Welcome to the Interview Project. Our website is Planephd.com, and if you google an aircraft, such as “Cessna 172d” you will find us in the #1 position, above Wikipedia. The test project is one which I did myself a few weeks ago. It probably took me well over a day, but I did not have any background in the techniques I used. The data will not be perfect; that’s ok. The goal of this exercise is to see what you can achieve within only a few hours of your time. It’s ok if you only have the beginnings of a solution. We simply want to see how you work.

Attached are three files:

* This one
* GIV.csv which is a text version of advertisements for a specific model of aircraft for sale.
* TargetData.xslx – This file shows the target fields in Red and some example data I got. This file only shows the target fields. The objective is to get these fields for both the Base Table and the Engine Table for all 700+ aircraft in the GIV file. Again, we don’t expect a finished result. Show us what you have after a few hours of work.

You can find examples of what the original advertisements looked like at websites like:

<https://www.controller.com/listings/search?Category=3&Model=GIV&Manufacturer=GULFSTREAM>

There are other similar websites with for sale. DO NOT SCRAPE ANY WEBSITES. I simply wish for you to take our data in the files supplied and extract with as much accuracy and completeness as you are able, the fields colored in Red. This is not intended to be a manual exercise, as this is a test. The real data will comprise 200,000 ads initially, which cannot be done manually.

Note that the GIV.csv file has a field named body. Body is a very long text field which describes all details of a single aircraft for sale. It’s only text, so it is completely unstructured.

The engines: The purpose of learning the details of the engines is to understand the mechanical state of them. Engines require an Overhaul every 8000 hours of use (depending on the engine). Usually around 4000 hours they will have an HSI or mid-life inspection to ensure there are no problems.

The overhaul for a single engine costs millions of dollars, and so it’s important to understand how much time is left in these engines. We are doing fleet wide analysis of aircraft, and we attempt to value each aircraft with a proprietary algorithm. For that, we need clean data to understand the value in the engine. How much time in hours or years is remaining before an overhaul.

Some engines are on a Maintenance Plan. This is where the aircraft owner will pay someone like GE, JSSI, or Rolls Royce a fee for every hour that the engine is run. When it’s time to overhaul the engine, that company will pay for the overhaul. This is like a pre-paid maintenance plan, and it also acts as insurance in case of early engine damage. It’s very important to know whether an engine is on a maintenance plan since then the engine maintenance is prepaid.

I did this task myself recently using the Gulfstream G-IV as a test case, so I have an idea of what to expect. This is only one model. We have over a thousand models, and over 700,000 aircraft in our database. I would ask that you try to understand the aviation terms by googling them yourself to familiarize yourself with the concepts, but I’m here if you have questions. There isn’t a hard time limit, but we would expect that you should not spend more than a few hours on this, and then show us what you’ve accomplished. We won’t stop you from doing more, but it’s not required.

Good luck!

Markus