2 - Report Notes - Dummy Data

Tags

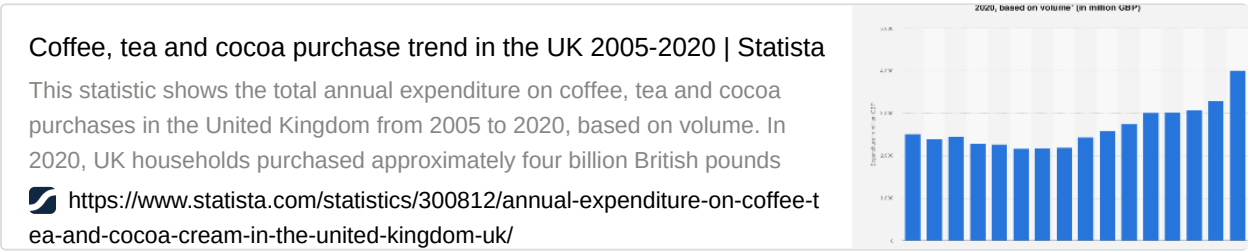
Dummy Data Generation

please have a look at the workflow-10

The Idea

Use The following datasets

- **Annual expenditure on coffee, tea and cocoa in the United Kingdom from 2005 to 2020, based on volume***



They show a slight increase during the pandemic (this is not something unexpected. The food retail industry were one of the few industries were remained largely unaffected by the pandemic).

We will take the the average difference between 2019-2020 of those two datasets, and apply the increase into sales, budgets,

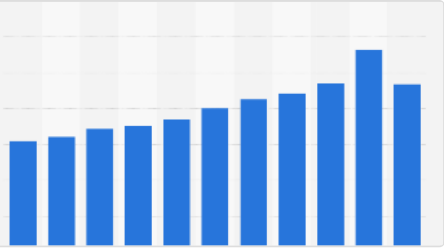
Marketing

- **Advertising expenditure in the United Kingdom from 2010 to 2020**

Advertising spending in the UK 2020 | Statista

Advertising expenditure in the United Kingdom in 2020 amounted to 23.46 billion British pounds, down by 7.2 percent compared to previous year. Ad spend in the country is expected to grow by 15.2 percent in

https://www.statista.com/statistics/262754/advertising-revenue-in-the-uk/



Drop in marketing!!

Inventory

Change in Inventories: Retail: £M: CP: SA

Change in Inventories: Retail: £M: CP: SA

economy/grossdomesticproductgdp/timeseries/faha/pn2

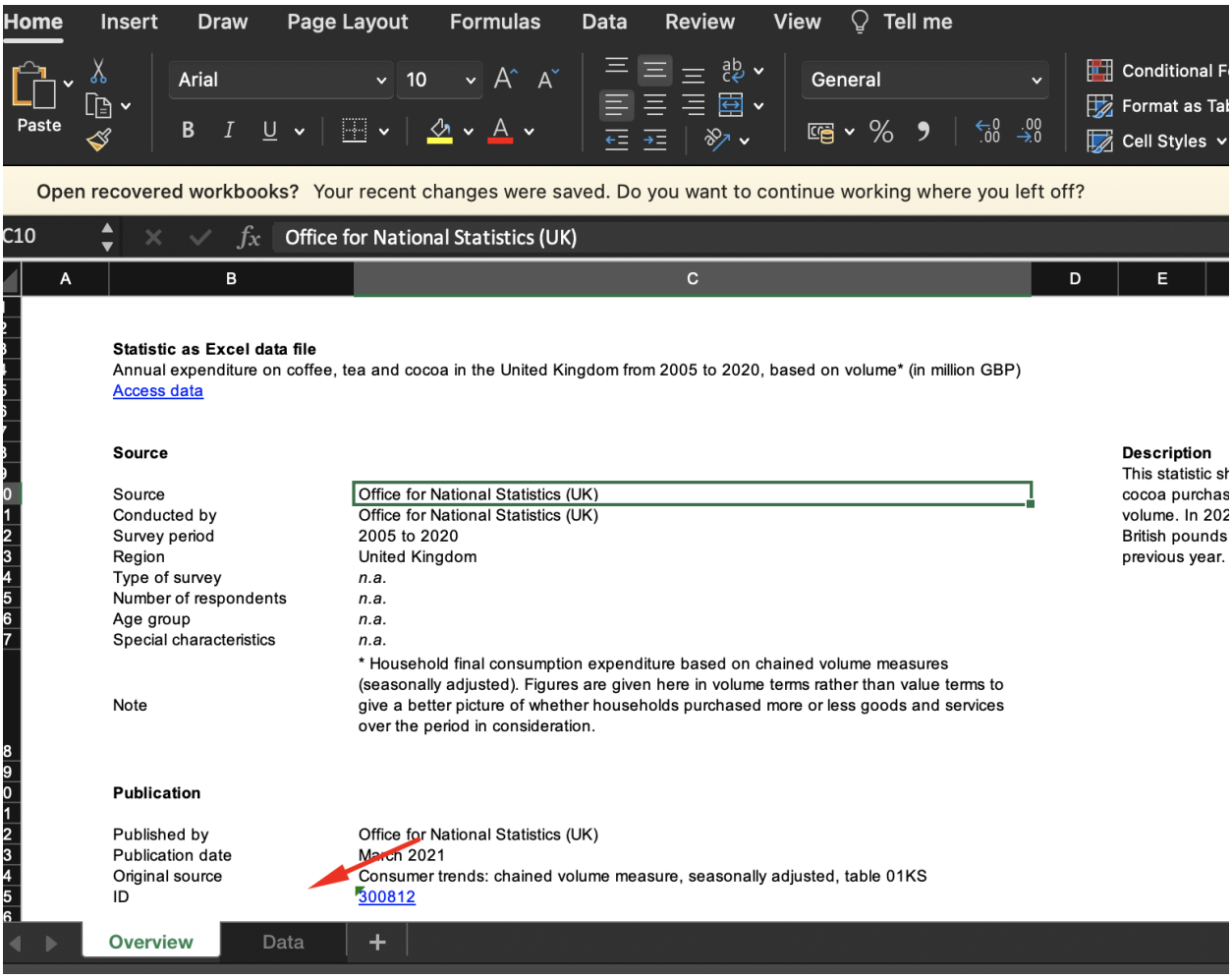
https://www.ons.gov.uk/economy/grossdomesticproductgdp/timeseries/faha/pn2

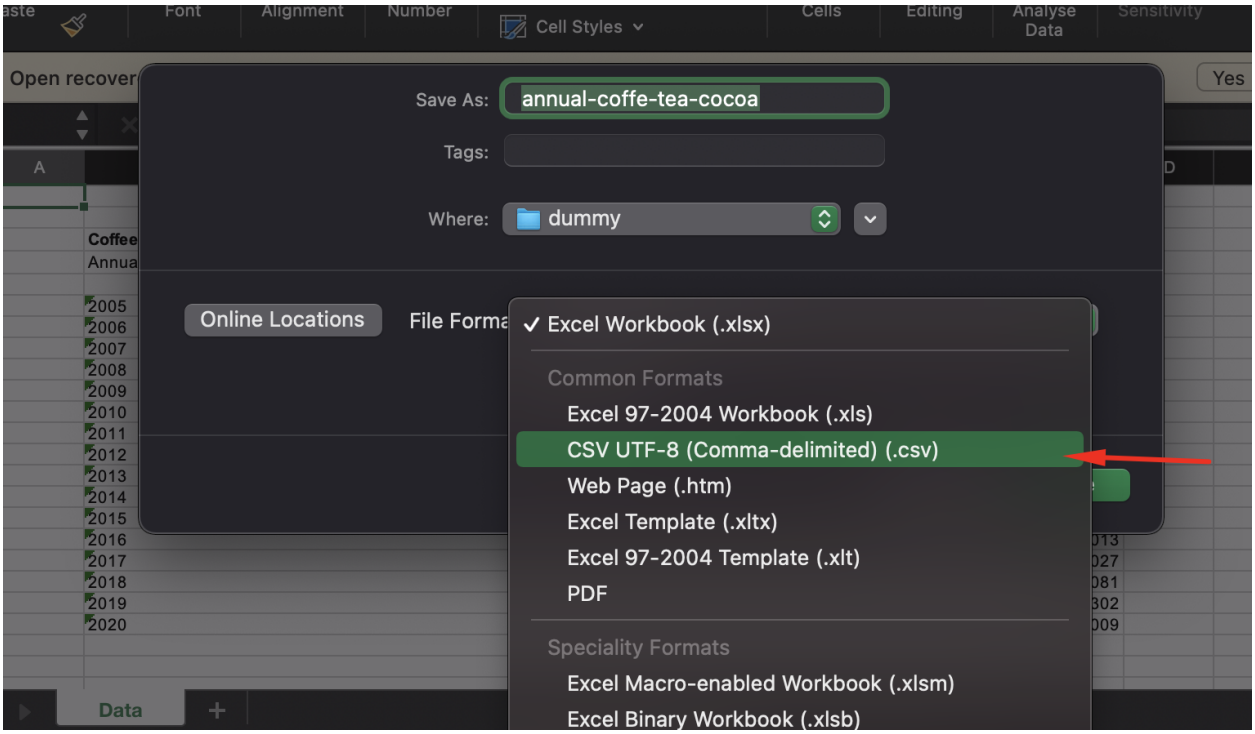
The process

Step 1 : Convert .xls files to .csv using Microsoft Excel

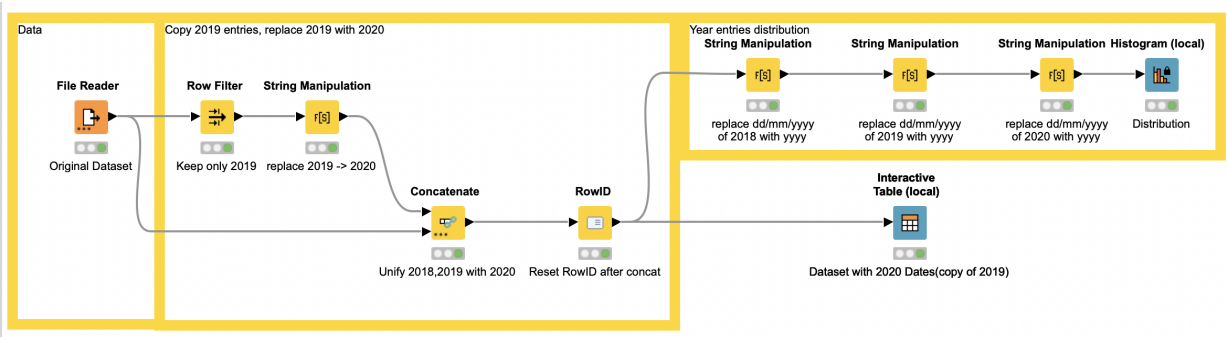
The following Datasets are of type .xls (Excel files) . So i opened them with excel and save them as .csv to be opened by KNIME(also, i deleted the first sheet from every one of them, as this is the introductory sheet with various irrelevant info (sheet name, publisher etc etc)

Example


Home Insert Draw Page Layout Formulas Data Review View Tell me
Paste
Arial
10
General
Conditional F
Format as Tab
Cell Styles
Open recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?
C10
Office for National Statistics (UK)
A B C D E
Statistic as Excel data file
Annual expenditure on coffee, tea and cocoa in the United Kingdom from 2005 to 2020, based on volume* (in million GBP)
Access data
Source
Source Office for National Statistics (UK)
Conducted by Office for National Statistics (UK)
Survey period 2005 to 2020
Region United Kingdom
Type of survey n.a.
Number of respondents n.a.
Age group n.a.
Special characteristics n.a.
Note * Household final consumption expenditure based on chained volume measures (seasonally adjusted). Figures are given here in volume terms rather than value terms to give a better picture of whether households purchased more or less goods and services over the period in consideration.
Publication
Published by Office for National Statistics (UK)
Publication date March 2021
Original source Consumer trends: chained volume measure, seasonally adjusted, table 01KS
ID 300812
Description
This statistic sh
cocoa purchas
volume. In 202
British pounds
previous year.
Overview
Data



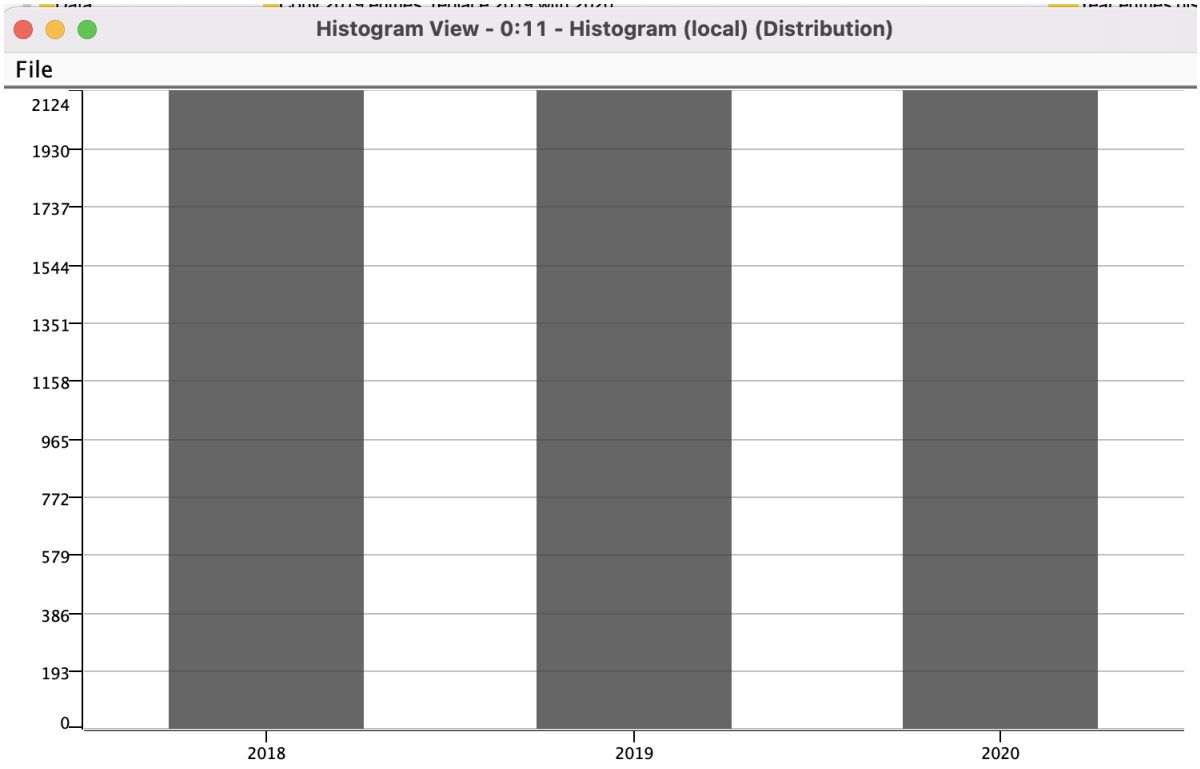
Step 2 : Create entries for 2020 (copy of 2019) (workflow-dummy)



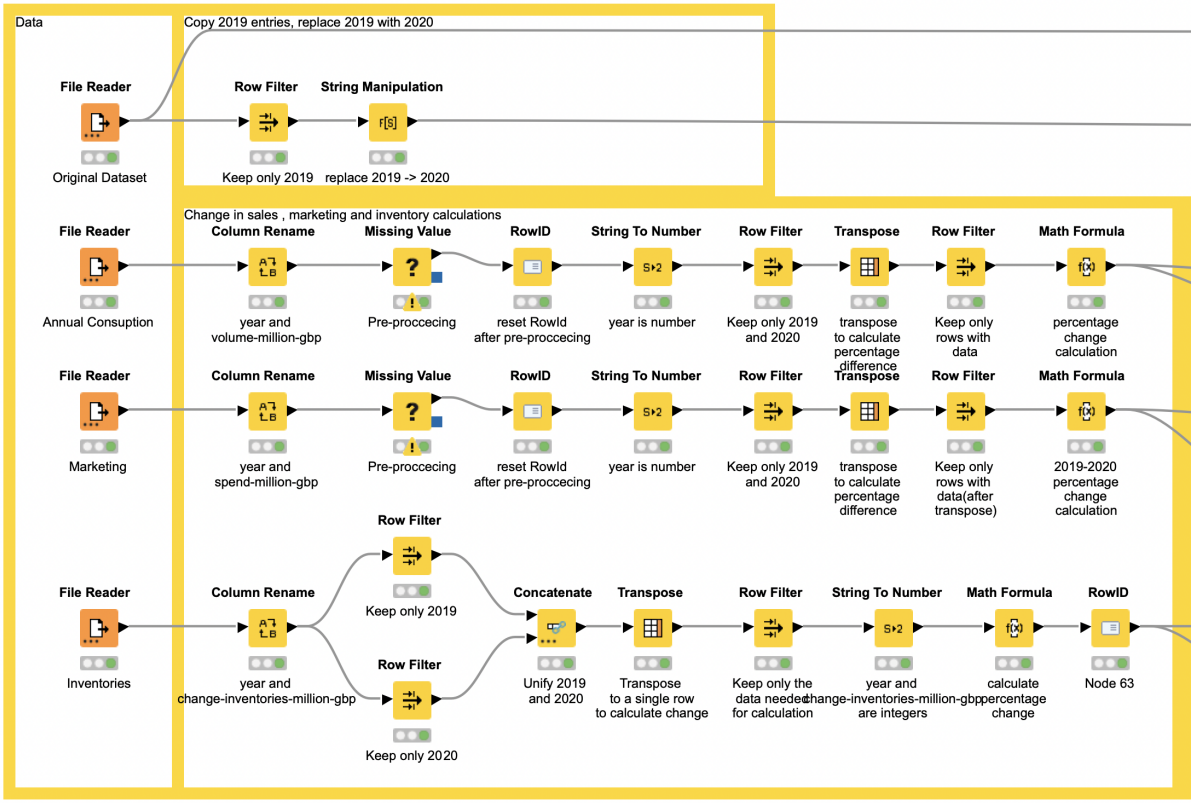
Logic is as follows

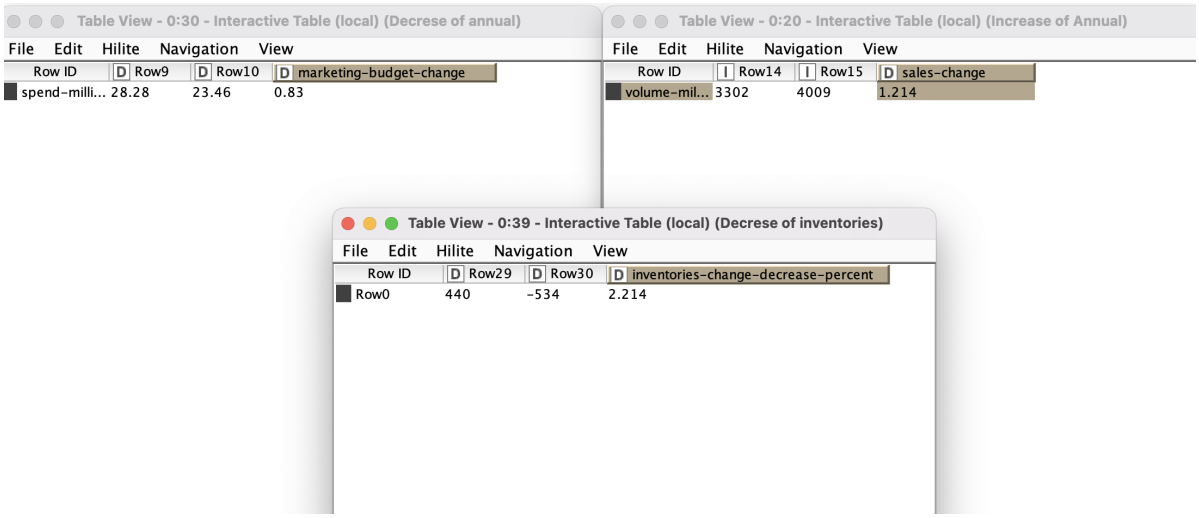
- Take all datarows
- filter by date (year == 2019)
- map date (date.year = 2020)
- concat original with the copied version

- Now we have , a dataset with 2018, 2019 and 2020 which 2020 is just a copy over 2019
- The distribution of rows is given below, no entry missing, thats good data (we can mention it)



Step 3 : Load aforementioned .csv files onto KNIME and make nessesary transformations to calculate annual percentage change in consumption, marketing and inventories



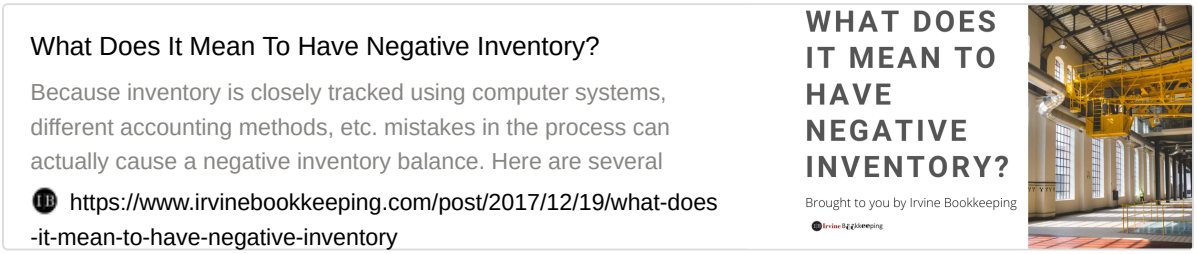


All of those are basic transformations, and the only logic applied is the

> Divide 2019 number with 2020 number

open workflow-dummy, i have left numerous comments in the nodes

The Inventories calculation is slighly more complicated, as it allows for negative values(see why negative inventories make sence here →



The calculation i used for calculating the decrease is as follows

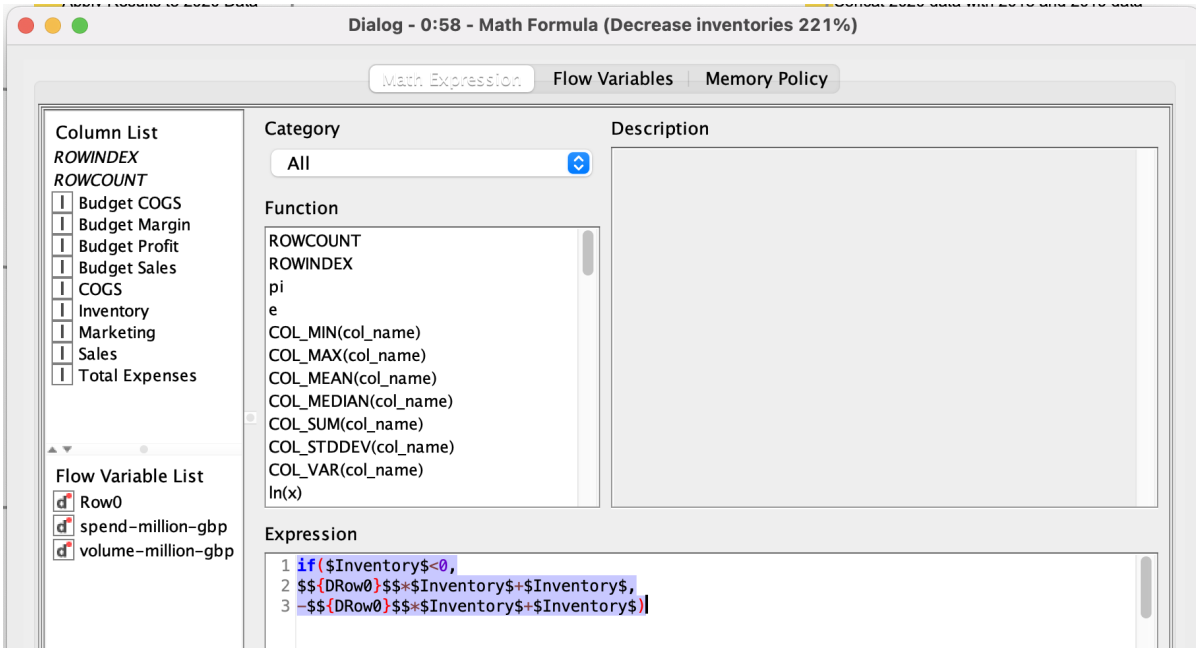
$$\frac{x - x_{ref}}{|x_{ref}|} \cdot 100 = -221\%$$

$$-\frac{x - x_{ref}}{|x_{ref}|} \cdot 1 = 2.21$$

$$-x - x_{ref} \cdot 1 = 2.21 \cdot |x_{ref}|$$

$$x = -2.21 \cdot |x_{ref}| + x_{ref}$$

$$x = \begin{cases} -2.21 \cdot x_{ref} + x_{ref}, & \text{if } x_{ref} \geq 0 \\ 2.21 \cdot x_{ref} + x_{ref}, & \text{otherwise} \end{cases}$$



(tested with random sampling and this tool https://www.omnicalculator.com/math/percentage-change)

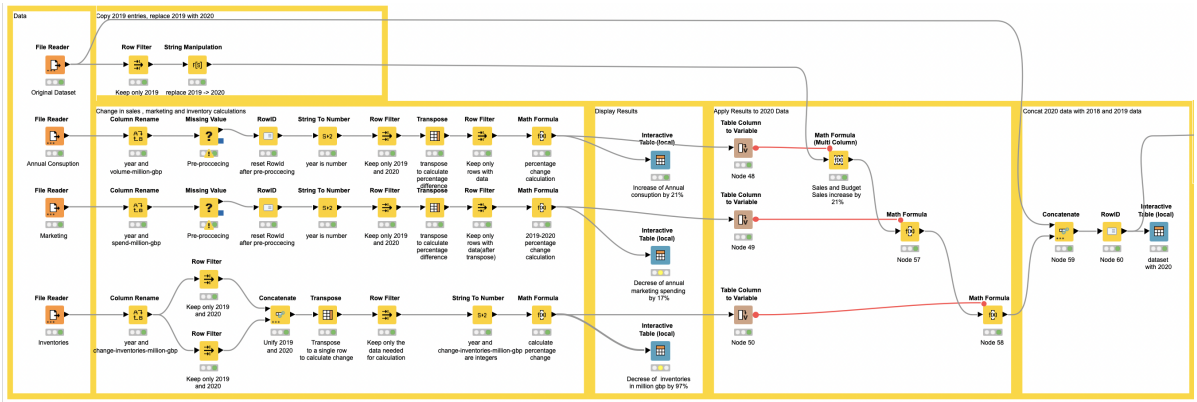
(please include the calculations if you like on the final report, search ‘how to add latex formulas on google docs’ και θα σου δωσω εγω τις φορμουλες σε ενα .txt

What we will implement?

- Sales and Budget sales will be increased by 21% (source → https://www.statista.com/statistics/300812/annual-expenditure-on-coffee-tea-and-cocoa-cream-in-the-united-kingdom-uk/)
- Marketing will be decreased by 17% (source → https://www.statista.com/statistics/262754/advertising-revenue-in-the-uk/)
- Inventory will be decreased by 97% (source → https://www.ons.gov.uk/economy/grossdomesticproductgdp/timeseries/fbid/ukea)

> Warning → all those datasets are for UK region, that choice is because of the fact that lupita coffe operates sorely on UK (we learned that on EDA proccess)

Step 4 :Apply percentages on the relevant variables on 2020



And voila, we have dummy data!

> Warning : Please write on the assumptions, that 2020 dummy data do not account for the 2020's inflation, thats for simplicity, : Inflation 2020 is on record high 7% (Source ons : https://www.pewresearch.org/fact-tank/2022/01/24/as-inflation-soars-a-look-at-whats-inside-the-consumer-price-index/