РЕТ – проєкт

на тему:

«Аналіз ефективності роботи мережі кав’ярень»

Виконавець:

Бразилевич Ірина

1. **Вступ**

Вдала кав'ярня – це магічне місце. Воно може бути місцем тусовки, в якій будуть збиратися компанії на поговорити, фрілансери попрацювати, або будуть приїздити сім'ями на сніданок і ранкову каву. А може бути місцем затишку і спокою, де можна насолодитися смачною кавою, спокійно посидіти на одинці або з приємним співбесідником. Як правило, кав’ярня може виконувати ці функції разом і або ці тенденції можуть змінюватися впродовж дня, в тому і магія. Тому важливо наповнити асортимент такого закладу продуктами, які мають високий попит, зможуть задовольнити будь – який смак і будуть мати добру якість. Але й не треба забувати про інтереси бізнесу: дохідність і ефективність. Тому й мене зацікавив датасет мережі кав’ярень для аналізу.

Мій РЕТ-проєкт спрямований на те, щоб дослідити ефективність роботи мережі кав’ярень. Виявити найбільш популярні продукти і продукти, які не користуються попитом. Дослідити пікові часи загрузки кав’ярень. Проаналізувати структуру пропозицій. При необхідності, надати рекомендації власникам бізнесу.

1. **Методологія**

Для аналізу планую використовувати такі технології:

* Excel таблиці та їхні функції для обробки та первинного аналізу даних;
* Google BigQuery та SQL-запити для обробки інформації і розрахунків;
* Tableau – для візуалізації даних;
* Python - статистичний аналіз для отримання інсайтів.

1. **Збір та очищення даних**

Для отримання даних для аналізу використовувала платформу Kaggle. Мені сподобався сет вигаданої кав'ярні [Maven Roasters](https://www.kaggle.com/datasets/ahmedabbas757/coffee-sales/data), що працює у трьох точках Нью-Йорка. Набір даних включає дату та час транзакції, місцезнаходження, а також відомості на рівні продукту. Дані для аналізу включають період січень – червень 2023р. Мені цікаво було дослідити, як змінювалися продажі Maven Roasters з часом. Що приносить найбільший прибуток бізнесу?

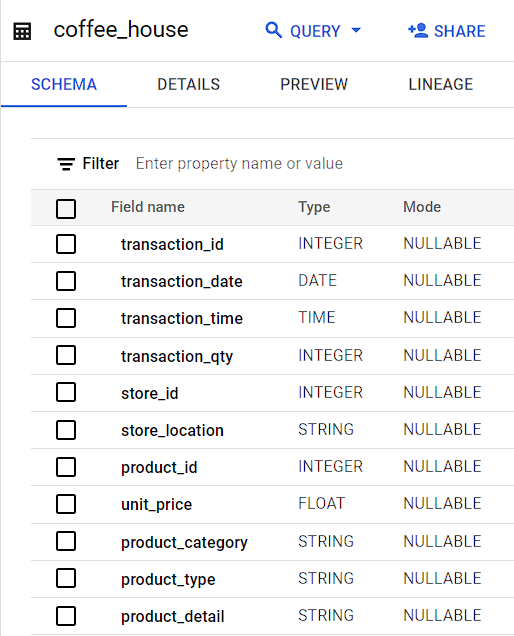
Сет було завантажено у форматі CSV-файлу. Для перевірки і очищення даних файл було завантажено до Power Query. Після всіх необхідних кроків перевірки не було виявлено помилок, дублікатів тощо. Таблицю було збережено у форматі xlsx для подальшого аналізу. А CSV-файл було завантажено у BigQuery.

1. **Аналіз даних**

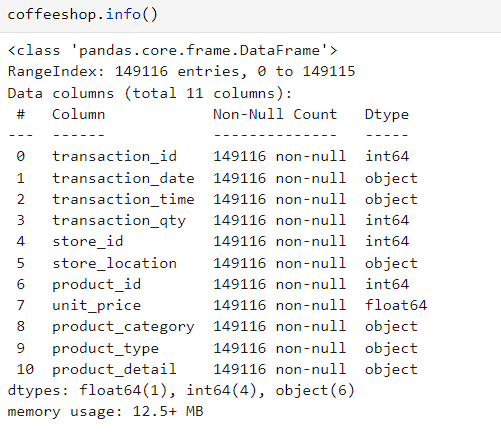
За допомогою Excel визначила кількість рядків в таблиці, яка відповідає кількості транзакцій за визначений період (149116) та визначила кількість стовбців (11), що відповідає заявленим параметрам таблиці, взятої з платформи Kaggle. В подальшому звіряла ці дані з даними, завантаженими у BigQuery та Tableau для відповідності повноти завантажених даних.

Тож маємо таку структуру таблиці:

|  |
| --- |
| *import pandas as pd*  *coffeeshop = pd.read\_csv('coffee-shop-sales-revenue1.csv')*  *print(type(coffeeshop))*  *coffeeshop* |

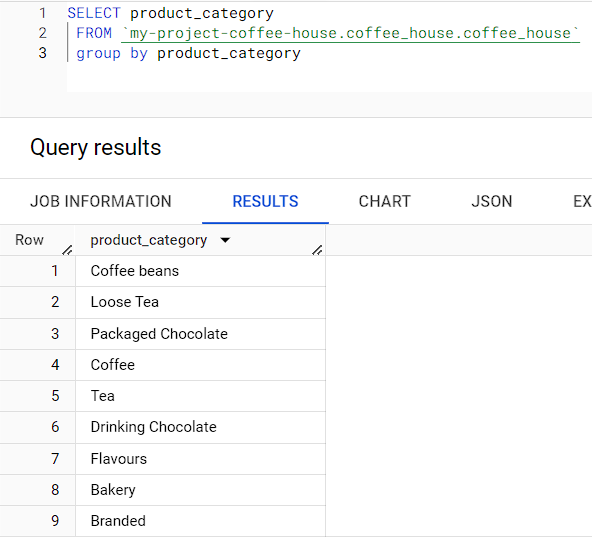


Також зробила перевірку в Python:

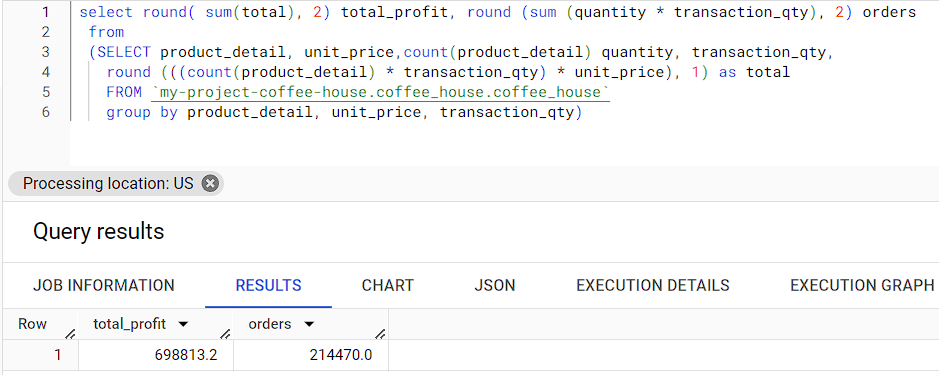


За допомогою SQL запитів та сортування в таблиці Excel визначила:

1. кількість локацій кав’ярень в мережі (store\_location) – 3;
2. кількість категорій продуктів product\_category - 9;
3. кількість типів продуктів (product\_type) – 29;
4. кількість продуктів в асортименті (product\_detail) – 80.



За допомогою SQL запиту обрахувала загальний об’єм продажу (total\_profit) за аналізований період та загальну кількість заказів продуктів (orders). В подальшому при розрахунках і створенні візуалізацій орієнтувалась на отримані результати для запобіганню помилок.



1. **Візуалізація даних**

Для візуалізації даних застосовувала Tableau і Python.

За допомогою Python провела статистичний аналіз впливання ціни продукту на кількість продажів. Для цього за допомогjю SQL запиту було сформовано таблицю з даними по кількості, ціні та загального продажу по кожному продукту. На підставі цих даних зроблено розрахунок кореляції та побудований графік лінійної регресії залежності продажів кількості продукту та загальних продажів від ціни продукту. Від’ємна кореляція -0,3193 в першому випадку та -0,0486 у другому свідчить про те, що влив є, тобто ми можемо зафіксувати вплив ціни продукту на продажі, але враховуючи кількісний показник кореляції, цей плив не дуже великий, але є. І це має враховуватися при формуванні асортименту.

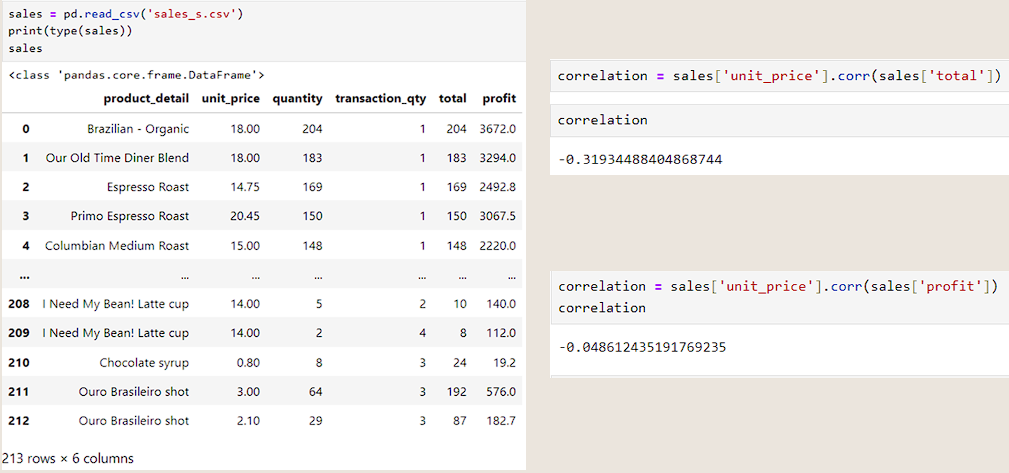
SQL запит: SELECT product\_detail, unit\_price, count(product\_detail) quantity,

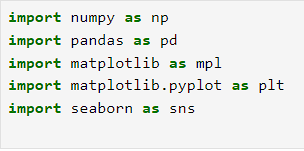
  transaction\_qty, (count(product\_detail)\*transaction\_qty) as total,

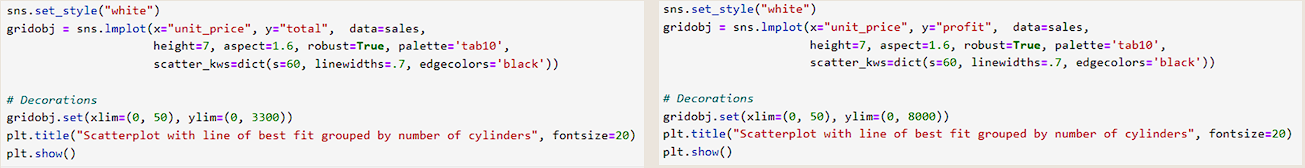
  round (((count(product\_detail) \* transaction\_qty) \* unit\_price), 1) as profit

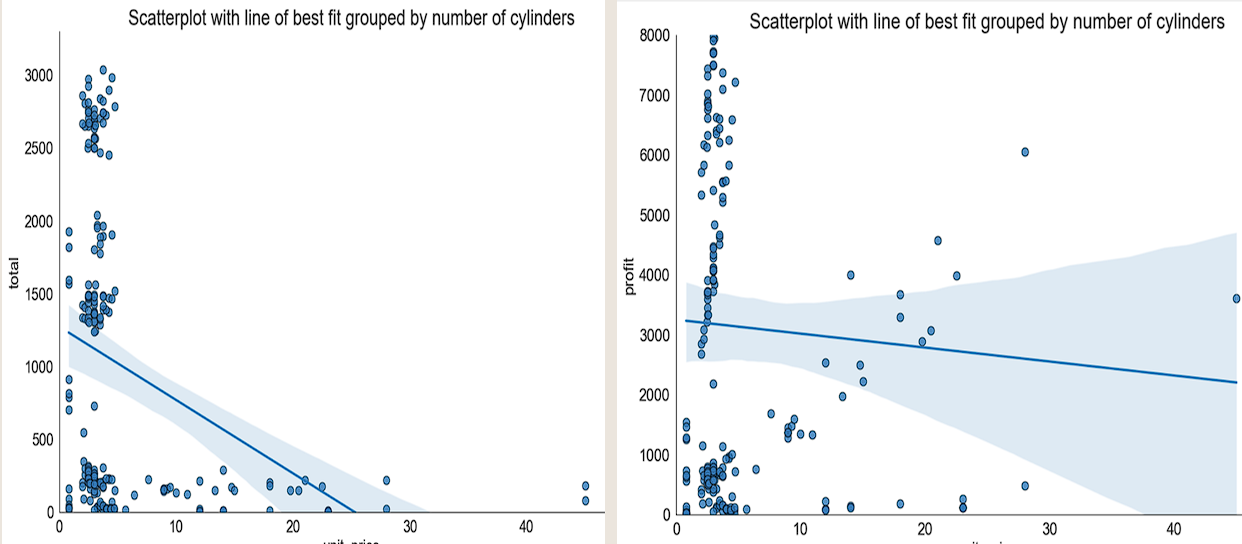
FROM `my-project-coffee-house.coffee\_house.coffee\_house`

  group by  unit\_price, transaction\_qty,product\_detail







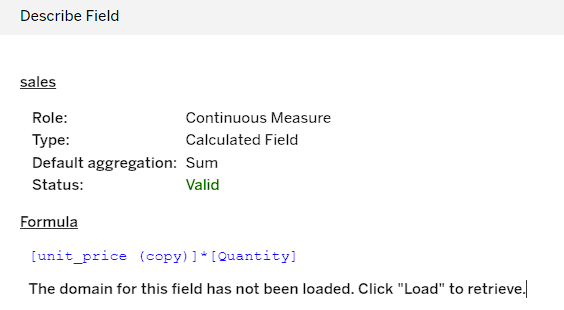


*price price*

[Візуалізаціі в Tableau](https://public.tableau.com/app/profile/irina.brazilevich/viz/coffee-shop_17134777495120/Dashboard5#1).

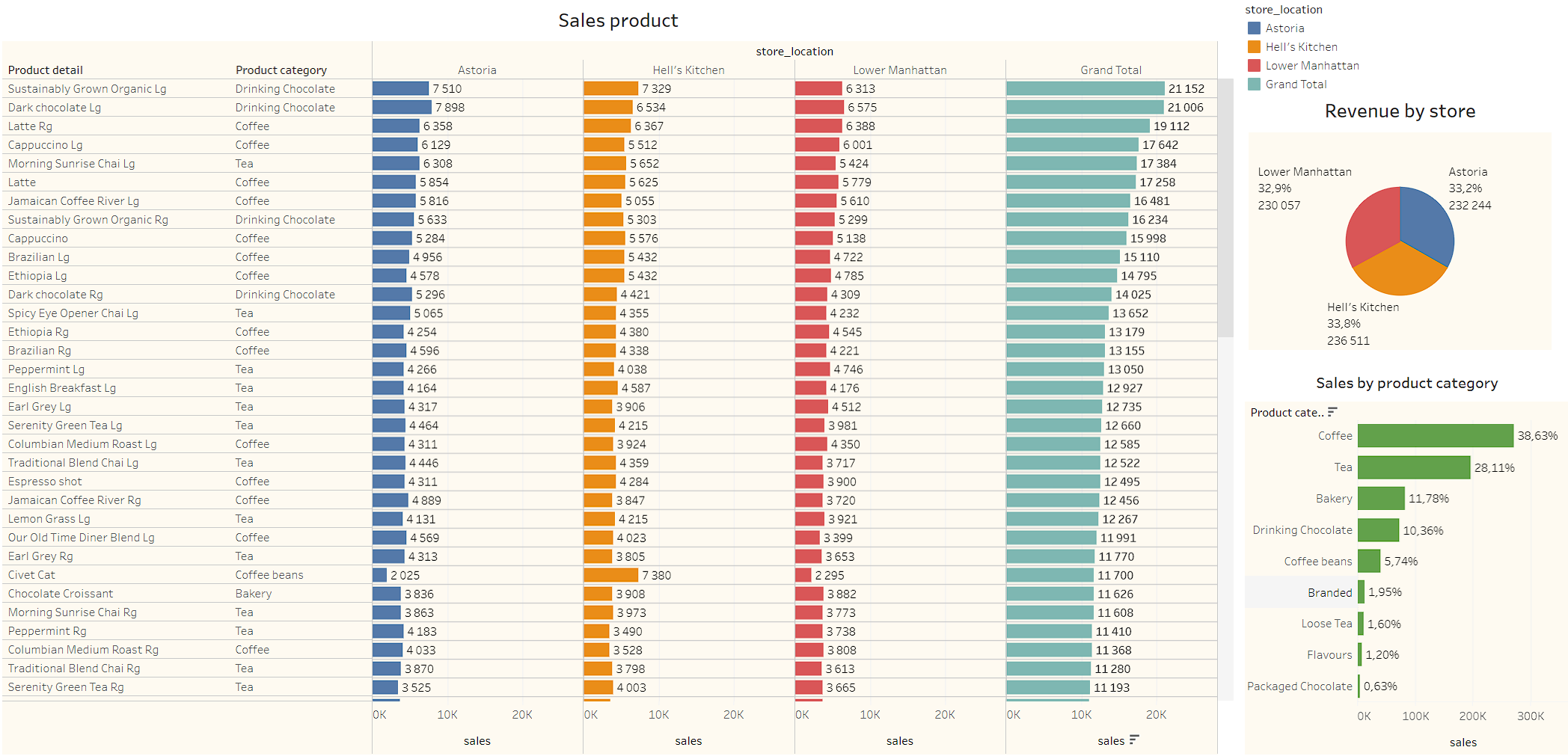
Загрузила CSV – файл з даними обраного датасету в Tableau для створення графіків та дашбрдів, що допоможуть у аналізі даних.

Для створення візуалізацій аналізу об’єму продажів, було створено розрахункове поле sales, що рахує об’єм продажу по кожному продукту.

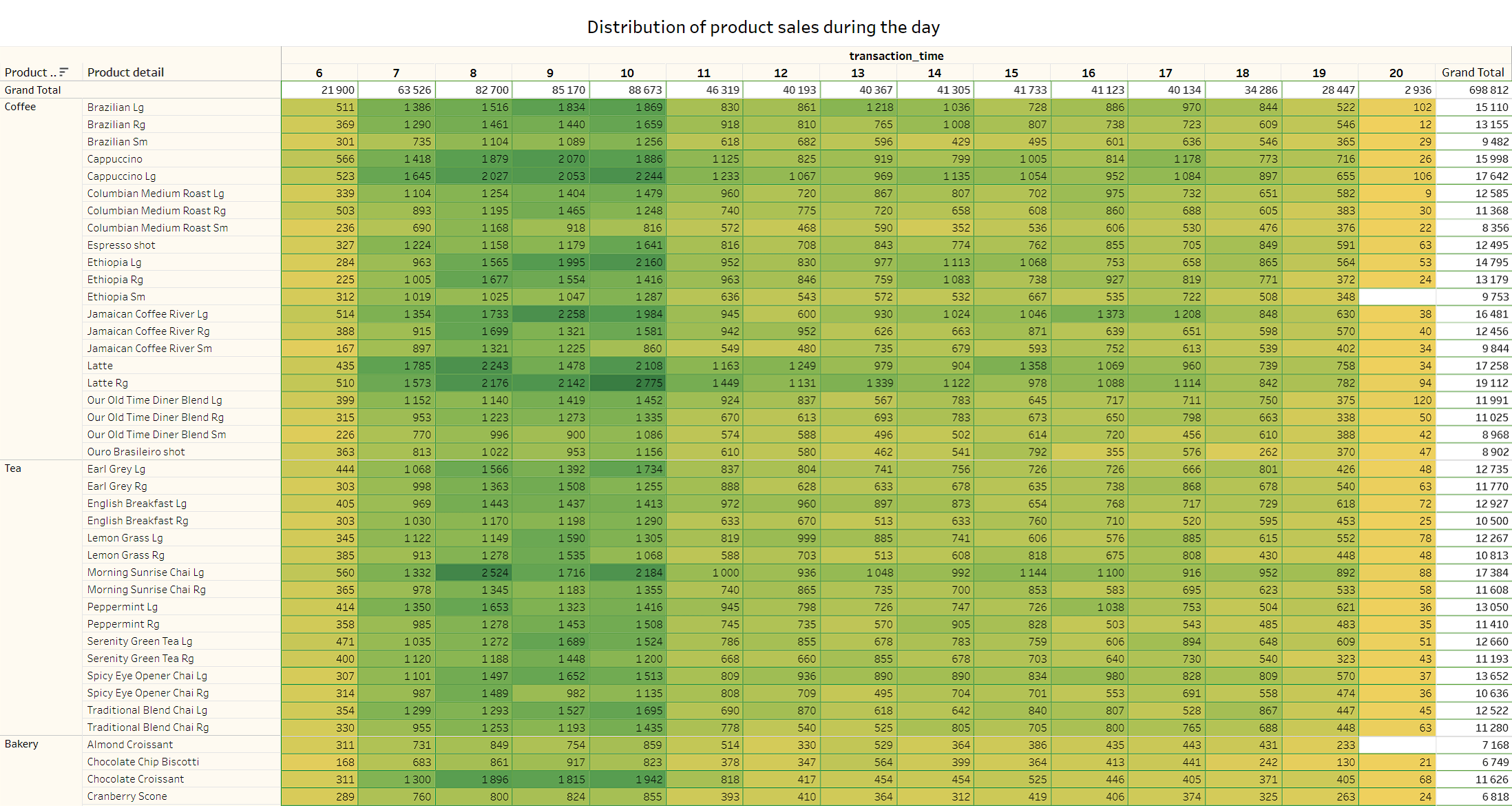


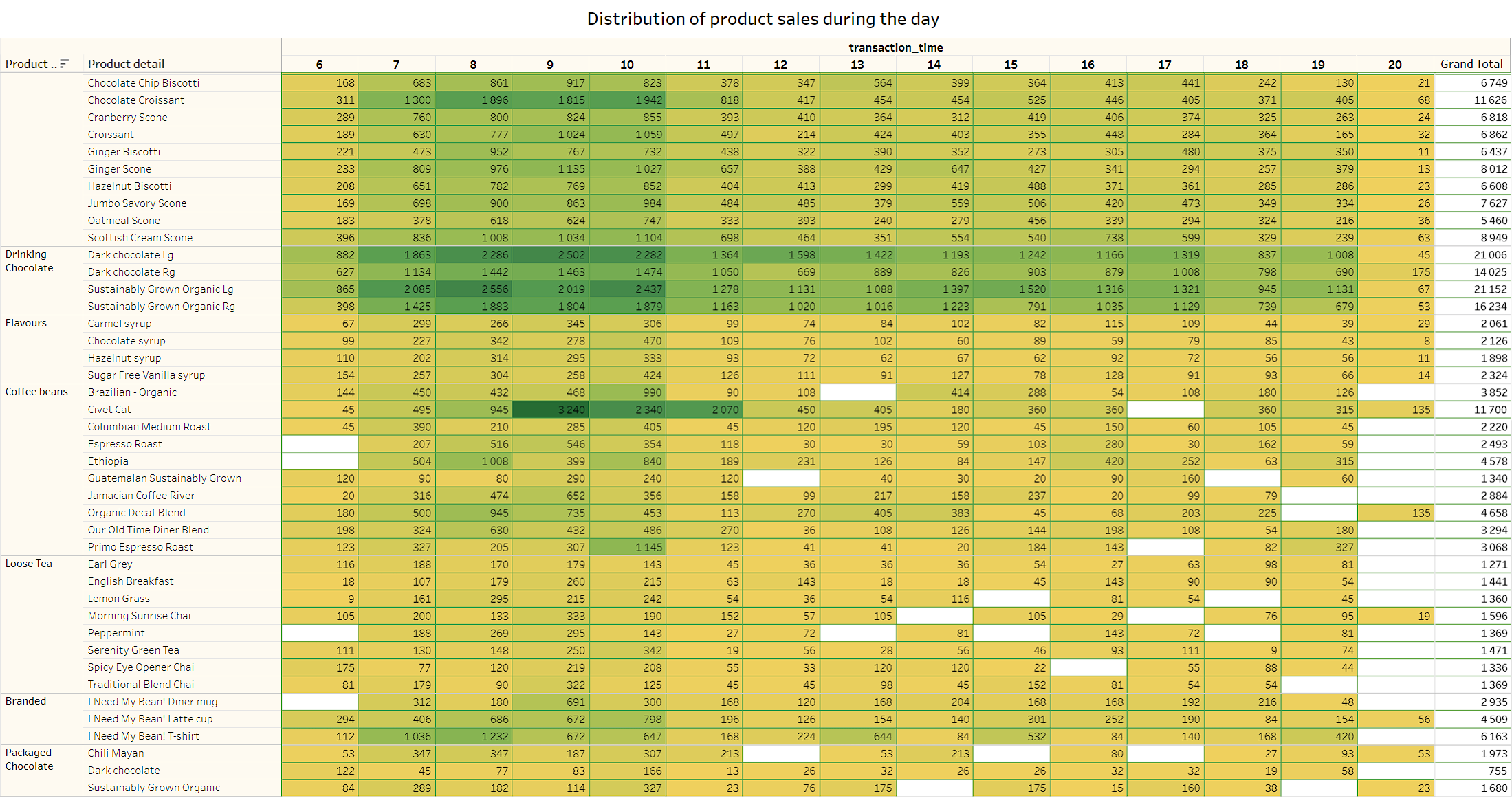
Параметр Quantity, це перейменований стовпець transaction\_qty, котрий показує, скільки продукту було замовлено за одну транзакцію.

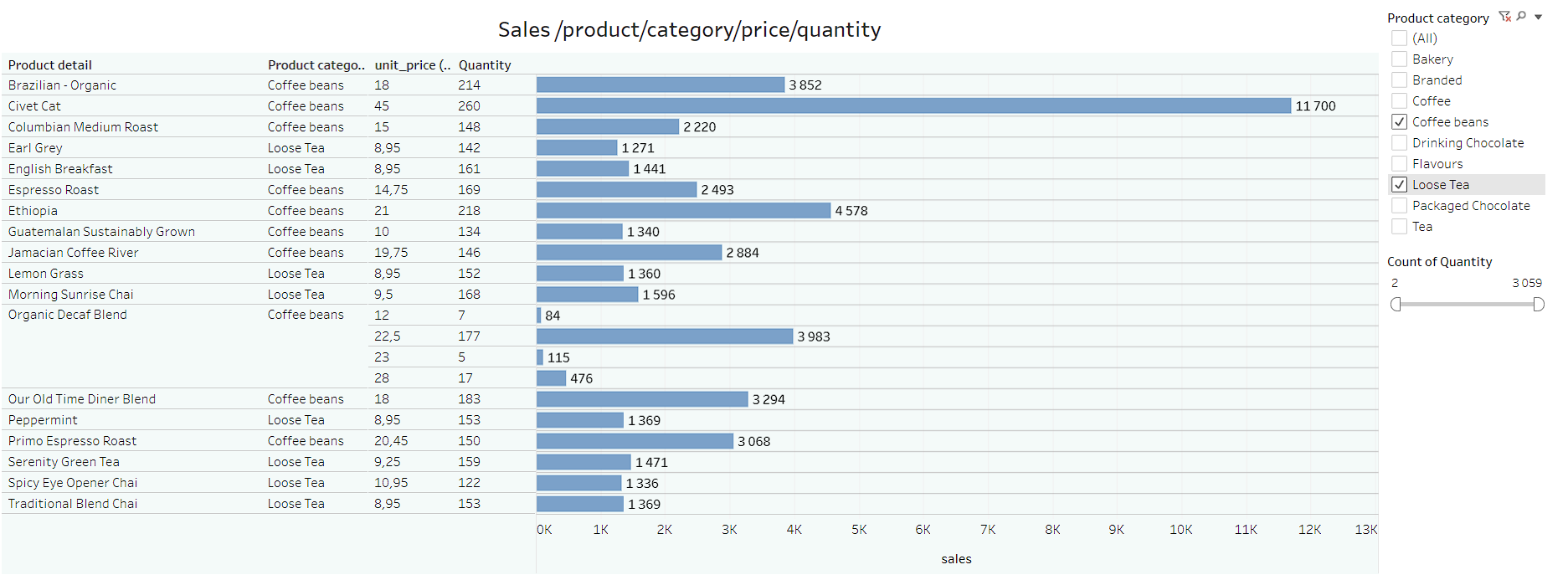
Створила дашборд SALES, де представлені візуалізації, на якій можемо побачити об’єми продажу по кожному продукту в розрізі трьох локацій кав’ярень та загальний об’єм продажу. Використавши функцію Actions, маємо змогу аналізувати продажі по кожній кав’ярні і у розрізі категорій. На круговій діаграмі бачимо розподіл загального об’єму продажів. По кожній точці об’єм продажу приблизно однаковий, що свідчить про професійний підхід до організації бізнесу. Також проаналізовано розподіл об’єму продажів в розрізі категорій. Бачимо, що найбільш користуються попитом категорії Coffee, Tea, Bakery та Drinking Chocolate. Категорії Branded , Coffee beans та Loose Tea не є популярними та мають бути розглянуті на доцільність їхнього перебування в асортименті. Завдяки такій візуалізації маємо змогу виділити найпопулярніші продукти і продукти з низьким попитом. Гістограма Sales by product category показує розподіл у відсотковому відношенні загального об’єму продажів в розрізі категорій.



Створила дашборд Distribution, в якому представлена візуалізація у вигляді таблиці в якій представлений розподіл продажу продуктів в залежності від часу дня. Бачимо, які продукти мають попит, які ні. Також таблиця містить дані загального об’єму продажу, об’єму продажу по кожному продукту за весь аналізований період. Ще є дані об’єму продажу за кожній часовий проміжок у продовж дня теж за весь період. Бачимо, які продукти мають попит, які ні. Часи роботи кав’ярень з 6 ранку до 8 години вечора. Бачимо, що пікове навантаження припадає на час з 7 до 10 ранку. Після 10 до 12 йде спадання продажів, з 12 до 17 години спостерігаємо більш сталий період, а з 17 до 20 вечора йде спадання.

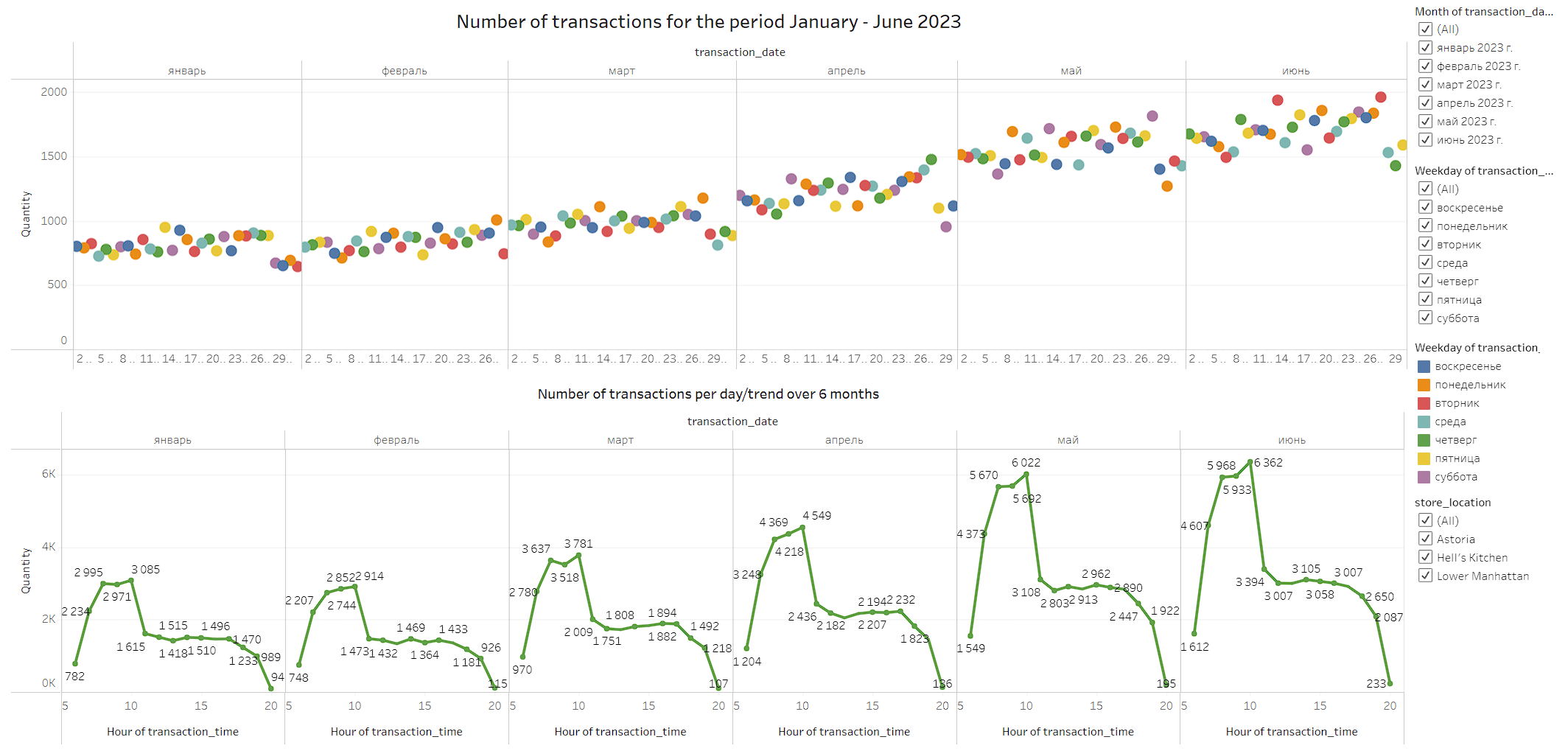




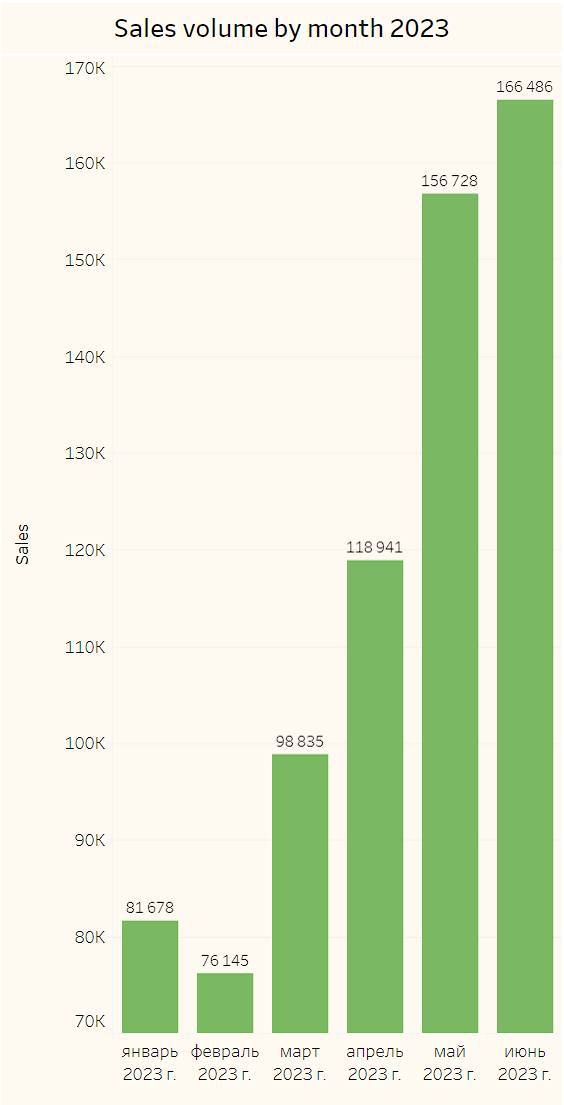
Створила дашборд Sales /product/category/price/quantity, на якій маємо візуалізацію у вигляді гістограми, яка надає інформацію по продуктам, їхнім категоріям, ціні та кількості продажів. Застосувавши фільтрацію можемо чітко визначити конкретні продукти, які мають низький попит. 

Створила дашборд Quantity by time, на якому два графіки. На першому графіку кількість транзакцій в розрізі днів тижня. Кожен колір відповідає дню тижня. Бачимо, що кількість транзакцій не залежить від цього параметру. Не має ніякої тенденції. Всі кружечки мають досить хаотичній характер розташування біля якогось середнього значення. Завдяки фільтрації можемо обирати дані по кожному місяцю або по розподілу об’єму транзакцій по дню тижня.

На другому графіку розглядаємо кількість транзакцій по всім кав’ярням в розрізі часу дня. Завдяки фільтрації можемо аналізувати дані по кожній кав’ярні.



Створила гістограму по якій можемо аналізувати зміну об’єму загальних продажів за аналізований період. За 6 місяців роботи кав’ярень спостерігається постійне зростання кількості транзакцій. А також вдвічі зріс об’єм продажів, що свідчить про гарне розташування кав’ярень, добре підібраний асортимент, кваліфікований персонал і гарно налаштовану організаційну структуру.



1. **Висновки та рекомендації.**

Висновки

Проведено аналіз роботи бізнесу Maven Roasters за період з 1 по 6 місяць 2023р., який має мережу кав’ярень з трьох закладів. Маю такі висновки: мережа добре організована, розташування закладів вдале, продажі постійно зростають. Заклади мають постійну загрузку, яка змінюється впродовж дня з однаковою тенденцією.

Рекомендації

При аналізі було виявлені категорії продуктів, які не користуються попитом., це категорії Coffee beans та Loose Tea. Окрім де-яких позицій з категорії Coffee beans, які зазначені вище і мають своїх прихильників, пропоную виключити повністю виключити категорію Loose Tea та більшість продуктів з категорії Coffee beans з асортименту.

Категорія Packaged chocolate має дуже низькі продажі, об’єм складає всього 0,63% від загального об’єму, тож пропоную цю категорію теж прибрати.

З категорії Bakery виключити продукти, які не мають попиту. Також пропоную додати продукти, які б мали доповнити асортимент цієї категорії на той період часу дня, коли клієнтам хотілось би з’їсти не солодкий продукт, а, наприклад, сендвіч з шинкою чи пиріг з не солодкою начинкою.

До категорії Flavours додати нових смаків сиропів.

В категорії Coffee переглянути позицію Ouro Brazileiro shot за ціною 2.1$, бо має дуже низький попит.

В категорії Branded, яка займає біля 2%, переглянути асортимент. Якщо бізнесу важливо залишити цю категорію, то вилучити товари, що не користуються попитом. Можливо додати нові товари, які будуть цікаві клієнтам.

Також хочу запропонувати ввести нову категорію продуктів, які б включали такі продукти, як горіхи, фруктові чіпси, печиво, тобто корисні перекуси, які можна взяти з собою.

І ще одна пропозиція, ввести в асортимент продукти вихідного дня, як рекламний захід, що були б цікаві і постійним клієнтам і зацікавили нових.