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:

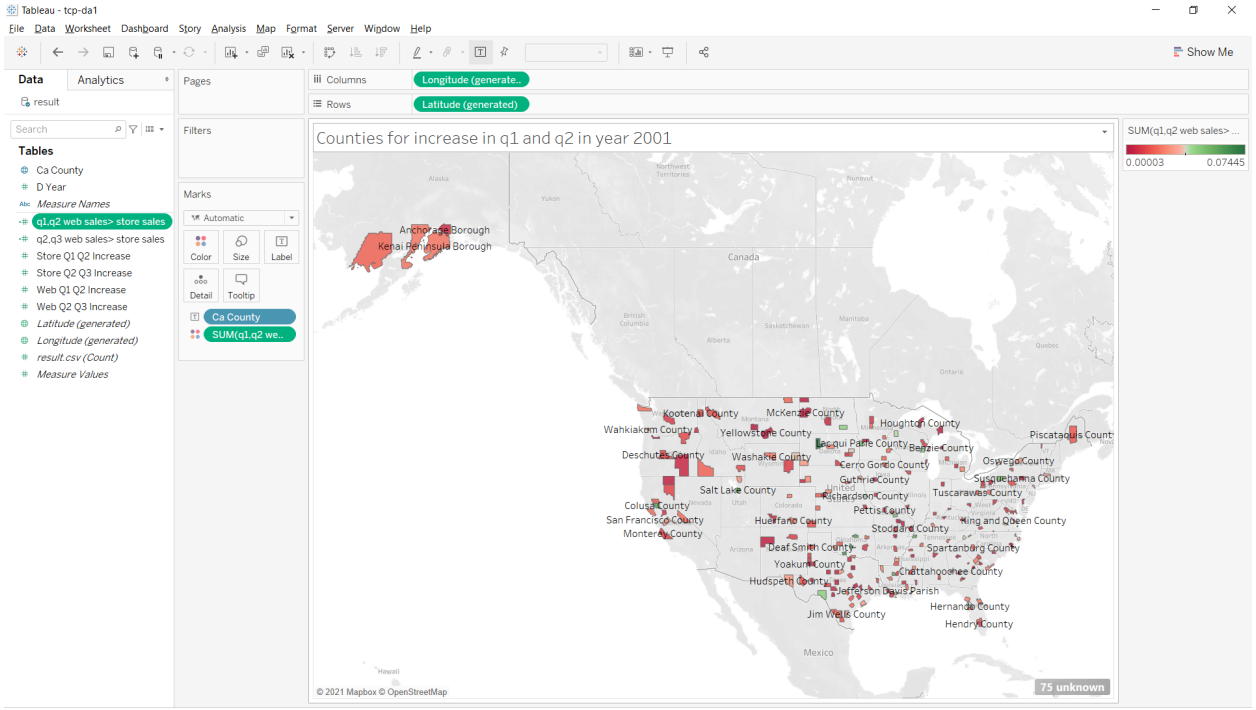
Tableau - tcp-da1
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data Analytics
result
Search
Tables
Ca County
D Year
Measure Names
q1,q2 web sales> store sales
q2,q3 web sales> store sales
Store Q1 Q2 Increase
Store Q2 Q3 Increase
Web Q1 Q2 Increase
Web Q2 Q3 Increase
Latitude (generated)
Longitude (generated)
result.csv (Count)
Measure Values
Pages
Filters
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Ca County
SUM(q1,q2 we..
Columns
Longitude (generate..
Rows
Latitude (generated)
Counties for increase in q1 and q2 in year 2001
SUM(q1,q2 web sales> ...
0.00003
0.07445
Anchorage Borough
Kenai Peninsula Borough
Canada
Mexico
Kootenai County
McKenzie County
Houghton County
Wahkiakum County
Yellowstone County
Lac qui Parle County
Piscataquis County
Deschutes County
Washakie County
Cerro Gordo County
Oswego County
Guthrie County
Susquehanna County
Salt Lake County
Richardson County
Tuscarawas County
Colusa County
Pettis County
San Francisco County
Huerfano County
King and Queen County
Monterey County
Stoddard County
Deaf Smith County
Spartanburg County
Yoakum County
Chattahoochee County
Hudspeth County
Jefferson Davis Parish
Hernando County
Jim Wells County
Hendry County
© 2021 Mapbox © OpenStreetMap
75 unknown

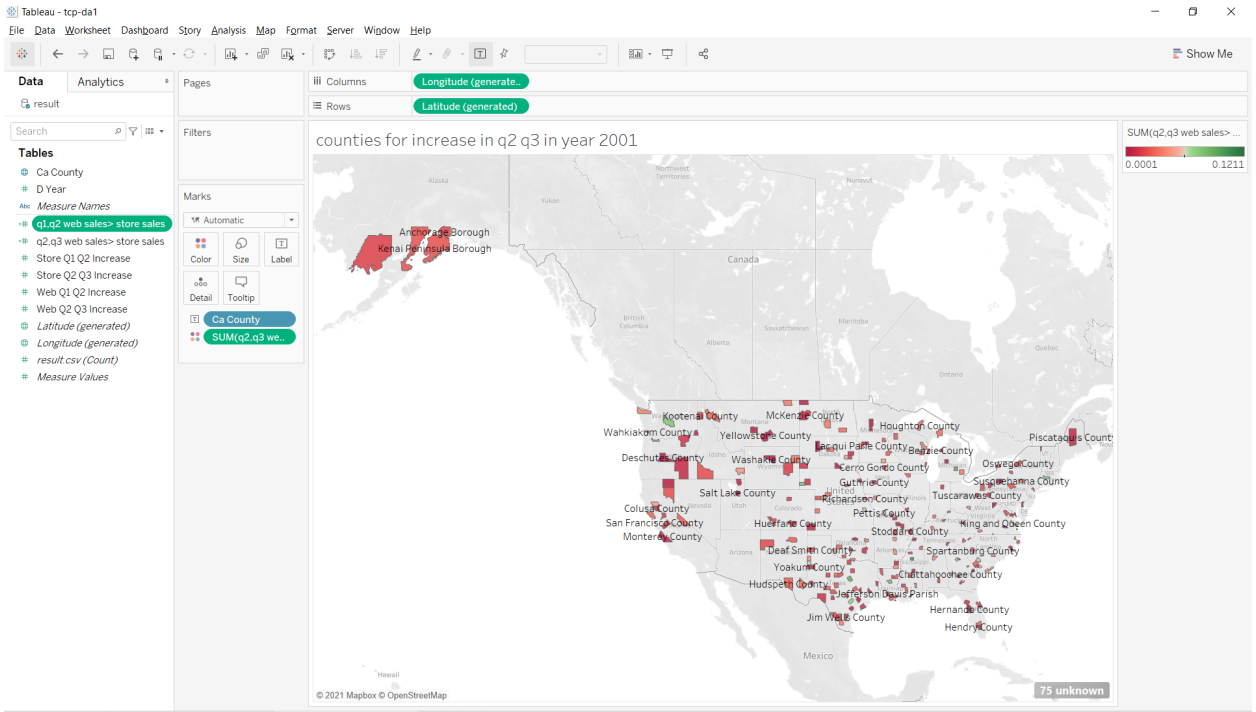
Tableau - tcp-da1
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data Analytics
result
Search
Tables
Ca County
D Year
Measure Names
q1,q2 web sales> store sales
q2,q3 web sales> store sales
Store Q1 Q2 Increase
Store Q2 Q3 Increase
Web Q1 Q2 Increase
Web Q2 Q3 Increase
Latitude (generated)
Longitude (generated)
result.csv (Count)
Measure Values
Pages
Filters
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Ca County
SUM(q2,q3 we..
Columns
Longitude (generate..
Rows
Latitude (generated)
counties for increase in q2 q3 in year 2001
SUM(q2,q3 web sales> ...
0.0001
0.1211
Anchorage Borough
Kenai Peninsula Borough
Canada
Mexico
Kootenai County
McKenzie County
Houghton County
Wahkiakum County
Yellowstone County
Lac qui Parle County
Piscataquis County
Deschutes County
Washakie County
Cerro Gordo County
Oswego County
Guthrie County
Susquehanna County
Salt Lake County
Richardson County
Tuscarawas County
Colusa County
Pettis County
San Francisco County
Huerfano County
King and Queen County
Monterey County
Stoddard County
Deaf Smith County
Spartanburg County
Yoakum County
Chattahoochee County
Hudspeth County
Jefferson Davis Parish
Hernando County
Jim Wells County
Hendry County
© 2021 Mapbox © OpenStreetMap
75 unknown

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-hH49VAAPlCtzL
3. Inserting data to snowflake through python and pandas.

```
import snowflake.connector

cnn = snowflake.connector.connect(
    user='mrunal',
    password='Welcome24@',
    account='fza88194.us-east-1'
    )
cs = cnn.cursor()
try:
    cs.execute("SELECT current_version()")
    one_row = cs.fetchone()
    print(one_row[0])
    print("Creating warehouse..")
    sq1 = "CREATE WAREHOUSE IF NOT EXISTS  project_warehouse"
    cs.execute(sq1)
    print("Creating database..")
    sq1 = "CREATE DATABASE IF NOT EXISTS project_database"
    cs.execute(sq1)
    print("Using Database")
    sq1 = "USE DATABASE project_database"
    cs.execute(sq1)
    print("Creating Schema..")
    sq1 = "CREATE SCHEMA IF NOT EXISTS project_schema"
    cs.execute(sq1)
    print('Creation complete..')
    sq1 = "USE WAREHOUSE project_warehouse"
    cs.execute(sq1)
    sq1 = "USE DATABASE project_database"
    cs.execute(sq1)
    sq1 = "USE SCHEMA project_schema"
    cs.execute(sq1)
    sq1 = ("CREATE OR REPLACE TABLE project_comments"
           "(ID integer, comments string)")
    cs.execute(sq1)
```

```python
import snowflake.connector
from snowflake.connector.pandas_tools import write_pandas
import pandas as pd

print('Opening..')
df = pd.read_csv('C:/Users/somir/Desktop/archive/Output.csv', sep=',', header=0, index_col=False)
df.reset_index(drop=True, inplace=True)
print(df.head(10))
print('Opening snowflake..')

cnn = snowflake.connector.connect(
    user='mrunal',
    password='Welcome24@',
    account='fza88194.us-east-1',
    warehouse='project_warehouse_assignment',
    database='project_database_assignment',
    schema='project_schema_assignment'
    )
success, nchunks, nrows, _ = write_pandas(cnn, df, 'project_assignment', quote_identifiers=False)
print(str(success) + ', '+ str(nchunks) + ', ' + str(nrows))
cnn.close()
print('done..')
```

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.

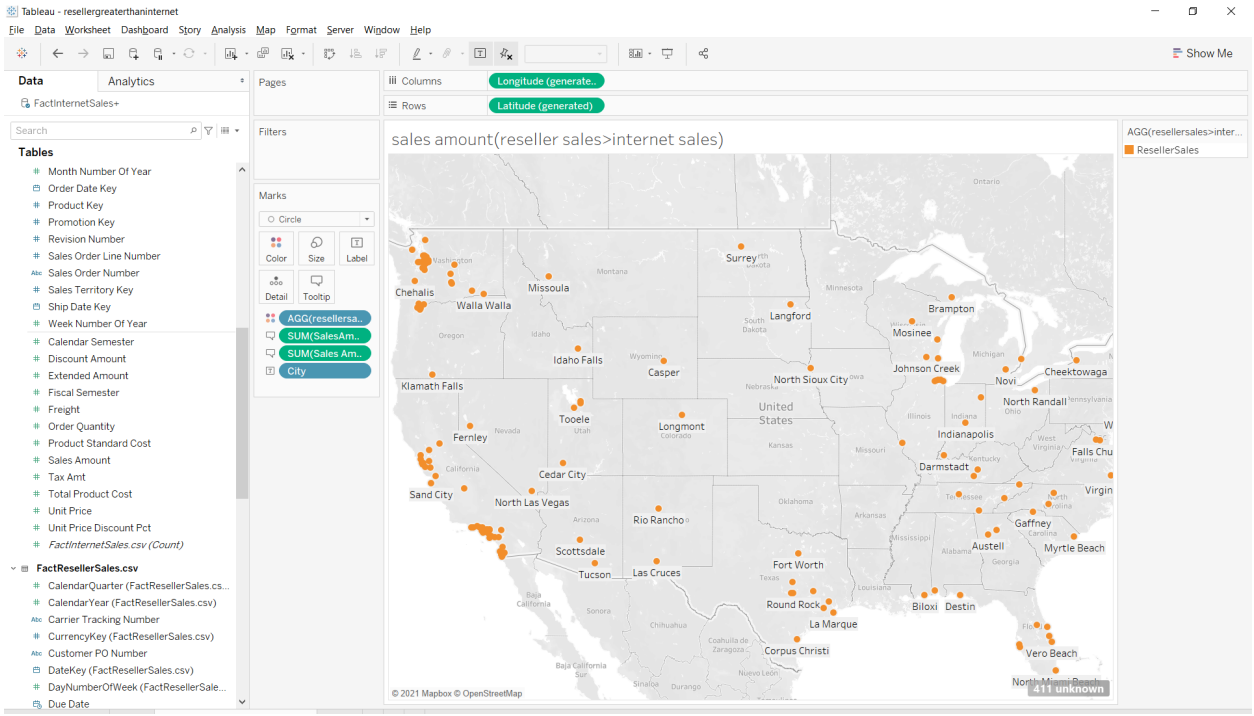
Tableau - resellergreaterthaninternet
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data
Analytics
FactInternetSales+
Search
Tables
Month Number Of Year
Order Date Key
Product Key
Promotion Key
Revision Number
Sales Order Line Number
Sales Order Number
Sales Territory Key
Ship Date Key
Week Number Of Year
Calendar Semester
Discount Amount
Extended Amount
Fiscal Semester
Freight
Order Quantity
Product Standard Cost
Sales Amount
Tax Amt
Total Product Cost
Unit Price
Unit Price Discount Pct
FactInternetSales.csv (Count)
FactResellerSales.csv
CalendarQuarter (FactResellerSales.cs...
CalendarYear (FactResellerSales.csv)
Carrier Tracking Number
CurrencyKey (FactResellerSales.csv)
Customer PO Number
DateKey (FactResellerSales.csv)
DayNumberOfWeek (FactResellerSale...
Due Date
Pages
Columns
Longitude (generate..
Rows
Latitude (generated)
Filters
Marks
Circle
Color
Size
Label
Detail
Tooltip
AGG(resellersa..
SUM(SalesAm..
SUM(Sales Am..
City
sales amount(reseller sales>internet sales)
AGG(resellersales>inter...
ResellerSales
Chehalis
Walla Walla
Missoula
Surrey
Langford
Brampton
Mosinee
Idaho Falls
Casper
North Sioux City
Johnson Creek
Novi
Cheektowaga
Klamath Falls
Tooele
Longmont
United States
North Randall
Indianapolis
Fernley
Cedar City
Darmstadt
Sand City
North Las Vegas
Rio Rancho
Gaffney
Scottsdale
Austell
Myrtle Beach
Tucson
Las Cruces
Fort Worth
Round Rock
Biloxi
Destin
La Marque
Corpus Christi
Vero Beach
© 2021 Mapbox © OpenStreetMap
411 unknown