

```
import pandas as pd
```

```
df=pd.read_csv("/bank.csv")  
df
```

```
⇒
```

	age	job	marital	education	default	balance	housing	loan	contact
0	59	admin.	Married	secondary	no	2343	yes	no	unknown
1	56	admin.	Married	secondary	no	45	no	no	unknown
2	41	technician	Married	secondary	no	1270	yes	no	unknown
3	55	services	Married	secondary	no	2476	yes	no	unknown
4	54	admin.	Married	tertiary	no	184	no	no	unknown
...	...	...	...	...	...	...	...	...	...
11157	33	blue-collar	single	primary	no	1	yes	no	cellular
11158	39	services	Married	secondary	no	733	no	no	unknown
11159	32	technician	single	secondary	no	29	no	no	cellular
11160	43	technician	Married	secondary	no	0	no	yes	cellular
11161	34	technician	Married	secondary	no	0	no	no	cellular

11162 rows × 10 columns

```
df.columns
```

```
⇒ Index(['age', 'job', 'marital', 'education', 'default', 'balance', 'housing',  
        'loan', 'contact', 'day', 'month', 'duration', 'campaign', 'pdays',  
        'previous', 'outcome', 'deposit'],  
        dtype='object')
```

```
df.index
```

```
⇒ RangeIndex(start=0, stop=11162, step=1)
```

```
df.size
```

```
⇒ 189754
```

```
df.shape
```

```
⇒ (11162, 10)
```

```
df.memory_usage()
```



0

---

Index	132
age	89296
job	89296
Marital	89296
education	89296
default	89296
balance	89296
housing	89296
loan	89296
contact	89296
day	89296
Month	89296
duration	89296
campaign	89296
pdays	89296
previous	89296
outcome	89296
deposit	89296

dtype: int64

df.ndim



2

df.head(10)



	age	job	marital	education	default	balance	housing	loan	contact
0	59	admin.	married	secondary	no	2343	yes	no	unknown
1	56	admin.	married	secondary	no	45	no	no	unknown
2	41	technician	married	secondary	no	1270	yes	no	unknown
3	55	services	married	secondary	no	2476	yes	no	unknown
4	54	admin.	married	tertiary	no	184	no	no	unknown
5	42	management	single	tertiary	no	0	yes	yes	unknown
6	56	management	married	tertiary	no	830	yes	yes	unknown
7	60	retired	divorced	secondary	no	545	yes	no	unknown
8	37	technician	married	secondary	no	1	yes	no	unknown
9	28	services	single	secondary	no	5090	yes	no	unknown

Next steps:

[Generate code with df](#)[View recommended plots](#)[New interactive sheet](#)

```
df.tail(8)
```



```
7 contact
5 cellul
5 cellul
0 cellul
0 cellul
0 unknow
0 cellul
5 cellul
0 cellul
```

```
df.at[1435,"marital"]
```



```
'married'
```

```
df.iat[1435,5]
```



```
np.int64(1293)
```

```
df.loc[1428:1433,"age"]
```



	age
1428	26
1429	33
1430	21
1431	30
1432	27
1433	30

*dtype: int64*

```
df.dtypes
```



	0
age	int64
job	object
Marital	object
education	object
default	object
balance	int64
housing	object
loan	object
contact	object
day	int64
Month	object
duration	int64
campaign	int64
pdays	int64
previous	int64
poutcome	object
deposit	object

*dtype: object*



```
df.describe()
```



	age	balance	day	duration	campaign	pda
count	11162.000000	11162.000000	11162.000000	11162.000000	11162.000000	11162.0000
mean	41.231948	1528.538524	15.658036	371.993818	2.508421	51.3304
std	11.913369	3225.413326	8.420740	347.128386	2.722077	108.7582
min	18.000000	-6847.000000	1.000000	2.000000	1.000000	-1.0000
25%	32.000000	122.000000	8.000000	138.000000	1.000000	-1.0000
50%	39.000000	550.000000	15.000000	255.000000	2.000000	-1.0000
75%	49.000000	1708.000000	22.000000	496.000000	3.000000	20.7500
max	95.000000	81204.000000	31.000000	3881.000000	63.000000	854.0000

```
df.select_dtypes(exclude=[object])
```



	age	balance	day	duration	campaign	pdays	previous	
0	59	2343	5	1042	1	-1	0	
1	56	45	5	1467	1	-1	0	
2	41	1270	5	1389	1	-1	0	
3	55	2476	5	579	1	-1	0	
4	54	184	5	673	2	-1	0	
...	...	...	...	...	...	...	...	
11157	33	1	20	257	1	-1	0	
11158	39	733	16	83	4	-1	0	
11159	32	29	19	156	2	-1	0	
11160	43	0	8	9	2	172	5	
11161	34	0	9	628	1	-1	0	
11162 rows x 7 columns								

```
df.select_dtypes(include=[int,float])
```



	age	balance	day	duration	campaign	pdays	previous
0	59	2343	5	1042	1	-1	0
1	56	45	5	1467	1	-1	0
2	41	1270	5	1389	1	-1	0
3	55	2476	5	579	1	-1	0
4	54	184	5	673	2	-1	0
...	...	...	...	...	...	...	...
11157	33	1	20	257	1	-1	0
11158	39	733	16	83	4	-1	0
11159	32	29	19	156	2	-1	0
11160	43	0	8	9	2	172	5
11161	34	0	9	628	1	-1	0



11162 rows x 7 columns

df.info()



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11162 entries, 0 to 11161
Data columns (total 17 columns):
#   Column      Non-Null Count  Dtype
---  -
0   age         11162 non-null  int64
1   job         11162 non-null  object
2   marital     11162 non-null  object
3   education   11162 non-null  object
4   default     11162 non-null  object
5   balance     11162 non-null  int64
6   housing     11162 non-null  object
7   loan        11162 non-null  object
8   contact     11162 non-null  object
9   day         11162 non-null  int64
10  month       11162 non-null  object
11  duration    11162 non-null  int64
12  campaign    11162 non-null  int64
13  pdays       11162 non-null  int64
14  previous    11162 non-null  int64
15  poutcome    11162 non-null  object
16  deposit     11162 non-null  object
dtypes: int64(7), object(10)
memory usage: 1.4+ MB
```

```
import numpy as np
```

```
np.unique(df['education'])
```

```
np.unique(df['job'])
```

```

⇒ array(['admin.', 'blue-collar', 'entrepreneur', 'housemaid', 'management',
        'retired', 'self-employed', 'services', 'student', 'technician',
        'unemployed', 'unknown'], dtype=object)

```

```
df.isnull().sum()
```

```
age_mean=df["age"].mean()
age_mean
```

```
df["age"].fillna(age_mean,inplace=True)
```

```
df
```

⇒ 'tmp/ipython-input-1031436070.py:6: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series consisting of zero or more existing
 rows. The behavior will change in pandas 3.0. This inplace method will never work because it does not
 modify the original DataFrame or Series. For example, when doing 'df[col].method(value, inplace=True)', try using 'df.loc[:,col].method(value, inplace=True)' instead.

```
df["age"].fillna(age_mean,inplace=True)
```

	age	job	marital	education	default	balance	housing	loan	contact
0	59	admin.	Married	secondary	no	2343	yes	no	unknown
1	56	admin.	Married	secondary	no	45	no	no	unknown
2	41	technician	Married	secondary	no	1270	yes	no	unknown
3	55	services	Married	secondary	no	2476	yes	no	unknown
4	54	admin.	Married	tertiary	no	184	no	no	unknown
...	...	...	...	...	...	...	...	...	...
11157	33	blue-collar	single	primary	no	1	yes	no	cellular
11158	39	services	Married	secondary	no	733	no	no	unknown
11159	32	technician	single	secondary	no	29	no	no	cellular
11160	43	technician	Married	secondary	no	0	no	yes	cellular
11161	34	technician	Married	secondary	no	0	no	no	cellular

1162 rows × 10 columns

Next steps:

[Generate code with df](#)
[View recommended plots](#)
[New interactive sheet](#)

```
df.isnull().sum()
```

```
# Calculate the mode of the 'job' column
job_mode = df['job'].mode()[0]
```

```
# Fill missing values in the 'job' column with the mode
df['job'].fillna(job_mode, inplace=True)
```

```
df.isnull().sum()
```



	0
<i>age</i>	0
<i>job</i>	0
<i>marital</i>	0
<i>education</i>	0
<i>default</i>	0
<i>balance</i>	0
<i>housing</i>	0
<i>loan</i>	0
<i>contact</i>	0
<i>day</i>	0
<i>month</i>	0
<i>duration</i>	0
<i>campaign</i>	0
<i>pdays</i>	0
<i>previous</i>	0
<i>outcome</i>	0
<i>deposit</i>	0

*dtype: int64*

```
df.isnull().sum()
```

```
df['duration'].value_counts()
```

```
max_counts=df['duration'].value_counts().index[0]  
max_counts
```

```
df['duration'].fillna(max_counts,inplace=True)
```



/tmp/ipython-input-1559020163.py:8: FutureWarning: A value is trying to be set