Lab-01

January 11, 2026

[1]:
```
import pandas as pd

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

[1]:
```
  age;"job";"marital";"education";"default";"balance";"housing";"loan";"contact"
;"day";"month";"duration";"campaign";"pdays";"previous";"poutcome";"y"
0  30;"unemployed";"married";"primary";"no";1787;…
1  33;"services";"married";"secondary";"no";4789;…
2  35;"management";"single";"tertiary";"no";1350;…
3  30;"management";"married";"tertiary";"no";1476…
4  59;"blue-collar";"married";"secondary";"no";0;…
```

[2]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
---------------------------------------------------------------------------
ModuleNotFoundError                       Traceback (most recent call last)
Cell In[2], line 4
      2 import pandas as pd
      3 import matplotlib.pyplot as plt
----> 4 import seaborn as sns

ModuleNotFoundError: No module named 'seaborn'
```

[3]:
```
df = pd.read_csv("bank.csv", sep = ';')
df.head()
```

[3]:

	age	job	marital	education	default	balance	housing	loan
0	30	unemployed	married	primary	no	1787	no	no
1	33	services	married	secondary	no	4789	yes	yes
2	35	management	single	tertiary	no	1350	yes	no
3	30	management	married	tertiary	no	1476	yes	yes
4	59	blue-collar	married	secondary	no	0	yes	no

```
    contact  day month  duration  campaign  pdays  previous poutcome   y
0  cellular   19   oct        79         1     -1         0  unknown  no
1  cellular   11   may       220         1    339         4  failure  no
2  cellular   16   apr       185         1    330         1  failure  no
3   unknown    3   jun       199         4     -1         0  unknown  no
4   unknown    5   may       226         1     -1         0  unknown  no
```

[4]:
```
df.tail()
```

[4]:
```
      age            job  marital  education default  balance housing loan  \
4516   33       services  married  secondary      no     -333     yes   no
4517   57  self-employed  married   tertiary     yes    -3313     yes  yes
4518   57     technician  married  secondary      no      295      no   no
4519   28    blue-collar  married  secondary      no     1137      no   no
4520   44   entrepreneur   single   tertiary      no     1136     yes  yes

       contact  day month  duration  campaign  pdays  previous poutcome   y
4516  cellular   30   jul       329         5     -1         0  unknown  no
4517   unknown    9   may       153         1     -1         0  unknown  no
4518  cellular   19   aug       151        11     -1         0  unknown  no
4519  cellular    6   feb       129         4    211         3    other  no
4520  cellular    3   apr       345         2    249         7    other  no
```

[5]:
```
def replace_marital(val):
  if val == 'single':
    return 0
  else:
    return 1

df['marital'] = df['marital'].apply(replace_marital)
df.head()
```

[5]:
```
   age          job  marital  education default  balance housing loan  \
0   30   unemployed        1    primary      no     1787      no   no
1   33     services        1  secondary      no     4789     yes  yes
2   35   management        0   tertiary      no     1350     yes   no
3   30   management        1   tertiary      no     1476     yes  yes
4   59  blue-collar        1  secondary      no        0     yes   no

    contact  day month  duration  campaign  pdays  previous poutcome   y
0  cellular   19   oct        79         1     -1         0  unknown  no
1  cellular   11   may       220         1    339         4  failure  no
2  cellular   16   apr       185         1    330         1  failure  no
3   unknown    3   jun       199         4     -1         0  unknown  no
4   unknown    5   may       226         1     -1         0  unknown  no
```

[6]:
```
df["housing"] = df["housing"].map({"yes": 1, "no": 0}.get)
df.head()
```

[6]:

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
0	30	unemployed	1	primary	no	1787	0	no	cellular	19	oct	79	1	-1	0	unknown	no
1	33	services	1	secondary	no	4789	1	yes	cellular	11	may	220	1	339	4	failure	no
2	35	management	0	tertiary	no	1350	1	no	cellular	16	apr	185	1	330	1	failure	no
3	30	management	1	tertiary	no	1476	1	yes	unknown	3	jun	199	4	-1	0	unknown	no
4	59	blue-collar	1	secondary	no	0	1	no	unknown	5	may	226	1	-1	0	unknown	no

[7]:
```
df["loan"] = df["loan"].replace({"yes": 1, "no": 0})
df.head()
```

```
/tmp/ipykernel_18086/1568368702.py:1: FutureWarning: Downcasting behavior in
`replace` is deprecated and will be removed in a future version. To retain the
old behavior, explicitly call `result.infer_objects(copy=False)`. To opt-in to
the future behavior, set `pd.set_option('future.no_silent_downcasting', True)`
  df["loan"] = df["loan"].replace({"yes": 1, "no": 0})
```

[7]:

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
0	30	unemployed	1	primary	no	1787	0	0	cellular	19	oct	79	1	-1	0	unknown	no
1	33	services	1	secondary	no	4789	1	1	cellular	11	may	220	1	339	4	failure	no
2	35	management	0	tertiary	no	1350	1	0	cellular	16	apr	185	1	330	1	failure	no
3	30	management	1	tertiary	no	1476	1	1	unknown	3	jun	199	4	-1	0	unknown	no
4	59	blue-collar	1	secondary	no	0	1	0	unknown	5	may	226	1	-1	0	unknown	no

[22]:
```
df["loan"] = df["loan"].replace({
"yes": 1,
"no": 0})
df.head()
```

[22]:

	age	job	marital	education	default	balance	housing	loan	contact
0	30	0.0	1	1.0	no	1787	0	0	cellular
1	33	0.0	1	2.0	no	4789	1	1	cellular

```
2   35  1.0        0        3.0      no     1350        1     0  cellular
3   30  1.0        1        3.0      no     1476        1     1   unknown
4   59  0.0        1        2.0      no        0        1     0   unknown

   day  month  duration  campaign  pdays  previous poutcome   y
0   19    NaN        79         1     -1         0  unknown  no
1   11    NaN       220         1    339         4  failure  no
2   16    NaN       185         1    330         1  failure  no
3    3    NaN       199         4     -1         0  unknown  no
4    5    NaN       226         1     -1         0  unknown  no
```

[9]:
```
df["job"].unique()
```

[9]:
```
array(['unemployed', 'services', 'management', 'blue-collar',
       'self-employed', 'technician', 'entrepreneur', 'admin.', 'student',
       'housemaid', 'retired', 'unknown'], dtype=object)
```

[18]:
```
df["job"] = df["job"].replace({'unemployed': 0,
                    'services': 0,
                    'management': 1,
                    'blue-collar': 0,
                    'self-employed': 0,
                    'technician': 1,
                    'entrepreneur': 1,
                    'admin.': 0,
                    'student': 1,
                    'housemaid': 0,
                    'retired': 0,
                    'unknown': np.nan})
df.head()
```

[18]:
```
   age  job  marital  education default  balance  housing  loan   contact  \
0   30  0.0        1        1.0      no     1787        0     0  cellular
1   33  0.0        1        2.0      no     4789        1     1  cellular
2   35  1.0        0        3.0      no     1350        1     0  cellular
3   30  1.0        1        3.0      no     1476        1     1   unknown
4   59  0.0        1        2.0      no        0        1     0   unknown

   day  month  duration  campaign  pdays  previous poutcome   y
0   19     10        79         1     -1         0  unknown  no
1   11      5       220         1    339         4  failure  no
2   16      4       185         1    330         1  failure  no
3    3      6       199         4     -1         0  unknown  no
4    5      5       226         1     -1         0  unknown  no
```

[11]:
```
df["month"].unique()
```

[11]:
```
array(['oct', 'may', 'apr', 'jun', 'feb', 'aug', 'jan', 'jul', 'nov',
       'sep', 'mar', 'dec'], dtype=object)
```

[20]:
```
df.month = df.month.map({
    'oct': 10,
    'may': 5,
    'apr': 4,
    'jun': 6,
    'feb': 2,
    'aug': 8,
    'jan': 1,
    'jul': 7,
    'nov': 11,
    'sep': 9,
    'mar': 3,
    'dec': 12
})
df.head(10)
```

[20]:

	age	job	marital	education	default	balance	housing	loan	contact	\
0	30	0.0	1	1.0	no	1787	0	0	cellular	
1	33	0.0	1	2.0	no	4789	1	1	cellular	
2	35	1.0	0	3.0	no	1350	1	0	cellular	
3	30	1.0	1	3.0	no	1476	1	1	unknown	
4	59	0.0	1	2.0	no	0	1	0	unknown	
5	35	1.0	0	3.0	no	747	0	0	cellular	
6	36	0.0	1	3.0	no	307	1	0	cellular	
7	39	1.0	1	2.0	no	147	1	0	cellular	
8	41	1.0	1	3.0	no	221	1	0	unknown	
9	43	0.0	1	1.0	no	-88	1	1	cellular	

	day	month	duration	campaign	pdays	previous	poutcome	y
0	19	NaN	79	1	-1	0	unknown	no
1	11	NaN	220	1	339	4	failure	no
2	16	NaN	185	1	330	1	failure	no
3	3	NaN	199	4	-1	0	unknown	no
4	5	NaN	226	1	-1	0	unknown	no
5	23	NaN	141	2	176	3	failure	no
6	14	NaN	341	1	330	2	other	no
7	6	NaN	151	2	-1	0	unknown	no
8	14	NaN	57	2	-1	0	unknown	no
9	17	NaN	313	1	147	2	failure	no

[13]:
```
df["education"].unique()
```

[13]:
```
array(['primary', 'secondary', 'tertiary', 'unknown'], dtype=object)
```

[14]:
```
df.education = df.education.map({
    'primary': 1,
    'secondary': 2,
    'tertiary': 3,
    'unknown': np.nan
})
```

[19]:
```
df.head(10)
```

[19]:

	age	job	marital	education	default	balance	housing	loan	contact	\
0	30	0.0	1	1.0	no	1787	0	0	cellular	
1	33	0.0	1	2.0	no	4789	1	1	cellular	
2	35	1.0	0	3.0	no	1350	1	0	cellular	
3	30	1.0	1	3.0	no	1476	1	1	unknown	
4	59	0.0	1	2.0	no	0	1	0	unknown	
5	35	1.0	0	3.0	no	747	0	0	cellular	
6	36	0.0	1	3.0	no	307	1	0	cellular	
7	39	1.0	1	2.0	no	147	1	0	cellular	
8	41	1.0	1	3.0	no	221	1	0	unknown	
9	43	0.0	1	1.0	no	-88	1	1	cellular	

	day	month	duration	campaign	pdays	previous	poutcome	y
0	19	10	79	1	-1	0	unknown	no
1	11	5	220	1	339	4	failure	no
2	16	4	185	1	330	1	failure	no
3	3	6	199	4	-1	0	unknown	no
4	5	5	226	1	-1	0	unknown	no
5	23	2	141	2	176	3	failure	no
6	14	5	341	1	330	2	other	no
7	6	5	151	2	-1	0	unknown	no
8	14	5	57	2	-1	0	unknown	no
9	17	4	313	1	147	2	failure	no

[23]:
```
df["poutcome"].unique()
```

[23]:
```
array(['unknown', 'failure', 'other', 'success'], dtype=object)
```

[24]:
```
df.poutcome = df.poutcome.map({
    'unknown': np.nan,
    'failure': 1,
    'other': 2,
    'success': 3
})
df.head(10)
```

[24]:

	age	job	marital	education	default	balance	housing	loan	contact	\
0	30	0.0	1	1.0	no	1787	0	0	cellular	

```
1   33  0.0         1         2.0       no     4789         1      1  cellular
2   35  1.0         0         3.0       no     1350         1      0  cellular
3   30  1.0         1         3.0       no     1476         1      1   unknown
4   59  0.0         1         2.0       no        0         1      0   unknown
5   35  1.0         0         3.0       no      747         0      0  cellular
6   36  0.0         1         3.0       no      307         1      0  cellular
7   39  1.0         1         2.0       no      147         1      0  cellular
8   41  1.0         1         3.0       no      221         1      0   unknown
9   43  0.0         1         1.0       no      -88         1      1  cellular

    day  month  duration  campaign  pdays  previous  poutcome   y
0    19    NaN        79         1     -1         0       NaN  no
1    11    NaN       220         1    339         4       1.0  no
2    16    NaN       185         1    330         1       1.0  no
3     3    NaN       199         4     -1         0       NaN  no
4     5    NaN       226         1     -1         0       NaN  no
5    23    NaN       141         2    176         3       1.0  no
6    14    NaN       341         1    330         2       2.0  no
7     6    NaN       151         2     -1         0       NaN  no
8    14    NaN        57         2     -1         0       NaN  no
9    17    NaN       313         1    147         2       1.0  no
```

[25]:
```
df["balance"] = df["balance"].apply(lambda v: (v - df["balance"].min())/␣
↪(df["balance"].max() - df["balance"].min()))
```

[26]:
```
df.head(10)
```

[26]:
```
   age  job  marital  education default   balance  housing  loan   contact  \
0   30  0.0        1        1.0      no  0.068455        0     0  cellular
1   33  0.0        1        2.0      no  0.108750        1     1  cellular
2   35  1.0        0        3.0      no  0.062590        1     0  cellular
3   30  1.0        1        3.0      no  0.064281        1     1   unknown
4   59  0.0        1        2.0      no  0.044469        1     0   unknown
5   35  1.0        0        3.0      no  0.054496        0     0  cellular
6   36  0.0        1        3.0      no  0.048590        1     0  cellular
7   39  1.0        1        2.0      no  0.046442        1     0  cellular
8   41  1.0        1        3.0      no  0.047436        1     0   unknown
9   43  0.0        1        1.0      no  0.043288        1     1  cellular

    day  month  duration  campaign  pdays  previous  poutcome   y
0    19    NaN        79         1     -1         0       NaN  no
1    11    NaN       220         1    339         4       1.0  no
2    16    NaN       185         1    330         1       1.0  no
3     3    NaN       199         4     -1         0       NaN  no
4     5    NaN       226         1     -1         0       NaN  no
5    23    NaN       141         2    176         3       1.0  no
6    14    NaN       341         1    330         2       2.0  no
```

```
7    6    NaN       151         2     -1         0       NaN  no
8   14    NaN        57         2     -1         0       NaN  no
9   17    NaN       313         1    147         2       1.0  no
```

[27]:
```
df["pdays"] = df["pdays"].apply(lambda v: (v - df["pdays"].min())/ (df["pdays"].
↪max() - df["pdays"].min()))
```

[28]:
```
df.head(10)
```

[28]:
```
   age  job  marital  education default   balance  housing  loan   contact  \
0   30  0.0        1        1.0      no  0.068455        0     0  cellular
1   33  0.0        1        2.0      no  0.108750        1     1  cellular
2   35  1.0        0        3.0      no  0.062590        1     0  cellular
3   30  1.0        1        3.0      no  0.064281        1     1   unknown
4   59  0.0        1        2.0      no  0.044469        1     0   unknown
5   35  1.0        0        3.0      no  0.054496        0     0  cellular
6   36  0.0        1        3.0      no  0.048590        1     0  cellular
7   39  1.0        1        2.0      no  0.046442        1     0  cellular
8   41  1.0        1        3.0      no  0.047436        1     0   unknown
9   43  0.0        1        1.0      no  0.043288        1     1  cellular

   day  month  duration  campaign     pdays  previous  poutcome   y
0   19    NaN        79         1  0.000000         0       NaN  no
1   11    NaN       220         1  0.389908         4       1.0  no
2   16    NaN       185         1  0.379587         1       1.0  no
3    3    NaN       199         4  0.000000         0       NaN  no
4    5    NaN       226         1  0.000000         0       NaN  no
5   23    NaN       141         2  0.202982         3       1.0  no
6   14    NaN       341         1  0.379587         2       2.0  no
7    6    NaN       151         2  0.000000         0       NaN  no
8   14    NaN        57         2  0.000000         0       NaN  no
9   17    NaN       313         1  0.169725         2       1.0  no
```

[29]:
```
from sklearn.preprocessing import MinMaxScaler
```

[30]:
```
scaler = MinMaxScaler()
df["duration"] = scaler.fit_transform(df[["duration"]])
df["pdays"] = scaler.fit_transform(df[["pdays"]])
df.head()
```

[30]:
```
   age  job  marital  education default   balance  housing  loan   contact  \
0   30  0.0        1        1.0      no  0.068455        0     0  cellular
1   33  0.0        1        2.0      no  0.108750        1     1  cellular
2   35  1.0        0        3.0      no  0.062590        1     0  cellular
3   30  1.0        1        3.0      no  0.064281        1     1   unknown
4   59  0.0        1        2.0      no  0.044469        1     0   unknown
```

	day	month	duration	campaign	pdays	previous	poutcome	y
0	19	NaN	0.024826	1	0.000000	0	NaN	no
1	11	NaN	0.071500	1	0.389908	4	1.0	no
2	16	NaN	0.059914	1	0.379587	1	1.0	no
3	3	NaN	0.064548	4	0.000000	0	NaN	no
4	5	NaN	0.073486	1	0.000000	0	NaN	no

[ ]: