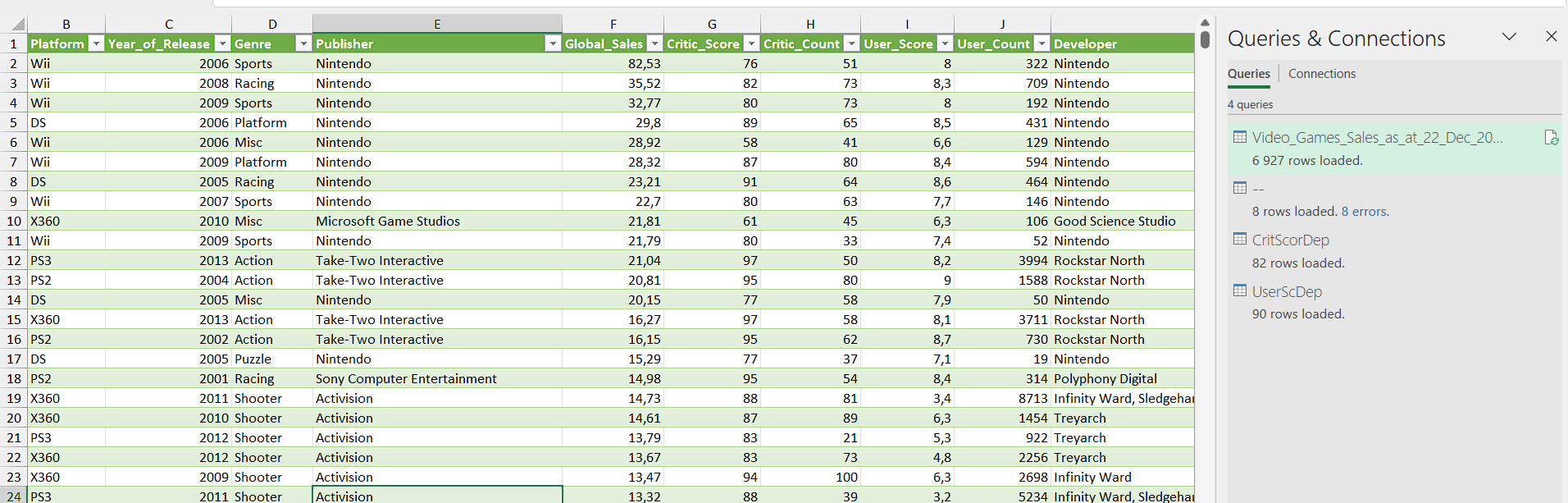


Змінимо типи колонок на зручні для аналізу і перейдемо до створення нових табличок для оцінки певних залежностей від рейтингу.

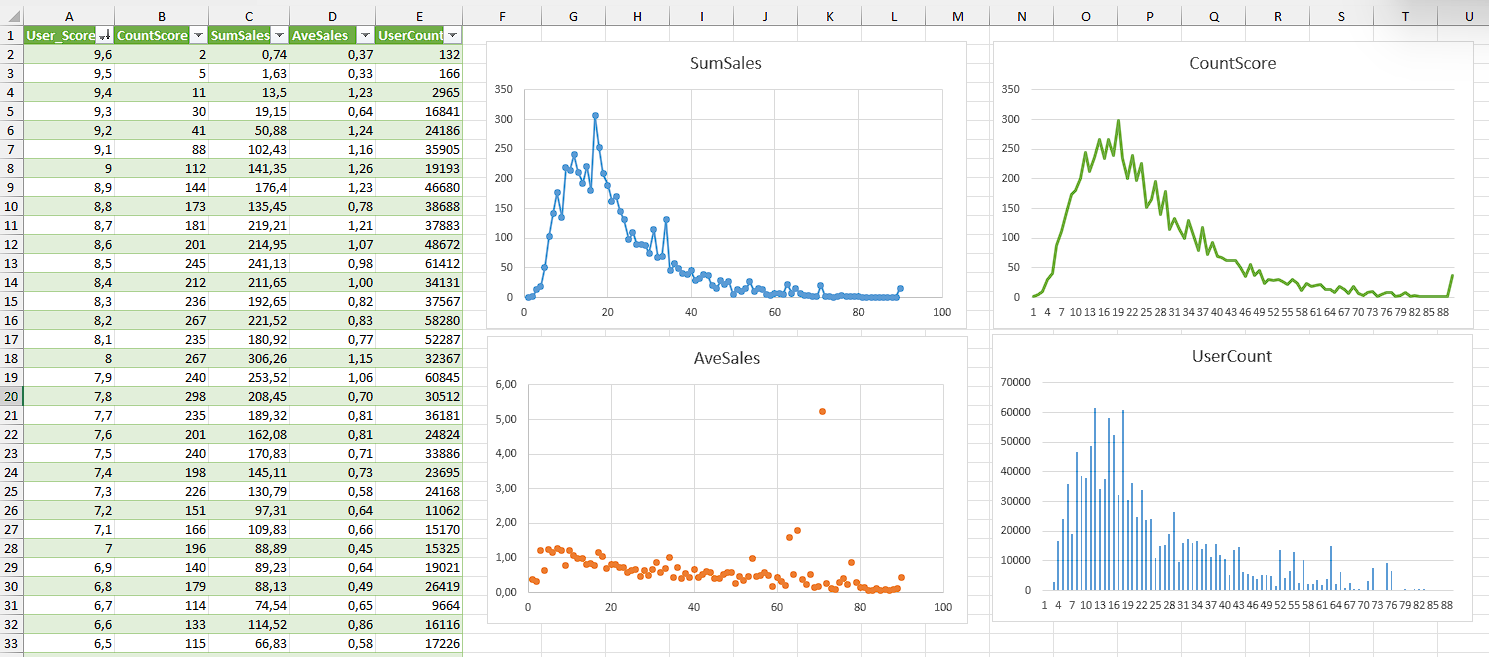


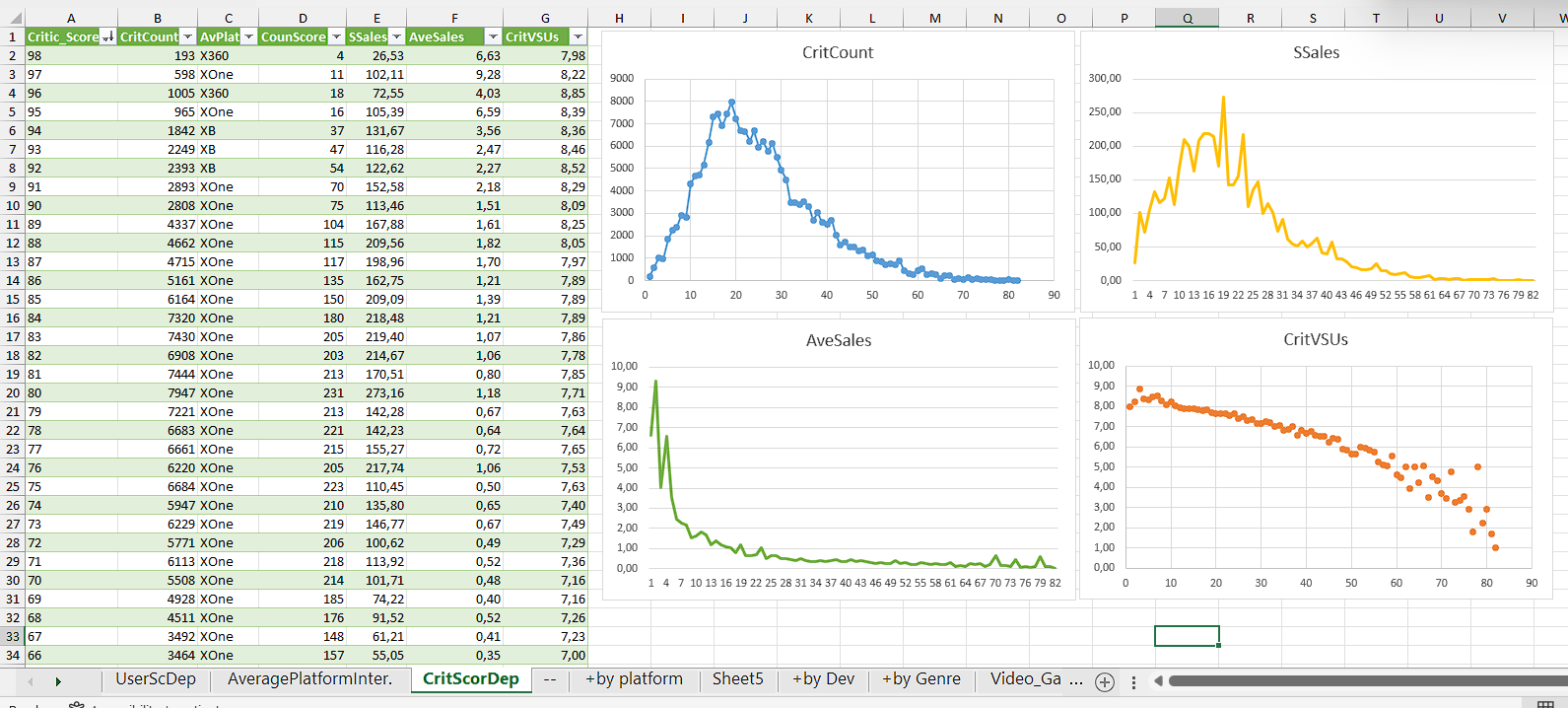


Далі завантажуємо дані з Power Query Editor в EXEL і можемо перейти до аналізу

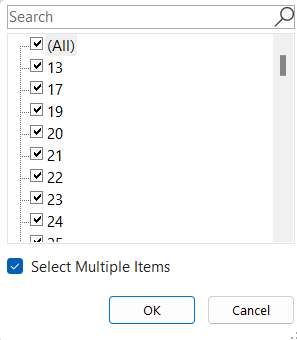
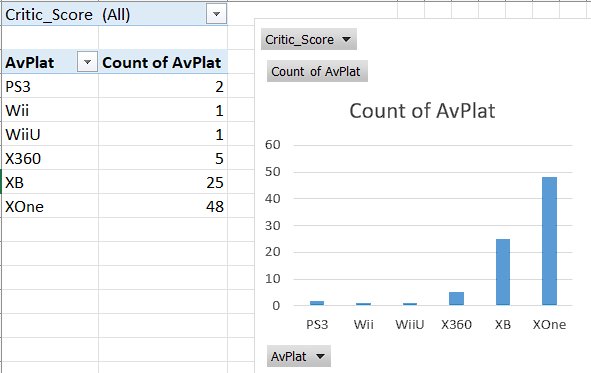


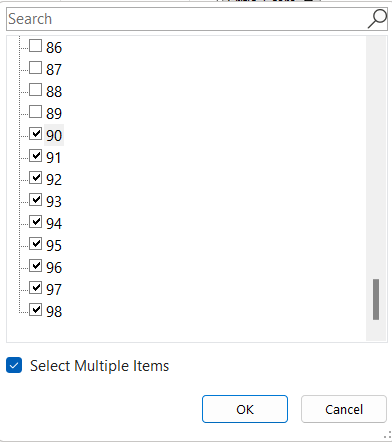
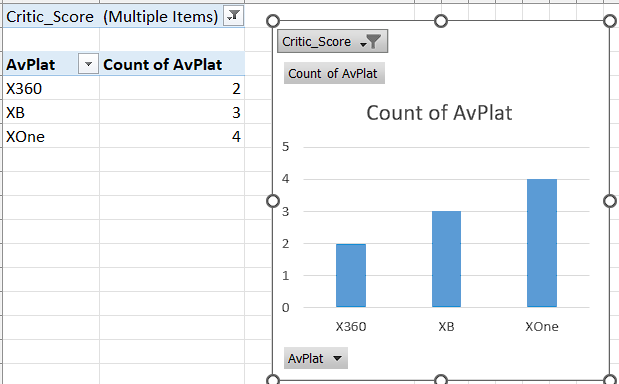
Додаємо візуальні елементи до таблиць і бачимо позитивну залежність від рейтингу більшості з отриманих розрахунків.

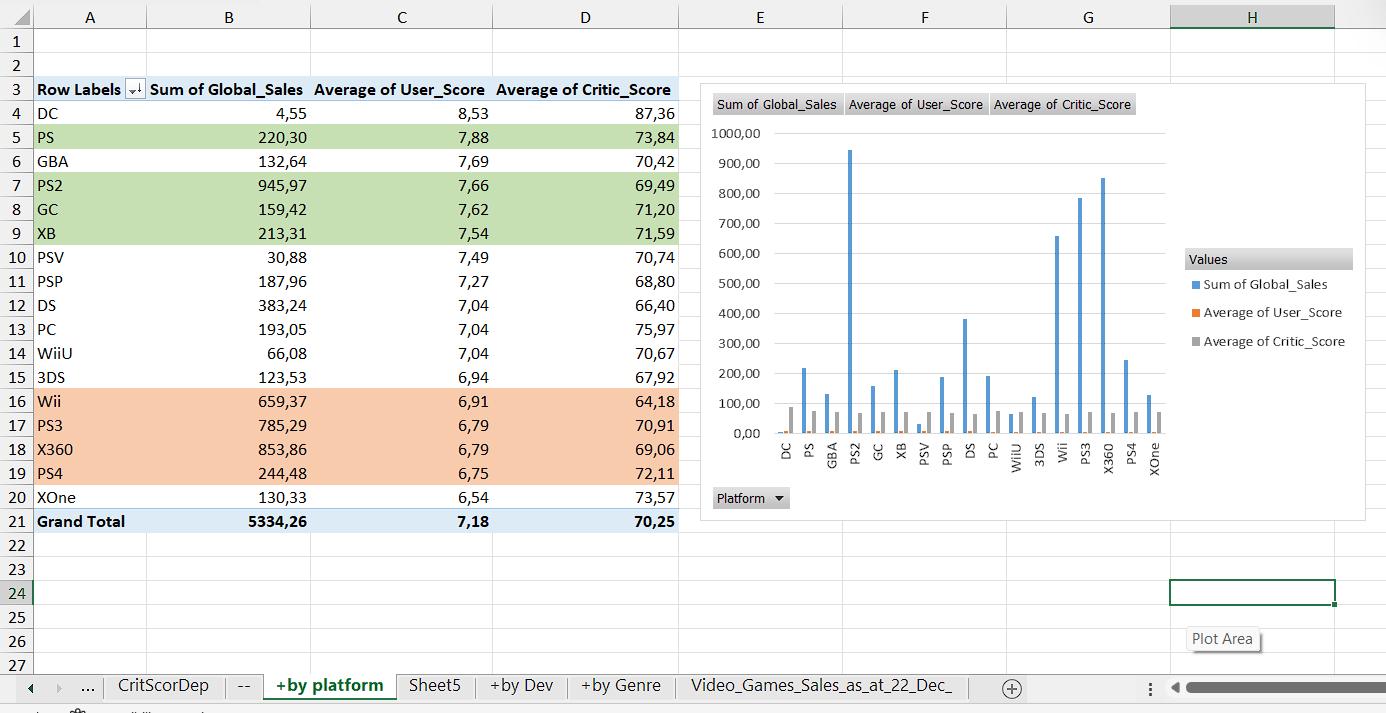
(за р. спож.)

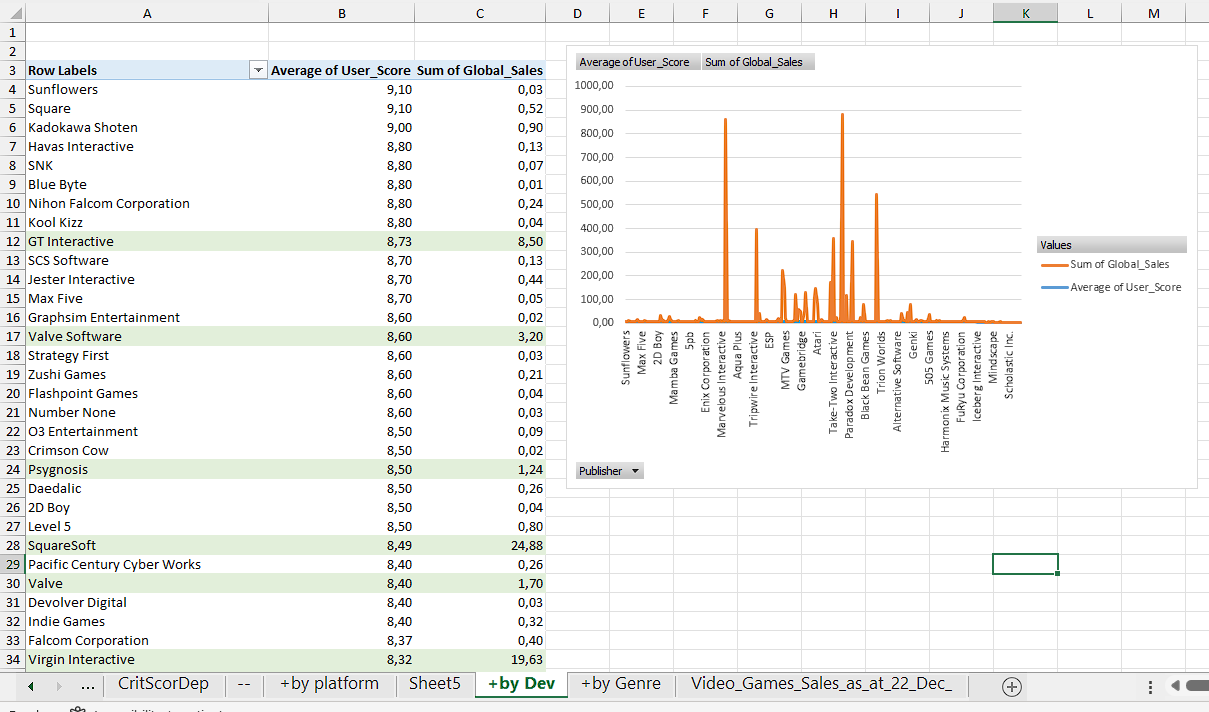
(за р. критиків)

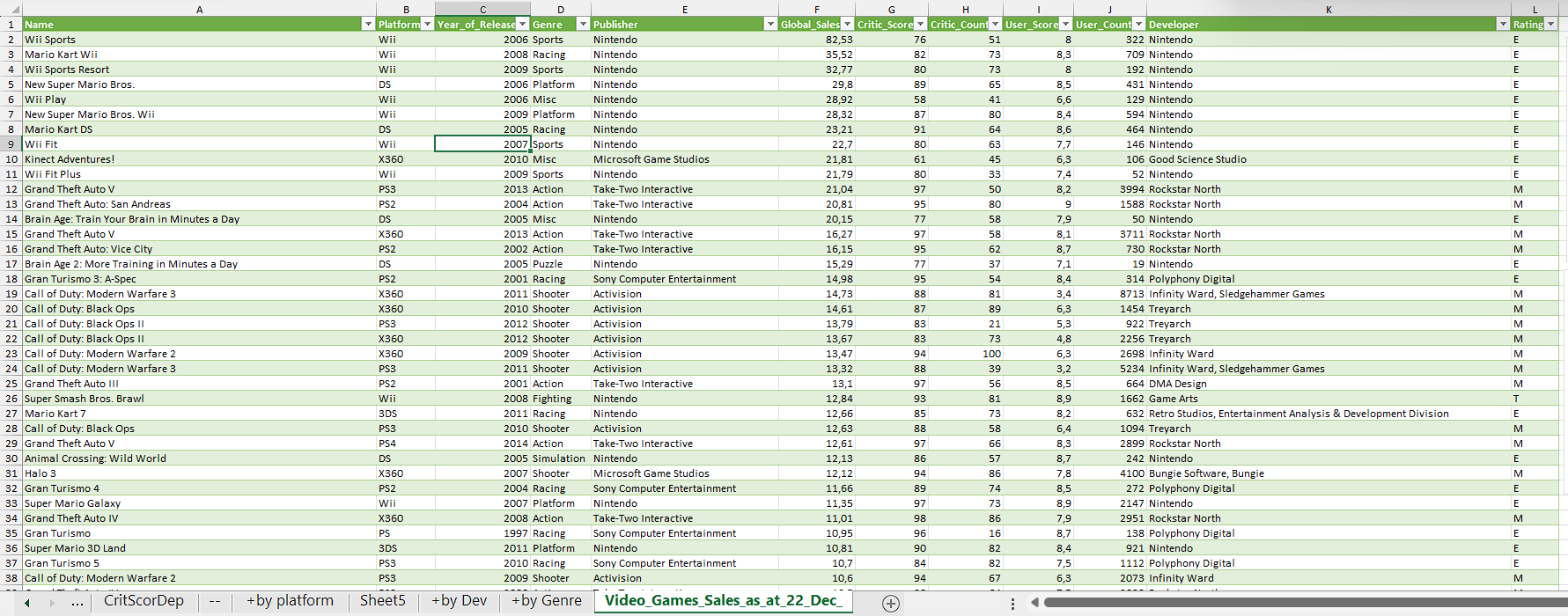
Також на прикладі платформ, можемо використати Pivot Table для додаткового аналізу з можливістю детальніше розглянути вибрану группу.

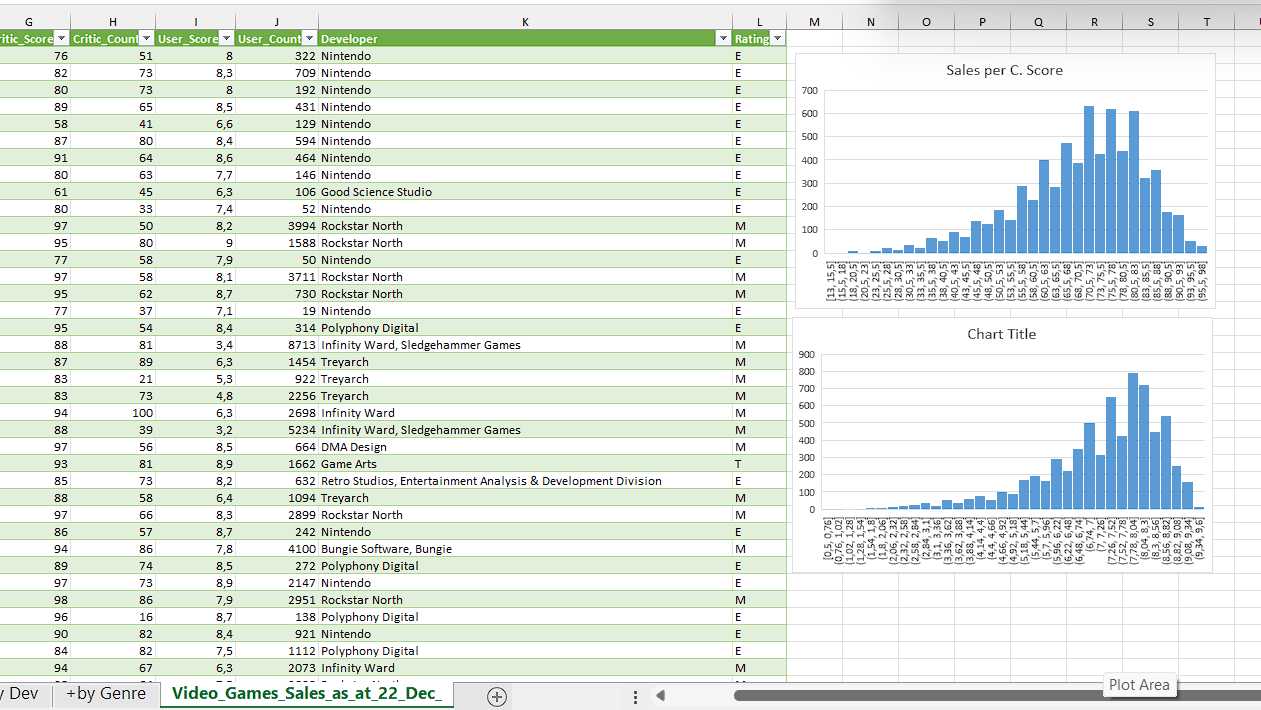
 





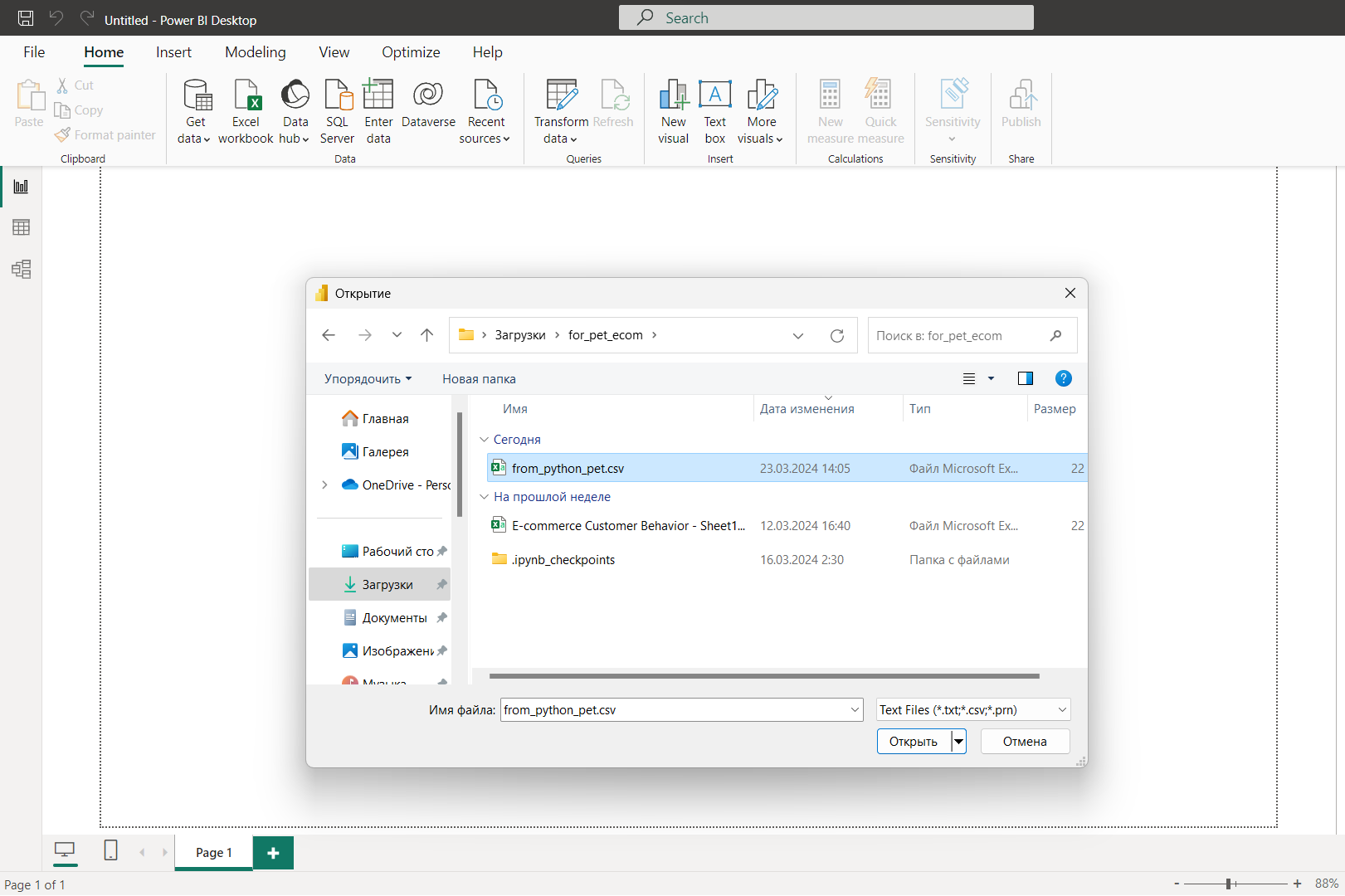




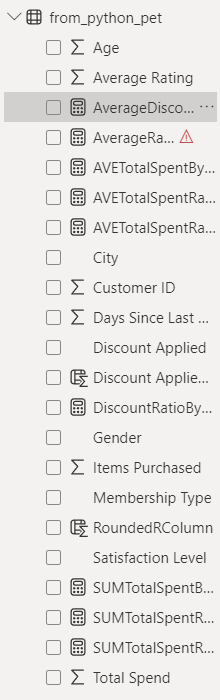
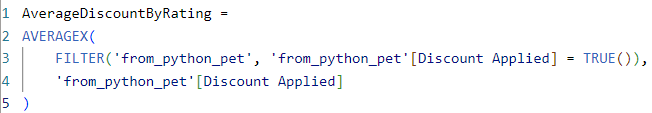
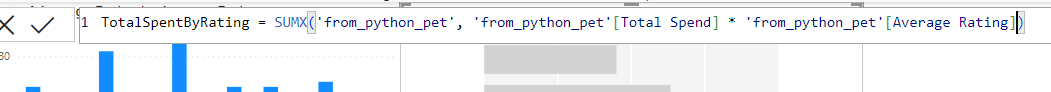
Провівши аналіз вибраних дата сетів, перейдемо до візуалізації отриманих інсайтів, де розберемо їх детальніше. До більш детального розгляду за темою доцільно вибрати датасети E-commerce Customer Behavior та S.Sales

**Power BI**

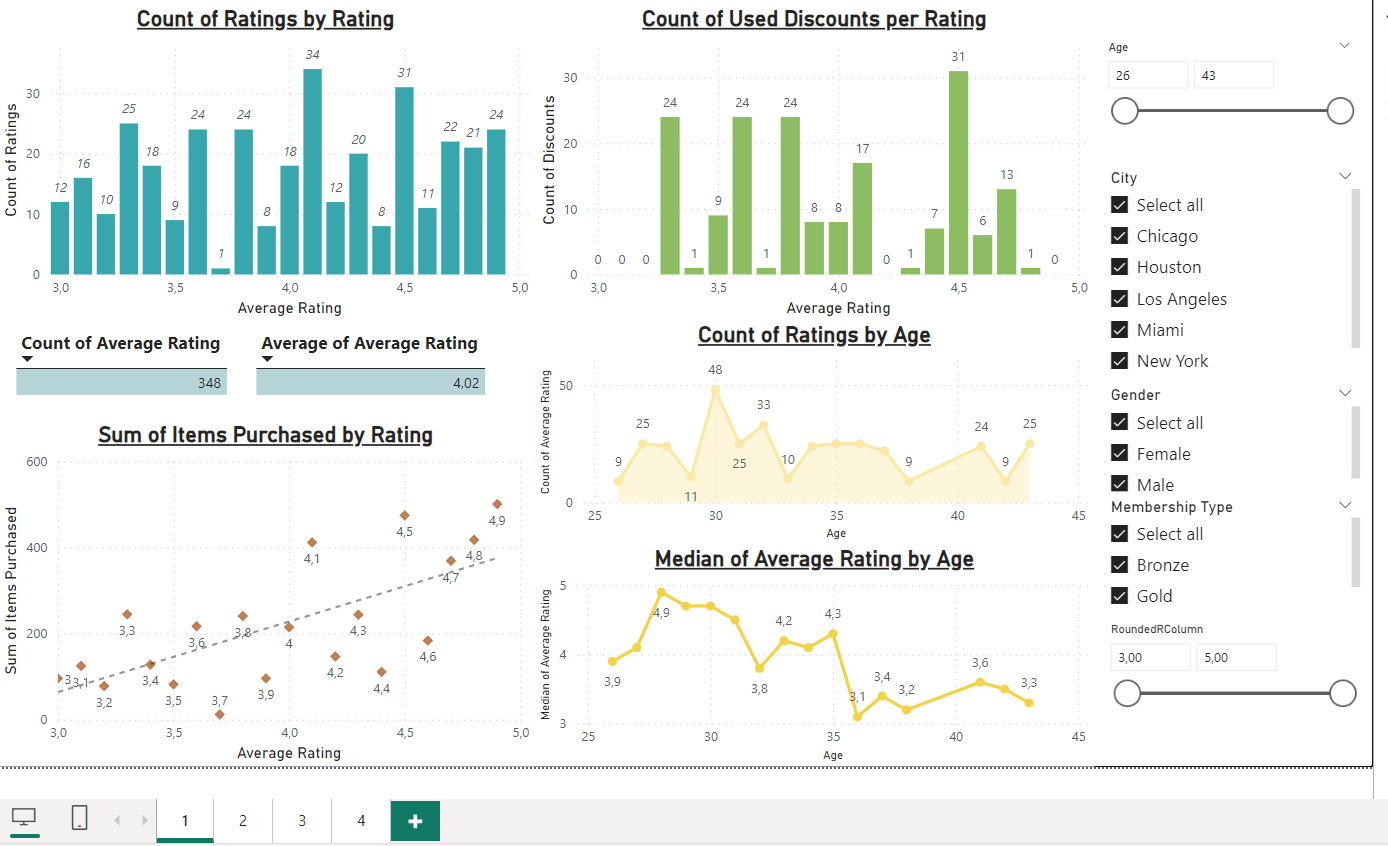
Очищений перший дата сет перенесемо в Power BI для створення візуалізацій.



Для створення потрібних візуалізацій додаємо потрібні стовпці.

 …

Збираємо візуалізації в дашборди, за ними ми можемо зробити деякі висновки, як загальні, так і на обрану групу за певними елементами.



(приклад)

Перша візуалізація на першому слайді показує кількість оцінок за рейтингом. Згідно з аналізом, у датасеті ми маємо 348 оцінок, середній рейтинг становить 4.02. Найбільша кількість оцінок (34) була у рейтингу 4.10, що становить 9.77% від загальної кількості оцінок. Другим за кількістю оцінок був рейтинг 4.50, а найнижчою кількістю оцінок (1) виявився рейтинг 3.70. Ці дані свідчать про те, що товари з рейтингом 4+ мають незначно проте більшу аудиторію серед користувачів, порівняно з іншими рейтингами. (Допомагає визначити які рейтинги найбільше привертають увагу споживачів, що є ключовим для розуміння їх вподобань та тенденцій).

Друга візуалізація показує кількість використаних знижок на рейтинг. За аналізом, рейтинг 4.50 має найбільшу суму використаних знижок (31), що становить 17.71% від загальної суми знижок. На другому місці з однаковою сумою знижок (24) розташувалися рейтинги 3.30 та 3.60. За цими даними можна зробити висновок, що товари з рейтингом 4.50 не лише мають високу кількість оцінок, але і активно використовуються зі знижками, що може впливати на їхню популярність серед споживачів.( Показує, як високий рейтинг впливає на використання знижок для товарів. Дозволяє оцінити, чи активно використовуються знижки для товарів з високим рейтингом, що може вказувати на їхню популярність та залучення покупців).

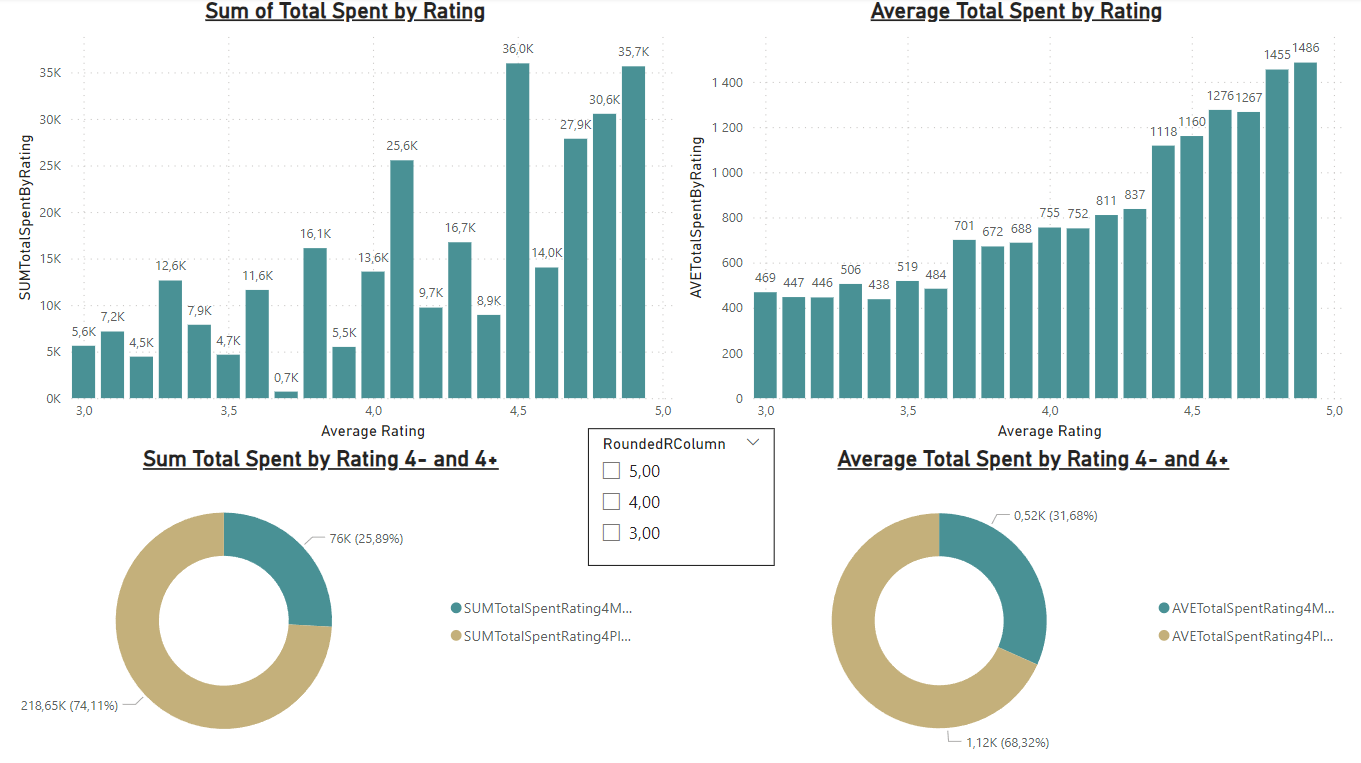
Третя візуалізація у формі скатерплоту за сумою куплених товарів по рейтингу ілюструє сильну позитивну залежність між рейтингом товару та обсягами продажів. Найвищий обсяг куплених товарів спостерігається у товарів з рейтингом 4.9, що мають 501 одиницю. Далі слідують товари з рейтингом 4.5 та 4.8. На іншому кінці спектру, товари з найнижчим рейтингом 3.7 та 3.2 мають лише 12 та 78 проданих товарів відповідно. Таким чином, можна підтвердити, що високі рейтинги супроводжуються більшими обсягами продажів, що важливо враховувати при зміні стратегій маркетингу та продажів.( Аналіз показує, що споживачі звертають увагу на рейтинг товару при прийнятті рішення про покупку, тому підтримка високого рейтингу стає важливим фактором у збільшенні продажів).

Четверта візуалізація, що відображає кількість рейтингів за віком, дозволяє отримати уявлення про те, як рейтинги розподіляються серед різних вікових груп споживачів. Згідно з аналізом, бачимо, що найбільша кількість рейтингів спостерігається у групи споживачів віком 30 років, де кількість оцінок сягає 48. Це свідчить про те, що у цій віковій групі товари отримують велику увагу та оцінки від користувачів. Наявність великої кількості рейтингів може бути позитивним сигналом для компаній у сегменті електроніки, оскільки це може вказувати на популярність та затребуваність продуктів серед споживачів цієї вікової групи.

П'ята візуалізація, що представляє середній рейтинг за віком, дозволяє оцінити якість продуктів у контексті різних вікових груп. За аналізом, бачимо, що найвищий середній рейтинг спостерігається у групи споживачів віком 28 років, де медіана середнього рейтингу досягає 4.90. Це свідчить про те, що в цій віковій групі товари отримують високу оцінку від користувачів, що може свідчити про якісні продукти або задоволення від їх використання. Низький середній рейтинг у певних вікових групах може вказувати на можливість покращення якості товарів або маркетингових стратегій для цих груп споживачів.

Загалом, ці візуалізації дозволяють краще розуміти залежність між віком споживачів, їхнім сприйняттям та оцінкою товарів у сегменті електроніки. Вони можуть бути важливими для розробки персоналізованих стратегій продажу та маркетингу, спрямованих на цільові вікові групи споживачів.

Перейдемо до наступного слайду.

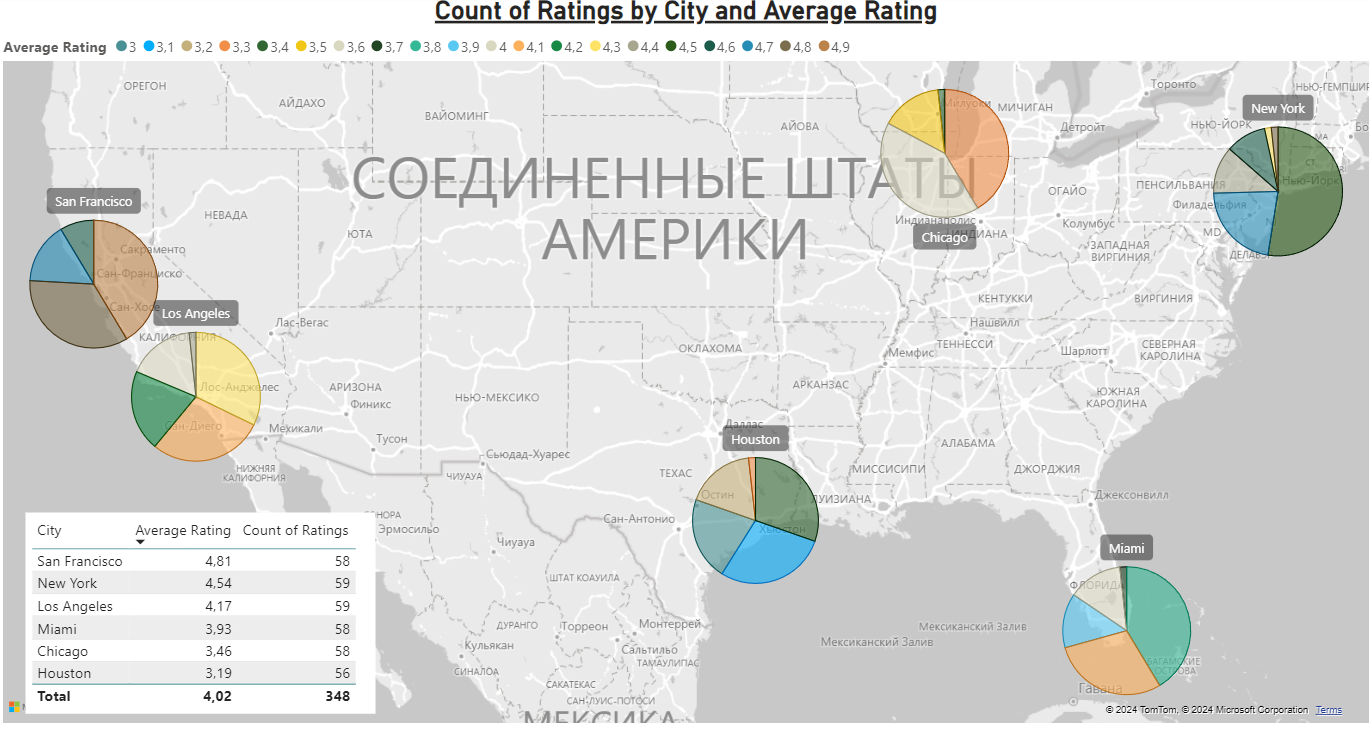


На цьому слайді ми спершу аналізуємо загальну суму витрат по рейтингу товарів, що дозволяє нам розуміти, як високий рейтинг впливає на загальний обсяг витрат споживачів. Згідно аналізу, ми бачимо, що товари з високим рейтингом (наприклад, 4.5) мають найвищу загальну суму витрат, що вказує на позитивний вплив високого рейтингу на витрати споживачів. Ця тенденція вказує на те, що споживачі готові інвестувати у товари з високим рейтингом, що відображається у збільшенні обсягів продажів та доходів компанії.

Отже, можна зробити висновок, що існує загально-позитивна залежність між рейтингом товару та загальною сумою витрат споживачів, що підкреслює важливість підтримки високих рейтингів для стимулювання продажів та прибутковості бізнесу.

На другому графіку ми аналізуємо середні витрати споживачів в залежності від рейтингу товарів. За результатами аналізу знову видно, що товари з високим рейтингом, зокрема рейтингом 4.9, характеризуються найвищими середніми витратами на покупку. Це свідчить про те, що споживачі, які обирають товари з високим рейтингом, готові витрачати більше коштів на ці товари порівняно з товарами з нижчим рейтингом.

Перейдемо до наступного слайду.



На цьому слайді представлена карта-візуалізація, де ми аналізуємо середні рейтинги по різних містах та кількість відгуків, отриманих від споживачів у кожному місті. Нижче наведені основні висновки за результатами аналізу:

Найвищий середній рейтинг товарів спостерігається у містах Los Angeles та San Francisco, де він становить 4.17 та 4.81 відповідно.

Найбільша кількість відгуків від споживачів отримана для міста New York, де рейтинги товарів складають у середньому 4.54.

У містах Chicago, Miami та New York виявлено найвищу кількість відгуків про товари з рейтингом 4.1, що свідчить про поширеність цього рейтингу в цих містах.

Усередині кожного міста різниця у середніх рейтингах товарів не є значною, що вказує на відносно однаковий рівень задоволеності споживачів товарами у цих містах.

Отже, аналіз демонструє, що існує певна варіабельність у рейтингах товарів за місцезнаходженням, проте загальна тенденція позитивна

І до останнього слайду цього дата сету.



На останньому слайді ми представляємо результати аналізу за допомогою трьох простих кругових діаграм. Основні висновки з аналізу:

Середній рейтинг товарів для чоловіків (4.31) вище, ніж для жінок (3.74), що може вказувати на певні різниці в уподобаннях та оцінках якості товарів між статями.

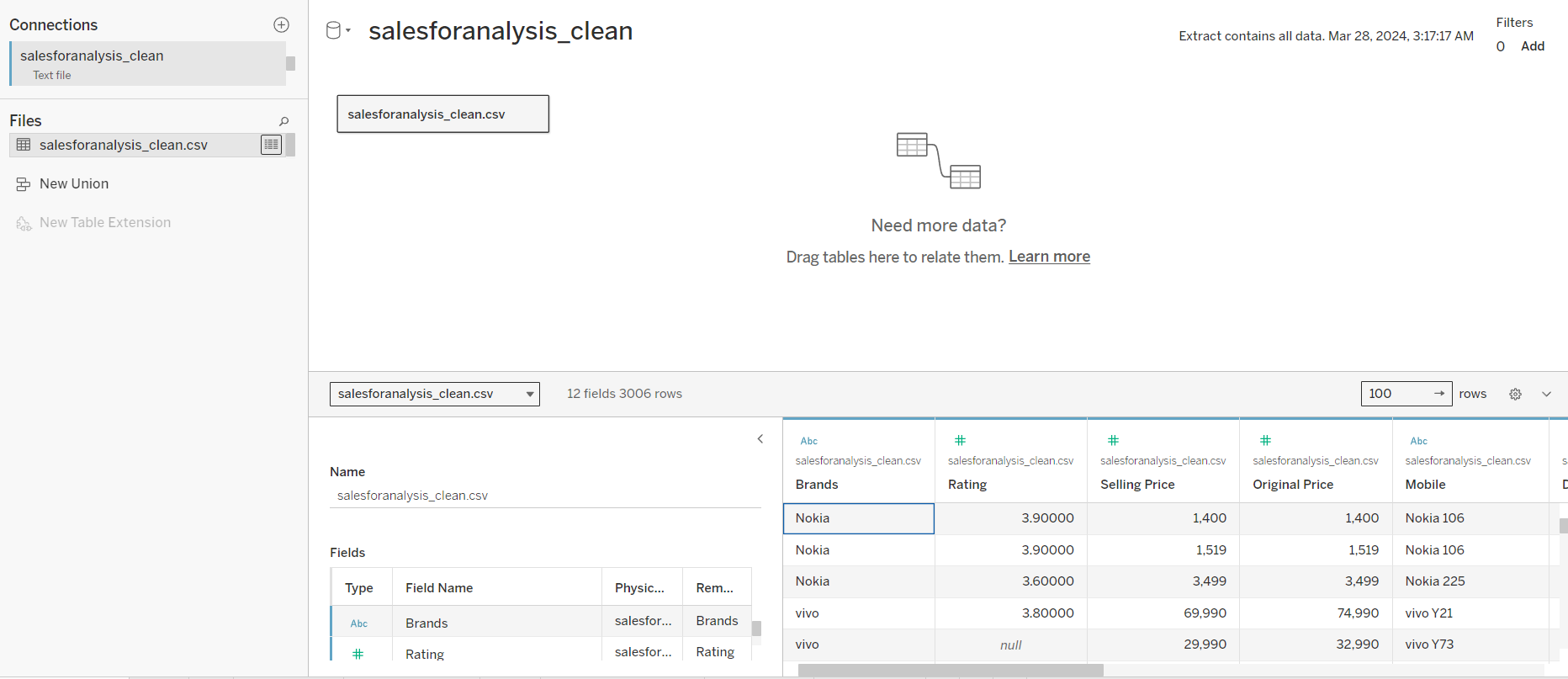
Категорія "Satisfied" має найвищий середній рейтинг товарів (4.65), що свідчить про високий рівень задоволеності споживачів цією категорією товарів. При цьому категорії "Unsatisfied" та "Neutral" мають нижчі середні рейтинги, що може вказувати на потребу в покращенні товарів цих категорій.

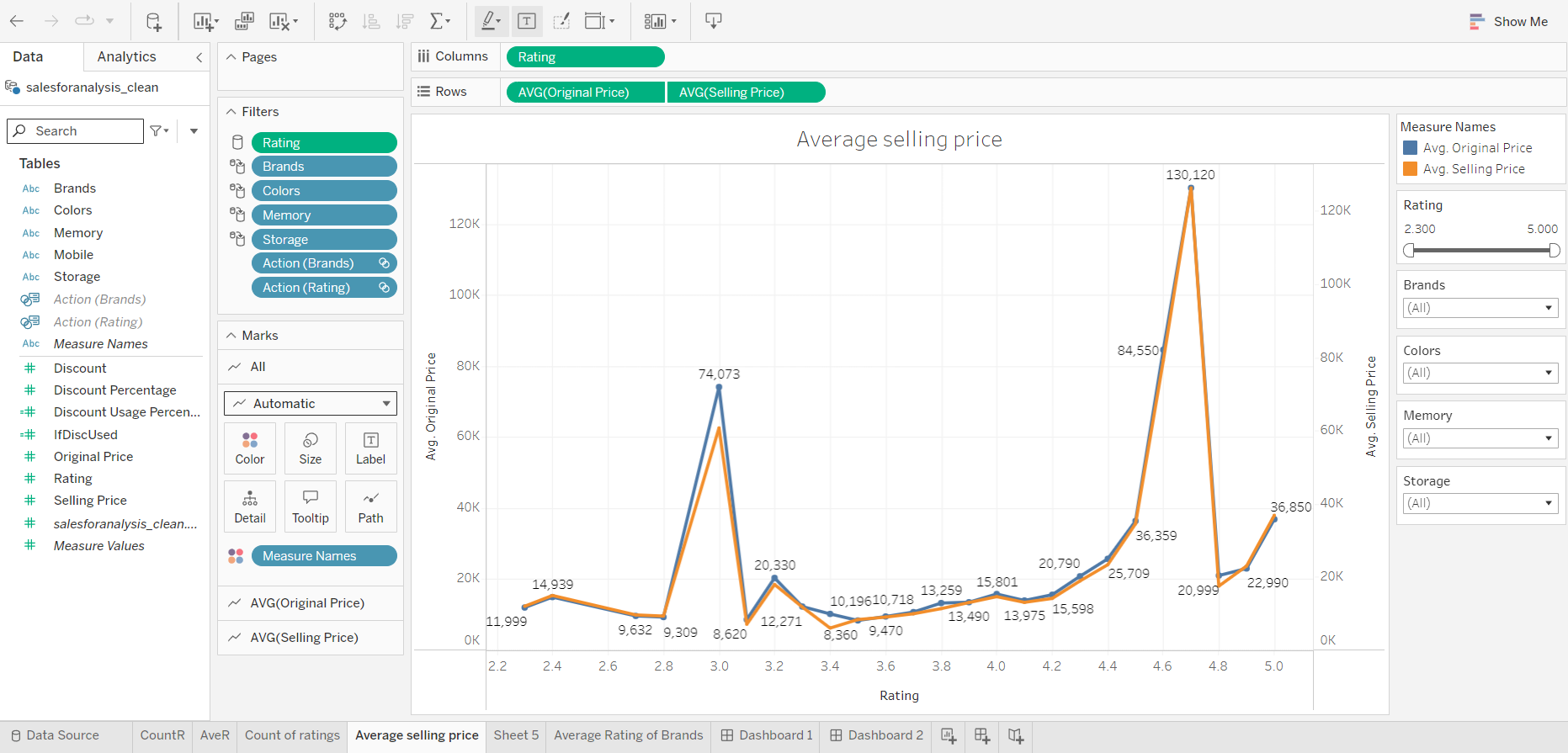
Товари категорії "Gold" мають найвищий середній рейтинг (4.68), що свідчить про їхню високу якість та популярність серед споживачів. При цьому категорії "Silver" та "Bronze" мають менші середні рейтинги, що може вказувати на різний рівень задоволеності споживачів цими категоріями товарів.

Загально, ці результати аналізу демонструють важливі відмінності у сприйнятті та оцінці товарів залежно від різних факторів, таких як стать, рівень задоволеності та категорія товару.

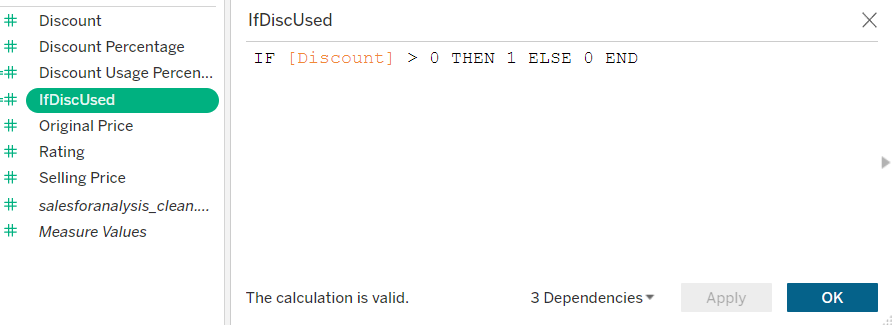
**Tableau (**[**Public**](https://public.tableau.com/views/pet2_fin/Dashboard1?:language=en-US&:sid=&:display_count=n&:origin=viz_share_link)**)**

Перенесемо очищені дані та перейдемо до створення візуалізацій

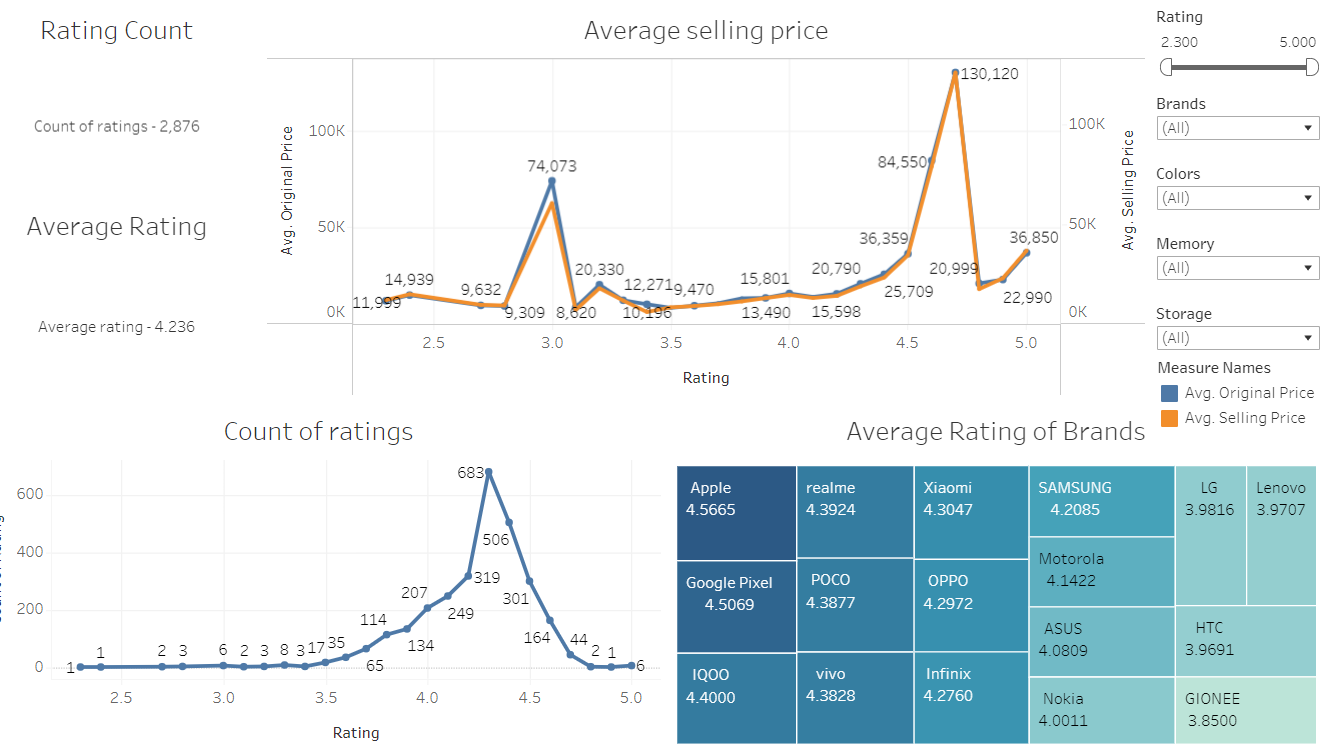


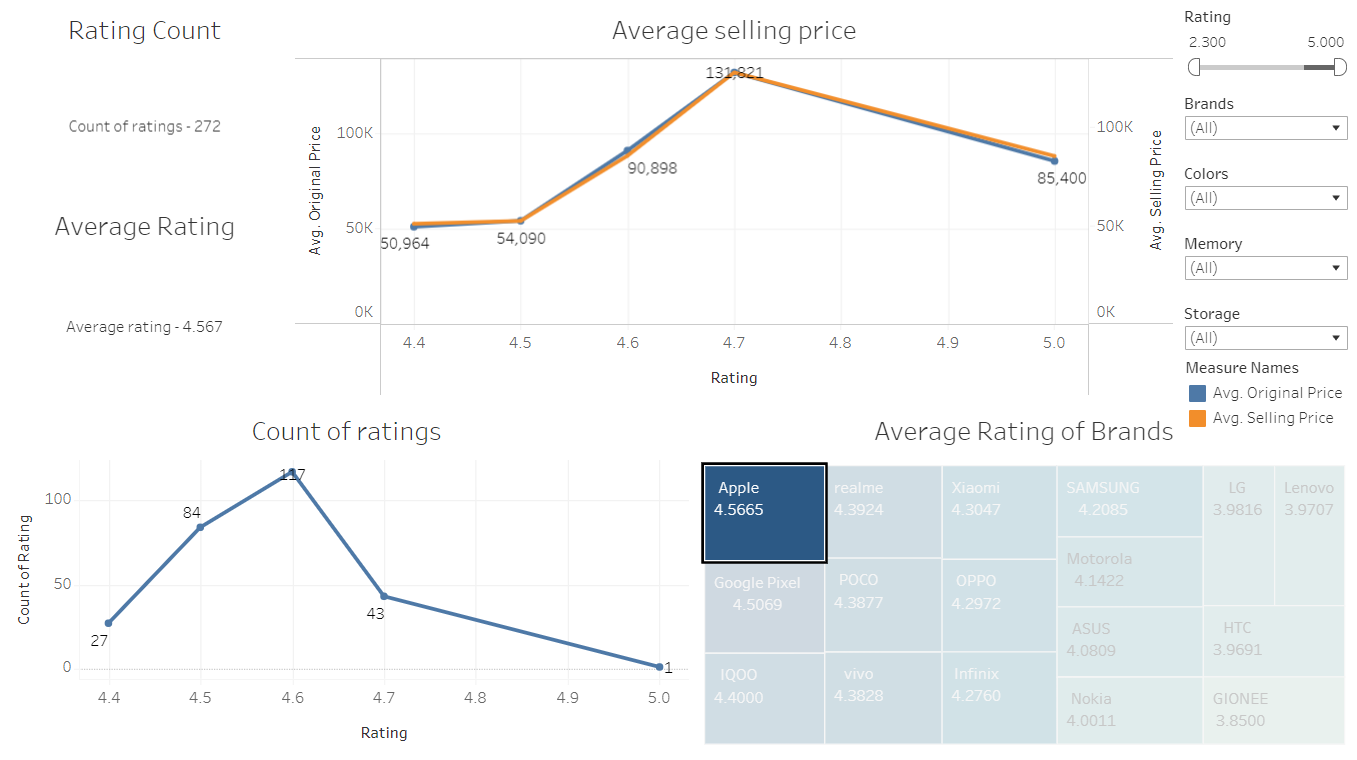
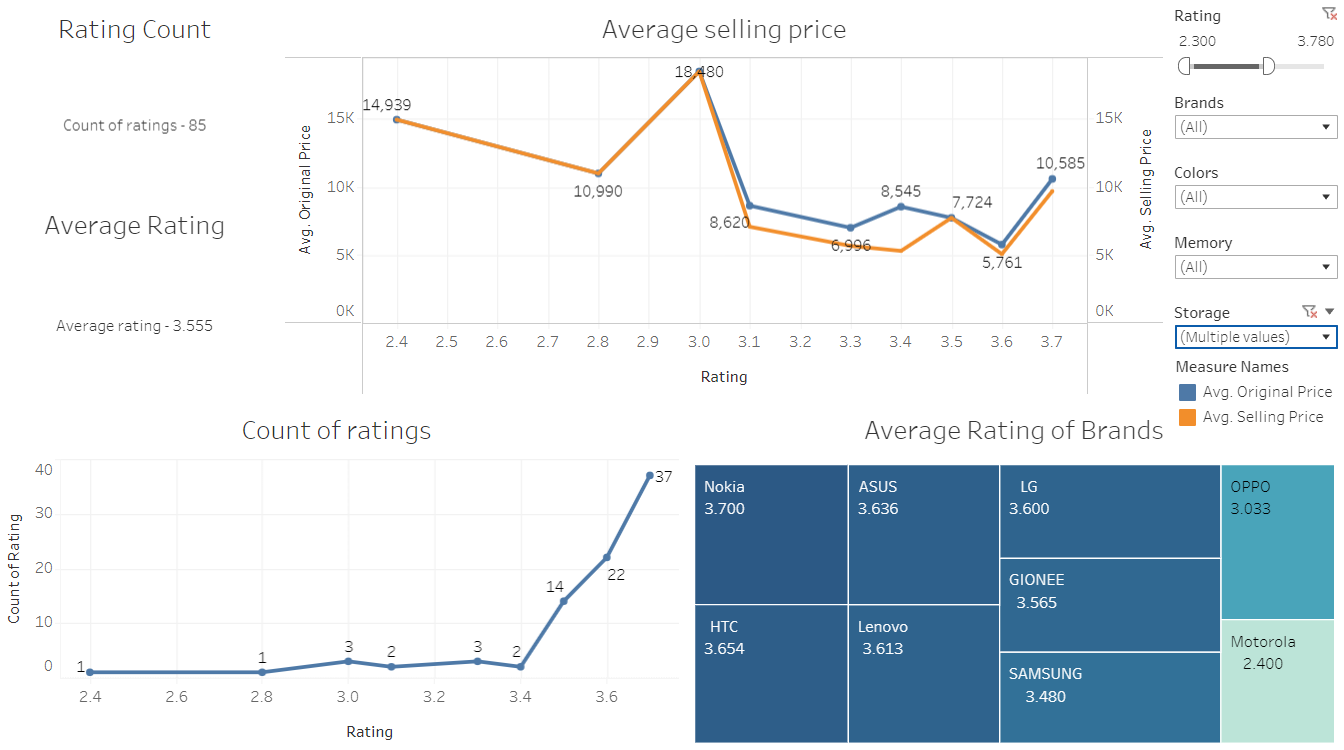


За потреби створимо нові поля для візуалізацій за прикладом



Знову збираемо візуалізації в дашборди, так як за ними ми можемо зробити більш детальний аналіз, як загальний, так і на певну групу за вибраними елементами завдяки простому фільтруванню, яке зберігається на обидва дашборди.



 (приклади)

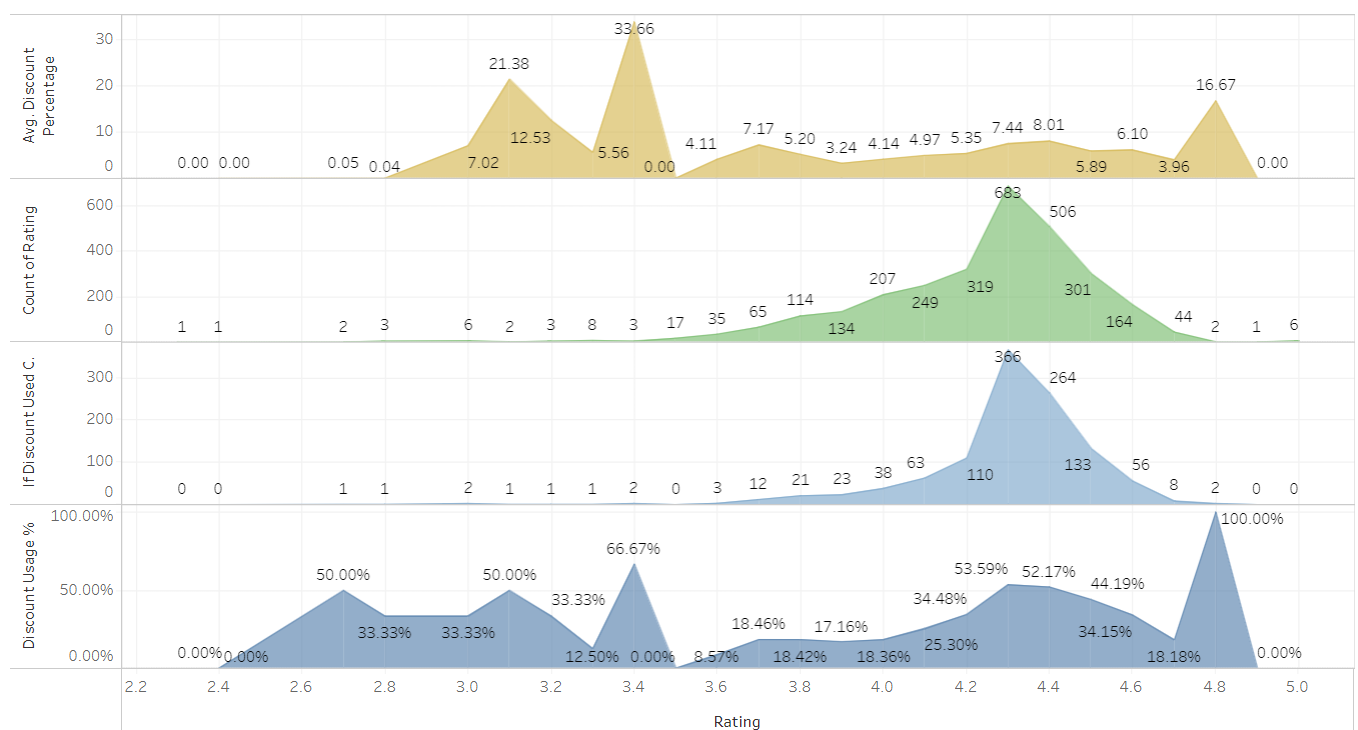
Отже, загально з цього дашборду можемо взяти, що встановили, що загальна кількість оцінок складає 2876, а середній рейтинг товарів 4.236. Далі висновок за результатами аналізу середньої ціни продажу та продажу по високій знижці:

Середня ціна продажу має середні значення приблизно 15000 з невеликим ростом. Видно два скачки цін на рівнях 3 та 4.7, що може вказувати на зміну стратегій ціноутворення або спеціальні пропозиції для певних категорій товарів.

В порівнянні зі звичайними продажами, продаж по високій знижці показує певну різницю. На рейтингу 3 різниця продажу зі знижкою складає близько 13000, що є помітно нижче в порівнянні зі звичайними продажами. На рівні 3.4 різниця продажу також має високе значення в 4000+. По рейтингу ж 4+, вплив знижок теж наявний, хоч і ненайбільший.

Розглядаючи візуалізацію кількості рейтингів, спостерігається низький рівень, близький до 5, проте від рівня 3.5 до 4.7 спостерігається значне зростання кількості оцінок, що в середньому коливається від 50 до 300, також виявлено пік на рівні 4.3 з найбільшою кількістю оцінок (683). Варто зазначити, що, що це може свідчити про підвищений інтерес споживачів до товарів з цими рейтингами близькими 4+.

Аналізуючи візуалізацію з брендами, можна побачити, що лідером за середнім рейтингом є бренд Apple (4.56), за яким слідують Google Pixel, IQOO, та найнижчі рейтинги мають бренди GIONEE (3.85), HTC і Lenovo. Хоча різниця у рейтингах брендів може бути невеликою, важливо звернути увагу на товари з низьким рейтингом, так як це може свідчити про певні проблеми або недоліки в їх якості або цікавості покупцям. Такі товари можуть потребувати додаткових маркетингових заходів для збільшення задоволеності споживачів та підвищення конкурентоспроможності на ринку.



Розглянемо другий дашборд. За цими даними можна сформулювати таке припущення:

Вцілому, середній показник використання знижок для товарів становить близько третини, що свідчить про те, що знижки відіграють важливу роль у стратегії продажу. Однак, товари з рейтингом 4.2 і вище мають високу концентрацію відгуків(і покупок), де в середньому половина з них було придбано зі знижками. І хоча ці знижки виявились невеликими, з середньою величиною 8%, така динаміка може свідчити про те, що споживачі більш схильні придбати вищеоцінені товари проте з знижкою, навіть без значних відсотків, вказуючи на деяку впевненість і задоволеність цими товарами.

На основі аналізу двох датасетів і враховуючи висновки з візуалізацій, можна створити наступні рекомендації для бізнесу:

Збільшення уваги до товарів з високим рейтингом: Посилення маркетингових заходів та акцентування уваги на продуктах з високими рейтингами (наприклад, 4.2 і вище) може бути ефективним стратегічним кроком, оскільки ці товари мають значну кількість відгуків і в середньому взято половина з них під час акцій зі знижками.

Оптимізація знижок для рейтингу 4.2+: Незважаючи на те, що в середньому знижки для товарів з високими рейтингами невеликі, їх вплив може бути значним. Варто розглянути можливість збільшення невеликих знижок для цієї категорії товарів, оскільки це може стимулювати покупців і підвищити конкурентоспроможність на ринку.

Увага до регіонів з високими показниками: Аналіз регіонального розподілу показників, таких як середній рейтинг чи кількість знижок, може допомогти визначити потенційно перспективні регіони для розвитку бізнесу. Наприклад, якщо виявлено, що у певному місті або регіоні товари з високим рейтингом мають високий попит, можна зосередити увагу на цьому ринку.

Стратегія ціноутворення: Розглянути можливість використання ціноутворення залежно від рейтингу товару. Наприклад, для товарів з високим рейтингом можна встановлювати невеликі, але стимулюючі знижки, що може привернути увагу покупців і збільшити обсяги продажів.

Підвищення якості та маркування продуктів: Продовжувати підвищувати якість продукції і зосередити увагу на позитивних відгуках та рейтингах, оскільки це може позитивно вплинути на споживачів і збільшити їхню лояльність до бренду.

Аналіз ефективності знижок: Провести детальний аналіз ефективності різноманітних знижок на підставі даних, щоб визначити оптимальні стратегії знижок для різних категорій товарів та рівнів рейтингу.

Дякую за увагу!