

Technical test

You have the OCR output for a couple of receipts, and we need to extract some relevant information such as vendor name, vendor address, line items and total. It is intended for you to develop a generalized way to extract this information for both receipts, bearing in mind that any receipt with the same format should be supported.

The OCR output is structured as:

```
<x1>,<y1>,<x2>,<y2>,<x3>,<y3>,<x4>,<y4>,<text>
```

Expected output for receipt 1:

```
{
  "company": "MR D.I.Y. (JOHOR) SDN BHD",
  "date": "12-01-19",
  "address": "LOT 1851-A & 1851-B, JALAN KPB 6, KAWASAN PERINDUSTRIAN BALAKONG, 43300 SERI KEMBANGAN, SELANGOR (MR DIY TESCO TERBAU)",
  "line_items": [
    {
      "sku": "8970669",
      "quantity": "1",
      "price": "19.00",
      "total": "19.00"
    },
    {
      "sku": "9066468",
      "quantity": "1",
      "price": "8.02",
      "total": "8.02"
    },
    {
      "sku": "9557031100236",
      "quantity": "1",
      "price": "3.02",
      "total": "3.02"
    },
    {
      "sku": "6935818350846",
      "quantity": "1",
      "price": "3.88",
      "total": "3.88"
    }
  ],
  "total": "33.90"
}
```

Send the link of the repository of your solution to jmgarzonv@veryfi.com.