

Project Development Phase

Project Development - Delivery of Sprint-1

Date	10 November 2022
Team ID	PNT2022TMID27276
Project Name	Project – Analytics for Hospitals' Health- Care Data

Task 1: Download the Dataset using Kaggle API (By uploading reference dataset in kaggle)

Step1: Login kaggle

Step2: Create new API token

Step3: Download the json file

Step4: Open Google Colab, new notebook

Step5: Procedure in Google Colab (Code)

```
!pip install -q Kaggle #Install

!mkdir ~/.kaggle #Creating a kaggle directory

!cp kaggle.json ~/.kaggle/ #Copying json file to folder

!chmod 600 ~/.kaggle/kaggle.json #Changing the permissions to json

!kaggle datasets download -d priyakumaran/healthcare #Copy pasting
the path from kaggle Downloading healthcare.zip to /content
 0% 0.00/6.54M [00:00, ?B/s]
100% 6.54M/6.54M [00:00<00:00, 122MB/s]

!unzip /content/healthcare.zip #To unzip
Archive: /content/healthcare.zip
  inflating: sample_sub.csv
  inflating: test_data.csv
  inflating: train_data.csv
  inflating: train_data_dictionary.csv
```

The screenshot displays a Google Colab notebook environment. The browser tabs at the top include 'Jira | Issue & Project Tracking', 'IBM', 'Healthcare | Kaggle', and 'Healthcare.ipynb - Colaboratory'. The address bar shows the URL: <https://colab.research.google.com/drive/1ecFKZWU1wsYWx5k0MXNgAYU0xapLBk4#scrollTo=06XQVcUbETR6>.

The notebook interface shows a file explorer on the left with the following files: `sample_data`, `healthcare.zip`, `kaggle.json`, `sample_sub.csv`, `test_data.csv`, `train_data.csv`, and `train_data_dictionary.csv`.

The code editor contains the following cells:

```
[1] !pip install -q kaggle
```

```
[2] !mkdir ~/.kaggle
```

```
[3] !cp kaggle.json ~/.kaggle/
```

```
[4] !chmod 600 ~/.kaggle/kaggle.json
```

```
[5] !kaggle datasets download -d priyakumaran/healthcare
```

The output for cell [5] shows the download progress:

```
Downloading healthcare.zip to /content
0% 0.00/6.54M [00:00<?, ?B/s]
100% 6.54M/6.54M [00:00<00:00, 122MB/s]
```

A terminal window is open, showing the unzipping of `healthcare.zip`:

```
!unzip /content/healthcare.zip
```

```
Archive: /content/healthcare.zip
  inflating: sample_sub.csv
  inflating: test_data.csv
  inflating: train_data.csv
  inflating: train_data_dictionary.csv
```

The status bar at the bottom indicates 'Disk 85.13 GB available' and 'completed at 10:41 PM'.

Step6: Download the csv files