

BIG DATA AND BUSINESS INTELLIGENCE

CIS4008-N-BF1-2020

YU SHING, LUI (A0201746)

TEESSIDE UNIVERSITY | 2020/21

Table of Contents

SECTION 1 – BUSINESS INTELLIGENCE DESIGN.....	3
1 Data Source Description and Business Questions.....	3
1.1 Dataset Information	3
1.2 Dataset Description	3
1.3 Business Questions	4
2 Data Pre-Processing and Data Cleansing.....	6
2.1 Renaming Columns	8
2.2 Changing Data Types	9
2.3 Removing Columns.....	11
2.4 Changing Format	13
3 Data Modelling – Star Schema Facts and Dimensions.....	17
3.1 Splitting tables.....	17
3.2 Remove Duplicate items	20
3.3 M Language and Create New Column	24
3.4 Data Model	28
3.5 Schema and Relationships	30
SECTION 2 – BUSINESS INTELLIGENCE SOLUTION	32
1 Executive Summary.....	33
2 Key Findings.....	33
3 Recommendations	34
4 Data Description	34
5 Data Model.....	34
6 Introduction.....	35
7 Finding based on analysis and evaluation – 1.....	35
7.1 Matrix and Key Finding From The Visual	35
7.1.1 DAX Expressions – 1 and New Calculated Measure Add Into Model	36
7.2 Key Performance Indicator (KPI)	39
7.2.1 Dax Expression - 2	40
8 Finding based on analysis and evaluation - 2.....	42
8.1 Infographic	42
8.2 Area Chart and Key Finding From The Visual	43
8.3 Treemap	44

8.4	Animated Chart	45
8.5	Map.....	46
8.6	Ask a Question Tool.....	48
9	<i>Data Analytics (Forecast, Analysing trends)</i>	<i>49</i>
9.1	Forecast.....	49
9.2	Artificial Intelligence And Key Finding From The Visual	51
10	<i>Structure of the Dashboard</i>	<i>52</i>
11	<i>Conclusion.....</i>	<i>54</i>
12	<i>Recommendations.....</i>	<i>54</i>
13	<i>Personal Conclusions</i>	<i>54</i>
14	<i>Reference</i>	<i>55</i>

SECTION 1 – BUSINESS INTELLIGENCE DESIGN

1 Data Source Description and Business Questions

1.1 Dataset Information

The dataset is about coffee house chain in United States of America. The dataset focus of this project as it contains important information with over 4,000 rows and 20 columns. The columns of the dataset is including Area Code, Date, Market, Market Size, Product, Product Line, Product Type, State, Type, Budget Cost of Goods Sold, Budget Margin, Budget Profit, Budget Sales, Cost of Goods Sold, Inventory, Margin, Marketing, Profit, Sales, and Total Expenses.

Area Code	Date	Market	Market Size	Product	Product Line	Product Type	State
563	40909	Central	Small Market	Caffe Mocha	Beans	Espresso	Iowa
563	40909	Central	Small Market	Decaf Espresso	Beans	Espresso	Iowa
978	40940	East	Major Market	Green Tea	Leaves	Tea	Massachusetts
603	40940	East	Small Market	Green Tea	Leaves	Tea	New Hampshire
702	40940	West	Small Market	Caffe Latte	Beans	Espresso	Nevada

Type	Budget COGS	Budget Margin	Budget Profit	Budget Sales	COGS	Inventory	Margin
Regular	220	330	250	550	239	1262	325
Regular	270	400	280	670	284	2108	428
Regular	110	160	120	270	104	871	152
Regular	110	160	120	270	104	821	152
Regular	110	-50	-180	60	121	1673	-52

Marketing	Profit	Sales	Total Expenses
66	235	564	90
99	276	712	152
29	112	256	40
29	111	256	41
109	-195	69	143

1.2 Dataset Description

The dataset is for the coffee chain that take place in the different states of United States of America and each type of product's statistics in 2012-2013. It is taken from the webpage, www.dataplusscience.com, and the total weight of the data is 403kb in excel file.

The dataset is downloaded as below linkage.
<http://www.dataplusscience.com/files/Coffee.xlsx>

1.3 Business Questions

Why did you select this specific dataset?

This specific dataset is common in the commercial industry to analysis the company's sales and revenue in different sectors. We believe that it can train us to use Power BI to solve lots of problems once we face the same issue in our career.

Will this dataset help you in developing specific business skills?

This dataset can develop specific business skills, which is like Data Analysis, Communication, Problem Solving.

Data Analysis – The dataset will translate into actionable information, then the management team can make decision that will enhance profitability as much as possible. It will involve large amount of data and it can help to practice analytical skills.

Communication – Practicing how to describe this data, explain the result from data analysis, and then provide possible solutions and forecast. It is one of the most critical soft skill in commercial industry.

Problem Solving – Analysing this data is just part of the data analyst's skill, who also offer solutions to management team based on the dataset. Therefore, it is necessary come up with clear suggestions and high accuracy forecast to assist the company make better decisions.

What questions do you seek to answer with your BI project?

- What is the overall on an average sale of the products?
- What is the sales variance of each products?
- What are the total expenses versus profit for each product?
- What is the market size for each state with their sales?
- How is the key performance indicator of sales, profit and cost of goods sold?
- How is the sales forecast in 2014?

Which specific features are you going to focus on?

Interactive Power BI Desktop – It allows to build reports by accessing the data rapidly, even people do not have advanced skills to create a report with Power BI. This tool is free to download, and allow everyone to create report without processing any technical knowledge.

Customized Visualization – Power BI has a default standard, and it can customize some functions to process some complex data depends on everyone required. It can access the custom library of visualization easily and design a visualization that serves the people need.

Visibility – Making the data meaningful is the main purpose. Power BI can help to deliver actionable insights and simplify the task of data collaboration, analysis, and makes it to next high-standard level.

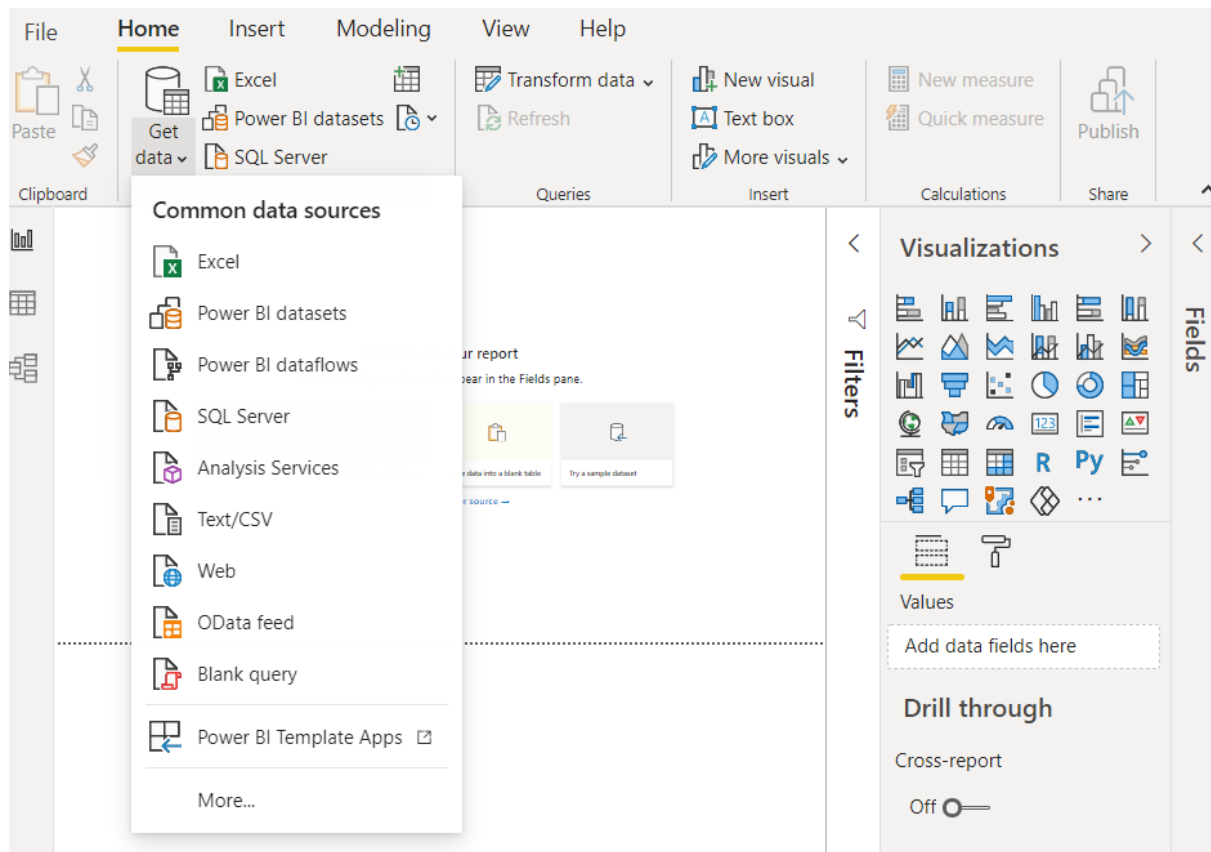
Does this dataset address the Big Data problem?

This dataset can address the big data problem. It requires to clean, edit and organise this data, then manage this dataset in Power BI and turn data into visualisation for the company. In addition, use it to provide valuable insights to management team and they can make decisions.

2 Data Pre-Processing and Data Cleansing

Power BI Desktop is able to input with different data source, which is like Excel, Text/CSV, Web, etc.

Loading data to the Power BI



Selecting the data file 'Sheet 1' and click the load button to load the data.

Navigator

Display Options ▾

- ☑ Coffee.xlsx [3]
 - ☒ Sheet1
 - ☐ Sheet2
 - ☐ Sheet3

Sheet1

Area Code	Date	Market	Market Size	Product	Product
719	40909	Central	Major Market	Amaretto	Bear
970	40909	Central	Major Market	Colombian	Bear
970	40909	Central	Major Market	Decaf Irish Cream	Bear
303	40909	Central	Major Market	Green Tea	Leav
303	40909	Central	Major Market	Caffe Mocha	Bear
720	40909	Central	Major Market	Decaf Espresso	Bear
970	40909	Central	Major Market	Chamomile	Leav
719	40909	Central	Major Market	Lemon	Leav
970	40909	Central	Major Market	Mint	Leav
719	40909	Central	Major Market	Darjeeling	Leav
303	40909	Central	Major Market	Earl Grey	Leav
217	40909	Central	Major Market	Colombian	Bear
309	40909	Central	Major Market	Decaf Irish Cream	Bear
309	40909	Central	Major Market	Caffe Mocha	Bear
630	40909	Central	Major Market	Decaf Espresso	Bear
312	40909	Central	Major Market	Chamomile	Leav
630	40909	Central	Major Market	Lemon	Leav
773	40909	Central	Major Market	Mint	Leav
217	40909	Central	Major Market	Darjeeling	Leav
708	40909	Central	Major Market	Earl Grey	Leav
319	40909	Central	Small Market	Amaretto	Bear
641	40909	Central	Small Market	Colombian	Bear
712	40909	Central	Small Market	Decaf Irish Cream	Bear

Load Transform Data Cancel

2.1 Renaming Columns

Once the data is loaded, double-click specify columns to rename in the “Fields” section. Now change the name of columns “Budget COGS” to “Budget Cost of Goods Sold”. Then use the same steps to change the two columns “COGS” and “Type” to “Cost of Goods Sold” and “Caffeine Type”.

The screenshot shows the Tableau interface with the 'Structure' tab selected. The columns are: Product, Product Line, Product Type, State, Type, and Budget COGS. The 'Fields' pane on the right shows a search bar and a list of fields including Area Code, Budget COGS, Budget Margin, Budget Profit, Budget Sales, COGS, Date, Inventory, Margin, and Market.

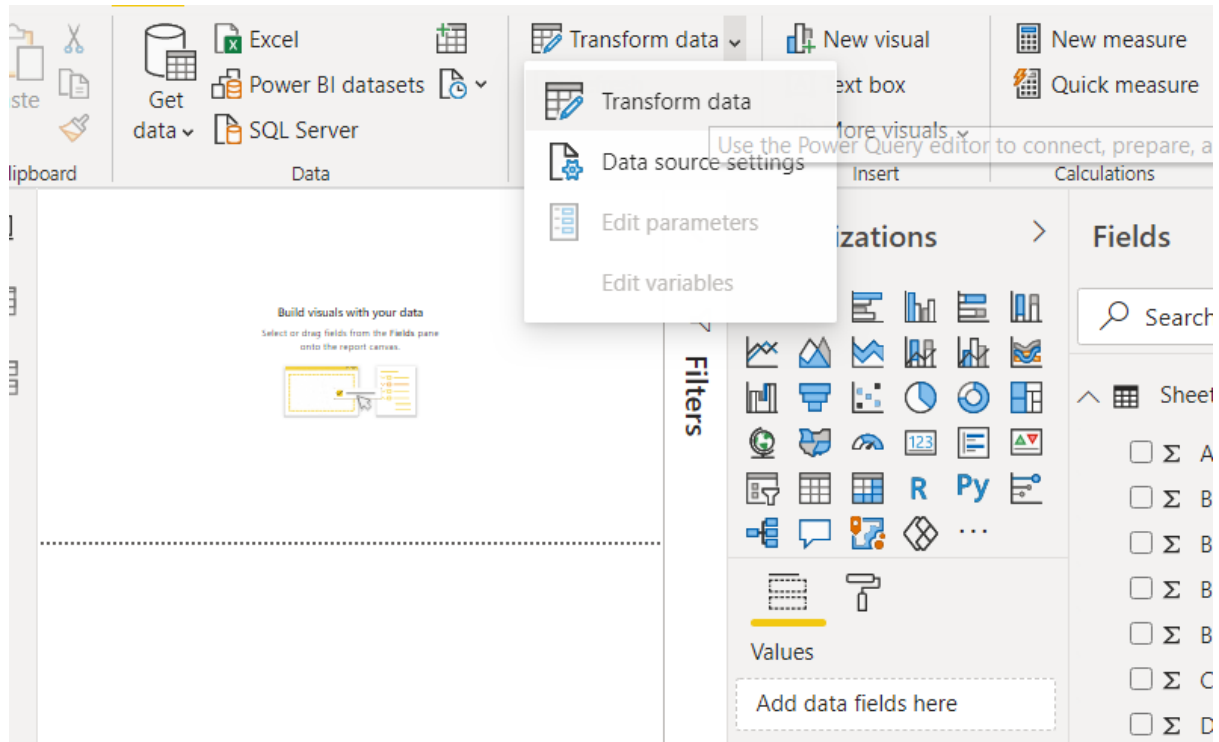
Product	Product Line	Product Type	State	Type	Budget COGS
Cafe Mocha	Beans	Espresso	Nevada	Regular	20
Earl Grey	Leaves	Tea	Washington	Regular	20
Caffe Latte	Beans	Espresso	Washington	Regular	20
Caffe Mocha	Beans	Espresso	Iowa	Regular	20
Green Tea	Leaves	Tea	Connecticut	Regular	20
Darjeeling	Leaves	Tea	Massachusetts	Regular	30
Green Tea	Leaves	Tea	New Hampshire	Regular	0
Caffe Latte	Beans	Espresso	Utah	Regular	30
Caffe Latte	Beans	Espresso	Nevada	Regular	0
Earl Grey	Leaves	Tea	Washington	Regular	30
Caffe Latte	Beans	Espresso	Washington	Regular	30
Colombian	Beans	Coffee	Iowa	Regular	20
Darjeeling	Leaves	Tea	Missouri	Regular	20
Lemon	Leaves	Herbal Tea	New Mexico	Decaf	10
Chamomile	Leaves	Herbal Tea	Ohio	Decaf	30
Decaf Espresso	Beans	Espresso	Wisconsin	Decaf	20

The screenshot shows the Tableau interface after renaming the columns. The columns are: Product Line, Product Type, State, Type, and Budget Cost of Goods Sold. The 'Fields' pane on the right shows a search bar and a list of fields including Area Code, Budget Cost of Goods Sold, Budget Margin, Budget Profit, Budget Sales, Cost of Goods Sold, Date, Inventory, Margin, and Market.

Product Line	Product Type	State	Type	Budget Cost of Goods Sold
Beans	Espresso	Iowa	Regular	
Beans	Espresso	Iowa	Decaf	
Leaves	Tea	Massachusetts	Regular	
Leaves	Tea	New Hampshire	Regular	
Beans	Espresso	Nevada	Regular	
Leaves	Tea	Connecticut	Regular	
Leaves	Tea	New Hampshire	Regular	
Beans	Espresso	Nevada	Regular	
Beans	Espresso	Iowa	Regular	
Leaves	Tea	Connecticut	Regular	
Leaves	Tea	New Hampshire	Regular	

2.2 Changing Data Types

After changing the name of columns, the next step is change the data type in “Transform Data”.



Click the logo which is located in the left of the name's column and change to "Date".

	123 Area Code	123 Date	ABC Market	ABC Market Size
1	719	1.2 Decimal Number	tral	Major Market
2	970	\$ Fixed decimal number	tral	Major Market
3	970	123 Whole Number	tral	Major Market
4	303	% Percentage	tral	Major Market
5	303	Date/Time	tral	Major Market
6	720	Date	tral	Major Market
7	970	Time	tral	Major Market
8	719	Date/Time/Timezone	tral	Major Market
9	970	Duration	tral	Major Market
10	719	ABC Text	tral	Major Market
11	303	True/False	tral	Major Market
12	217	Binary	tral	Major Market
13	309	Using Locale...	tral	Major Market
14	309		tral	Major Market
15	630		40909 Central	Major Market

	123 Area Code	Date	ABC Market	ABC Market Size	A
1	719	01/01/2012	Central	Major Market	A
2	970	01/01/2012	Central	Major Market	C
3	970	01/01/2012	Central	Major Market	D
4	303	01/01/2012	Central	Major Market	G
5	303	01/01/2012	Central	Major Market	C
6	720	01/01/2012	Central	Major Market	D
7	970	01/01/2012	Central	Major Market	C
8	719	01/01/2012	Central	Major Market	Li
9	970	01/01/2012	Central	Major Market	M
10	719	01/01/2012	Central	Major Market	D
11	303	01/01/2012	Central	Major Market	E
12	217	01/01/2012	Central	Major Market	C
13	309	01/01/2012	Central	Major Market	D
14	309	01/01/2012	Central	Major Market	C

2.3 Removing Columns

Also, the value of “Marketing” column is not necessary and we should delete it. Right-click of the column and select “Remove”. After that, click “Close & Supply” to save the changes.

	1 ² ₃ Margin	1 ² ₃ Marketing	1 ² ₃ Profit	1 ² ₃ Sales	1 ² ₃ Total Expenses
1	777	130			219
2	623	107			190
3	821	139			234
4	623	56			100
5	456	80			134
6	558	108			180
7	1091	171			341
8	435	87			150
9	336	80			140
10	338	72			130
11	965	76			140
12	862	201			345
13	608	139			234
14	1310	312			546
15	1459	228			456
16	777	130			219
17	623	107			190
18	821	139			234
19	456	80			134
20	558	108			180
21	821	27			45
22	965	34	8	5	62
23	623	31	7	12	54
24	777	26	4	11	43
25	777	26	4	10	43

1 ² Margin	1 ² Profit	1 ² Sales	1 ² Total Expenses
777	130	94	219
623	107	68	190
821	139	101	234
623	56	30	100
456	80	54	134
558	108	53	180
1091	171	99	341
435	87	0	150
336	80	33	140
338	72	17	130
965	76	36	140
862	201	111	345
608	139	87	234
1310	312	203	546
1459	228	140	456
777	130	95	219
623	107	68	190
821	139	101	234
456	80	53	134
558	108	54	180
821	27	11	45
965	34	5	62
623	31	12	54
777	26	11	43
777	26	10	43
1310	312	202	546

Query Settings

PROPERTIES

Name
coffee_data_all

[All Properties](#)

APPLIED STEPS

- Source
- Navigation
- Promoted Headers
- Changed Type
- Renamed Columns
- Changed Type1
- Renamed Columns1
- Removed Columns

Close & Apply | New Source | Recent Sources | Enter Data | Data source settings | Manage Parameters | Refresh Preview

Close & Apply | Apply | Close | Sheet1

1 ² Cost of Goods Sold
226
227
228
229
230
231

2.4 Changing Format

After all the adjustments, move to select “Data” section on the left and click to highlight “Date” column. Select “Formatting” and revise the format to “(mmmm yyyy)”.

The screenshot shows the Microsoft Excel interface. The 'Date' column is selected in the 'Structure' pane on the left. The 'Format' dropdown menu is open, showing a list of date formats. The 'March 2001 (mmmm yyyy)' format is highlighted. The background shows a data table with columns for Date, Market, and Product.

Date	Market	Product
01 January 2012	Central	Small Market
01 January 2012	Central	Small Market
01 February 2012	East	Major Market
01 February 2012	East	Small Market
01 February 2012	West	Small Market
01 March 2012	East	Small Market
01 March 2012	East	Small Market
01 March 2012	West	Small Market
01 April 2012	Central	Small Market
01 April 2012	East	Small Market
01 April 2012	East	Small Market
01 April 2012	West	Small Market
01 May 2012	East	Small Market
01 May 2012	East	Major Market
01 May 2012	East	Small Market
01 May 2012	West	Small Market
01 June 2012	Central	Small Market
01 June 2012	Central	Small Market
01 June 2012	East	Major Market
01 June 2012	East	Small Market
01 June 2012	South	Small Market

	Date		Formatting		Properties		Sort by column		Data groups		Manage relationships		New column
	Date												
Structure													
													Fields
	Date	Market	Market Size	Product	Product Line	Product Type	Search						
	Sheet1												
	Budget Cost of G... Budget Margin Budget Profit Budget Sales Cost of Goods Sold Date Inventory Margin Market Market Size Marketing Product Product Line												
	January 2012	Central	Small Market	Caffe Mocha	Beans	Espresso							
	January 2012	Central	Small Market	Decaf Espresso	Beans	Espresso							
	February 2012	East	Major Market	Green Tea	Leaves	Tea							
	February 2012	East	Small Market	Green Tea	Leaves	Tea							
	February 2012	West	Small Market	Caffe Latte	Beans	Espresso							
	March 2012	East	Small Market	Green Tea	Leaves	Tea							
	March 2012	East	Small Market	Green Tea	Leaves	Tea							
	March 2012	West	Small Market	Caffe Latte	Beans	Espresso							
	April 2012	Central	Small Market	Caffe Mocha	Beans	Espresso							
	April 2012	East	Small Market	Green Tea	Leaves	Tea							
	April 2012	East	Small Market	Green Tea	Leaves	Tea							
	April 2012	West	Small Market	Caffe Latte	Beans	Espresso							
	May 2012	East	Small Market	Green Tea	Leaves	Tea							
	May 2012	East	Major Market	Darjeeling	Leaves	Tea							
	May 2012	East	Small Market	Green Tea	Leaves	Tea							
	May 2012	West	Small Market	Caffe Latte	Beans	Espresso							
	June 2012	Central	Small Market	Caffe Mocha	Beans	Espresso							
	June 2012	Central	Small Market	Amaretto	Beans	Coffee							

Changing the format of “Budget Cost of Goods Sold” column from “Whole Number” to “Currency” and add common currency symbols to “\$ English (United States)”. Also, change decimal places to “0”. Repeat the same step for the rest of columns of Budget Margin, Budget Profit, Budget Sales, Cost of Goods Sold, Inventory, Margin, Marketing, Profit, Sales, and Total Expenses.

Format dropdown menu options:

- General
- Currency
- Decimal number
- Whole number
- Percentage
- Scientific

Fields pane search results:

- Budget Cost of Go...
- Budget Margin
- Budget Profit
- Budget Sales
- Cost of Goods Sold
- Date
- Inventory
- Margin
- Market
- Market Size
- Marketing
- Product
- Product Line

Format dropdown menu options:

- Currency
- \$
- %
- Auto

Fields pane search results:

- Budget Cost of Go...
- Budget Margin
- Budget Profit
- Budget Sales
- Cost of Goods Sold
- Date
- Inventory
- Margin
- Market
- Market Size
- Marketing
- Product
- Product Line

Format Currency

Change the number of decimal places shown for this value.

Type	Budget Cost of Goods Sold	Profit	Budget Sales	Cost
Regular	\$10	40	30	50
Decaf	\$10	40	30	50
Regular	\$20	40	30	60
Regular	\$0	40	30	40
Regular	\$0	40	30	40
Regular	\$20	40	30	60
Regular	\$0	40	30	40
Regular	\$0	40	30	40
Regular	\$10	40	30	50
Regular	\$20	40	30	60
Regular	\$0	40	30	40
Regular	\$0	40	30	40
Regular	\$20	40	30	60
Regular	\$30	40	30	70
Regular	\$0	40	30	40
Regular	\$0	40	30	40
Regular	\$20	40	30	60

Fields

Search

Sheet1

- Budget Cost of Go...
- Budget Margin
- Budget Profit
- Budget Sales
- Cost of Goods Sold
- Date
- Inventory
- Margin
- Market
- Market Size
- Marketing
- Product

Budget Cost of Goods Sold	Budget Margin	Budget Profit	Budget Sales	Cost of Goods Sold	Inventory	Margin	Marketing	Profit	Sales	Total Expenses
\$10	\$40	\$30	\$50	\$17	\$777	\$26	\$4	\$11	\$43	\$15
\$10	\$40	\$30	\$50	\$17	\$777	\$26	\$4	\$10	\$43	\$16
\$20	\$40	\$30	\$60	\$22	\$452	\$34	\$6	\$17	\$56	\$17
\$0	\$40	\$30	\$40	\$0	\$774	\$43	\$0	\$32	\$43	\$11
\$0	\$40	\$30	\$40	\$0	\$774	\$43	\$0	\$31	\$43	\$12
\$20	\$40	\$30	\$60	\$25	\$769	\$37	\$7	\$18	\$62	\$19
\$0	\$40	\$30	\$40	\$0	\$731	\$43	\$0	\$32	\$43	\$11
\$0	\$40	\$30	\$40	\$0	\$731	\$43	\$0	\$31	\$43	\$12
\$10	\$40	\$30	\$50	\$17	\$776	\$26	\$4	\$10	\$43	\$16
\$20	\$40	\$30	\$60	\$25	\$775	\$37	\$7	\$18	\$62	\$19
\$0	\$40	\$30	\$40	\$0	\$688	\$43	\$0	\$32	\$43	\$11
\$0	\$40	\$30	\$40	\$0	\$688	\$43	\$0	\$32	\$43	\$11
\$20	\$40	\$30	\$60	\$23	\$787	\$34	\$6	\$17	\$57	\$17
\$30	\$40	\$30	\$70	\$29	\$835	\$42	\$8	\$22	\$71	\$20
\$0	\$40	\$30	\$40	\$0	\$645	\$43	\$0	\$32	\$43	\$11
\$0	\$40	\$30	\$40	\$0	\$645	\$43	\$0	\$31	\$43	\$12
\$20	\$40	\$30	\$60	\$21	\$798	\$32	\$5	\$15	\$53	\$17
\$20	\$40	\$30	\$60	\$23	\$843	\$35	\$6	\$18	\$58	\$17
\$20	\$40	\$30	\$60	\$24	\$856	\$36	\$6	\$19	\$60	\$17
\$20	\$40	\$30	\$60	\$23	\$843	\$35	\$6	\$17	\$58	\$18
\$0	\$40	\$30	\$40	\$0	\$602	\$43	\$0	\$32	\$43	\$11
\$30	\$40	\$30	\$70	\$24	\$856	\$36	\$6	\$19	\$60	\$17
\$0	\$40	\$30	\$40	\$0	\$602	\$43	\$0	\$31	\$43	\$12
\$20	\$40	\$30	\$60	\$25	\$854	\$36	\$7	\$17	\$61	\$19
\$20	\$40	\$30	\$60	\$23	\$868	\$34	\$6	\$16	\$57	\$18
\$20	\$40	\$30	\$60	\$25	\$854	\$36	\$7	\$17	\$61	\$19
\$0	\$40	\$30	\$40	\$0	\$559	\$43	\$0	\$31	\$43	\$12

Fields

Search

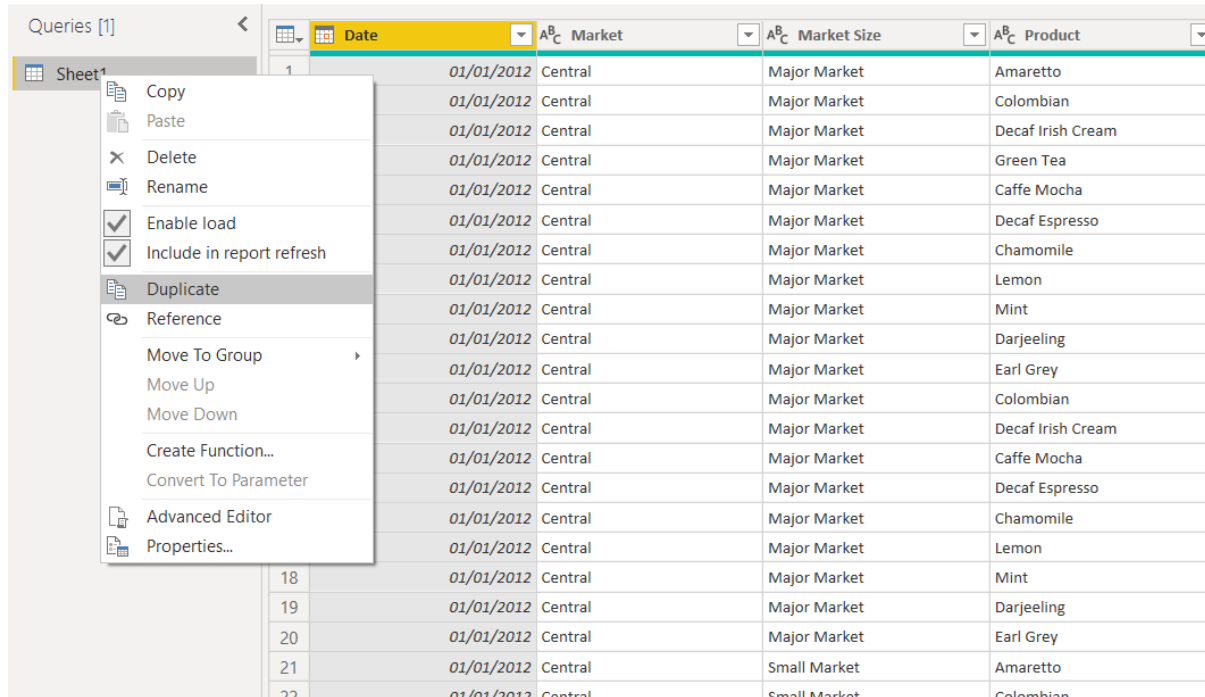
Sheet1

- Budget Cost of Go...
- Budget Margin
- Budget Profit
- Budget Sales
- Cost of Goods Sold
- Date
- Inventory
- Margin
- Market
- Market Size
- Marketing
- Product
- Product Line
- Product Type
- Profit
- Sales
- State
- Total Expenses
- Type

3 Data Modelling – Star Schema Facts and Dimensions

3.1 Splitting tables

In this dataset, try to split seven tables to normalise data. So, click “Transform data” and the “Power Query Editor” come up. After that, right-click “Sheet1” in Queries section and duplicate six tables for further usage.



The screenshot shows the Power Query Editor interface. On the left, the 'Queries' pane lists 'Sheet1'. A right-click context menu is open over 'Sheet1', with the 'Duplicate' option highlighted. The main area displays a table with the following columns: Date, Market, Market Size, and Product. The table contains 22 rows of data.

Date	Market	Market Size	Product
01/01/2012	Central	Major Market	Amaretto
01/01/2012	Central	Major Market	Colombian
01/01/2012	Central	Major Market	Decaf Irish Cream
01/01/2012	Central	Major Market	Green Tea
01/01/2012	Central	Major Market	Caffe Mocha
01/01/2012	Central	Major Market	Decaf Espresso
01/01/2012	Central	Major Market	Chamomile
01/01/2012	Central	Major Market	Lemon
01/01/2012	Central	Major Market	Mint
01/01/2012	Central	Major Market	Darjeeling
01/01/2012	Central	Major Market	Earl Grey
01/01/2012	Central	Major Market	Colombian
01/01/2012	Central	Major Market	Decaf Irish Cream
01/01/2012	Central	Major Market	Caffe Mocha
01/01/2012	Central	Major Market	Decaf Espresso
01/01/2012	Central	Major Market	Chamomile
01/01/2012	Central	Major Market	Lemon
01/01/2012	Central	Major Market	Mint
01/01/2012	Central	Major Market	Darjeeling
01/01/2012	Central	Major Market	Earl Grey
01/01/2012	Central	Small Market	Amaretto
01/01/2012	Central	Small Market	Colombian

Queries [7]

- Sheet1
- Sheet1 (2)
- Sheet1 (3)
- Sheet1 (4)
- Sheet1 (5)
- Sheet1 (6)
- Sheet1 (7)

	Area Code	Date	Market	Market Size	Product
1	719	01/01/2012	Central	Major Market	Amaro
2	970	01/01/2012	Central	Major Market	Colombian
3	970	01/01/2012	Central	Major Market	Decaf
4	303	01/01/2012	Central	Major Market	Green Tea
5	303	01/01/2012	Central	Major Market	Caffe Moch
6	720	01/01/2012	Central	Major Market	Decaf Espre
7	970	01/01/2012	Central	Major Market	Chamomile
8	719	01/01/2012	Central	Major Market	Lemon
9	970	01/01/2012	Central	Major Market	Mint
10	719	01/01/2012	Central	Major Market	Darjeeling
11	303	01/01/2012	Central	Major Market	Earl Grey
12	217	01/01/2012	Central	Major Market	Colombian
13	309	01/01/2012	Central	Major Market	Decaf Irish
14	309	01/01/2012	Central	Major Market	Caffe Moch
15	630	01/01/2012	Central	Major Market	Decaf Espre
16	312	01/01/2012	Central	Major Market	Chamomile
17	630	01/01/2012	Central	Major Market	Lemon
18	773	01/01/2012	Central	Major Market	Mint
19	217	01/01/2012	Central	Major Market	Darjeeling
20	708	01/01/2012	Central	Major Market	Earl Grey
21	319	01/01/2012	Central	Small Market	Amaro

Major Market

Right-click “Sheet1” in the Queries section and choose “Rename”. Change the name from “Sheet1” to “coffee_data_all”.

Queries [7]

- Sheet1
- Sheet1 (2)
- Sheet1 (3)
- Sheet1 (4)
- Sheet1 (5)
- Sheet1 (6)
- Sheet1 (7)

	Area Code	Date	Market	Market Size	Product
1	719	01/01/2012	Central	Major Market	Amaretto
2	970	01/01/2012	Central	Major Market	Colombian
3	970	01/01/2012	Central	Major Market	Decaf Irish
4	303	01/01/2012	Central	Major Market	Green Tea
5	303	01/01/2012	Central	Major Market	Caffe Moch
6	720	01/01/2012	Central	Major Market	Decaf Espre
7	970	01/01/2012	Central	Major Market	Chamomile
8	719	01/01/2012	Central	Major Market	Lemon
9	970	01/01/2012	Central	Major Market	Mint
10	719	01/01/2012	Central	Major Market	Darjeeling
11	303	01/01/2012	Central	Major Market	Earl Grey
12	217	01/01/2012	Central	Major Market	Colombian
13	309	01/01/2012	Central	Major Market	Decaf Irish
14	309	01/01/2012	Central	Major Market	Caffe Moch
15	630	01/01/2012	Central	Major Market	Decaf Espre
16	312	01/01/2012	Central	Major Market	Chamomile
17	630	01/01/2012	Central	Major Market	Lemon
18	773	01/01/2012	Central	Major Market	Mint
19	217	01/01/2012	Central	Major Market	Darjeeling
20	708	01/01/2012	Central	Major Market	Earl Grey
21	319	01/01/2012	Central	Small Market	Amaretto

Major Market

Queries [7]

- coffee_data_all
- Sheet1 (2)
- Sheet1 (3)
- Sheet1 (4)
- Sheet1 (5)
- Sheet1 (6)
- Sheet1 (7)

	Area Code	Date	Market	Market Size	Pro
1	719	01/01/2012	Central	Major Market	Amarett
2	970	01/01/2012	Central	Major Market	Colombi
3	970	01/01/2012	Central	Major Market	Decaf Iri
4	303	01/01/2012	Central	Major Market	Green T
5	303	01/01/2012	Central	Major Market	Caffe M
6	720	01/01/2012	Central	Major Market	Decaf Es
7	970	01/01/2012	Central	Major Market	Chamon
8	719	01/01/2012	Central	Major Market	Lemon
9	970	01/01/2012	Central	Major Market	Mint
10	719	01/01/2012	Central	Major Market	Darjeeli
11	303	01/01/2012	Central	Major Market	Earl Gre
12	217	01/01/2012	Central	Major Market	Colombi
13	309	01/01/2012	Central	Major Market	Decaf Iri
14	309	01/01/2012	Central	Major Market	Caffe M
15	630	01/01/2012	Central	Major Market	Decaf Es
16	312	01/01/2012	Central	Major Market	Chamon
17	630	01/01/2012	Central	Major Market	Lemon
18	773	01/01/2012	Central	Major Market	Mint
19	217	01/01/2012	Central	Major Market	Darjeeli
20	708	01/01/2012	Central	Major Market	Earl Gre
21	319	01/01/2012	Central	Small Market	Amarett

Repeat all the same steps and change the rest of sheet's name to "product", "product_type", "product_caffeine", "Market_size", "Area_code", "State".

Queries [7]

- coffee_data_all
- product
- product_type
- product_caffeine
- Market_size
- Area_code
- State

	Area Code	Date	Market	Market Size	Pro
1	719	01/01/2012	Central	Major Market	Am
2	970	01/01/2012	Central	Major Market	Col
3	970	01/01/2012	Central	Major Market	Dec
4	303	01/01/2012	Central	Major Market	Gre
5	303	01/01/2012	Central	Major Market	Caf
6	720	01/01/2012	Central	Major Market	Der
7	970	01/01/2012	Central	Major Market	Ch
8	719	01/01/2012	Central	Major Market	Ler
9	970	01/01/2012	Central	Major Market	Mir
10	719	01/01/2012	Central	Major Market	Dar

3.2 Remove Duplicate items

Once select “product” column, there are many duplicates items in the “Product” column and need to remove all of it because it can prevent many-to-many relationship happen. So, right-click “Product” column and choose “Remove Duplicates”.

The screenshot shows the Power BI Desktop interface. On the left, the 'Queries' pane lists several queries, with 'product' selected. The main area displays a table with columns: 'Product', 'Product Type', and 'Caffeine Type'. The 'Product' column contains 27 rows of data, including 'Amaretto', 'Colombian', 'Decaf Irish Cream', 'Green Tea', 'Caffe Mocha', 'Decaf Espresso', 'Chamomile', 'Lemon', 'Mint', 'Darjeeling', 'Earl Grey', and others. A right-click context menu is open over the 'Product' column, with 'Remove Duplicates' highlighted. Other options include 'Copy', 'Remove', 'Remove Other Columns', 'Duplicate Column', 'Add Column From Examples...', 'Remove Errors', 'Change Type', 'Transform', 'Replace Values...', 'Replace Errors...', 'Split Column', 'Group By...', 'Fill', 'Unpivot Columns', 'Unpivot Other Columns', 'Unpivot Only Selected Columns', 'Rename...', 'Move', 'Drill Down', and 'Add as New Query'.

Product	Product Type	Caffeine Type
Amaretto	Coffee	Regular
Colombian	Coffee	Regular
Decaf Irish Cream	Coffee	Decaf
Green Tea	Tea	Regular
Caffe Mocha	Espresso	Regular
Decaf Espresso	Espresso	Decaf
Chamomile	Herbal Tea	Decaf
Lemon	Herbal Tea	Decaf
Mint	Herbal Tea	Decaf
Darjeeling	Tea	Regular
Earl Grey	Tea	Regular
Colombian	Coffee	Regular
Decaf Irish Cream	Coffee	Decaf
Caffe Mocha	Espresso	Regular
Decaf Espresso	Espresso	Decaf
Chamomile	Herbal Tea	Decaf
Lemon	Herbal Tea	Decaf
Mint	Herbal Tea	Decaf
Darjeeling	Tea	Regular
Earl Grey	Tea	Regular
Amaretto	Coffee	Regular
Colombian	Coffee	Regular
Decaf Irish Cream	Coffee	Decaf
Caffe Mocha	Espresso	Regular
Decaf Espresso	Espresso	Decaf
Chamomile	Herbal Tea	Decaf
Lemon	Herbal Tea	Decaf

The screenshot shows the same Power BI Desktop interface after removing duplicates from the 'Product' column. The 'Product' column now contains 13 unique items. The 'Product Line' column has been added, showing categories like 'Beans' and 'Leaves'. The 'Product Type' and 'Caffeine Type' columns remain the same.

Product	Product Line	Product Type	Caffeine Type
Amaretto	Beans	Coffee	Regular
Colombian	Beans	Coffee	Regular
Decaf Irish Cream	Beans	Coffee	Decaf
Green Tea	Leaves	Tea	Regular
Caffe Mocha	Beans	Espresso	Regular
Decaf Espresso	Beans	Espresso	Decaf
Chamomile	Leaves	Herbal Tea	Decaf
Lemon	Leaves	Herbal Tea	Decaf
Mint	Leaves	Herbal Tea	Decaf
Darjeeling	Leaves	Tea	Regular
Earl Grey	Leaves	Tea	Regular
Regular Espresso	Beans	Espresso	Regular
Caffe Latte	Beans	Espresso	Regular

Repeat the same steps for the rest of columns. Then click “Close & Apply” to save all the changes.

Queries [7]	<	<table><tr><td></td><td>ABC Product Type</td><td>▼</td></tr><tr><td>1</td><td>Coffee</td><td></td></tr><tr><td>2</td><td>Tea</td><td></td></tr><tr><td>3</td><td>Espresso</td><td></td></tr><tr><td>4</td><td>Herbal Tea</td><td></td></tr></table>		ABC Product Type	▼	1	Coffee		2	Tea		3	Espresso		4	Herbal Tea	
	ABC Product Type	▼															
1	Coffee																
2	Tea																
3	Espresso																
4	Herbal Tea																
coffee_data_all																	
product																	
product_type																	
product_caffeine																	
Market_size																	
Area_code																	
State																	

Queries [7]	<	<table><tr><td></td><td>ABC Caffeine Type</td><td>▼</td></tr><tr><td>1</td><td>Regular</td><td></td></tr><tr><td>2</td><td>Decaf</td><td></td></tr></table>		ABC Caffeine Type	▼	1	Regular		2	Decaf	
	ABC Caffeine Type	▼									
1	Regular										
2	Decaf										
coffee_data_all											
product											
product_type											
product_caffeine											
Market_size											
Area_code											
State											

Queries [7]	<	<table><tr><td></td><td>ABC Market Size</td><td>▼</td></tr><tr><td>1</td><td>Major Market</td><td></td></tr><tr><td>2</td><td>Small Market</td><td></td></tr></table>		ABC Market Size	▼	1	Major Market		2	Small Market	
	ABC Market Size	▼									
1	Major Market										
2	Small Market										
coffee_data_all											
product											
product_type											
product_caffeine											
Market_size											
Area_code											
State											

Queries [7]			
coffee_data_all			
product			
product_type			
product_caffeine			
Market_size			
Area_code			
State			
	1 ² ₃ Area Code	A ^B _C State	
1	719	Colorado	
2	970	Colorado	
3	303	Colorado	
4	720	Colorado	
5	217	Illinois	
6	309	Illinois	
7	630	Illinois	
8	312	Illinois	
9	773	Illinois	
10	708	Illinois	
11	319	Iowa	
12	641	Iowa	
13	712	Iowa	
14	563	Iowa	
15	636	Missouri	
16	573	Missouri	
17	417	Missouri	
18	660	Missouri	
19	314	Missouri	
20	740	Ohio	
21	614	Ohio	
22	216	Ohio	
23	937	Ohio	
24	567	Ohio	
25	262	Wisconsin	
26	414	Wisconsin	
27	715	Wisconsin	

Queries [7]	<	<div><div>ABC State</div><div></div></div>
<div>coffee_data_all</div>		1 Colorado
<div>product</div>		2 Illinois
<div>product_type</div>		3 Iowa
<div>product_caffeine</div>		4 Missouri
<div>Market_size</div>		5 Ohio
<div>Area_code</div>		6 Wisconsin
<div>State</div>		7 Connecticut
		8 Florida
		9 Massachusetts
		10 New Hampshire
		11 New York
		12 Louisiana
		13 New Mexico
		14 Oklahoma
		15 Texas
		16 Utah
		17 California
		18 Nevada
		19 Oregon
		20 Washington

3.3 M Language and Create New Column

The following paragraph will describe a few steps to create a calendar table by day, year, quarter, week number, month number, month, and day of week.

In the left-hand side of queries, right-click and create "Blank Query". Choose "Advanced Editor" and write the following code in Query1.

```
#create date dimension
(StartDate as date, EndDate as date)=>

let
#date ranges from parameters
StartDate = #date(Date.Year(StartDate), Date.Month(StartDate),
Date.Day(StartDate)),
EndDate = #date(Date.Year(EndDate), Date.Month(EndDate),
Date.Day(EndDate)),

#the number of dates required for the table
GetDateCount = Duration.Days(EndDate - StartDate),

#Take count of dates and get it into a list of dates
GetDateList = List.Dates(StartDate, GetDateCount,
#duration(1,0,0,0)),

#Convert the list into a table
DateListToTable = Table.FromList(GetDateList,
Splitter.SplitByNothing(), {"Date"}, null, ExtraValues.Error),

#Add year column
YearNumber = Table.AddColumn(DateListToTable, "Year",
each Date.Year([Date])),

#Add quarter number column
QuarterNumber = Table.AddColumn(YearNumber, "Quarter",
each "Q" & Number.ToText(Date.QuarterOfYear([Date]))),

#Add week number column
WeekNumber= Table.AddColumn(QuarterNumber, "Week Number",
each Date.WeekOfYear([Date])),

#Add month number column
MonthNumber = Table.AddColumn(WeekNumber, "Month Number",
each Date.Month([Date])),
```

#Add month column

**MonthName = Table.AddColumn(MonthNumber , "Month",
each Date.ToText([Date],"MMMM")),**

**in
DayOfWeek**

Queries [7] fx = Table.RemoveColumns(#"Renamed Columns1",{"Marketing"})

	Area Code	Date	Market	Market Size	Product
1	719	01/01/2012	Central	Major Market	Amaretto
2	970	01/01/2012	Central	Major Market	Colombian
3	970	01/01/2012	Central	Major Market	Decaf Irish Cream
4	303	01/01/2012	Central	Major Market	Green Tea
5	303	01/01/2012	Central	Major Market	Caffe Mocha
6	720	01/01/2012	Central	Major Market	Decaf Espresso
7	970	01/01/2012	Central	Major Market	Chamomile
8	719	01/01/2012	Central	Major Market	Lemon
9	970	01/01/2012	Central	Major Market	Mint
10	719	01/01/2012	Central	Major Market	Darjeeling
11	303	01/01/2012	Central	Major Market	Earl Grey
12	217	01/01/2012	Central	Major Market	Colombian
13	309	01/01/2012	Central	Major Market	Decaf Irish Cream
	309	01/01/2012	Central	Major Market	Caffe Mocha
		01/01/2012	Central	Major Market	Decaf Espresso
		01/01/2012	Central	Major Market	Chamomile
		01/01/2012	Central	Major Market	Lemon
		01/01/2012	Central	Major Market	Mint
		01/01/2012	Central	Major Market	Darjeeling
		01/01/2012	Central	Major Market	Earl Grey
21		01/01/2012	Central	Small Market	Amaretto
22		01/01/2012	Central	Small Market	Colombian
23		01/01/2012	Central	Small Market	Decaf Irish Cream
24		01/01/2012	Central	Small Market	Caffe Mocha
25		01/01/2012	Central	Small Market	Decaf Espresso
26	641	01/01/2012	Central	Small Market	Chamomile

Close & Apply * New Source * Recent Sources * Enter Data * Data source settings * Manage Parameters * Refresh Preview * Properties * Advanced Editor * Choose Columns * Remove Columns * Keep Rows * Remove Rows * Split Column * Group By * Data Type: Any * Merge Queries * Append Queries * Combine Files * Text Analytics * Vision * Azure Machine Learning

Close New Query

Queries [8]

- coffee_data_all
- product
- product_type
- product_caffeine
- Market_size
- Area_code
- State
- Query1

Advanced Editor

Query1

```
let
    Source = ""
in
    Source
```

Display Options

✓ No syntax errors have been detected.

Done Cancel

Advanced Editor

Query1

Display Options ?

```
(StartDate as date, EndDate as date)=>
let
    StartDate = #date(Date.Year(StartDate), Date.Month(StartDate),
        Date.Day(StartDate)),
    EndDate = #date(Date.Year(EndDate), Date.Month(EndDate),
        Date.Day(EndDate)),

    GetDateCount = Duration.Days(EndDate - StartDate),

    GetDateList = List.Dates(StartDate, GetDateCount,
        #duration(1,0,0,0)),

    DateListToTable = Table.FromList(GetDateList,
        Splitter.SplitByNothing(), {"Date"}, null, ExtraValues.Error),
    YearNumber = Table.AddColumn(DateListToTable, "Year",
        each Date.Year([Date])),
    QuarterNumber = Table.AddColumn(YearNumber, "Quarter",
        each "Q" & Number.ToText(Date.QuarterOfYear([Date]))),
    WeekNumber = Table.AddColumn(QuarterNumber, "Week Number",
        each Date.WeekOfYear([Date])),
    MonthNumber = Table.AddColumn(WeekNumber, "Month Number",
        each Date.Month([Date])),
    MonthName = Table.AddColumn(MonthNumber, "Month",
        each Date.ToText([Date], "MMMM")),
    DayOfWeek = Table.AddColumn(MonthName, "Day of Week",
        each Date.ToText([Date], "dddd"))
in
    DayOfWeek
```

✓ No syntax errors have been detected.

Done Cancel

Once click “Done”, parameters are required to input the date. According to our dataset, the start date is 01/01/2012, and the end date is 31/12/2013.

Queries [8]

- coffee_data_all
- product
- product_type
- product_caffeine
- Market_size
- Area_code
- State
- Query1**

fx = (StartDate as date, EndDate as date)=>

Enter Parameters

StartDate
01/01/2012

EndDate
31/12/2013

Invoke Clear

function (StartDate as date, EndDate as date) as any

The new table is created in “Queries” and rename the table as “Calendar”. Change the first column of type to “Date” for further usage.

Queries [9] = Query1(#date(2012, 1, 1), #date(2013, 12, 31))

	ABC 123 Date	ABC 123 Year	ABC 123 Quarter	ABC 123 Week Number	ABC 123 Month Number
1	01/01/2012	2012	Q1		1
2	02/01/2012	2012	Q1		2
3	03/01/2012	2012	Q1		2
4	04/01/2012	2012	Q1		2
5	05/01/2012	2012	Q1		2
6	06/01/2012	2012	Q1		2
7	07/01/2012	2012	Q1		2
8	08/01/2012	2012	Q1		2
9	09/01/2012	2012	Q1		3
10	10/01/2012	2012	Q1		3
11	11/01/2012	2012	Q1		3
12	12/01/2012	2012	Q1		3
13	13/01/2012	2012	Q1		3
14	14/01/2012	2012	Q1		3
15	15/01/2012	2012	Q1		3
16	16/01/2012	2012	Q1		4
17	17/01/2012	2012	Q1		4
18	18/01/2012	2012	Q1		4
19	19/01/2012	2012	Q1		4
20	20/01/2012	2012	Q1		4
21	21/01/2012	2012	Q1		4
22	22/01/2012	2012	Q1		4
23	23/01/2012	2012	Q1		5
24	24/01/2012	2012	Q1		5
25	25/01/2012	2012	Q1		5
26	26/01/2012	2012	Q1		5

Queries [9] = Query1(#date(2012, 1, 1), #date(2013, 12, 31))

	ABC 123 Date	ABC 123 Year	ABC 123 Quarter	ABC 123 Week Number	ABC 123 Month Number
1	1.2 Decimal Number		2012 Q1		1
2	\$ Fixed decimal number		2012 Q1		2
3	1²³ Whole Number		2012 Q1		2
4	% Percentage		2012 Q1		2
5	Date/Time		2012 Q1		2
6	Date		2012 Q1		2
7	Time		2012 Q1		2
8	Date/Time/Timezone		2012 Q1		2
9	Duration		2012 Q1		3
10	Text		2012 Q1		3
11	True/False		2012 Q1		3
12	Binary		2012 Q1		3
13	Using Locale...		2012 Q1		3
14			2012 Q1		3
15	15/01/2012		2012 Q1		3
16	16/01/2012		2012 Q1		4
17	17/01/2012		2012 Q1		4
18	18/01/2012		2012 Q1		4
19	19/01/2012		2012 Q1		4
20	20/01/2012		2012 Q1		4
21	21/01/2012		2012 Q1		4
22	22/01/2012		2012 Q1		4
23	23/01/2012		2012 Q1		5
24	24/01/2012		2012 Q1		5

3.4 Data Model

Now look at the “Model” section, which are eight tables. It shows their structure for each table as below.

1) Market_size

Column	Description
Market Size	Market size of USA

2) State

Column	Description
State	State name of USA

3) Area_code

Column	Description
Area Code	Area code of USA
State	State name of USA

4) Product_caffeine

Column	Description
Caffeine Type	Type of caffeine of products

5) product_type

Column	Description
Product Type	Type of products

6) product

Column	Description
Product	Name of products
Product Line	Original material of products
Product Type	Type of products
Caffeine Type	Type of caffeine of products

7) Calendar

Column	Description
Date	Date of purchasing
Year	Year of purchasing
Quarter	Quarter of purchasing

Week Number	Week number of the year
Month Number	Month number of the year
Month	Month of the year
Day of Week	Day of week of the year

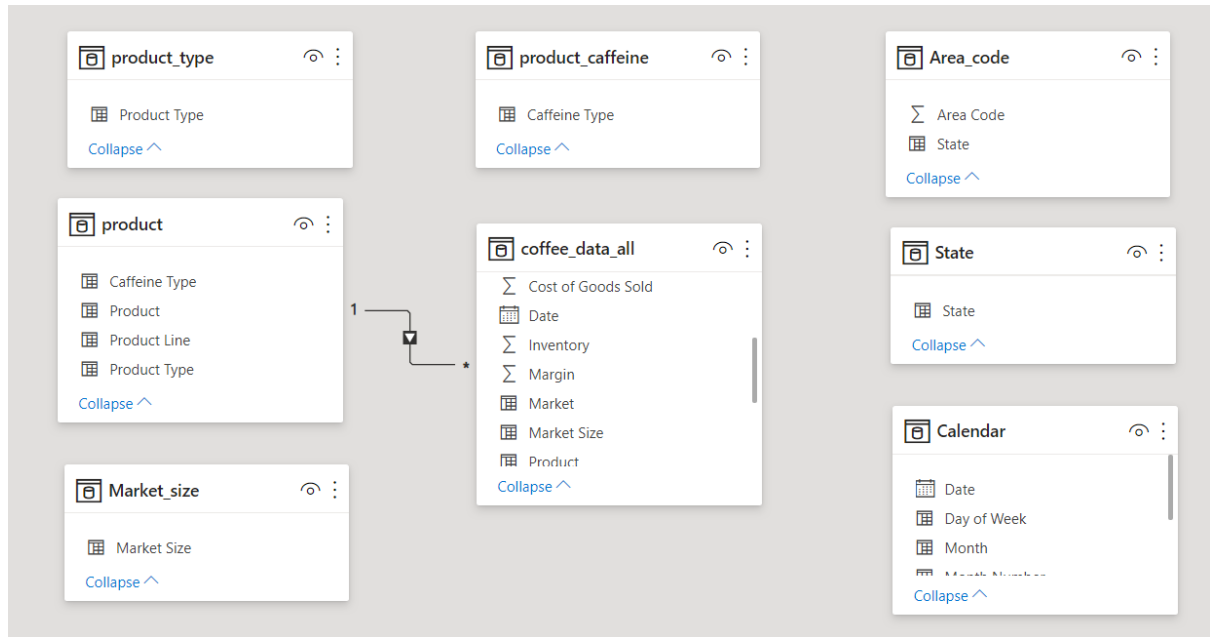
8) coffee_data_all

Column	Description
Date	Date of purchasing
Market	Location of market
Market Size	Size of market
Product	Name of products
Product Line	Original material of products
Product Type	Type of products
State	Location of state in USA
Caffeine Type	Type of caffeine of products
Budget Cost of Goods Sold	Budget cost of goods sold
Budget Margin	Budget margin of products
Budget Profit	Budget profit of products
Budget Sales	Budgets sales of products
Cost of Goods Sold	Cost of goods sold
Inventory	Inventory of products
Margin	Margin of products
Marketing	Marketing expenses of products
Profit	Profit of products
Sales	Sales of products
Total Expenses	Total expenses of products

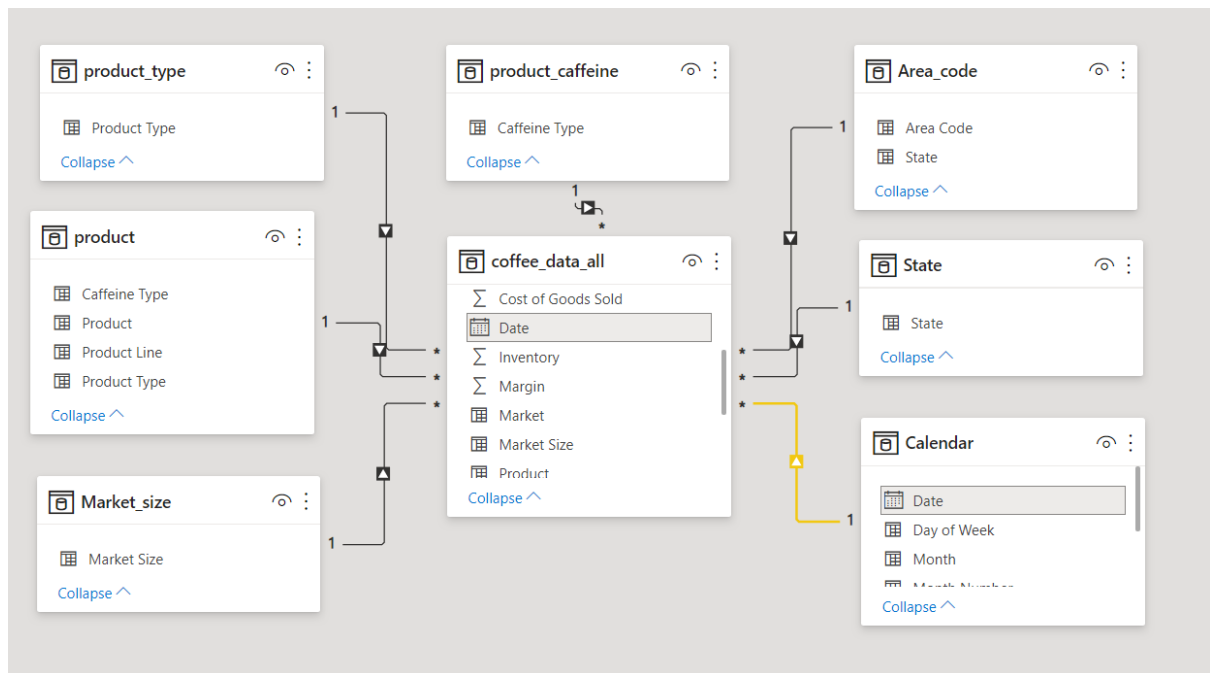
3.5 Schema and Relationships

We can find that all tables have not any relationship and Star schema format will be used to create relationship and link up together. In the dataset, “coffee_data_all” table is the main one and it relates with all other tables.

For creating relationship with “product” and “coffee_data_all”, click “Product” of “product” table, then hold and drag to “Product” of “coffee_data_all” table.



Use the same steps for the rest of the model, and look at the relationship as below. As all tables of duplicated items are removed before, so it do not have any many-to-many relationship issue happened.



SECTION 2 – BUSINESS INTELLIGENCE SOLUTION

COFFEE CHAIN ANALYSIS IN 2012-2013

1 Executive Summary

Background – Coffee is one of most important beverages for most people during the day. Believe that it should be a huge market over the world and specially people who live in North America. The dataset of coffee chain house is the best choice and create data visualisation to identify data patterns. Exploring this dataset and find out an underlying issue. Sales amount, profit, expense for each product in different states will be a main concern which it will focus on.

The aim of this report will provide vital sights on how to improve the sales volume and profit. Sales and marketing team can setup marketing plan and sales strategy based on this business report to improve overall revenue.

2 Key Findings

- Most of tea product's sales volume is better than expected and it has over 15% positive variance. However, coffee product's performance is totally out of expectation and some product's sales volume are also overestimated.
- The profit margin of coffee products is much better than tea products. The profit of Colombian and Regular Espresso nearly double of total expenses.
- The state of Iowa is a small market compare with other big cities, but the profit volume is much better the other major market, like Florida and New York.
- Caffè Mocha is the most popular product in the major market of Illinois in 2012 and 2013.

PRODUCT SALES ANALYSIS			
Product	Budget Sales	Sales	Variance %
⊕ Earl Grey	\$50,900	\$66,772	24%
⊕ Green Tea	\$25,340	\$32,850	23%
⊕ Darjeeling	\$57,360	\$73,151	22%
⊕ Mint	\$28,320	\$35,710	21%
⊕ Lemon	\$78,300	\$95,926	18%
⊕ Chamomile	\$63,840	\$75,578	16%
⊕ Caffè Latte	\$30,540	\$35,899	15%
⊕ Regular Espresso	\$22,620	\$24,031	6%
⊕ Decaf Espresso	\$75,720	\$78,162	3%
⊕ Caffè Mocha	\$84,600	\$84,904	0%
⊕ Amaretto	\$27,200	\$26,269	-4%
⊕ Colombian	\$134,380	\$128,311	-5%
⊕ Decaf Irish Cream	\$67,040	\$62,248	-8%
Total	\$746,160	\$819,811	9%

3 Recommendations

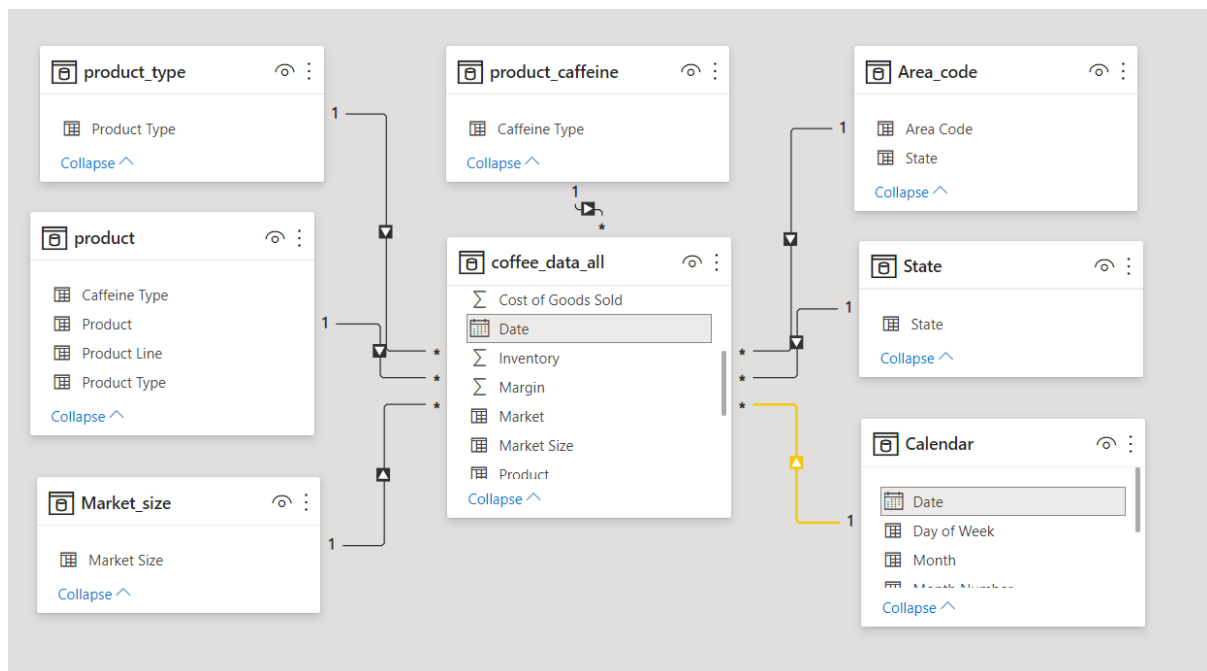
- Boosting sales of coffee products.
- Finding the solutions to control the cost of tea products.
- Providing suggestion for the market team about the marketing plan.

4 Data Description

The dataset is for the coffee chain that take place in the different states of United States of America and each type of product's statistics in 2012-2013. It is taken from the webpage, www.dataplusscience.com, and the total weight of the data is 403kb in excel file.

5 Data Model

The data model is shown as below and the table of “coffee_data_all” is the key element with all the other table. Basically, it separates with products type, product, market size, caffeine type, area code, state, and calendar.



6 Introduction

The business report is including three sections, which are sales summary, sales insight, and sales forecast and artificial intelligence. The following section will describe of this dataset and provide some basic understanding to target audience. Then the other section will compare the data between 2012 and 2013, which it will find some insights of different states. For the last section, it tries to forecast the upcoming sales volume and use artificial intelligence tool to find out some relationships with sales aspect.

7 Finding based on analysis and evaluation – 1

7.1 Matrix and Key Finding From The Visual

PRODUCT SALES ANALYSIS			
Product	Budget Sales	Sales	Variance %
⊕ Earl Grey	\$50,900	\$66,772	24%
⊕ Green Tea	\$25,340	\$32,850	23%
⊕ Darjeeling	\$57,360	\$73,151	22%
⊕ Mint	\$28,320	\$35,710	21%
⊕ Lemon	\$78,300	\$95,926	18%
⊕ Chamomile	\$63,840	\$75,578	16%
⊕ Caffè Latte	\$30,540	\$35,899	15%
⊕ Regular Espresso	\$22,620	\$24,031	6%
⊕ Decaf Espresso	\$75,720	\$78,162	3%
⊕ Caffè Mocha	\$84,600	\$84,904	0%
⊕ Amaretto	\$27,200	\$26,269	-4%
⊕ Colombian	\$134,380	\$128,311	-5%
⊕ Decaf Irish Cream	\$67,040	\$62,248	-8%
Total	\$746,160	\$819,811	9%

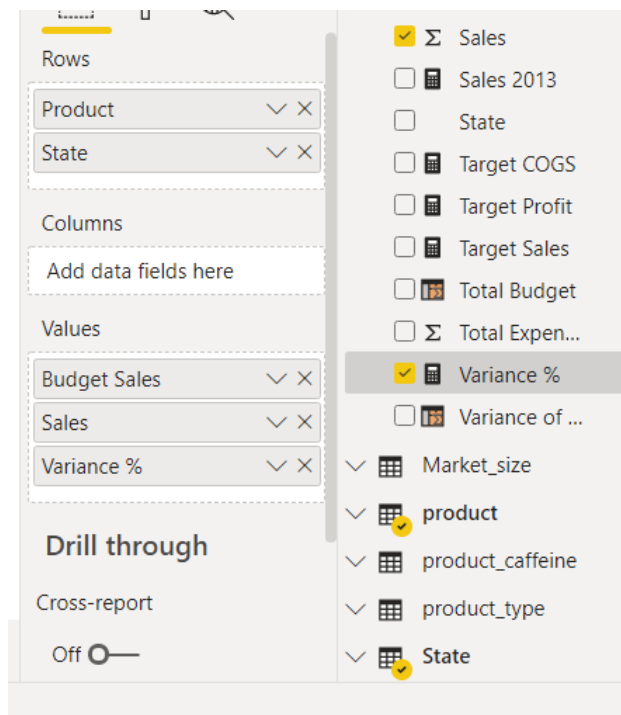
It mentioned before in the key finding for this chart. It shows the variance of budget sales and sales with each product. Most of tea product's sales volume is better than expected and it has over 10% positive variance. However, coffee product's performance is totally out of expectation and some product's sales volume are also overestimated.

7.1.1 DAX Expressions – 1 and New Calculated Measure Add Into The Model

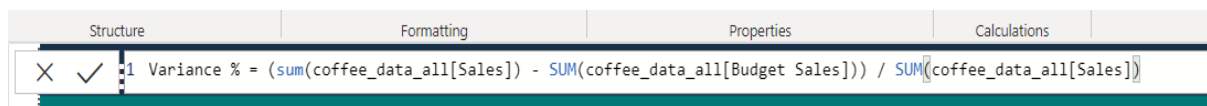
The percentage of variance is good to illustrate about the difference. However, it requires to create a new column of the model in the Data section and use DAX language to calculate the variance of sales. The formula shows the value of sales subtracts budget sales and it is shown as below.

Variance of Sales = coffee_data_all[Sales]-coffee_data_all[Budget Sales]

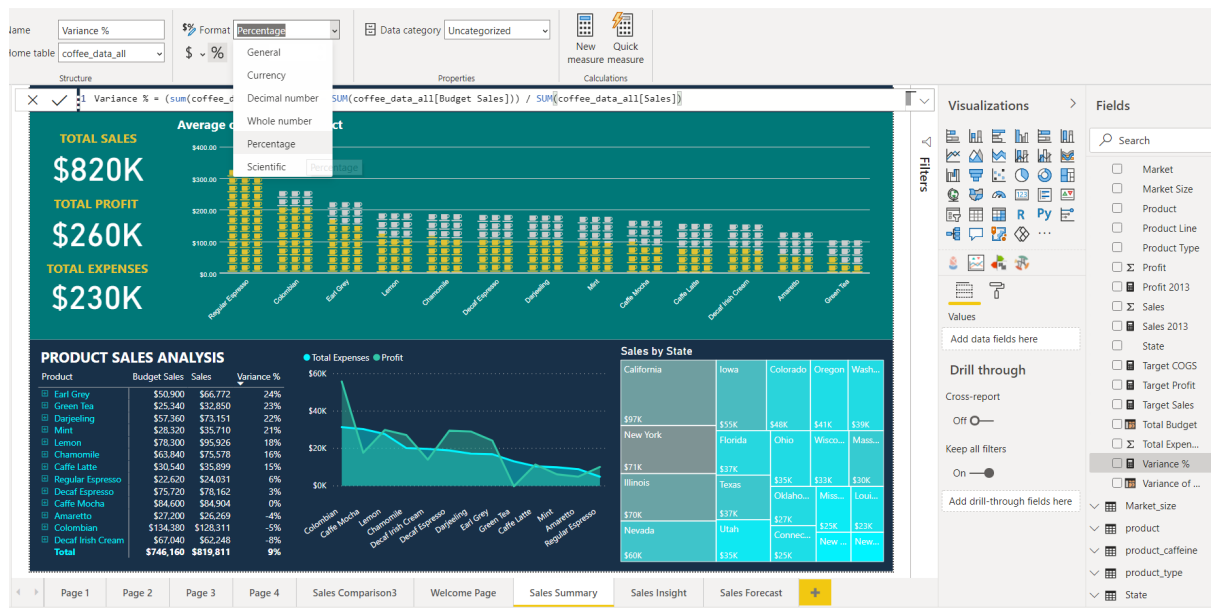
Then, move back to report section that create a new measure and input the formula to calculate the percentage of variance between the budget sales and sales. The formula is shown as below.



Variance % = (sum(coffee_data_all[Sales]) - SUM(coffee_data_all[Budget Sales])) / SUM(coffee_data_all[Sales])



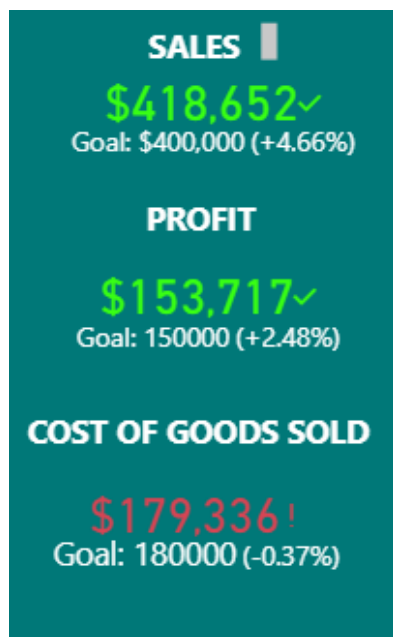
We have to sum up the total of sales and subtract the sum of the budget sales, then divide by the sum of sales again. After that, change the format from “General” to “Percentage” for this measure.



7.2 Key Performance Indicator (KPI)

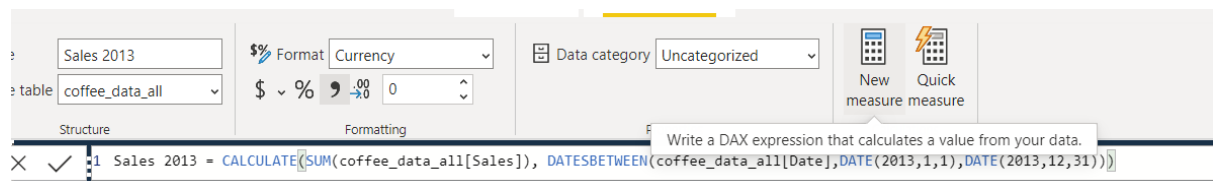
Key Performance Indicator (KPI) can evaluate whether the current value is fulfilled our target or not. For a KPI visual, setup the target measure or value is necessary and compare the current value.

For the next graph, it will investigate a KPI of the Sales, Profit and Cost of goods sold in 2013. It will set the goal for each of value and it also have to write a formula in DAX language to filter the data for 2013 only.



7.2.1 Dax Expression - 2

There are three KPI graphs to illustrate Sales and Profit are completed the goal and Cost of goods Sold is almost fulfilled the target. For those KPI setting, we create three measures and input the formulas as below:

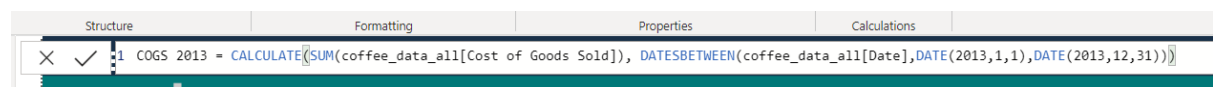
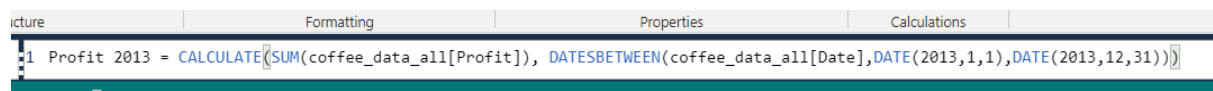
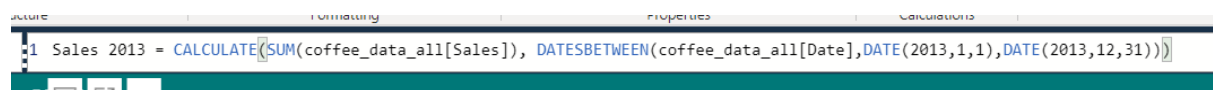


**Sales 2013 = CALCULATE(SUM(coffee_data_all[Sales]),
DATESBETWEEN(coffee_data_all[Date],DATE(2013,1,1),DATE(2013,12,31)))**

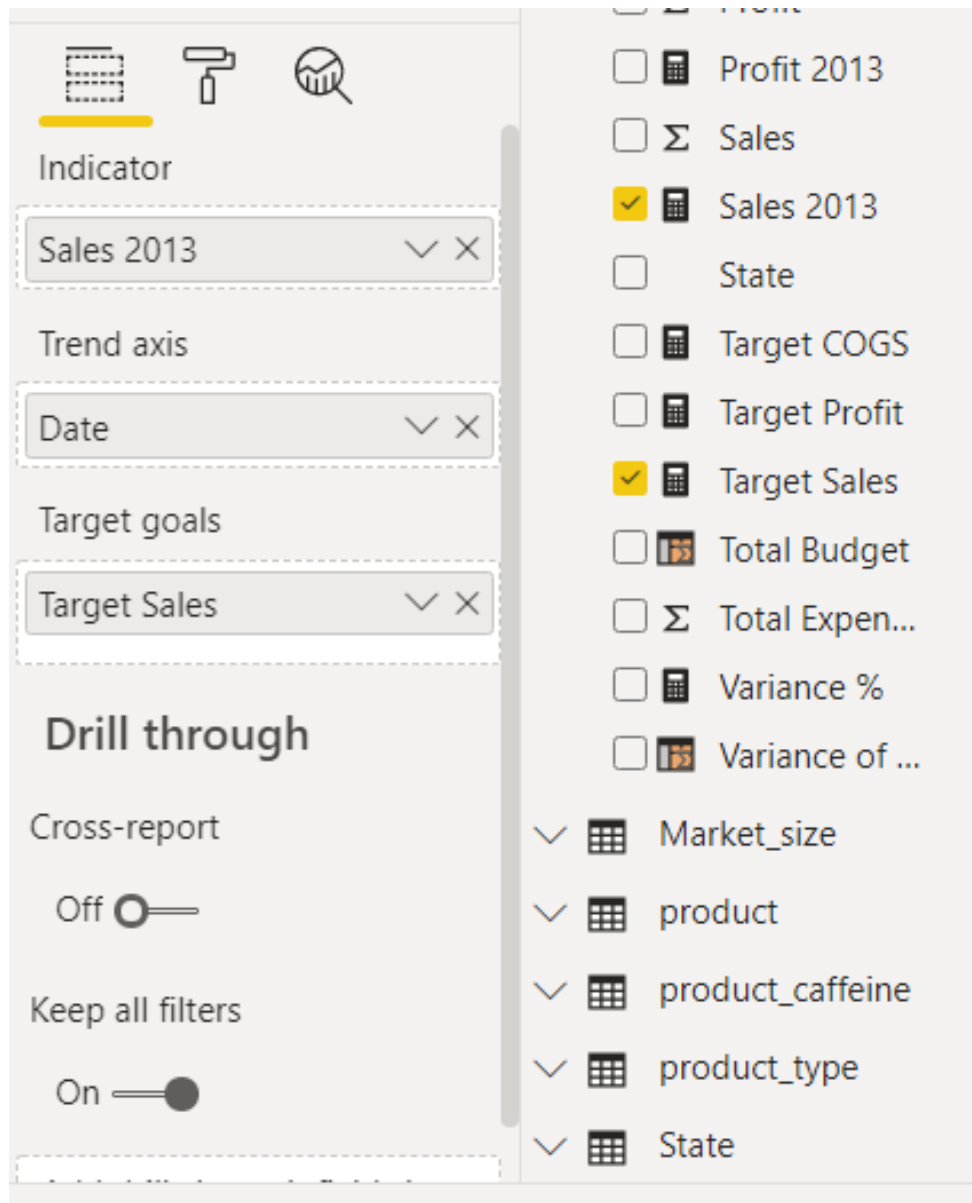
**Profit 2013 = CALCULATE(SUM(coffee_data_all[Profit]),
DATESBETWEEN(coffee_data_all[Date],DATE(2013,1,1),DATE(2013,12,31)))**

**COGS 2013 = CALCULATE(SUM(coffee_data_all[Cost of Goods Sold]),
DATESBETWEEN(coffee_data_all[Date],DATE(2013,1,1),DATE(2013,12,31)))**

Those formulas calculate the sum of Sales, Profit and Cost of goods Sold. Then use “DATEBETWEEN” function to filter the period as we required.

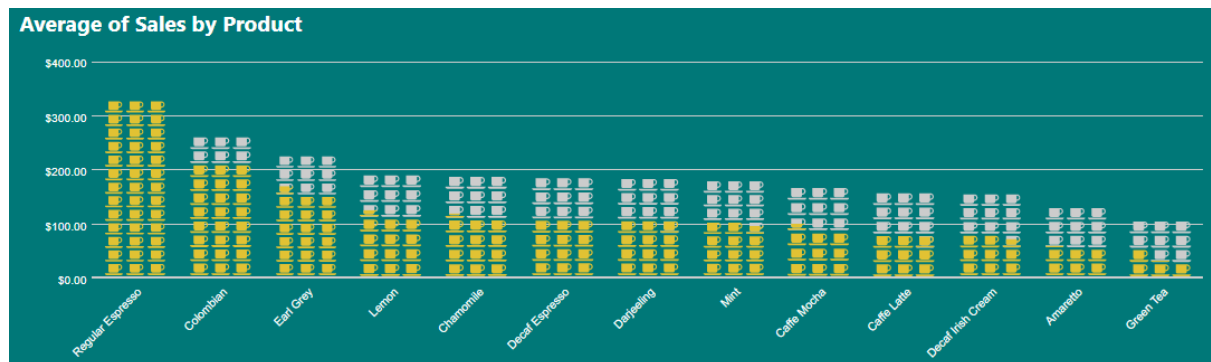


After that, it should be put those measures into “Indicator” and remember to set the target goal as required for KPI.



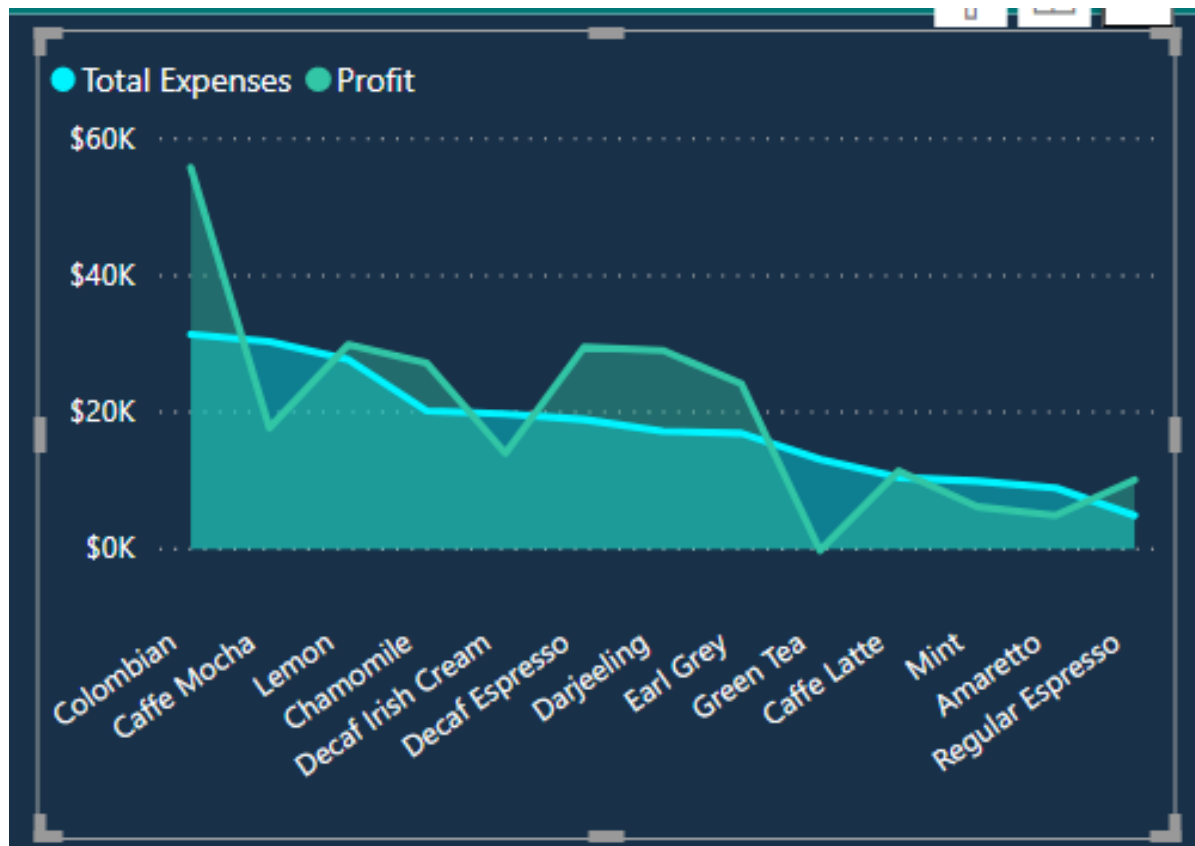
8 Finding based on analysis and evaluation - 2

8.1 Infographic



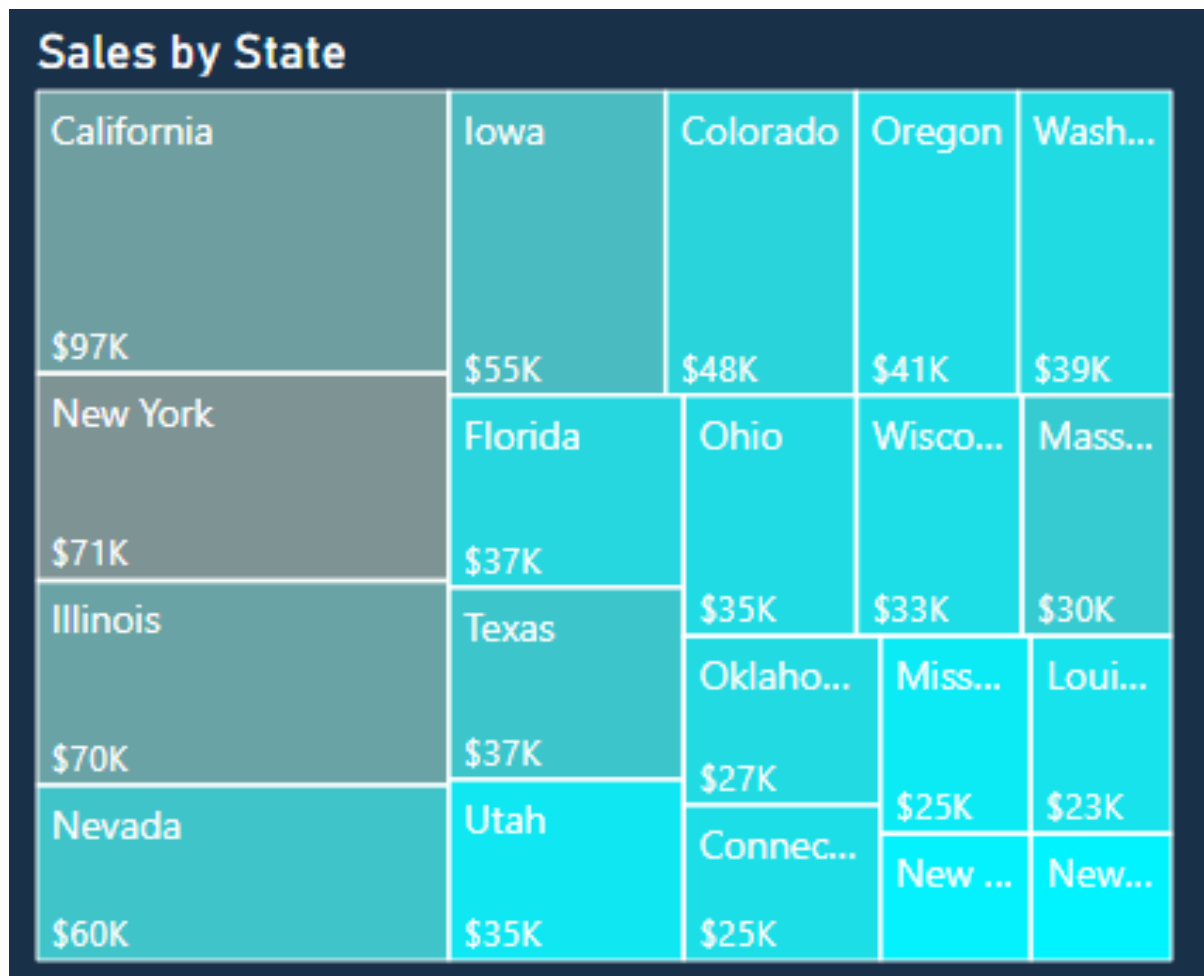
It can show about the information of the average of sales by product. Regular espresso is very popular product, which the average sales is over \$300 USD. However, green tea is not many people to buy and it has around \$100 USD sales only.

8.2 Area Chart and Key Finding From The Visual



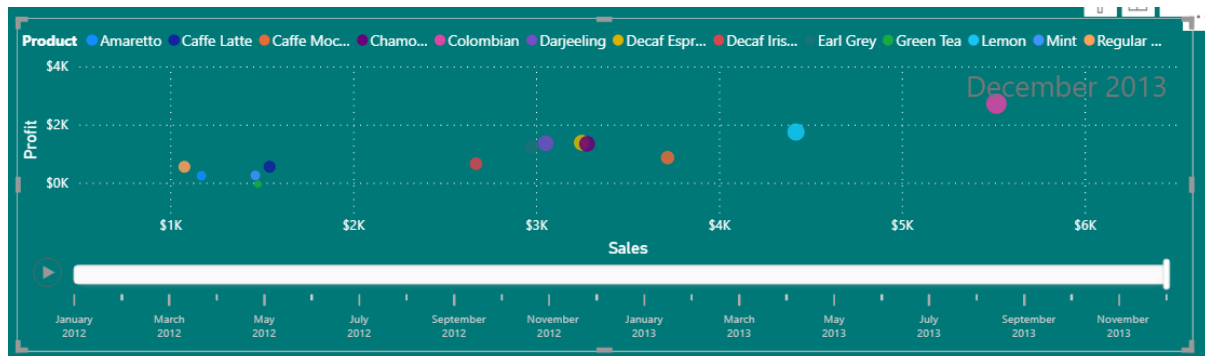
This data is also mentioned in the key finding before. Looking at the relationship with total expenses and profit in this chart. The profit margin of coffee products is much better than tea products. The profit of Colombian and Regular Espresso nearly double of total expenses. Also, it indicates some products lose profit, which is Caffè Mocha, Decaf Irish Cream, or Green Tea etc.

8.3 Treemap



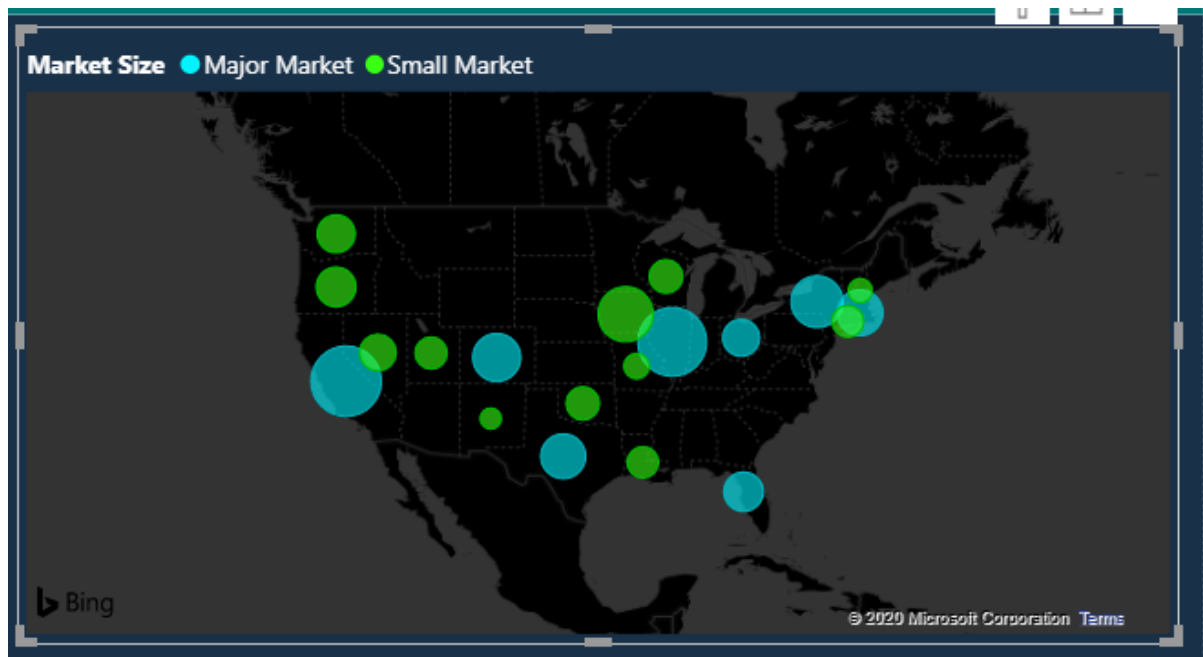
It represents the relationship with Sales and State in the treemap. California, New York, and Illinois are the highest sales in USA. If decision maker desires to find the other states, they can start exploring from left to right and up to down with descending order of sales.

8.4 Animated Chart



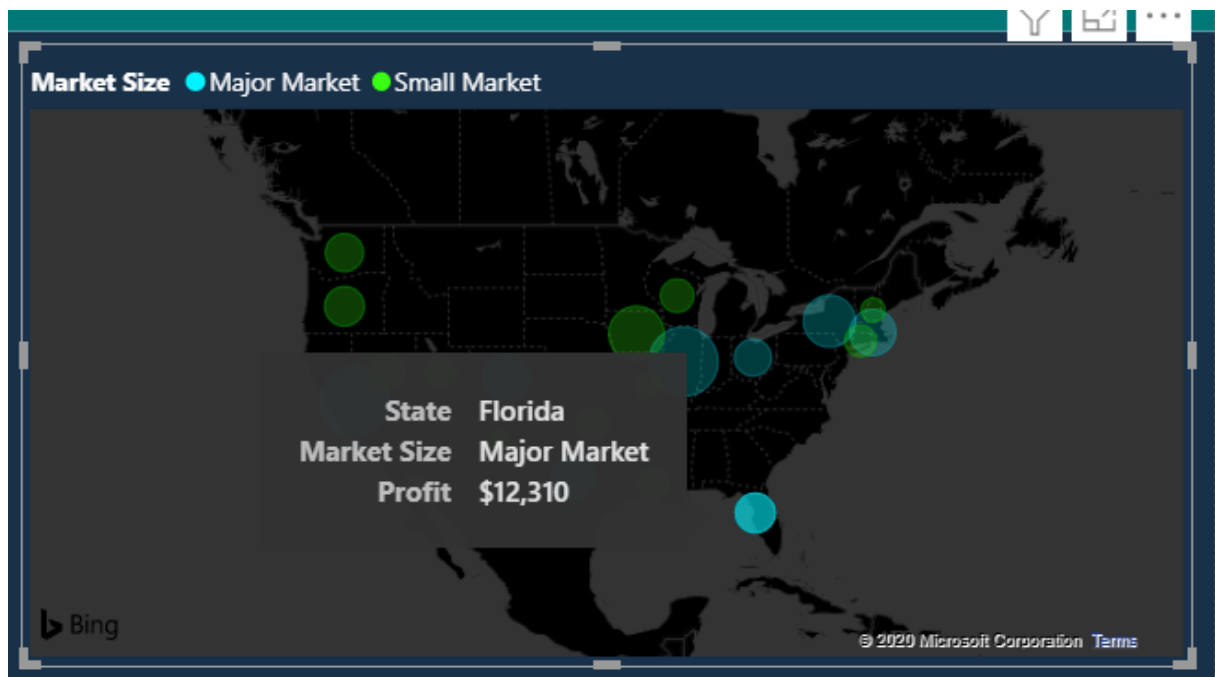
The main reason to use animated chart that it can draw the trend line by relationship changes over the period. It attempts to find the relationship between sales and profit with each product.

8.5 Map

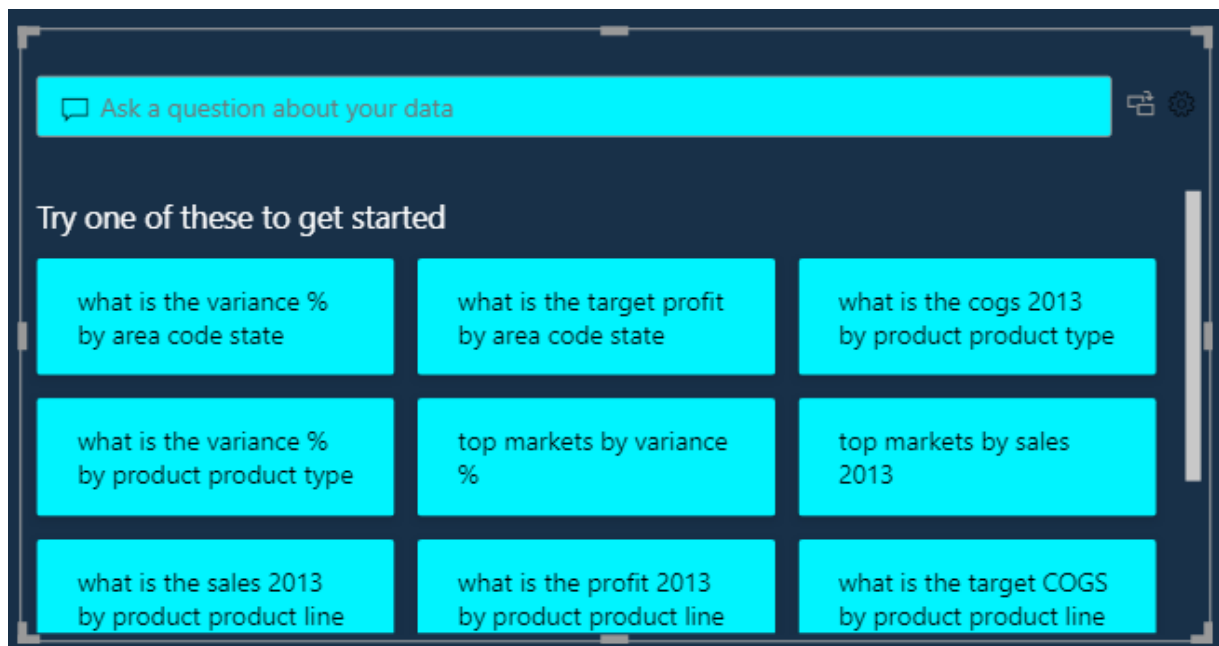


The map can show the distribution of market size and profit by state. It can be easy to find the size of profit depends on the bubble's size. The state of Iowa even is a small market compare with other big cities, but the profit volume is much better the other major market, like Florida and New York.





8.6 Ask a Question Tool

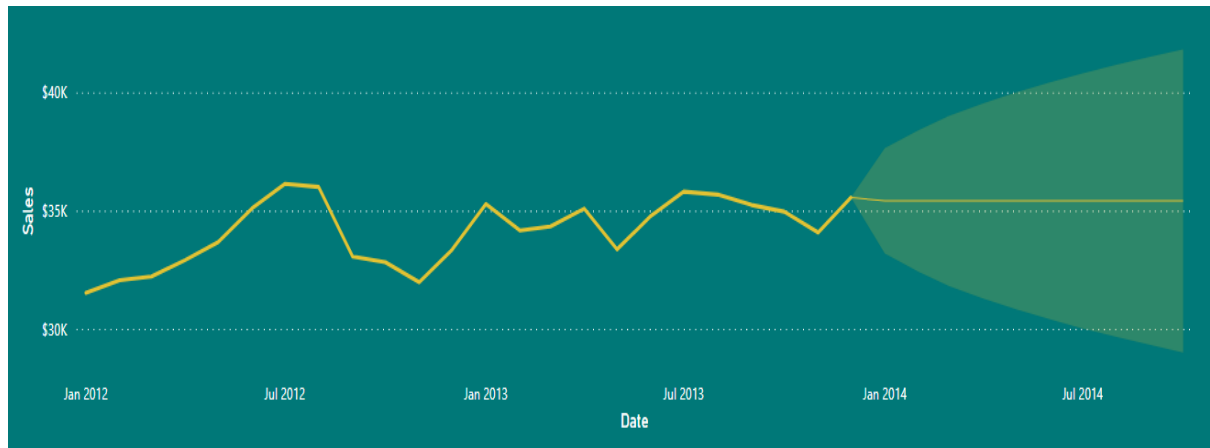


Ask a question tool is embedded into our dashboard and it is possible to type a question or click a question into the question tool to display the result that someone looking for.



9 Data Analytics (Forecast, Analysing trends)

9.1 Forecast



Sales volume is one the most important thing that target audience would like to concern, as they desire to forecast the value for further action. Power BI can be able to apply forecasting function into the chart. It can use the data to analysis and predict the coming value for our reference.

Once created the chart, it is available to choose “Analytics” and add “Forecast”. Then, it has the extension at the end of the result, and it is a prediction for your reference. Also, look at the table with numbers which may use it for more details.

The screenshot shows a forecasting tool interface. At the top, there are three icons: a grid, a funnel, and a magnifying glass. Below them is a search bar. The main section is titled 'Forecast 1' with a close button. Inside this section, there are several settings: 'Forecast length' set to 10 with a dropdown menu showing 'Point(s)'; 'Ignore last' set to 0 with a dropdown menu showing 'Point(s)'; 'Confidence interval' set to 95% with a dropdown menu; and 'Seasonality' set to 'Auto' with a dropdown menu showing 'Point(s)'.

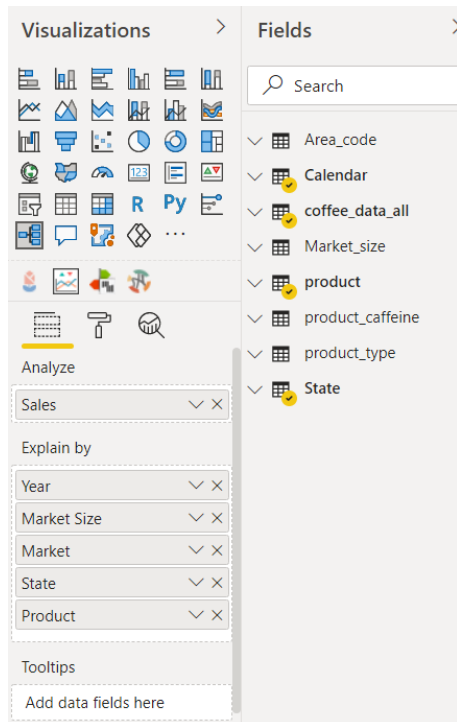
Date	Sales	forecastValue	confidenceHighBound	confidenceLowBound
October 2013	\$34,103			
November 2013	\$34,103			
December 2013	\$35,580	35580	35580	35580
January 2014		35441	37668	33215
February 2014		35441	38437	32446
March 2014		35441	39045	31838
April 2014		35441	39565	31318
May 2014		35441	40026	30857
June 2014		35441	40444	30438
July 2014		35441	40831	30052
August 2014		35441	41191	29692
September 2014		35441	41530	29353
October 2014		35441	41851	29032

9.2 Artificial Intelligence And Key Finding From The Visual



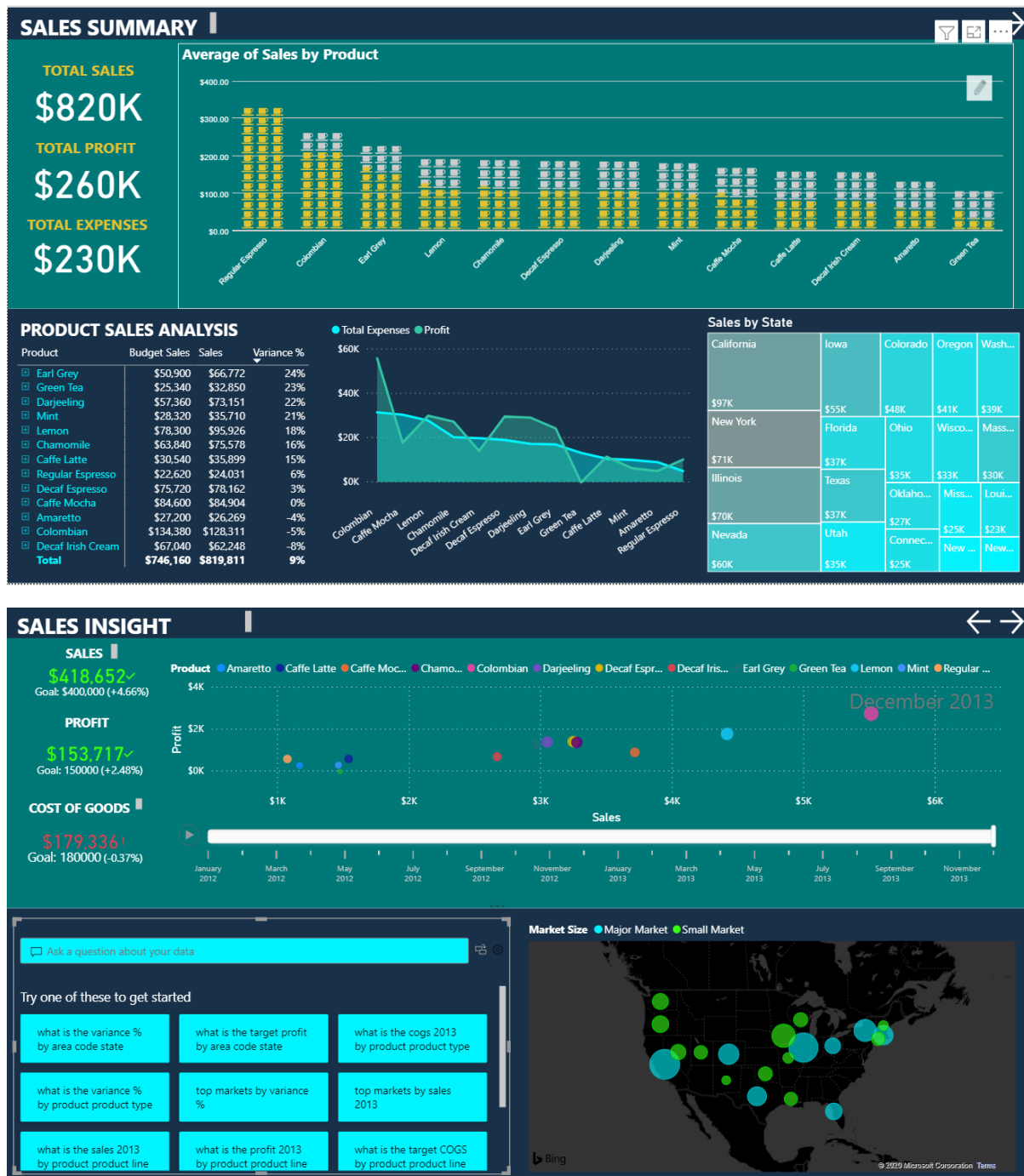
The decomposition tree can visualise data through many dimensions. It can aggregate data and be able to drill down into the dimensions in any order automatically. It is an artificial intelligence visualisation, which finds the next dimension to drill down into based on specific criteria.

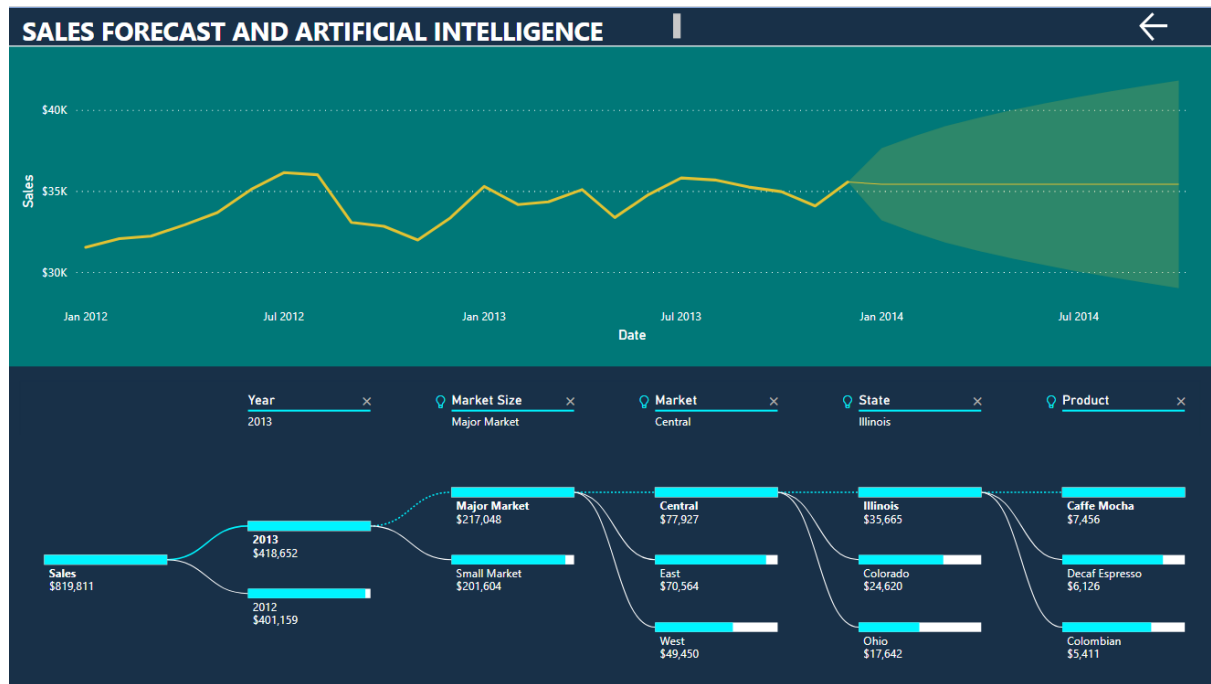
In the “Explain by” column, choose our own pattern and put the items that desire to analysis in “Analyze” column. As this graph mentioned in the key finding before, Caffe Mocha is the most popular product in the major market of Illinois in 2012 and 2013. Also, it is possible to find many relationships from decomposition tree with different mix and match.



10 Structure of the Dashboard

Try to structure the dashboard and the purpose to make it clean and easy to read. It attempts to keep colour and background consistently. Also, it is necessary to put the most important chart on the top and place the others for the rest of area, which it can capture audience attention.





11 Conclusion

In this report, it can find some insights with different attributes, which is like the average of sales by product, the variance between budget sales and sales. Also, it illustrates about the relationship with total expenses and profit, and key performance indicator in the dashboard. Sales forecast and artificial intelligence can provide different ideas for management team to make decision. Also, the structure of dashboard can make target audience easy to understand the dataset in a short time.

12 Recommendations

From the report, it indicates some products are losing money, which is most of tea products are facing this situation. The possible suggestions may discuss with suppliers, which it may try to control the cost of tea products, or it may consider changing the other products if it does not have good sales volume. A marketing team may also think about a promotion to have a bundle package with coffee and tea, maybe it can boost the sales volume and use the other profit to cover the loss.

13 Personal Conclusions

Thank you so much for my module leader Annalisa Occhipinti and module tutor Vishalkumar Thakor with their valuable support and let me know how to make my ICA better.

When I started to do the element one, I faced the problem about making relationship with models. I tried to search from internet and watched the video from Blackboard, but it cannot make me understand to apply in my case. The problem was still not yet solved until my tutor gave me some advices. As I did not have any experience about database, so it was a tough mission for me, and I tried to enquire my tutor. He was using many examples to let me understand and he did not mind following up my process all the time. He makes me learn a lot from his advice and make me feel more comfortable for the ICA.

On the other hand, finding dataset is also one of the most difficult tasks in ICA. The main problem is my first task to do Power BI report and it will be very challenge if I find a dataset's structure is not well organised and data volume is too large. I attempt to look for the dataset from Microsoft website, but most of datasets for education use only. It is clean and tidy, which means it will be too easy for us to make a report and it cannot fit on the ICA requirement.

14 Reference

Microsoft Corporation. 2020. Create and view decomposition tree visuals in Power BI. [Online]. [1 December 2020]. Available from: <https://docs.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-decomposition-tree>

Microsoft Corporation. 2020. Create key performance indicator (KPI) visualizations. [Online]. [5 December 2020]. Available from: <https://docs.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-kpi>

Science direct. 2020. Big Data Problem. [Online]. [5 December 2020]. Available from: <https://www.sciencedirect.com/topics/computer-science/big-data-problem>

April, R. 2013. Big Data Integration. [Online]. [10 December 2020]. Available from: <http://www.sciencedirect.com/science/article/pii/B9780123971678000212>

Microsoft Corporation. 2019. Understand star schema and the importance for Power BI. [Online]. [12 December 2020]. Available from: <https://docs.microsoft.com/en-us/power-bi/guidance/star-schema>

Microsoft corporation. 2019. Introduction to dashboards for Power BI designers. [Online]. [12 December 2020]. Available from: <https://docs.microsoft.com/en-us/power-bi/create-reports/service-dashboards>

Nicholas, G. 2020. 15 Best Power BI Dashboard Examples. [Online]. [15 December 2020]. Available from: <https://rigorousthemes.com/blog/best-power-bi-dashboard-examples/>