

Overview:

Each trade happening involves multiple players known as “Agents”. Shipper sends goods via air or ocean to a Consignee, by buying space from Airline or Shipping line. The agent who orchestrates this procurement of freight is called the Forwarder. Each party has its own Forwarder (referred as the Local Agent) and the opposite Forwarder (referred as the Overseas Agent).

Airlines have authorised Forwarders to issue proof of carriage, by issuing documents known as Air Waybill (AWB). This document carries name of all the stakeholders. Further, there are two such documents: one is called Master Air Waybill or MAWB and the other is known by House Air Waybill or HAWB. Two mandatory fields in this document are: “Shipper Name & Address” and “Consignee Name & Address”. HAWB carries the actual names of the parties in the corresponding fields. While MAWB carries the respective Forwarder’s name in each of the two fields.

There’s another document called the Air Cargo Manifest or ACM, which contains summary of both MAWB and HAWB. It is issued by the Shipper’s Forwarder. It can be found in the OTH-M folder, with the same label.

Data:

You are provided with 4 files:

1. oag_codes.json : ids of all Shipper’s Forwarders
2. identify_codes.json : is a list of masters with
 - a. ID - primary key
 - b. OagCode - code of the Shipper’s Forwarder
 - c. houses: is a list of hawb each having a Shipper and Consignee Code
3. document_labels.json: is a list of all jobs with unique ID and attachments.
 - a. Each attachment will have labels of certain pages of the pdf file, whose Path is also given
4. AICST.zip - contains documents whose Paths (described in 3.a)

Challenge:

The task is to predict the Shipper and Consignee Code masked in the file identify_codes.json

Remarks:

* OTH-H and PS-M are beyond the scope of this Challenge.

** Multiple tuples in House represent the scenario of Consolidation (CONSOL) where 1 or more Shippers could be shipping goods to 1 or more Consignees.

*** This is just a fraction of the data. You may not find all pdf files mentioned in document_labels.json