LAB : CLUSTERING

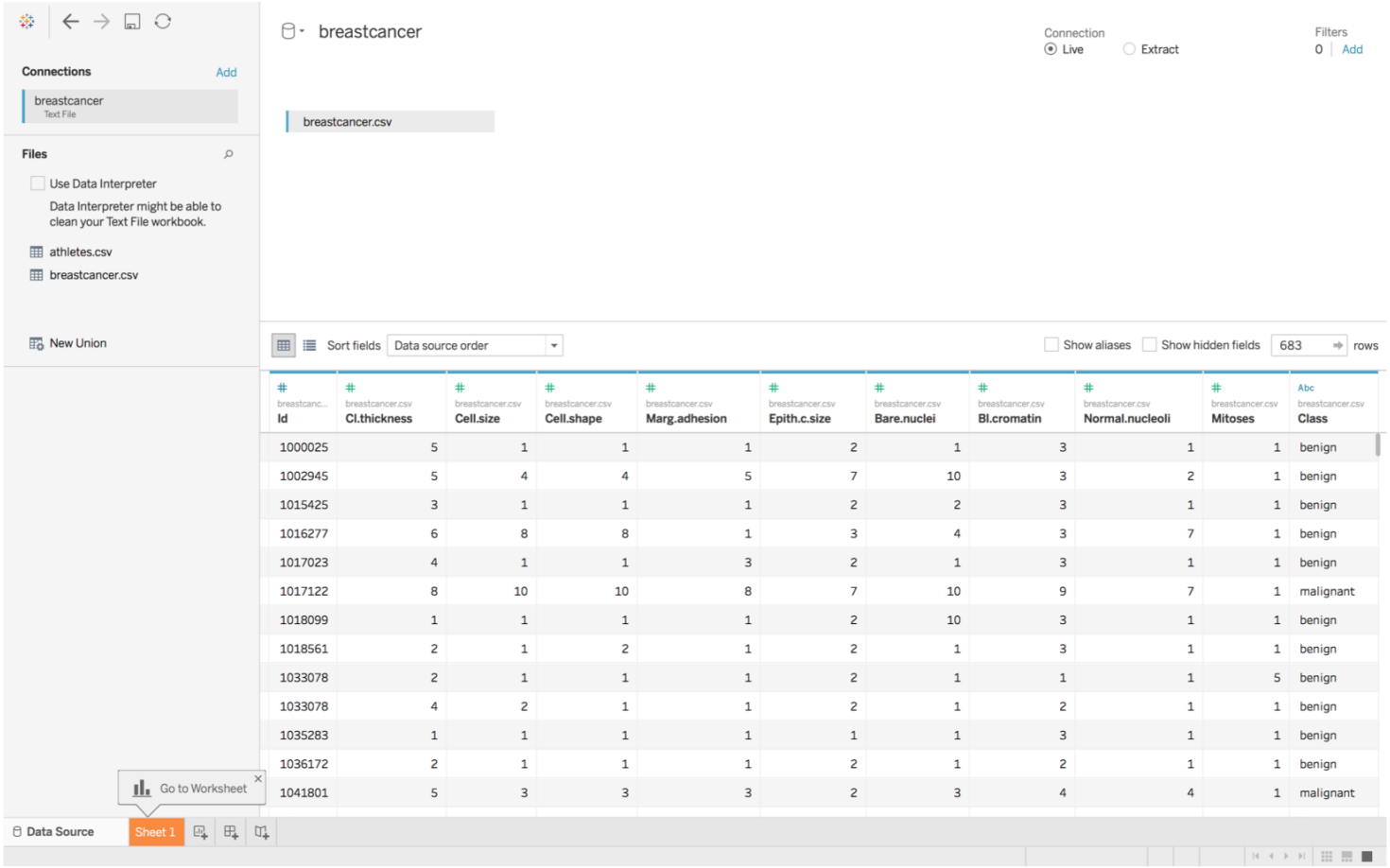
Clustering in Tableau

- Choose a dataset
- Choose the attributes for clustering
- Choose number of clusters

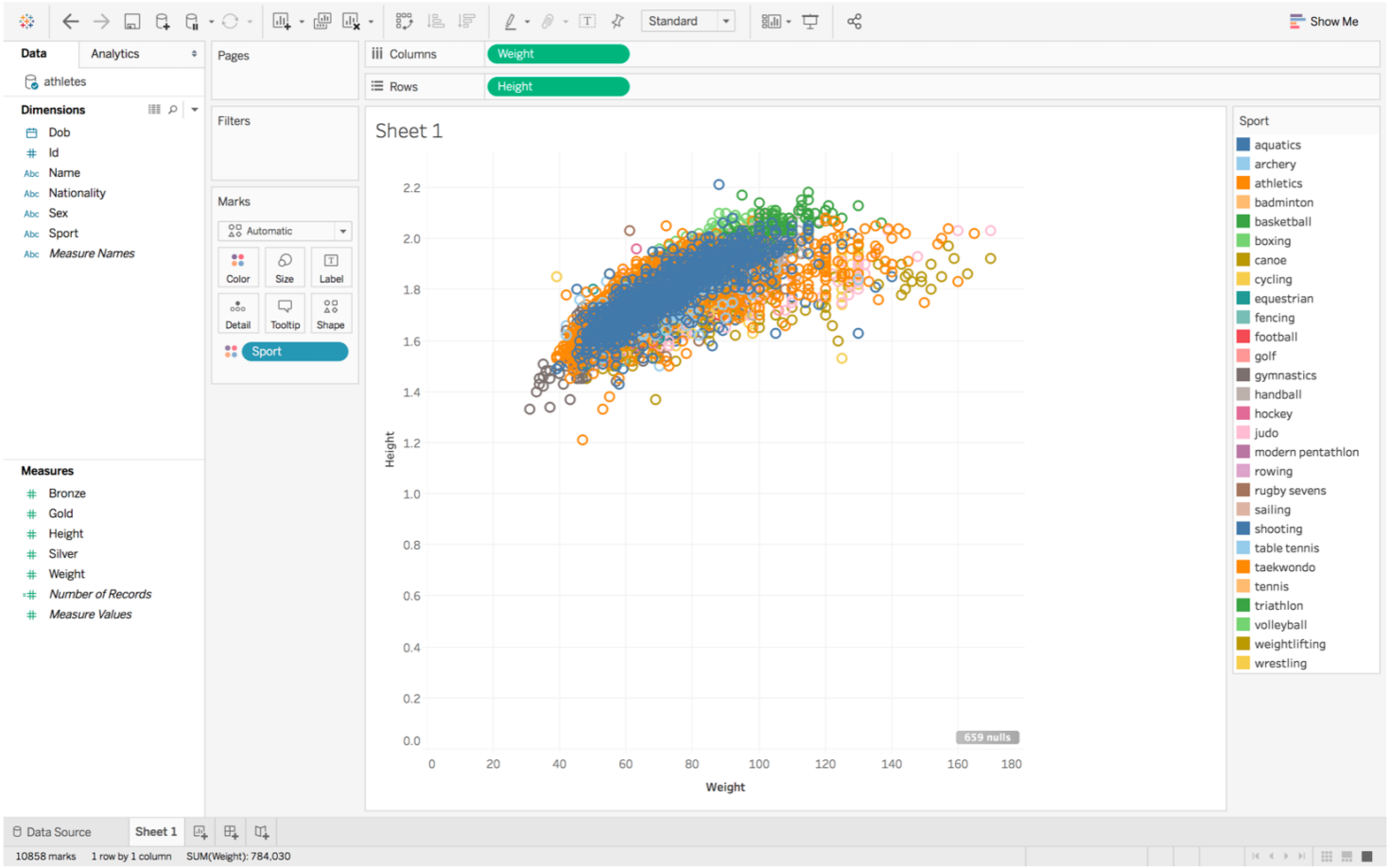
(Uses K-Means Clustering)

Clustering in Tableau

(using athletes.csv file)

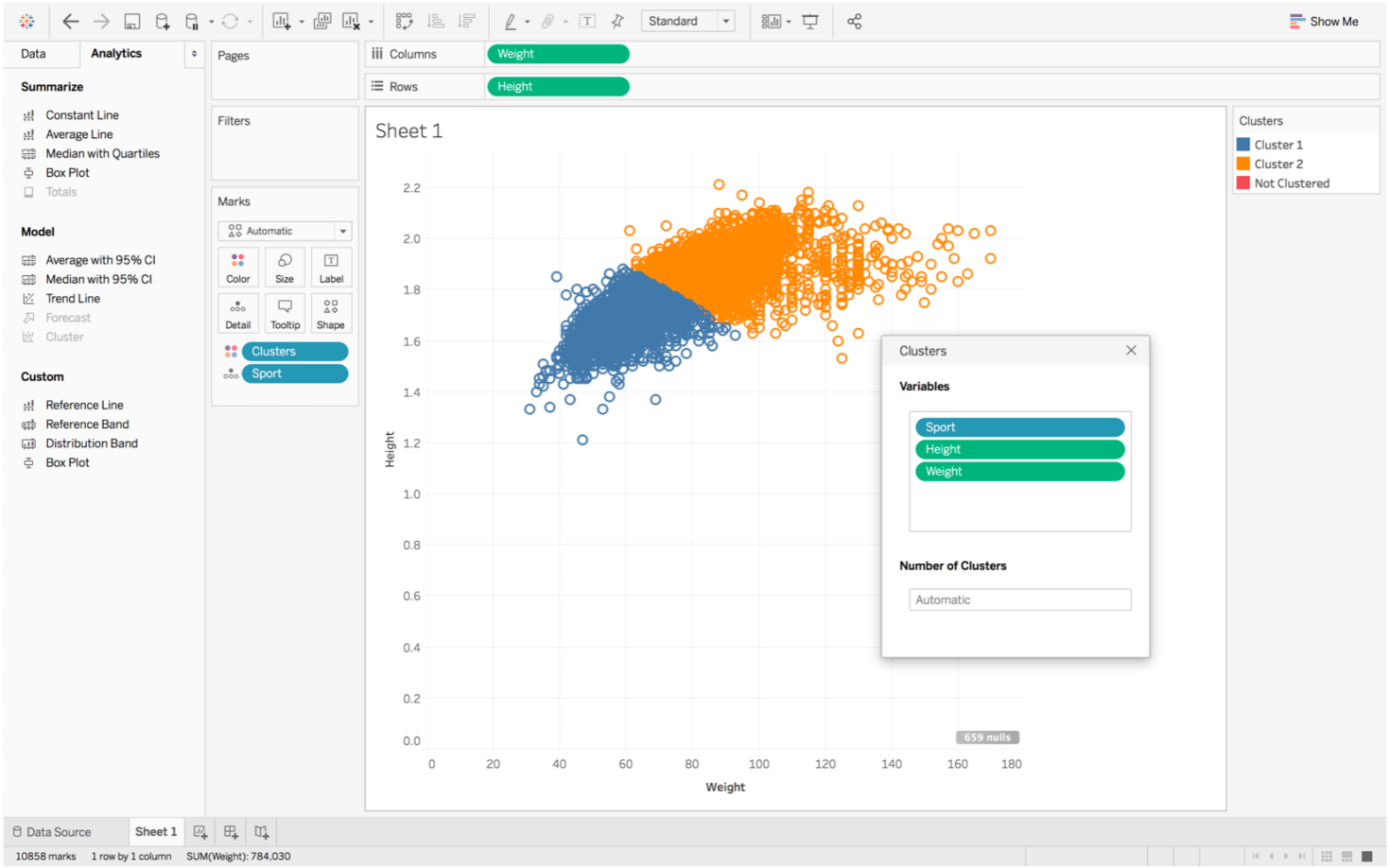


Clustering in Tableau



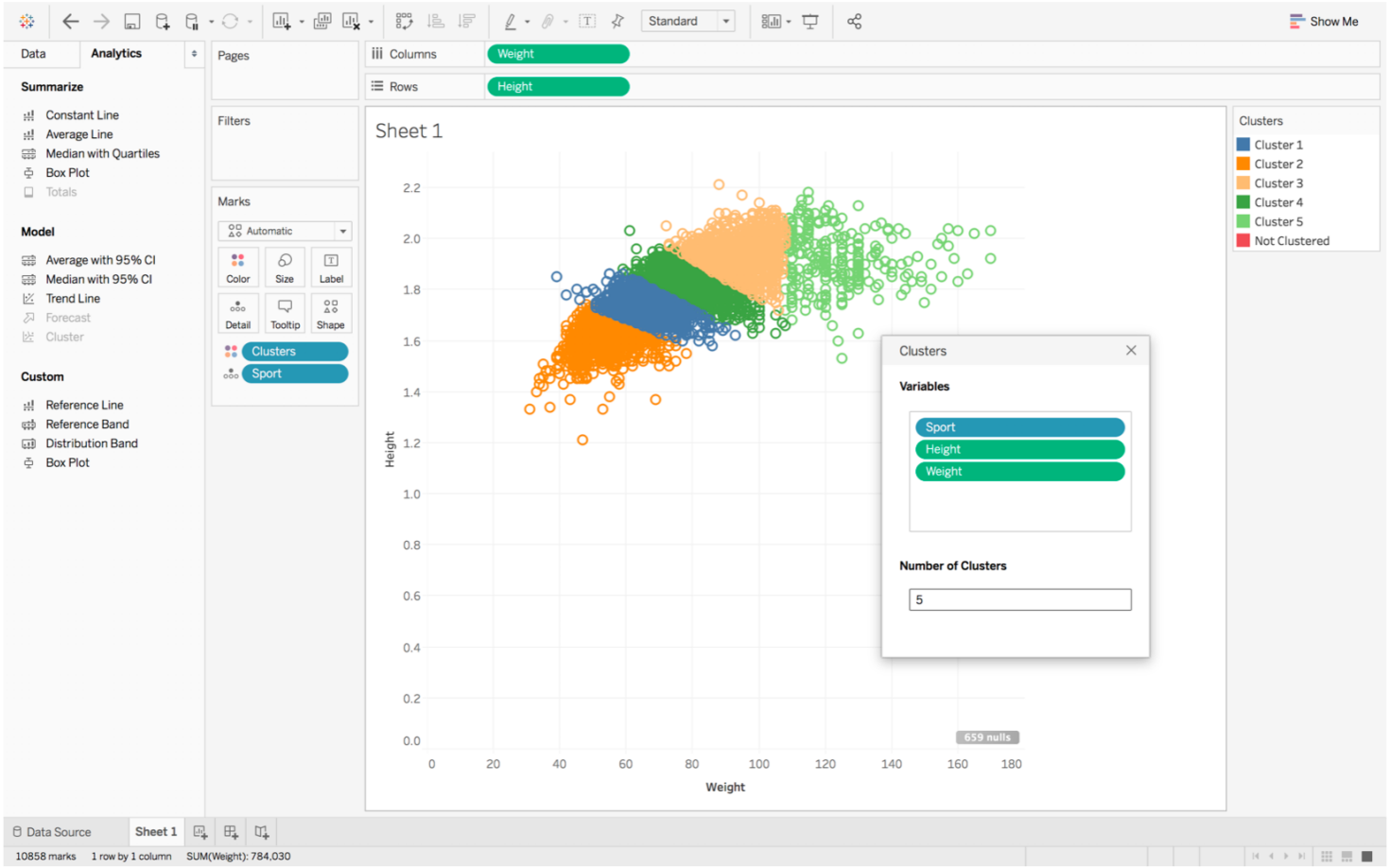
Clustering in Tableau

(Double click on clusters in Analytics tab)



Clustering in Tableau

(Changing number of clusters)



Hierarchical clustering

- Look into the Jupyter notebook: ‘Lab Clustering.ipynb’

Tabpy and DB Scan

https://github.com/tableau/TabPy

Tabpy Installation (Mac/Linux)

- Create a new conda environment

>>conda create --name Tableau-Python-Server python=3.5 anaconda

>>source activate Tableau-Python-Server

>>pip install tabpy-server

>>pip install --upgrade tabpy-server

Go to this directory: (For example) /Users/username/anaconda/envs/Tableau-Python-Server/lib/python3.5/site-packages/tabpy_server. You should find setup.sh file, just run it

>>./setup.sh

let this run……. It is just that you are running your server, don't stop it. (You can stop it by using control+C)

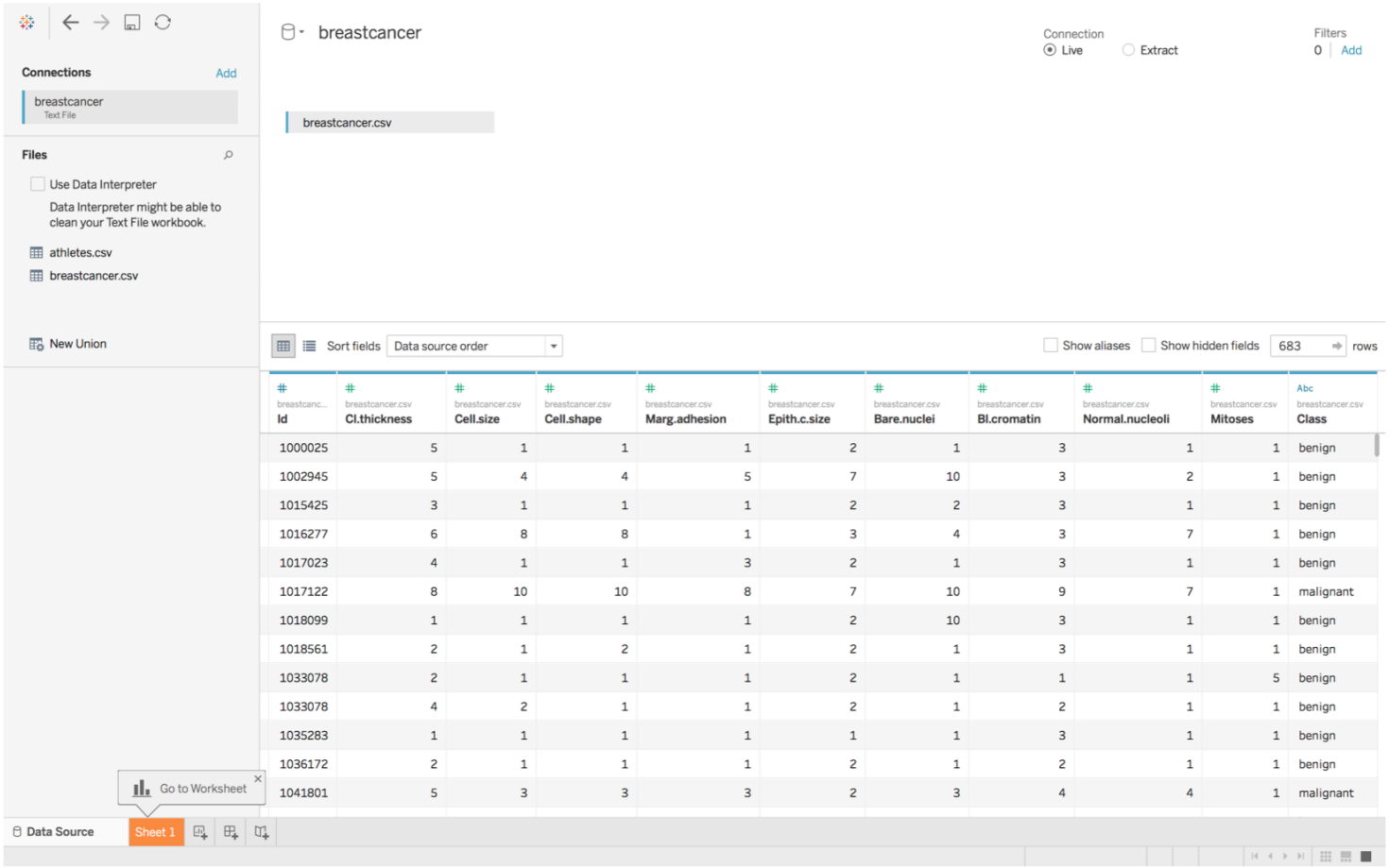
Execute below command in another termal

>>pip install tabpy-client

Tabpy

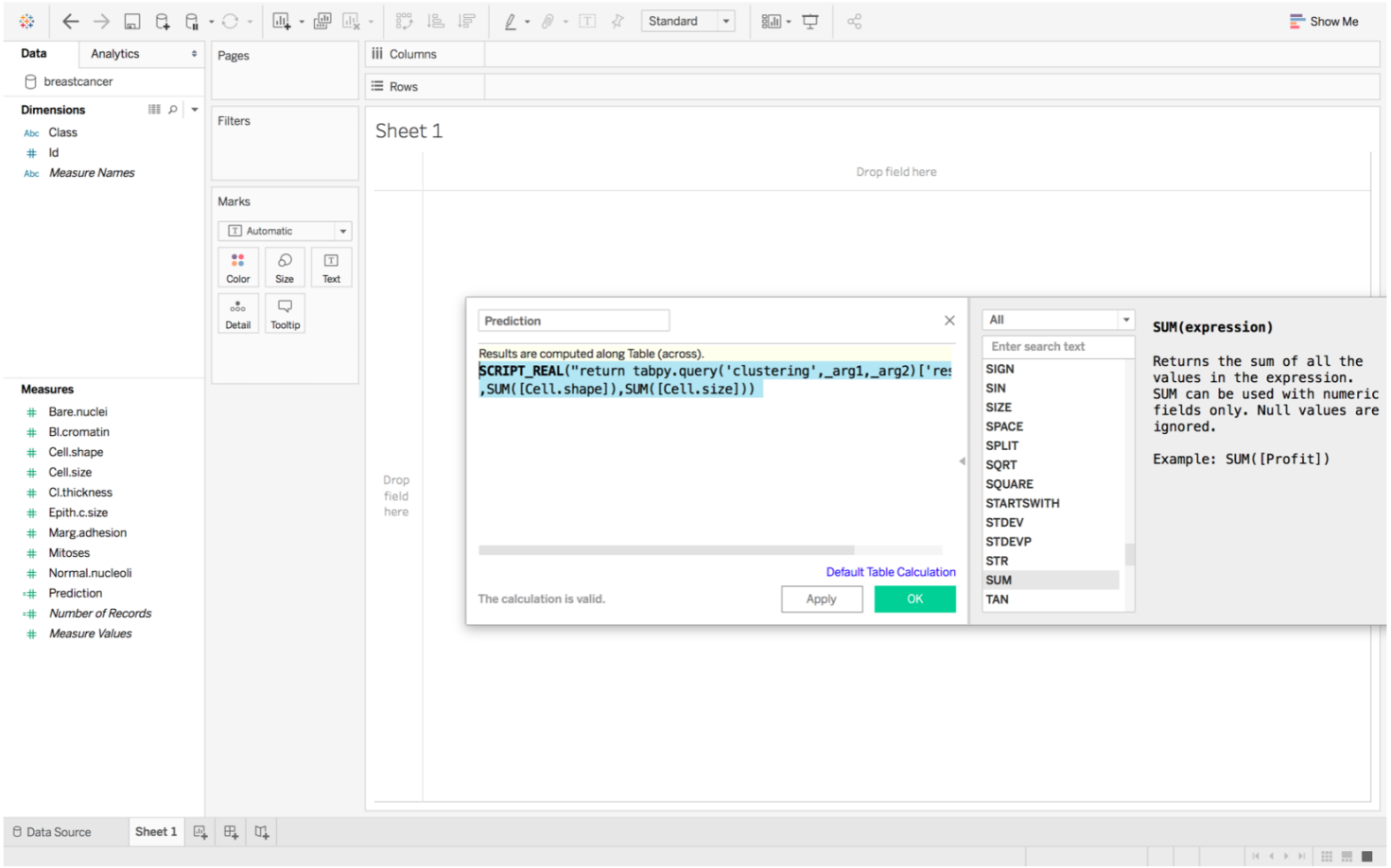
- Look into ‘tabpy_clustering.ipynb’

Using breastcancer.csv file



Create new calculated field ‘Prediction’

SCRIPT_REAL("return tabpy.query('clustering',_arg1,_arg2)['response']",SUM([Cell.shape]),SUM([Cell.size]))



Prediction field in action