```
from google.colab import drive;
drive.mount('/content/drive','ACTIVITY.txt')
```

```
Mounted at /content/drive
```

```
import pandas as pd
import numpy as np
exam_data = {
'name': ['Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'MattheW','Laura', 'Kevin', 'Jonas'],
'score': [12.5, 9, 16.5, np.nan, 9, 20, 14.5, np.nan, 8, 19],
'attempts': [1, 3, 4, 3, 5, 3, 6, 1, 7, 1]
}
df = pd.DataFrame(exam_data)
df
select=df[['name','score']]
print(select)
```

```
         name  score
0  Anastasia   12.5
1       Dima    9.0
2  Katherine   16.5
3      James    NaN
4      Emily    9.0
5    Michael   20.0
6    MattheW   14.5
7      Laura    NaN
8      Kevin    8.0
9      Jonas   19.0
```

```
import numpy as np
import pandas as pd
df=pd.DataFrame(exam_data)
select=df[df['attempts']>3]
print(select)
```

```
         name  score  attempts
2  Katherine   16.5         4
4      Emily    9.0         5
6    MattheW   14.5         6
8      Kevin    8.0         7
```

```
import numpy as np
import pandas as pd
data = {
'name': ['Alice', 'Bob', 'Charlie', 'Dave'],
'age': [25, 35, 40, 28],
'gender': ['F', 'M', 'M', 'M'],
'salary': [50000, 70000, 60000, 80000]
}
df=pd.DataFrame(data)
df
```

	name	age	gender	salary
0	Alice	25	F	50000
1	Bob	35	M	70000
2	Charlie	40	M	60000
3	Dave	28	M	80000

Next steps: Generate code with df | View recommended plots | New interactive sheet

```
import numpy as np
import pandas as pd
df=pd.DataFrame(data)
select=df[df['age']>30]
print(select)
```

```
       name  age gender  salary
1       Bob   35      M   70000
2   Charlie   40      M   60000
```

```
import numpy as np
import pandas as pd
df=pd.DataFrame(data)
select=df[df['name'].str.contains('e')]
print(select)
```

```
      name  age gender  salary
0    Alice   25      F   50000
2  Charlie   40      M   60000
3     Dave   28      M   80000
```

```
import numpy as np
import pandas as pd
df=pd.DataFrame(data)
select=df[(df['gender']=='M') & (df['salary']>65000)]
print(select)
```

```
    name  age gender  salary
1   Bob   35      M   70000
3  Dave   28      M   80000
```

```
import numpy as np
import pandas as pd
df=pd.DataFrame(data)
select=df[['name','age']]
print(select)
```

```
      name  age
0    Alice   25
1      Bob   35
2  Charlie   40
3     Dave   28
```

```
from google.colab import drive;
drive.mount('/content/drive','TRAIN.csv')
```

```
Mounted at /content/drive
```

```
import pandas as pd
df=pd.read_csv('/content/drive/My Drive/TRAIN.csv')
df
```

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous
0	76	retired	married	secondary	no	2302.0	no	no	telephone	5	feb	110	1	87	2
1	66	retired	divorced	unknown	no	53.0	no	no	cellular	12	jul	562	4	-1	0
2	51	management	married	tertiary	no	2455.0	yes	no	cellular	21	jul	553	1	-1	0
3	41	blue-collar	married	secondary	no	356.0	yes	no	cellular	14	may	90	5	-1	0
4	51	technician	married	secondary	no	-1944.0	yes	no	cellular	7	may	623	1	-1	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4461	33	management	married	tertiary	no	133.0	yes	no	unknown	26	may	308	4	-1	0
4462	39	services	divorced	secondary	no	687.0	yes	no	cellular	9	jul	869	1	-1	0
4463	40	admin.	single	secondary	no	2040.0	yes	no	cellular	18	may	906	2	350	2
4464	31	technician	single	secondary	no	628.0	yes	no	unknown	12	may	1083	2	-1	0
4465	70	retired	divorced	primary	no	383.0	no	no	cellular	28	apr	50	2	-1	0

4466 rows × 17 columns

Next steps: Generate code with df | View recommended plots | New interactive sheet

```
import pandas as pd
df=pd.read_csv('/content/drive/My Drive/TRAIN.csv')
select=df[(df['education']=='primary') & (df['deposit']=='yes')]
print(select)
```

```
      age          job   marital education default  balance housing loan  \
29     39  blue-collar  divorced   primary      no   1317.0     yes   no
39     31   unemployed    single   primary      no    163.0      no   no
56     49  blue-collar    single   primary      no    566.0     yes   no
66     53  blue-collar   married   primary     yes   -462.0      no   no
103    42  blue-collar    single   primary      no   4930.0      no   no
...   ...          ...       ...       ...     ...      ...     ...  ...
4411   55    housemaid   married   primary      no      0.0     yes   no
4422   80      retired   married   primary      no   1468.0      no   no
4451   41  blue-collar   married   primary      no    143.0     yes  yes
4452   53  blue-collar   married   primary      no    421.0     yes   no
4458   32  blue-collar   married   primary      no   -454.0     yes  yes

       contact  day month  duration  campaign  pdays  previous poutcome  \
29    cellular   20   nov       543         1    170         4    other
```

```
39    cellular   30   jan       707         2      2         1    other
56    cellular   25   jul       979         2     -1         0  unknown
66    cellular   29   jan       470         1     -1         0  unknown
103    unknown   18   jun       973         1     -1         0  unknown
...        ...  ...   ...       ...       ...    ...       ...      ...
4411  cellular   17   jul      1303         2     -1         0  unknown
4422  cellular   13   jan       330         3     -1         0  unknown
4451   unknown    2   jun       659         2     -1         0  unknown
4452  cellular   20   nov       677         1     -1         0  unknown
4458  cellular   18   may       801         5    355         2  failure

     deposit
29       yes
39       yes
56       yes
66       yes
103      yes
...      ...
4411     yes
4422     yes
4451     yes
4452     yes
4458     yes

[243 rows x 17 columns]
```

```
import pandas as pd
df=pd.read_csv('/content/drive/My Drive/TRAIN.csv')
select=df[df['deposit']=='no']
print(select)
```

```
       age          job   marital  education default  balance housing loan  \
0       76      retired   married  secondary      no   2302.0      no   no
3       41  blue-collar   married  secondary      no    356.0     yes   no
6       59      retired   married  secondary      no    136.0      no   no
7       34  blue-collar   married    primary      no   5299.0     yes   no
9       44  blue-collar   married  secondary      no    879.0     yes   no
...    ...          ...       ...        ...     ...      ...     ...  ...
4457    43   management   married   tertiary      no   1336.0     yes  yes
4460    54      retired   married  secondary      no    522.0      no  yes
4461    33   management   married   tertiary      no    133.0     yes   no
4464    31   technician    single  secondary      no    628.0     yes   no
4465    70      retired  divorced    primary      no    383.0      no   no

         contact  day month  duration  campaign  pdays  previous poutcome  \
0      telephone    5   feb       110         1     87         2  failure
3       cellular   14   may        90         5     -1         0  unknown
6       cellular    6   aug       301         4     -1         0  unknown
7        unknown   26   jun        75         5     -1         0  unknown
9       cellular    3   apr       383         1     -1         0  unknown
...          ...  ...   ...       ...       ...    ...       ...      ...
4457    cellular   27   may        82         2    309         1  failure
4460    cellular   14   jul        81         3     -1         0  unknown
4461     unknown   26   may       308         4     -1         0  unknown
4464     unknown   12   may      1083         2     -1         0  unknown
4465    cellular   28   apr        50         2     -1         0  unknown

     deposit
0         no
3         no
6         no
7         no
9         no
...      ...
4457      no
4460      no
4461      no
4464      no
4465      no

[2354 rows x 17 columns]
```

```
import pandas as pd
df=pd.read_csv('/content/drive/My Drive/TRAIN.csv')
select=df[(df['deposit'] == 'yes') & ((df['housing'] == 'yes') | (df['loan'] == 'yes'))]
print(select)
```

```
       age          job   marital  education default  balance housing loan  \
2       51   management   married   tertiary      no   2455.0     yes   no
4       51   technician   married  secondary      no  -1944.0     yes   no
15      37   management    single   tertiary      no    455.0     yes   no
17      24        admin.   single   tertiary      no      0.0     yes   no
21      33        admin.  married   tertiary      no     79.0     yes   no
...    ...          ...       ...        ...     ...      ...     ...  ...
4454    30  blue-collar    single  secondary      no    155.0     yes  yes
4458    32  blue-collar   married    primary      no   -454.0     yes  yes
4459    37   technician    single  secondary      no   3326.0     yes   no
```

```
4462   39     services  divorced  secondary      no    687.0     yes   no
4463   40       admin.    single  secondary      no   2040.0     yes   no

       contact  day month  duration  campaign  pdays  previous poutcome  \
2     cellular   21   jul       553         1     -1         0  unknown
4     cellular    7   may       623         1     -1         0  unknown
15    cellular   13   aug       904         6     -1         0  unknown
17    cellular   27   may       122         2     -1         0  unknown
21    cellular    5   may       389         1    195         4  success
...        ...  ...   ...       ...       ...    ...       ...      ...
4454  cellular    9   jul      1426         3     -1         0  unknown
4458  cellular   18   may       801         5    355         2  failure
4459   unknown   21   may       799         1     -1         0  unknown
4462  cellular    9   jul       869         1     -1         0  unknown
4463  cellular   18   may       906         2    350         2  failure

     deposit
2        yes
4        yes
15       yes
17       yes
21       yes
...      ...
4454     yes
4458     yes
4459     yes
4462     yes
4463     yes

[893 rows x 17 columns]
```

```
import pandas as pd
df=pd.read_csv('/content/drive/My Drive/TRAIN.csv')
select=df[(df['education']=='secondary') & (df['deposit']=='no')]
print(select)
```

```
       age          job  marital  education default  balance housing loan  \
0      76      retired  married  secondary      no   2302.0      no   no
3      41  blue-collar  married  secondary      no    356.0     yes   no
6      59      retired  married  secondary      no    136.0      no   no
9      44  blue-collar  married  secondary      no    879.0     yes   no
10     34     services  married  secondary      no   1637.0     yes   no
...   ...          ...      ...        ...     ...      ...     ...  ...
4446   35     services  married  secondary      no      0.0     yes   no
4453   31     services  married  secondary      no    505.0      no   no
4456   54  blue-collar  married  secondary      no   -102.0     yes   no
4460   54      retired  married  secondary      no    522.0      no  yes
4464   31   technician   single  secondary      no    628.0     yes   no

        contact  day month  duration  campaign  pdays  previous poutcome  \
0     telephone    5   feb       110         1     87         2  failure
3      cellular   14   may        90         5     -1         0  unknown
6      cellular    6   aug       301         4     -1         0  unknown
9      cellular    3   apr       383         1     -1         0  unknown
10     cellular   21   nov       107         4     -1         0  unknown
...         ...  ...   ...       ...       ...    ...       ...      ...
4446   cellular   24   jul       810         1     -1         0  unknown
4453   cellular   11   jul       773         3     -1         0  unknown
4456   cellular   27   aug       164         7     -1         0  unknown
4460   cellular   14   jul        81         3     -1         0  unknown
4464    unknown   12   may      1083         2     -1         0  unknown

     deposit
0         no
3         no
6         no
9         no
10        no
...      ...
4446      no
4453      no
4456      no
4460      no
4464      no

[1229 rows x 17 columns]
```

```
import pandas as pd
df=pd.read_csv('/content/drive/My Drive/TRAIN.csv')
select=df[(df['deposit']=='yes') & (df['poutcome']=='success')]
print(select)
```

```
       age            job   marital  education default  balance housing loan  \
19     76  self-employed   married    unknown      no   4984.0      no   no
21     33         admin.   married   tertiary      no     79.0     yes   no
45     71        retired  divorced  secondary      no      0.0      no   no
51     68        retired   married  secondary      no   1146.0      no   no
```

```
52     46     management   married   tertiary      no    273.0     yes   no
...   ...           ...       ...        ...     ...      ...     ...  ...
4338   38         admin.  divorced  secondary      no     19.0     yes   no
4372   20        student    single  secondary      no    215.0      no   no
4376   42     technician   married  secondary      no    994.0     yes   no
4408   29      housemaid    single   tertiary      no     19.0      no   no
4448   27    blue-collar    single  secondary      no    535.0      no   no

         contact  day month  duration  campaign  pdays  previous poutcome  \
19     telephone   28   apr       403         1    182         1  success
21      cellular    5   may       389         1    195         4  success
45      cellular   26   feb       771         1    171         1  success
51      cellular   13   may       356         1     71         5  success
52      cellular   18   mar       910         2    184         4  success
...          ...  ...   ...       ...       ...    ...       ...      ...
4338    cellular    5   feb      1130         3    251         2  success
4372    cellular   24   feb       175         1     92         6  success
4376    cellular   12   nov       227         3     93         6  success
4408    cellular    4   may       268         1     88         4  success
4448    cellular   16   aug       265         3     95         4  success

     deposit
19       yes
21       yes
45       yes
51       yes
52       yes
...      ...
4338     yes
4372     yes
4376     yes
4408     yes
4448     yes

[392 rows x 17 columns]
```

```
import pandas as pd
df=pd.read_csv('/content/drive/My Drive/TRAIN.csv')
select=df[(df['job']=='unemployed') & (df['deposit']=='no')]
print(select)
```

```
4066      no
4102      no
4125      no
4130      no
4140      no
4179      no
4181      no
4271      no
4289      no
4395      no
```

```
import pandas as pd
df=pd.read_csv('/content/drive/My Drive/TRAIN.csv')
select = df.loc[df['age'] <= 30, ['education', 'balance']]
print(select)
```

```
       education  balance
17      tertiary      0.0
22       primary    544.0
26     secondary     30.0
27     secondary    195.0
40     secondary    743.0
...          ...      ...
4440    tertiary    674.0
4448   secondary    535.0
4449   secondary     81.0
4454   secondary    155.0
4455    tertiary    265.0

[809 rows x 2 columns]
```