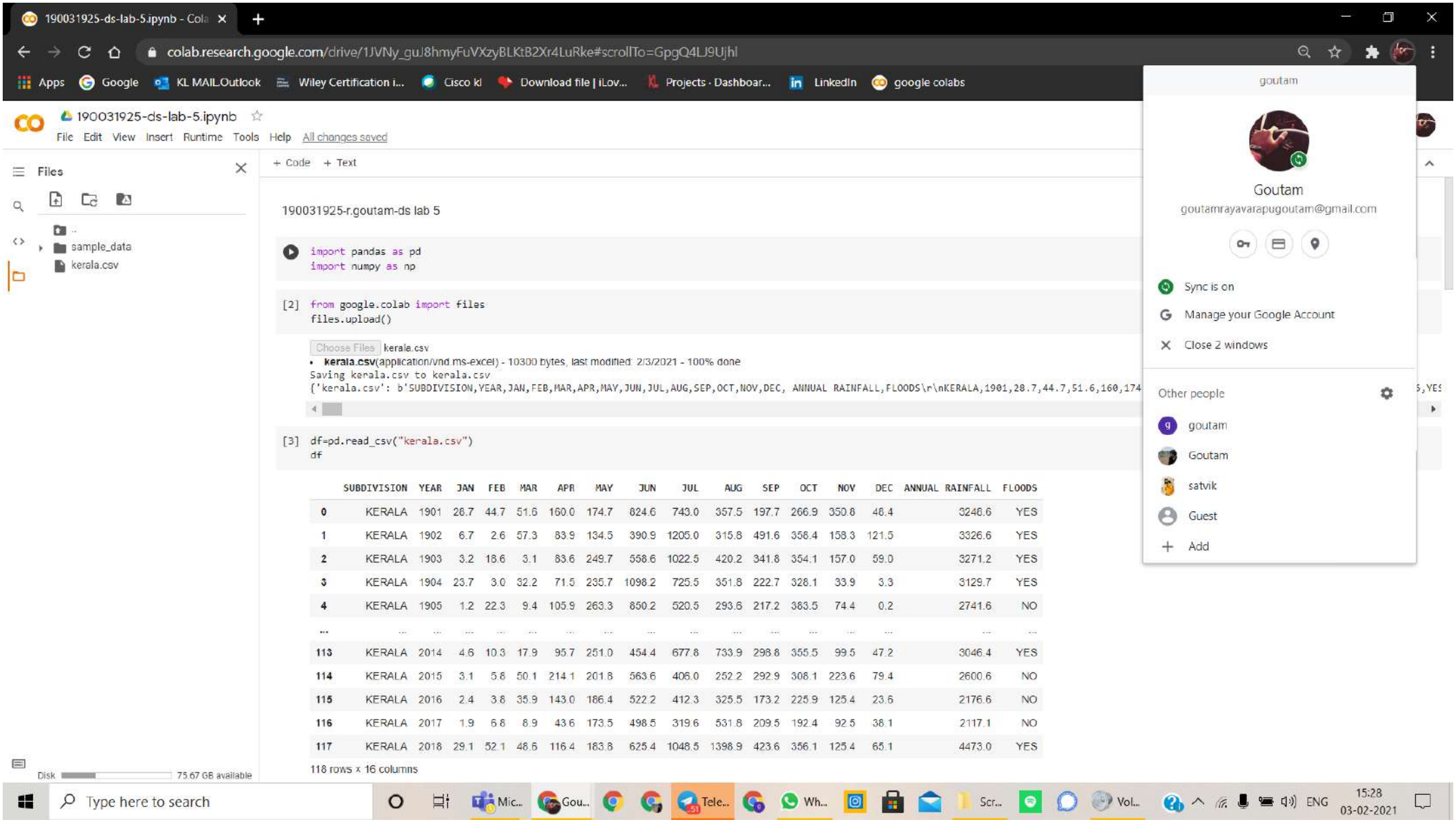
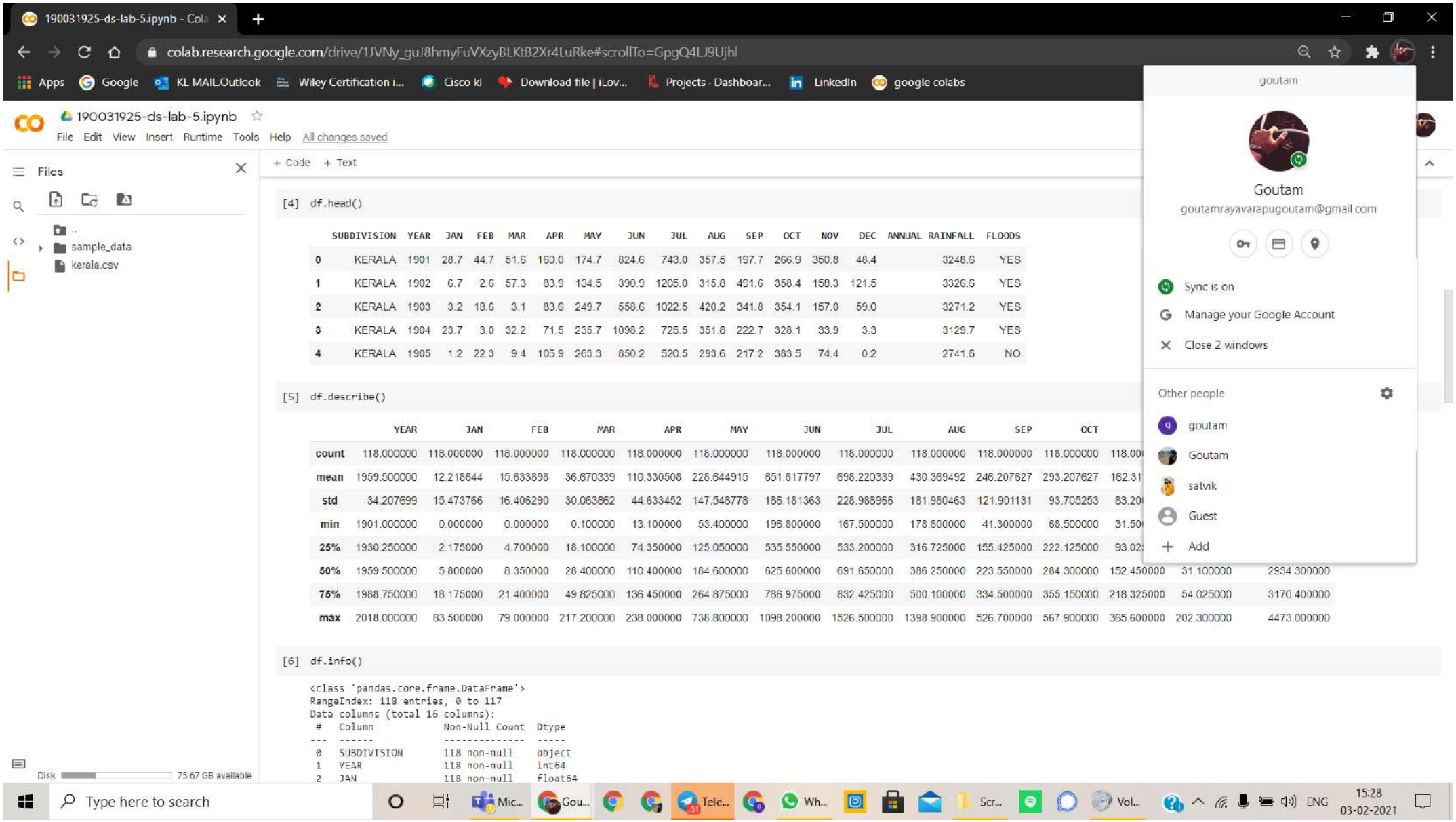
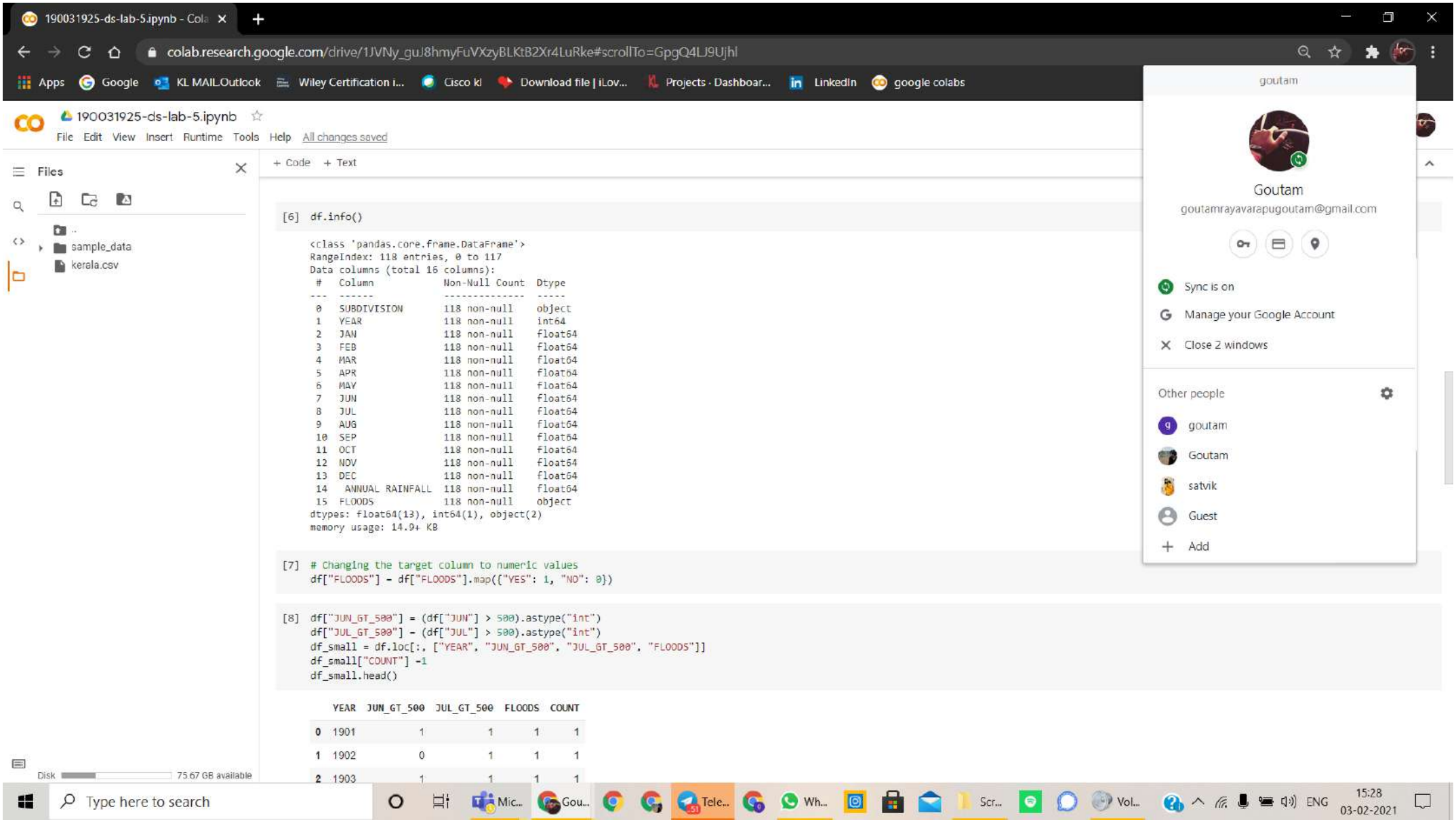

190031925-ds-lab-5.ipynb - Cola
colab.research.google.com/drive/1JVNy_guJ8hmyFuVXzyBLKtB2Xr4LuRke#scrollTo=GpgQ4LJ9Ujhl
Apps
Google
KL MAIL.Outlook
Wiley Certification i...
Cisco kl
Download file | iLov...
Projects - Dashboar...
LinkedIn
google colabs
190031925-ds-lab-5.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Files
sample_data
kerala.csv
+ Code + Text
190031925-r.goutam-ds lab 5
import pandas as pd
import numpy as np
[2] from google.colab import files
files.upload()
Choose Files kerala.csv
kerala.csv(application/vnd.ms-excel) - 10300 bytes, last modified: 2/3/2021 - 100% done
Saving kerala.csv to kerala.csv
{'kerala.csv': b'SUBDIVISION,YEAR,JAN,FEB,MAR,APR,MAY,JUN,JUL,AUG,SEP,OCT,NOV,DEC, ANNUAL RAINFALL,FLOODS\r\nKERALA,1901,28.7,44.7,51.6,160,174
[3] df=pd.read_csv("kerala.csv")
df
SUBDIVISION YEAR JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC ANNUAL RAINFALL FLOODS
0 KERALA 1901 28.7 44.7 51.6 160.0 174.7 824.6 743.0 357.5 197.7 266.9 350.8 48.4 3246.6 YES
1 KERALA 1902 6.7 2.6 57.3 83.9 134.5 390.9 1205.0 315.8 491.6 358.4 158.3 121.5 3326.6 YES
2 KERALA 1903 3.2 18.6 3.1 83.6 249.7 558.6 1022.5 420.2 341.8 354.1 157.0 59.0 3271.2 YES
3 KERALA 1904 23.7 3.0 32.2 71.5 235.7 1098.2 725.5 351.8 222.7 328.1 33.9 3.3 3129.7 YES
4 KERALA 1905 1.2 22.3 9.4 105.9 263.3 850.2 520.5 293.6 217.2 383.5 74.4 0.2 2741.6 NO
113 KERALA 2014 4.6 10.3 17.9 95.7 251.0 454.4 677.8 733.9 298.8 355.5 99.5 47.2 3046.4 YES
114 KERALA 2015 3.1 5.8 50.1 214.1 201.8 563.6 406.0 252.2 292.9 308.1 223.6 79.4 2600.6 NO
115 KERALA 2016 2.4 3.8 35.9 143.0 186.4 522.2 412.3 325.5 173.2 225.9 125.4 23.6 2176.6 NO
116 KERALA 2017 1.9 6.8 8.9 43.6 173.5 498.5 319.6 531.8 209.5 192.4 92.5 38.1 2117.1 NO
117 KERALA 2018 29.1 52.1 48.6 116.4 183.8 625.4 1048.5 1398.9 423.6 356.1 125.4 65.1 4473.0 YES
118 rows x 16 columns
Disk 75.67 GB available
goutam
Goutam
goutamrayavarapugoutam@gmail.com
Sync is on
Manage your Google Account
Close 2 windows
Other people
goutam
Goutam
satvik
Guest
Add
Type here to search
ENG
15:28
03-02-2021



	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL RAINFALL	FLOODS
0	KERALA	1901	28.7	44.7	51.6	160.0	174.7	824.6	743.0	357.5	197.7	266.9	350.8	48.4	3248.6	YES
1	KERALA	1902	6.7	2.6	57.3	83.9	134.5	390.9	1205.0	315.8	491.6	358.4	158.3	121.5	3326.6	YES
2	KERALA	1903	3.2	18.6	3.1	83.6	249.7	558.6	1022.5	420.2	341.8	354.1	157.0	59.0	3271.2	YES
3	KERALA	1904	23.7	3.0	32.2	71.5	235.7	1098.2	725.5	351.8	222.7	328.1	33.9	3.3	3129.7	YES
4	KERALA	1905	1.2	22.3	9.4	105.9	263.3	850.2	520.5	293.6	217.2	383.5	74.4	0.2	2741.6	NO

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT			
count	118.000000	118.000000	118.000000	118.000000	118.000000	118.000000	118.000000	118.000000	118.000000	118.000000	118.000000	118.00		
mean	1959.500000	12.218644	15.633898	36.670339	110.330508	228.644915	651.617797	698.220339	430.369492	246.207627	293.207627	162.31		
std	34.207699	15.473766	16.406290	30.063862	44.633452	147.548778	186.161363	228.988966	181.980463	121.901131	93.705253	83.20		
min	1901.000000	0.000000	0.000000	0.100000	13.100000	53.400000	196.800000	167.500000	178.600000	41.300000	68.500000	31.50		
25%	1930.250000	2.175000	4.700000	18.100000	74.350000	125.050000	535.550000	533.200000	316.725000	155.425000	222.125000	93.02		
50%	1959.500000	5.800000	8.350000	28.400000	110.400000	184.600000	625.600000	691.650000	386.250000	223.550000	284.300000	152.450000	31.100000	2934.300000
75%	1988.750000	18.175000	21.400000	49.825000	136.450000	264.875000	786.975000	832.425000	500.100000	334.500000	355.150000	218.325000	54.025000	3170.400000
max	2018.000000	83.500000	79.000000	217.200000	238.000000	738.800000	1098.200000	1526.500000	1398.900000	526.700000	567.900000	365.600000	202.300000	4473.000000


190031925-ds-lab-5.ipynb - Cola
colab.research.google.com/drive/1JVNy_guJ8hmyFuVXzyBLKtB2Xr4LuRke#scrollTo=GpgQ4LJ9Ujhl
Apps
Google
KL MAIL.Outlook
Wiley Certification i...
Cisco kl
Download file | iLov...
Projects · Dashboar...
LinkedIn
google colabs
190031925-ds-lab-5.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Files
sample_data
kerala.csv
+ Code + Text
[6] df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 118 entries, 0 to 117
Data columns (total 16 columns):
Column Non-Null Count Dtype
0 SUBDIVISION 118 non-null object
1 YEAR 118 non-null int64
2 JAN 118 non-null float64
3 FEB 118 non-null float64
4 MAR 118 non-null float64
5 APR 118 non-null float64
6 MAY 118 non-null float64
7 JUN 118 non-null float64
8 JUL 118 non-null float64
9 AUG 118 non-null float64
10 SEP 118 non-null float64
11 OCT 118 non-null float64
12 NOV 118 non-null float64
13 DEC 118 non-null float64
14 ANNUAL RAINFALL 118 non-null float64
15 FLOODS 118 non-null object
dtypes: float64(13), int64(1), object(2)
memory usage: 14.9+ KB
[7] # Changing the target column to numeric values
df["FLOODS"] = df["FLOODS"].map({"YES": 1, "NO": 0})
[8] df["JUN_GT_500"] = (df["JUN"] > 500).astype("int")
df["JUL_GT_500"] = (df["JUL"] > 500).astype("int")
df_small = df.loc[:, ["YEAR", "JUN_GT_500", "JUL_GT_500", "FLOODS"]]
df_small["COUNT"] =1
df_small.head()
YEAR JUN_GT_500 JUL_GT_500 FLOODS COUNT
0 1901 1 1 1 1
1 1902 0 1 1 1
2 1903 1 1 1 1
goutam
Goutam
goutamrayavarapugoutam@gmail.com
Sync is on
Manage your Google Account
Close 2 windows
Other people
goutam
Goutam
satvik
Guest
Add
Disk 75.67 GB available
Type here to search
ENG
15:28
03-02-2021

190031925-ds-lab-5.ipynb - Cola

colab.research.google.com/drive/1JVNy_guJ8hmyFuVXzyBLKtB2Xr4LuRke#scrollTo=GpgQ4LJ9Ujhl

Apps | Google | KL MAIL.Outlook | Wiley Certification i... | Cisco kl | Download file | iLov... | Projects - Dashboar... | LinkedIn | google colabs

190031925-ds-lab-5.ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

Files

..
sample_data
kerala.csv

```
[8] df["JUN_GT_500"] = (df["JUN"] > 500).astype("int")
    df["JUL_GT_500"] = (df["JUL"] > 500).astype("int")
    df_small = df.loc[:, ["YEAR", "JUN_GT_500", "JUL_GT_500", "FLOODS"]]
    df_small["COUNT"] =1
    df_small.head()
```

	YEAR	JUN_GT_500	JUL_GT_500	FLOODS	COUNT
0	1901	1	1	1	1
1	1902	0	1	1	1
2	1903	1	1	1	1
3	1904	1	1	1	1
4	1905	1	1	0	1

```
[9] df_small.shape
```

```
(118, 5)
```

```
[10] # Creating the tabular data based on the counts
     pd.crosstab(df_small["FLOODS"], df_small["JUN_GT_500"])
```

JUN_GT_500	0	1
FLOODS		
0	19	39
1	6	54

```
[11] pd.crosstab(df_small["FLOODS"], df_small["JUN_GT_500"])
```

JUN_GT_500	0	1
FLOODS		
0	19	39

goutam

Goutam

goutamrayavarapugoutam@gmail.com

Sync is on

Manage your Google Account

Close 2 windows

Other people

goutam

Goutam

satvik

Guest

Add

Disk 75.67 GB available

Type here to search

ENG 15:28 03-02-2021

190031925-ds-lab-5.ipynb - Cola

colab.research.google.com/drive/1JVNy_guJ8hmyFuVXzyBLKtB2Xr4LuRke#scrollTo=GpgQ4LJ9Ujhl

Apps | Google | KL MAIL.Outlook | Wiley Certification i... | Cisco kl | Download file | iLov... | Projects - Dashboar... | LinkedIn | google colabs

190031925-ds-lab-5.ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

Files

..
sample_data
kerala.csv

Disk 75.67 GB available

```
[11] pd.crosstab(df_small["FLOODS"], df_small["JUN_GT_500"])
```

JUN_GT_500	0	1
FLOODS		
0	19	39
1	6	54

```
[12] P_F = (6 + 54) / (6 + 54 + 19 + 39)
     P_J = (39 + 54) / (6 + 54 + 19 + 39)
     P_F_intersect_J = 54 / (6 + 54 + 19 + 39)
     print(f"P(F): {P_F}")
     print(f"P(J): {P_J}")
     print(f"P(F AND J): {P_F_intersect_J}")
```

```
P(F): 0.5084745762711864
P(J): 0.788135593220339
P(F AND J): 0.4576271186440678
```

```
[13] # Now calculate probaillitity of flood given it rained more than 500 mm in June (P(A|B))
     P_F_J = P_F_intersect_J / P_J
     print(f"P(F|J): {P_F_J}")
```

```
P(F|J): 0.5806451612903226
```

```
[14] # Probability of rain more than 500 mm in June given it flooded that year (P(B|A))
     P_J_F = (P_F_J * P_J) / P_F
     print(f"P(J|F): {P_J_F}")
```

```
P(J|F): 0.9000000000000001
```

```
[15] # We can simllarly do it for july
     pd.crosstab(df_small["FLOODS"], df_small["JUL_GT_500"])
```

JUL_GT_500	0	1
FLOODS		
0	19	39

goutam

Goutam
goutamrayavarapugoutam@gmail.com

Sync is on
Manage your Google Account
Close 2 windows

Other people

goutam
Goutam
satvik
Guest
Add

Type here to search

ENG 15:28 03-02-2021

190031925-ds-lab-5.ipynb - Cola

colab.research.google.com/drive/1JVNy_guJ8hmyFuVXzyBLKtB2Xr4LuRke#scrollTo=GpgQ4LJ9Ujhl

Apps | Google | KL MAIL.Outlook | Wiley Certification i... | Cisco kl | Download file | iLov... | Projects - Dashboar... | LinkedIn | google colabs

190031925-ds-lab-5.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

..
sample_data
kerela.csv

+ Code + Text

```
[15] # We can similarly do it for july
     pd.crosstab(df_small["FLOODS"], df_small["JUL_GT_500"])
```

JUL_GT_500	0	1
FLOODS		
0	19	39
1	3	57

```
[16] P_F = (3 + 57) / (3 + 57 + 19 + 39)
     P_J = (39 + 57) / (3 + 57 + 19 + 39)
     P_F_intersect_J = 57 / (3 + 57 + 19 + 39)
     print(f"P(F): {P_F}")
     print(f"P(J): {P_J}")
     print(f"P(F AND J): {P_F_intersect_J}")
```

```
P(F): 0.5084745762711864
P(J): 0.8135593220338984
P(F AND J): 0.4830508474576271
```

```
[17] # Now calculate probailitity of flood given it rained more than 500 mm in July
     P_F_J = P_F_intersect_J / P_J
     print(f"P(F|J): {P_F_J}")
```

```
P(F|J): 0.59375
```

```
# Probability of rain more than 500 mm in July given it flooded that year (P(B|A))
P_J_F = (P_F_J * P_J) / P_F
print(f"P(J|F): {P_J_F}")
```

```
P(J|F): 0.9500000000000002
```

goutam

Goutam
goutamrayavarapugoutam@gmail.com

Sync is on
Manage your Google Account
Close 2 windows

Other people

goutam
Goutam
satvik
Guest
Add

Disk 75.67 GB available

Type here to search

ENG 15:28 03-02-2021