

To document the process of loading and preprocessing a dataset using Python and pandas in a PDF, you can follow these steps:

1. ****Write a Colab Notebook****:

Start by creating a Colab Notebook (a .ipynb file) in your preferred Colab Notebook environment. You can use Colab Notebook, Colab Lab, or any other suitable tool. In this notebook, write down the code for loading and preprocessing the dataset, and include explanations and comments to make it clear and informative.

2. ****Add Markdown Cells****:

Use Markdown cells to add explanations, titles, and section headings. You can create a Markdown cell by selecting "Markdown" from the cell type dropdown in Colab Notebook.

For example, you can have sections like:

- Introduction
- Loading the Dataset
- Exploring the Dataset
- Preprocessing the Data
- Saving the Preprocessed Data

3. ****Include Code Cells****:

In the code cells, add the Python code for loading and preprocessing the dataset, as well as any additional code necessary for data analysis. Make sure to add comments to explain each step of the code. Use Markdown for explanations within or after the code cells.

```
```python
Loading the Dataset
import pandas as pd
```

```
Replace 'URL' with the actual dataset URL

url = 'https://tn.data.gov.in/resource/marginal-workers-classified-age-industrial-category-and-sex-scheduled-caste-2011-tamil'

data = pd.read_csv(url)

...
```

#### 4. **\*\*Run the Notebook\*\***:

Execute each code cell to ensure that the code runs without errors and that the output, if any, is correctly displayed in the notebook.

#### 5. **\*\*Export to PDF\*\***:

Once you have completed your Colab Notebook and are satisfied with the content and formatting, you can export it to a PDF file. To do this, follow these steps:

- In Colab Notebook, select "File" > "Export Notebook As" > "PDF".
- Choose the page layout and style options that suit your preferences.
- Click the "Export" button.

Colab Notebook will generate a PDF file containing the content of your notebook.

#### 6. **\*\*Review and Share\*\***:

Review the PDF to ensure it accurately documents the process of loading and preprocessing the dataset. If everything looks good, you can share the PDF document with others.

#### 7. **\*\*Upload to GitHub\*\***:

As per your assignment instructions, upload both the Colab Notebook (.ipynb) and the generated PDF to your private GitHub repository.

Make sure to grant access to the evaluators by adding their email addresses as collaborators on your GitHub repository, so they can access and evaluate your work.

✓  
0s

```
[1] import pandas as pd
```

✓  
0s

```
[2] # Replace 'URL' with the actual dataset URL
url = '/content/drive/MyDrive/DDW_B06SC_3300_State_TAMIL_NADU-2011.csv'
data = pd.read_csv(url)

Alternatively, you can download the dataset and load it from a local file:
data = pd.read_csv('your_local_dataset.csv')
```