```
import pandas as pd
```

```
bankfull = pd.read_excel("/content/drive/MyDrive/Colab Notebooks/bank-full2.xlsx")
```

```
bankfull.head(n=30)
```

0	58	management	married	tertiary	no	2143	yes	no	unknown	5	may	261	1	-1
1	44	technician	single	secondary	no	29	yes	no	unknown	5	may	151	1	-1
2	33	entrepreneur	married	secondary	no	2	yes	yes	unknown	5	may	76	1	-1
3	47	blue-collar	married	unknown	no	1506	yes	no	unknown	5	may	92	1	-1
4	33	unknown	single	unknown	no	1	no	no	unknown	5	may	198	1	-1
5	35	management	married	tertiary	no	231	yes	no	unknown	5	may	139	1	-1
6	28	management	single	tertiary	no	447	yes	yes	unknown	5	may	217	1	-1
7	42	entrepreneur	divorced	tertiary	yes	2	yes	no	unknown	5	may	380	1	-1
8	58	retired	married	primary	no	121	yes	no	unknown	5	may	50	1	-1
9	43	technician	single	secondary	no	593	yes	no	unknown	5	may	55	1	-1
10	41	admin.	divorced	secondary	no	270	yes	no	unknown	5	may	222	1	-1
11	29	admin.	single	secondary	no	390	yes	no	unknown	5	may	137	1	-1
12	53	technician	married	secondary	no	6	yes	no	unknown	5	may	517	1	-1
13	58	technician	married	unknown	no	71	yes	no	unknown	5	may	71	1	-1
14	57	services	married	secondary	no	162	yes	no	unknown	5	may	174	1	-1
15	51	retired	married	primary	no	229	yes	no	unknown	5	may	353	1	-1

```
print(bankfull.shape)
```

```
(45211, 17)
```

18	60	retired	married	primary	no	60	yes	no	unknown	5	may	219	1	-1

```
bankfull.isnull().any().any()
```

```
False
```

21	56	management	married	tertiary	no	779	yes	no	unknown	5	may	164	1	-1

```
video = bankfull.dropna(axis=0)
```

```
print(bankfull.shape)
```

```
(45211, 17)
```

```
banksimple = pd.read_excel("/content/drive/MyDrive/Colab Notebooks/banksimple.xlsx")
```

```
banksimple.head(n=30)
```

	age	job	marital	education	default	balance	housing	loan	contact	day	month	durati
0	30	unemployed	married	primary	no	1787	no	no	cellular	19	oct	
1	33	services	married	secondary	no	4789	yes	yes	cellular	11	may	2
2	35	management	single	tertiary	no	1350	yes	no	cellular	16	apr	1
3	30	management	married	tertiary	no	1476	yes	yes	unknown	3	jun	1
4	59	blue-collar	married	secondary	no	0	yes	no	unknown	5	may	2
5	35	management	single	tertiary	no	747	no	no	cellular	23	feb	1
6	36	self-employed	married	tertiary	no	307	yes	no	cellular	14	may	3
7	39	technician	married	secondary	no	147	yes	no	cellular	6	may	1
8	41	entrepreneur	married	tertiary	no	221	yes	no	unknown	14	may	
9	43	services	married	primary	no	-88	yes	yes	cellular	17	apr	3
10	39	services	married	secondary	no	9374	yes	no	unknown	20	may	2
11	43	admin.	married	secondary	no	264	yes	no	cellular	17	apr	1
12	36	technician	married	tertiary	no	1109	no	no	cellular	13	aug	3
13	20	student	single	secondary	no	502	no	no	cellular	30	apr	2
14	31	blue-collar	married	secondary	no	360	yes	yes	cellular	29	jan	
15	40	management	married	tertiary	no	194	no	yes	cellular	29	aug	1

```
print(banksimple.shape)
```

```
(4521, 17)
```

```
bankfull.isnull().any().any()
```

False

```
video = banksimple.dropna(axis=0)
print(banksimple.shape)
```

```
(4521, 17)
```

```
filter1 = bankfull['job'] != 'unemployed'
```

```
print(bankfull[filter1].shape)
```

```
(43908, 17)
```

29 53 admin married secondary no 105 no yes cellular 21 aug

```
bankfull[filter1].head(n=30)
```

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duratio
0	58	management	married	tertiary	no	2143	yes	no	unknown	5	may	26
1	44	technician	single	secondary	no	29	yes	no	unknown	5	may	15
2	33	entrepreneur	married	secondary	no	2	yes	yes	unknown	5	may	7
3	47	blue-collar	married	unknown	no	1506	yes	no	unknown	5	may	9
4	33	unknown	single	unknown	no	1	no	no	unknown	5	may	19
5	35	management	married	tertiary	no	231	yes	no	unknown	5	may	13
6	28	management	single	tertiary	no	447	yes	yes	unknown	5	may	21
7	42	entrepreneur	divorced	tertiary	yes	2	yes	no	unknown	5	may	38
8	58	retired	married	primary	no	121	yes	no	unknown	5	may	5
9	43	technician	single	secondary	no	593	yes	no	unknown	5	may	5
10	41	admin.	divorced	secondary	no	270	yes	no	unknown	5	may	22
11	29	admin.	single	secondary	no	390	yes	no	unknown	5	may	13
12	53	technician	married	secondary	no	6	yes	no	unknown	5	may	51
13	58	technician	married	unknown	no	71	yes	no	unknown	5	may	7
14	57	services	married	secondary	no	162	yes	no	unknown	5	may	17
15	51	retired	married	primary	no	229	yes	no	unknown	5	may	35
16	45	admin.	single	unknown	no	13	yes	no	unknown	5	may	9
17	57	blue-collar	married	primary	no	52	yes	no	unknown	5	may	3
18	60	retired	married	primary	no	60	yes	no	unknown	5	may	21
19	33	services	married	secondary	no	0	yes	no	unknown	5	may	5

```
import numpy as np
```

```
bankfull['abletoapply'] = np.where (
    bankfull['poutcome'] != 'failure',
    'yes',
    'no'
)
```

```
bankfull.head(n=26792)
```

marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	pre
married	tertiary	no	2143	yes	no	unknown	5	may	261	1	-1	
single	secondary	no	29	yes	no	unknown	5	may	151	1	-1	
married	secondary	no	2	yes	yes	unknown	5	may	76	1	-1	
married	unknown	no	1506	yes	no	unknown	5	may	92	1	-1	
single	unknown	no	1	no	no	unknown	5	may	198	1	-1	
...	...	...	...	...	...	...	...	...	...	...	...	
divorced	tertiary	yes	10	no	no	cellular	20	nov	75	3	-1	
married	secondary	no	655	no	no	cellular	20	nov	986	6	-1	
married	primary	no	1379	no	yes	cellular	20	nov	172	2	-1	
married	secondary	no	142	yes	no	cellular	20	nov	315	2	196	
married	secondary	no	392	yes	no	cellular	20	nov	117	3	-1	

```
bankfull[filter1].head(n=26792)
```

	age	job	marital	education	default	balance	housing	loan	contact	day	month	dur
0	58	management	married	tertiary	no	2143	yes	no	unknown	5	may	
1	44	technician	single	secondary	no	29	yes	no	unknown	5	may	
2	33	entrepreneur	married	secondary	no	2	yes	yes	unknown	5	may	
3	47	blue-collar	married	unknown	no	1506	yes	no	unknown	5	may	
4	33	unknown	single	unknown	no	1	no	no	unknown	5	may	
...	...	...	...	...	...	...	...	...	...	...	...	
27418	40	management	married	tertiary	no	1584	yes	no	telephone	21	nov	
27419	52	management	married	tertiary	no	1065	yes	yes	cellular	21	nov	
27420	36	management	single	tertiary	no	7	no	no	unknown	21	nov	
27421	49	services	married	secondary	no	66	yes	yes	cellular	21	nov	
27422	36	management	single	tertiary	no	72	yes	yes	cellular	21	nov	

26792 rows × 18 columns

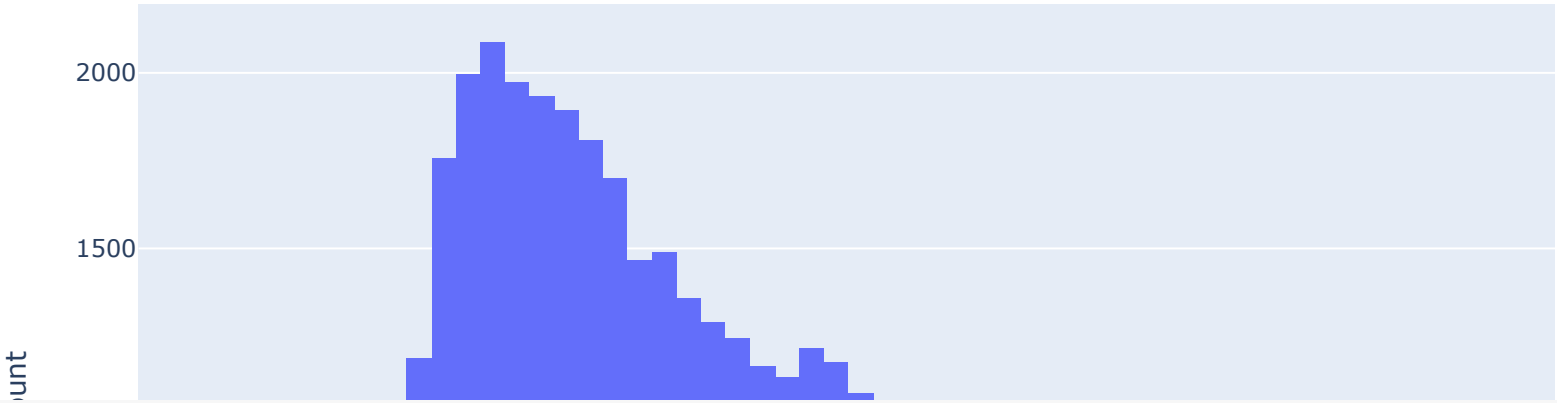
```
print(bankfull[filter1].shape)
```

```
(43908, 18)
```

```
bankfull.groupby(['education', 'age', 'job']).sum()
```

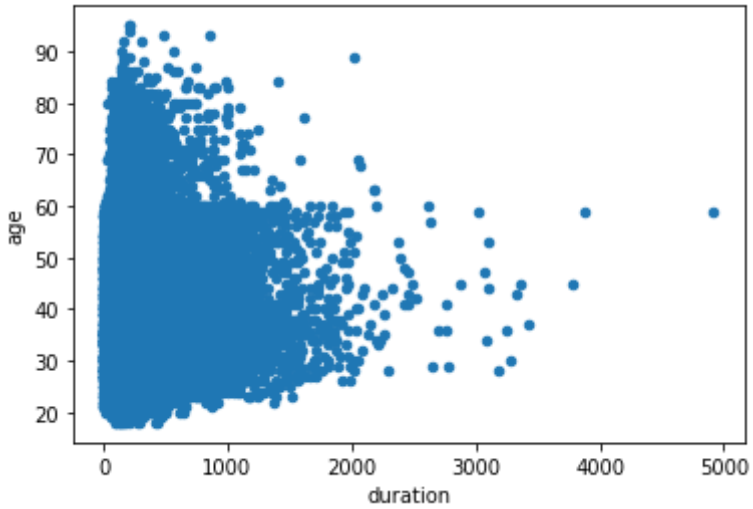
education	age	job	balance	day	duration	campaign	pdays	previous
primary	18	student	3160	35	599	5	91	1
	19	student	2599	131	1394	17	477	10
	20	blue-collar	-172	19	238	3	-1	0
		student	7950	55	690	13	548	15
	21	blue-collar	-640	29	553	4	359	3
...	...	...	...	...	...	...	...	...
unknown	83	retired	4216	12	92	1	-1	0
	85	retired	1934	12	140	1	-1	0
	86	retired	157	7	147	1	-1	0
	92	retired	1550	48	477	7	95	3

```
import plotly.express as px
px.histogram(bankfull, x = 'age')
```




```
import matplotlib.pyplot as plt
bankfull.plot.scatter('duration', 'age')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f8120155990>
```



Colab paid products - Cancel contracts here

✓ 0s completed at 9:18 PM