

Download and install the vscode: <https://code.visualstudio.com/download>

Download and install anaconda: <https://www.anaconda.com/download/success>

Python download and installation link: <https://www.python.org/downloads/>

**Note:** Download python version at least 3.10 or more than that

**Alternative solution for the notebook:** Open the colab using this link:  
<https://colab.research.google.com/>

If python and vs code and anaconda is not ready inside your system you can work with colab

=====

Steps which you need to perform:

1. Create a folder inside your local system(any drive)
2. Open this folder inside your vscode
3. Open the vscode terminal(you can see this at your header of vscode)
4. Select the cmd and make sure base env(it is from the anaconda) is there
5. **Special Note:** Go and search python in your vs code extension and download the microsoft python extension

**6. Special Note: if base env is not coming the follow below step**

Click on view > open command pallet > select python interpreter > select base interpreter(this interpreter from anaconda)

7. Steps for creating and activating the virtual env:

```
Conda create -p venv python=3.10 -y
```

```
Conda activate <path of your virtual env>
```

8. We have to create and install the requirement file :  

```
pip install -r requirements.txt
```

This is the command for loading the data:

```
data=pd.read_csv("https://raw.githubusercontent.com/Ankit152/IMDB-sentiment-analysis/master/IMDB-Dataset.csv")
```

### Text with html=

```
"""<!DOCTYPE html><html lang="en"><head><meta
charset="UTF-8"><meta http-equiv="X-UA-Compatible" content="IE=edge"><meta
name="viewport" content="width=device-width,
initial-scale=1.0"><title>Welcome to My
Website</title><style>body{font-family:'Arial',sans-serif;background-color
:#f0f0f0;color:#333;margin:20px}h1{color:#007bff}p{line-height:1.5}</style
></head><body><header><h1>Welcome to My Awesome
Website!</h1></header><main><p>This is a sample HTML document created for
demonstration purposes.</p><p>Feel free to explore and enjoy the content
on this website.</p></main><footer><p>&copy; 2024 My Website. All rights
reserved.</p></footer></body></html>"""
```

Link for the regular expression: <https://docs.python.org/3/library/re.html>

For interview perspective this is not imp

```
def remove_html_tag(text):
    pattern=re.compile('<.*?>')
    return pattern.sub("",text)
```

```
data["review"]=data["review"].apply(remove_html_tag)
```

```
def remove_url(text):
    pattern=re.compile(r'https?://\S+|www\.\S+')
    return pattern.sub("",text)
```

sat: next session will discuss remaining techniques of preprocessing

sun: in sun's session will talk about the encoding and embedding

=====