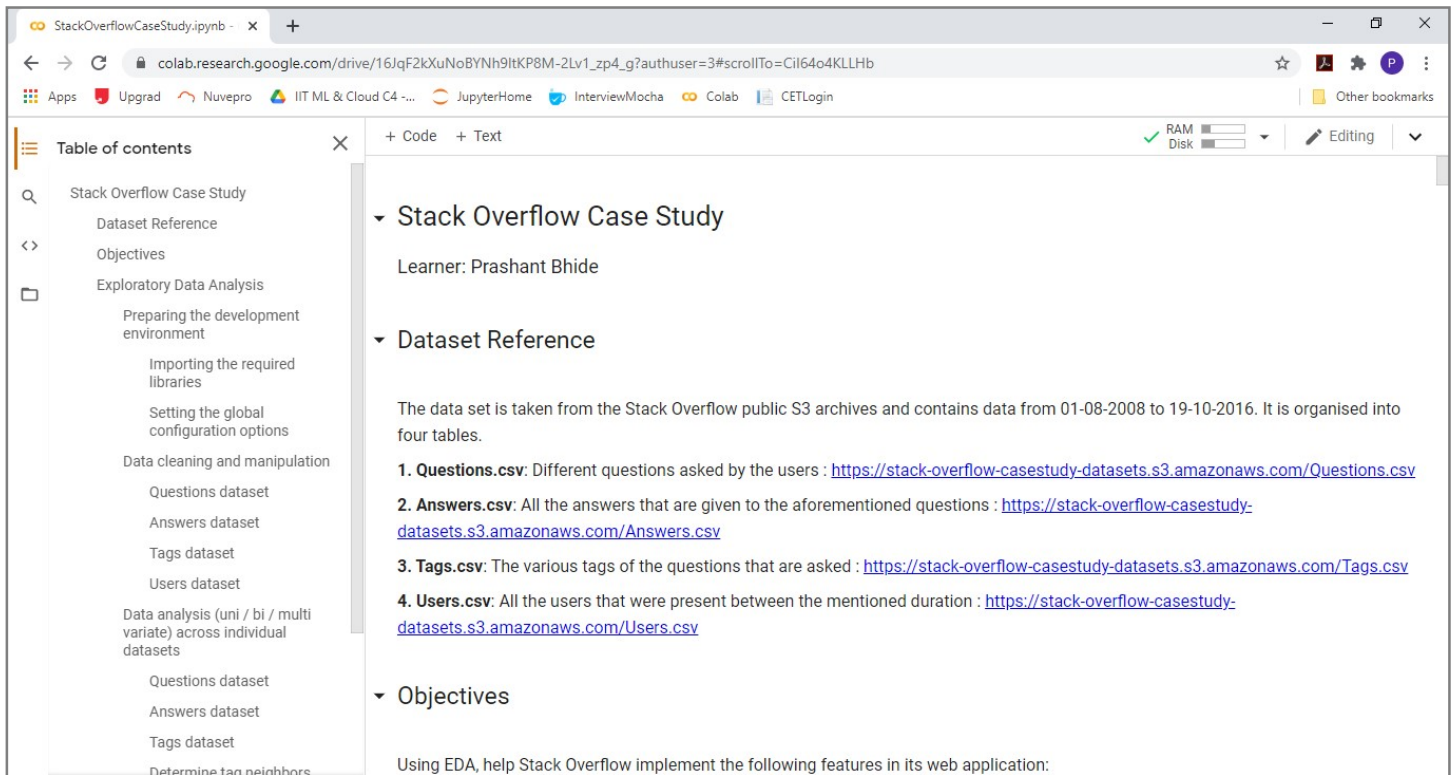


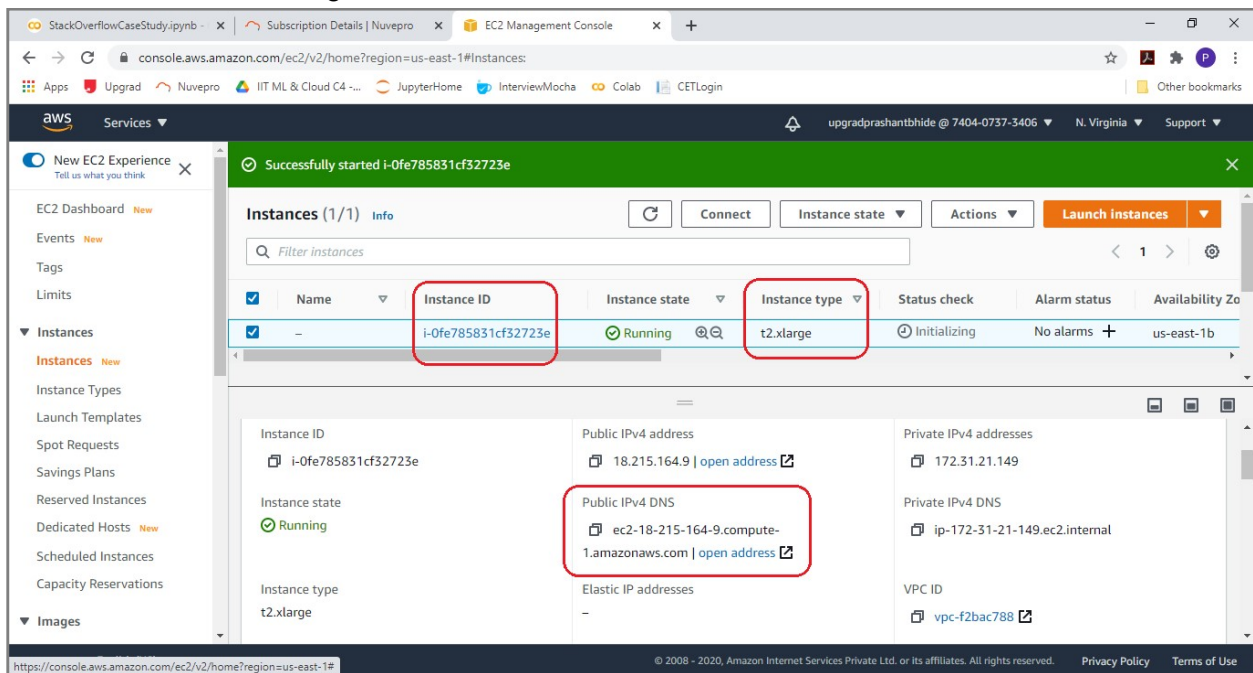
1. The public link of ipython file on the Google Colab where the EDA is performed.

- https://colab.research.google.com/drive/16JqF2kXuNoBYNh9ItKP8M-2Lv1_zp4_g?usp=sharing
- Screenshot of Google Colab running ipython file



Launched EC2 instance: **t2.xlarge** (higher configuration with 16 GB RAM chosen to allow processing the Stack Overflow Case Study multi-GB datasets)

2. Screenshot of the running EC2 instance



3. Screenshot of Anaconda Installation on EC2

```
ec2-user@ip-172-31-21-149:~
no change      /home/ec2-user/anaconda3/condabin/conda
no change      /home/ec2-user/anaconda3/bin/conda
no change      /home/ec2-user/anaconda3/bin/conda-env
no change      /home/ec2-user/anaconda3/bin/activate
no change      /home/ec2-user/anaconda3/bin/deactivate
no change      /home/ec2-user/anaconda3/etc/profile.d/conda.sh
no change      /home/ec2-user/anaconda3/etc/fish/conf.d/conda.fish
no change      /home/ec2-user/anaconda3/shell/condabin/Conda.ps1
no change      /home/ec2-user/anaconda3/shell/condabin/conda-hook.ps1
no change      /home/ec2-user/anaconda3/lib/python3.8/site-packages/xontrib/conda.xsh
no change      /home/ec2-user/anaconda3/etc/profile.d/conda.csh
modified       /home/ec2-user/.bashrc

==> For changes to take effect, close and re-open your current shell. <==

If you'd prefer that conda's base environment not be activated on startup,
    set the auto_activate_base parameter to false:

conda config --set auto_activate_base false

Thank you for installing Anaconda3!

=====

Working with Python and Jupyter notebooks is a breeze with PyCharm Pro,
designed to be used with Anaconda. Download now and have the best data
tools at your fingertips.

PyCharm Pro for Anaconda is available at: https://www.anaconda.com/pycharm

[ec2-user@ip-172-31-21-149 ~]$
```

4. Screenshot of the terminal running the jupyter notebook on EC2

```
[screen 0: ec2-user@ip-172-31-21-149:~/Notebooks]
(base) [ec2-user@ip-172-31-21-149 Notebooks]$ echo $USER
ec2-user
(base) [ec2-user@ip-172-31-21-149 Notebooks]$ sudo chown $USER:$USER /home/ec2-u
ser/certs/mycert.pem
(base) [ec2-user@ip-172-31-21-149 Notebooks]$ jupyter notebook
[I 10:00:20.577 NotebookApp] JupyterLab extension loaded from /home/ec2-user/ana
conda3/lib/python3.8/site-packages/jupyterlab
[I 10:00:20.577 NotebookApp] JupyterLab application directory is /home/ec2-user/
anaconda3/share/jupyter/lab
[I 10:00:20.580 NotebookApp] Serving notebooks from local directory: /home/ec2-u
ser/Notebooks
[I 10:00:20.580 NotebookApp] Jupyter Notebook 6.1.4 is running at:
[I 10:00:20.580 NotebookApp] https://ip-172-31-21-149.ec2.internal:8888/
[I 10:00:20.580 NotebookApp] Use Control-C to stop this server and shut down all
kernels (twice to skip confirmation).
```

5. Screenshot of the Jupyter Notebook running on EC2 instance server

StackOverflowCaseStudy Last Checkpoint: a few seconds ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
In [102]: # Finally write top 10 rows of the FinalDataset to the DB
FinalDataset[:10].to_sql(name='FinalDataset', con=database_connection, method='multi', if_exists='replace')

In [103]: # Read the FinalDataset data from DB back into a new dataframe and display
NewDF = pd.read_sql_table('FinalDataset', con=database_connection)
NewDF

Out[103]:
```

index	Q_Id	Q_Userid	Q_CreationDate	Q_Score	A_Id	A_Userid	A_CreationDate	A_Score	A_CreationYear	AnswerTimeInMinutes	Tag	TagCategory
0	0	260	2008-08-01T12:22:00Z	49	260	91	2008-08-01T12:43:31Z	4	2008	27	c#	Top10NonD
1	1	650	2008-08-03T11:12:52Z	79	655	91	2008-08-03T11:41:38Z	69	2008	29	c#	Top10NonD
2	2	10190	2008-08-13T18:38:02Z	10	10194	91	2008-08-13T18:40:32Z	8	2008	2	sql	DataScieno
3	3	17870	2008-08-20T17:58:51Z	11	17873	91	2008-08-20T17:58:51Z	7	2008	1	mysql	Top10NonD

6. Screenshot of the summary page of the database which you used to store the results.

Amazon RDS

DB identifier: test-instance

CPU: 2.62%

Status: Available

Class: db.t2.micro

Role: Instance

Current activity: 0 Connections

Engine: MySQL Community

Region & AZ: us-east-1b

Connectivity & security

Endpoint & port

Endpoint: test-instance.crqdduy2ogq.us-east-1.rds.amazonaws.com

Port: 3306

Networking

Availability zone: us-east-1b

VPC: vpc-f2bac788

Subnet group: default

Security

VPC security groups: default (sg-e3784abf) (active)

Public accessibility: Yes

Certificate authority: rds-ca-2019

Learner: Prashant Bhide