# Filtering:

For the filtering MapReduce Pattern, I decided to use the Forest Fires dataset provide here <https://archive.ics.uci.edu/ml/datasets/forest+fires>

Preparing the dataset was nice and simple, the data was provided as a .csv file, so formatting the original data was not need, I would have accepted a .data, .csv or a .txt file as long as the .txt file was separated by a consistent value.

The only preparing I needed to do before the MapReduce Pattern was splitting the file into two separate files to speed up the process.

The question I wanted answered with my MapReduce Pattern was “What are the top ten windspeeds documented on forest fires? And what month are they in?”.

I wanted to find the ten highest wind speeds from forest fires and the month they happened in to see if there is any correlation to them

Using a Mapper file, I was able to retrieve ten of the highest wind speeds from each of the

Split files I had created and passed them into an output file.

Then using the Reducer file, I was able to narrow the twenty speeds I had gotten previously into ten and their months.

The output is as followed (all speeds are in km/h):

A black sign with white text

Description automatically generated

As you can see, although the fasted speeds recorded are within spring/early summer, December seems to receive high speeds the most consistently.

This data could be used to track when to consistently stack up on supplies, i.e. come November we should be stacking up on water supplies as higher winds tend to spread forest fires quicker.

# Summarization:

For the summarization MapReduce Pattern, I decided to use the Adult dataset provided here <https://archive.ics.uci.edu/ml/datasets/Adult>

Preparing the dataset was nice and easy again as the file given was a .data file, as previously stated this is what I would prefer.

The file however was very large, roughly thirty-two thousand lines, however I did not feel this was too big to warrant a reduce, so I just split the file in two.

The question I wanted to ask was “What is the average, min and max ages of the adult dataset provided?”

A black sign with white text

Description automatically generated

After running through the Mapper and Reduced I found the min, max and average age.

I was very surprised to find someone as old as 90 and as young as 17 in the data.