AIR POLLUTION

CCDS 311 – Data Visualization
Course Instructors: Dr Ines Boufateh

GROUP WORK:
GHADI ALSHEARI 1905145
RANEEM MUKHTAR 1911807

Table of Content

01 Introduction: General Description

02 List of additional variables and hierarchies

03 Selected variables

04 Questions

05 Fix Data Types

07 Convert to measure or dimension.

08 Create additional variables

10 Create hierarchies

11 single variables visualization

15 two continuous variables Visualize

16 Visualize variables of different types

18 LOD Expression

19 Maps

23 *visual attributes in the visualizations*

General Description

As the importance of pollutant standards in air. Our dataset today is about "air quality" due to the big importance to different domains such as health, business..etc.What knowledge will be extracted from this dataset will help a lot in Decision making in all different domains.

Comparing the data with The national ambient air quality standards 'NAAQS' is important but our job today is to help decision makers in all different domains associated with different cases and problems.

All of this data in dataset cames from EPA’s Air Quality System (AQS). mention of attributes will used it in our visualization (Latitude, Longitude, AQI,.....and more) in the next pages will mention in it all.

List of additional variables and hierarchies

First Additional variables we decide it for our visualizes:

1-population Standard Deviation

2-Sample Standard Deviation

Which will contain these fields below:

-Observation Count: The number of observations (samples) taken during the day.

-Observation Percent: The percent representing the number of observations taken (only certain parameters).

-Arithmetic Mean: The average (arithmetic mean) value for the year.

Second Hierarchies we decide it for our visualizes:

1-Address Number

2-Address String

Which will contain these fields below:

-State Name: The name of the state where the monitoring site is located.

-County Name: The name of the county where the monitoring site is located.

-City Name: The name of the city where the monitoring site is located.

-State Code: The FIPS code of the state in which the monitor resides.

-County Code: The FIPS code of the county in which the monitor resides.

-Site Num: A unique number within the county identifying the site.

Selected variables

- Site Num : A unique number within the county identifying the site.
- Latitude: The monitoring site’s angular distance north of the equator measured in decimal degrees.
- Longitude: The monitoring site’s angular distance east of the prime meridian measured in decimal degrees.
- Date Local: The calendar date for the summary. All daily summaries are for the local standard day (midnight to midnight) at the monitor.
- AQI: The Air Quality Index for the day for the pollutant, if applicable.
- Observation Percent: The percent representing the number of observations taken to be taken during the day.
- 1st Max Value: The highest value for the day.
- Arthimtec Mean: The average value for the day.
- CBSA Name: The name of the area metropolitan area where the monitoring site is located.
- Date of Last Change: The date the last time any numeric values in this record were updated in the AQS data system.
- State Name: The name of the state where the monitoring site is located.
- Event Type: Indicates whether data measured during exceptional events are included in the summary.
- Local Site Name : The name of the site given by the State, local, or tribal air pollution control agency that operates it.
- Address : The approximate street address of the monitoring site.
- County Name :The name of the county where the monitoring site is located.
- City Name :The name of the city where the monitoring site is located.

questions

1-Determine the air quality index per day?

Air Quality Index field Depending on a local date field to show all local records for each day.

2-Find out the distribution of data is normal or skewed?

Using Population/Sample Standard Deviation to measure how dispersed the data is in relation to the mean. Low standard deviation means data are clustered around the mean, and high standard deviation indicates data are more spread out

3-What is the benefit of the Air Quality index depending on the monitoring site is located "CBSA"?

The Air Quality Index is daily recorded in the local history, any change occurs; The base statistical area where the monitoring site is located will be accessed, and update the date for the last modification date.

4-How can we derive useful information from determining the 1st max value for the number of observations taken for certain parameters?

We can see the maximum number of outages and exceptional events listed in the summary that affect air quality.

5-report the Air Pollution Control Agency (local site name) showing the AQI percentage for each local site.

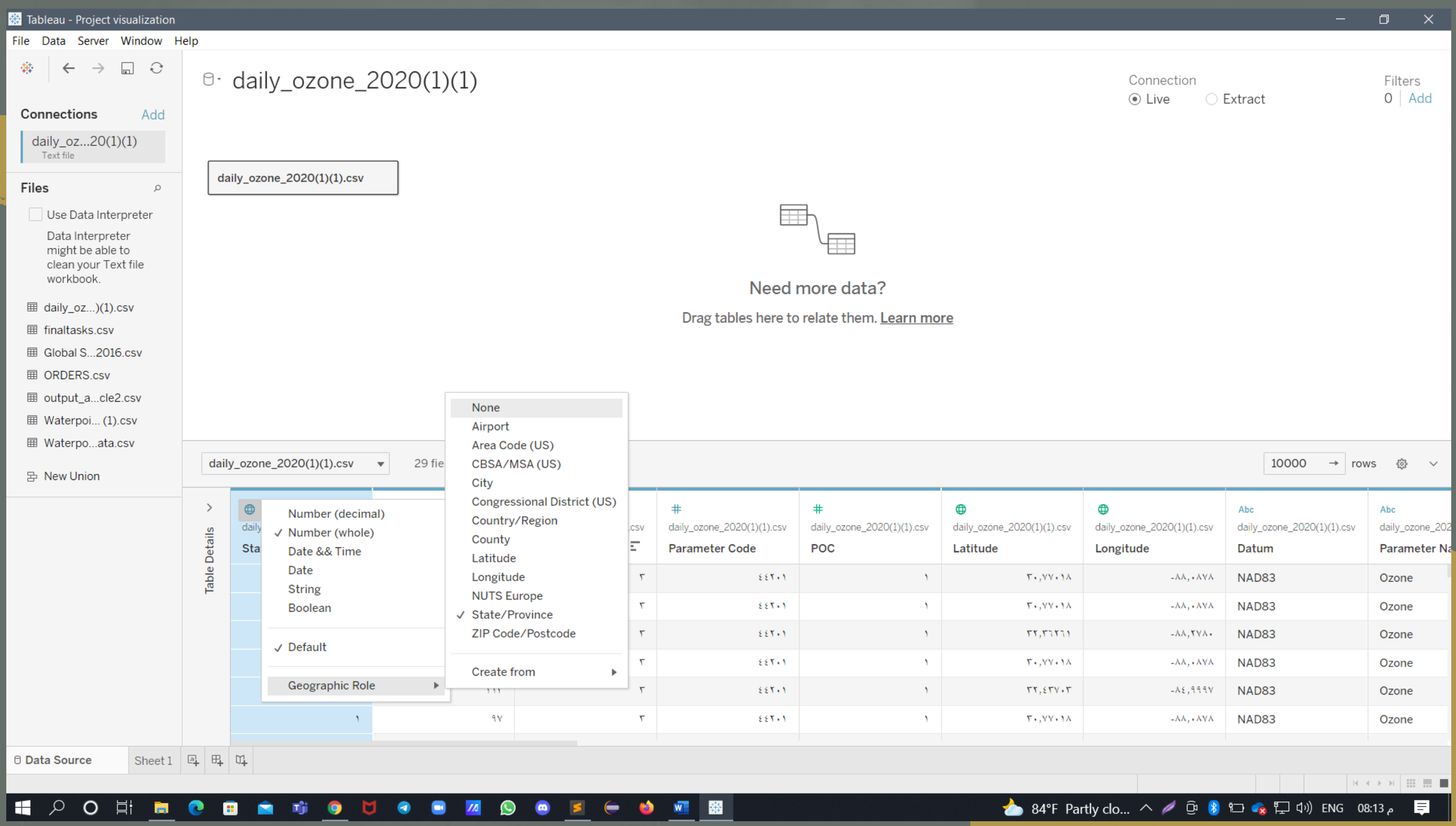
6-Determine the air quality index for each state?

Air quality index field based on state name field to show all air quality ratio values for each state.

Fix data types

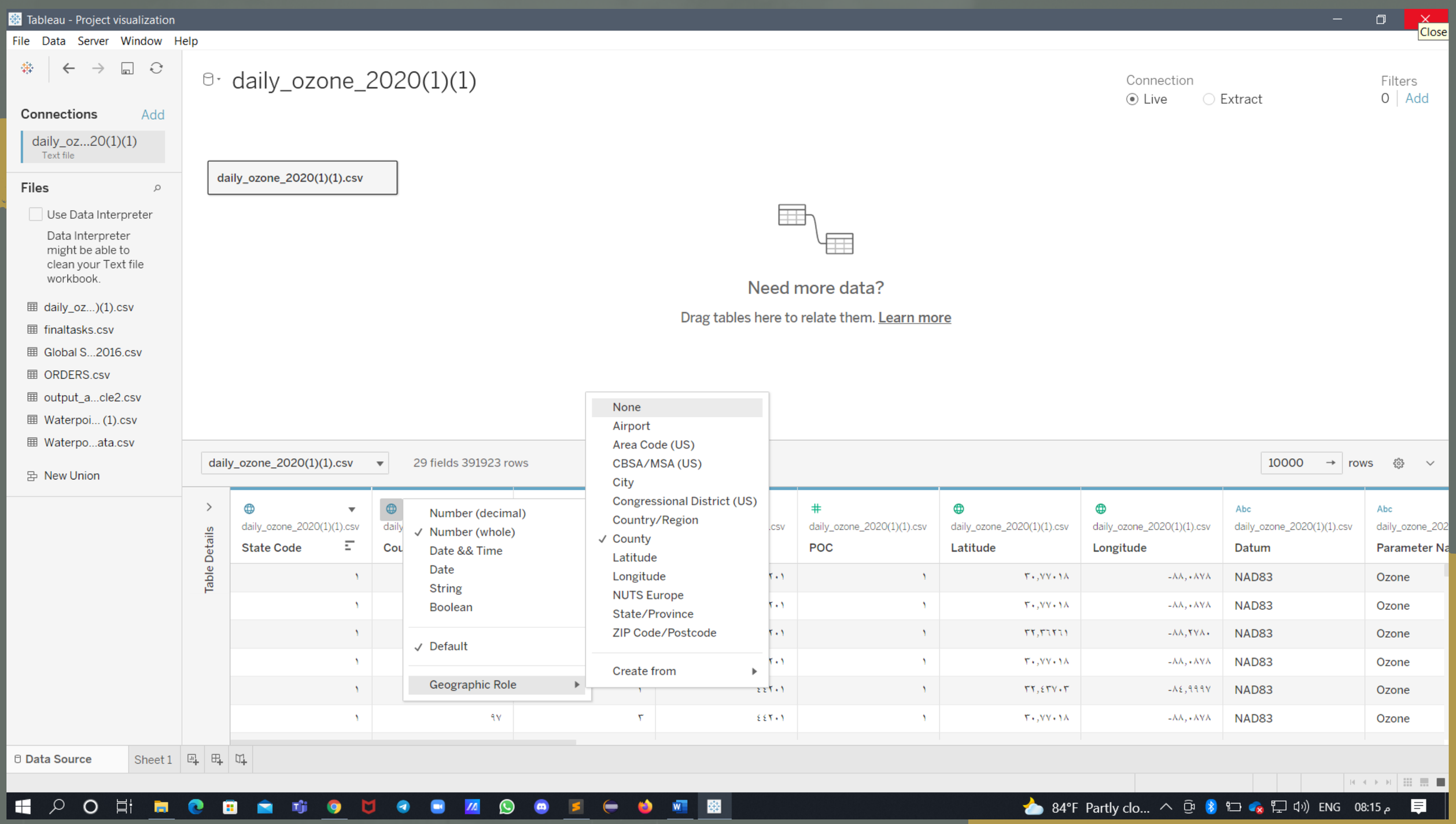
We changed some data types according to our need in visualizing. All changes only in Geographic Role to suit are needed, all other fields has a correct data type so it doesn't need to fix it.

- **Changed the Geographic Role in the state code field from "None" to "state/Province"**



Fix data types

- changed the Geographic Role in the County code field from "None" to "County".

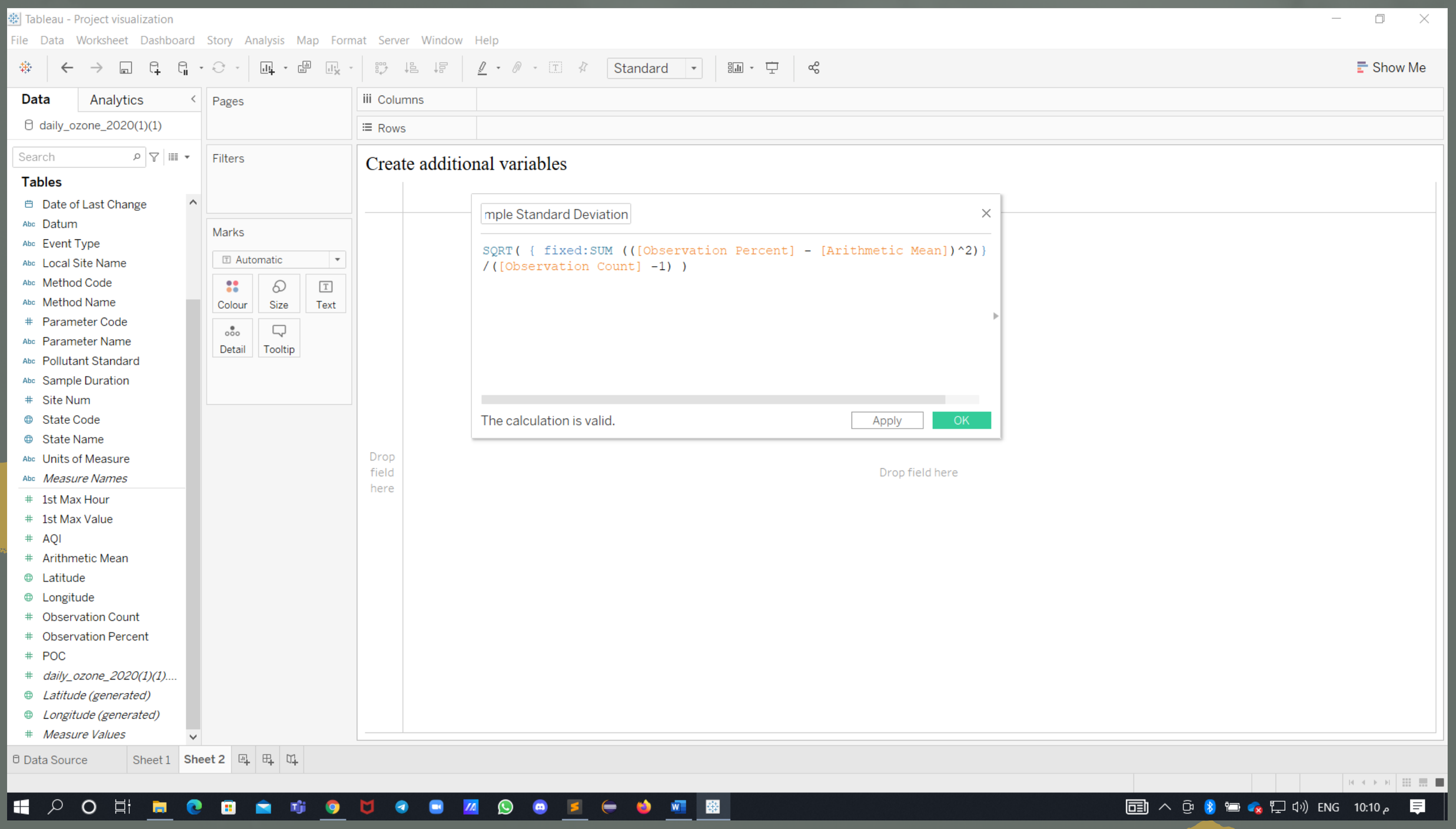


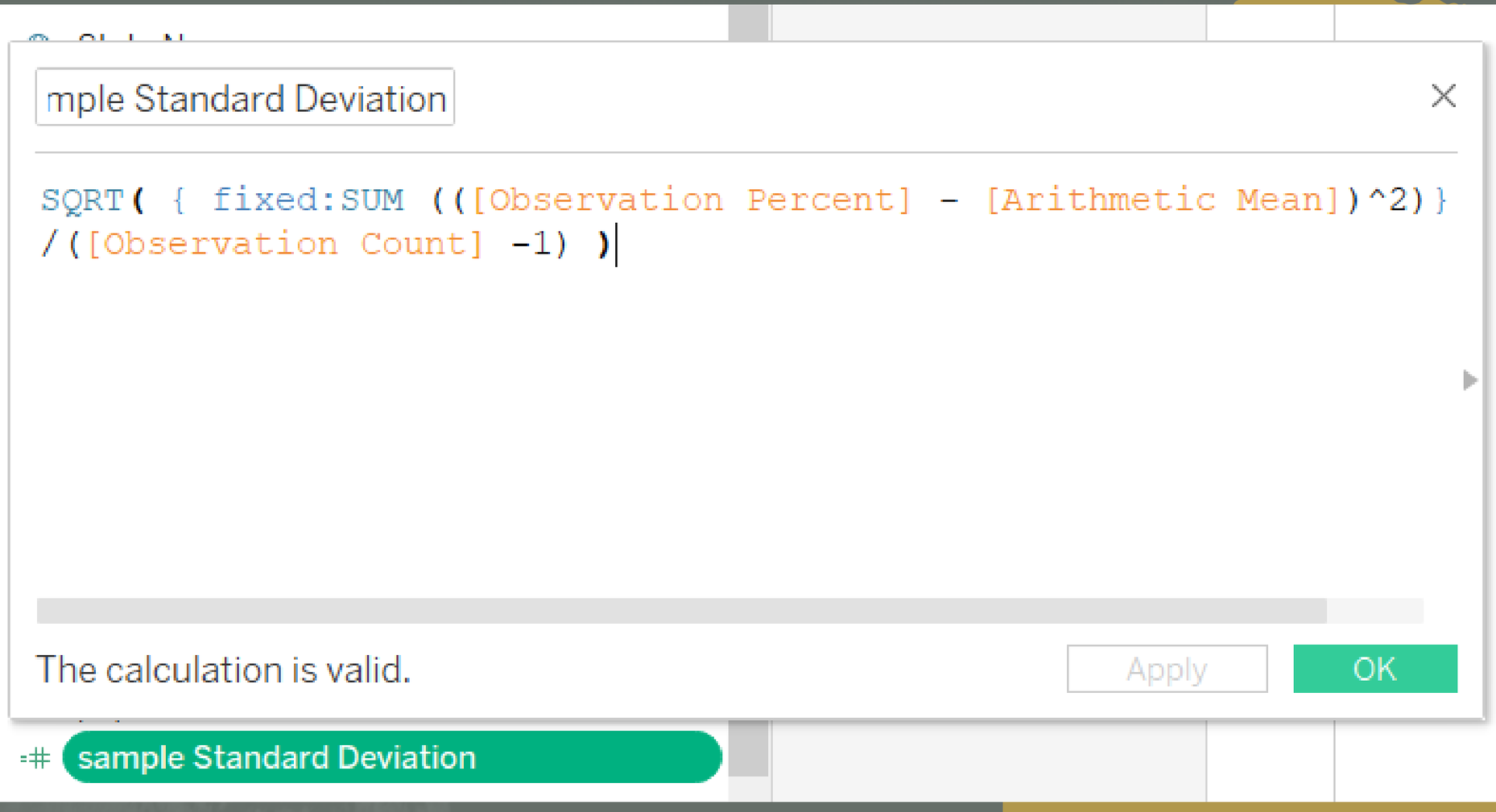
Convert to measure or dimension.

- **All fields were correct furthermore, we didn't need to convert any field.**

Create additional variables

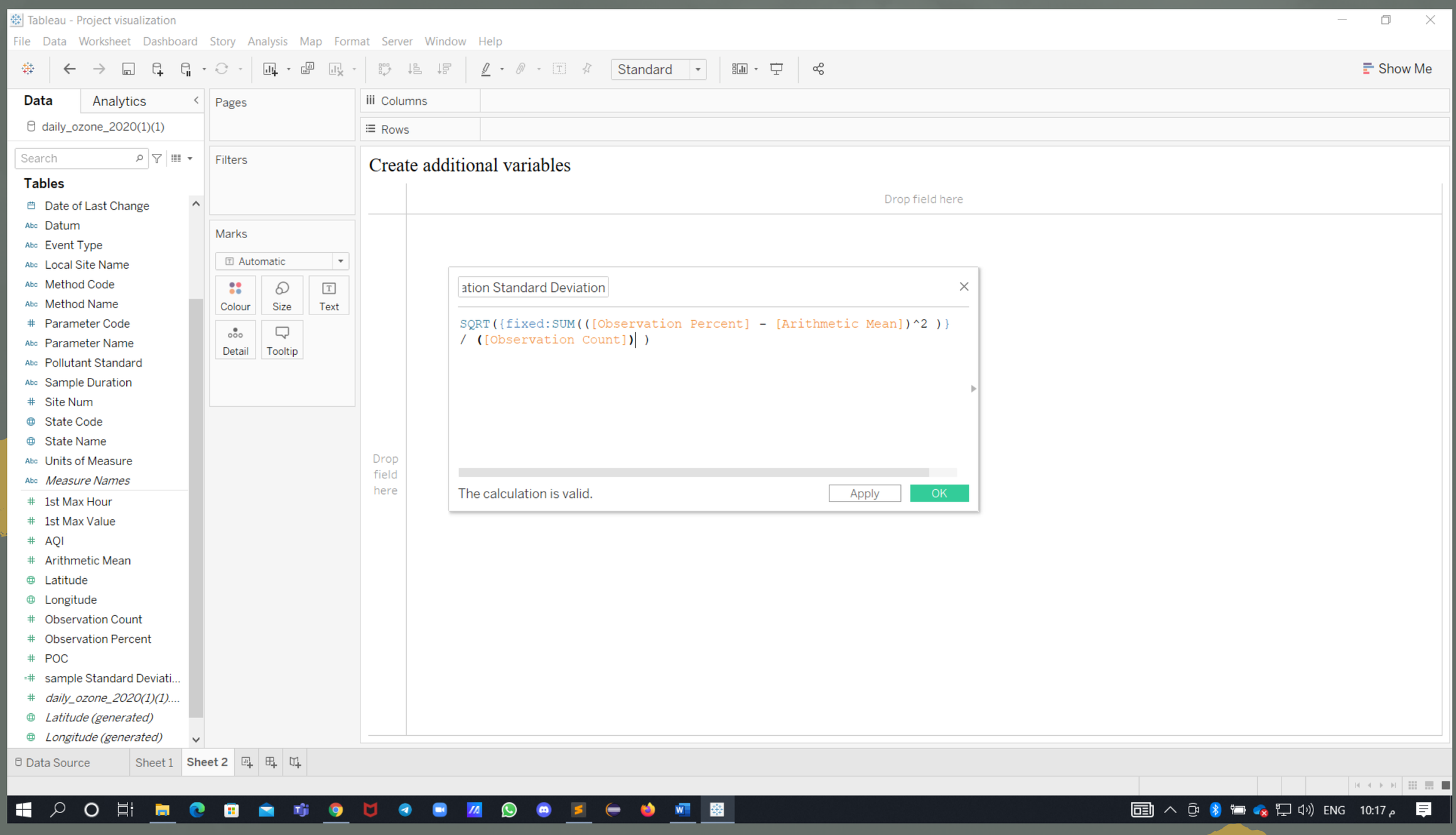
Create Sample Standard deviation additional variable to compute the Sample Standard Deviation method.

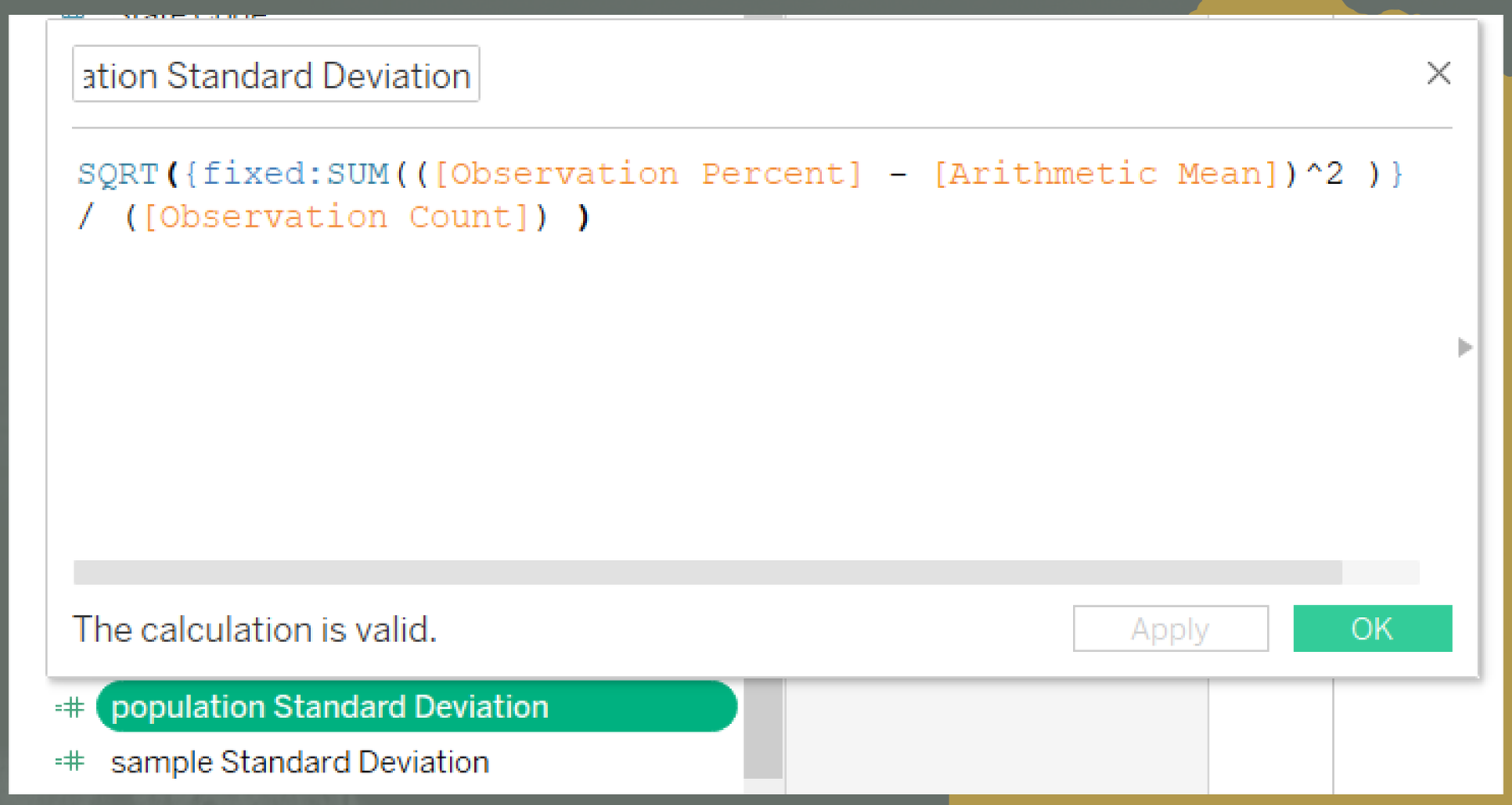




Create additional variables

Create Population Standard deviation additional variable to compute the population Standard Deviation method.



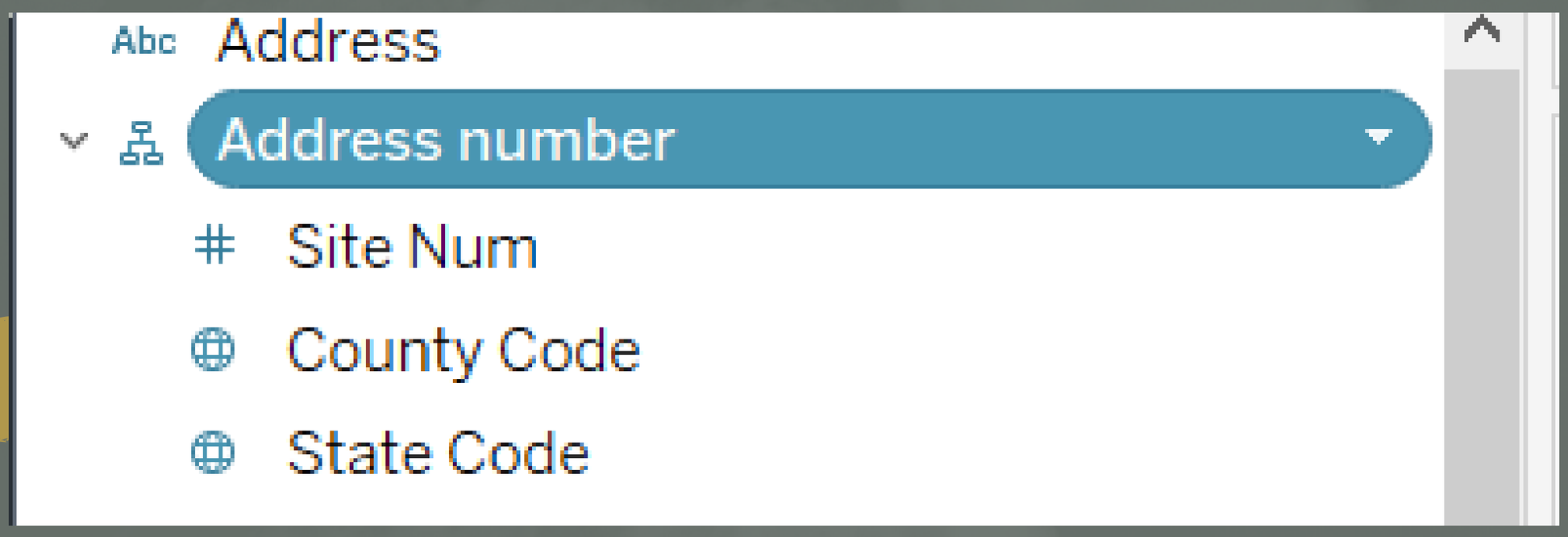


Create hierarchies

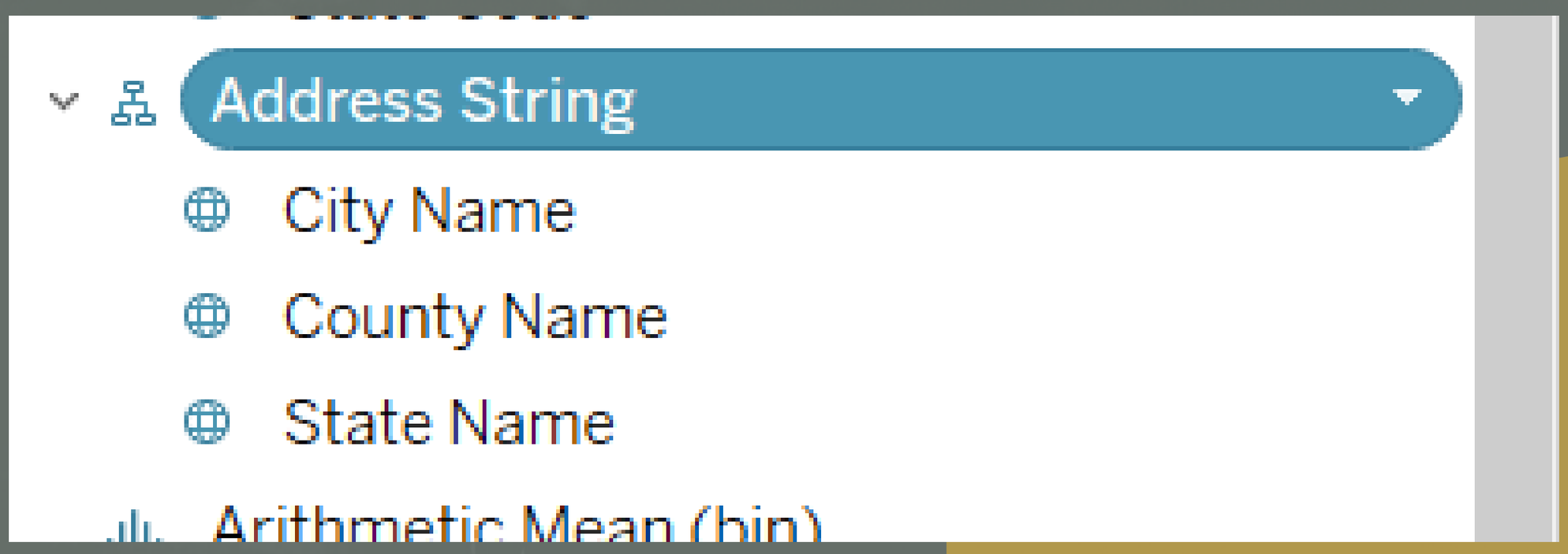
To create our hierarchies:

In the Data pane, drag a field and drop it directly on top of another field.

1

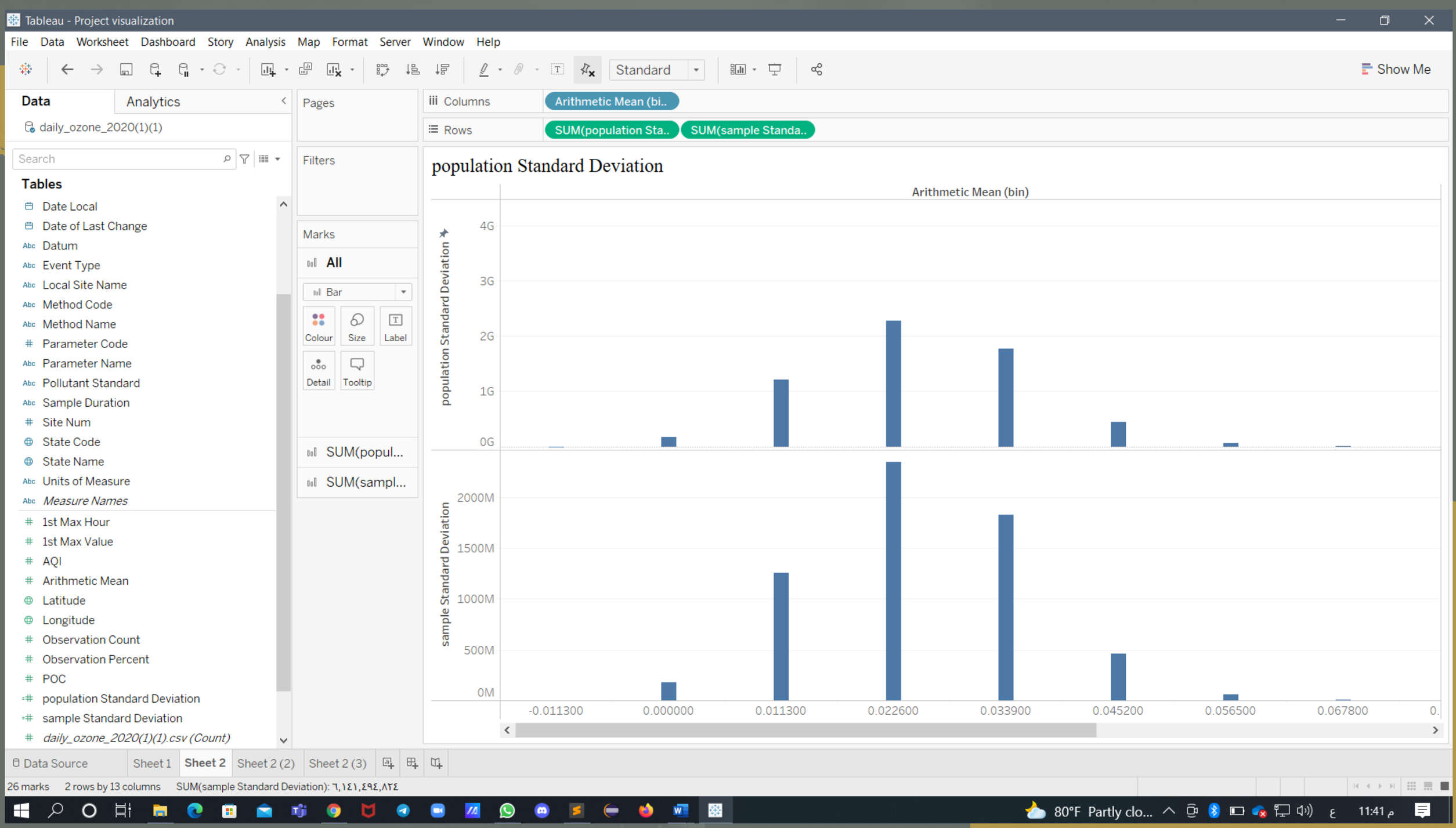


2



single variables visualization

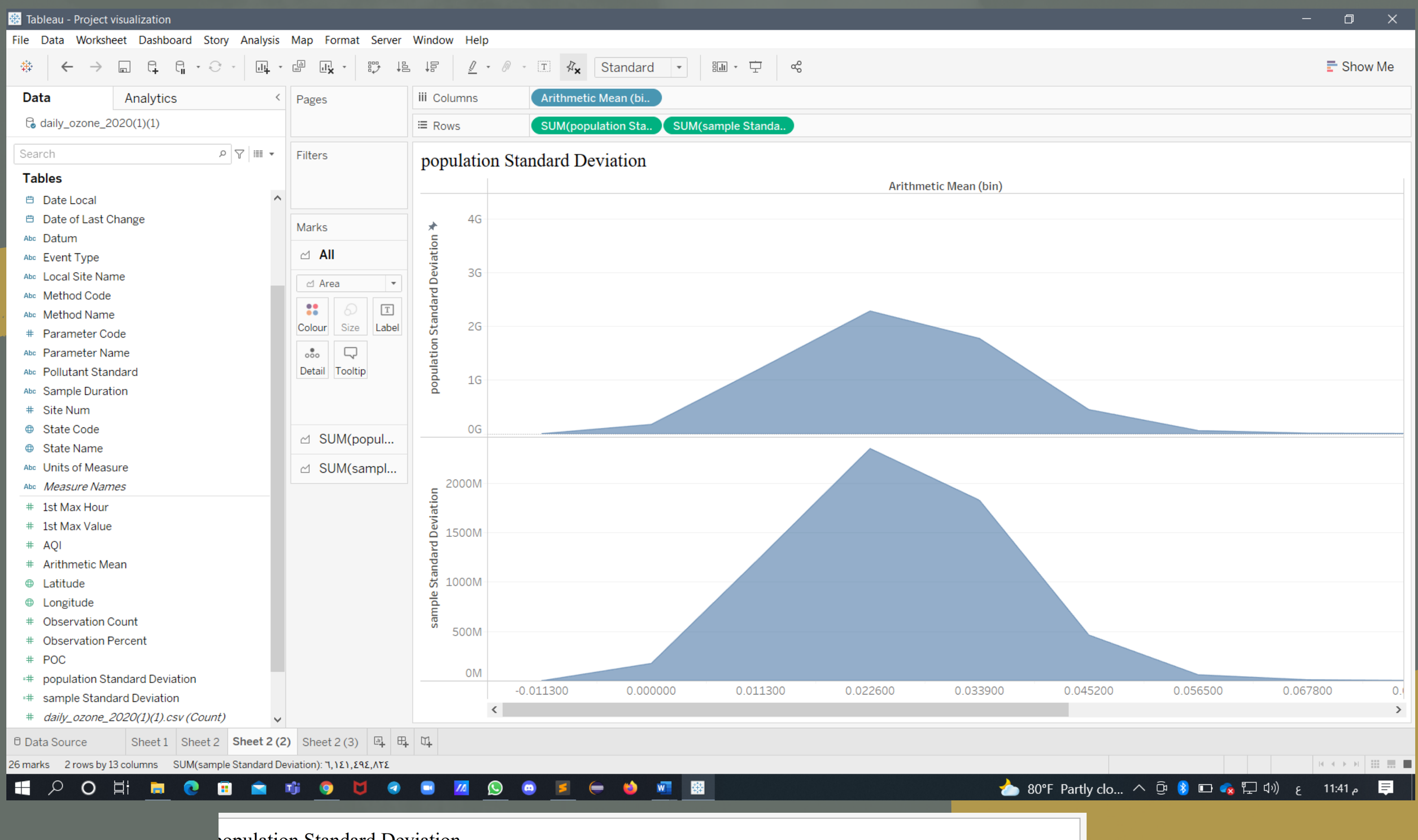
- **Plotting this data shows whether it's distributed normally or skewed.,We used the additional variables created before Answering**

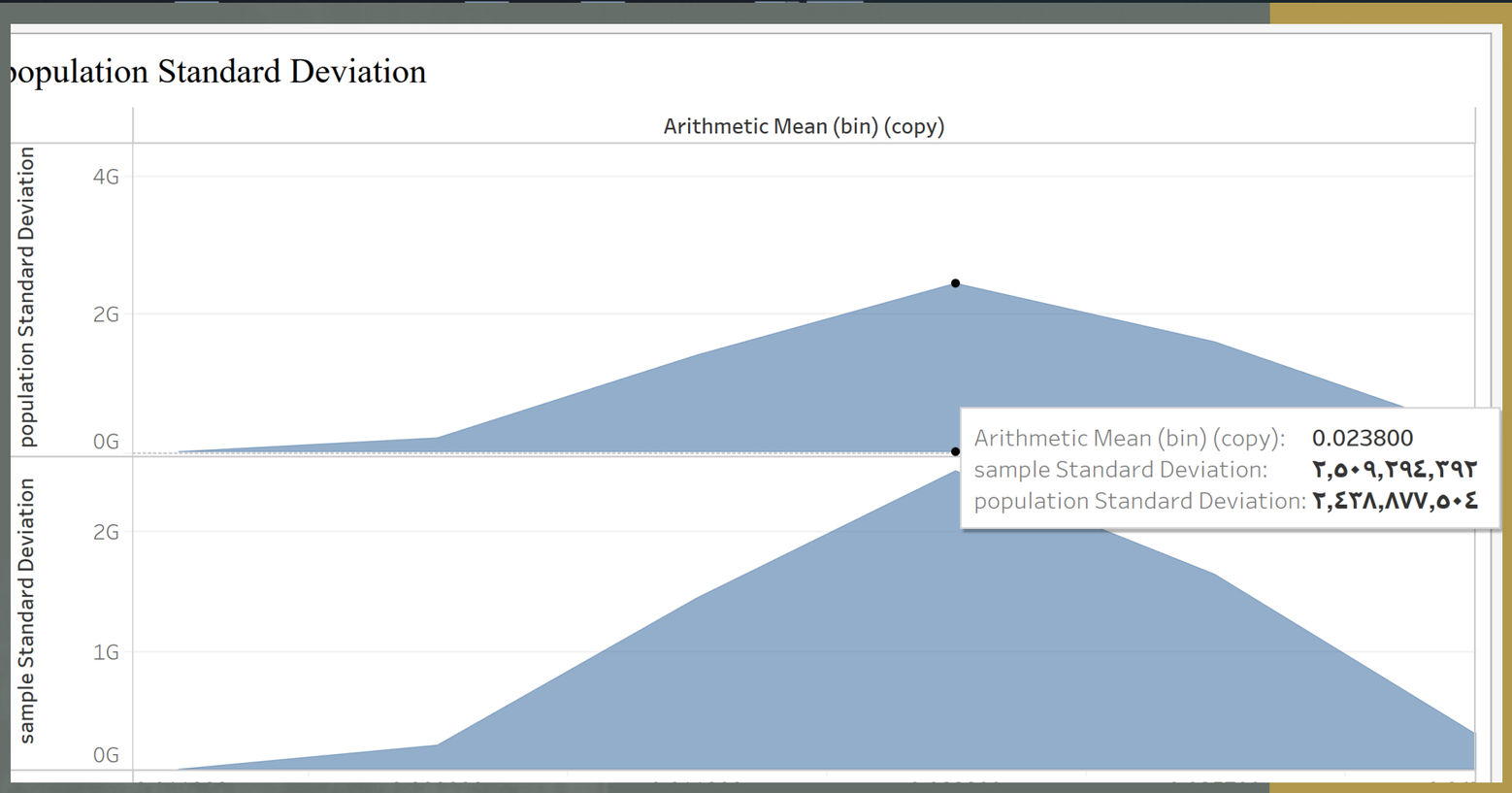


single variables visualization

In this Area Graph, We used the additional variables created before Answering the Question Find out the distribution of data is normal or skewed?.

As seeing the data is has a high standard deviation which indicates data are more spread out. This gives an answer..the greater the value of the standard deviation, the greater the value of the dispersion.

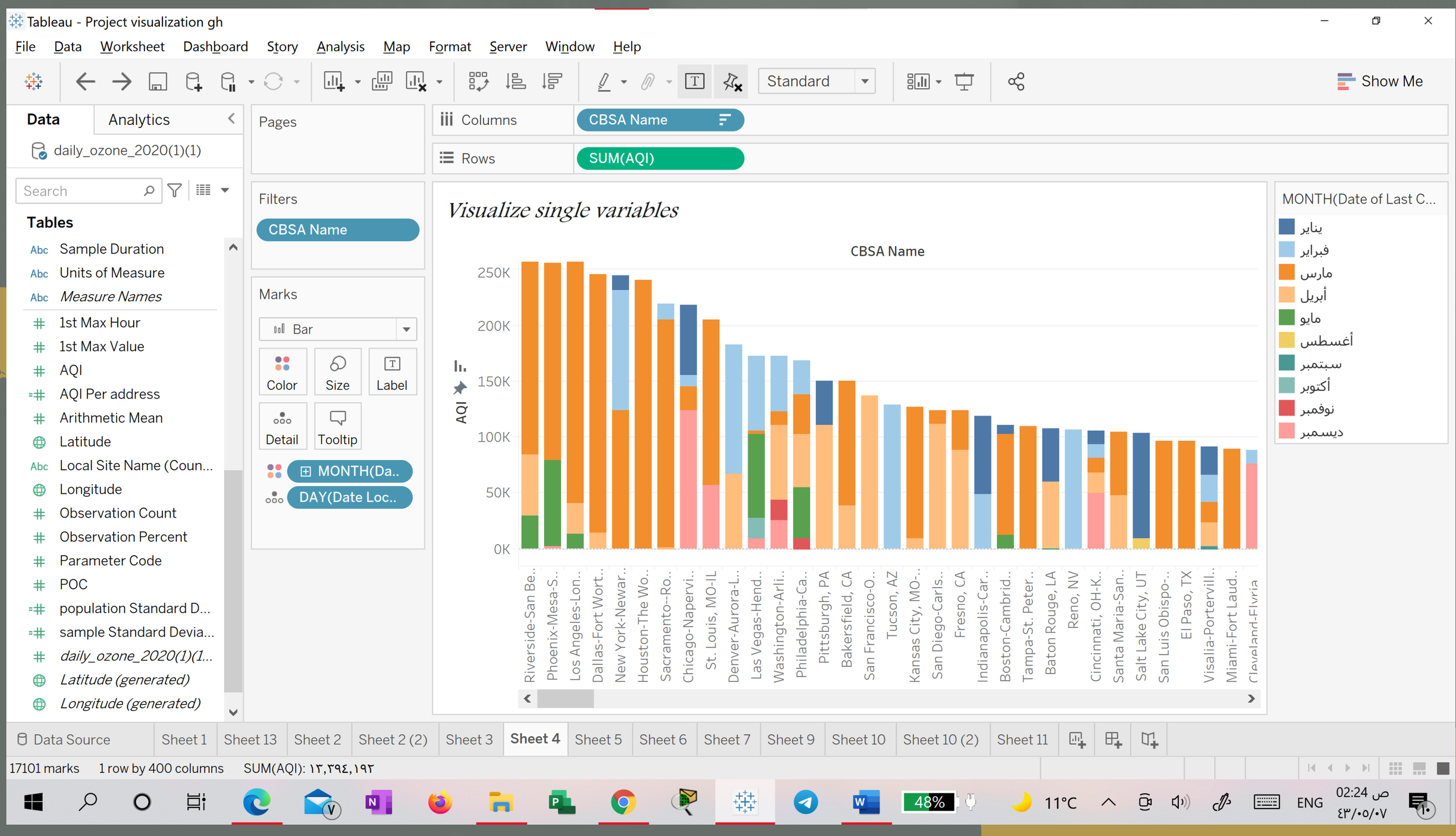


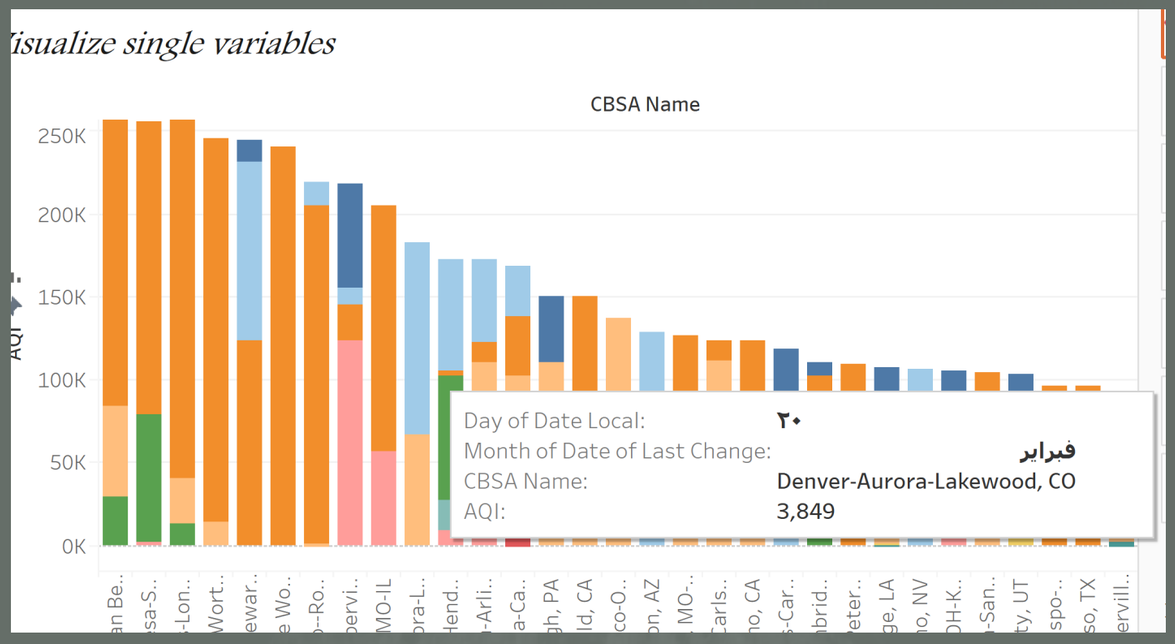


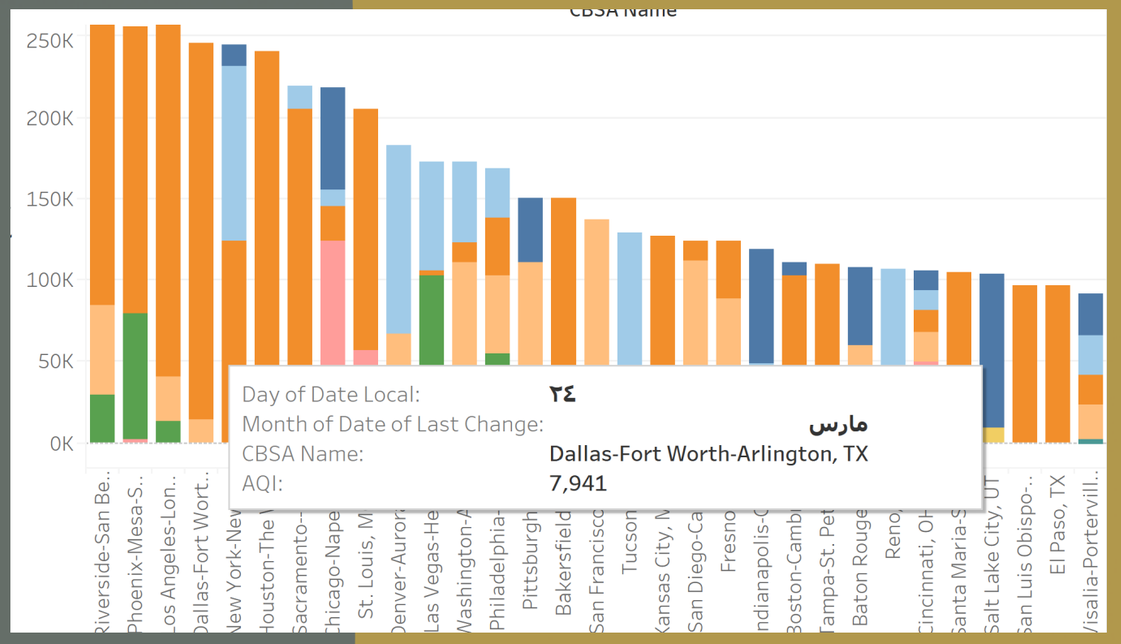
single variables visualization

This Bar plot Answering the Question of What is the benefit of the Air Quality index depending on the monitoring site is located "CBSA"? Using one categorical and one continuous attribute and Filter Nulls in CBSA.

the visualize Show us the daily SUM(AQI) for each CBSA per month(last modified date) which helps in a lot of decision making.





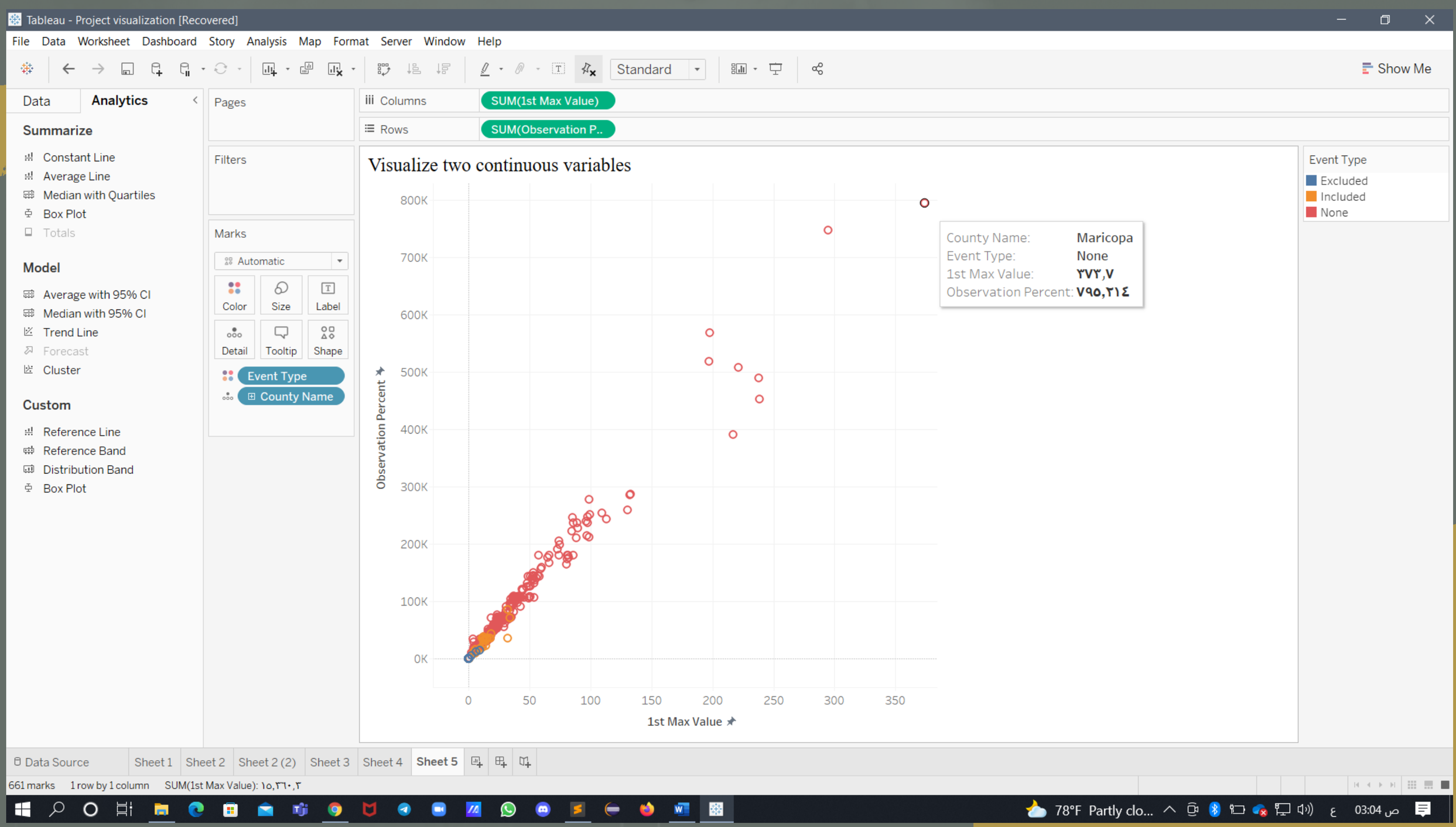


Single Variable visulize

This Line Chart Answer the Question Determine the air quality index per day? which also we used our hierarchies Address Number and Address String created before. to see the air quality index per day and determine the highest rates

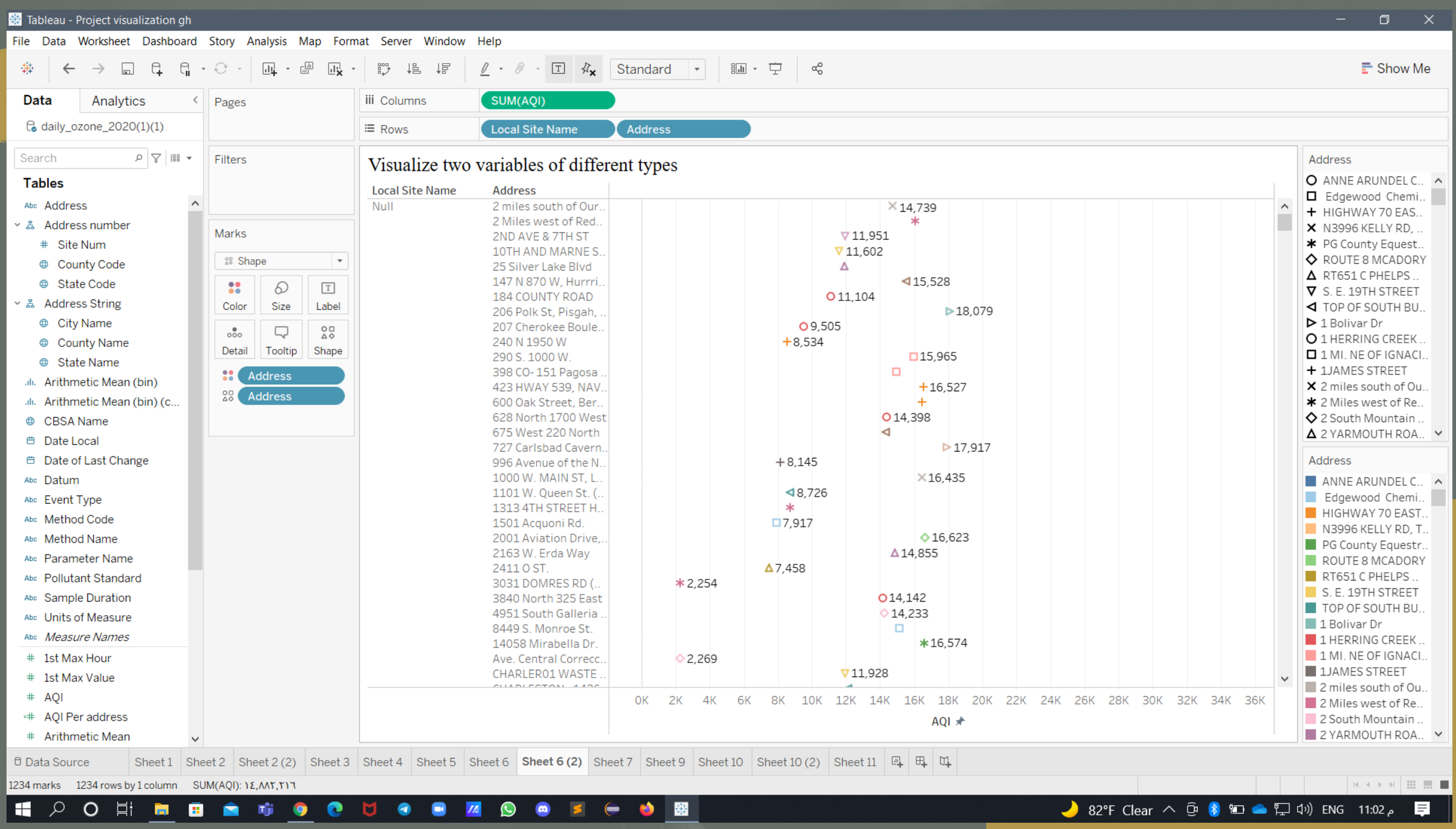
two continuous variables Visualize

This Scatter Plot Answering the Question of determine the highest value recorded from the number of observations taken during the day. Using two continuous variables Visualize.



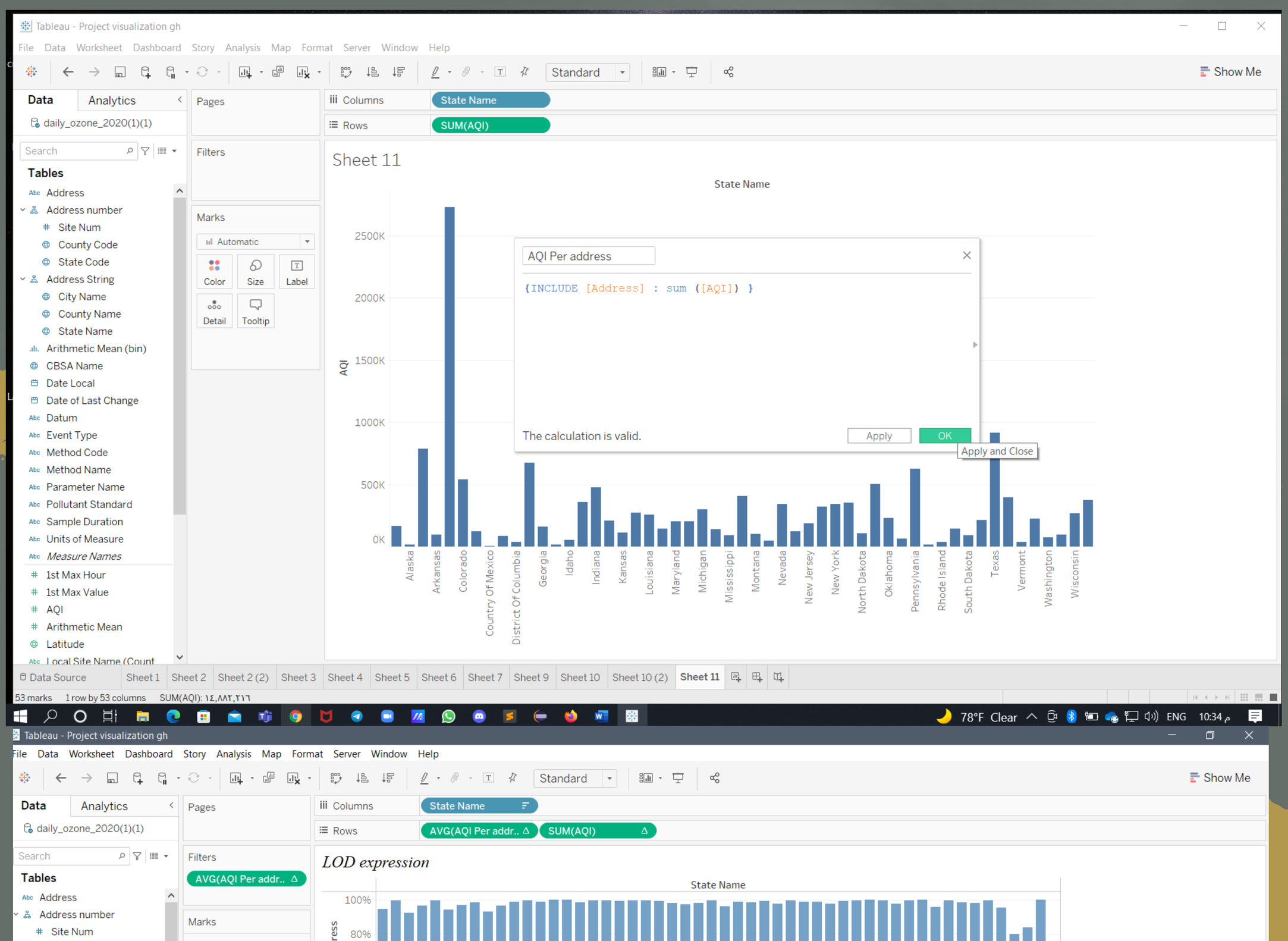
Visualize two variables of different types

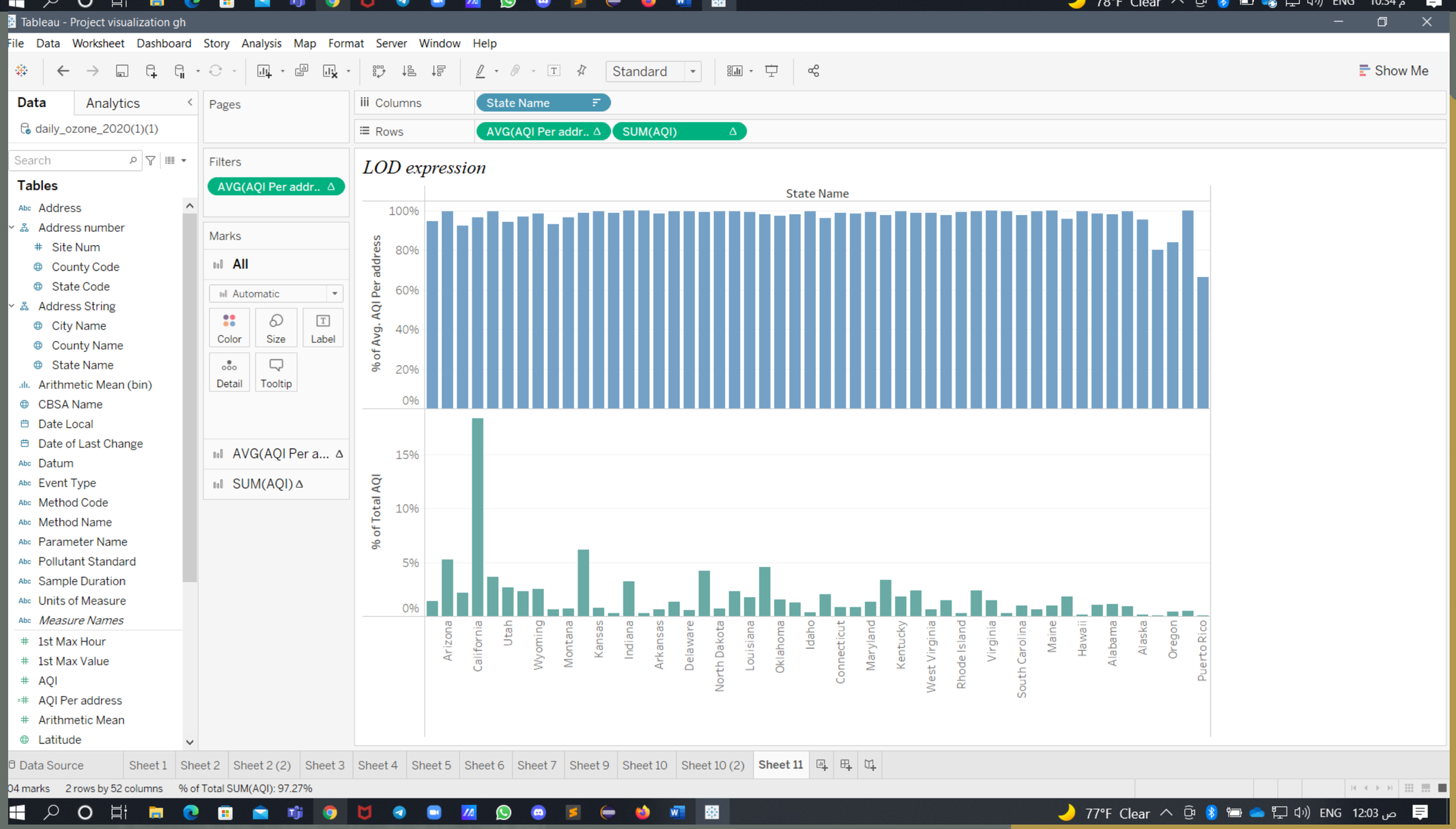
this can take advantage of the local Site name provided by the Air Pollution Control Agency and the addresses to determine the quality of the air indicator



LOD expression

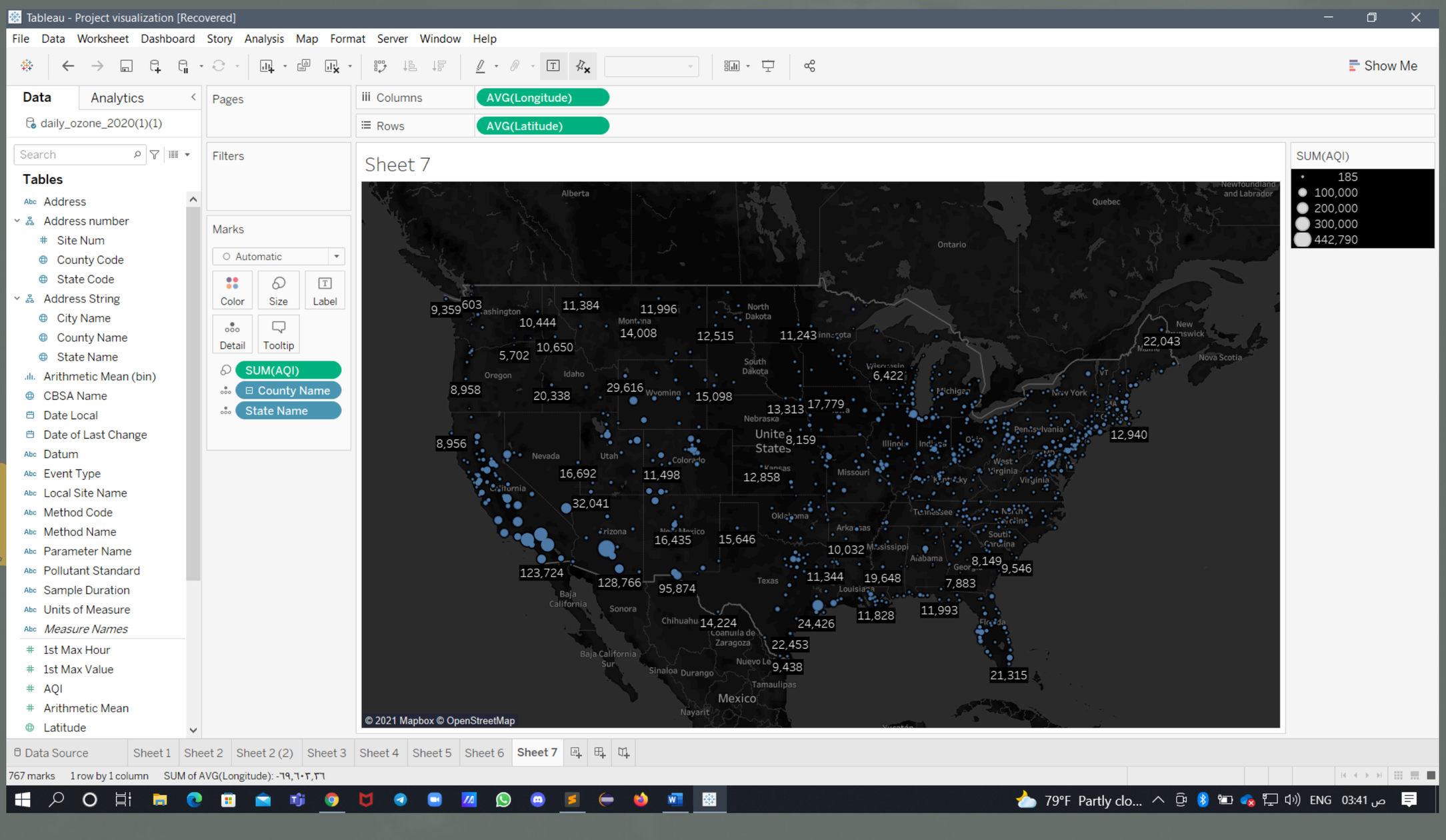
- *This LOD the Answering the Question of determining the air quality index for each state. Using Two calculations for sum AQI per address and AQI.*

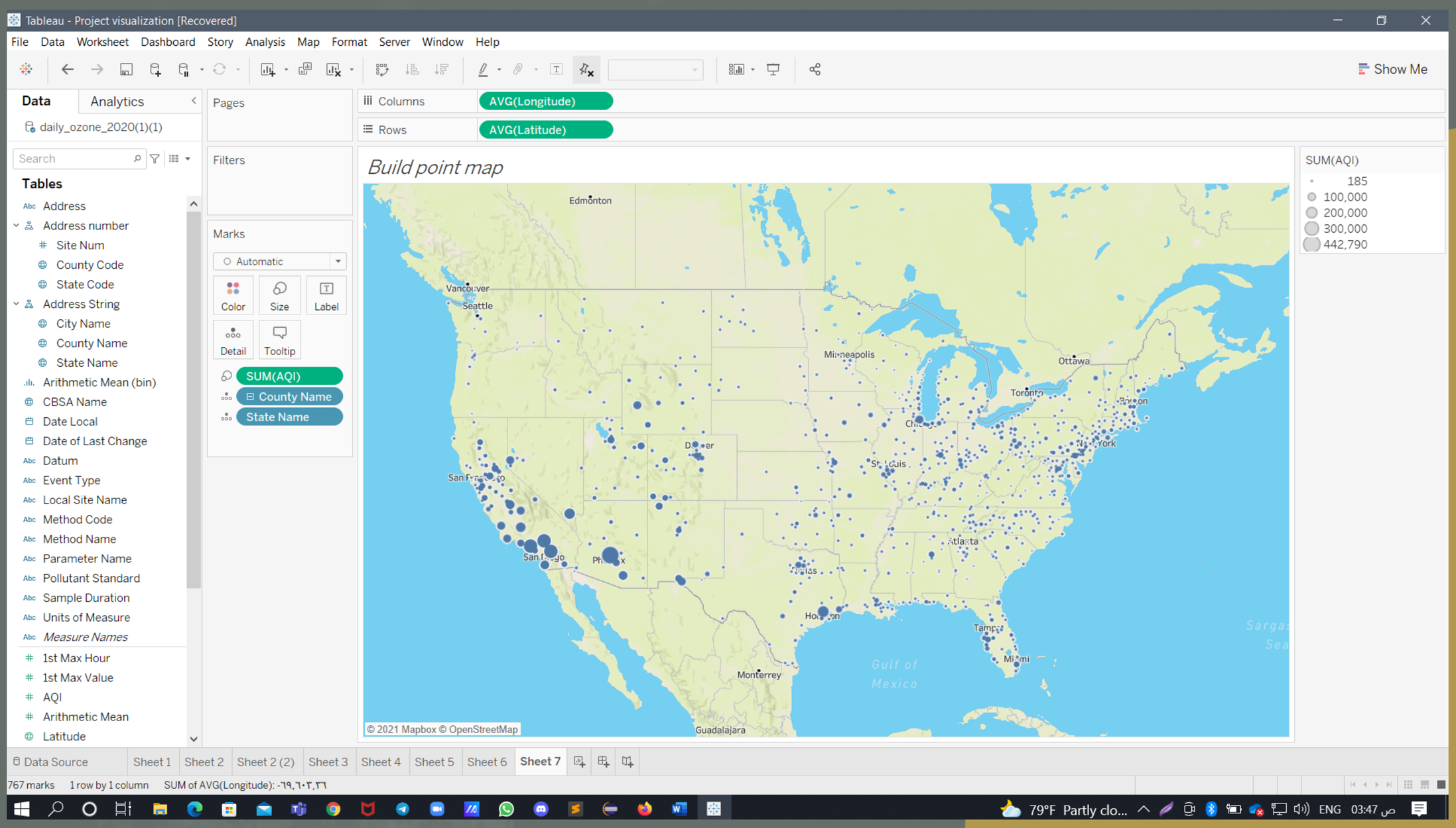




Build point map

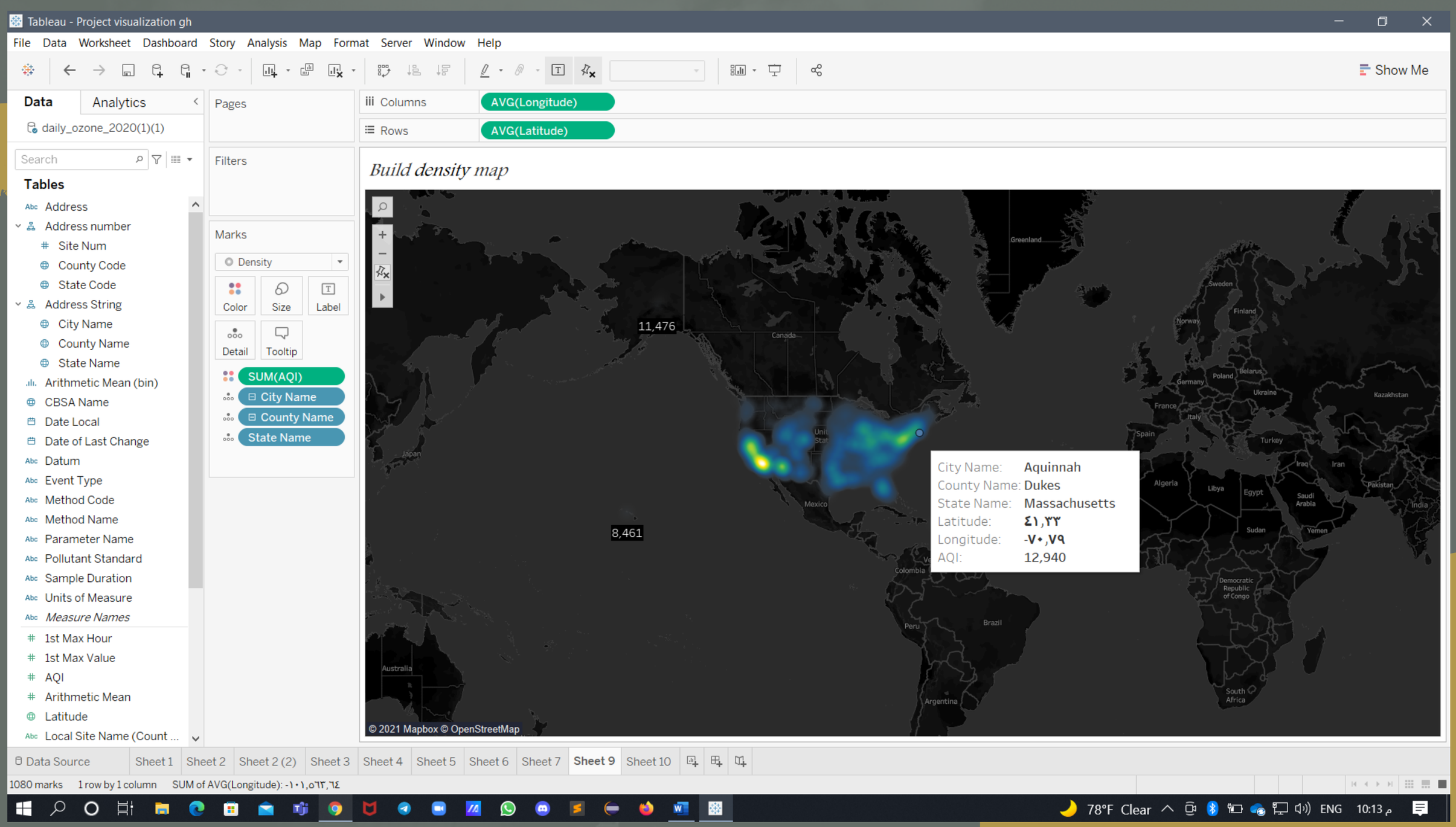
- this Maps , Answering Question of Determine the air quality index for each state . but by using some different details





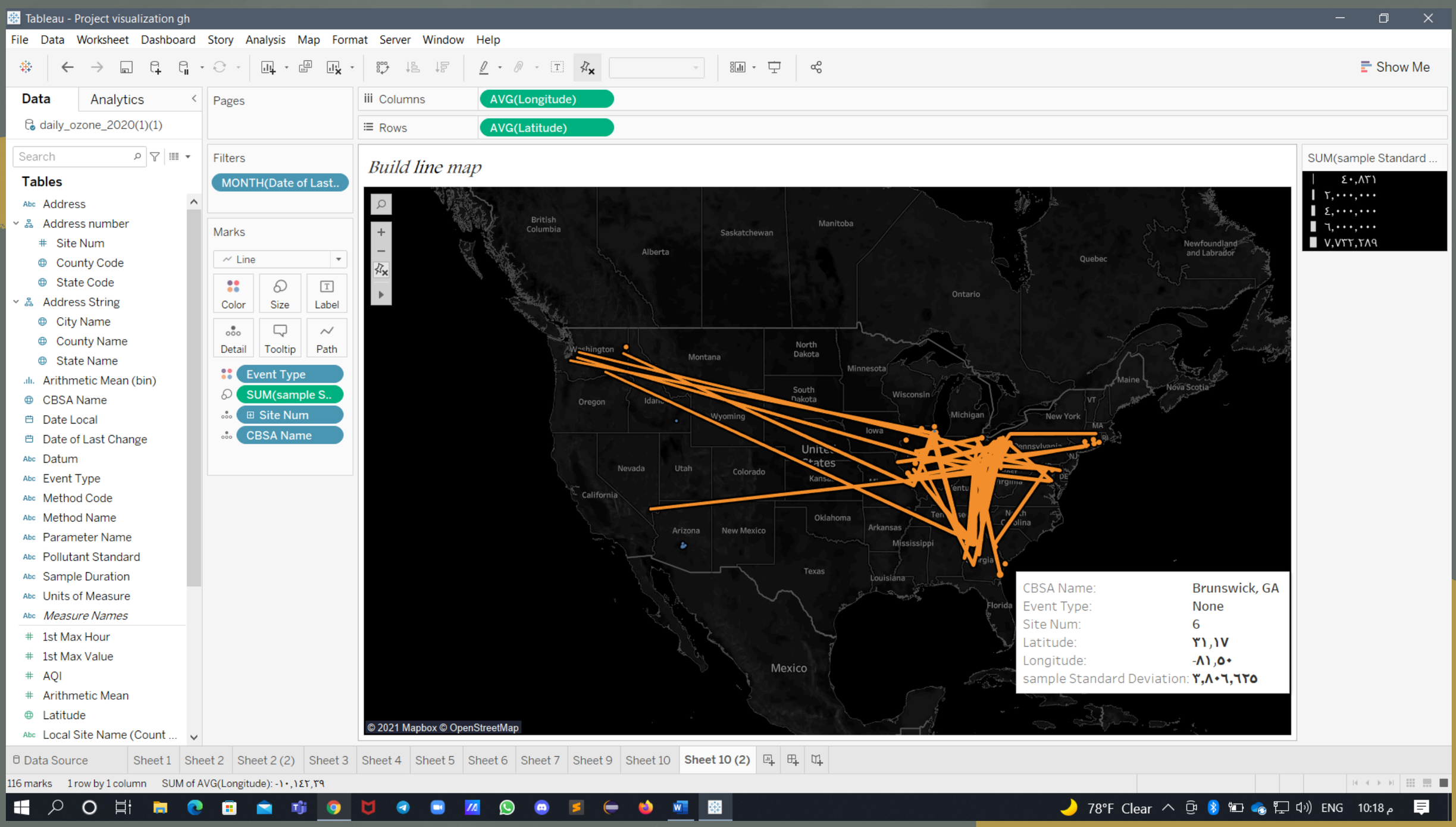
Build density map

- this Maps , Answering Question of Determine the air quality index for each state . but by using some different details



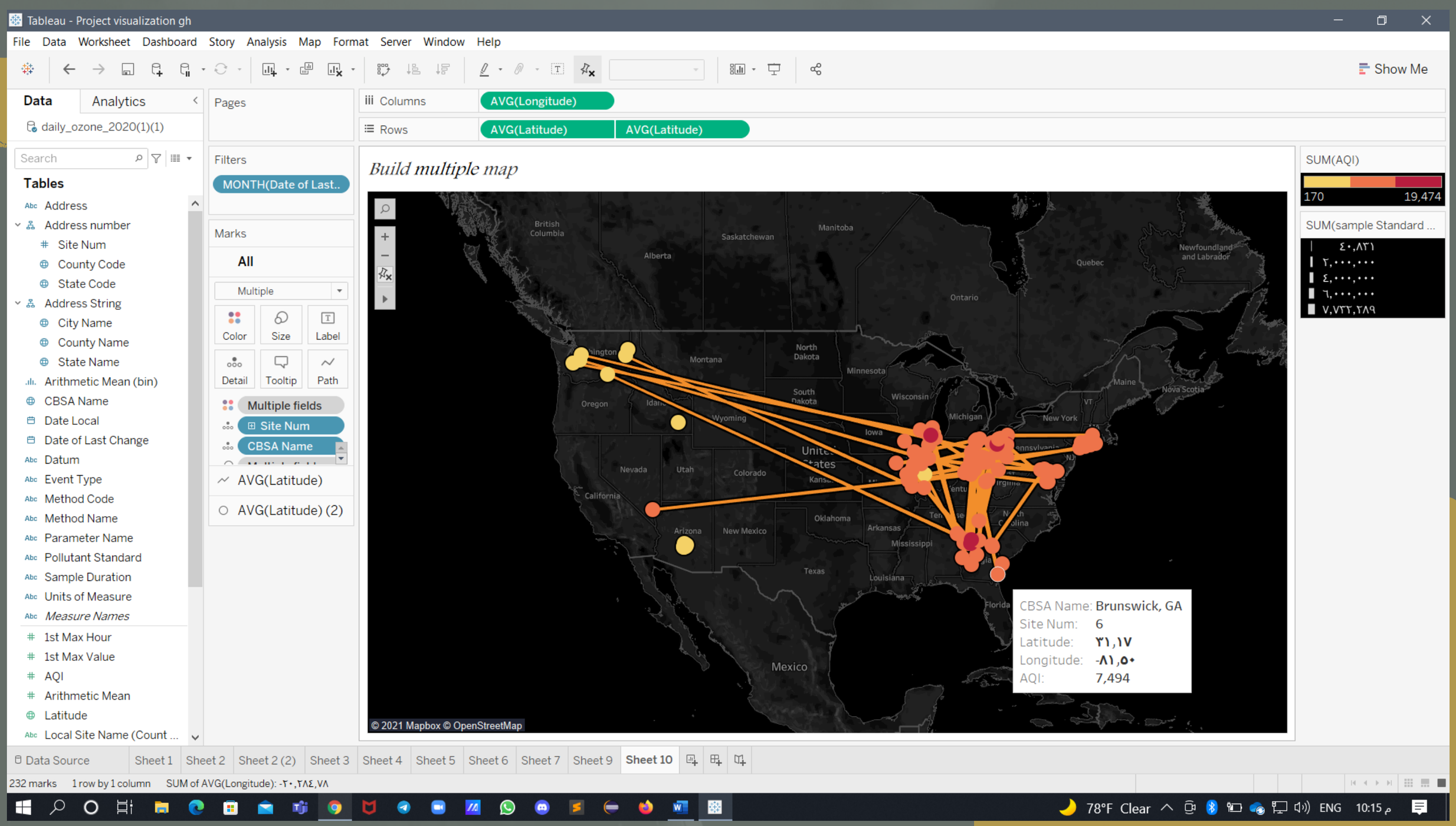
Build line map

- this Maps , Answering Question of Determine the air quality index for each state . but by using some different details



Build multiple map

- *this Maps , Answering Question of Determine the air quality index for each state . but by using some different details*



visual attributes in the visualizations

Visual variables we used:

1-Color

2-Size

3-Postion

4-Shape