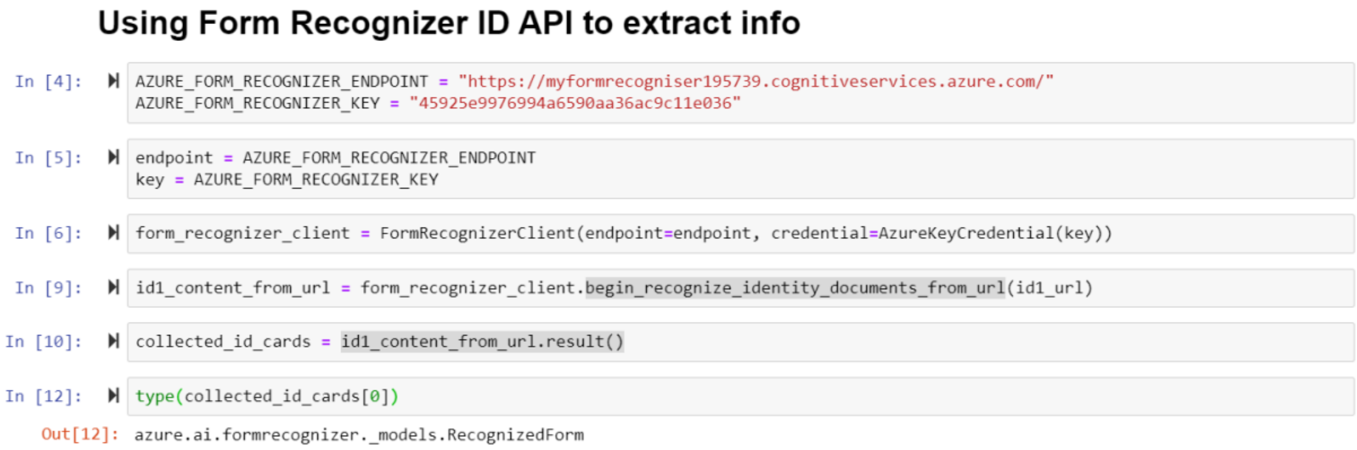
Step 2 - Text Data Extraction

Text Data from Digital ID

Form recognizer prebuild ID's API is used to extract information from the digital ID



Using Form Recognizer ID API to extract info

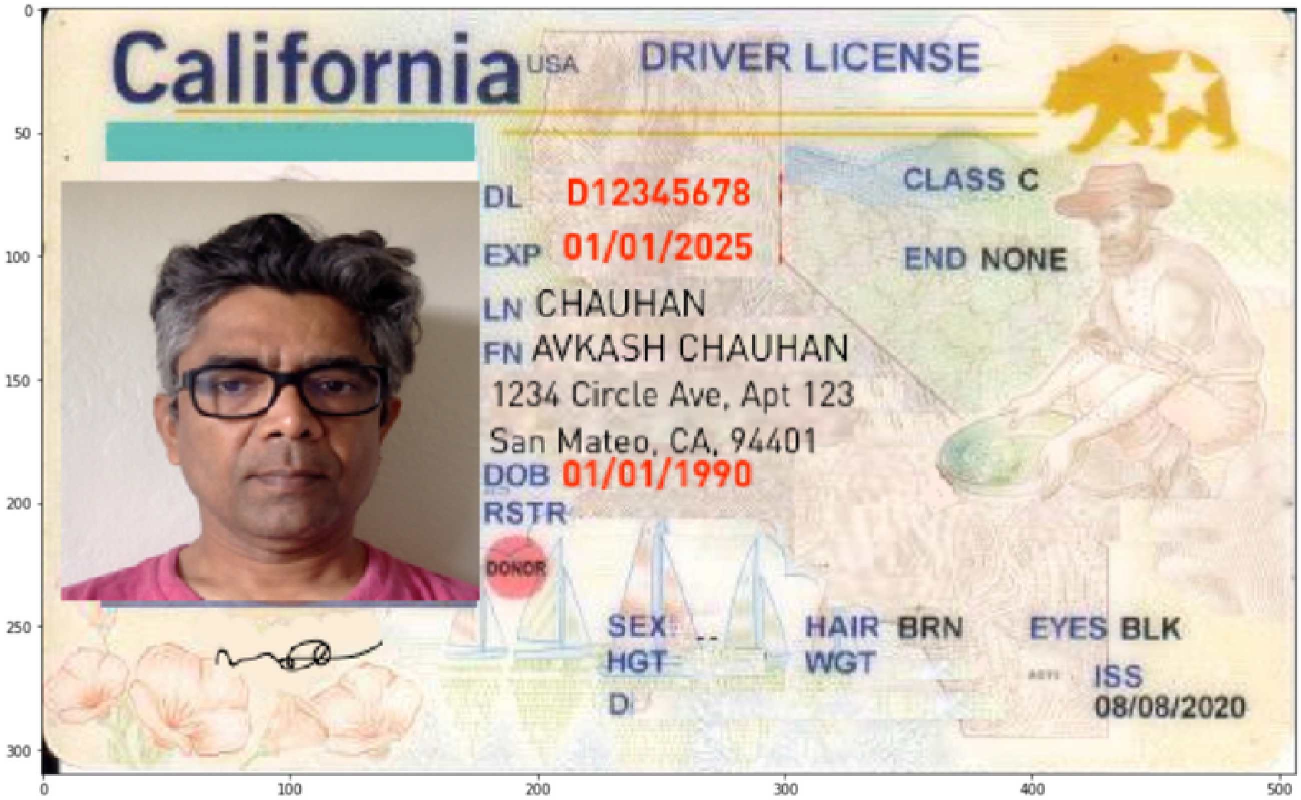
```
In [4]:  AZURE_FORM_RECOGNIZER_ENDPOINT = "https://myformrecogniser195739.cognitiveservices.azure.com/"
         AZURE_FORM_RECOGNIZER_KEY = "45925e9976994a6590aa36ac9c11e036"

In [5]:  endpoint = AZURE_FORM_RECOGNIZER_ENDPOINT
         key = AZURE_FORM_RECOGNIZER_KEY

In [6]:  form_recognizer_client = FormRecognizerClient(endpoint=endpoint, credential=AzureKeyCredential(key))

In [9]:  id1_content_from_url = form_recognizer_client.begin_recognize_identity_documents_from_url(id1_url)

In [10]: collected_id_cards = id1_content_from_url.result()

In [12]: type(collected_id_cards[0])

Out[12]: azure.ai.formrecognizer._models.RecognizedForm
```

Comparing the extracted result with the actual ID

```
for index_id, id_card in enumerate(collected_id_cards):
    print("Displaying identity card details ....... # {}".format(index_id+1))
    get_id_card_details(id_card)
    print("---------------- EOL ------------------------")
```

```
Displaying identity card details ....... # 1
First Name: AVKASH CHAUHAN has confidence: 0.76
Last Name: CHAUHAN has confidence: 0.883
Document Number: D1234578 has confidence: 0.995
Date of Birth: 1990-01-01 has confidence: 0.995
Date of Expiration: 2025-01-01 has confidence: 0.992
Sex: X has confidence: 0.161
Address: 1234 Circle Ave, Apt 123 San Mateo, CA, 94401 has confidence: 0.585
Country/Region: USA has confidence: 0.99
Region: California has confidence: 0.984
---------------- EOL ------------------------
```

```
show_image_in_cell(id1_url)
```



California USA DRIVER LICENSE

DL D12345678 CLASS C

EXP 01/01/2025 END NONE

LN CHAUHAN

FN AVKASH CHAUHAN

1234 Circle Ave, Apt 123

San Mateo, CA, 94401

DOB 01/01/1990

RSTR

DONOR

SEX .. HAIR BRN EYES BLK

HGT WGT ISS

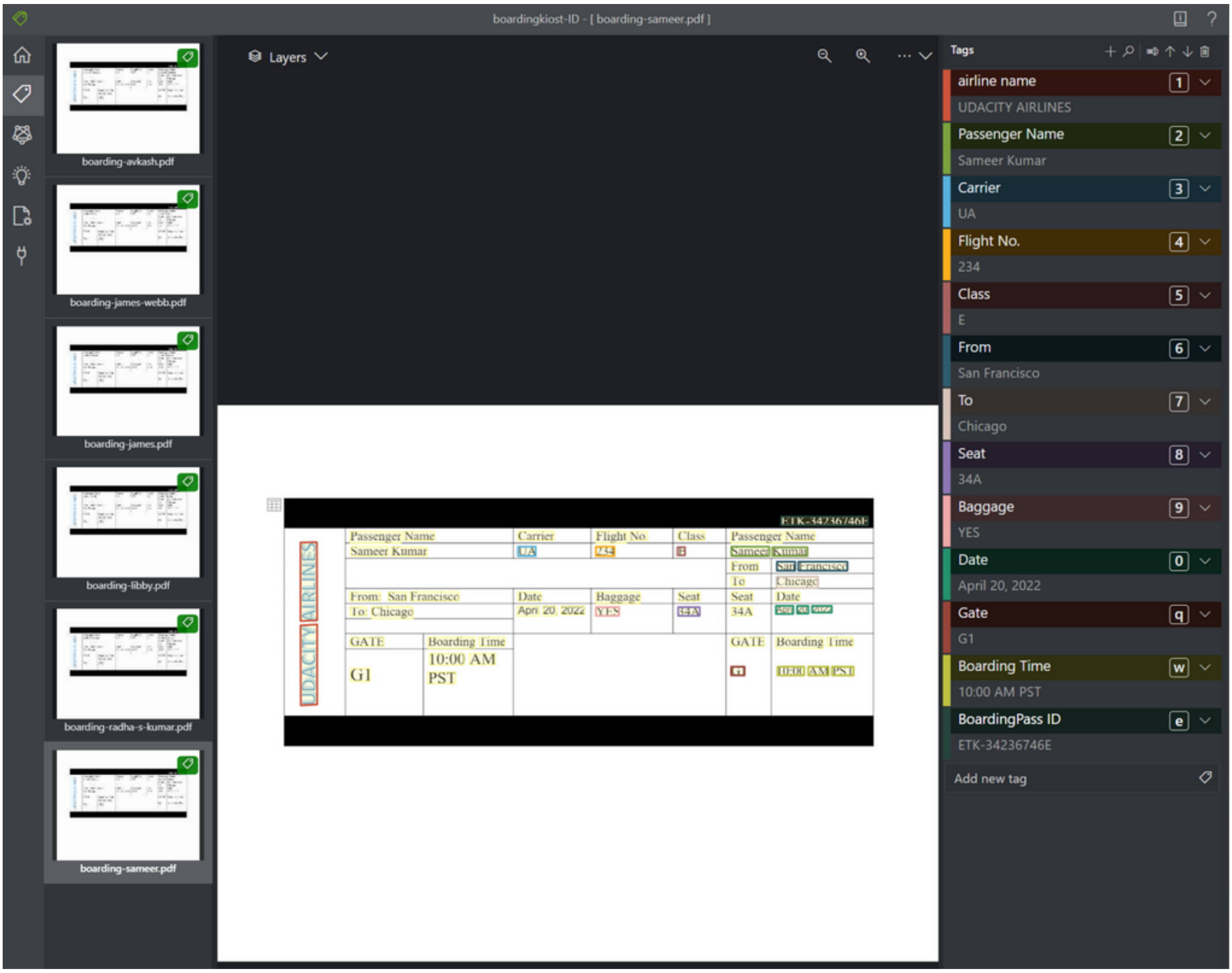
DI 08/08/2020

0
50
100
150
200
250
300
0
100
200
300
400
500

Text Data from Boarding Pass

Build a Custom Boaring Pass Recognizer Model

Step1 - Label the boarding pass



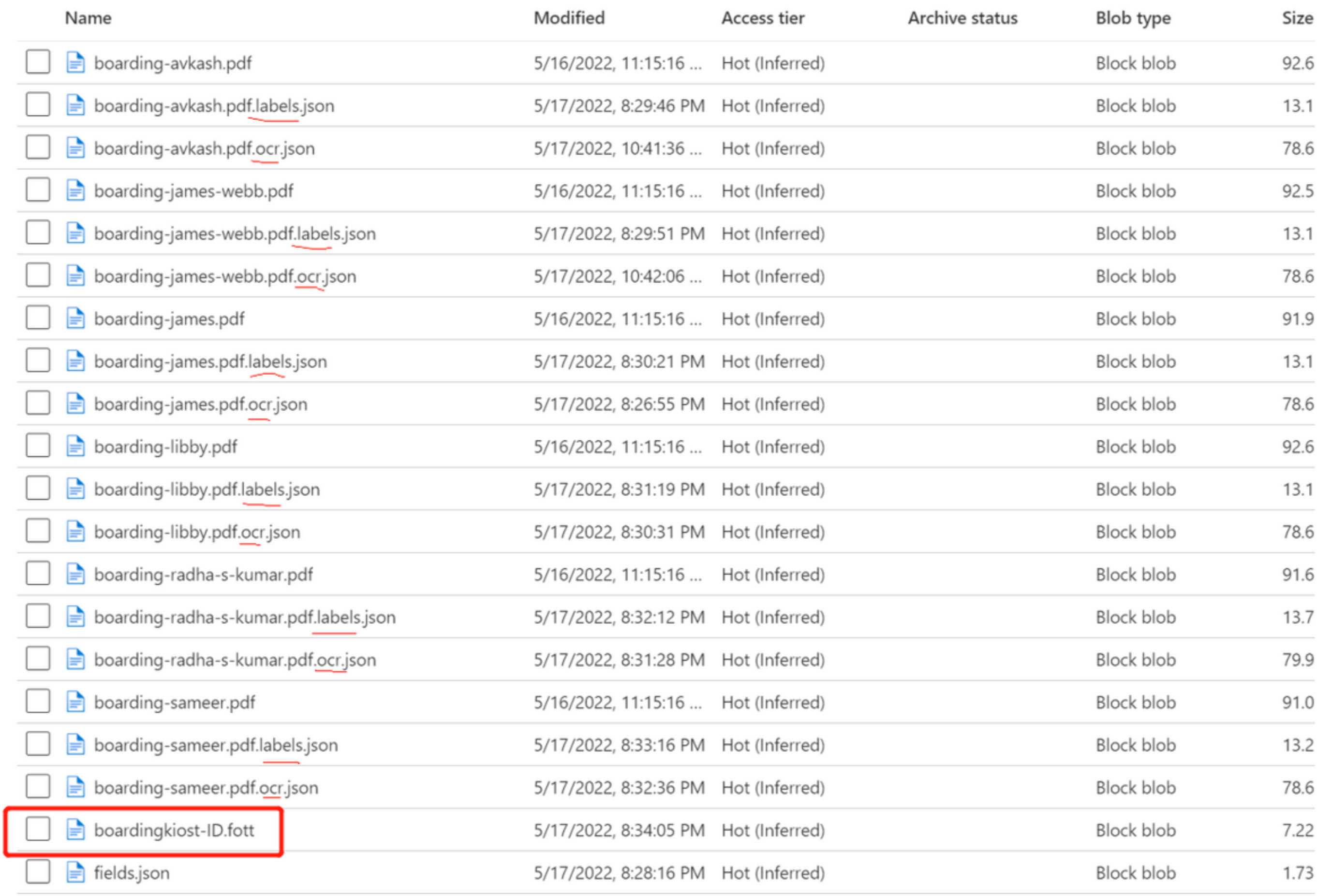
ETK-34236746E

UDACITY AIRLINES	Passenger Name	Carrier	Flight No.	Class	Passenger Name
	Sameer Kumar	UA	234	E	Sameer Kumar
					From San Francisco
					To Chicago
	From: San Francisco	Date	Baggage	Seat	Seat 34A Date Apr 20 2022
	To: Chicago	Apri 20, 2022	YES	34A	
	GATE G1	Boarding Time 10:00 AM PST			GATE G1 Boarding Time 10:00 AM PST

Tag	Key	Value
airline name	1	UDACITY AIRLINES
Passenger Name	2	Sameer Kumar
Carrier	3	UA
Flight No.	4	234
Class	5	E
From	6	San Francisco
To	7	Chicago
Seat	8	34A
Baggage	9	YES
Date	0	April 20, 2022
Gate	q	G1
Boarding Time	w	10:00 AM PST
BoardingPass ID	e	ETK-34236746E

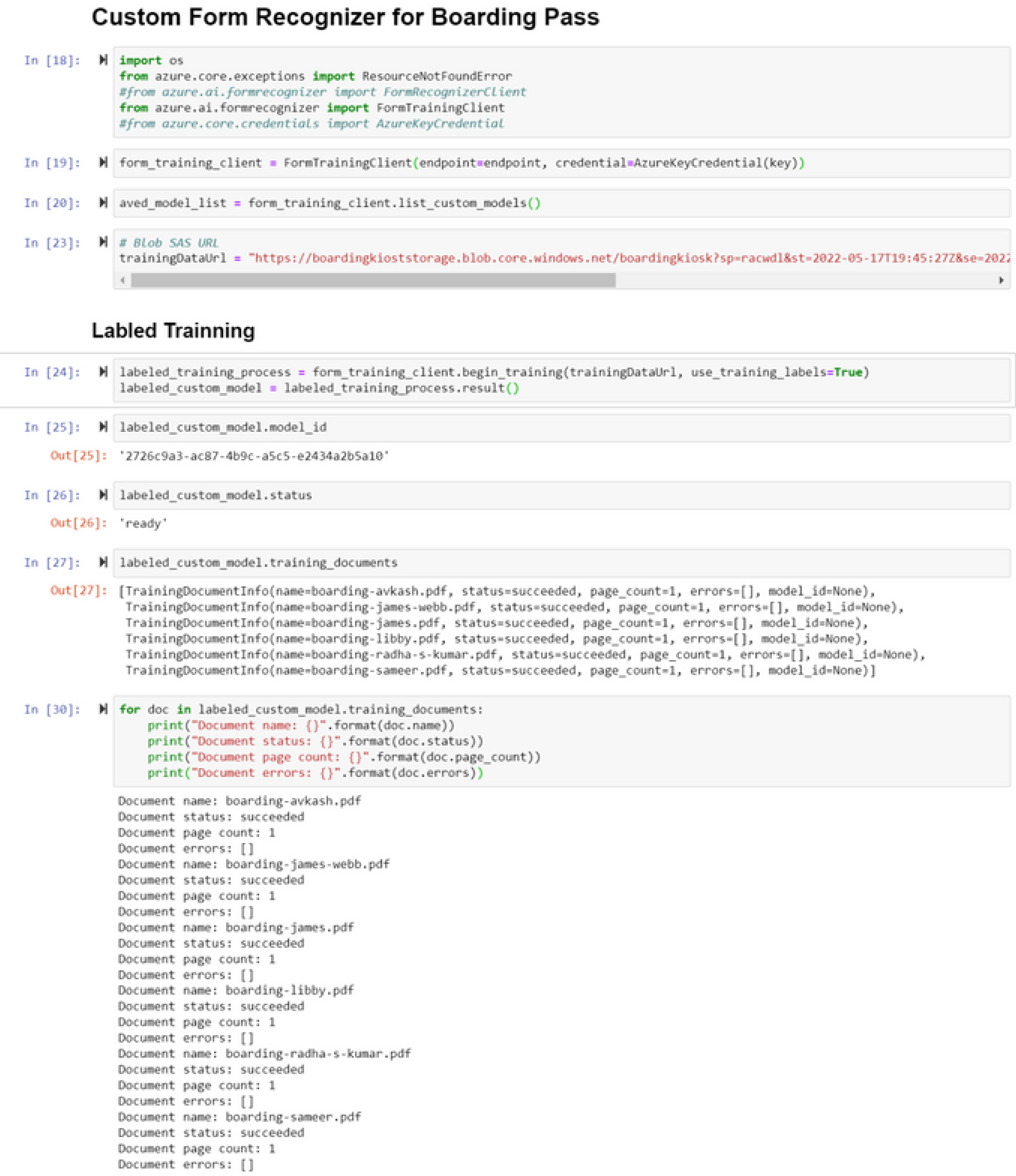
After I label the fields at the Form Recognizer portal, `labels.json` files along with `ocr.json` files are added to my blob container.

I have labelled 5 files and left 1 file unlabelled.



	Name	Modified	Access tier	Archive status	Blob type	Size
□	boarding-avkash.pdf	5/16/2022, 11:15:16 ...	Hot (Inferred)		Block blob	92.6
□	boarding-avkash.pdf.labels.json	5/17/2022, 8:29:46 PM	Hot (Inferred)		Block blob	13.1
□	boarding-avkash.pdf.ocr.json	5/17/2022, 10:41:36 ...	Hot (Inferred)		Block blob	78.6
□	boarding-james-webb.pdf	5/16/2022, 11:15:16 ...	Hot (Inferred)		Block blob	92.5
□	boarding-james-webb.pdf.labels.json	5/17/2022, 8:29:51 PM	Hot (Inferred)		Block blob	13.1
□	boarding-james-webb.pdf.ocr.json	5/17/2022, 10:42:06 ...	Hot (Inferred)		Block blob	78.6
□	boarding-james.pdf	5/16/2022, 11:15:16 ...	Hot (Inferred)		Block blob	91.9
□	boarding-james.pdf.labels.json	5/17/2022, 8:30:21 PM	Hot (Inferred)		Block blob	13.1
□	boarding-james.pdf.ocr.json	5/17/2022, 8:26:55 PM	Hot (Inferred)		Block blob	78.6
□	boarding-libby.pdf	5/16/2022, 11:15:16 ...	Hot (Inferred)		Block blob	92.6
□	boarding-libby.pdf.labels.json	5/17/2022, 8:31:19 PM	Hot (Inferred)		Block blob	13.1
□	boarding-libby.pdf.ocr.json	5/17/2022, 8:30:31 PM	Hot (Inferred)		Block blob	78.6
□	boarding-radha-s-kumar.pdf	5/16/2022, 11:15:16 ...	Hot (Inferred)		Block blob	91.6
□	boarding-radha-s-kumar.pdf.labels.json	5/17/2022, 8:32:12 PM	Hot (Inferred)		Block blob	13.7
□	boarding-radha-s-kumar.pdf.ocr.json	5/17/2022, 8:31:28 PM	Hot (Inferred)		Block blob	79.9
□	boarding-sameer.pdf	5/16/2022, 11:15:16 ...	Hot (Inferred)		Block blob	91.0
□	boarding-sameer.pdf.labels.json	5/17/2022, 8:33:16 PM	Hot (Inferred)		Block blob	13.2
□	boarding-sameer.pdf.ocr.json	5/17/2022, 8:32:36 PM	Hot (Inferred)		Block blob	78.6
□	boardingkiost-ID.fott	5/17/2022, 8:34:05 PM	Hot (Inferred)		Block blob	7.22
□	fields.json	5/17/2022, 8:28:16 PM	Hot (Inferred)		Block blob	1.73

Step 2 - Labeled Training



Custom Form Recognizer for Boarding Pass

```
In [18]:  import os
          from azure.core.exceptions import ResourceNotFoundError
          #from azure.ai.formrecognizer import FormRecognizerClient
          from azure.ai.formrecognizer import FormTrainingClient
          #from azure.core.credentials import AzureKeyCredential
```

```
In [19]:  form_training_client = FormTrainingClient(endpoint=endpoint, credential=AzureKeyCredential(key))
```

```
In [20]:  aved_model_list = form_training_client.list_custom_models()
```

```
In [23]:  # Blob SAS URL
          trainingDataUrl = "https://boardingkioststorage.blob.core.windows.net/boardingkiosk?sp=racwdl&st=2022-05-17T19:45:27Z&se=2022
```

Labled Trainning

```
In [24]:  labeled_training_process = form_training_client.begin_training(trainingDataUrl, use_training_labels=True)
          labeled_custom_model = labeled_training_process.result()
```

```
In [25]:  labeled_custom_model.model_id

Out[25]:  '2726c9a3-ac87-4b9c-a5c5-e2434a2b5a10'
```

```
In [26]:  labeled_custom_model.status

Out[26]:  'ready'
```

```
In [27]:  labeled_custom_model.training_documents

Out[27]:  [TrainingDocumentInfo(name=boarding-avkash.pdf, status=succeeded, page_count=1, errors=[], model_id=None),
           TrainingDocumentInfo(name=boarding-james-webb.pdf, status=succeeded, page_count=1, errors=[], model_id=None),
           TrainingDocumentInfo(name=boarding-james.pdf, status=succeeded, page_count=1, errors=[], model_id=None),
           TrainingDocumentInfo(name=boarding-libby.pdf, status=succeeded, page_count=1, errors=[], model_id=None),
           TrainingDocumentInfo(name=boarding-radha-s-kumar.pdf, status=succeeded, page_count=1, errors=[], model_id=None),
           TrainingDocumentInfo(name=boarding-sameer.pdf, status=succeeded, page_count=1, errors=[], model_id=None)]
```

```
In [30]:  for doc in labeled_custom_model.training_documents:
              print("Document name: {}".format(doc.name))
              print("Document status: {}".format(doc.status))
              print("Document page count: {}".format(doc.page_count))
              print("Document errors: {}".format(doc.errors))
```

```
Document name: boarding-avkash.pdf
Document status: succeeded
Document page count: 1
Document errors: []
Document name: boarding-james-webb.pdf
Document status: succeeded
Document page count: 1
Document errors: []
Document name: boarding-james.pdf
Document status: succeeded
Document page count: 1
Document errors: []
Document name: boarding-libby.pdf
Document status: succeeded
Document page count: 1
Document errors: []
Document name: boarding-radha-s-kumar.pdf
Document status: succeeded
Document page count: 1
Document errors: []
Document name: boarding-sameer.pdf
Document status: succeeded
Document page count: 1
Document errors: []
```

Step 3 - Testing

Testing

```
In [37]:  nt.com/yuanfresa/Azure-AI-Engineer-Nanodegree-Project-Portfolio/main/Data%20Preparation/boarding_pass/boarding_pass_test.png"
```

```
In [38]:  labeled_custom_test_action = form_recognizer_client.begin_recognize_custom_forms_from_url(model_id=labeled_custom_model.model
```

```
In [40]:  labeled_custom_test_action.status()

Out[40]:  'succeeded'
```

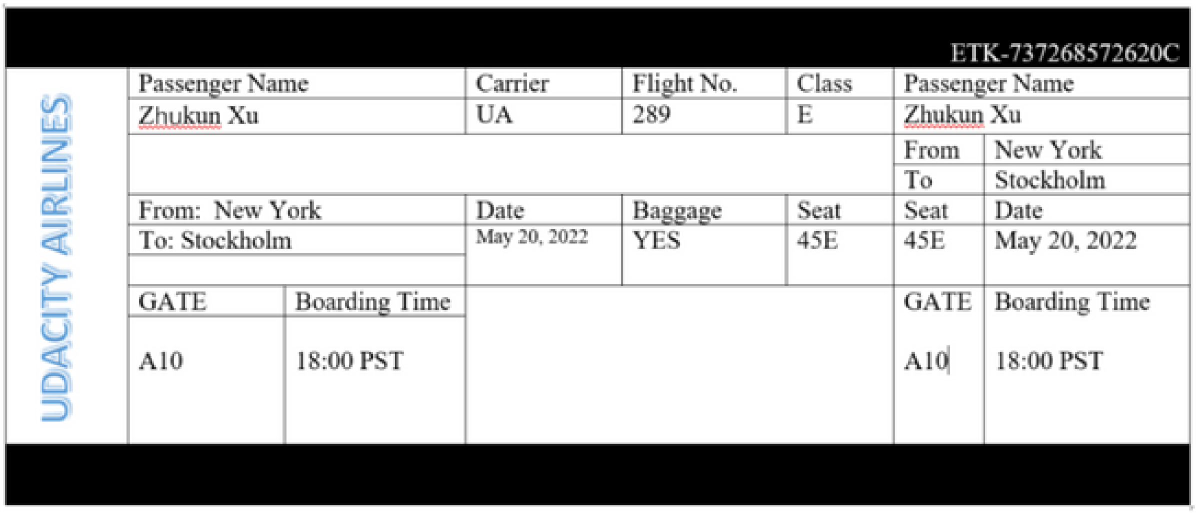
```
In [41]:  labeled_custom_test_action_result = labeled_custom_test_action.result()
```

```
In [42]:  for recognized_content in labeled_custom_test_action_result:
              print("Form type: {}".format(recognized_content.form_type))
              for name, field in recognized_content.fields.items():
                  print("Field '{}' has label '{}' with value '{}' and a confidence score of {}".format(
                      name,
                      field.label_data.text if field.label_data else name,
                      field.value,
                      field.confidence
                  ))
```

```
Form type: custom:2726c9a3-ac87-4b9c-a5c5-e2434a2b5a10
Field 'Passenger Name' has label 'Passenger Name' with value 'Zhukun Xu' and a confidence score of 0.99
Field 'airline name' has label 'airline name' with value 'UDACITY AIRLINES' and a confidence score of 0.99
Field 'Boarding Time' has label 'Boarding Time' with value 'None' and a confidence score of 0.906
Field 'Baggage' has label 'Baggage' with value 'YES' and a confidence score of 0.99
Field 'From' has label 'From' with value 'New York' and a confidence score of 0.99
Field 'Carrier' has label 'Carrier' with value 'UA' and a confidence score of 0.99
Field 'To' has label 'To' with value 'Stockholm' and a confidence score of 0.99
Field 'Date' has label 'Date' with value 'May 20, 2022' and a confidence score of 0.989
Field 'Class' has label 'Class' with value 'E' and a confidence score of 0.99
Field 'BoardingPass ID' has label 'BoardingPass ID' with value 'ETK-737268572620C' and a confidence score of 0.99
Field 'Gate' has label 'Gate' with value 'A10' and a confidence score of 0.99
Field 'Seat' has label 'Seat' with value '45E' and a confidence score of 0.021
Field 'Flight No.' has label 'Flight No.' with value '289' and a confidence score of 0.99
```

```
In [43]:  Image.open(requests.get(new_test_url, stream=True).raw)
```

Out[43]:



ETK-737268572620C

UDACITY AIRLINES	Passenger Name		Carrier	Flight No.	Class	Passenger Name	
	Zhukun Xu		UA	289	E	Zhukun Xu	
						From	New York
						To	Stockholm
	From: New York		Date	Baggage	Seat	Seat	Date
	To: Stockholm		May 20, 2022	YES	45E	45E	May 20, 2022
	GATE	Boarding Time				GATE	Boarding Time
	A10	18:00 PST				A10	18:00 PST