PROJECT REPORT

1.INTRODUCTION

1.1 Overview:

Thyroid Disease Classification Using ML

Project Description:

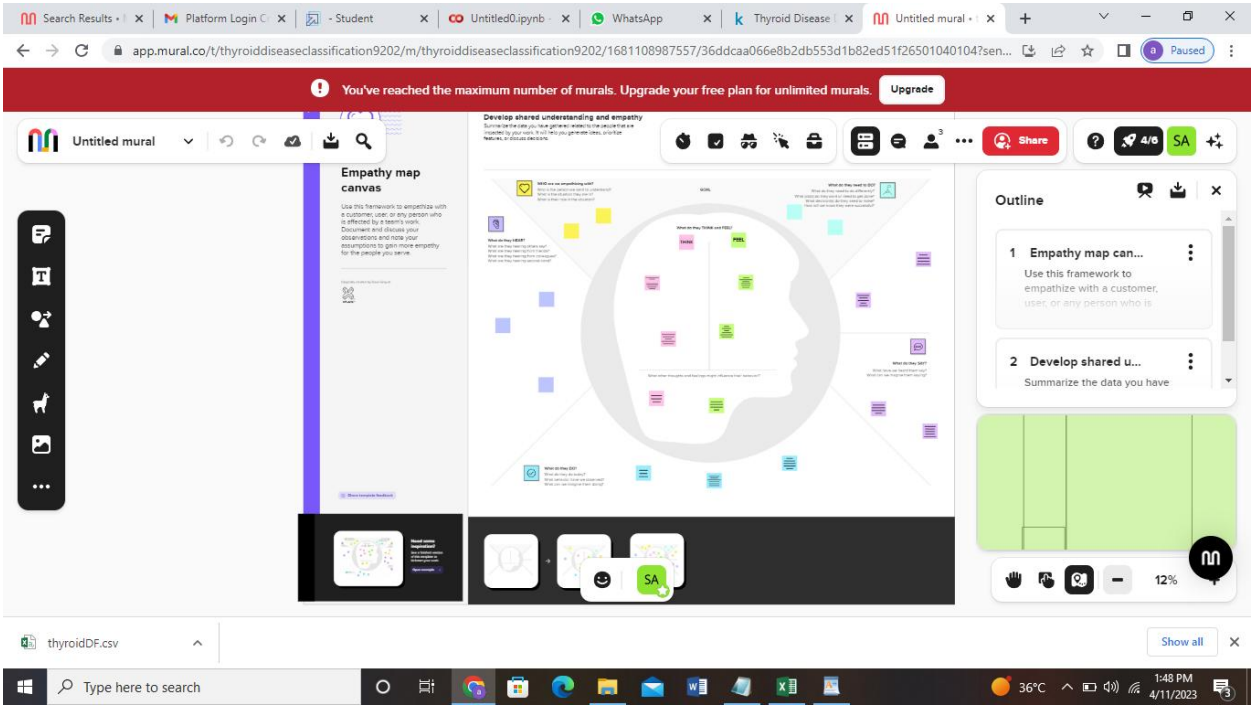
- Thyroid disease is a general term for a medical condition that keeps your thyroid from making the right amount of hormones. Your thyroid typically makes hormones that keep your body functioning normally. When the thyroid makes too much thyroid hormone, your body uses energy too quickly. This is called hyperthyroidism.

1.2purpose

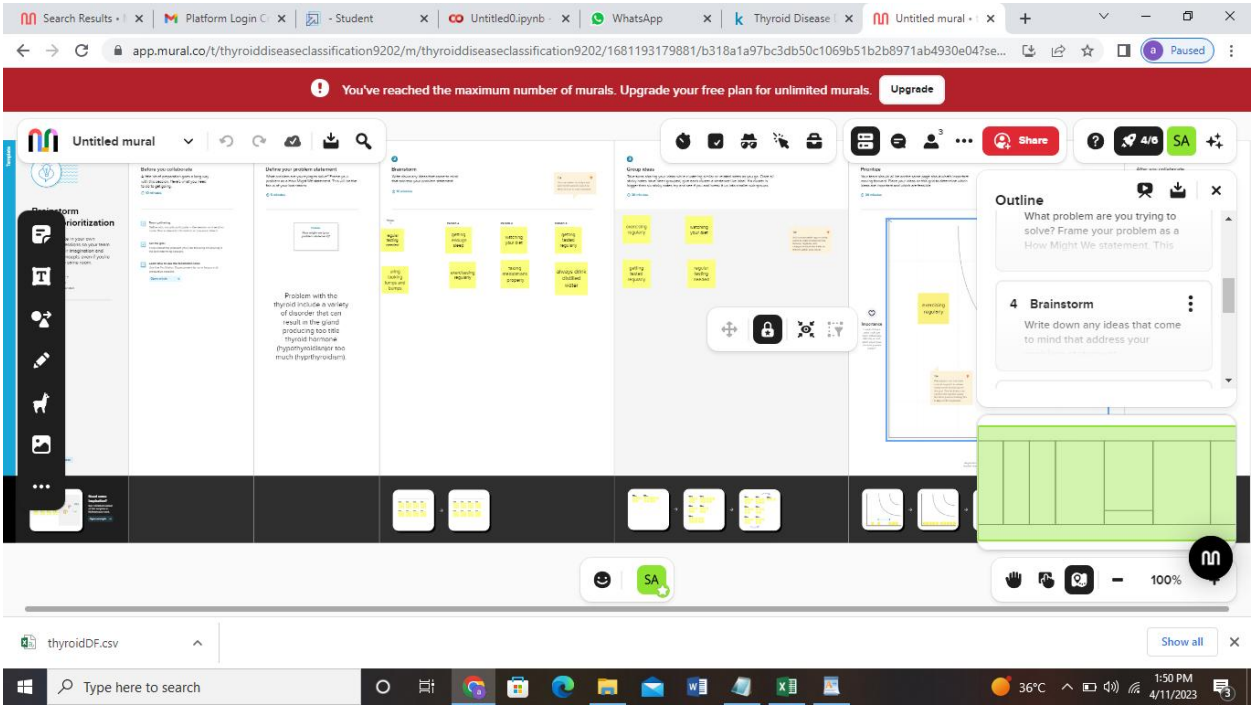
- The thyroid gland is avital hormone gland: It plays a major role in the metabolism , growth and development of the human body.it helps to regulate many body functions by constantly releasing a steady amount of thyroid hormones into the bloodstream.

Problem Definition & Design Thinking

2.1 Empathy Map

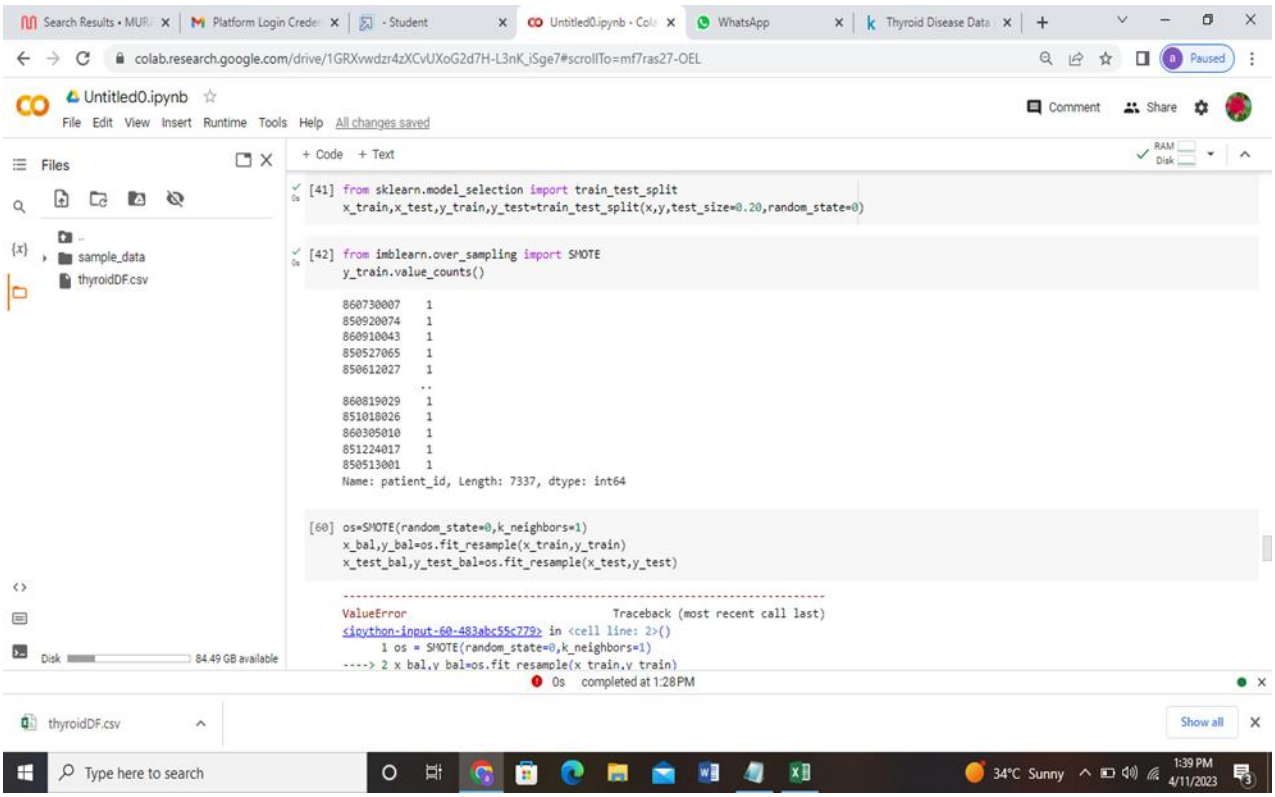


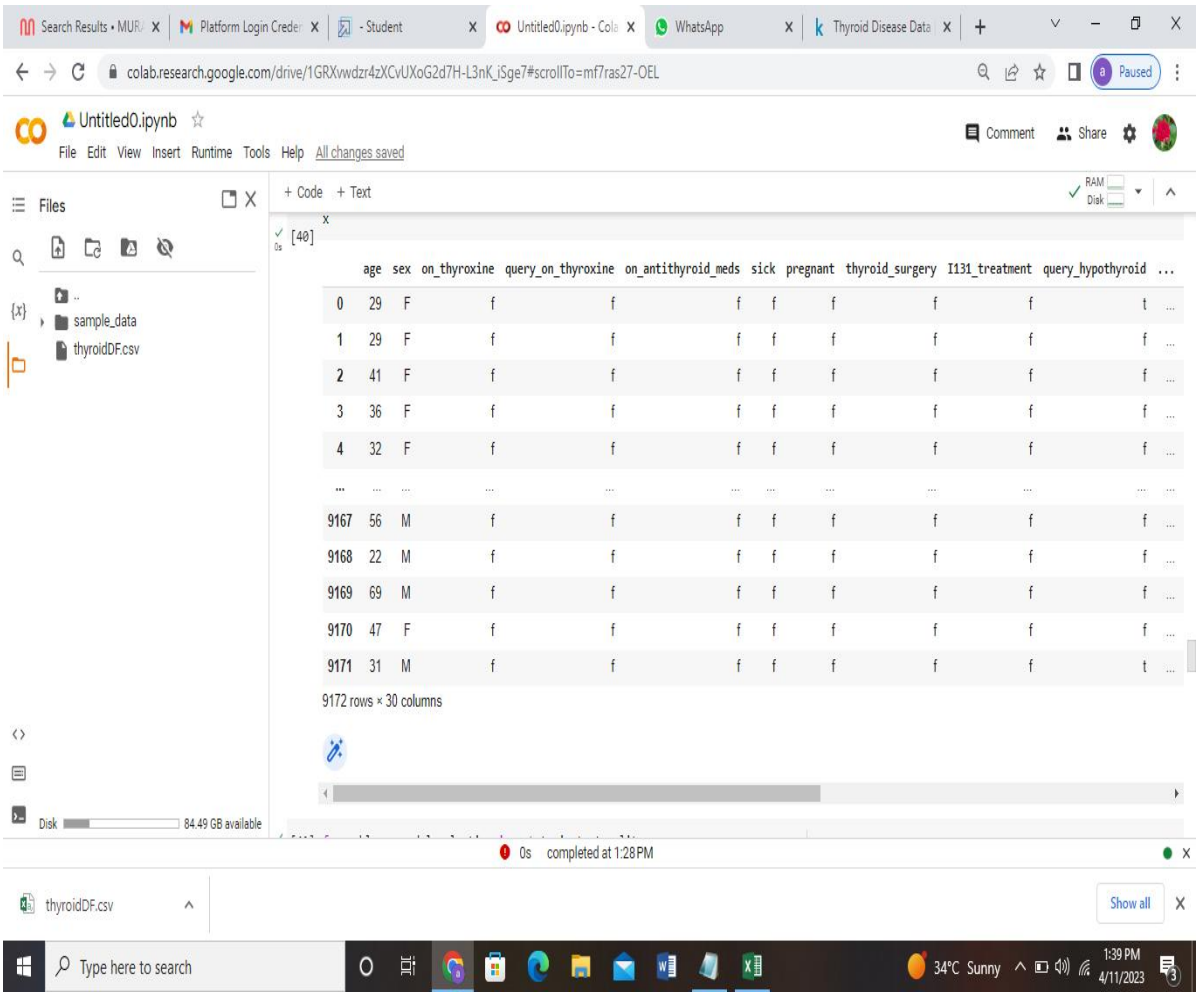
2.2 ideation & Brainstorming Map



3.RESULT

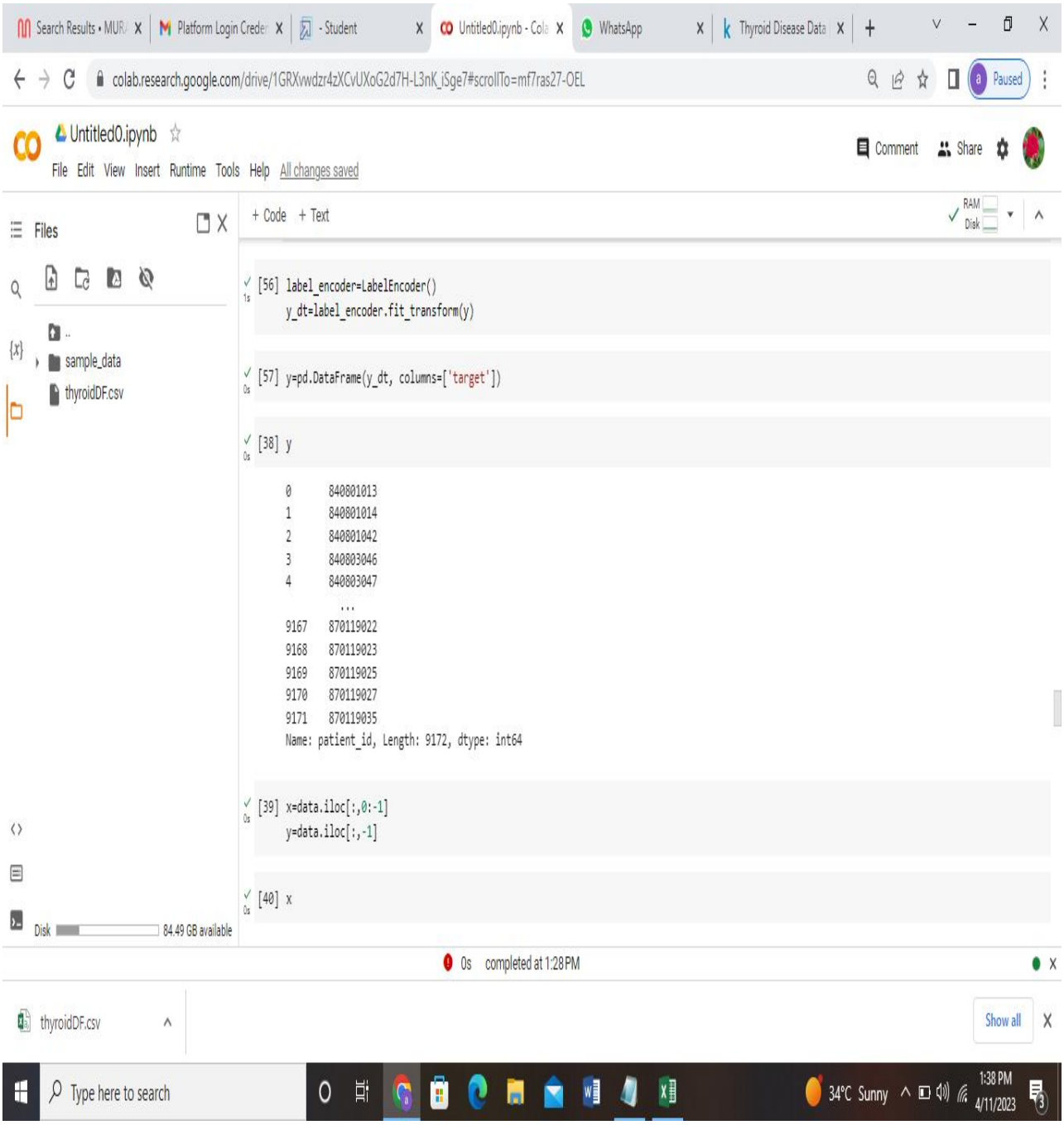
Thyroid Disease Analysis Coding Output

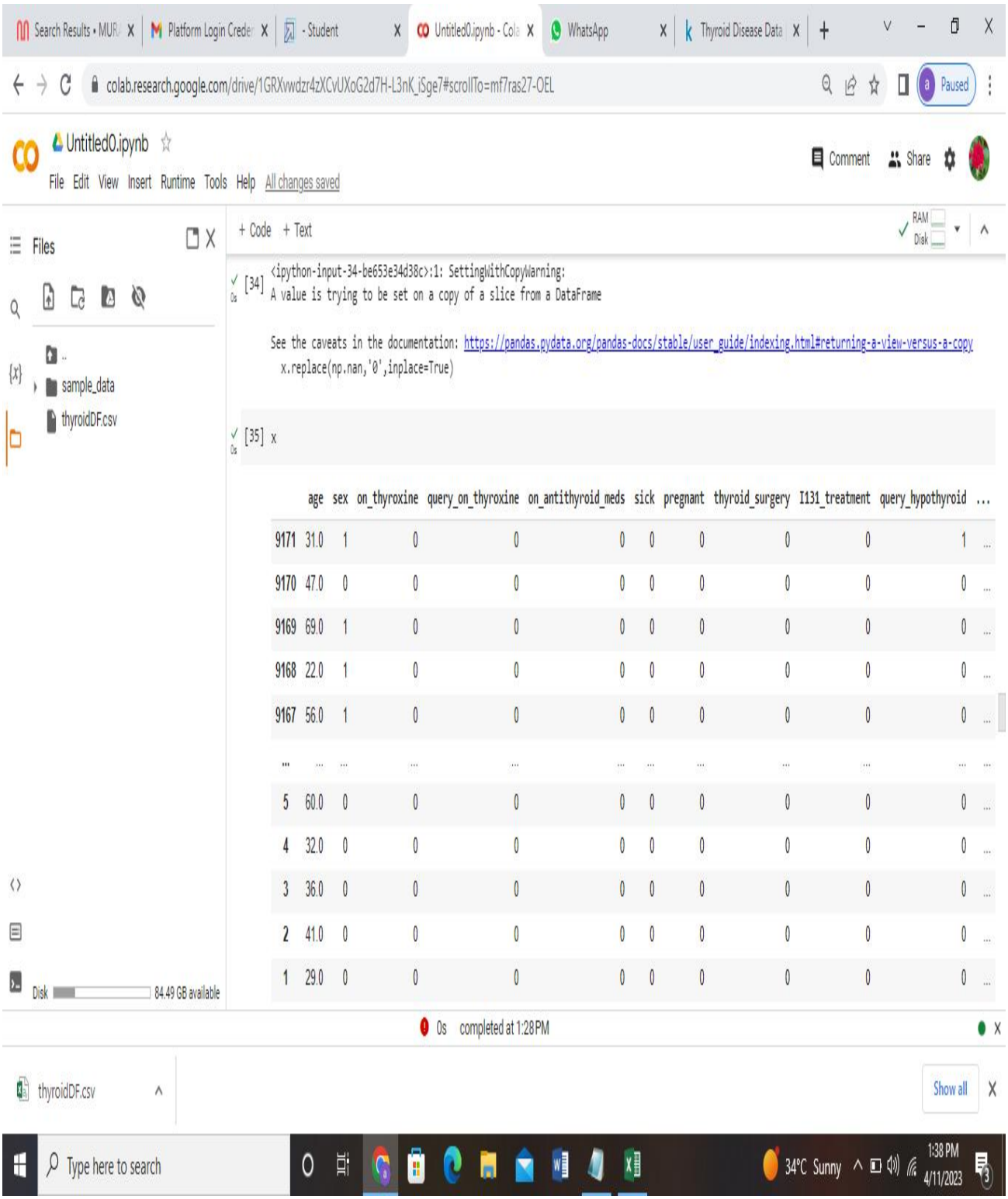




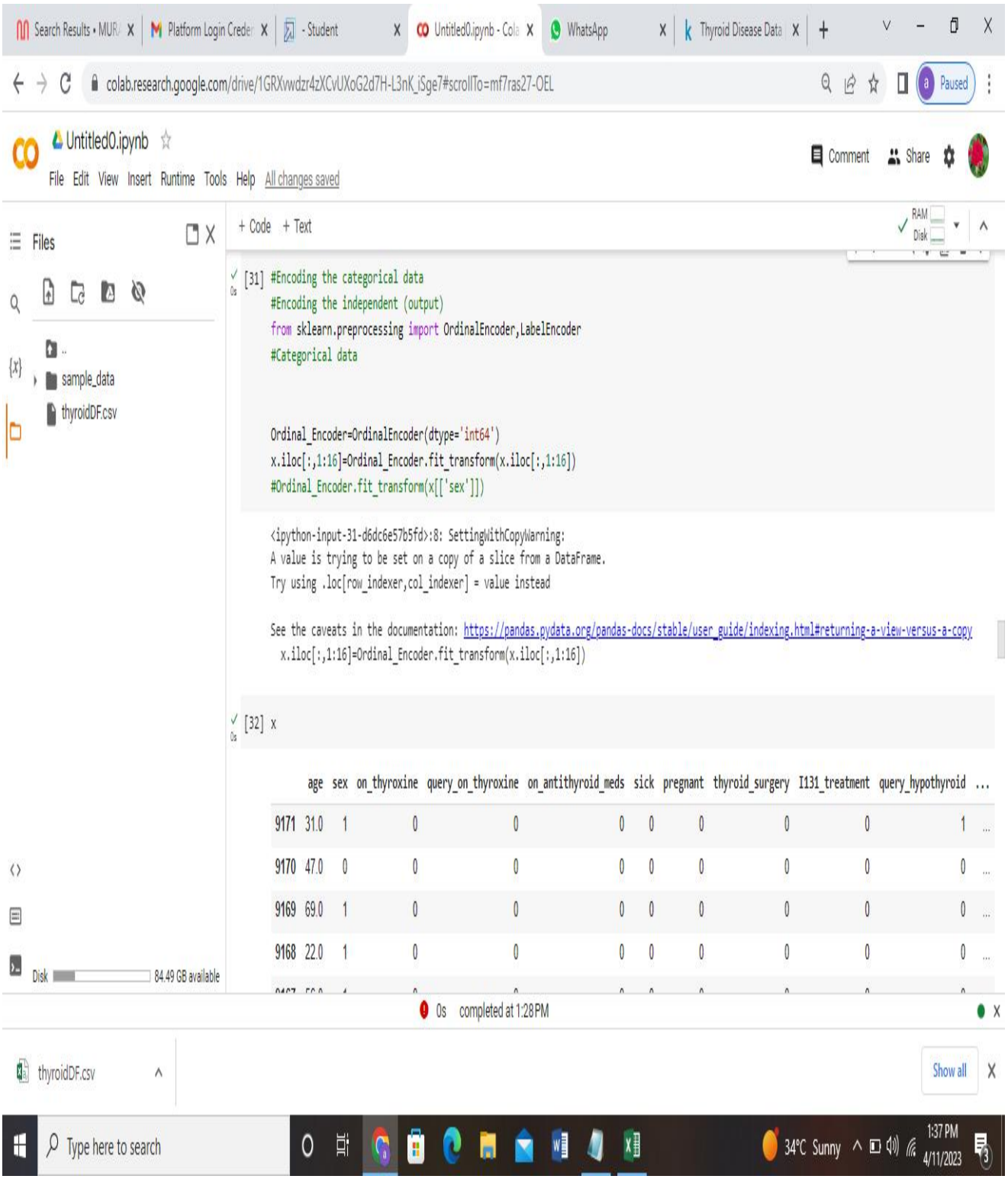
	age	sex	on_thyroxine	query_on_thyroxine	on_antithyroid_meds	sick	pregnant	thyroid_surgery	I131_treatment	query_hypothyroid	...
0	29	F	f	f	f	f	f	f	f	t	...
1	29	F	f	f	f	f	f	f	f	f	...
2	41	F	f	f	f	f	f	f	f	f	...
3	36	F	f	f	f	f	f	f	f	f	...
4	32	F	f	f	f	f	f	f	f	f	...
...	...	...	...	...	...	...	...	...	...	...	...
9167	56	M	f	f	f	f	f	f	f	f	...
9168	22	M	f	f	f	f	f	f	f	f	...
9169	69	M	f	f	f	f	f	f	f	f	...
9170	47	F	f	f	f	f	f	f	f	f	...
9171	31	M	f	f	f	f	f	f	f	t	...

9172 rows × 30 columns


Search Results • MUR
Platform Login Creder
- Student
Untitled0.ipynb - Cola
WhatsApp
Thyroid Disease Data
colab.research.google.com/drive/1GRXwdzr4zXCvUXoG2d7H-L3nK_iSge7#scrollTo=mf7ras27-OEL
Paused
Untitled0.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
sample_data
thyroidDF.csv
[56] label_encoder=LabelEncoder()
y_dt=label_encoder.fit_transform(y)
[57] y=pd.DataFrame(y_dt, columns=['target'])
[38] y
0 840801013
1 840801014
2 840801042
3 840803046
4 840803047
...
9167 870119022
9168 870119023
9169 870119025
9170 870119027
9171 870119035
Name: patient_id, Length: 9172, dtype: int64
[39] x=data.iloc[:,0:-1]
y=data.iloc[:,-1]
[40] x
Disk 84.49 GB available
0s completed at 1:28PM
thyroidDF.csv
Show all
Type here to search
34°C Sunny
1:38 PM
4/11/2023



	age	sex	on_thyroxine	query_on_thyroxine	on_antithyroid_meds	sick	pregnant	thyroid_surgery	I131_treatment	query_hypothyroid	...
9171	31.0	1	0	0	0	0	0	0	0	1	...
9170	47.0	0	0	0	0	0	0	0	0	0	...
9169	69.0	1	0	0	0	0	0	0	0	0	...
9168	22.0	1	0	0	0	0	0	0	0	0	...
9167	56.0	1	0	0	0	0	0	0	0	0	...
...	...	...	...	...	...	...	...	...	...	...	...
5	60.0	0	0	0	0	0	0	0	0	0	...
4	32.0	0	0	0	0	0	0	0	0	0	...
3	36.0	0	0	0	0	0	0	0	0	0	...
2	41.0	0	0	0	0	0	0	0	0	0	...
1	29.0	0	0	0	0	0	0	0	0	0	...



Search Results • MUR | Platform Login Creden | - Student | Untitled0.ipynb - Cola | WhatsApp | Thyroid Disease Data

colab.research.google.com/drive/1GRXvwdzr4zXCvUXoG2d7H-L3nK_iSge7#scrollTo=mf7ras27-OEL

Untitled0.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share Paused

Files

..
sample_data
thyroidDF.csv

Disk 84.49 GB available

+ Code + Text

RAM Disk

[31] #Encoding the categorical data
#Encoding the independent (output)
from sklearn.preprocessing import OrdinalEncoder,LabelEncoder
#Categorical data

Ordinal_Encoder=OrdinalEncoder(dtype='int64')
x.iloc[:,1:16]=Ordinal_Encoder.fit_transform(x.iloc[:,1:16])
#Ordinal_Encoder.fit_transform(x[['sex']])

<ipython-input-31-d6dc6e57b5fd>:8: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
x.iloc[:,1:16]=Ordinal_Encoder.fit_transform(x.iloc[:,1:16])

[32] x

	age	sex	on_thyroxine	query_on_thyroxine	on_antithyroid_meds	sick	pregnant	thyroid_surgery	I131_treatment	query_hypothyroid	...
9171	31.0	1	0	0	0	0	0	0	0	1	...
9170	47.0	0	0	0	0	0	0	0	0	0	...
9169	69.0	1	0	0	0	0	0	0	0	0	...
9168	22.0	1	0	0	0	0	0	0	0	0	...

0s completed at 1:28PM

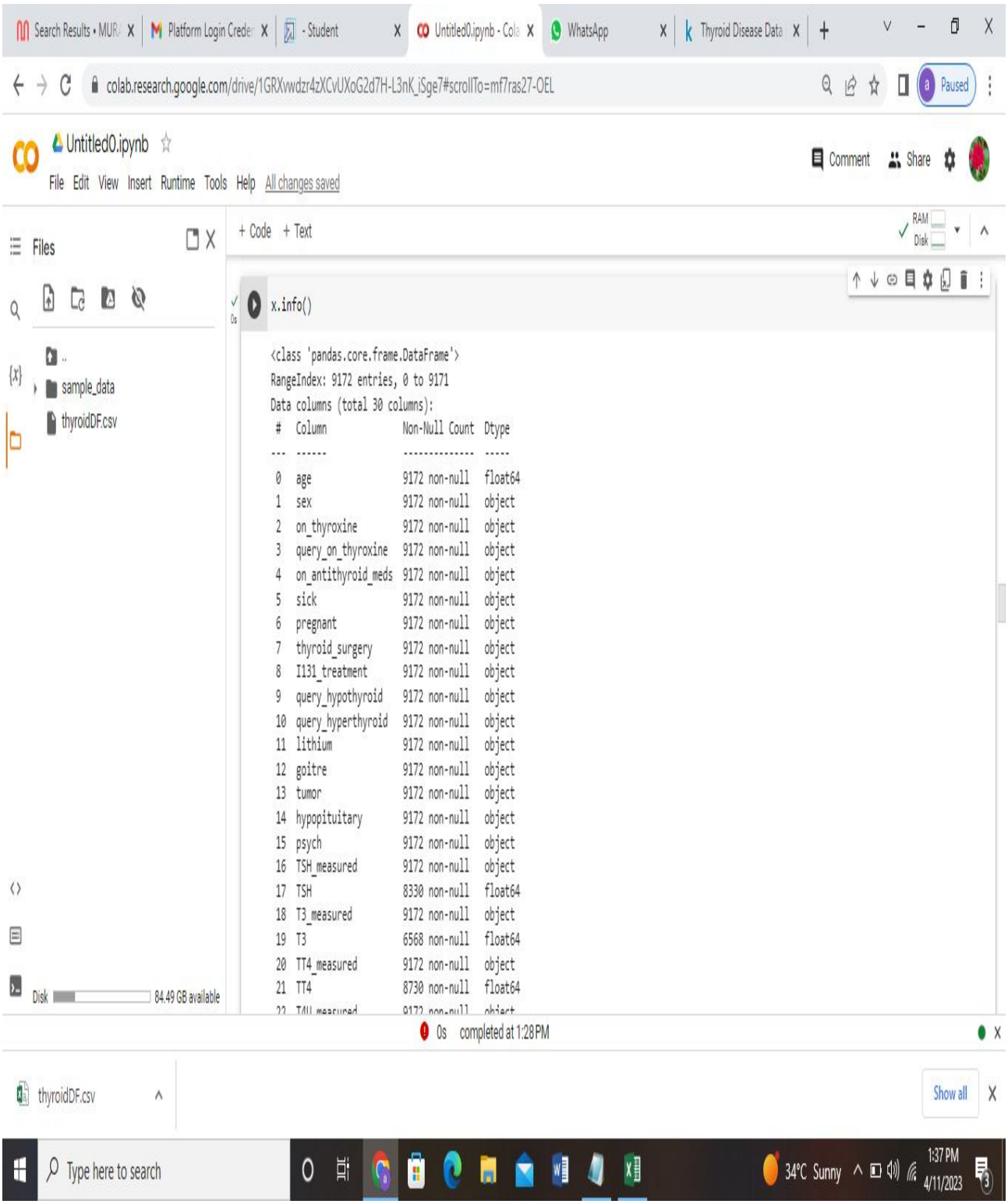
thyroidDF.csv

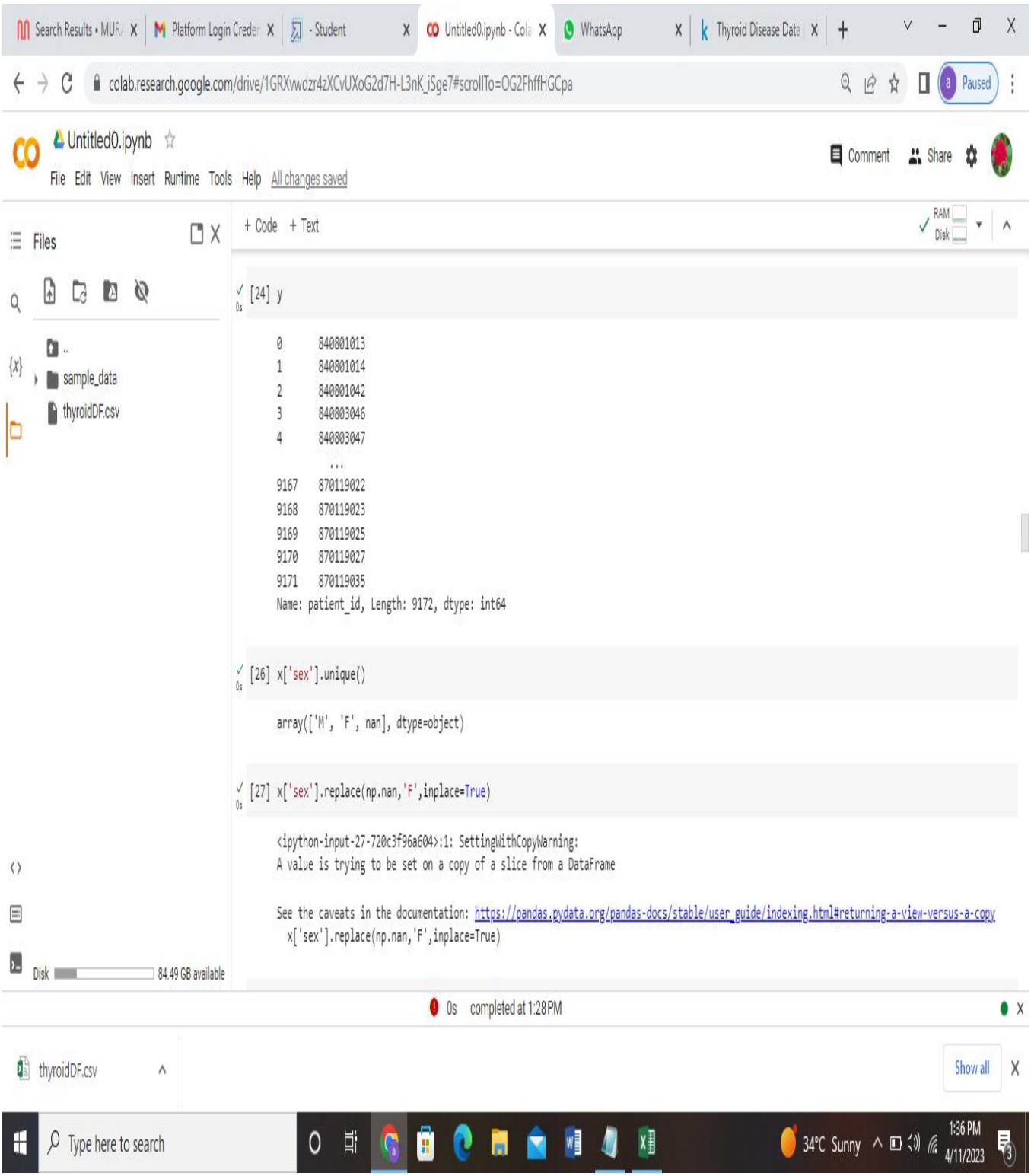
Show all

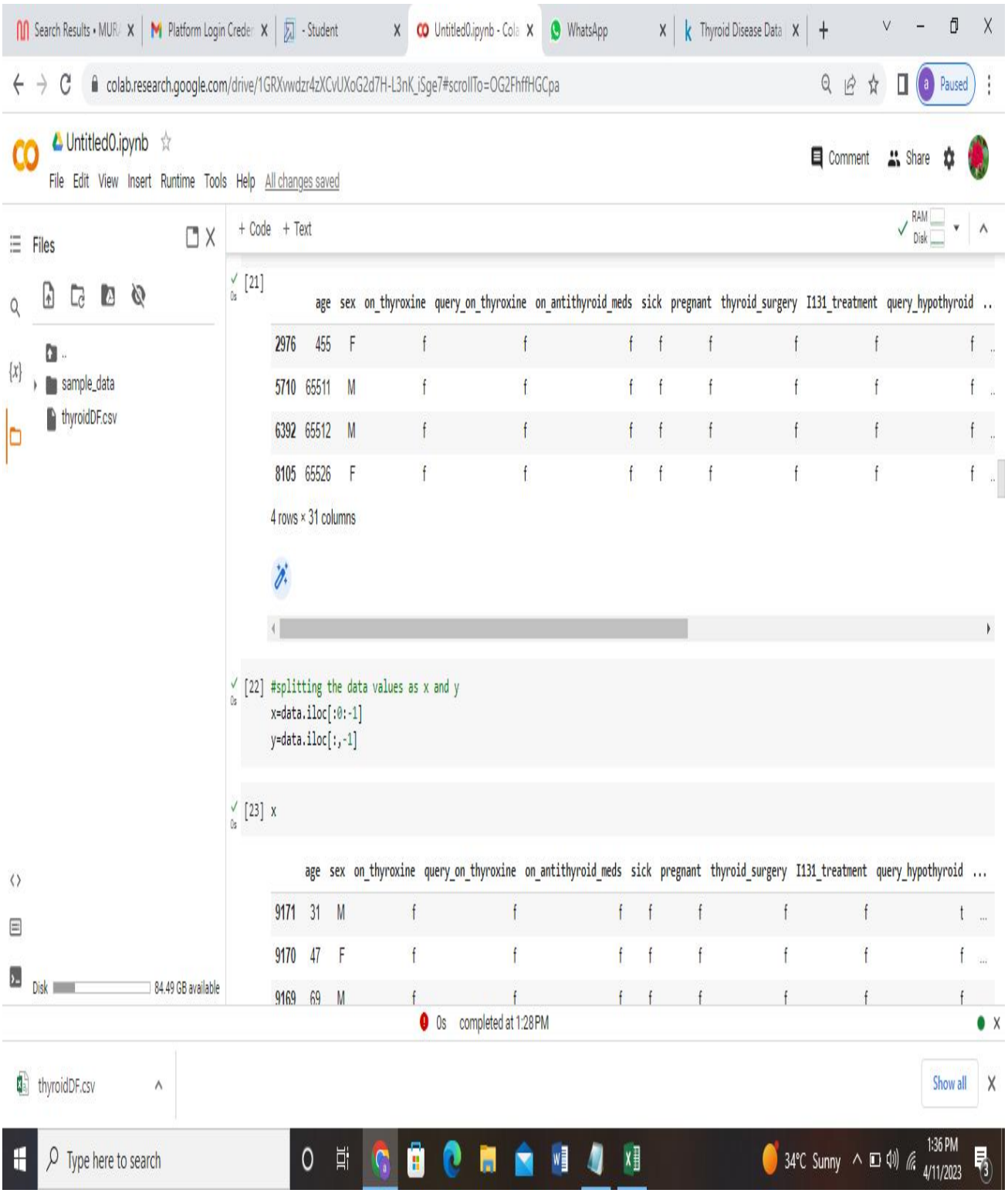
Type here to search

34°C Sunny

1:37 PM 4/11/2023


Search Results • MUR
Platform Login Creden
- Student
Untitled0.ipynb - Cola
WhatsApp
Thyroid Disease Data
colab.research.google.com/drive/1GRXvwdzr4zXCvUXoG2d7H-L3nK_iSge7#scrollTo=mf7ras27-OEL
Paused
Untitled0.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
x.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9172 entries, 0 to 9171
Data columns (total 30 columns):
Column Non-Null Count Dtype
0 age 9172 non-null float64
1 sex 9172 non-null object
2 on_thyroxine 9172 non-null object
3 query_on_thyroxine 9172 non-null object
4 on_antithyroid_meds 9172 non-null object
5 sick 9172 non-null object
6 pregnant 9172 non-null object
7 thyroid_surgery 9172 non-null object
8 I131_treatment 9172 non-null object
9 query_hypothyroid 9172 non-null object
10 query_hyperthyroid 9172 non-null object
11 lithium 9172 non-null object
12 goitre 9172 non-null object
13 tumor 9172 non-null object
14 hypopituitary 9172 non-null object
15 psych 9172 non-null object
16 TSH_measured 9172 non-null object
17 TSH 8330 non-null float64
18 T3_measured 9172 non-null object
19 T3 6568 non-null float64
20 TT4_measured 9172 non-null object
21 TT4 8730 non-null float64
sample_data
thyroidDF.csv
Disk 84.49 GB available
0s completed at 1:28PM
thyroidDF.csv
Show all
Type here to search
34°C Sunny
1:37 PM
4/11/2023

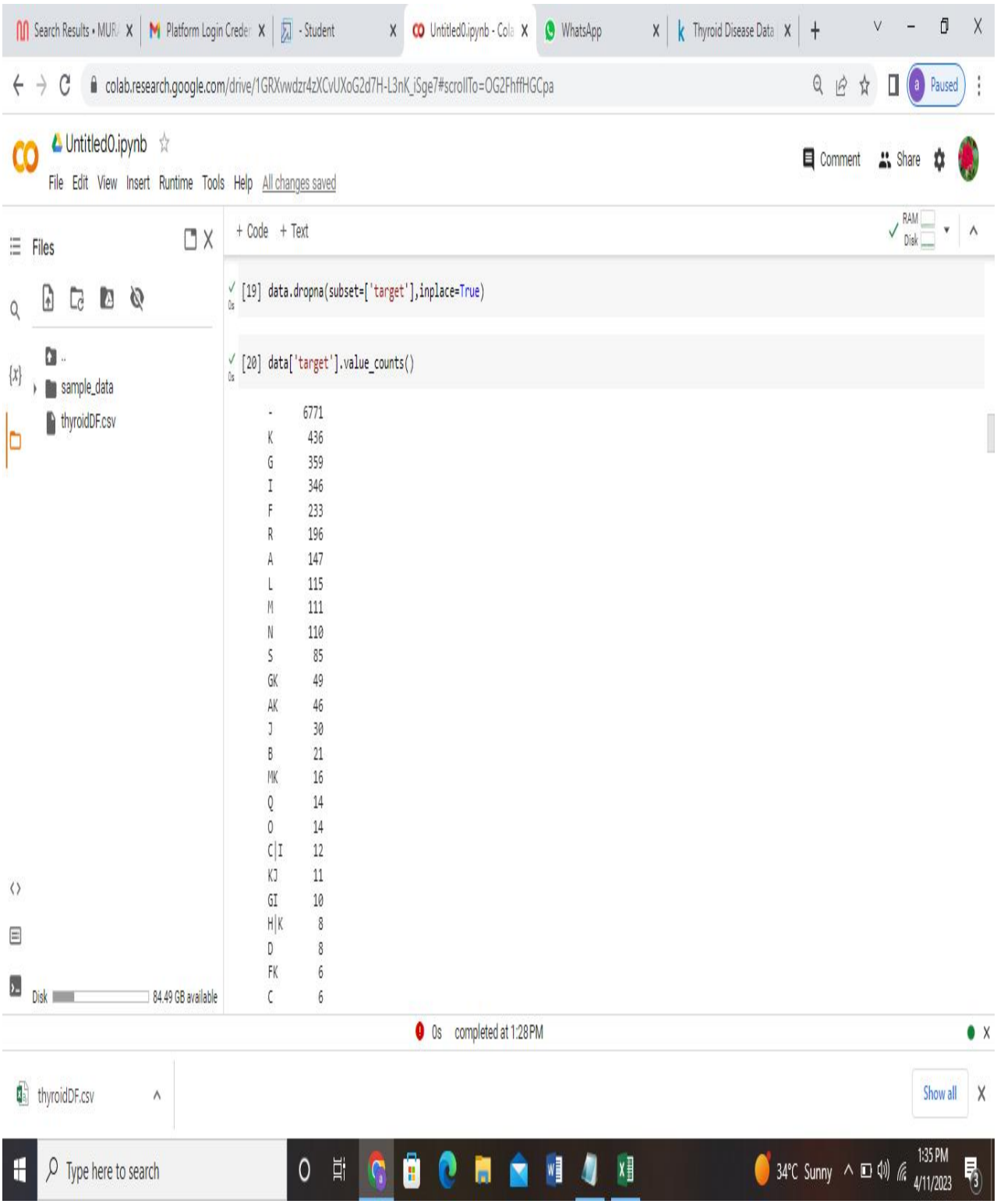

Search Results • MUR
Platform Login Creden
- Student
Untitled0.ipynb - Cola
WhatsApp
Thyroid Disease Data
colab.research.google.com/drive/1GRXvwdzr4zXCvUXoG2d7H-L3nK_iSge7#scrollTo=OG2FhffHGCpa
Paused
Untitled0.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
..
sample_data
thyroidDF.csv
[24] y
0 840801013
1 840801014
2 840801042
3 840803046
4 840803047
...
9167 870119022
9168 870119023
9169 870119025
9170 870119027
9171 870119035
Name: patient_id, Length: 9172, dtype: int64
[26] x['sex'].unique()
array(['M', 'F', nan], dtype=object)
[27] x['sex'].replace(np.nan,'F',inplace=True)
<ipython-input-27-720c3f96a604>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
x['sex'].replace(np.nan,'F',inplace=True)
Disk 84.49 GB available
0s completed at 1:28 PM
thyroidDF.csv
Show all
Type here to search
34°C Sunny
1:36 PM
4/11/2023

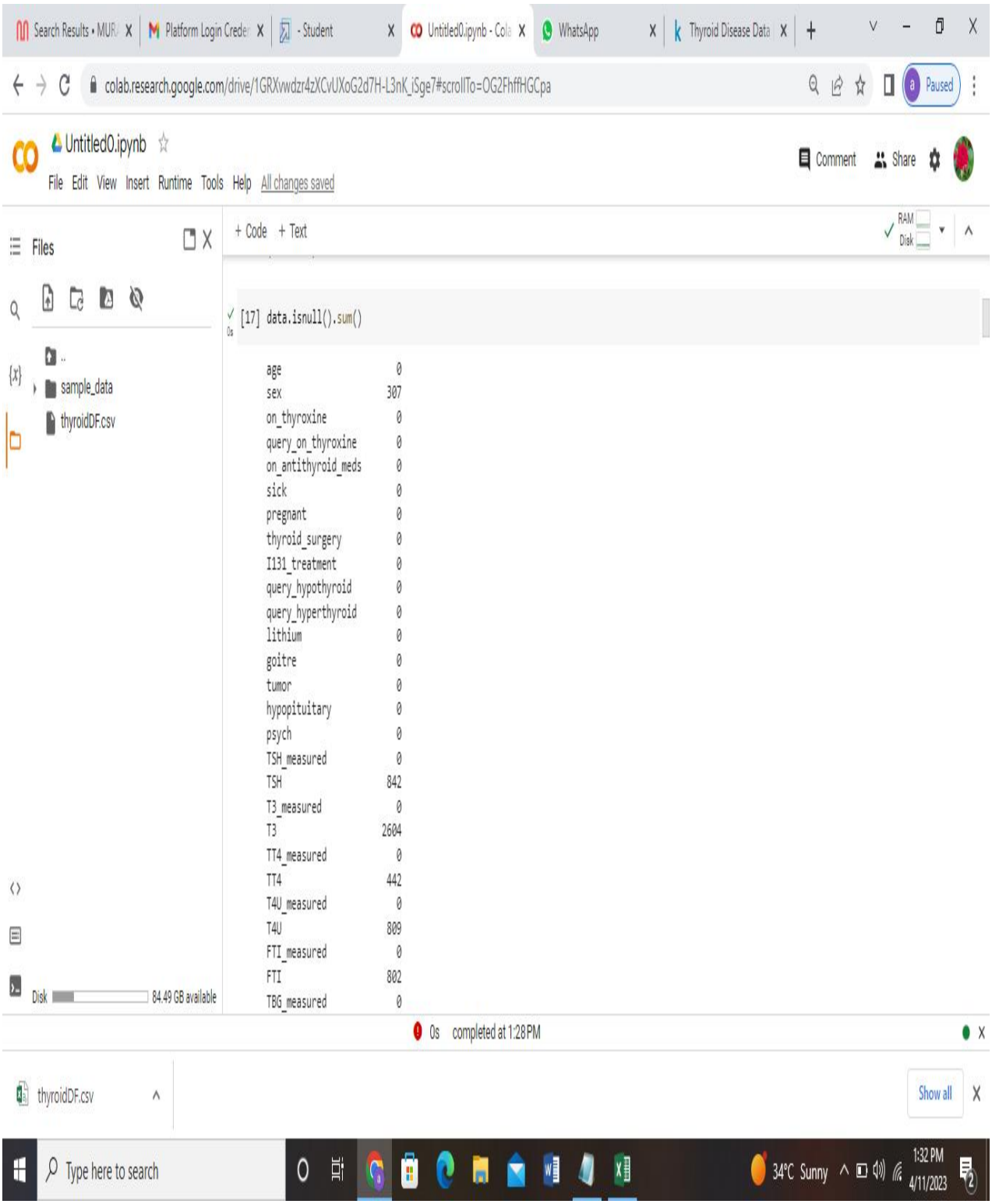


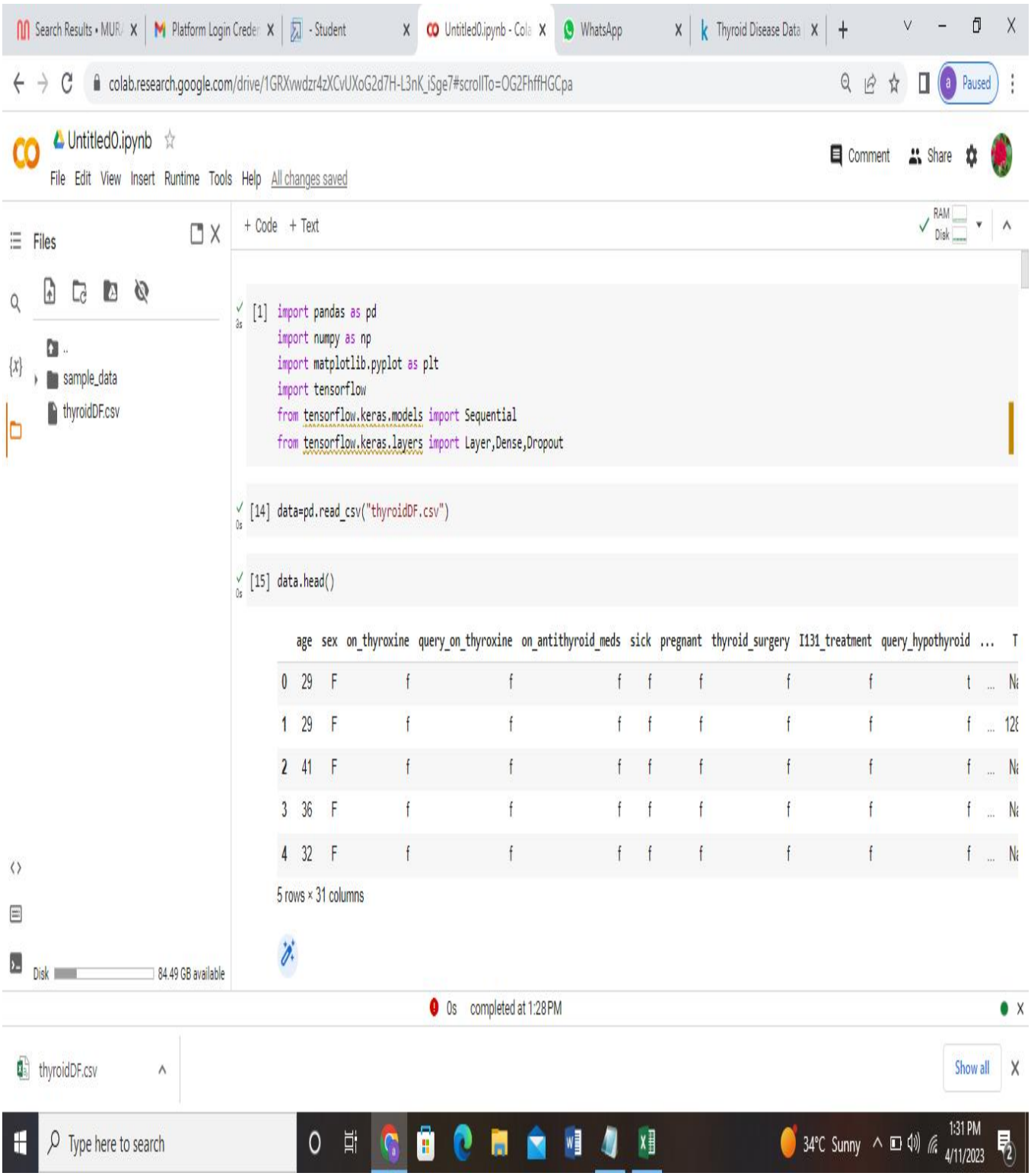
	age	sex	on_thyroxine	query_on_thyroxine	on_antithyroid_meds	sick	pregnant	thyroid_surgery	I131_treatment	query_hypothyroid	..
2976	455	F	f	f	f	f	f	f	f	f	..
5710	65511	M	f	f	f	f	f	f	f	f	..
6392	65512	M	f	f	f	f	f	f	f	f	..
8105	65526	F	f	f	f	f	f	f	f	f	..

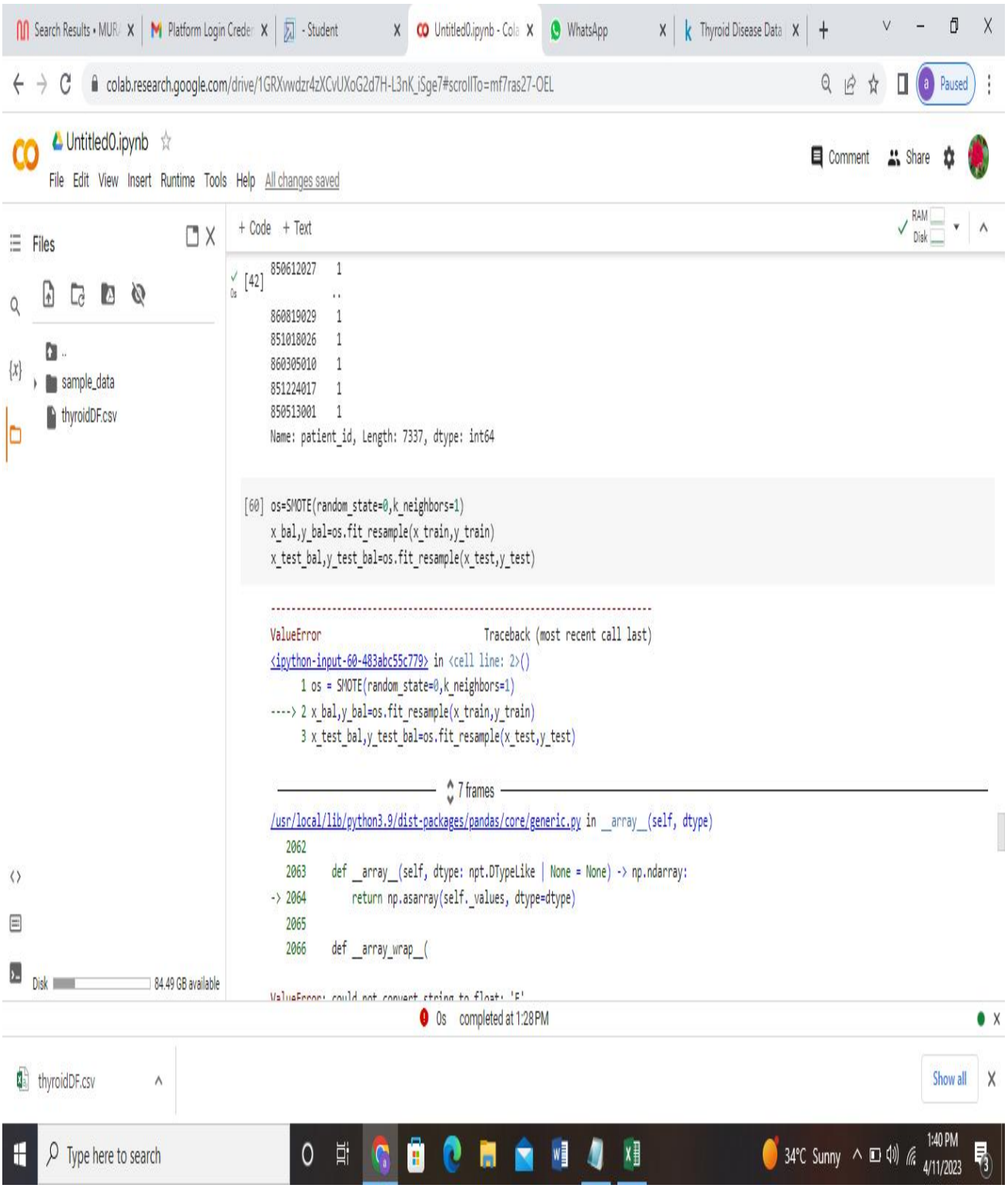
4 rows × 31 columns

	age	sex	on_thyroxine	query_on_thyroxine	on_antithyroid_meds	sick	pregnant	thyroid_surgery	I131_treatment	query_hypothyroid	...
9171	31	M	f	f	f	f	f	f	f	t	...
9170	47	F	f	f	f	f	f	f	f	f	...
9169	69	M	f	f	f	f	f	f	f	f	


Search Results • MUR
Platform Login Creden
- Student
Untitled0.ipynb - Cola
WhatsApp
Thyroid Disease Data
colab.research.google.com/drive/1GRXvwdzr4zXCvUXoG2d7H-L3nK_iSge7#scrollTo=OG2FhffHGCpa
Paused
Untitled0.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
[19] data.dropna(subset=['target'],inplace=True)
[20] data['target'].value_counts()
- 6771
K 436
G 359
I 346
F 233
R 196
A 147
L 115
M 111
N 110
S 85
GK 49
AK 46
J 30
B 21
MK 16
Q 14
O 14
C|I 12
KJ 11
GI 10
H|K 8
D 8
FK 6
C 6
sample_data
thyroidDF.csv
Disk
84.49 GB available
0s completed at 1:28PM
thyroidDF.csv
Show all
Type here to search
34°C Sunny
1:35 PM
4/11/2023


Search Results • MUR
Platform Login Creden
- Student
Untitled0.ipynb - Cola
WhatsApp
Thyroid Disease Data
colab.research.google.com/drive/1GRXvwdzr4zXCvUXoG2d7H-L3nK_iSge7#scrollTo=OG2FhffHGCpa
Paused
Untitled0.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
[17] data.isnull().sum()
0s
age 0
sex 307
on_thyroxine 0
query_on_thyroxine 0
on_antithyroid_meds 0
sick 0
pregnant 0
thyroid_surgery 0
I131_treatment 0
query_hypothyroid 0
query_hyperthyroid 0
lithium 0
goitre 0
tumor 0
hypopituitary 0
psych 0
TSH_measured 0
TSH 842
T3_measured 0
T3 2604
TT4_measured 0
TT4 442
T4U_measured 0
T4U 809
FTI_measured 0
FTI 802
TBG_measured 0
..
sample_data
thyroidDF.csv
Disk 84.49 GB available
0s completed at 1:28PM
thyroidDF.csv
Show all
Type here to search
34°C Sunny
1:32 PM
4/11/2023


Search Results • MUR
Platform Login Creder
- Student
Untitled0.ipynb - Cola
WhatsApp
Thyroid Disease Data
colab.research.google.com/drive/1GRXvwdzr4zXCvUXoG2d7H-L3nK_iSge7#scrollTo=OG2FhffHGCpa
Paused
Untitled0.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
sample_data
thyroidDF.csv
[1] import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import tensorflow
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Layer,Dense,Dropout
[14] data=pd.read_csv("thyroidDF.csv")
[15] data.head()
age sex on_thyroxine query_on_thyroxine on_antithyroid_meds sick pregnant thyroid_surgery I131_treatment query_hypothyroid ... T
0 29 F f f f f f f f t ... Na
1 29 F f f f f f f f f ... 128
2 41 F f f f f f f f f ... Na
3 36 F f f f f f f f f ... Na
4 32 F f f f f f f f f ... Na
5 rows × 31 columns
Disk 84.49 GB available
0s completed at 1:28 PM
thyroidDF.csv
Show all
Type here to search
34°C Sunny
1:31 PM
4/11/2023


Search Results • MUR
Platform Login Crede
- Student
Untitled0.ipynb - Cola
WhatsApp
Thyroid Disease Data
colab.research.google.com/drive/1GRXwdzr4zXCvUXoG2d7H-L3nK_iSge7#scrollTo=mf7ras27-OEL
Paused
Untitled0.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
sample_data
thyroidDF.csv
[42]
850612027 1
860819029 1
851018026 1
860305010 1
851224017 1
850513001 1
Name: patient_id, Length: 7337, dtype: int64
[60] os=SMOTE(random_state=0,k_neighbors=1)
x_bal,y_bal=os.fit_resample(x_train,y_train)
x_test_bal,y_test_bal=os.fit_resample(x_test,y_test)
ValueError Traceback (most recent call last)
<ipython-input-60-483abc55c779> in <cell line: 2>()
1 os = SMOTE(random_state=0,k_neighbors=1)
----> 2 x_bal,y_bal=os.fit_resample(x_train,y_train)
3 x_test_bal,y_test_bal=os.fit_resample(x_test,y_test)
7 frames
/usr/local/lib/python3.9/dist-packages/pandas/core/generic.py in __array__(self, dtype)
2062
2063 def __array__(self, dtype: npt.DTypeLike | None = None) -> np.ndarray:
-> 2064 return np.asarray(self._values, dtype=dtype)
2065
2066 def __array_wrap__(
Disk 84.49 GB available
0s completed at 1:28PM
thyroidDF.csv
Show all
Type here to search
34°C Sunny
1:40 PM
4/11/2023

4. Thyroid disease Advantages and Disadvantages

Advantages	Disadvantages
No recurrent hyperthyroidism	Risk of postoperative hypoparathyroidism
No radiation risk	Risk of recurrent nerve palsy
Rapid control of hyperthyroidism	Permanent hypothyroidism
No reported detrimental effect on the course of Graves' orbitopathy	Risks related to anesthesia or surgery
	Hospitalization
	Costs
	Permanent scar

5.APPLICATION

- The thyroid gland is a vital hormone gland: It plays a major role in the metabolism, growth and development of the human body. It helps to regulate many body functions by constantly releasing a steady amount of thyroid hormones into the bloodstream.
- Hospitals.
- Medical research Department.

6.CONCLUSION

- The thyroid gland maintains the level of metabolism in the tissues that is optimal for their normal function. Thyroid hormone stimulates the o2 consumption of most of the cells in the body, regulates lipid and carbohydrate metabolism, and is also necessary for normal growth and maturation.

7.FUTURE SCOPE

- Hospitals.
- Specialty Clinics.
- Medical Research Department.
- Patient(Body).
- Getting enough Sleep.
- Exercising regularly.
- Taking Medications properly.
- Getting tested regularly.
- Watching your diet.

8.APPENDIX

Source code

```
# -*- coding: utf-8 -*-
"""Untitled0.ipynb

Automatically generated by Colaboratory.

Original file is located at
    https://colab.research.google.com/drive/1GRXvwdzr4zXCvUXoG2d7H-
L3nK_iSge7
"""
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import tensorflow
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Layer,Dense,Dropout

data=pd.read_csv("thyroidDF.csv")

data.head()

data.shape

data.isnull().sum()

data.drop(['TSH_measured','T3_measured','TT4_measured','T4U_measured','TBG
_measured','referral_source','patient_id'])

#remapping target valueas ton diagnostic group
dignoses={'A':'hypothyroid conditions',
        'B':'hypothyroid conditions',
        'C':'hypothyroid conditions',
        'D':'hypothyroid conditions',
        'E':'hypothyroid conditions',
        'F':'hypothyroid conditions',
        'G':'hypothyroid conditions',
        'H':'hypothyroid conditions',
        'I':'binding protein',
        'J':'binding protein',
        'K':'binding protein',
        'L':'replacement therapy',
        'M':'replacement therapy',
        'N':'replacement therapy',
        'O':'antithyroid treatment',
        'P':'antithyroid treatment',
        'Q':'antithyroid treatment',
        'R':'miscellaneous',
        'S':'miscellaneous',
        'T':'miscellaneous'}
data['target']=data['target'].map(diagnoses)#remapping

data.dropna(subset=['target'],inplace=True)

data['target'].value_counts()

data[data.age>100]

#splitting the data values as x and y
x=data.iloc[:0:-1]
y=data.iloc[:,-1]

x
```

```
y

x['sex'].unique()

x['sex'].replace(np.nan,'F',inplace=True)

x['sex'].value_counts()

#Converting the data type

x['age']=x['age'].astype('float')
x['TSH']=x['TSH'].astype('float')
x['T3']=x['T3'].astype('float')
x['TT4']=x['TT4'].astype('float')
x['T4U']=x['T4U'].astype('float')
x['FTI']=x['FTI'].astype('float')
x['TBG']=x['TBG'].astype('float')

x.info()

#Encoding the categorical data
#Encoding the independent (output)
from sklearn.preprocessing import OrdinalEncoder,LabelEncoder
#Categorical data


Ordinal_Encoder=OrdinalEncoder(dtype='int64')
x.iloc[:,1:16]=Ordinal_Encoder.fit_transform(x.iloc[:,1:16])
#Ordinal_Encoder.fit_transform(x[['sex']])

x

x.replace(np.nan,'0',inplace=True)

x

label_encoder=LabelEncoder()
y_dt=label_encoder.fit_transform(y)

y=pd.DataFrame(y_dt, columns=['target'])

y

x=data.iloc[:,0:-1]
y=data.iloc[:,-1]

x

from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.20,random_s
tate=0)

from imblearn.over_sampling import SMOTE
y_train.value_counts()
```

```
os=SMOTE(random_state=0,k_neighbors=1)
x_bal,y_bal=os.fit_resample(x_train,y_train)
x_test_bal,y_test_bal=os.fit_resample(x_test,y_test)

from sklearn.preprocessing import StandardScalar
sc=StandardScalar()
x_bal=sc.fit_transform(x_bal)
x_test_bal=sc.transform(x_test_bal)

x_bal

x_test_bal=pd.DataFrame(x_test_bal,columns=columns)

x_bal=pd.DataFrame(x_bal,columns=columns)

x_bal

from sklearn.inspection import permutation_importance
results=permutation_importance(rtf,x_bal,y_bal,scoring='accuracy')
```