Stage 4: Annual totals & Filter

We now have a four annual time series tracking departures and passengers for F Class in total at 813 airports in the United States. The files (for pax) look like this:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

We now have next final two steps to take. First one is to create annual totals for each year (2019, 2020, 2021, 2022) and then filter the dataset down to airports excluding Alaska, Hawaii, Puerto Rico, Virgin Island and Trust territories and then include only airports that had more than 10,000 annual departing passengers either in 2019 or 2022. This accomplished in the notebook annual\_totals\_pax

1) Using the list of airports in 'airport\_codes\_master\_analysis.csv' it loops through the annual time\_series\_pax\_airport\_{year} files and creates an output\_df dataframe with annual level data on 813 (some have passenger data only for some years)

2) As an intermediate step it saves the annual totals as 'total\_annual\_pax\_airports.csv'

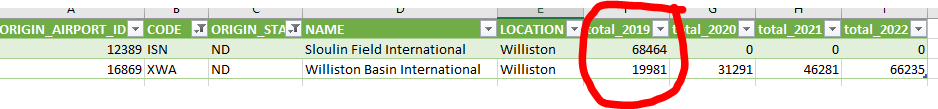
3) It then takes output\_df and excludes Alaska, Hawaii, Puerto Rico, Virgin Island and Trust territories and includes only airports that had more than 10,000 annual departing passengers either in 2019 or 2022.

4) Output\_df is now down 355 airports. The notebook saves it as 'above\_10K\_total\_annual\_pax\_airports.csv'

5) One important final step: city of Williston is represented in the data by two airports: Sloulin Field and Williston. That is because Sloulin was closed in 2019 and the new Williston replaced. It is logical that Williston is represented only by a single airport, and to make numerical analysis meaningful:

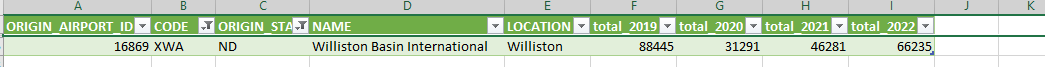
1. Sloulin’s data for 2019 must be added to Williston’s data
2. Sloulin must removed, and the result is 354 airport’s

This is done by hand in Excel, and the output file is saved by the same name.



A screenshot of a computer

Description automatically generated



Williston correction must be always done when raw data is processed to this point and you’re creating a baseline dataset for the 354 airports to be analyzed and processed. The matter is discussed in detail in the Williston\_update Word document.

7. Save the list as above10K\_total\_annual\_pax\_airports.csv. You have the annual passenger totals for the 354 airports.

Create a reference list of above10K airports for future use

1. Remove all other columns except ORIGIN\_AIRPORT\_ID and CODE, and you have list of the 354 airports. You can use to create dataframes in pandas when you want to filter raw/processed data to created datasets exclusive to the 354 airports. I created two lists, narrow above10K\_airports\_list.csv with just ORIGIN\_AIRPORT\_ID and broad with all the airport data above10K\_airports\_list\_broad.csv

Create annual totals for departures at all airport and at 354

1. You start off with same procedure as for the pax, but because of the Williston correction, you need to do one manual correction to the output csv file

2. Use the code in the notebook total\_annual\_deps

The output will be two csv files, the intermediate with all the 813 airports and the 354 above10K airports.

3. Now you have to just the Williston correction, that is add the Sloulin’s 2019 figure to Williston’s 2019 figure. Open both csvs, do the calculation, amend the Williston figure in above10K\_total\_annual\_deps\_airports and you’re done.

A close-up of a table

Description automatically generated

