

Upgrade

Open in app

<b>Superman</b> <b>Net \$3,956</b> <b>12/20/2019</b> <b>12/01/19 - 12/14/19</b>	<b>Regular</b> <b>OT</b>	<b>\$45</b> <b>\$67.50</b>	<b>80</b> <b>10</b>	<b>\$3,600</b> <b>\$675</b>
<b>Batman</b> <b>Net \$1,856</b> <b>12/20/2019</b> <b>12/01/19 - 12/14/19</b>	<b>Regular</b> <b>OT</b>	<b>\$50</b> <b>\$75.00</b>	<b>55</b> <b>0</b>	<b>\$2,750</b> <b>\$0</b>

(Created by Author)

There are few steps we need to take to transform the data into panel format.

- **Step 1: Import PDF data as a DataFrame**

Like data in a structured format, we also use `tb.read_pdf` to import the unstructured data.

This time, we need to specify extra options to properly import the data.

```
file = 'payroll_sample.pdf'
df= tb.read_pdf(file, pages = '1', area = (0, 0, 1000, 1000), columns
= [200, 265, 300, 320], pandas_options={'header': None}, stream=True)
[0]
```

*Area and Columns:* I've talked about `area` above. Here we will also need to use `columns` to identify the locations all relevant columns (one column on the left section and four

