Stage\_5: Exploration & Analysis

Once you created a simple spreadsheet/dataframe containing the 354 largest airports, you start adding analytical column heads as shown below. This can be done in various ways:

A screenshot of a computer

Description automatically generated

- using Pandas in a Jupyter notebook

- using functions in Excel

- clustering was done simply by adding a column and the sorting the airports and assign them into correct cluster

- adding data from other sources such hub category and EAS airports. Those datasets are in the Stage\_1 folder

- creating new datasets such as calculating the share for 50-seater regional jets of all the departures in the airport

- the geolocation data has been generated with the Geocode by Awesome Table extension in Google Sheets. Note that some airport names are open to various interpretations by the extension. For example, “Jackson Regional” can refer to several locations. So you’ll have to check each location and correct faulty ones manually

All the datasets and code notebooks used in creating them are in the folder above10K. You can explore them or create your own datasets