Task For Data Science

Problem Statement:

Our Objective is to Clare Non available/blank Data from Our Dataset

Data:

* The file named “Data.xlsx” contains four different data sets of different medicine names.
* Each sheet has data from different website

1. Sheet 1 : Medlife Data
2. Sheet 2 : Netmed Data
3. Sheet 3 : PharmEasy Data
4. Sheet 4 : Onemg Data

* Each sheet has multiple attributes

Tips:

Use Business Rules for getting blank Data

Ex: med\_name:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| medName | manufacturer | Ingredients | pack\_form | strength | medlife\_price | medicineType | Pack Size |
| A TO Z NS NEW | ALKEM LABORATORIES LTD | ASCORBIC |  | Not Available | 89.25 | TABLET | 15 |

In above example you can observe that pack\_form is empty we need to fill that

***As you can see we have MedicineType which is Tablet so we can say that pack form for this med\_name is Strip***

There are many rules like this

* For strength check name of medicine if its has XX MG in its name then the strength is XX MG

Project Deliverables –

**Project Deliverables –**

1. Send us a copy of the code/jupyter notebook used to map the medicines.

2. Send us the combined data made by you in xlsx or csv format.

**Software to be used** – Excel/ Python/ R.

**Deadline** - Kindly complete the task in 24 hours (1 day).

**Dataset**: <https://drive.google.com/file/d/1A6JZ1jpUEkqOe377Xpr5LYoddq-7s3KP/view?usp=sharing>

**Note**: This is an open-euyuwnded Task where your approach to tackle the problems faced by you will be evaluated along with the success of the task too.