Dataset Cleaning

# COVID\_Dataset

* Only row 2768 2769 is relevant to what we need from this dataset, as that’s the row that represents el paso.
* Manipulation will be required to filter only information from that row as well as ensuring that we correctly filter out the days that are from march and july.
* We may also have to filter out inconsequential data from both the first row (that shows the dates) and row 2768 2769 (which shows el paso information)
* Another thing I wanted to note: it seems like this dataset only shows the total number of infections (including the ones who have recovered, or worse I presume). Not sure if it’ll matter but maybe we might need a dataset that contains purely the current number of active cases in el paso as well?

# Flight Dataset1

* We once again need to filter out for all flights that have their destination as El Paso
  + Perhaps even those departing from el paso as well just to be on the safe side (i.e if the analysis team decides that that would be useful information later down the line).
* Out of those flights that are to or from EL Paso: we also need to ensure that they are not cancelled

Done:

* March Flights: Flights\_Dataset1\_March.txt
* July Flights: Flights\_Dataset1\_July.txt
* ELP March Flights: elp\_flights\_dataset1\_march.txt -> includes in/outcoming flights
* ELP July Flights: elp\_flights\_dataset1\_july.txt -> includes in/outcoming flights

# Flight Dataset2

* .ASC can be read by a traditional text editor but would likely need to be converted to be digested and manipulated by python or something else easily.
* This dataset differs from the other dataset as it actually has the number of passengers on each flight, meaning that we could potentially extrapolate a heat map of most passengers per day sometime later down the line
* Since this also includes origin and destination: we can filter it out as we did the first flight dataset.

Done:

* March Flights: Flights\_Dataset2\_March.txt
* July Flights: Flights\_Dataset2\_July.txt
* ELP March Flights: elp\_flights\_dataset2\_march.txt -> includes in/outcoming flights
* ELP July Flights: elp\_flights\_dataset2\_july.txt -> includes in/outcoming flights