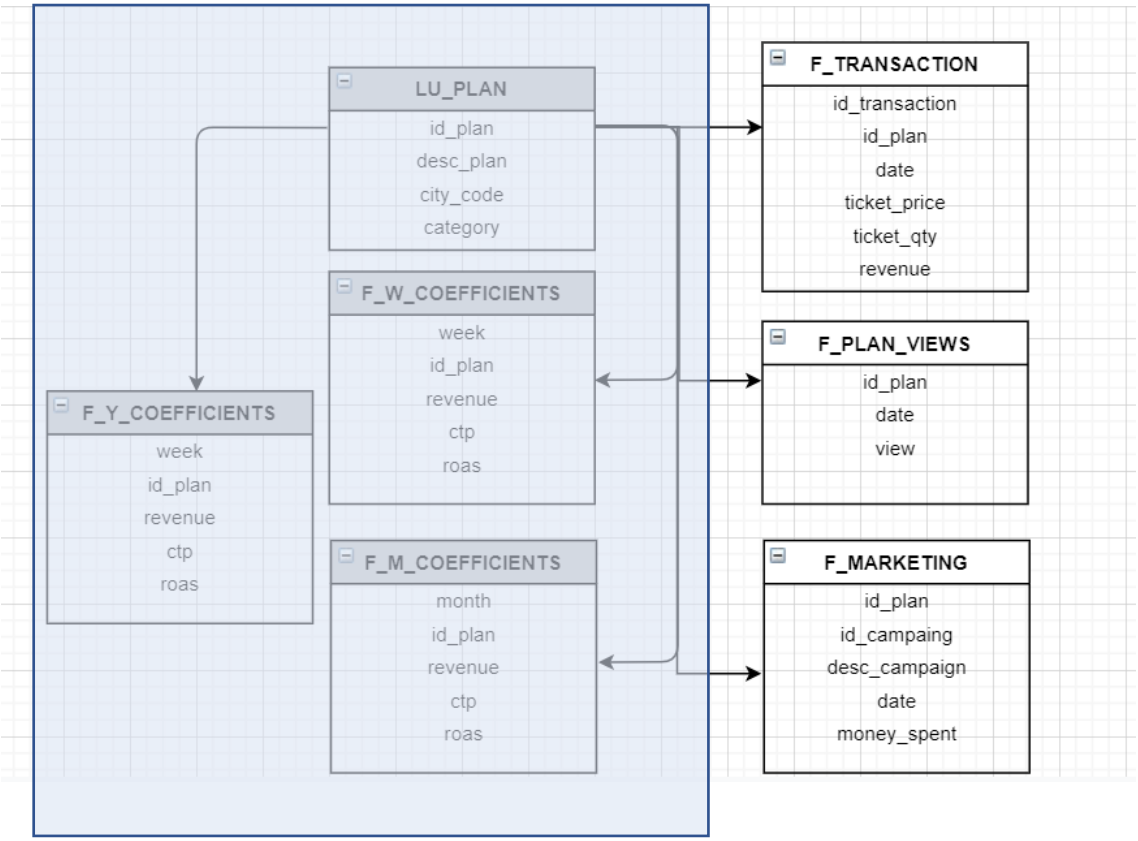
Exercise 1: Dimensional Model



Given the provided information we have opted for a Star Schema. We consider a look-up table for the plan, and we consider the other tables fact_tables. We take several decisions.

- No look up tables for attributes like category, marketing or city. In these cases, as not a lot of additional information is provided, we prefer redundancy in the fact tables. As mainly have ids for these objects, in case of a look up needed, it would be easily added.

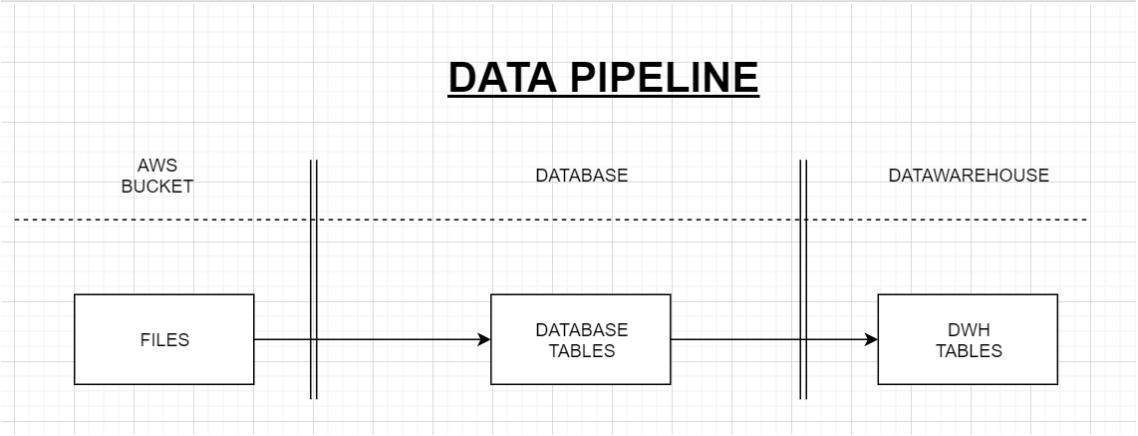
- About the date, we maintain the timestamp format in the original fact tables because in a future, it could be needed the time to make dossiers by time of the day. With microstrategy, we don't have problems using the timestamp as a normal date attribute using the objects available in the tool.

- Monthly, Weekly and Yearly tables because we are working with coefficients. For sake of the required calculations is better to do the coefficients in the database.

As the question is for the optimal model, it could be interpreted that we are going to model just the needed part for answering the questions. In that case, we should only consider in the model the tables LU_PLAN, F_W_COEFFICIENTS, F_M_COEFFICIENTS and F_Y_COEFFICIENTS. If we consider optimal in terms of scalability, then all the tables should be added to the DWH, because we would be able to answer more questions.

In this case, we will consider optimal the model with just the look up table for plans and the three facts tables.

Exercise 2: Pipeline



We have a very simple pipeline.

DATABASE INGESTION

First, we get the files from the AWS bucket to the database tables. This is a normal ingestión of files. Depending of the type of generated files we can overwrite or update the tables of the database.

This could be made with a script in Python or with any normal ingestión tool.

In this case we will have four tables.

tmp_test_micro_transactions

- id
- plan_id
- purchase_date
- ticket_price
- ticket_qty

tmp_test_micro_plans

- id
- description
- city_code
- category

tmp_test_micro_plan_views

- plan_id
- date

tmp_test_micro_marketing_spend

- campaign_id
- campaing_name
- date
- marketing_spend

DATAWHAREHOUSE

From the database tables we load the datawarehouse. If we choose the option of only the needed tables we would update four tables in the datawarehouse. If we try the more complete model, we would update the seven tables.

These could be made with powercenter or with other ETL tool. We would have the following tables. In green there`s the origin of the field.

LU_PLAN

- id_plan (tmt_test_micro_plans.id)
- desc_plan (tmt_test_micro_plans.description)
- city_code (tmt_test_micro_plans.city_code)
- category (tmt_test_micro_plans.category)

F_W_COEFFICIENTS

- id_plan (tmt_test_micro_plans.id)
- week (extracted from date)
- revenue (ticket_price * ticket_qty)
- ctp (nº of transactions / nº of plan views)
- roas (revenue / marketing spen)

F_M_COEFFICIENTS

- id_plan (tmt_test_micro_plans.id)
- month (extracted from date)
- revenue (ticket_price * ticket_qty)
- ctp (nº of transactions / nº of plan views)
- roas (revenue / marketing spen)

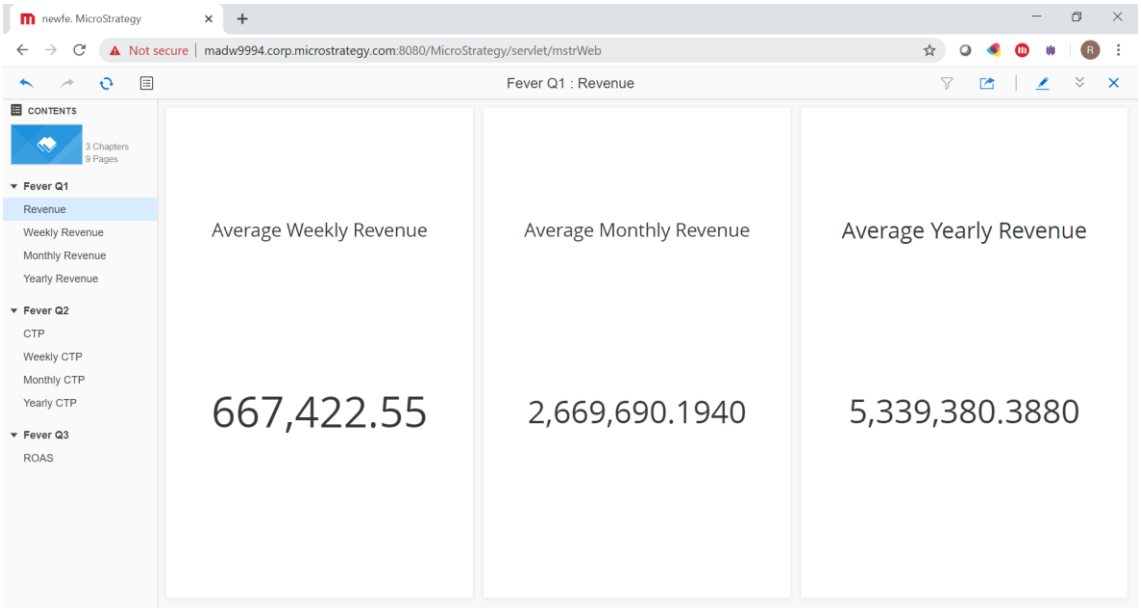
F_Y_COEFFICIENTS

- id_plan (tmt_test_micro_plans.id)
- year (extracted from date)
- revenue (ticket_price * ticket_qty)
- ctp (nº of transactions / nº of plan views)
- roas (revenue / marketing spen)

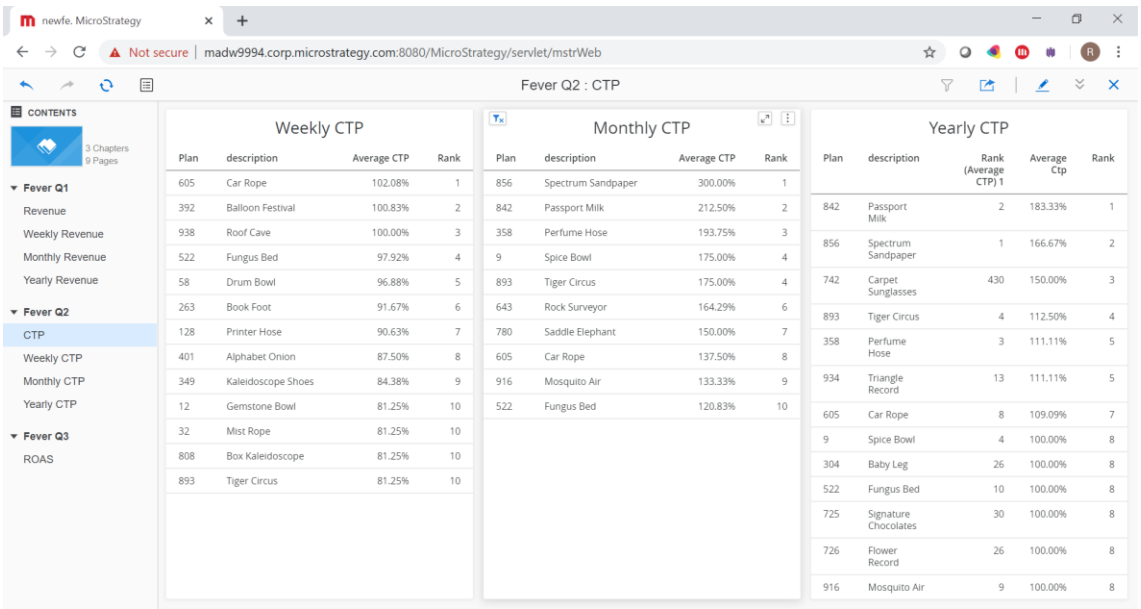
Exercise 4: Answers

As I wasn't able to build the pipeline, this microstrategy dossiers are made from datasets loaded from the provided csv files. I built some datasets with the same structure as the tables I mention in the exercise 3 and I worked with them.

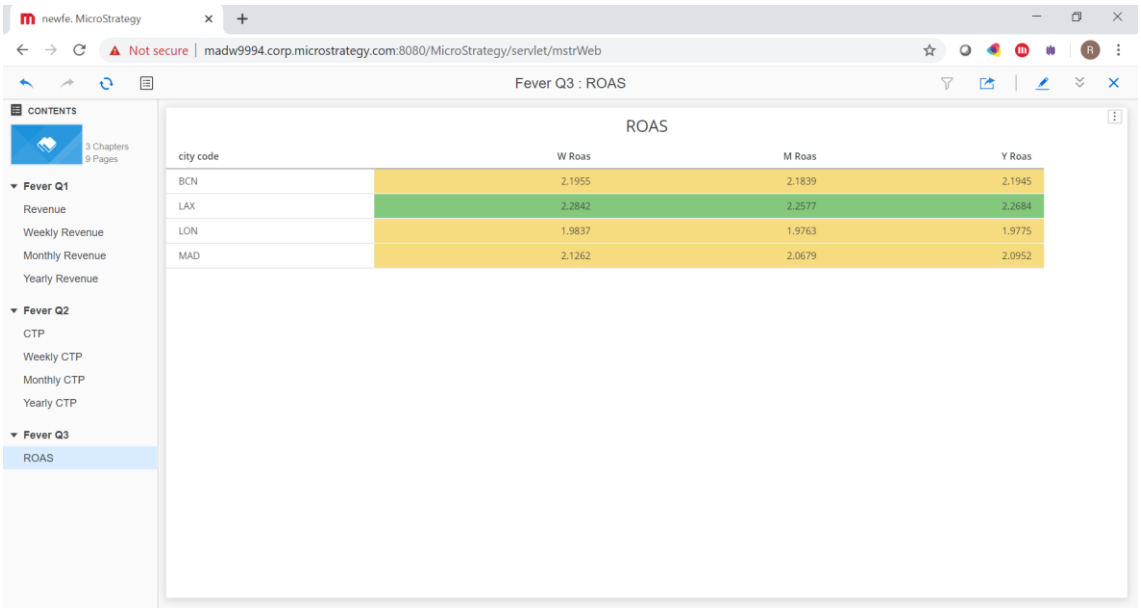
Question 1:



Question 2:

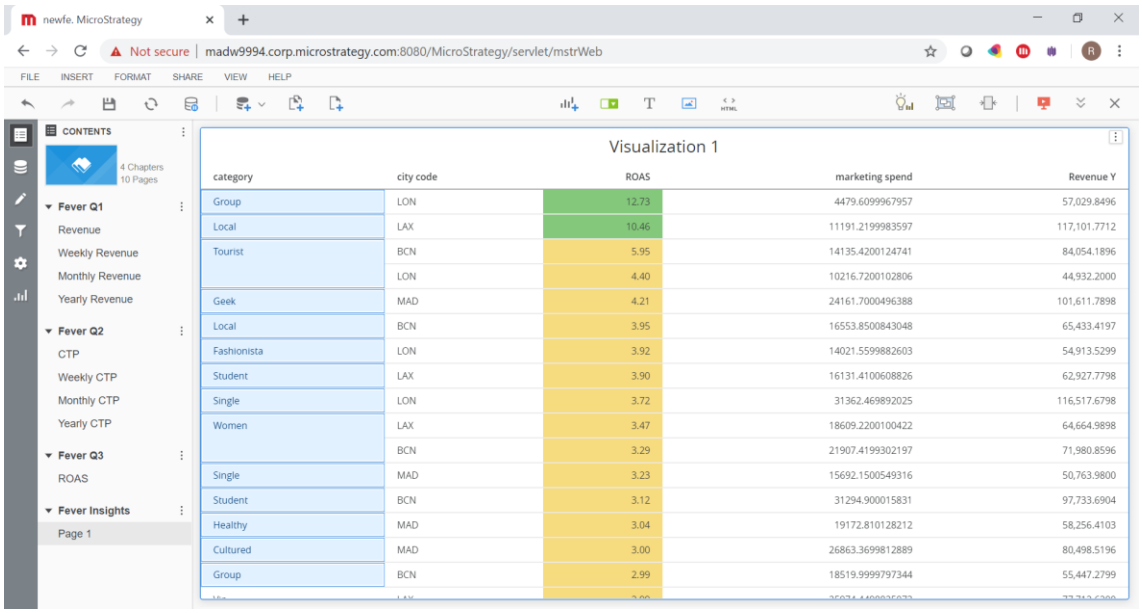
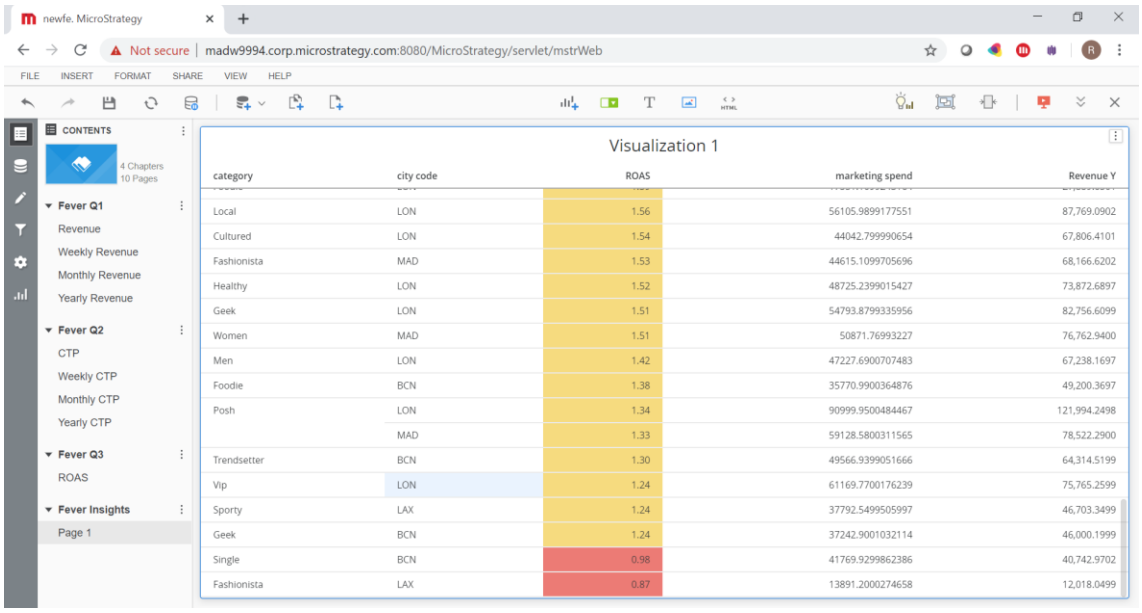


Question 3:


newfe. MicroStrategy
Not secure | madw9994.corp.microstrategy.com:8080/MicroStrategy/servlet/mstrWeb
Fever Q3 : ROAS
CONTENTS
3 Chapters
9 Pages
Fever Q1
Revenue
Weekly Revenue
Monthly Revenue
Yearly Revenue
Fever Q2
CTP
Weekly CTP
Monthly CTP
Yearly CTP
Fever Q3
ROAS
ROAS
city code
W Roas
M Roas
Y Roas
BCN
2.1955
2.1839
2.1945
LAX
2.2842
2.2577
2.2684
LON
1.9837
1.9763
1.9775
MAD
2.1262
2.0679
2.0952

Exercise 4: Answers

It’s more profitable invest in marketing for activities of group in London or local in Los Angeles.



It’s not profitable to invest in marketing for activities of single in Barcelona or fashionista in Los Angeles.