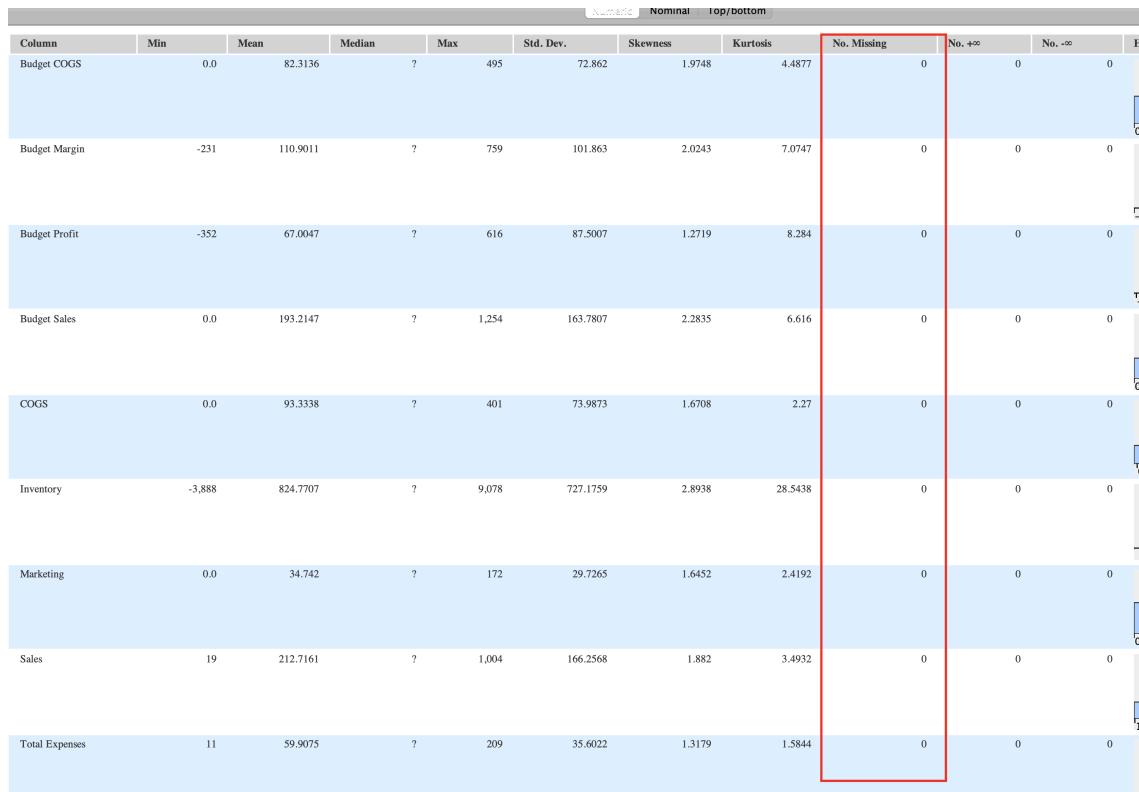


1 - Report Notes - EDA

Tags

Initial file was excel

- Open with excel → convert to CSV, in order to be able to read with KNIME easily
- Data quality assurance
 - No values missing (workflow-1)



- No duplicate values(workflow-2) so no need to remove duplicated

Table View - 0:10 - Interactive Table (local)

File	Edit	Hilite	Navigation	View
Row ID	Duplicate Removed	Original Number of Rows		
Number Rows	4248	?		
Number Columns	18	?		
Number Rows-	?	4248		
Number Columns-	?	18		

- Initial EDA statistics (workflow -1)

EDA : Exploratory Data Analysis

- Quick Notes
 - Number of Dimensions : 18
 - Indexes : 9
 - Measures : 9
 - General Description of the data. 'Every store per day per product vs COGS Budget'
- Important facts about the data
 - Country is always 'United Kingdom' . So company operates solely on UK
 - Beans = coffee product always
 - Leaves = tea products always
 - Coffees always have Decaf and Regular
 - Tea is not the same.
 - Herbal tea : Decaf always
 - Regular Tea : Regular Brew type always
 - Fields Hierarchies
 - Date Info (Date)

- Location Info(Country,State,City)
- Product Type(Product,Product Type,Product Line,Type)
- Total Product Sold (COGS,Sales,Total Expenses,Marketing)
- Expected Product Sold(Budget COGS,Budget Sales,Budget Margin,Budget Profit)
- Other - (Market Size,Inventory)
 - IDEA : Market Size as drillable category
 - Inventory as calculated field (low,high) and then drillable category(?)
- Transformations Suggestion(the tableau file contains these changes)
 - No Duplicates or Values Missing, so no prepossessing required
 - Create Hierarchies
 - Country,State,City | Location
 - Product,Product Type,Type,Product Line | Product
 - Delete/Hide of 'Country' field. Constant and not useless
 - Rename of the following Fields for more clarity
 - Product → Product Name
 - Type = [Regular, Decaf] → Brew Type (something more descriptive)
 - Total Expenses → Additional Expenses(See description, that's more descriptive)
 - Calculated Fields
 - [Given from coursework]Margin = Sales – COGS
 - [Given from coursework]Profits = Sales – COGS – Total Expenses(Additional Expenses)
 - [Ours] Budget additional expenses = Margin - Profit

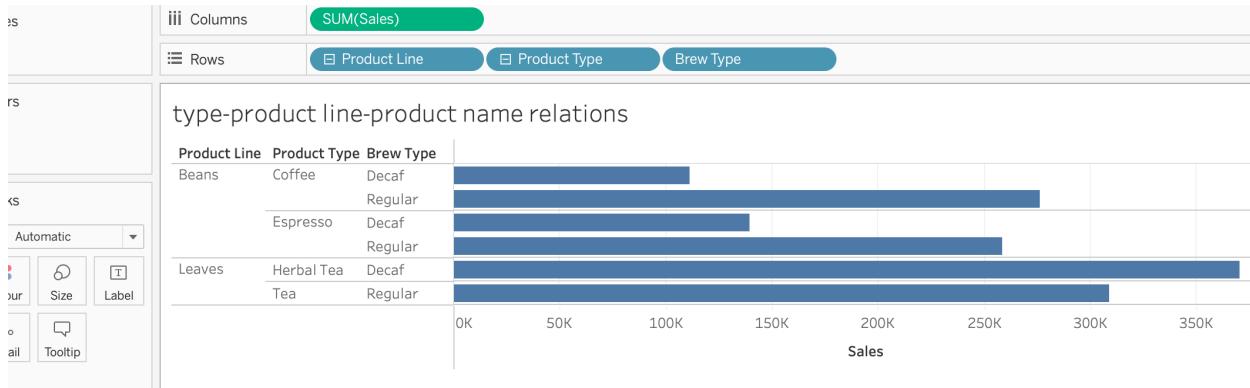
Fields and Descriptions

Aa Name	Tags	Description	Reference
<u>Date</u>	Useful	Date in dd/mm/yyyy	
<u>Country</u>	Not-Useful	Categorical variable with only one category 'United Kingdom' . Always constant, we should ignore and justify . On the part on 'understanding the company' we should mention this	Go to coursework/tableau/EDA.twb file, on country-constant Sheet for proof
<u>State</u>	Useful	England,Scotland,Wales	
<u>City</u>	Useful	Verification if the data is correct, i checked if every single city mentioned are indeed, in their respective state - check workflow-3	Open knime with main folder at coursework/ . Open workflow-3
<u>Market Size</u>	Kinda-Useful	Categorical , Small and Major only	proof : Open knime with main folder at coursework/ Open workflow-4
<u>Product</u>	Useful	Categorical, see next paragraph for a set of facts for this variable in conjuction with others	proof : Open knime with main folder at coursework/ Open workflow-5
<u>Product line</u>	Useful	Categorical, Leaf and Beans,see next paragraph for a set of facts for this variable in conjuction with others	proof : Open knime with main folder at coursework/ Open workflow-6
<u>Product Type</u>	Useful	Categorical, Coffee Espresso Herbal Tea Tea,see next paragraph for a set of facts for this variable in conjuction with others	proof : Open knime with main folder at coursework/ Open workflow-7
<u>Type</u>	Kinda-Useful	Categorical, Decaf or Regular. ,see next paragraph for a set of facts for this variable in conjuction with others	proof : Open knime with main folder at coursework/ Open workflow-8

Aa Name	Tags	Description	Reference
<u>Budget COGS</u>	Useful	expected costs for producing the products to be sold at given date	
<u>Budget Sales</u>	Useful	expected amount of sales based on the sales plan	
<u>Budget Margin</u>	Redundant Useful	Budget Sales - Budget COGS = Budget Margin , is just a calculated field, you can mention that on report	proof : Open knime with main folder at coursework/ Open workflow-9
<u>Budget Profit</u>	Useful	Budgeted financial gain. It is the difference between the amount earned and the amount spent in buying, operating, or producing something : Hence Margin - Profit = Budget Other expenses (maybe marketing?)	
<u>COGS</u>	Useful	Total costs for producing the products to be sold at given date	
<u>Sales</u>	Useful	Total amount of sales	
<u>Total Expenses</u>	Useful	Additional expenses to complete the sale	
<u>Marketing</u>	Useful	Amount dedicated to activities a company undertakes to promote the buying or selling of a product or service. Seems Marketing figures are not counted towards the 'margin'	proof : Open knime with main folder at coursework/ Open workflow-10
<u>Inventory</u>	Useful	Goods and materials that a business holds for the ultimate goal of resale	

ProductXXX class of variables

proof of the following statements can be found on /tableau/EDA.twb file, on 'type-product line-product name relations' Sheet



Beans = coffee product always

Leaves = tea products always

Coffees always have Decaf and Regular

Tea is not the same.

Herbal tea : Decaf always

Regular Tea : Regular Brew type always

Another miscellaneous stuff that i found

not every city sells all the products possible, please open /tableu/EDA.twb file, on 'type-product line-product name relations' Sheet

Product Line	Product Type	Brew Type	City																
			Aberdeen	Barry	Basingst..	Bath	Birming..	Bristol	Briton Ferry	Cardiff	Clydach	Cumbern..	Dundee	Edinburgh	Glasgow	Leeds	Liverpool	London	
Beans	Coffee	Decaf	11,122	5,016	5,586	3,171	6,796	13,293	1,870	6,463	11,175	23,710	11,067	32,513	4,740	7,622	6,674		
		Regular	12,369	18,413	4,696	5,723	5,786	22,609	18,918	5,179	7,622	10,190	10,190	23,710	11,067	32,513	10,843	11,530	37,370
	Espresso	Decaf	5,067	7,202	5,020	3,978	10,190	8,280	21,399	1,908	4,118	8,723	8,723	21,399	1,908	4,118	13,923	4,830	26,045
		Regular	6,255	8,280	12,478	7,959	13,667	18,819	26,045	2,254	4,830	10,522	18,940	7,959	28,956	10,735	10,692	35,111	
Leaves	Herbal Tea	Decaf	14,085	24,719	14,196	5,179	16,229	10,287	29,535	39,338	10,580	17,167	5,188	3,532	33,873	8,997	16,653	40,527	
		Tea	7,704	16,323					15,482	47,170	10,475	9,076	5,663	4,075	31,019	23,857	11,937	27,099	