# 

# 

ITCS453 Data Warehouse and Data Mining

Project 1: Starbucks

Instructor

Asst. Prof. Dr. Songsri Tangsripairok

By

6388010 Teerapat Burasotikul

6388084 Rathapol Kittirudeekul

6388103 Viphu Sopanakitkosol

6388109 Suravich Thammapatee

6388159 Settanun Virattinun

Faculty of Information and Communication Technology

Mahidol University, Semester 2/2023

**Executive Summary**

Starbucks is a global coffeehouse chain with over 30,000 stores in over 80 countries. Recently, the company has faced various challenges in managing its inventory and operations, such as stock declines and supply chain disruptions. To address these problems, Starbucks aims to implement a data warehouse that can integrate data from multiple sources, such as point-of-sale systems, inventory systems, and suppliers.

The main objective of this project is to study the case of Starbucks and demonstrate how the company can use a data warehouse to solve its inventory management problem and enhance its operational efficiency and effectiveness. With the help of designing schemas in the data warehouse that will benefit the company to reflect its business processes and objectives, such as sales analysis, inventory optimization, supplier performance evaluation, and expansion. The data warehouse will also facilitate an ETL (extract-transform-load) process using Tableau Prep Builder that can clean, transform, and combine data from different sources into a unified format. The ETL process will ensure the quality and consistency of the data warehouse. Furthermore, the data are analyzed and visualized using Tableau Desktop. The analysis and visualization will provide insights into key performance indicators (KPIs), such as sales trends, inventory levels, stock-out rates, and expansion opportunities. The insights will help Starbucks make informed decisions to improve its inventory management and operations.

## Table of Contents

[Table of Contents](#_10kd4ooar5i1) 2

[**Overview**](#_tb10v13h5zao) **4**

[Starbuck overview](#_7qedug4z27w5) 4

[Project objective](#_i064053ij0pr) 4

[Scope of the Project](#_83et3lpdcfmx) 4

[Business requirement](#_a7gd8734p9b3) 5

[How data warehouse and BI important to Starbucks](#_f3ig3qvk98yt) 5

[Data sources](#_u4n1703gjzh2) 7

[Operational data (OLTP) dictionaries](#_yyhmsjdstdu3) 8

[Branch](#_i0lugkfpb6fu) 16

[Supplier](#_mj5rthywhsaj) 17

[Menu](#_ppowis2g8vv1) 18

[Stock](#_6k3g86tpx3q5) 19

[Order](#_8mcftj4grf9z) 20

[Best seller products](#_ukdxb4tqcd0u) 23

[Sell volume](#_ddlpcsnpnus4) 24

[Branch Location](#_cg2k8ic1re3q) 25

[Most used ingredient](#_n2h09hpkf0tb) 25

[Branch Inventory](#_xevbky87ld5i) 26

[Area suitable for new branches](#_fkuknlev1rpq) 27

# Overview

## Starbuck overview

Starbucks is a global coffeehouse chain founded in Seattle, Washington in 1971. The company operates over 30,000 stores worldwide in more than 80 markets. Starbucks is known for its premium coffee, handcrafted beverages, and extensive food menu.

In addition to its physical stores, Starbucks also operates a successful mobile app and rewards program, with over 23 million active members in the United States alone. The company is committed to sustainability and ethical sourcing practices, and it has established partnerships with farmers worldwide to ensure that it sources high-quality, responsibly grown coffee.

Starbucks has been at the forefront of innovation in the coffee industry, introducing new products such as cold brew and nitro cold brew and expanding its food offerings to include breakfast sandwiches, pastries, and salads. The company has also embraced technology, offering mobile ordering and payment and integrating artificial intelligence and machine learning to personalize the customer experience.

Starbucks has established itself as a leading brand in the global coffee industry, known for its high-quality products, innovative offerings, and commitment to sustainability and social responsibility.

## Project objective

The objective is to undertake an analysis and design process for a data warehouse tailored to a specific business case study. The next step is to implement this design by implementing the data warehouse for business intelligence purposes. This implementation will be done using Tableau Desktop and Tableau Prep, two powerful software tools that can help to organize and visualize complex data sets in a user-friendly manner. By leveraging these tools, businesses can gain valuable insights into their data that can inform decision-making processes and ultimately drive better outcomes.

## Scope of the Project

As Starbucks Thailand is expanding its business by opening new branches, the company has run into an issue, in this case, the ingredient shortage. It appears that the problem comes from the supply chain management process, in which the stock distributor cannot keep up with the varying demand of each branch as the customers have different preferences for the products. Consequently, some stores are encountering an issue where they cannot deliver some menu items to the customer since the ingredients ran out. Moreover, while new locations of Starbucks are open, they are further from the existing supplier. Since the company relies heavily on local ingredients sources, information on local suppliers within proximity of the store is required. This problem hinders the company's goal of extending its business to other parts of Thailand.

## Business requirement

| No. | Description |
| --- | --- |
|  | BI should provide the information of weekly ingredients consumption of every province. |
|  | BI should provide the information of ingredient inventory level of every branch |
|  | BI should be able to display the best seller products of each branch. |
|  | BI should provide the area that a new branch can be opened based on the local suppliers and customer groups. |
|  | BI should provide the information of the ingredient that is used the most. |
|  | BI should be able to display the weekly sell volume of each product in each branch. |

## How data warehouse and BI important to Starbucks

Data warehouse and business intelligence (BI) are critical tools for Starbucks as they help the company make informed business decisions based on data insights. Here are some ways in which data warehouse and BI are important to Starbucks:

* Sales and revenue analysis: Starbucks uses its data warehouse and BI tools to analyze sales and revenue data, including transaction data from its mobile app and point-of-sale systems. This data helps the company understand customer buying behavior, identify trends, and forecast future sales.
* Inventory management: Starbucks uses data from its data warehouse to monitor inventory levels and analyze product usage patterns. This information helps the company optimize inventory management and ensure sufficient inventory to meet customer demand.
* Customer analytics: Starbucks leverages data from its rewards program and other sources to gain insights into customer behavior and preferences. This information helps the company personalize its marketing efforts and improve the customer experience.
* Supply chain optimization: Starbucks uses data from its data warehouse to track supplier performance, monitor transportation and logistics, and optimize its supply chain operations. This helps the company ensure it has the right products in the right places at the right time.
* Store performance analysis: Starbucks uses data from its data warehouse to analyze store performance, including sales, foot traffic, and customer satisfaction. This information helps the company identify opportunities to improve store operations and customer experience.

Data warehouse and BI are critical to Starbucks as they help the company make data-driven decisions, improve efficiency, and enhance the customer experience. By leveraging data insights, Starbucks can continue to innovate and maintain its position as a leader in the global coffee industry.

# Designing databases

## Data sources

The data required for this project are the operational data, consisting of data on every branch opened in Thailand from Keggle, data on every menu item sold in Starbucks, data on every ingredient used by Starbucks, and the ordering data in each branch, as well as customer information. Other data are from external sources, including every local ingredient supplier in Thailand, their production quantity, and their location. Lastly, since the Starbuck customer groups are around 20 to 60 years old, additional data, such as the population in Thailand by age, is also acquired.

| **Data** | **Source** |
| --- | --- |
| **Beverage Products & ingredient** | <https://www.starbucks.co.uk/sites/starbucks-uk/files/2023-01/WIN23_UK_AllergenBook_CORE_BEVERAGE_v07.pdf> |
| **Branch location** | <https://www.kaggle.com/datasets/kukuroo3/starbucks-locations-worldwide-2021-version> |
| **Ordering data** | generated |
| **Customer information** | generated |
| **Dairy production information** | <https://data.go.th/dataset/item_2a1f6f83-4721-4f3d-894e-f886a6d6f157>, generated |
| **Coffee production information** | <https://data.go.th/dataset/coffee62>, generated |
| **Sugar production information** | <https://data.go.th/dataset/sugar-factory>, generated |
| **Number of Population by Age and province** | <http://statbbi.nso.go.th/staticreport/Page/sector/EN/report/sector_01_11101_EN_.xlsx> |
| **All provinces name in Thailand** | <https://github.com/kongvut/thai-province-data/blob/master/csv/thai_provinces.csv> |

## 

## Operational data (OLTP) dictionaries

Ingredient

| Column | Data types | Description | Example |
| --- | --- | --- | --- |
| id\_ingredient | int | Unique identifier for ingredient | 1234 |
| ingredient\_name | String | Name of the ingredient | Caramel drizzle |
| stock | int | Amount of ingredient available | 35 |

The Ingredient table has an id\_ingredient as an integer that uniquely identifies each ingredient, ingredient\_name as a string that stores the name of the ingredient, and a stock as a numeric value that indicates the amount of the ingredient available. The id\_ingredient is a primary key of this table

Product

| Column | Data types | Description | Example |
| --- | --- | --- | --- |
| id\_product | int | Unique identifier for product | 38 |
| product\_name | String | Name of the product | Vanilla Crème |
| product\_type | String | Type of the product | Frappuccino |

In the product table, the primary key is id\_product, an integer that uniquely identifies each product. Other fields are product\_name, a string that stores the name of the product, and product\_type, a string that stores the type of the product, which can be Espressos, Iced coffee, Mocha & hot chocolate, teas, as well as Frappuccino.

Branch

| Column | Data types | Description | Example |
| --- | --- | --- | --- |
| id\_branch | String | Unique identifier for a branch | 33728-152607 |
| branch\_name | String | Name of the branch | camp-davis-88 |
| latitude | float | Latitude location of the branch | 13.721278 |
| longitude | float | Longitude location of the branch | 100.566011 |
| address | String | Street address of the branch | Sukhumvit 24 Soi |
| city | String | Branch city | Bangkok |
| id\_subdivision | int | Branch subdivision number | 10 |
| zipcode | int(5) | Postal code of the city | 10110 |

Branch table contains id\_branch as a primary key. This field holds a string that uniquely identifies each branch. Other fields are branch\_name, a string that stores the name of the branch, latitude and longitude, float values that define a location of the branch, address for storing a string of an address of the branch, city for storing a string of the city that the branch is located, id\_subdivision, an integer that stores a subdivision number, and a zipcode for storing a zip code.

Customer

| Column | Data types | Description | Example |
| --- | --- | --- | --- |
| id\_customer | int | Unique identifier for a customer | 43 |
| first\_name | String | Customer’s first name | John |
| last\_name | String | Customer’s last name | Smith |
| phone\_number | String | Customer’s phone number | 0987654321 |

In the customer table, the primary key is id\_customer, which contains an integer that uniquely identifies each customer. Other fields are first\_name and last\_name, which store strings of customers’ names. The last field is a phone\_number, containing a string of each customer’s phone number.

Supplier

| Column | Data types | Description | Example |
| --- | --- | --- | --- |
| id\_supplier | int | Unique identifier for a supplier | 27 |
| sup\_name | String | Supplier’s name | Thai-Denmark |
| sup\_type | String | Type of product supplied | Dairy |
| production | float | Amount of supplier’s production(tons) | 23.4 |
| city | String | Supplier’s city location | Saraburi |

In the supplier table, the primary key is id\_supplier which contains an integer that uniquely identifies each supplier. Other fields are sup\_name, sup\_type and city, which store strings of the supplier name, the type of product, and the supplier city’s name. The production field store float which is the amount of the production in tons.

Order\_product

| Column | Data types | Description | Example |
| --- | --- | --- | --- |
| id\_order | int | Unique identifier for an order | 971 |
| id\_product | int | Unique identifier for a product | 30 |
| quantity | int | Quantity of a product | 2 |

In the order product table, the primary key is id\_order which contains an integer that uniquely identifies each order. Other fields are id\_product and quantity, which store integers that uniquely identify product and show the quantity of product.

Recipe

| Column | Data types | Description | Example |
| --- | --- | --- | --- |
| id\_product | int | Unique identifier for a product | 6 |
| id\_ingredient | int | Unique identifier for a ingredient | 20 |

In the recipe table, the primary key is id\_product which contains an integer that uniquely identifies each product. Another field is id\_ingredient which contains an integer that uniquely identifies each ingredient.

Stock

| Column | Data types | Description | Example |
| --- | --- | --- | --- |
| id\_branch | String | Unique identifier for a branch | 33713-146284 |
| id\_ingredient | int | Unique identifier for a ingredient | 21 |
| quantity | int | Quantity of an ingredient | 291 |

In the stock table, the primary key is id\_branch which contains a string that uniquely identifies each branch. Other fields are id\_ingredient and quantity, which both store integers that uniquely identify ingredient and show quantity of ingredient respectively

branch\_supplier

| Column | Data types | Description | Example |
| --- | --- | --- | --- |
| id\_branch | String | Unique identifier for a branch | 20245-200745 |
| id\_supplier | int | Unique identifier for a supplier | 79 |

In the branch supplier table, the primary key is id\_branch which contains a string that uniquely identifies each branch. Another field is id\_supplier which contains integer that uniquely identifies each supplier

Population

| Column | Data types | Description | Example |
| --- | --- | --- | --- |
| region | String | Unique identifier for a region | Central |
| province | String | Unique identifier for a province | Bangkok |
| sex | String | Unique identifier for sex | Male |
| population | int | Quantity of population | 200,000 |

In the population table, the primary key is region which contains a string that uniquely identifies the region. Other fields are province and sex, which contain strings that uniquely identify province and sex respectively. The last field is population which contains integers that show the quantity of the population.

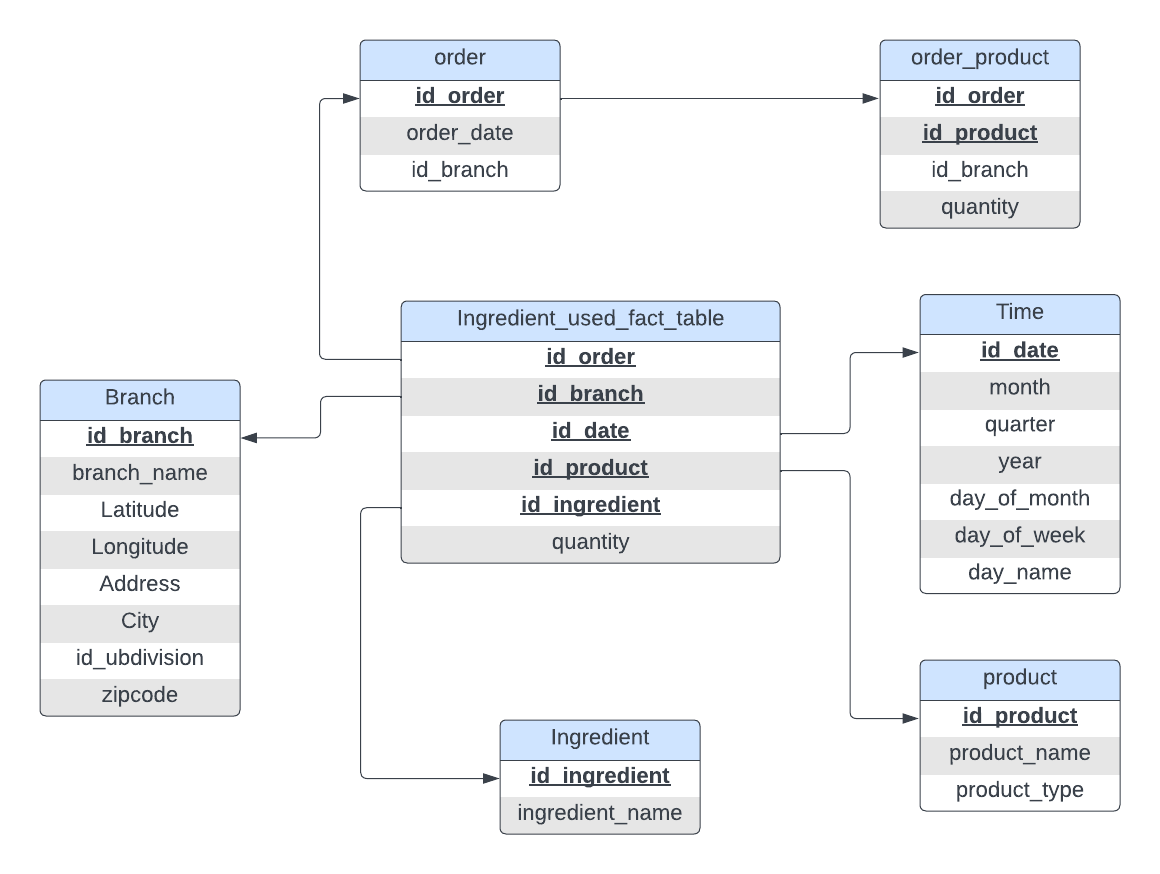
Order

| Column | Data types | Description | Example |
| --- | --- | --- | --- |
| id\_order | int | Unique identifier for order | 047 |
| order\_date | datetime | Date and time of the order | 28/02/2023 |
| id\_customer | int | Unique identifier for customer | 47 |
| id\_branch | int | Unique identifier for branch | 20245-200745 |

In the order table, the primary key is id\_order which contains an integer that uniquely identifies each order. Other fields are order\_date which contains datetime that show the date and time of the order. The last two fields are id\_customer and id\_branch which contain integers that uniquely identify customer and branch respectively.

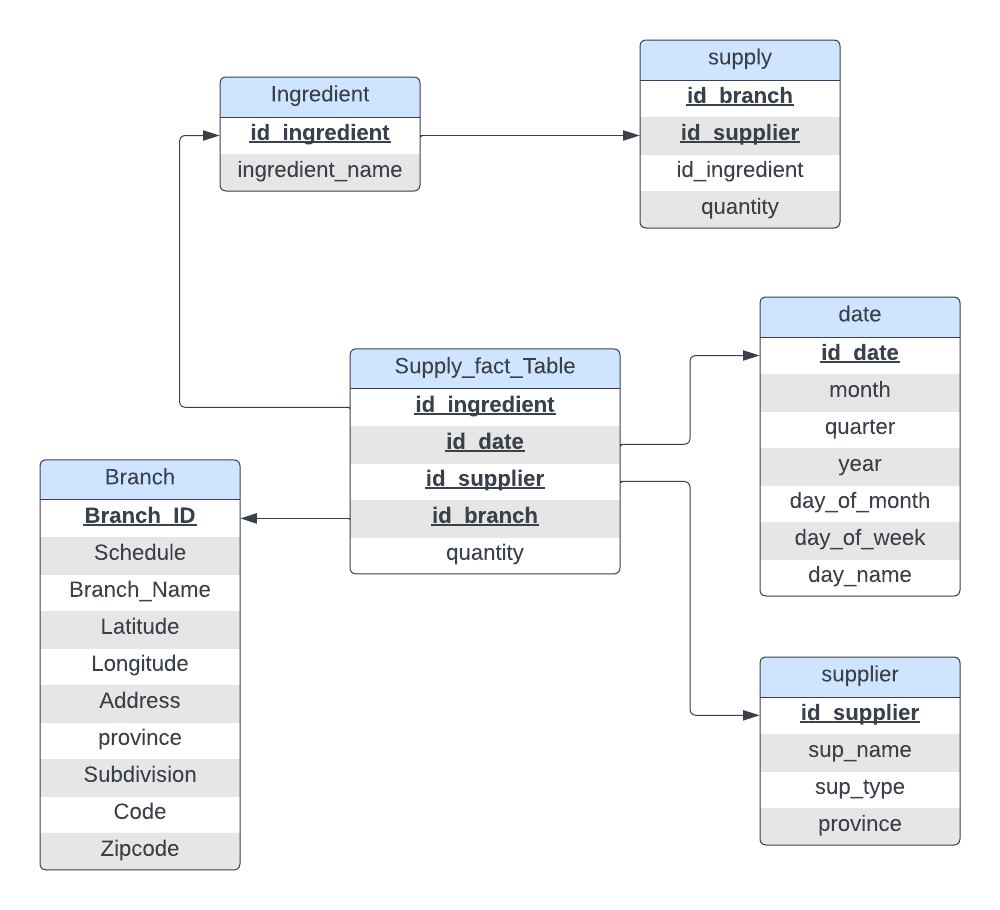
# Data warehouse design

**Ingredient\_used fact table**

****

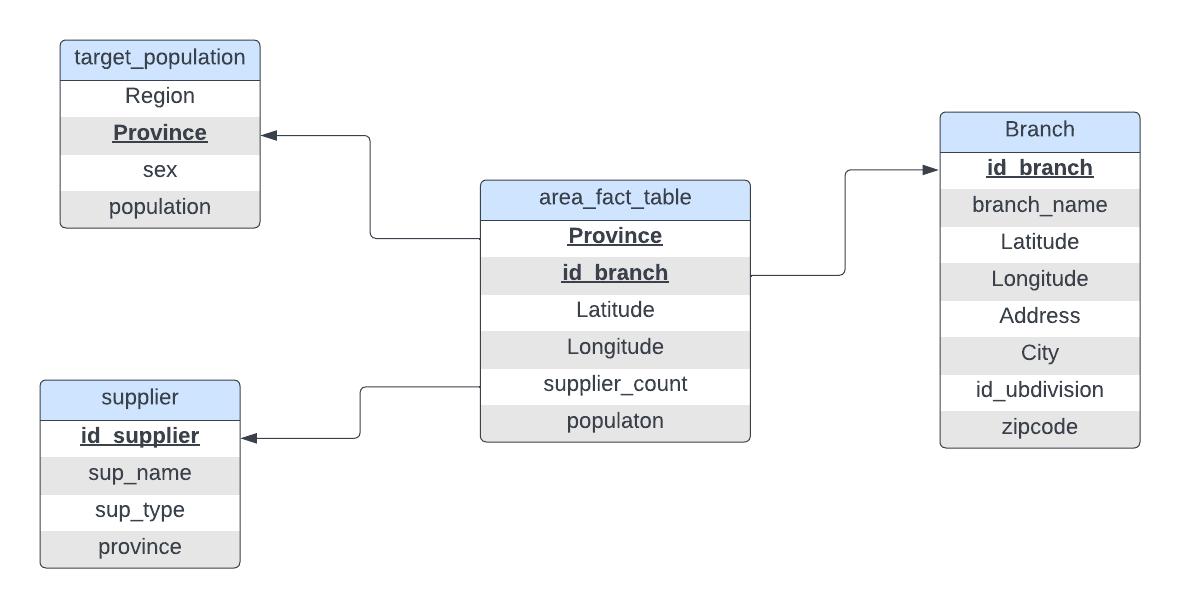
The first star schema is an ingredient used fact table. This fact table represents the usage of ingredients for each branch in Thailand. In this fact table, there are five dimensions. Order dimension represents the sale to our customer, which will link to order\_product, representing the menu in that order. The time dimension represents the time that orders have been sold. The product represents the product sold in that transaction. In the product dimension, it will provide the product name and type. The ingredient dimension represents the ingredient used in that transaction. Lastly, the branch name represents which branch performs that transaction.

**Supply fact table**

****

The following snowflake schema is the supply fact table. In the supply fact table, there are four dimensions. The first dimension is the Ingredient. In the ingredient dimension, there are ingredient name attributes, and the ingredient dimension is linked to the supply dimension, which represents data when the branch refill ingredient in their stock. Finally, the date dimension represents the date a branch fills its stock. The date dimension includes month, quarter, year, day of the month, day of the week, and day name. All of the attributes in date are used for the roll-up and drill-down process. The supplier dimension represents the supplier of that Ingredient. In the supplier dimension, there are three attributes: sup\_name as supplier name, sup\_type as supplier type, and the province as the province of that supplier. Finally, the branch dimension represents the branch that was refilled supply.

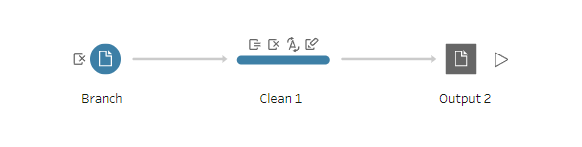
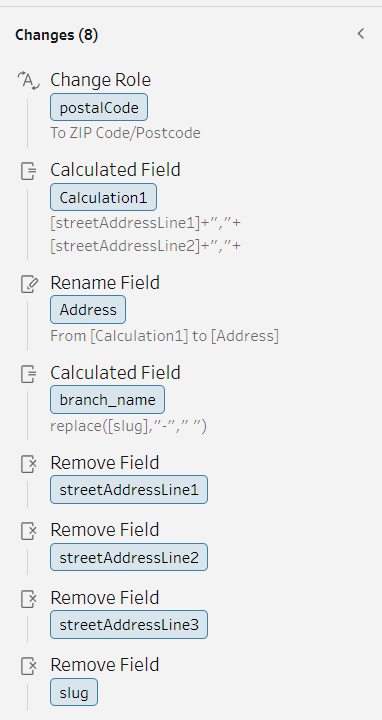
**Area fact table**



The area fact table evaluates the population in different regions of Thailand to determine the ideal location for the next branch. This star schema comprises three dimension tables - target population, supplier, and branch dimensions. The target population dimension provides insights into the individuals residing in each region, while the supplier dimension focuses on the location of suppliers. Finally, the branch dimension will primarily be used to study the location of existing Starbucks branches to determine where the next one should be established.

# ETL process

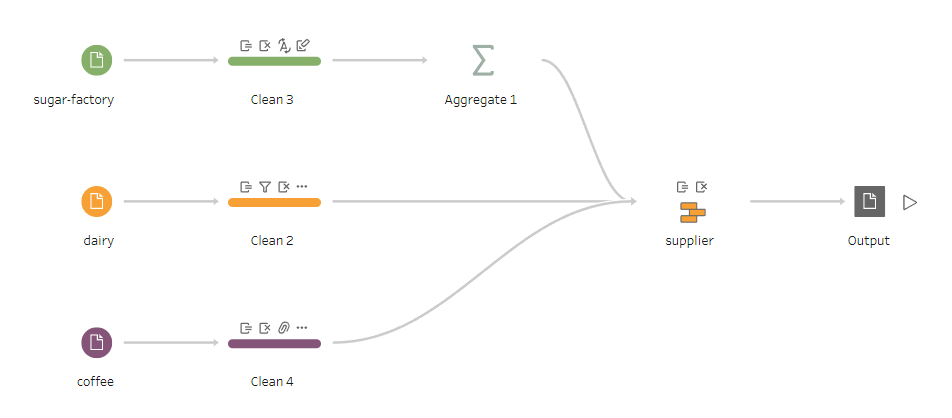
## Branch

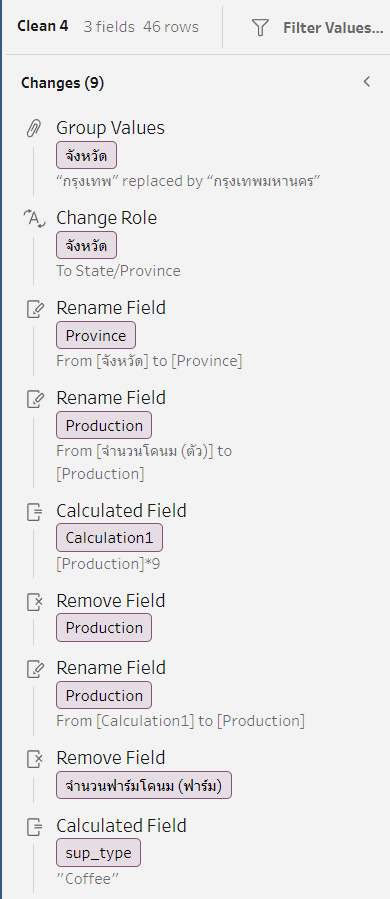
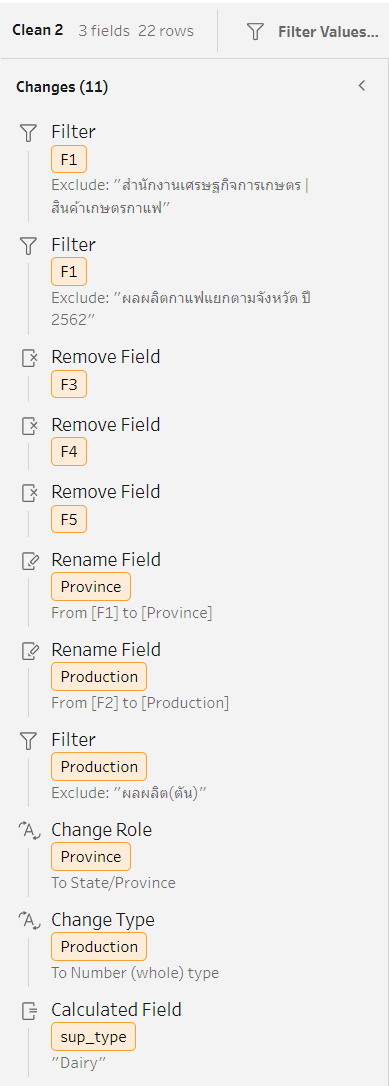
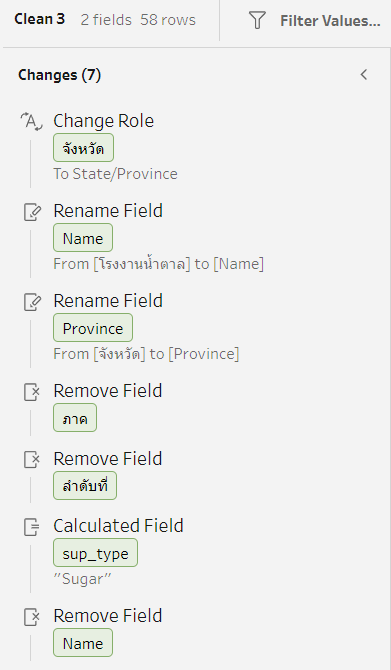


The ELT process for Branch data:

1. Cleaning Branch file
   1. Change row data type from int to location (Zip Code/Postcode)
   2. Create a new calculated field by concatenating between streetAddressLine1, streetAddressLine2, and streetAddressLine3.
   3. Rename calculate field to Address.
   4. Create branch\_name calculated field by using data form slug field but replace “-” with space.
   5. Remove field streetAddressLine1, streetAddressLine2, streetAddressLine3, and slug.
2. Output file as a csv.

## Supplier

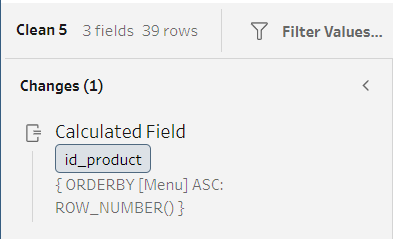
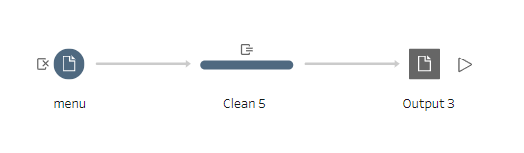




The ELT process for Supplier data:

1. Cleaning sugar-factory file
   1. Change จังหวัด field data type from stringto State/Province
   2. Rename โรงงานน้ำตาล field name to Name
   3. Rename จังหวัด field name to Province
   4. Remove field ภาค, and ลำดับ
   5. Create field name sub\_type that contain string data which all row are Sugar
   6. Remove Name field
2. Cleaning dairy data file
   1. Remove data “สำนักงานเศรษฐกิจการเกษตร | สินค้าเกษตรกาแฟ”, and “ผลผลิตกาแฟแยกตามจังหวัดปี 2562” from field F1
   2. Remove field F3, F4, F5
   3. Rename F1 to Province
   4. Rename F2 to Production
   5. Remove data “ผลผลิต(ตัน)” from Production field
   6. Change field Province data type from string to State/Province
   7. Change field Production data type from string to Integer
   8. Create field name sub\_type that contain string data which all row are Dairy
3. Cleaning coffee file
   1. Replace data “กรุงเทพ” in field จังหวัด by “กรุงเทพมหานคร”
   2. Change field จังหวัด data type from string to State/Province data type
   3. Rename field “จำนวนโคนม (ตัว)” to Production
   4. Create field name sub\_type that contain string data which all row are Coffee
4. Union between sugar-factory file, dairy file, and coffee file.

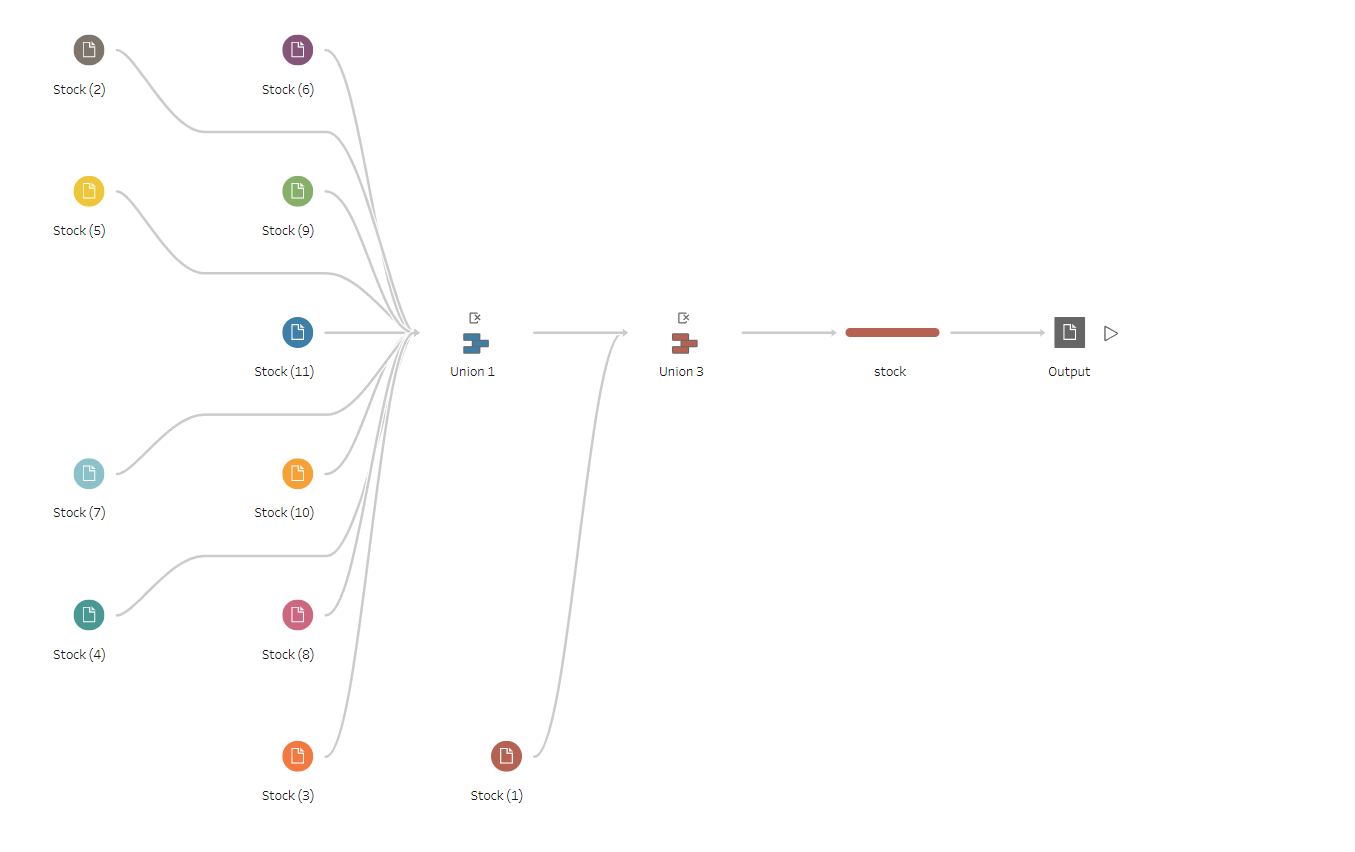
## Menu

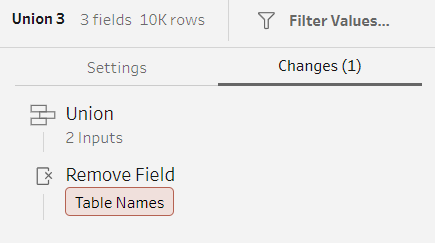
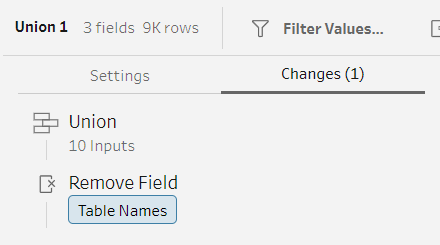


The ELT process for menu data:

1. Create id\_product field by using row number as the data in id\_product field

## Stock

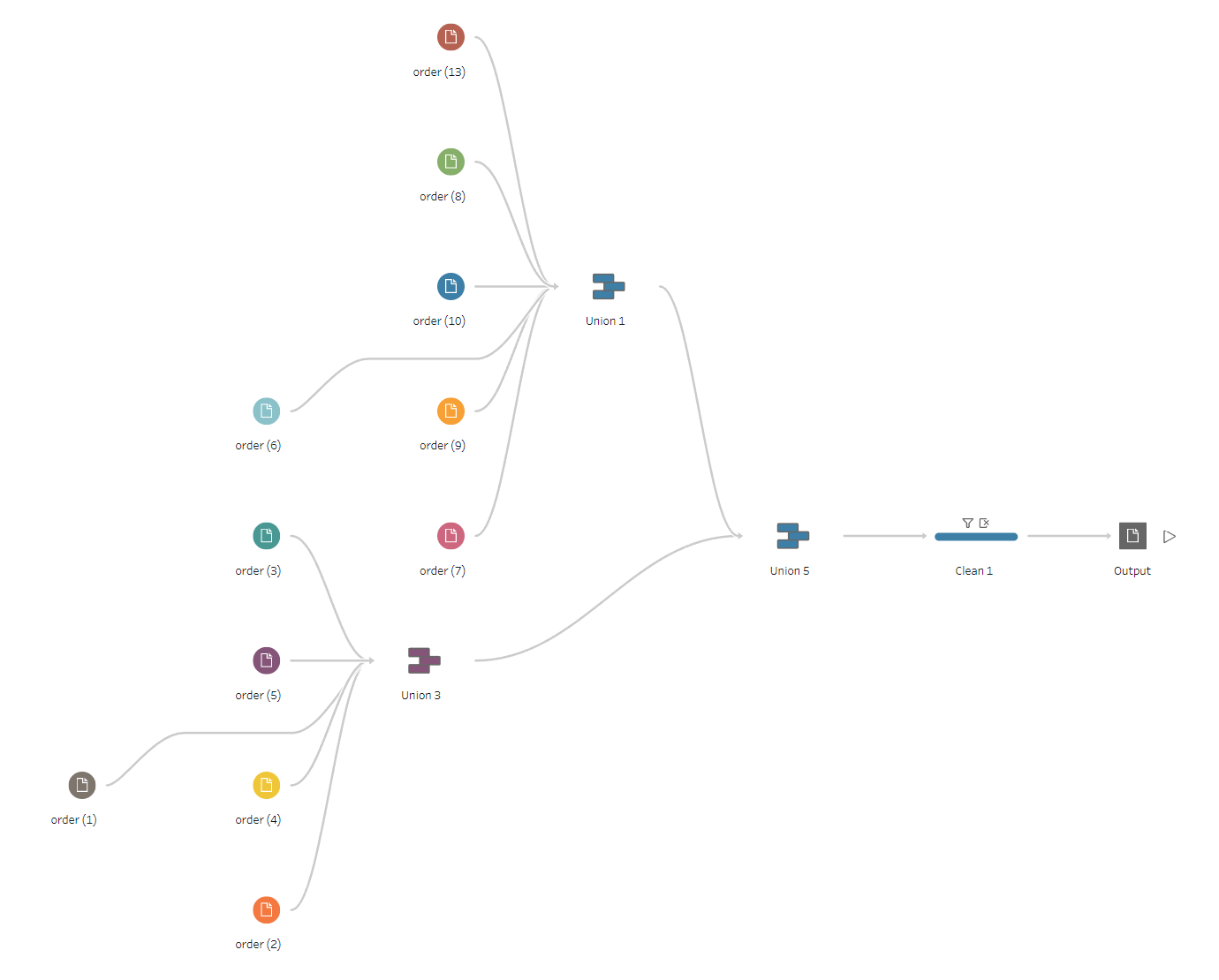
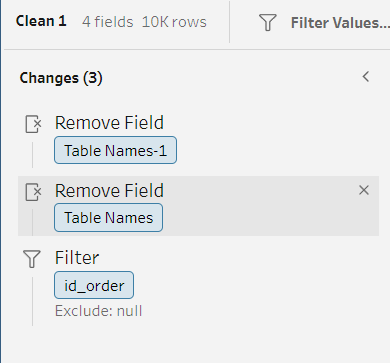
****

****

The ELT process for Stock data:

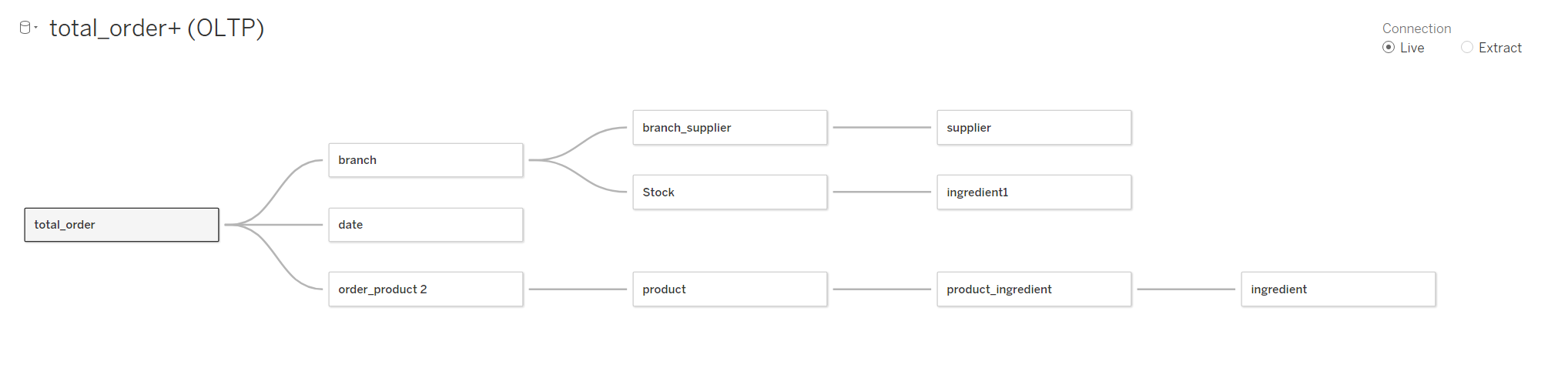
1. Union all stock file in to one file
2. Remove table name field

## Order

****

The ELT process for Order:

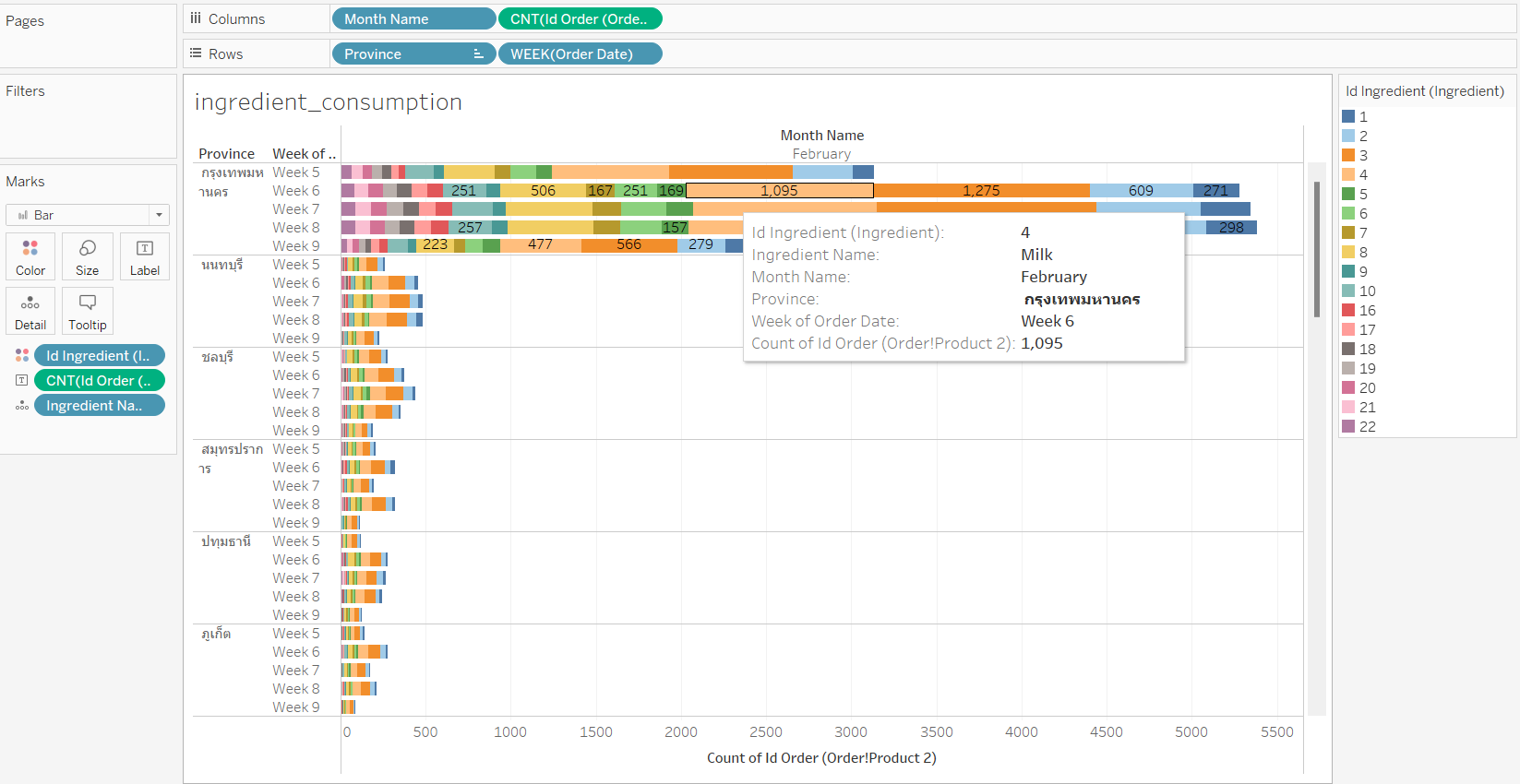
1. Union all Order file together
2. Remove Table Name-1 field after union all Order file
3. Remove Table Name field after union all Order file
4. Exclude data NULL from id\_order field after union all Order file



Integrate data from operational database to Tableau

# Analysis and Visualization reports

**Ingredient consumption**

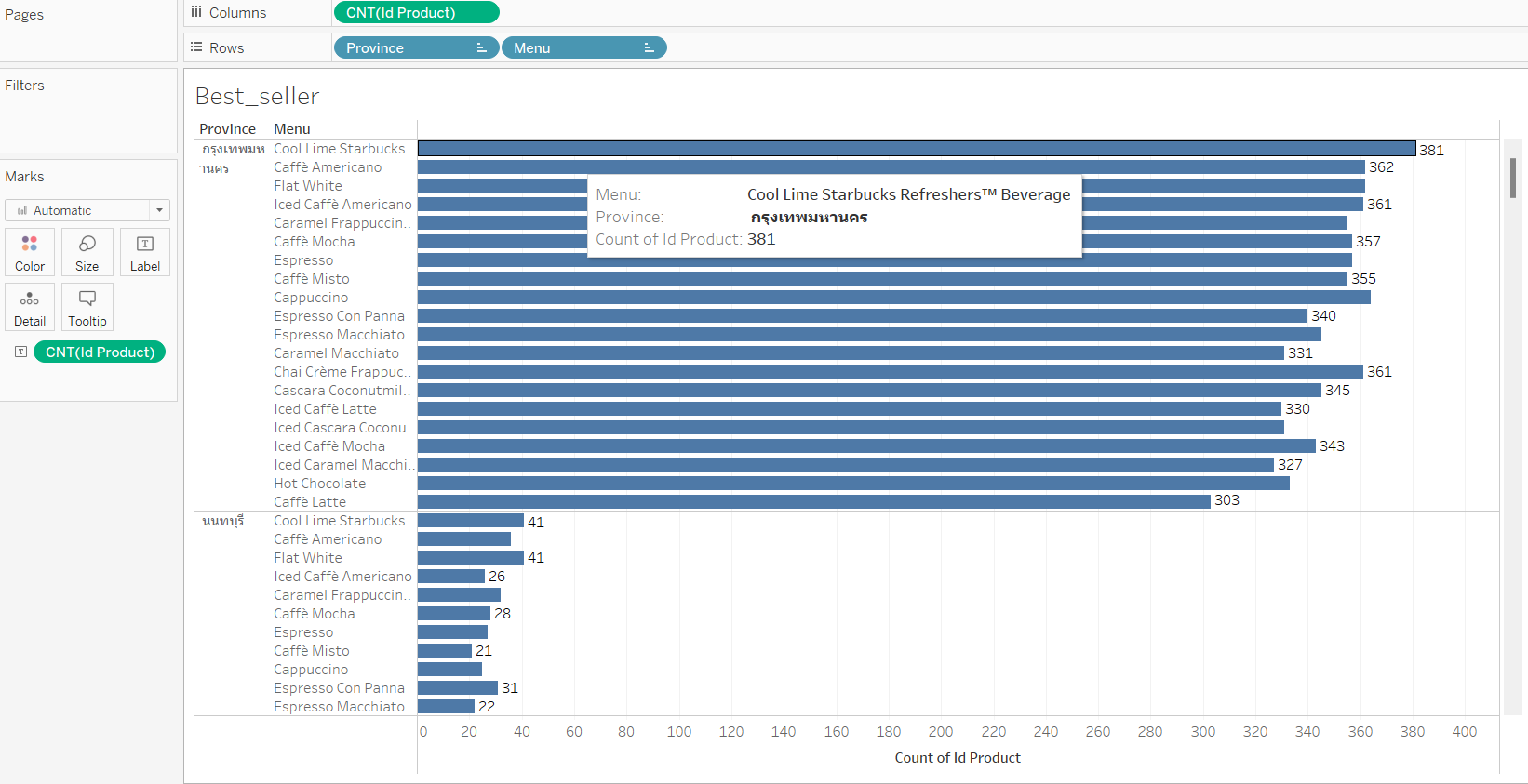
****

This is a bar chart that visualizes the overall Ingredient consumption with order count as a measure on the row and province and week as dimensions on the column. This shows how many ingredients were used in each week of each province.

The color masks the ingredient, which means each bar will be split into different color segments based on different ingredients. When hovering over any segment of a bar, the ingredient name will be shown. Each segment also has a number of ingredients used shown inside.

With this chart, Starbucks will know the information of weekly ingredients used in each province. This information will assist the decision making on inventory management and supply logistics.

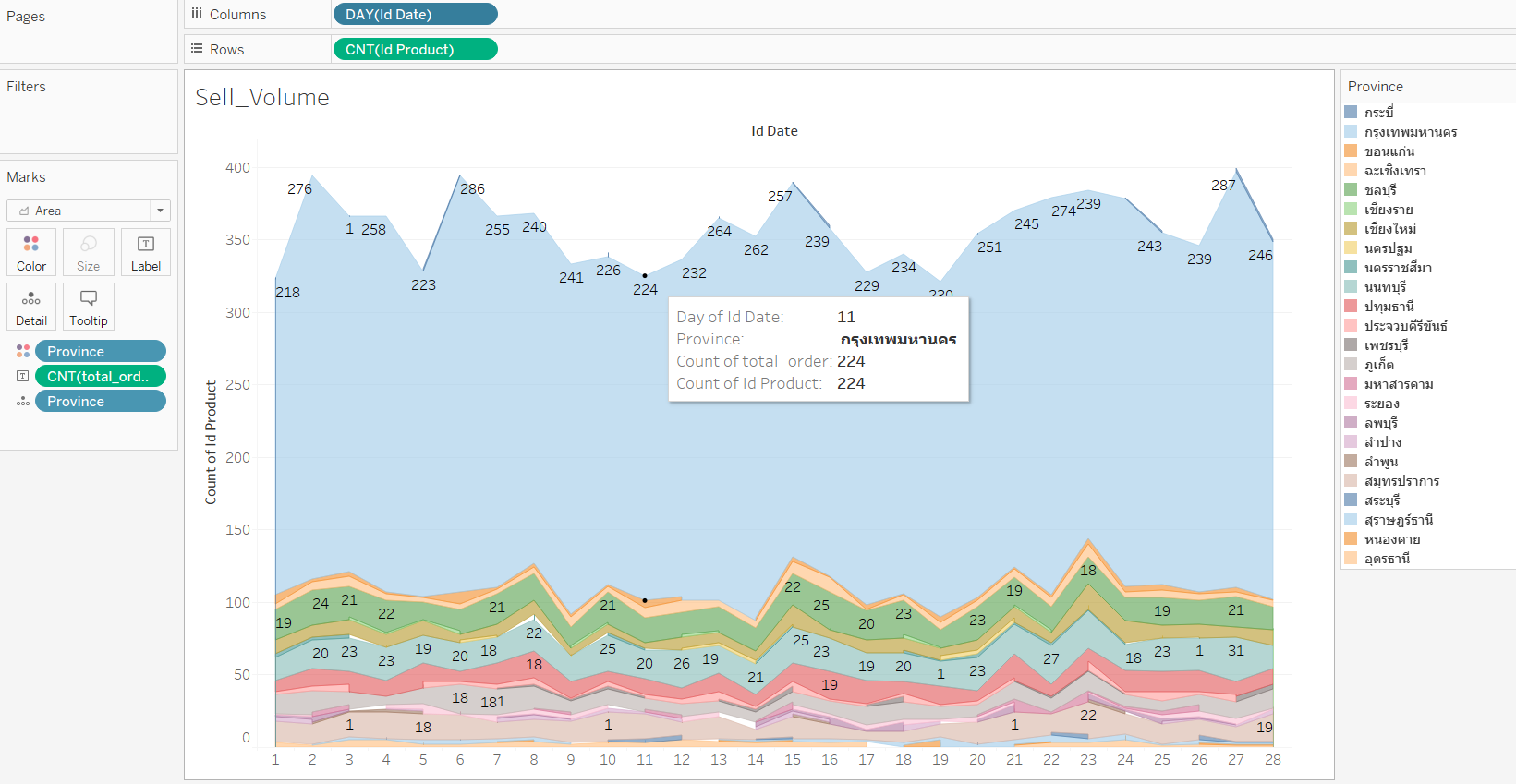
## Best seller products

****

This is bar chart that show the demand for each product in each province with product count as a measure on the columns and province and product on the row. This show how many product have been sold in each province.

With this chart, Starbucks will know the information of weekly product sold in each province. This information will assist the decision making on what type of product should Starbucks focus on.

## Sell volume

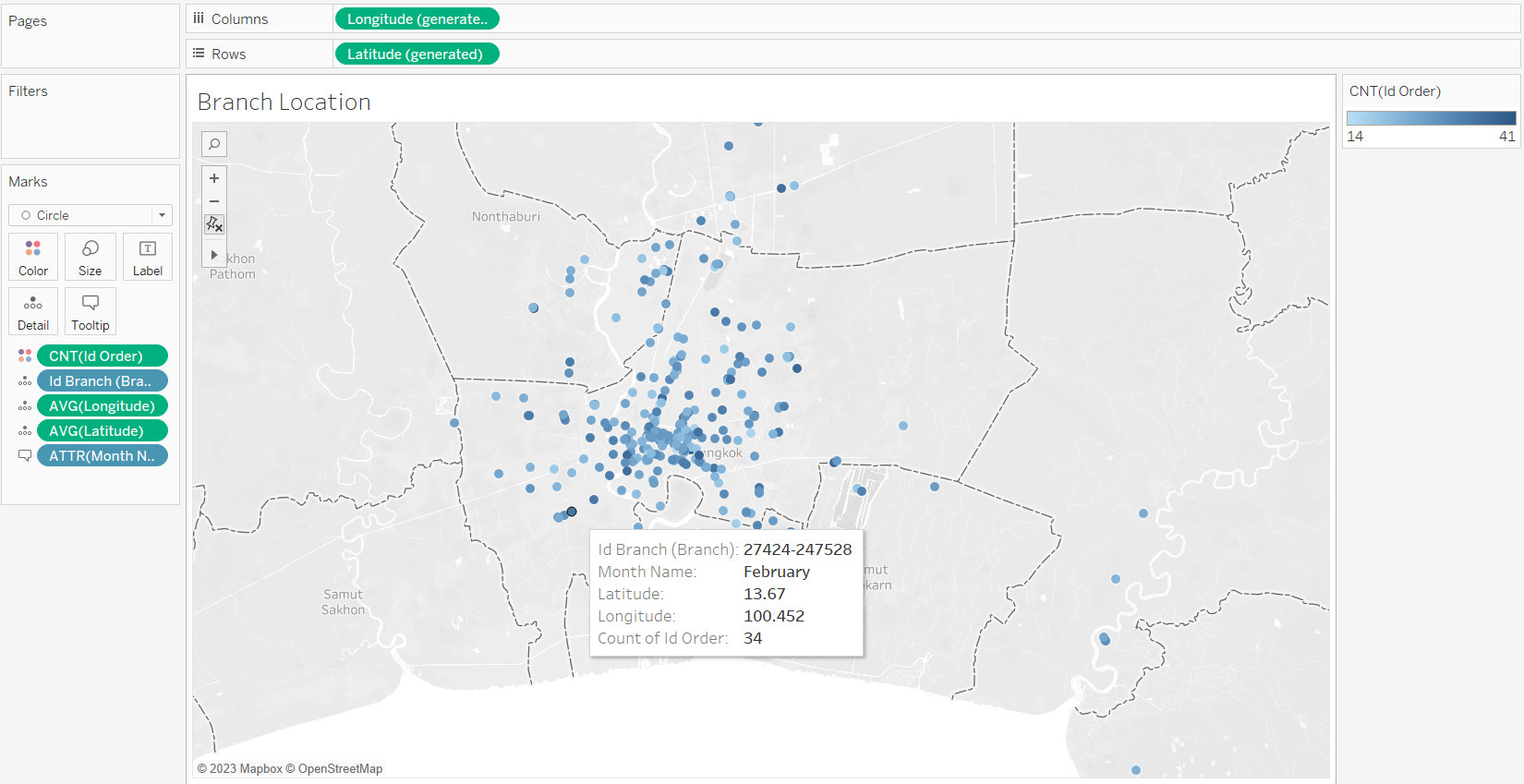
****

This chart visualizes the sell volume on each day in February with the number of products sold as a measure on the row and data on February on the column. This shows how many sales each day in February.

The color masks the province, which means the area will be split into different color segments based on the volume of each province. When hovering over any point of an area, the date, province, and sell volume will be shown.

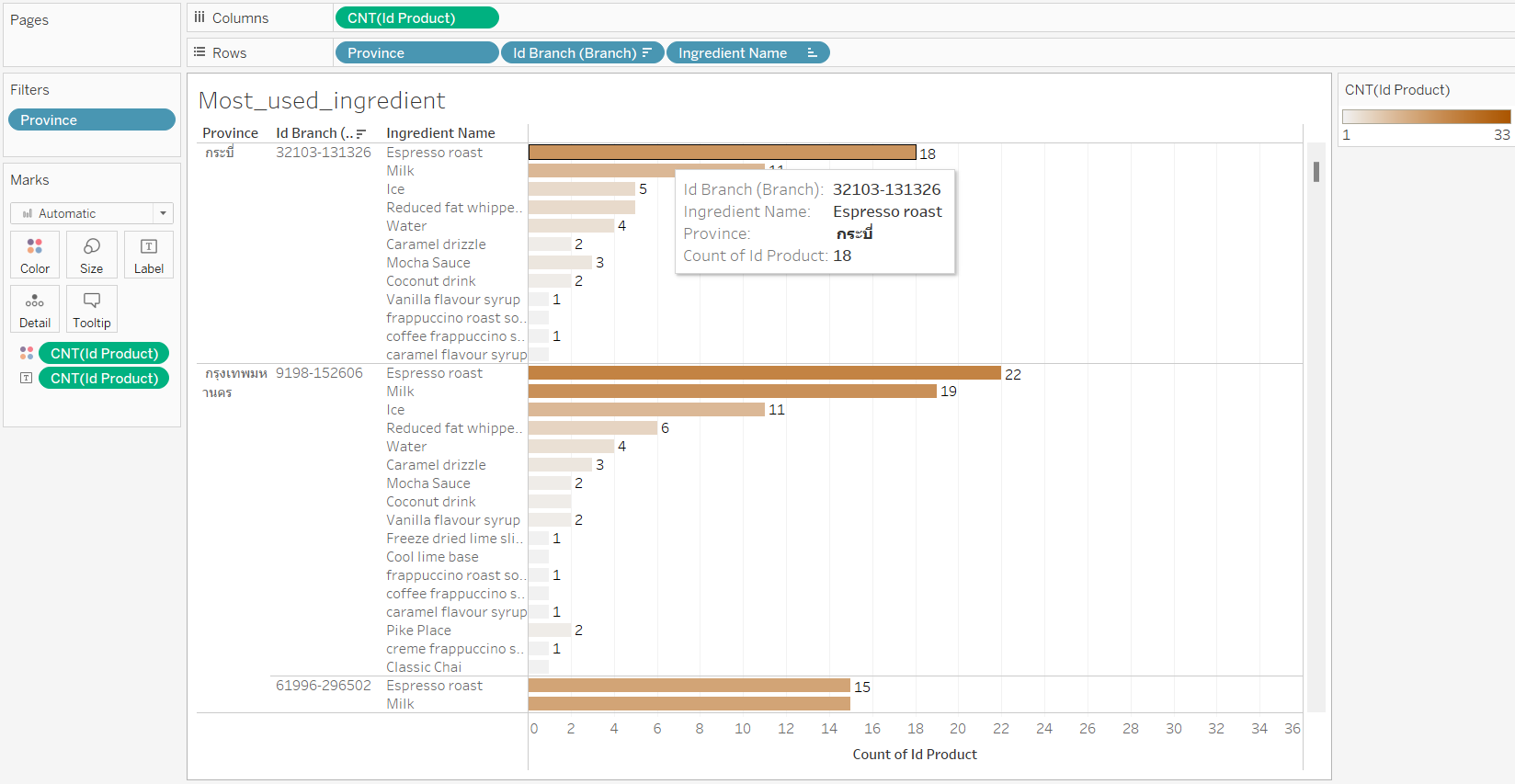
With this chart, Starbucks will know the information of sales volume for each province in February. This information will assist the decision making on how to distribute suppile to each branch.

## Branch Location

****

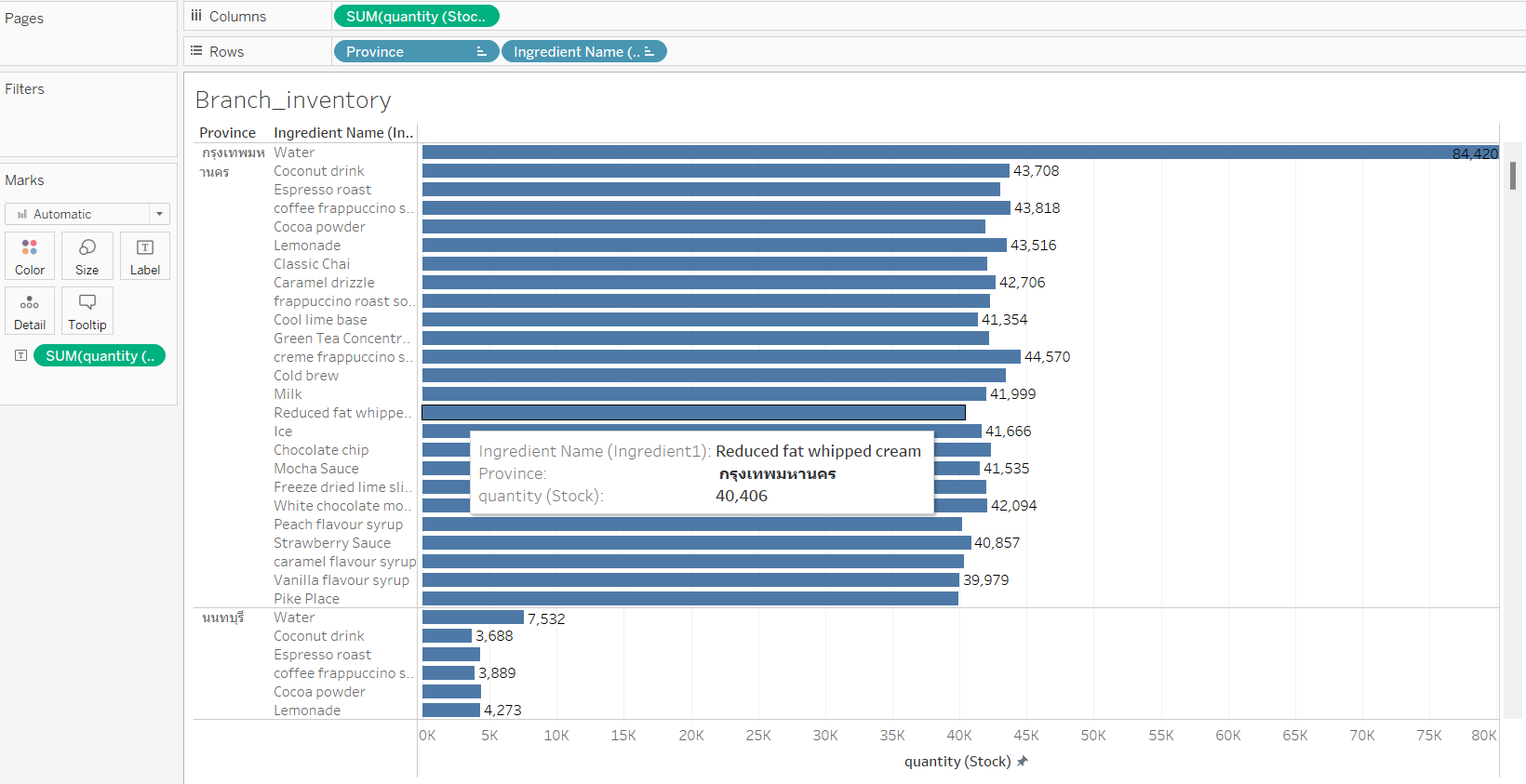
This map displays each branch location and number of orders for the company to look at the performance of each branch to support decision making on opening and closing branches.

## Most used ingredient

****

From this display we show how many ingredients were used and what is the most used ingredient of each branch for each branch to be able to keep up with their ingredient stock.

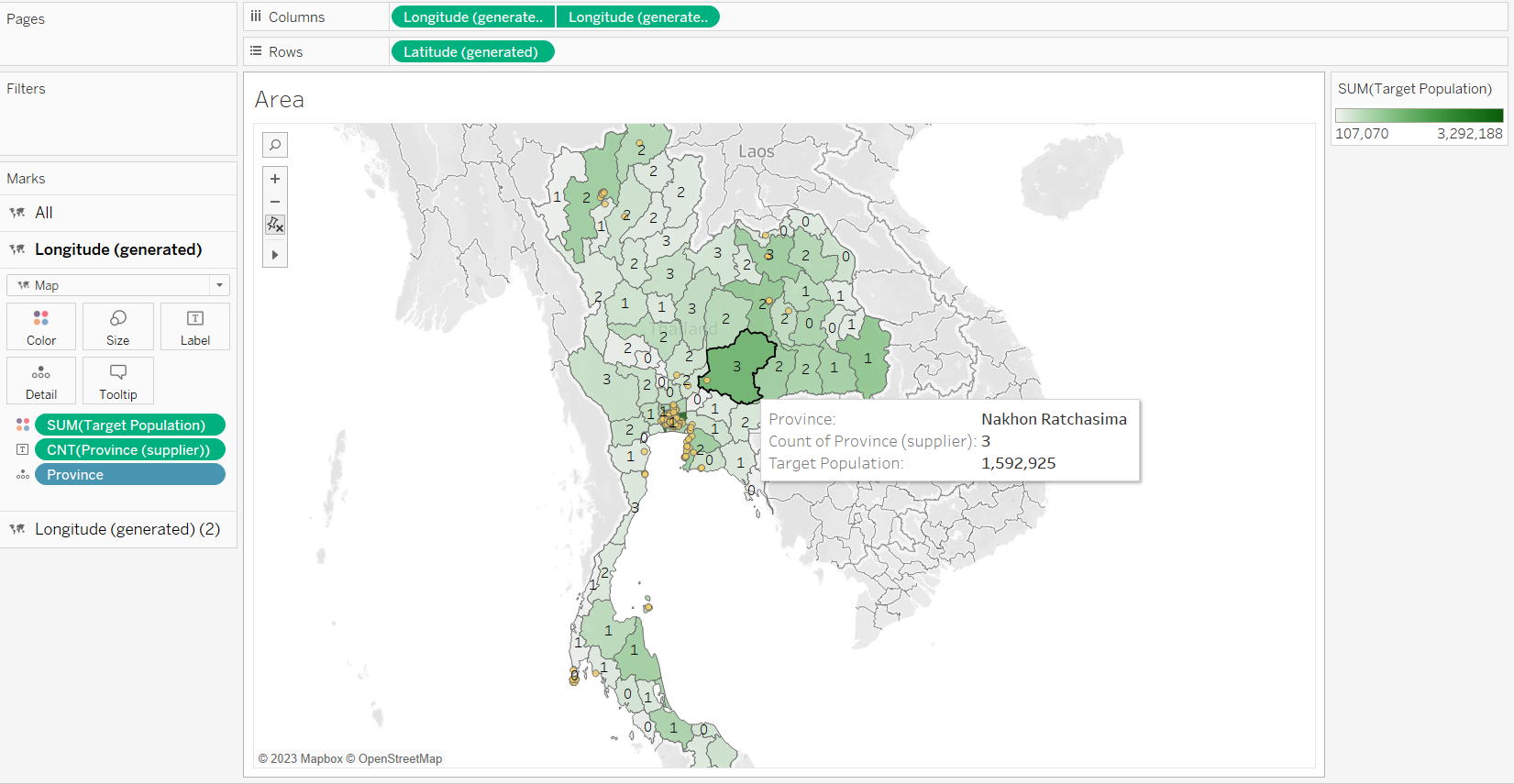
## Branch Inventory

****

This chart visualizes the overall ingredient quantity in stock of each province in which province and ingredient name as a row and ingredient quantity in stock as a measurement in column. This shows the overall quantity over all ingredients in stock of each province. This chart is ordered by ingredient quantity in stock in descending order.

With this chart, Starbucks will know the information of ingredient quantity in stock of each province. This information will assist the decision making on inventory management.

## Area suitable for new branches

****

This map is used for the company to look out for a suitable place to build a branch close to both the supplier and the target customer. All existing branches are shown as a yellow circle and the number of target customers is displayed as the intensity of the green color and the number of suppliers are in the center of each province in the map.

**Discussion and conclusion**

Discussion

The outcomes of this research have provided insight into how to implement data warehouses to solve Starbucks’ issues of supply chain management . However, the results should be interpreted with caution due to the limitations to the data. This chapter provides a reflection on the research process. The limitations and potential consequences of the design are discussed, as well as the implications for the interpretation of the results. The chapter ends with several recommendations for future research.

conclusion

In conclusion, to solve Starbucks Thailand challenges with supply chain management, resulting in ingredient shortages and hindering the company's growth. By analyzing sales and revenue data, inventory levels, customer behavior and preferences, supplier performance, and store performance, Starbucks can optimize its supply chain, improve inventory management, personalize its marketing efforts, enhance the customer experience, and drive better outcomes, Starbucks can leverage data warehouse and business intelligence tools to gain insights into its operations and make data-driven decisions.

**References**

* A. Soni, “Role of supply chain function in Starbucks coffee by Anesh Soni 22/08/2020,” *LinkedIn*. [Online]. Available: https://www.linkedin.com/pulse/role-supply-chain-function-starbucks-coffee-anesh-soni-anesh-soni. [Accessed: 25-Feb-2023].
* “demography population and Housing Branch,” *Oracle Business Intelligence*. [Online]. Available: http://statbbi.nso.go.th/staticreport/page/sector/en/01.aspx. [Accessed: 26-Feb-2023].
* Kongvut, “Thai-province-data/thai\_provinces.CSV at master · Kongvut/Thai-Province-Data,” *GitHub*. [Online]. Available: https://github.com/kongvut/thai-province-data/blob/master/csv/thai\_provinces.csv. [Accessed: 26-Feb-2023].
* kukuroo3, “Starbucks locations worldwide 2021 version,” *Kaggle*, 03-Jan-2022. [Online]. Available: https://www.kaggle.com/datasets/kukuroo3/starbucks-locations-worldwide-2021-version. [Accessed: 25-Feb-2023].
* N. M. Wire, “Starbucks ingredient shortage means 'various items' may be unavailable to make popular drinks,” *WANE 15*, 09-Jun-2021. [Online]. Available: https://www.wane.com/news/starbucks-ingredient-shortage-means-various-items-may-be-unavailable-to-make-popular-drinks/. [Accessed: 24-Feb-2023].
* “Nutrition & Allergen Guide - Starbucks.co.uk.” [Online]. Available: https://www.starbucks.co.uk/sites/starbucks-uk/files/2023-01/WIN23\_UK\_AllergenBook\_PROMO\_v02%20%28002%29.pdf. [Accessed: 24-Mar-2023].
* “Out of key ingredients, Starbucks' store shortages reflect national supply chain disruptions,” *Thomasnet® - Product Sourcing and Supplier Discovery Platform - Find North American Manufacturers, Suppliers and Industrial Companies*. [Online]. Available: https://www.thomasnet.com/insights/out-of-key-ingredients-starbucks-store-shortages-reflect-national-supply-chain-disruptions/. [Accessed: 24-Feb-2023].
* “Starbucks facing supply chain shortages,” *QSR magazine*. [Online]. Available: https://www.qsrmagazine.com/fast-food/starbucks-facing-supply-chain-shortages. [Accessed: 24-Feb-2023].
* “ข้อมูลเกษตรกรผู้เลี้ยงโคนม - open government data of Thailand.” [Online]. Available: https://data.go.th/dataset/item\_2a1f6f83-4721-4f3d-894e-f886a6d6f157. [Accessed: 25-Feb-2023].
* “ผลผลิตกาแฟแยกตามจังหวัด ปี 2562 - open government data of Thailand,” *imgHeader*. [Online]. Available: https://data.go.th/dataset/coffee62. [Accessed: 25-Feb-2023].
* “สถานที่ตั้งโรงงานน้ำตาลในประเทศไทย - Open Government Data of Thailand,” *imgHeader*. [Online]. Available: https://data.go.th/dataset/sugar-factory. [Accessed: 25-Feb-2023].

Link to a video presentation: <https://drive.google.com/file/d/1TGkRZDXc0N-A65noDbW8fo12w3Ameh3p/view?usp=share_link>