

# ds-as01-1

August 15, 2024

```
[1]: import pandas as pd

try:
    df = pd.read_csv('bank_train.csv')
    print(df.head())
except FileNotFoundError as e:
    print("File not found. Please check the file path.")
    print(e)
```

	age	job	marital	education	default	balance	housing	loan	\
0	76	retired	married	secondary	no	2302.0	no	no	
1	66	retired	divorced	unknown	no	53.0	no	no	
2	51	management	married	tertiary	no	2455.0	yes	no	
3	41	blue-collar	married	secondary	no	356.0	yes	no	
4	51	technician	married	secondary	no	-1944.0	yes	no	

  

	contact	day	month	duration	campaign	pdays	previous	poutcome	deposit
0	telephone	5	feb	110	1	87	2	failure	no
1	cellular	12	jul	562	4	-1	0	unknown	yes
2	cellular	21	jul	553	1	-1	0	unknown	yes
3	cellular	14	may	90	5	-1	0	unknown	no
4	cellular	7	may	623	1	-1	0	unknown	yes

```
[2]: import pandas as pd

df = pd.read_csv('bank_train.csv')
```

```
[3]: df.head()
```

```
[3]:
```

	age	job	marital	education	default	balance	housing	loan	\
0	76	retired	married	secondary	no	2302.0	no	no	
1	66	retired	divorced	unknown	no	53.0	no	no	
2	51	management	married	tertiary	no	2455.0	yes	no	
3	41	blue-collar	married	secondary	no	356.0	yes	no	
4	51	technician	married	secondary	no	-1944.0	yes	no	

  

	contact	day	month	duration	campaign	pdays	previous	poutcome	deposit
0	telephone	5	feb	110	1	87	2	failure	no

1	cellular	12	jul	562	4	-1	0	unknown	yes
2	cellular	21	jul	553	1	-1	0	unknown	yes
3	cellular	14	may	90	5	-1	0	unknown	no
4	cellular	7	may	623	1	-1	0	unknown	yes

```
[6]: filtered_df = df[(df['education'] == 'primary') & (df['deposit'] == 'yes')]
```

```
[7]: filtered_df.head()
```

```
[7]:
```

	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	

  

	contact	day	month	duration	campaign	pdays	previous	poutcome	deposit
29	cellular	20	nov	543	1	170	4	other	yes
39	cellular	30	jan	707	2	2	1	other	yes
56	cellular	25	jul	979	2	-1	0	unknown	yes
66	cellular	29	jan	470	1	-1	0	unknown	yes
103	unknown	18	jun	973	1	-1	0	unknown	yes

```
[8]: not_subscribed_df = df[df['deposit'] != 'yes']
```

```
[9]: not_subscribed_df.head()
```

```
[9]:
```

	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	

  

	contact	day	month	duration	campaign	pdays	previous	poutcome	deposit
0	telephone	5	feb	110	1	87	2	failure	no
3	cellular	14	may	90	5	-1	0	unknown	no
6	cellular	6	aug	301	4	-1	0	unknown	no
7	unknown	26	jun	75	5	-1	0	unknown	no
9	cellular	3	apr	383	1	-1	0	unknown	no

```
[10]: filtered_df = df[
    (df['deposit'] == 'yes') &
    ((df['housing'] == 'yes') | (df['loan'] == 'yes'))
]
```

```
[11]: filtered_df.head()
```

```
[11]:
```

	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	

  

	contact	day	month	duration	campaign	pdays	previous	poutcome	deposit
2	cellular	21	jul	553	1	-1	0	unknown	yes
4	cellular	7	may	623	1	-1	0	unknown	yes
15	cellular	13	aug	904	6	-1	0	unknown	yes
17	cellular	27	may	122	2	-1	0	unknown	yes
21	cellular	5	may	389	1	195	4	success	yes

```
[12]: filtered_df = df[
        (df['education'] == 'secondary') &
        (df['deposit'] == 'no')
    ]
```

```
[13]: filtered_df.head()
```

```
[13]:
```

	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	

  

	contact	day	month	duration	campaign	pdays	previous	poutcome	deposit
0	telephone	5	feb	110	1	87	2	failure	no
3	cellular	14	may	90	5	-1	0	unknown	no
6	cellular	6	aug	301	4	-1	0	unknown	no
9	cellular	3	apr	383	1	-1	0	unknown	no
10	cellular	21	nov	107	4	-1	0	unknown	no

```
[16]: filtered_df = df[
        (df['deposit'] == 'yes') &
        (df['poutcome'] == 'success')
    ]
```

```
[17]: filtered_df.head()
```

```
[17]:
```

	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
----	----	------------	---------	----------	----	-------	-----	----

  

	contact	day	month	duration	campaign	pdays	previous	poutcome	deposit
19	telephone	28	apr	403	1	182	1	success	yes
21	cellular	5	may	389	1	195	4	success	yes
45	cellular	26	feb	771	1	171	1	success	yes
51	cellular	13	may	356	1	71	5	success	yes
52	cellular	18	mar	910	2	184	4	success	yes

```
[18]: filtered_df = df[
      (df['job'] == 'unemployed') &
      (df['deposit'] == 'no')
    ]
```

```
[19]: filtered_df.head()
```

```
[19]:
```

	age	job	marital	education	default	balance	housing	loan	\
74	37	unemployed	single	secondary	no	48.0	no	no	
304	48	unemployed	married	secondary	no	855.0	yes	no	
404	57	unemployed	married	primary	no	0.0	yes	no	
464	47	unemployed	divorced	secondary	no	947.0	no	no	
494	45	unemployed	married	tertiary	no	1148.0	no	no	

  

	contact	day	month	duration	campaign	pdays	previous	poutcome	deposit
74	cellular	31	jul	40	10	-1	0	unknown	no
304	unknown	9	may	241	2	-1	0	unknown	no
404	cellular	20	nov	231	2	-1	0	unknown	no
464	cellular	5	feb	124	2	-1	0	unknown	no
494	cellular	19	nov	782	2	-1	0	unknown	no

```
[20]: filtered_df = df[df['age'] <= 30][['education', 'balance']]
```

```
[21]: filtered_df.head()
```

```
[21]:
```

	education	balance
17	tertiary	0.0
22	primary	544.0
26	secondary	30.0
27	secondary	195.0
40	secondary	743.0

```
[22]: import pandas as pd
```

```
[23]: import numpy as np
```

```
[24]: 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]
}

```

```
[25]: df = pd.DataFrame(exam_data)
```

```
[26]: selected_columns = df[['name', 'score']]
```

```
[27]: data = {
    'name': ['Alice', 'Bob', 'Charlie', 'Dave'],
    'age': [25, 35, 40, 28],
    'gender': ['F', 'M', 'M', 'M'],
    'salary': [50000, 70000, 60000, 80000]
}

```

```
[28]: df = pd.DataFrame(data)
```

```
[29]: age_gt_30 = df[df['age'] > 30]
```

```
[30]: name_contains_e = df[df['name'].str.contains('e')]
```

```
[31]: gender_m_salary_gt_65000 = df[(df['gender'] == 'M') & (df['salary'] > 65000)]
```

```
[32]: name_age_columns = df[['name', 'age']]
```

```
[33]: print("a. Rows where age is greater than 30:")
print(age_gt_30)
```

a. Rows where age is greater than 30:

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

```
[34]: print("\nb. Rows where name contains 'e':")
print(name_contains_e)
```

b. Rows where name contains 'e':

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

```
[35]: print("\nc. Rows where gender is 'M' and salary is greater than 65000:")
print(gender_m_salary_gt_65000)
```

c. Rows where gender is 'M' and salary is greater than 65000:

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

```
[36]: print("\nd. Selected columns 'name' and 'age':")  
      print(name_age_columns)
```

d. Selected columns 'name' and 'age':

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

```
[ ]:
```