

RINEX

Machine Learning

Mini Project: 1

12 Operations on a Dataset

Harsh Singh

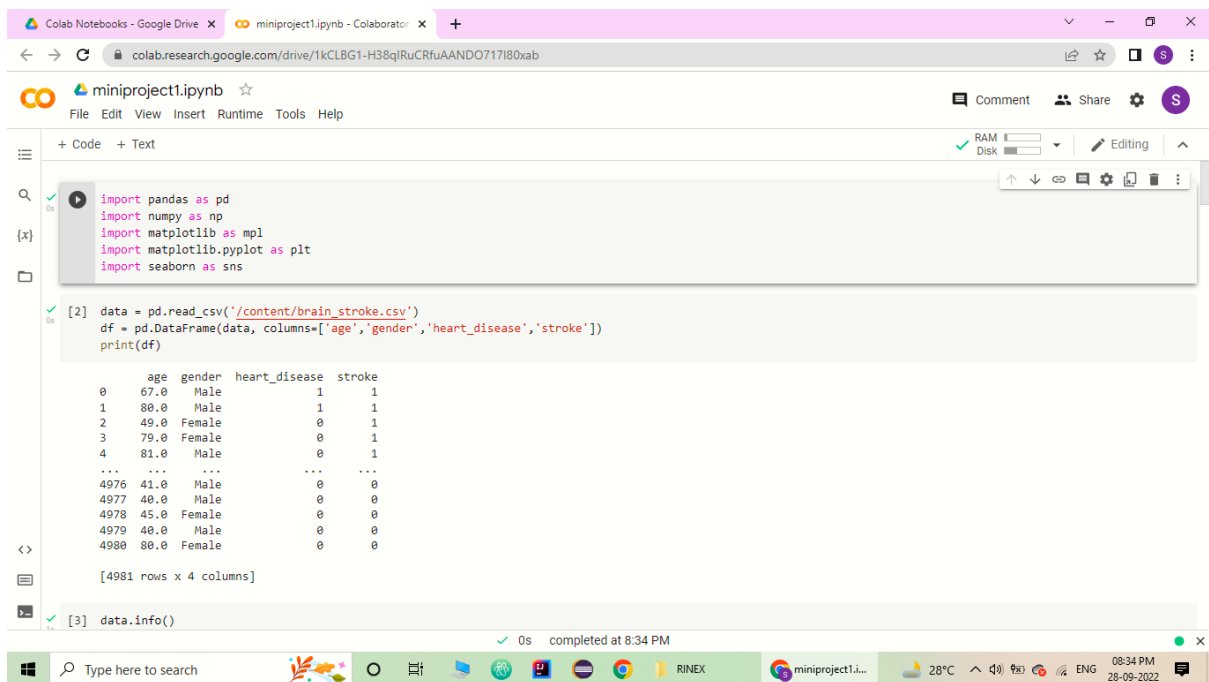
Thakur College of Engineering & Technology-(TCET)

B.E. Electronics

Third Year-[TE]

Dataset: Brain Stroke (Kaggle)

Screenshots of Code & Outputs:



The screenshot shows a Google Colab notebook interface. The browser address bar displays the URL: `colab.research.google.com/drive/1kCLBG1-H38qIRuCRfuAANDO71780xab`. The notebook title is `miniproject1.ipynb`. The code editor contains the following Python code:

```
import pandas as pd
import numpy as np
import matplotlib as mpl
import matplotlib.pyplot as plt
import seaborn as sns

[2] data = pd.read_csv('/content/brain_stroke.csv')
df = pd.DataFrame(data, columns=['age', 'gender', 'heart_disease', 'stroke'])
print(df)
```

The output of the code is a preview of the DataFrame:

	age	gender	heart_disease	stroke
0	67.0	Male	1	1
1	80.0	Male	1	1
2	49.0	Female	0	1
3	79.0	Female	0	1
4	81.0	Male	0	1
...
4976	41.0	Male	0	0
4977	40.0	Male	0	0
4978	45.0	Female	0	0
4979	40.0	Male	0	0
4980	80.0	Female	0	0

Below the preview, it indicates `[4981 rows x 4 columns]`. The next code cell contains `[3] data.info()`. The status bar at the bottom shows the code completed at 8:34 PM.

RINEX

Colab Notebooks - Google Drive X miniproject1.ipynb - Collaborator X +

colab.research.google.com/drive/1kCLBG1-H38qIRuCRfuAANDO717180xab

miniproject1.ipynb ☆

File Edit View Insert Runtime Tools Help

+ Code + Text

RAM 100% Disk 100%

Editing

[3] data.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4981 entries, 0 to 4980
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  --
0   gender                4981 non-null   object
1   age                   4981 non-null   float64
2   hypertension          4981 non-null   int64
3   heart_disease         4981 non-null   int64
4   ever_married          4981 non-null   object
5   work_type             4981 non-null   object
6   Residence_type        4981 non-null   object
7   avg_glucose_level     4981 non-null   float64
8   bmi                   4981 non-null   float64
9   smoking_status        4981 non-null   object
10  stroke                4981 non-null   int64
dtypes: float64(3), int64(3), object(5)
memory usage: 428.2+ KB
```

[4] data.head(5)

	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
0	Male	67.0	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
1	Male	80.0	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1

0s completed at 8:34 PM

Type here to search

RINEX miniproject1.i... 28°C 08:34 PM 28-09-2022

Colab Notebooks - Google Drive X miniproject1.ipynb - Collaborator X +

colab.research.google.com/drive/1kCLBG1-H38qIRuCRfuAANDO717180xab

miniproject1.ipynb ☆

File Edit View Insert Runtime Tools Help

+ Code + Text

RAM 100% Disk 100%

Editing

[4] data.head(5)

	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
0	Male	67.0	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
1	Male	80.0	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
2	Female	49.0	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
3	Female	79.0	1	0	Yes	Self-employed	Rural	174.12	24.0	never smoked	1
4	Male	81.0	0	0	Yes	Private	Urban	186.21	29.0	formerly smoked	1

[5] data.tail(3)

	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
4978	Female	45.0	1	0	Yes	Govt_job	Rural	95.02	31.8	smokes	0
4979	Male	40.0	0	0	Yes	Private	Rural	83.94	30.0	smokes	0
4980	Female	80.0	1	0	Yes	Private	Urban	83.75	29.1	never smoked	0

[6] data.shape

(4981, 11)

0s completed at 8:34 PM

Type here to search

RINEX miniproject1.i... 28°C 08:34 PM 28-09-2022

RINEX

Colab Notebooks - Google Drive

miniproject1.ipynb - Collaborator

+

colab.research.google.com/drive/1kCLBG1-H38qIRuCRfuAANDO717180xab

miniproject1.ipynb

File Edit View Insert Runtime Tools Help

Comment Share

RAM Disk

Editing

+ Code + Text

[6] data.shape

(4981, 11)

[7] data.size

54791

[8] data.isnull()

	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
0	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False
...
4976	False	False	False	False	False	False	False	False	False	False	False
4977	False	False	False	False	False	False	False	False	False	False	False
4978	False	False	False	False	False	False	False	False	False	False	False

0s completed at 8:34 PM

Type here to search

28°C 08:35 PM 28-09-2022

Colab Notebooks - Google Drive

miniproject1.ipynb - Collaborator

+

colab.research.google.com/drive/1kCLBG1-H38qIRuCRfuAANDO717180xab

miniproject1.ipynb

File Edit View Insert Runtime Tools Help Saving...

Comment Share

RAM Disk

Editing

+ Code + Text

[9] data.describe()

	age	hypertension	heart_disease	avg_glucose_level	bmi	stroke
count	4981.000000	4981.000000	4981.000000	4981.000000	4981.000000	4981.000000
mean	43.419859	0.096165	0.055210	105.943562	28.498173	0.049789
std	22.662755	0.294848	0.228412	45.075373	6.790464	0.217531
min	0.080000	0.000000	0.000000	55.120000	14.000000	0.000000
25%	25.000000	0.000000	0.000000	77.230000	23.700000	0.000000
50%	45.000000	0.000000	0.000000	91.850000	28.100000	0.000000
75%	61.000000	0.000000	0.000000	113.860000	32.600000	0.000000
max	82.000000	1.000000	1.000000	271.740000	48.900000	1.000000

[10] data['stroke'].value_counts()

0	4733
1	248

Name: stroke, dtype: int64

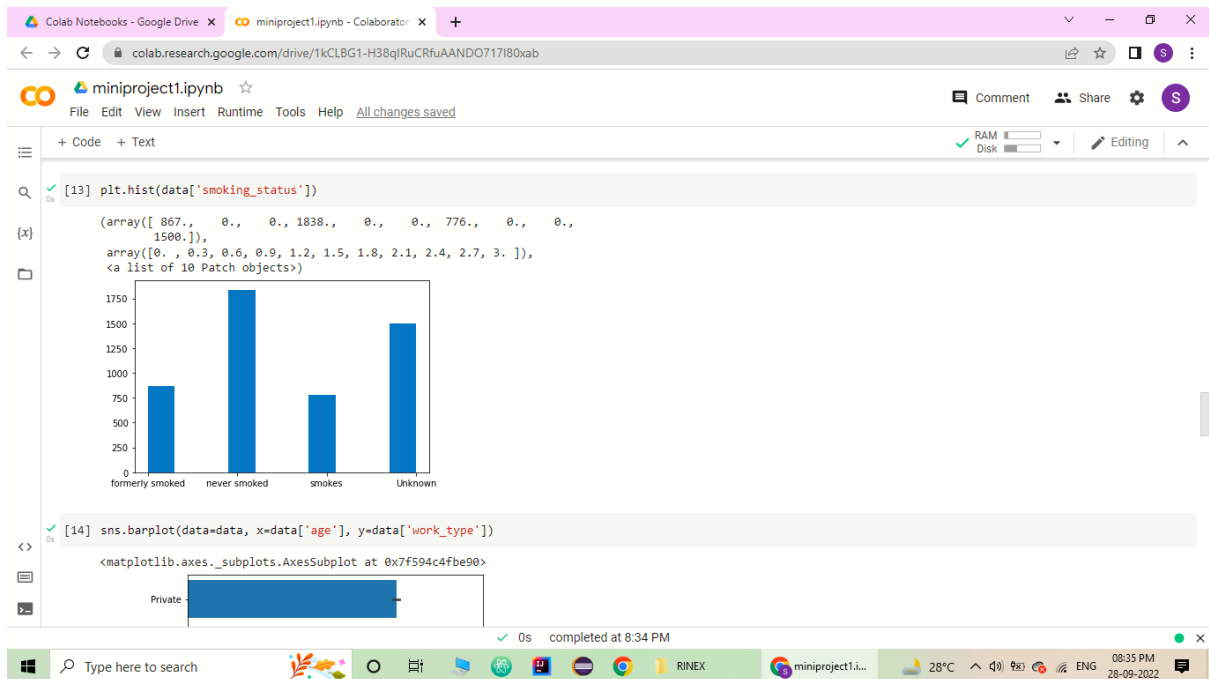
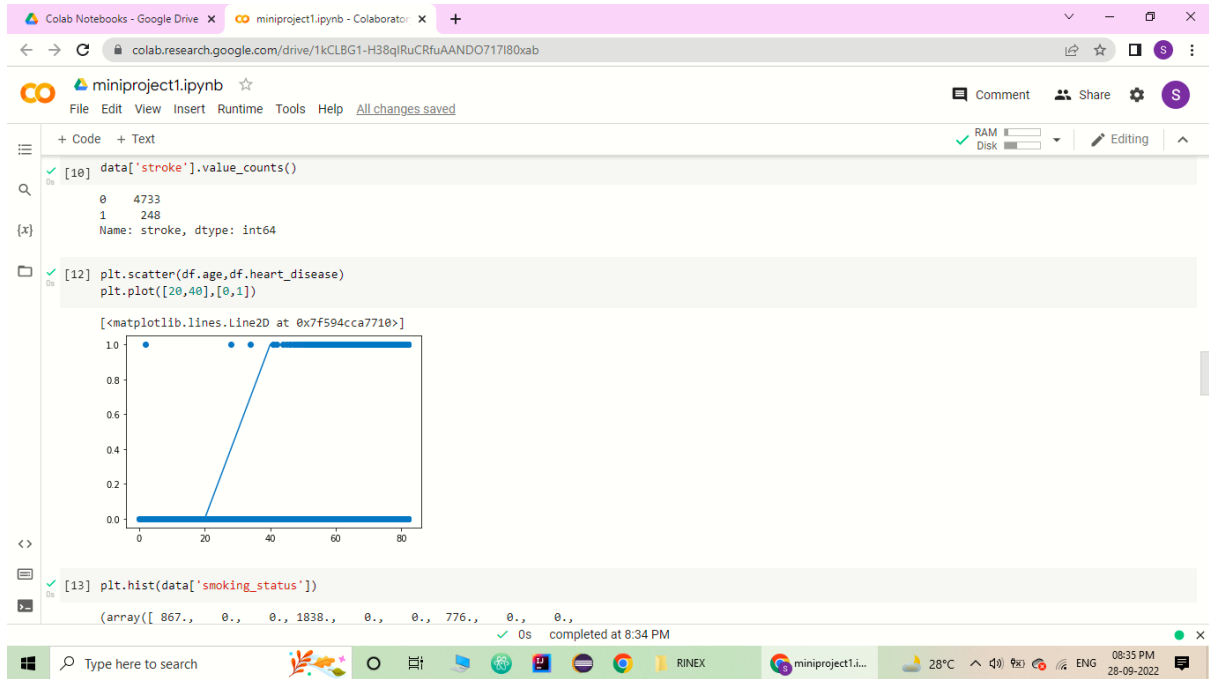
[12] plt.scatter(df.age,df.heart_disease)

0s completed at 8:34 PM

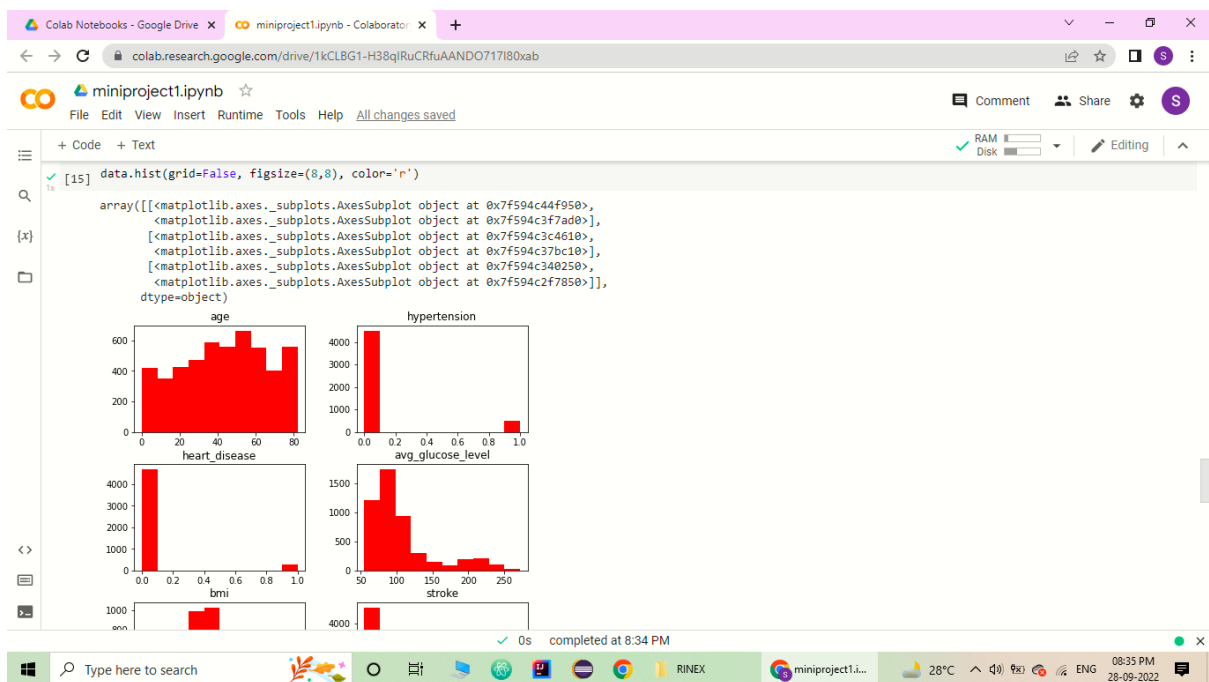
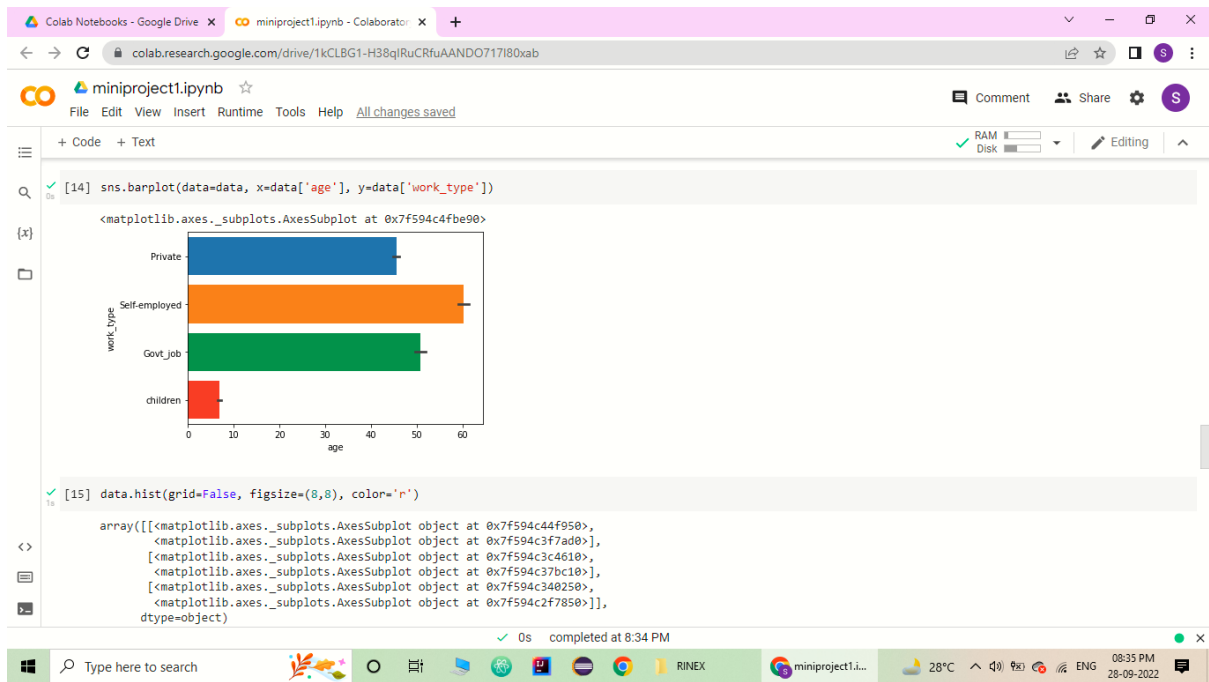
Type here to search

28°C 08:35 PM 28-09-2022

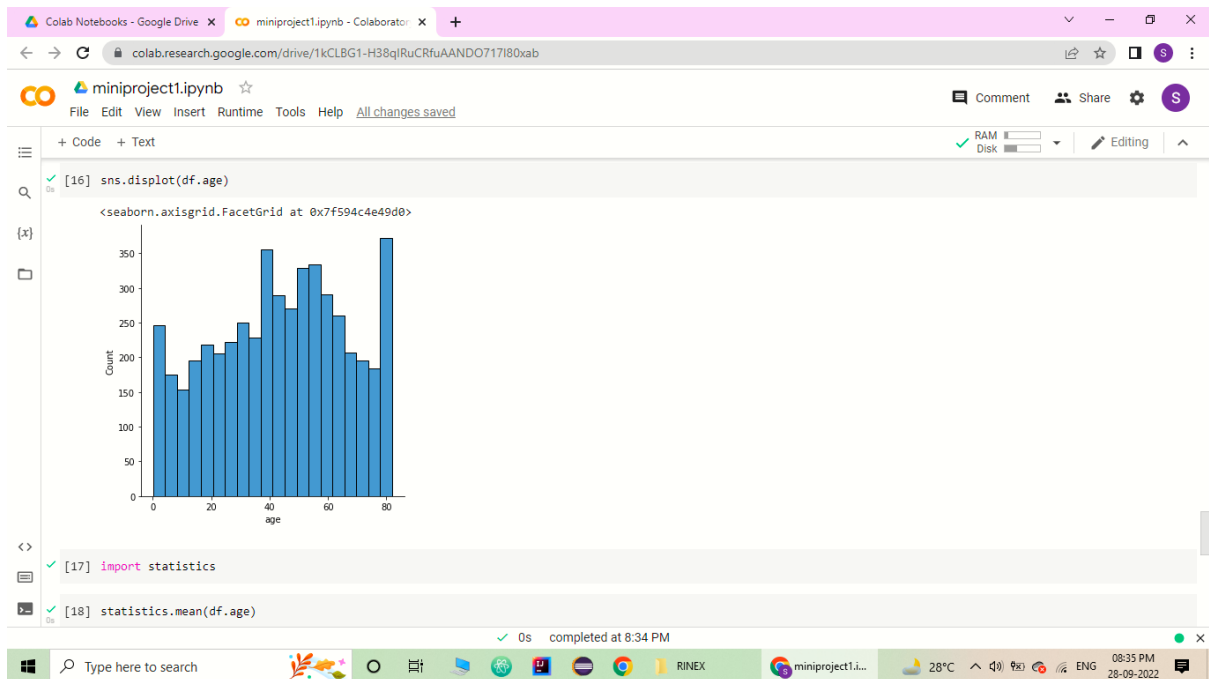
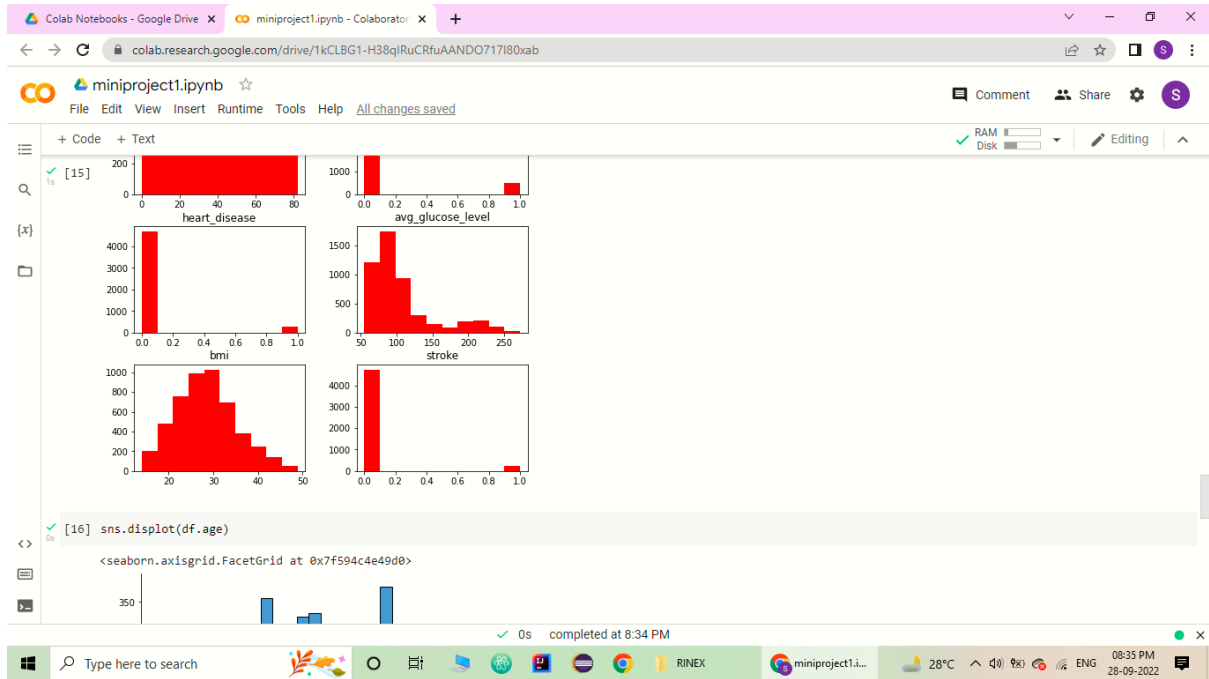
RINEX



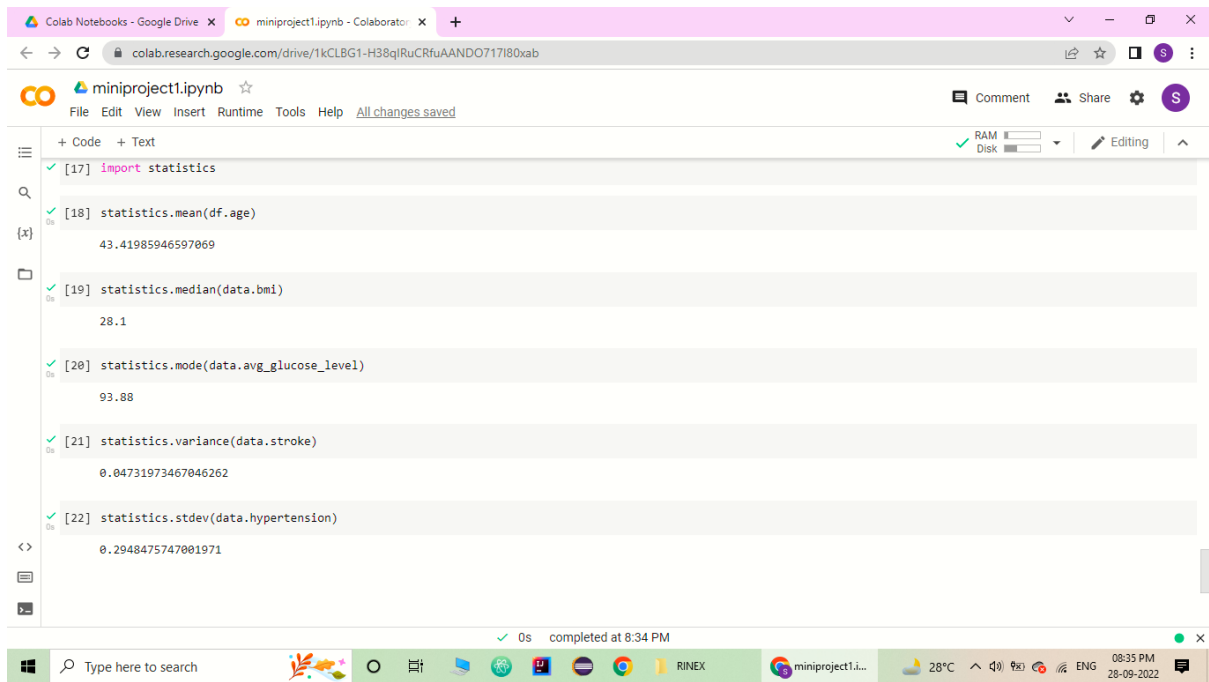
RINEX



RINEX



RINEX



The screenshot shows a Google Colab notebook interface. The browser address bar displays the URL: colab.research.google.com/drive/1kCLBG1-H38qIRuCRfuAANDO717I80xab. The notebook title is 'miniproject1.ipynb'. The menu bar includes File, Edit, View, Insert, Runtime, Tools, and Help. The toolbar shows 'All changes saved' and icons for Comment, Share, and a user profile. The code editor displays the following Python code cells:

```
[17] import statistics
[18] statistics.mean(df.age)
43.41985946597069
[19] statistics.median(data.bmi)
28.1
[20] statistics.mode(data.avg_glucose_level)
93.88
[21] statistics.variance(data.stroke)
0.04731973467046262
[22] statistics.stdev(data.hypertension)
0.2948475747001971
```

The status bar at the bottom indicates '0s completed at 8:34 PM'. The Windows taskbar at the very bottom shows the search bar, task view button, and several application icons, including RINEX and miniproject1.i... The system tray shows a temperature of 28°C, network status, and the date/time: 08:35 PM, 28-09-2022.

Google Drive Link: https://drive.google.com/drive/folders/1wNawEyjY5Q-J_6Q1PRjcH5h9yjgOP_jz