**Business Context**

Onfido helps hundreds of businesses across the globe verify their customers. In most cases, this is done using a government-issued legal document and matching this with a biometric element (i.e. a selfie/video). Processing a diverse set of identity documents poses a major challenge for building a reliable AI engine that catches fraud accurately. The Onfido technology team continuously strives towards building a better risk engine while not compromising on automation rates.

Just to provide technical context, from the point a document is submitted to the Onfido risk engine, AI needs to first identify the document, extract appropriate fields and document features for fraud detection and come up with a result. However, in some cases the automation may not be able to create a conclusive result and a human agent may need to finalise the verification process.

At Onfido, our goal is to provide the highest pass rate (a document report with ‘Clear’ result) in the fastest possible turn-around time without letting fraudulent customers through our controls.

**Data**

You are provided with 3 datasets:

Document Report

This dataset contains information about document reports processed by Onfido.

Human-Agent Data Set

When a document verification fails to complete automatically, it gets split into parts or tasks. We have three main tasks:

* 1. Classification task: the agent needs to classify a document (for instance they will have to state it’s a UK Driving Licence)
  2. Extraction task: the agent needs to extract personal information from the document (first name, last name, date of birth, etc)
  3. Fraud Assessment task: the agent needs to assess whether the document submitted is fraudulent.

This dataset contains details of each manual task processed by Onfido.

Clients

This dataset contains information about Onfido’s clients.

**Questions**

Document Reports overview

* Explain the business and document report performance trends

Focus on the tasks performance

Now, we want to focus on the impact on task performance (in particular the turn-around time).

* Explain the main drivers of the agencies’ performance
* Suggest a few improvements to reduce the turnaround time
* Estimate the impact of these improvements on Onfido’s business (e.g. in terms of revenues, cost, number of tasks created)

**Expected output**

Assume you will have 15-20 mins to present this when you come onsite.

We expect the exercise to take a maximum of 4-5 hours to complete.

This task aims to assess your capacity to dive into datasets, get insights from them applying relevant analytics techniques and be able to tell a story in a clear and logical manner.

Feel free to make assumptions if you think it is needed.

We hope you will enjoy it. Good Luck!