# Start with the TPC-DS Dataset from Snowflake. Design a dashboard that will leverage queries from Snowflake to build the dashboard.

Query:

query31.tpl List counties where the percentage growth in web sales is consistently higher compared to the percentage growth in store sales in the first three consecutive quarters for a given year.

with ss as

(select ca\_county,d\_qoy, d\_year,sum(ss\_ext\_sales\_price) as store\_sales

from store\_sales,date\_dim,customer\_address

where ss\_sold\_date\_sk = d\_date\_sk

and ss\_addr\_sk=ca\_address\_sk

group by ca\_county,d\_qoy, d\_year),

ws as

(select ca\_county,d\_qoy, d\_year,sum(ws\_ext\_sales\_price) as web\_sales

from web\_sales,date\_dim,customer\_address

where ws\_sold\_date\_sk = d\_date\_sk

and ws\_bill\_addr\_sk=ca\_address\_sk

group by ca\_county,d\_qoy, d\_year)

select

ss1.ca\_county

,ss1.d\_year

,ws2.web\_sales/ws1.web\_sales web\_q1\_q2\_increase

,ss2.store\_sales/ss1.store\_sales store\_q1\_q2\_increase

,ws3.web\_sales/ws2.web\_sales web\_q2\_q3\_increase

,ss3.store\_sales/ss2.store\_sales store\_q2\_q3\_increase

from

ss ss1

,ss ss2

,ss ss3

,ws ws1

,ws ws2

,ws ws3

where

ss1.d\_qoy = 1

and ss1.d\_year = 2001

and ss1.ca\_county = ss2.ca\_county

and ss2.d\_qoy = 2

and ss2.d\_year = 2001

and ss2.ca\_county = ss3.ca\_county

and ss3.d\_qoy = 3

and ss3.d\_year = 2001

and ss1.ca\_county = ws1.ca\_county

and ws1.d\_qoy = 1

and ws1.d\_year = 2001

and ws1.ca\_county = ws2.ca\_county

and ws2.d\_qoy = 2

and ws2.d\_year = 2001

and ws1.ca\_county = ws3.ca\_county

and ws3.d\_qoy = 3

and ws3.d\_year =2001

and case when ws1.web\_sales > 0 then ws2.web\_sales/ws1.web\_sales else null end

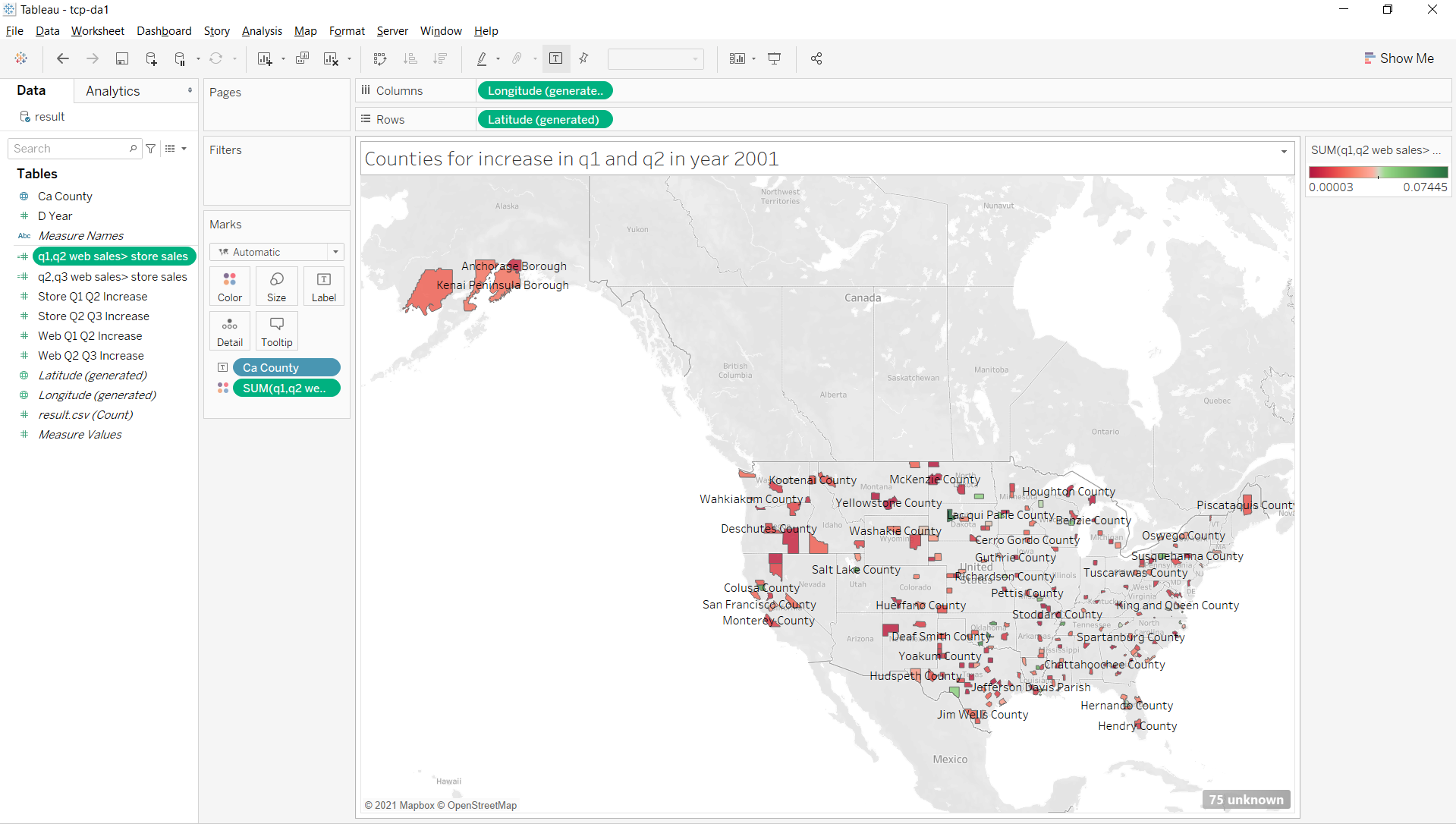
> case when ss1.store\_sales > 0 then ss2.store\_sales/ss1.store\_sales else null end

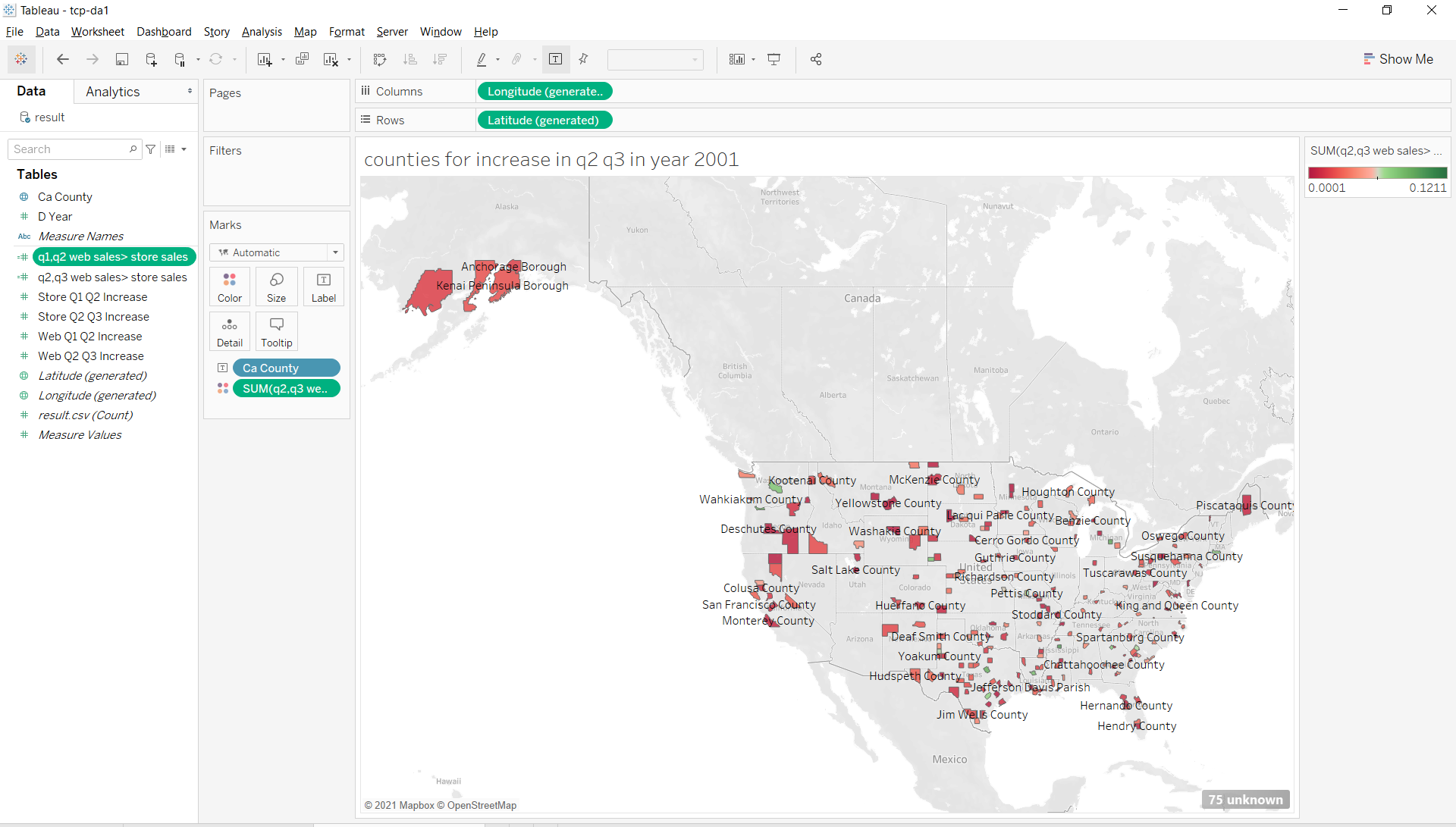
and case when ws2.web\_sales > 0 then ws3.web\_sales/ws2.web\_sales else null end

> case when ss2.store\_sales > 0 then ss3.store\_sales/ss2.store\_sales else null end

order by web\_q2\_q3\_increase;

List of counties for percentage growth in web sales is consistently higher compared to the percentage growth in store sales in the first three consecutive quarters for a given year:





# Discuss who is this dashboard targeted towards and the use cases you will accomplish with it:

Dashboard is designed for Sales optimization, using which we can distribute the products across stores accordingly. The following are the use cases with this kind of optimization:

* Suggestions on discounts on store specific products.
* Promotions based on the sales data of web and store.
* Proposal for store closures which are in a negative revenue generation state.
* Promotions and appreciation for the employees based on the store specific performance.
* Transfer of products from one store to another if the other is doing good sales and again it depends on logistics.

# Describe your design on how would you onboard the dataset

Warehouse database schema on table created. Through snowflake python connector by using

pandas connector package inserted data into table which we already created with the same

column names as we are on-boarding the dataset through pandas.

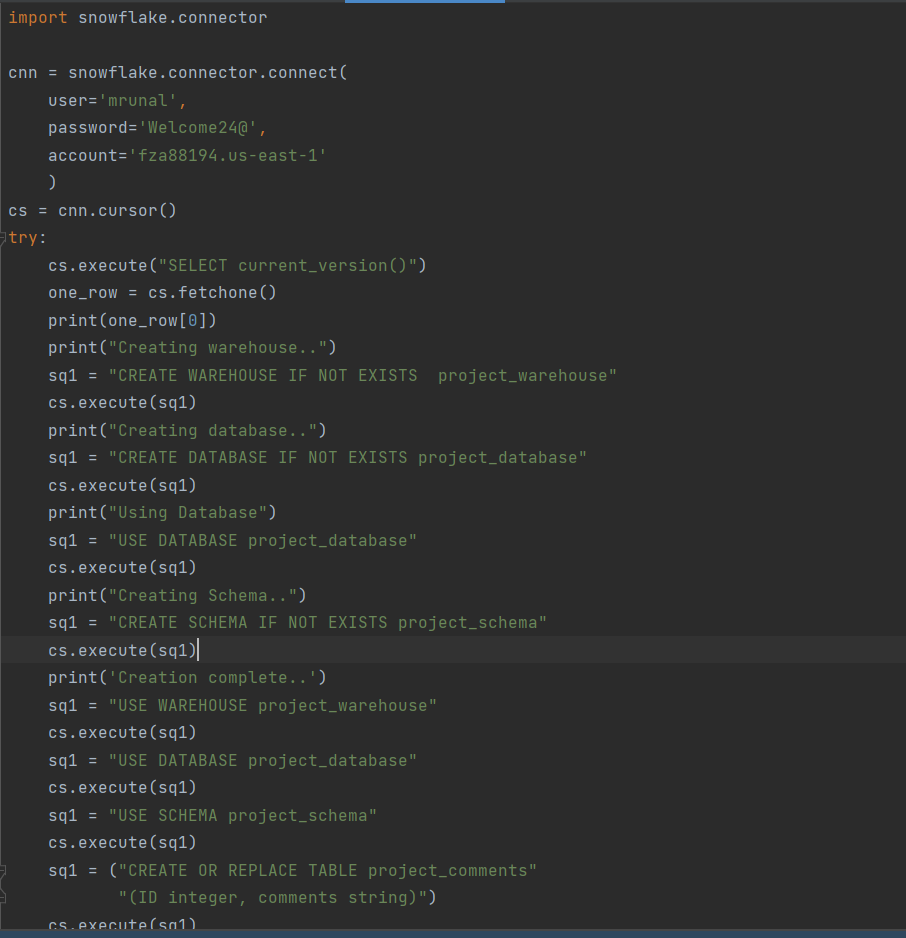
# Describe what tools (xsv, Python) will be used for data cleanup:

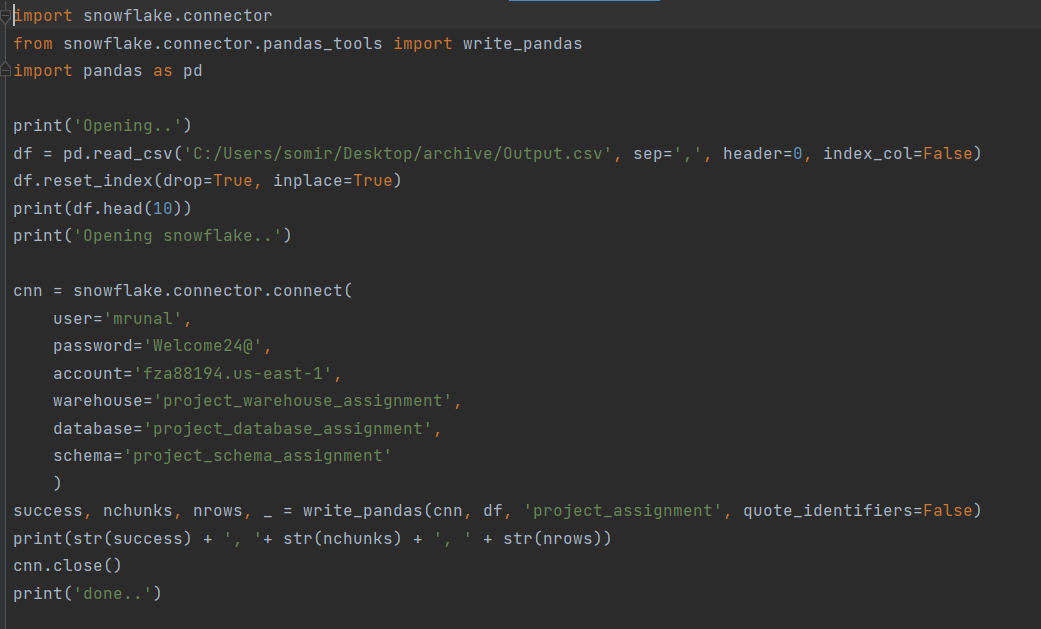
We used python (pandas, NumPy) to data cleaning:

* Handling missing values
* Converting Data types
* Dealing with duplicates
* Dropping unnecessary columns

# Prototype your application

1. Have used adventureworks DW: Dim date, dim geography, factResellerSales, factInternetSales.
2. Colab link for pre processing steps:https://colab.research.google.com/drive/1x-hH49VAAPlCtzL0NiDoGC3ErNJ964Aa#scrollTo=isFBWNHy3oFX
3. Inserting data to snowflake through python and pandas.





iv.

Show how this data can be visualized in an analytics tool of your choice.

Found out List cities where reseller sales is consistently higher compared to the internet sales

for a given year.

