

Data Extraction - Remote Assignment

Expense date (also called spend date) is the exact date on which a financial transaction or operation has happened and must be recognized in the accounting system.

For example, in the receipt below, the date underlined is the expense date.



The format of expense dates vary across merchants and the country of issue. In a few cases, the expense date may not be present in the receipt.

Sample receipts dataset:

<https://drive.google.com/drive/folders/1tsNzkpVvfbEQDVOFTuverce1CqBUWSYC?usp=sharing>

Build a RESTful service that extracts expense date from a receipt. Please feel free to use any language and framework of your choice. Deploy the service on any cloud platform like Heroku/AWS/GCP. The service should contain one API which has the following contract:

Request:

POST /extract_date

Payload:

```
{"base_64_image_content": <base_64_image_content>}
```

Response:

If date is present:

```
{"date": "YYYY-MM-DD"}
```

If date is not present:

```
{"date": null}
```

Calculate the accuracy of the service against the sample receipts dataset given. This can be calculated as follows:

$$\text{accuracy} = (\text{number of receipts for which the service extracted correct date} / \text{total receipts}) * 100\%$$