Vadaga Venkatesh Prasad

Boston university  COMMONWEALTH AVE, BOSTON

Sales forecast

Table of Contents

[Design 1](#_Toc57995741)

[Dataset 2](#_Toc57995742)

[Dimensions 2](#_Toc57995743)

[SCD 2](#_Toc57995744)

[Hierarchies 3](#_Toc57995745)

[Extract Transform Reload 3](#_Toc57995746)

[Extract 3](#_Toc57995747)

[Transform 3](#_Toc57995748)

load [7](#_Toc57995749)

[Date time dimensional table 7](#_Toc57995750)

[Insertion of Fact table 9](#_Toc57995751)

[Visualizations 9](#_Toc57995752)

[Assumptions 12](#_Toc57995754)

[Loaded Database images 12](#_Toc57995755)

# Figures Index

[Figure 1\_dataset image 2](#_Toc57996445)

[Figure 2 ERD using Lucid chart 2](#_Toc57996446)

[Figure 3 Hierarchy 3](#_Toc57996447)

[Figure 4 Extract 3](#_Toc57996448)

[Figure 5 transform 4](#_Toc57996449)

[Figure 6 Connect DB 4](#_Toc57996450)

[Figure 7 Reload into database 6](#_Toc57996451)

[Figure 8 Continent sales tableau 10](#_Toc57996452)

[Figure 9 item name and sales channel 10](#_Toc57996453)

[Figure 10 day of week filter 11](#_Toc57996454)

[Figure 11 drilling up in tableau 11](#_Toc57996455)

# Design

The structure of is a star schema and all the dimensions are connected to a signal fact table. The processing time is much lower in star schema. The data in dimension table also has a role to play larger the data higher the query execution time. We can extend dimensions from item dimension to business unit dimension to have snowflake schema.

# Dataset

The dataset has around 500,000 rows.

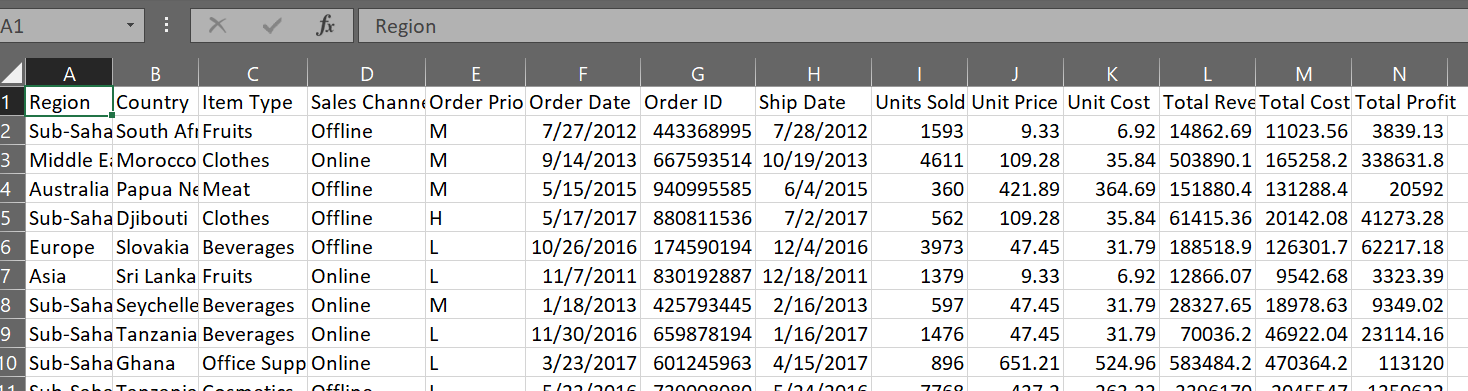


Figure 1\_dataset image

# Dimensions

The dimensions considered here are having a narrow scope growing. This is due to having additive measures more in the data. The dimensions in the data area as follows

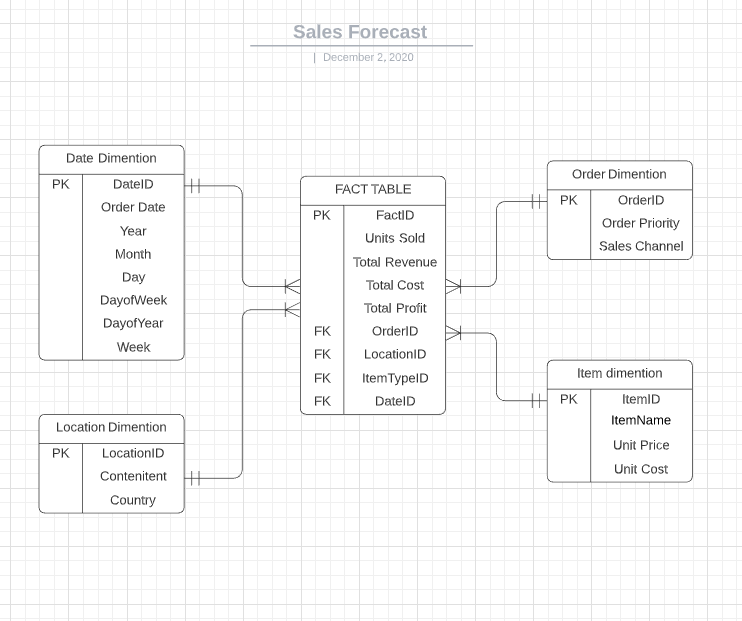


Figure 2 ERD using Lucid chart

# SCD

My order dimension is a fact less fact table. The slowly changing dimension here is Item name. Which could be associated to a business model change of the company to call products having prefixes like fresh, caramelized. This is a Type-2 SCD due to the previous records which are related would have a different reference.

# Hierarchies

The dimensions which contain hierarchy is date and region dimension. This helps to drill down the from overview to detailed location and patterns of consecutive dates, weekdays.

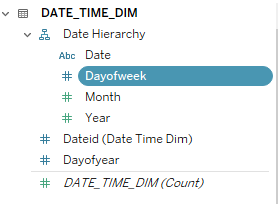


Figure 3 Hierarchy

# Extract Transform Reload

## Extract

Read the csv file and see the datatype of the

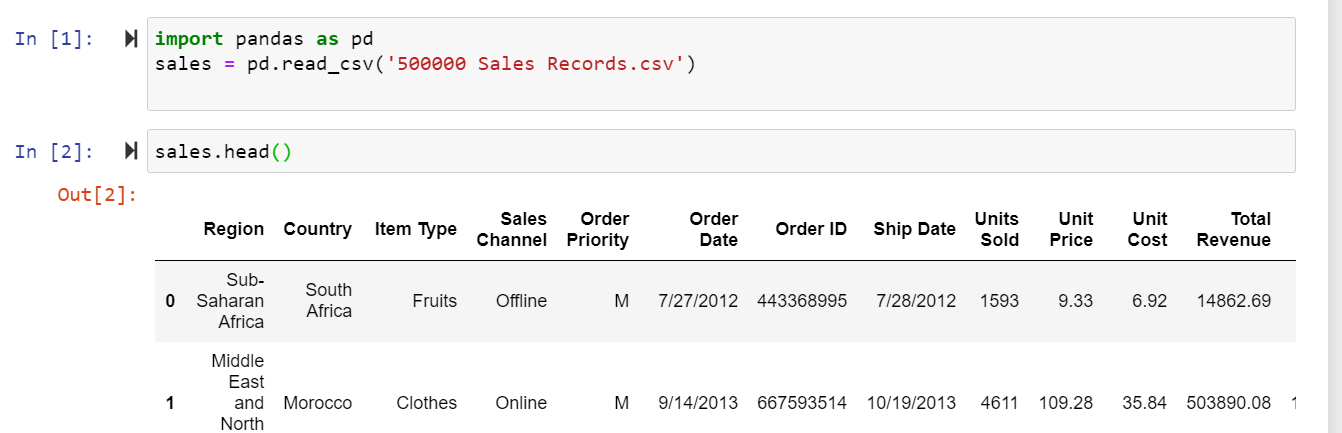


Figure 4 Extract

## Transform

Take the dimensions are related fields into data frames. Compute the required fields and populate the main data frame.

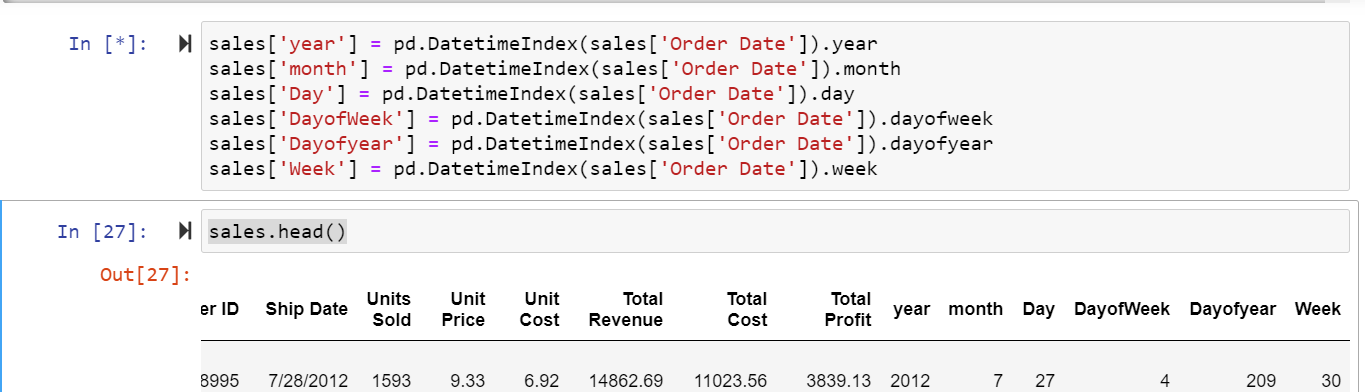
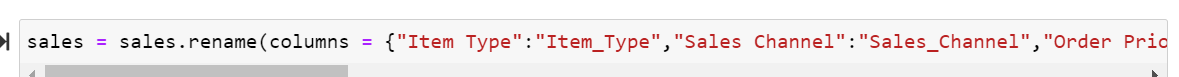


Figure 5 transform

Rename the column names such that it does have any spaces.



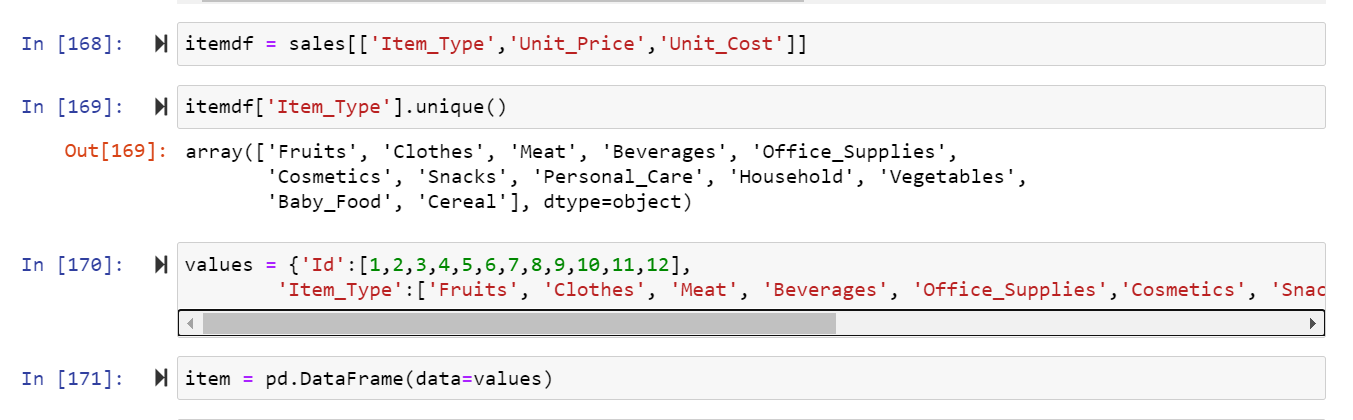
Connect the SQL Server with the database and have a cursor connected



Figure 6 Connect DB

The white spaces are unacceptable in SQL server during the insertion of the data. Replace them with underscore.

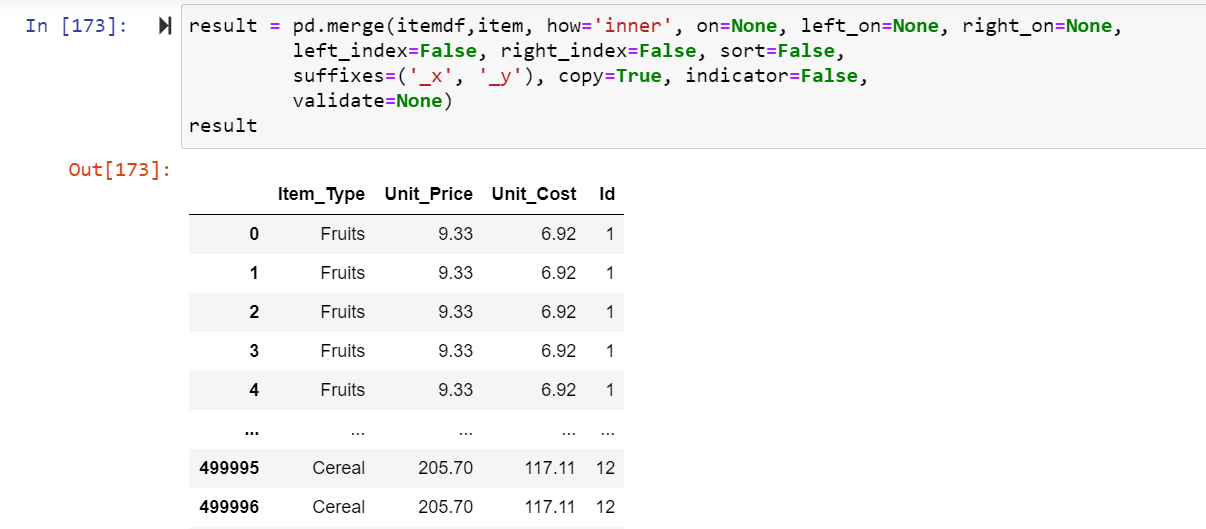
Create another data frame. Identify unique elements and assign them different values to each and load in another data frame



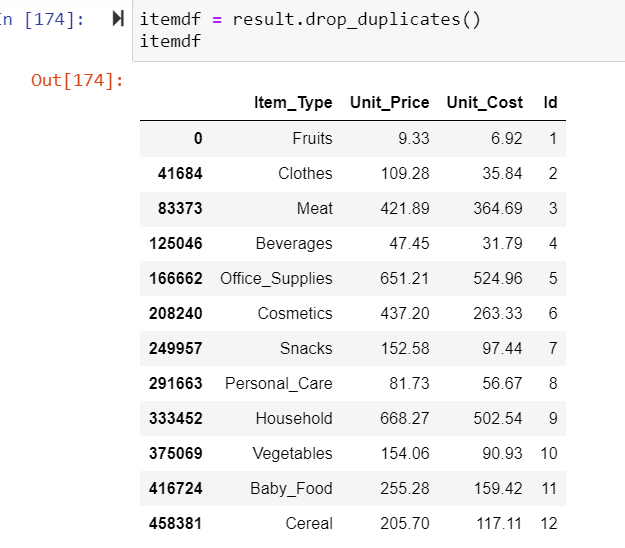
Item data frame



Merge the data frame item and itemdf which is stored in result



Drop the duplicates.



Insert into Item Type dimension table and commit

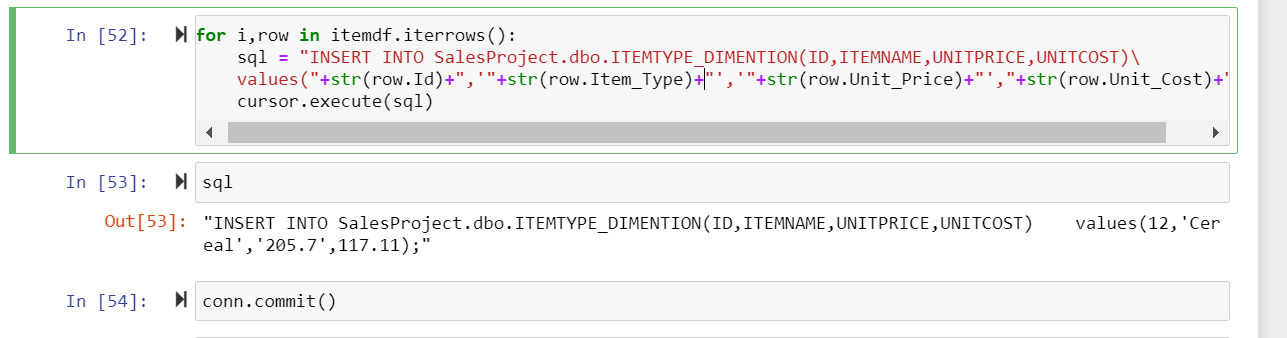
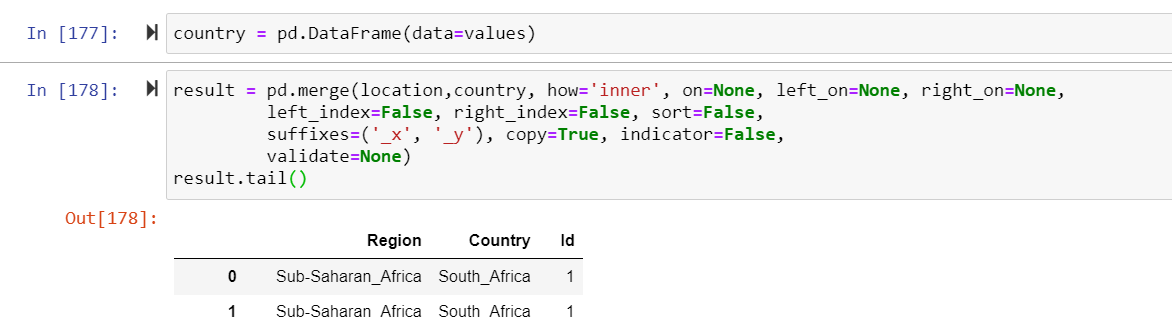


Figure 7 Reload into database

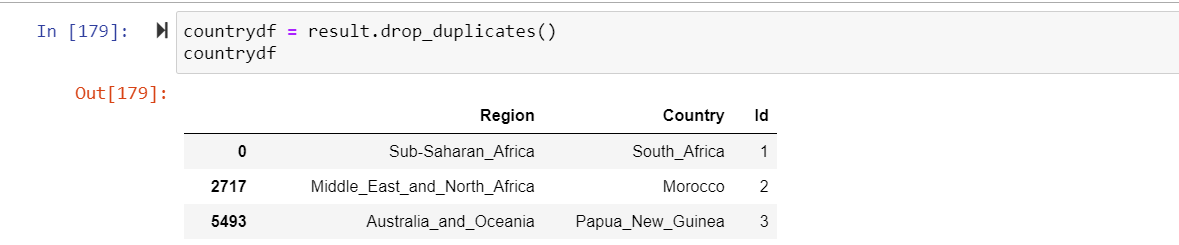
Insertion of Location dimension.



Remove white spaces, identify how many unique values and insert into a data frame ‘country’. Merge the data frames with inner.



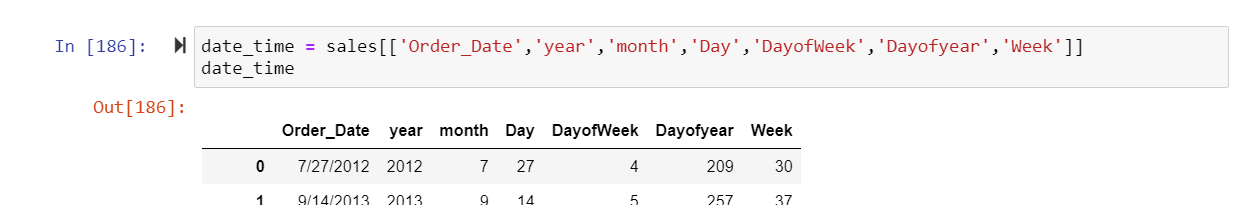
Drop duplicates and get 185 unique country names.



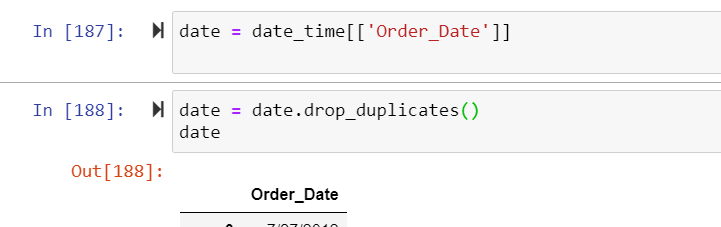
Insertion into database

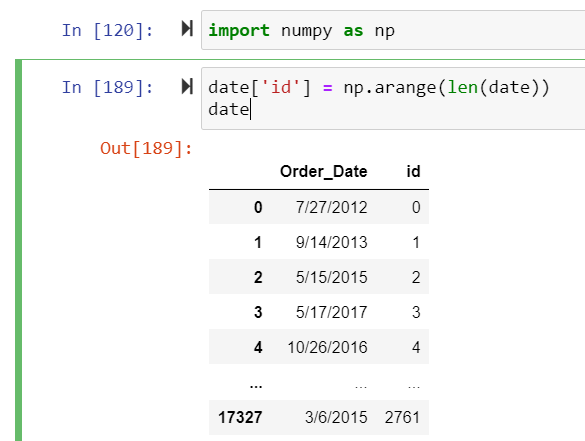
### 

### Date time dimensional table

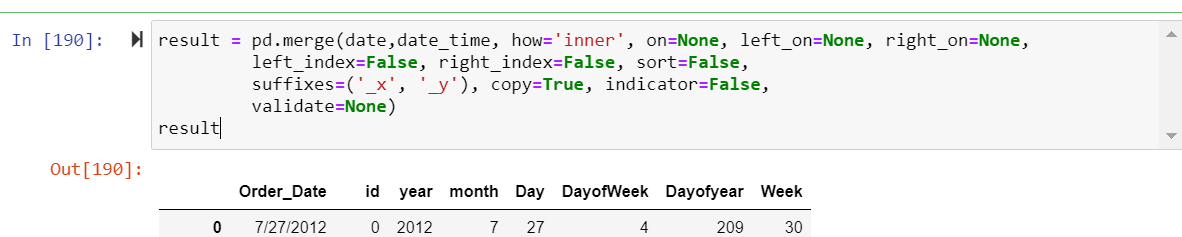


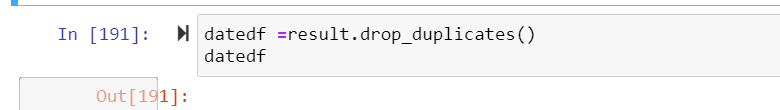
Take the order date and assign unique values to each date



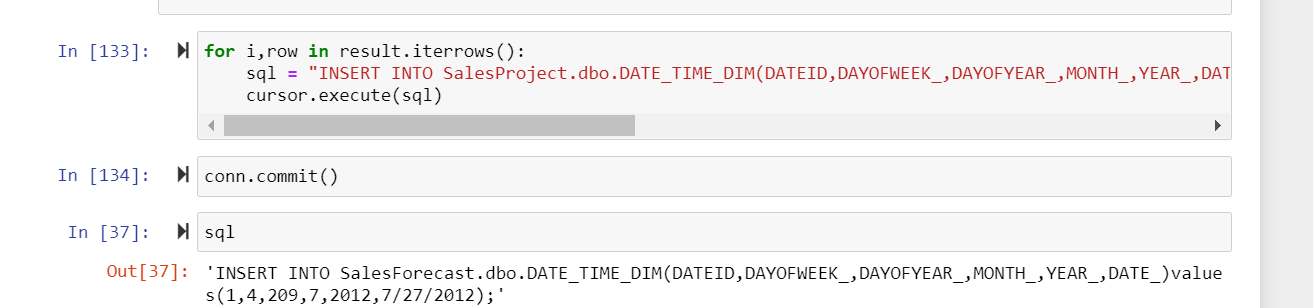


Merge and drop duplicates

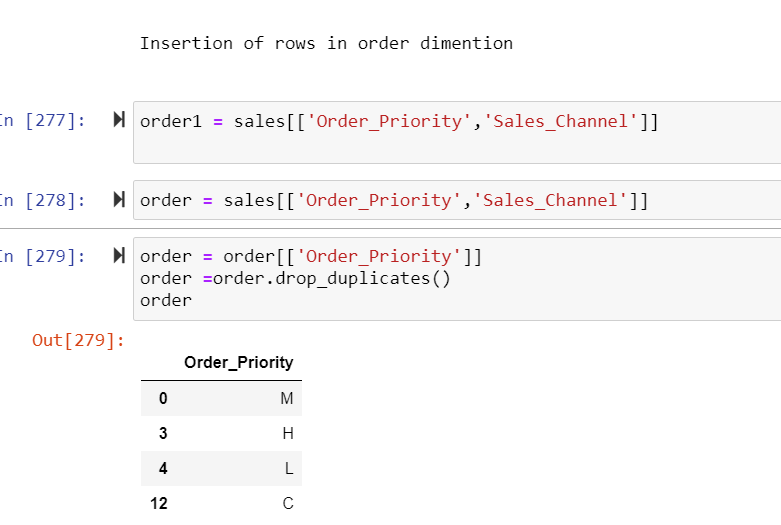




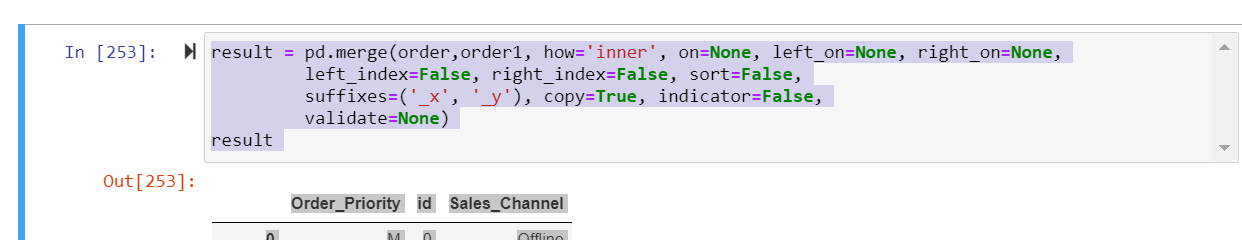
Insertion into database and commit

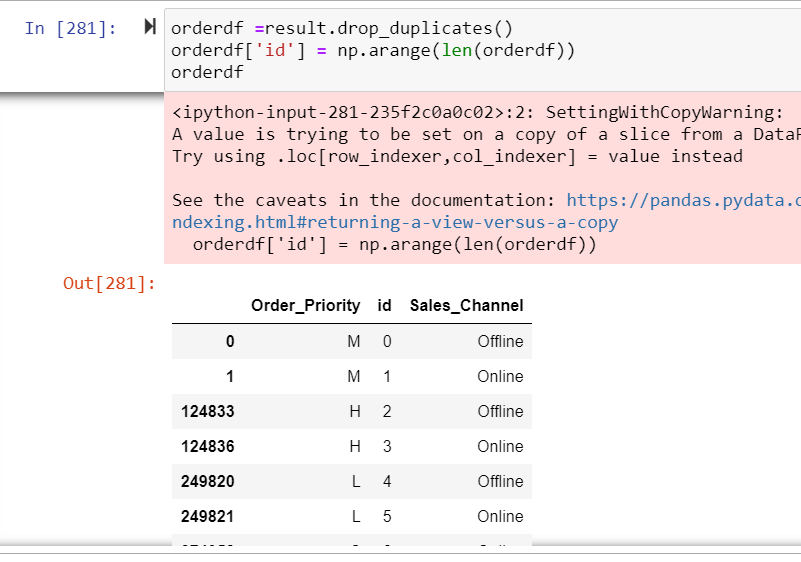


Order dimension data frame creation



Merge and drop duplicates

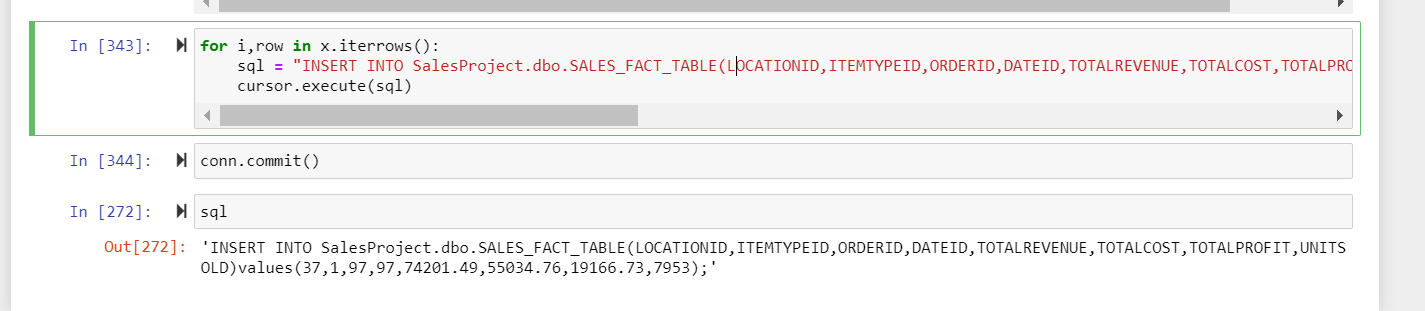
  
combination of order priority and sales channel to assign a unique value in the data frame



### Insertion of Fact table



Merge all and rename all continuingly and then insert into DB



# Visualizations

# Chart, line chart Description automatically generated

Figure 8 Continent sales

# 

The above tableau shows the continents sales from 2010 to 2017. The most total cost by the company is sub – Saharan Africa. The lowest is in north America. Asia comes at 9 billion.

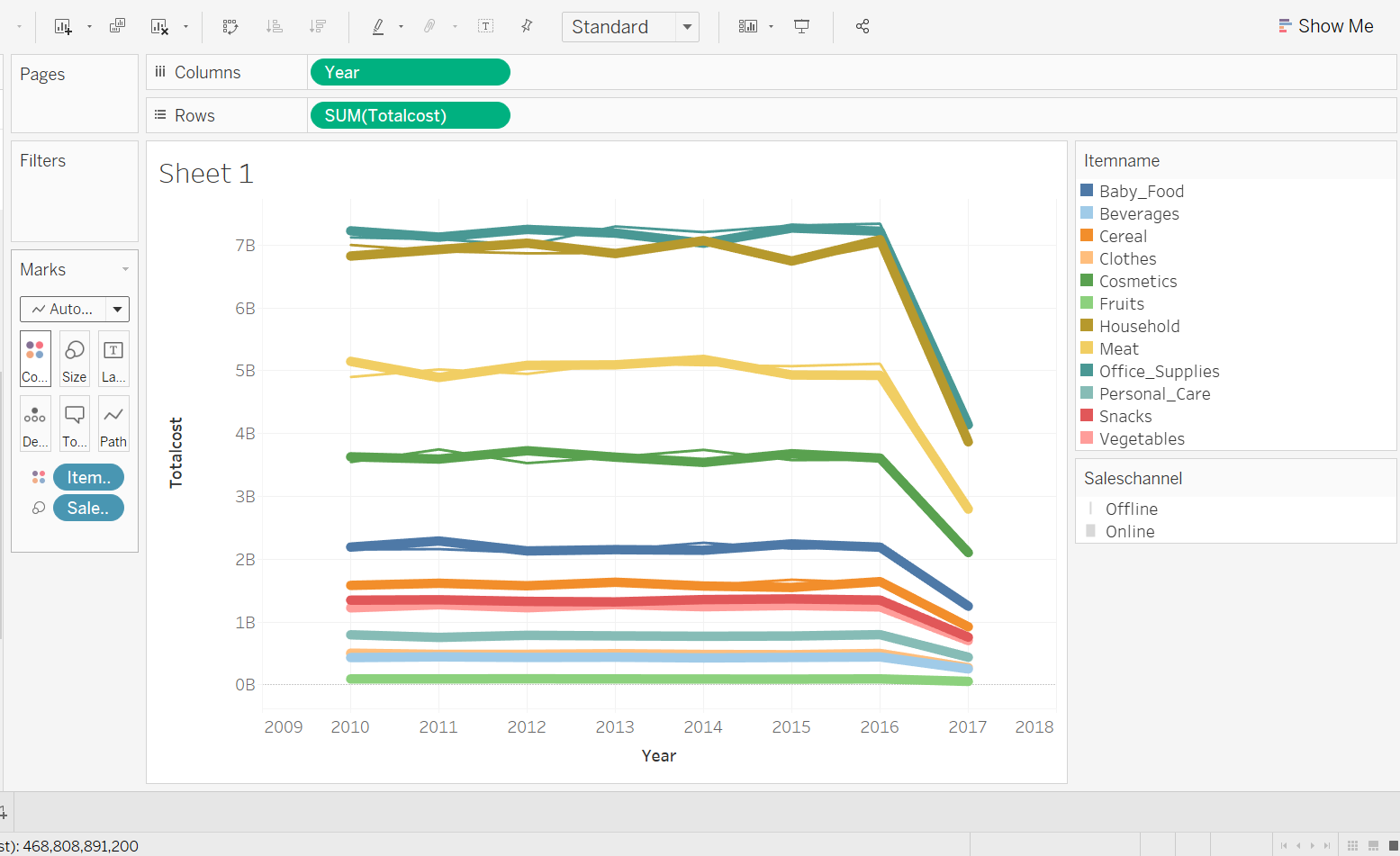


Figure 9 item name and sales channel

Here online and offline sales are similar to each other, office supplies and households are having high total cost.

Which day in the week has a more profit and total cost.

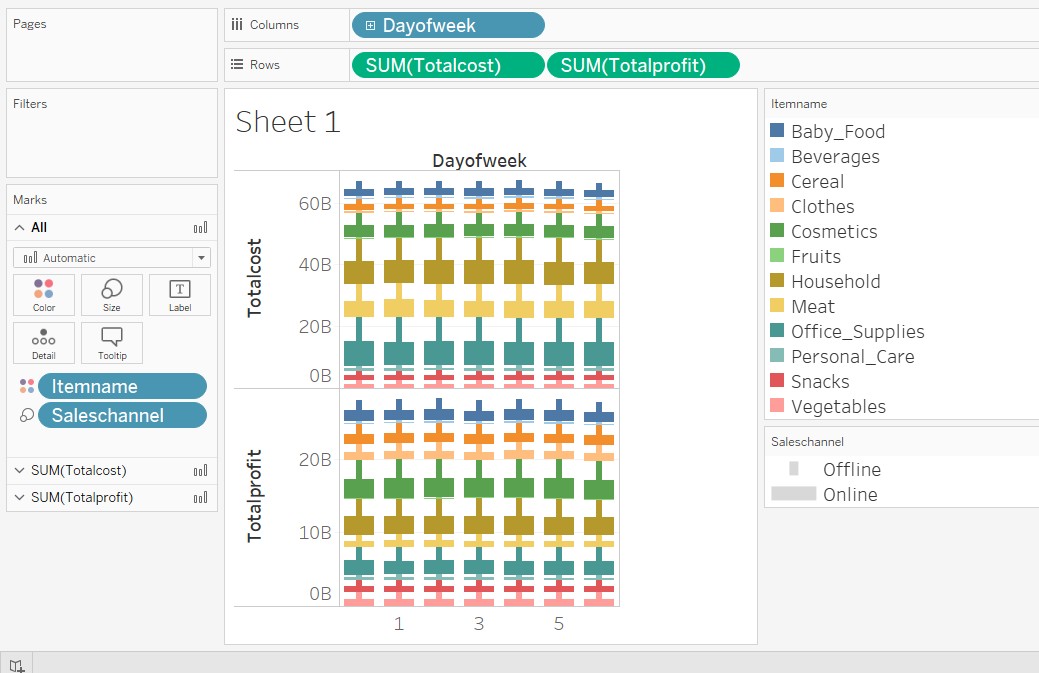


Figure 10 day of week

Drilling up from month to year

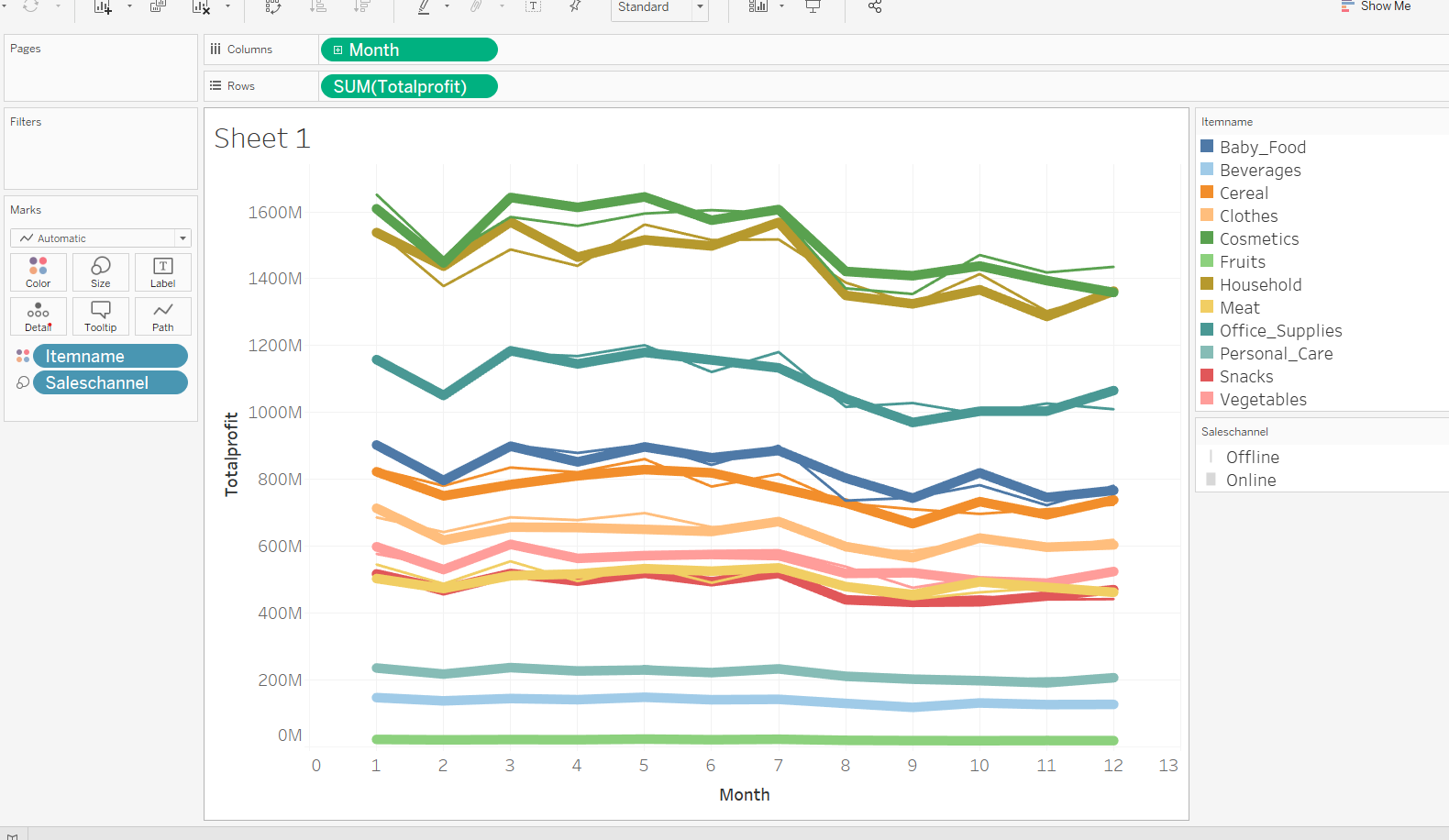
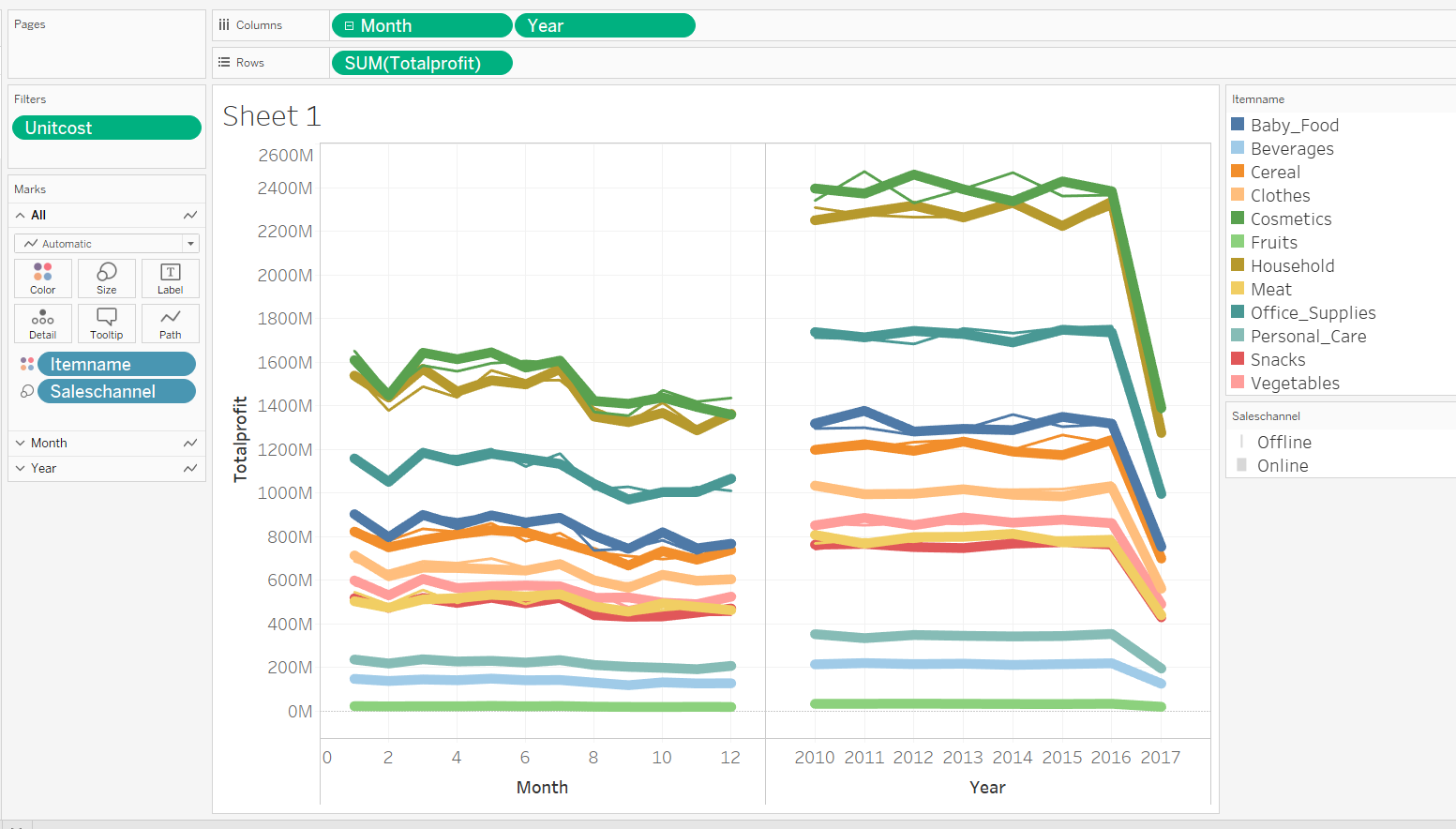


Figure 11 drilling up in tableau

Drilling UP



# Assumptions

The sales of cosmetics are high and the amount that the company pays for manufacturing is low, which fetches more profit. The household and office supplies are high as well. But the total cost that company expends on items like house and office supplies are higher. That means the risk associated to cosmetics is less when compared to office supplies and household products.

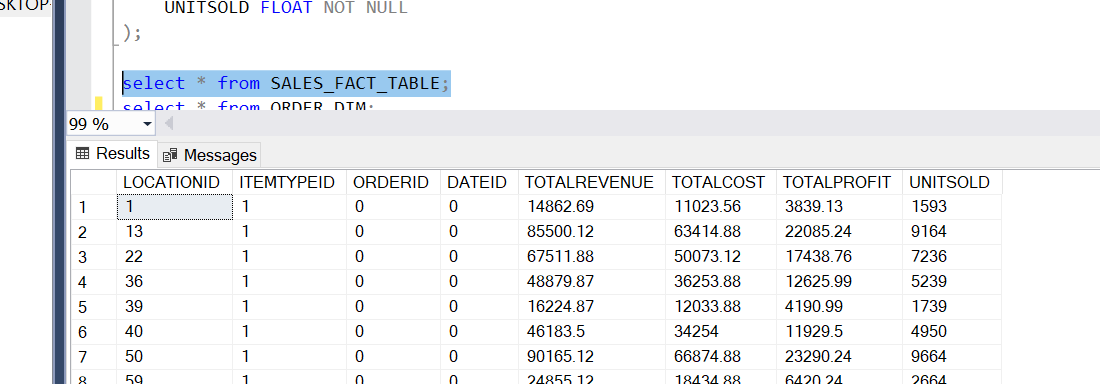
The company has a large business which is not operated in north America. The turnover for the company has its roots in Europe and Sub Saharan Africa. Fruits are the lowest profit for each year.

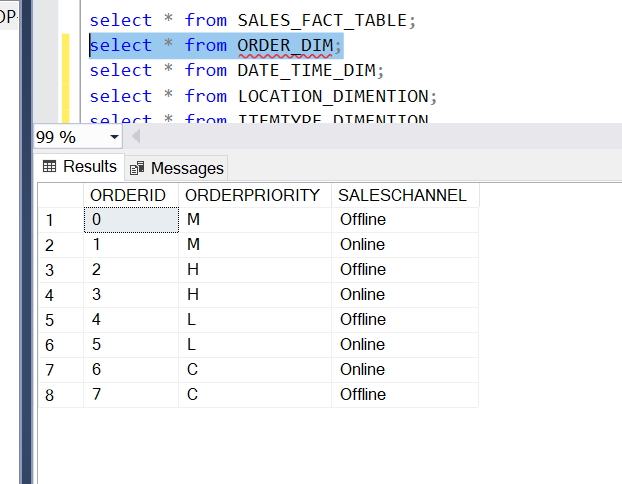
In 2016 the demand of cosmetics has reduced and in 2017 it is at all-time low.

The online and offline channels are having similar patterns which go together. Both having down word trend.

# 

# Loaded Database images





# References

# <https://stackoverflow.com/questions/tagged/sql-server>

<https://docs.python.org/3/reference/expressions.html#arithmetic-conversions>

<https://www.w3schools.com/sql/sql_ref_keywords.asp>

# 