```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns

df=pd.read_csv("IceAlex (1).csv",index_col=0)
df.head()
```

```
        Code               Email  Phone Number  Gender
Title  \
#

1  Trainer 1  trainer1@gmail.com    1253334345  Female    Business
Developer
2  Trainer 2  trainer1@gmail.com    1253334345    Male  Business
Consultant
3  Trainer 3  trainer1@gmail.com    1253334345  Female
Marketier
4  Trainer 4  trainer1@gmail.com    1253334345    Male  Business
Consultant
5  Trainer 5  trainer1@gmail.com    1253334345  Female       Project
manager

  Organization                          Area of expertise  Years of
Experience  \
#

1           Na                      Business Development
4
2           Na  (Business Model) Innovation Consultancy
4
3           Na                         Digital Marketing
5
4           Na                      Business consultancy
6
5           Na                         Impact Management
4

   Social Media  CV  ...                          Projcet 1   Projcet 1
Year  \
#                   ...

1           NaN  Na  ...                   Startups of Alex
2023
2           NaN  Na  ...                TIEC Pre-incubation
2024
3           NaN  Na  ...                    Talent4Startups
2024
4           NaN  Na  ...                   Startups of Alex
2023
```

```
5           NaN  Na  ...  Ladies Startups League Superstar
2023

   Project 1 hourly rate (EGP)  Project 2  Projcet 2 Year  \
#
1                          400        NaN             NaN
2                          600        NaN             NaN
3                          400        NaN             NaN
4                          500        NaN             NaN
5                          200        NaN             NaN

   Project 2 hourly rate (EGP)  Project 3  Projcet 3 Year  \
#
1                          NaN        NaN             NaN
2                          NaN        NaN             NaN
3                          NaN        NaN             NaN
4                          NaN        NaN             NaN
5                          NaN        NaN             NaN

   Project 3 hourly rate (EGP)  Satifaction rate %
#
1                          NaN              81.00%
2                          NaN              21.52%
3                          NaN              55.81%
4                          NaN              57.45%
5                          NaN              32.19%

[5 rows x 21 columns]
```

```
#columns names and data types
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 30 entries, 1 to 30
Data columns (total 21 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   Code                         30 non-null     object
 1   Email                        30 non-null     object
 2   Phone Number                 30 non-null     int64
 3   Gender                       30 non-null     object
 4   Title                        30 non-null     object
 5   Organization                 30 non-null     object
 6   Area of expertise            30 non-null     object
 7   Years of Experience          30 non-null     int64
 8   Social Media                 0 non-null      float64
 9   CV                           30 non-null     object
 10  No. Projects Joined          30 non-null     int64
 11   Projcet 1                   26 non-null     object
 12   Projcet 1 Year              30 non-null     int64
```

```
 13  Project 1 hourly rate (EGP)  30 non-null     int64
 14  Project 2                    6 non-null      object
 15   Projcet 2 Year              6 non-null      float64
 16  Project 2 hourly rate (EGP)  9 non-null      float64
 17  Project 3                    5 non-null      object
 18   Projcet 3 Year              2 non-null      float64
 19  Project 3 hourly rate (EGP)  5 non-null      float64
 20  Satifaction rate %           30 non-null     object
dtypes: float64(5), int64(5), object(11)
memory usage: 5.2+ KB
```

```
#Remove white spaces from column name
df.columns=df.columns.str.strip()
df.columns
```

```
Index(['Code', 'Email', 'Phone Number', 'Gender', 'Title',
'Organization',
       'Area of expertise', 'Years of Experience', 'Social Media',
'CV',
       'No. Projects Joined', 'Projcet 1', 'Projcet 1 Year',
       'Project 1 hourly rate (EGP)', 'Project 2', 'Projcet 2 Year',
       'Project 2 hourly rate (EGP)', 'Project 3', 'Projcet 3 Year',
       'Project 3 hourly rate (EGP)', 'Satifaction rate %'],
      dtype='object')
```

```
#Unpivot data from wide to long format
project=df.loc[:,['Projcet 1','Project 2','Project
3']].melt(value_name='Project').drop("variable",axis=1)
year=df.loc[:,['Projcet 1 Year','Projcet 2 Year','Projcet 3
Year']].melt(value_name='Project_year').drop("variable",axis=1)
rate=df.loc[:,['Project 1 hourly rate (EGP)','Project 2 hourly rate
(EGP)','Project 3 hourly rate
(EGP)']].melt(value_name='Project_hourly_rate').drop("variable",axis=1
)
```

```
#drop unpivoted column and concatenate new columns
df=pd.melt(df,id_vars=['Code', 'Email', 'Phone Number', 'Gender',
'Title', 'Organization','Area of expertise', 'Years of Experience',
'Social Media', 'CV',
       'No. Projects Joined','Projcet 1 Year',
       'Project 1 hourly rate (EGP)', 'Projcet 2 Year',
       'Project 2 hourly rate (EGP)', 'Projcet 3 Year',
       'Project 3 hourly rate (EGP)', 'Satifaction rate
%'],value_vars=['Projcet 1','Project 2','Project
3'],value_name="Project").drop('variable',axis=1)
df
```

```
          Code               Email  Phone Number  Gender  \
0    Trainer 1  trainer1@gmail.com    1253334345  Female
1    Trainer 2  trainer1@gmail.com    1253334345    Male
```

```
2    Trainer 3  trainer1@gmail.com    1253334345  Female
3    Trainer 4  trainer1@gmail.com    1253334345    Male
4    Trainer 5  trainer1@gmail.com    1253334345  Female
..         ...                 ...           ...     ...
85  Trainer 26  trainer1@gmail.com    1253334345    Male
86  Trainer 27  trainer1@gmail.com    1253334345    Male
87  Trainer 28  trainer1@gmail.com    1253334345    Male
88  Trainer 29  trainer1@gmail.com    1253334345  Female
89  Trainer 30  trainer1@gmail.com    1253334345    Male

                   Title Organization  \
0     Business Developer           Na
1   Business Consultant            Na
2              Marketier           Na
3   Business Consultant            Na
4        Project manager           Na
..                   ...          ...
85       Project manager           Na
86             Marketier           Na
87      Finance director           Na
88          Enterpreneur           Na
89          Enterpreneur           Na

                                  Area of expertise  Years of
Experience  \
0                              Business Development
4
1          (Business Model) Innovation Consultancy
4
2                                 Digital Marketing
5
3                              Business consultancy
6
4                                 Impact Management
4
..                                               ...                    ..
.
85  Senior Consultant & Learning Management Expert
8
86                                Digital Marketing
6
87                                           Finance
5
88               Curriculum Development Consultant
4
89                                             Legal
5

    Social Media  CV  No. Projects Joined  Projcet 1 Year  \
0            NaN  Na                    1            2023
```

```
1            NaN  Na                    1            2024
2            NaN  Na                    1            2024
3            NaN  Na                    1            2023
4            NaN  Na                    1            2023
..           ...  ..                  ...             ...
85           NaN  Na                    1            2022
86           NaN  Na                    2            2023
87           NaN  Na                    1            2023
88           NaN  Na                    1            2023
89           NaN  Na                    2            2023

    Project 1 hourly rate (EGP)  Projcet 2 Year  Project 2 hourly rate
(EGP)  \
0                           400             NaN
NaN
1                           600             NaN
NaN
2                           400             NaN
NaN
3                           500             NaN
NaN
4                           200             NaN
NaN
..                          ...             ...
...
85                          200             NaN
NaN
86                          100          2024.0
500.0
87                          400             NaN
NaN
88                          300             NaN
NaN
89                          240          2024.0
100.0

    Projcet 3 Year  Project 3 hourly rate (EGP) Satifaction rate %  \
0              NaN                         NaN             81.00%
1              NaN                         NaN             21.52%
2              NaN                         NaN             55.81%
3              NaN                         NaN             57.45%
4              NaN                         NaN             32.19%
..             ...                         ...                ...
85             NaN                         NaN             78.67%
86             NaN                         NaN             80.76%
87             NaN                         NaN             82.84%
88             NaN                         NaN             84.93%
89             NaN                         NaN             87.01%

                            Project
```

```
0                    Startups of Alex
1                TIEC Pre-incubation
2                    Talent4Startups
3                    Startups of Alex
4   Ladies Startups League Superstar
..                                ...
85                                NaN
86                                NaN
87                                NaN
88                                NaN
89                                NaN

[90 rows x 19 columns]
```

```
df=pd.melt(df,id_vars=['Code', 'Email', 'Phone Number', 'Gender',
'Title', 'Organization','Area of expertise', 'Years of Experience',
'Social Media', 'CV',
       'No. Projects Joined','Project 1 hourly rate (EGP)', 'Project 2
hourly rate (EGP)', 'Project 3 hourly rate (EGP)',
"Project",'Satifaction rate %']
          ,value_vars=['Projcet 1 Year','Projcet 2 Year','Projcet 3
Year'],value_name="Project_year").drop('variable',axis=1)
df
```

```
           Code                Email  Phone Number  Gender  \
0     Trainer 1  trainer1@gmail.com    1253334345  Female
1     Trainer 2  trainer1@gmail.com    1253334345    Male
2     Trainer 3  trainer1@gmail.com    1253334345  Female
3     Trainer 4  trainer1@gmail.com    1253334345    Male
4     Trainer 5  trainer1@gmail.com    1253334345  Female
..          ...                 ...           ...     ...
265  Trainer 26  trainer1@gmail.com    1253334345    Male
266  Trainer 27  trainer1@gmail.com    1253334345    Male
267  Trainer 28  trainer1@gmail.com    1253334345    Male
268  Trainer 29  trainer1@gmail.com    1253334345  Female
269  Trainer 30  trainer1@gmail.com    1253334345    Male

                    Title Organization  \
0      Business Developer           Na
1    Business Consultant            Na
2              Marketier            Na
3    Business Consultant            Na
4         Project manager           Na
..                    ...          ...
265       Project manager           Na
266             Marketier           Na
267      Finance director           Na
268          Enterpreneur           Na
269          Enterpreneur           Na
```

```
                                  Area of expertise  Years of
Experience  \
0                               Business Development
4
1          (Business Model) Innovation Consultancy
4
2                                  Digital Marketing
5
3                               Business consultancy
6
4                                  Impact Management
4
..                                                ...                    .
..
265  Senior Consultant & Learning Management Expert
8
266                                Digital Marketing
6
267                                          Finance
5
268               Curriculum Development Consultant
4
269                                            Legal
5

     Social Media  CV  No. Projects Joined  Project 1 hourly rate
(EGP)  \
0              NaN  Na                    1
400
1              NaN  Na                    1
600
2              NaN  Na                    1
400
3              NaN  Na                    1
500
4              NaN  Na                    1
200
..             ...  ..                  ...                           ..
.
265            NaN  Na                    1
200
266            NaN  Na                    2
100
267            NaN  Na                    1
400
268            NaN  Na                    1
300
269            NaN  Na                    2
240
```

```
     Project 2 hourly rate (EGP)  Project 3 hourly rate (EGP)  \
0                            NaN                          NaN
1                            NaN                          NaN
2                            NaN                          NaN
3                            NaN                          NaN
4                            NaN                          NaN
..                           ...                          ...
265                          NaN                          NaN
266                        500.0                          NaN
267                          NaN                          NaN
268                          NaN                          NaN
269                        100.0                          NaN

                                Project Satifaction rate %  Project_year

0                    Startups of Alex                81.00%        2023.0

1                  TIEC Pre-incubation               21.52%        2024.0

2                     Talent4Startups                55.81%        2024.0

3                    Startups of Alex                57.45%        2023.0

4    Ladies Startups League Superstar                32.19%        2023.0

..                                 ...                   ...           ...

265                                NaN               78.67%           NaN

266                                NaN               80.76%           NaN

267                                NaN               82.84%           NaN

268                                NaN               84.93%           NaN

269                                NaN               87.01%           NaN


[270 rows x 17 columns]
```

```
df=pd.melt(df,id_vars=['Code', 'Email', 'Phone Number', 'Gender',
'Title', 'Organization','Area of expertise', 'Years of Experience',
'Social Media', 'CV',
       'No. Projects Joined', "Project",'Project_year','Satifaction
rate %']
          ,value_vars=['Project 1 hourly rate (EGP)','Project 2 hourly
rate (EGP)','Project 3 hourly rate
(EGP)'],value_name="Project_yearly_rate").drop('variable',axis=1)
df
```

```
           Code                Email  Phone Number  Gender  \
0     Trainer 1  trainer1@gmail.com    1253334345  Female
1     Trainer 2  trainer1@gmail.com    1253334345    Male
2     Trainer 3  trainer1@gmail.com    1253334345  Female
3     Trainer 4  trainer1@gmail.com    1253334345    Male
4     Trainer 5  trainer1@gmail.com    1253334345  Female
..          ...                 ...           ...     ...
805  Trainer 26  trainer1@gmail.com    1253334345    Male
806  Trainer 27  trainer1@gmail.com    1253334345    Male
807  Trainer 28  trainer1@gmail.com    1253334345    Male
808  Trainer 29  trainer1@gmail.com    1253334345  Female
809  Trainer 30  trainer1@gmail.com    1253334345    Male

                    Title Organization  \
0      Business Developer           Na
1    Business Consultant            Na
2              Marketier            Na
3    Business Consultant            Na
4         Project manager           Na
..                    ...          ...
805       Project manager           Na
806             Marketier           Na
807      Finance director           Na
808          Enterpreneur           Na
809          Enterpreneur           Na

                                  Area of expertise  Years of
Experience  \
0                                Business Development
4
1           (Business Model) Innovation Consultancy
4
2                                   Digital Marketing
5
3                                Business consultancy
6
4                                   Impact Management
4
..                                                ...                   .
..
805  Senior Consultant & Learning Management Expert
8
806                                 Digital Marketing
6
807                                           Finance
5
808               Curriculum Development Consultant
4
809                                             Legal
5
```

```
     Social Media  CV  No. Projects Joined
Project  \
0              NaN  Na                    1                    Startups
of Alex
1              NaN  Na                    1                TIEC Pre-
incubation
2              NaN  Na                    1
Talent4Startups
3              NaN  Na                    1                    Startups
of Alex
4              NaN  Na                    1  Ladies Startups League
Superstar
..             ...  ..                  ...
...
805            NaN  Na                    1
NaN
806            NaN  Na                    2
NaN
807            NaN  Na                    1
NaN
808            NaN  Na                    1
NaN
809            NaN  Na                    2
NaN

     Project_year Satifaction rate %  Project_yearly_rate
0          2023.0              81.00%                400.0
1          2024.0              21.52%                600.0
2          2024.0              55.81%                400.0
3          2023.0              57.45%                500.0
4          2023.0              32.19%                200.0
..            ...                 ...                  ...
805           NaN              78.67%                  NaN
806           NaN              80.76%                  NaN
807           NaN              82.84%                  NaN
808           NaN              84.93%                  NaN
809           NaN              87.01%                  NaN

[810 rows x 15 columns]
```

Check Null and Duplicates

```
df.shape
```

```
(810, 15)
```

```
#check for duplicates
df.duplicated().sum()
```

```
481
```

```
#Drop Duplicate values
df.drop_duplicates(inplace=True)
df.duplicated().sum()
```

```
0
```

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

```
Code                      0
Email                     0
Phone Number              0
Gender                    0
Title                     0
Organization              0
Area of expertise         0
Years of Experience       0
Social Media            329
CV                        0
No. Projects Joined       0
Project                 138
Project_year            139
Satifaction rate %        0
Project_yearly_rate     112
dtype: int64
```

```
df.drop(["Social Media","CV","Organization"],axis=1,inplace=True)
```

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

```
Code                      0
Email                     0
Phone Number              0
Gender                    0
Title                     0
Area of expertise         0
Years of Experience       0
No. Projects Joined       0
Project                 138
Project_year            139
Satifaction rate %        0
Project_yearly_rate     112
dtype: int64
```

```
df.dropna(inplace=True)
df.isnull().sum()
```

```
Code                    0
Email                   0
Phone Number            0
```

```
Gender                  0
Title                   0
Area of expertise       0
Years of Experience     0
No. Projects Joined     0
Project                 0
Project_year            0
Satifaction rate %      0
Project_yearly_rate     0
dtype: int64
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 82 entries, 0 to 624
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   Code                  82 non-null     object 
 1   Email                 82 non-null     object 
 2   Phone Number          82 non-null     int64  
 3   Gender                82 non-null     object 
 4   Title                 82 non-null     object 
 5   Area of expertise     82 non-null     object 
 6   Years of Experience   82 non-null     int64  
 7   No. Projects Joined   82 non-null     int64  
 8   Project               82 non-null     object 
 9   Project_year          82 non-null     float64
 10  Satifaction rate %    82 non-null     object 
 11  Project_yearly_rate   82 non-null     float64
dtypes: float64(2), int64(3), object(7)
memory usage: 8.3+ KB
```

Fix Data types

```
df['Phone Number']=df['Phone Number'].astype(str)
df['Gender']=df['Gender'].astype('category')
df['Project_year']=df['Project_year'].astype(int)
df['Project_year']=df['Project_year'].astype(str)

df['Satifaction rate %']=df['Satifaction rate %'].str.replace("%","")
df['Satifaction rate %']=df['Satifaction rate %'].astype(float)

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 82 entries, 0 to 624
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
```

```
 0   Code                  82 non-null     object
 1   Email                 82 non-null     object
 2   Phone Number          82 non-null     object
 3   Gender                82 non-null     category
 4   Title                 82 non-null     object
 5   Area of expertise     82 non-null     object
 6   Years of Experience   82 non-null     int64
 7   No. Projects Joined   82 non-null     int64
 8   Project               82 non-null     object
 9   Project_year          82 non-null     object
 10  Satifaction rate %    82 non-null     float64
 11  Project_yearly_rate   82 non-null     float64
dtypes: category(1), float64(2), int64(2), object(7)
memory usage: 7.9+ KB
```

Analysis

```
df.describe()
```

```
       Years of Experience  No. Projects Joined  Satifaction rate %  \
count            82.000000            82.000000           82.000000
mean              6.451220             1.682927           61.734634
std               1.325462             0.664404           27.496569
min               3.000000             1.000000            2.060000
25%               6.000000             1.000000           56.190000
50%               7.000000             2.000000           66.160000
75%               7.750000             2.000000           82.380000
max               8.000000             3.000000           99.350000

       Project_yearly_rate
count            82.000000
mean            335.000000
std             143.812721
min             100.000000
25%             210.000000
50%             350.000000
75%             450.000000
max             600.000000
```

```
df.describe(include="object")
```

```
              Code                Email Phone Number
Title  \
count           82                   82           82
82
unique          26                    1            1
10
top     Trainer 25  trainer1@gmail.com   1253334345  Business
Consultant
freq             9                   82           82
```

```
24
```

```
        Area of expertise     Project Project_year
count                  82          82           82
unique                 19          12            4
top     Digital Marketing  Heya Raeda         2023
freq                   20          20           44
```

```
for col in df.columns:
    if df[col].dtype=="object":
        print(f'{col} unique values: {df[col].nunique()}')
        print(f"{df[col].value_counts()}")
        print()
```

```
Code unique values: 26
Trainer 25    9
Trainer 30    8
Trainer 27    8
Trainer 11    8
Trainer 20    8
Trainer 8     6
Trainer 9     6
Trainer 10    6
Trainer 18    6
Trainer 16    1
Trainer 29    1
Trainer 28    1
Trainer 26    1
Trainer 19    1
Trainer 17    1
Trainer 1     1
Trainer 15    1
Trainer 2     1
Trainer 13    1
Trainer 12    1
Trainer 7     1
Trainer 6     1
Trainer 5     1
Trainer 4     1
Trainer 3     1
Trainer 14    1
Name: Code, dtype: int64

Email unique values: 1
trainer1@gmail.com    82
Name: Email, dtype: int64

Phone Number unique values: 1
1253334345    82
Name: Phone Number, dtype: int64
```

```
Title unique values: 10
Business Consultant     24
Marketier               20
Finance director        10
Enterpreneur             9
Business Expert          6
Senior Consultant        6
Project manager          2
Frontend developer       2
Backend developer        2
Business Developer       1
Name: Title, dtype: int64

Area of expertise unique values: 19
Digital Marketing                                    20
Businesss/Marketing/Finance                           9
Legal                                                 8
Project Manager & Innovation Management Consultant    8
Business Consultant & Team Development                6
HR and Business Expert                                6
Senior Consultant                                     6
Business Development & Marketing Consultant           6
Backend Web Development                               2
Frontend Web Development                              2
(Business Model) Innovation Consultancy               1
Business Modeling Consultant                          1
Startups Advisor, Business Mentor & Trainer           1
Impact Management                                     1
Business consultancy                                  1
Senior Consultant & Learning Management Expert        1
Finance                                               1
Curriculum Development Consultant                     1
Business Development                                  1
Name: Area of expertise, dtype: int64

Project unique values: 12
Heya Raeda                      20
Startups of Alex                13
Talent4Startups                 12
TIEC Pre-incubation              9
Ladies Startups League Superstar 8
Talent 4 Startups                4
TIEC re-incubation               4
Safir                            3
Ladies Startups League           3
SwitchMED                        3
InvestMED                        2
Mentoring Lounge                 1
Name: Project, dtype: int64
```

```
Project_year unique values: 4
2023    44
2024    25
2022     7
2021     6
Name: Project_year, dtype: int64
```

```
trainer_project=df.groupby('Code')['No. Projects Joined'].count()
trainer_project
```

```
Code
Trainer 1     1
Trainer 10    6
Trainer 11    8
Trainer 12    1
Trainer 13    1
Trainer 14    1
Trainer 15    1
Trainer 16    1
Trainer 17    1
Trainer 18    6
Trainer 19    1
Trainer 2     1
Trainer 20    8
Trainer 25    9
Trainer 26    1
Trainer 27    8
Trainer 28    1
Trainer 29    1
Trainer 3     1
Trainer 30    8
Trainer 4     1
Trainer 5     1
Trainer 6     1
Trainer 7     1
Trainer 8     6
Trainer 9     6
Name: No. Projects Joined, dtype: int64
```

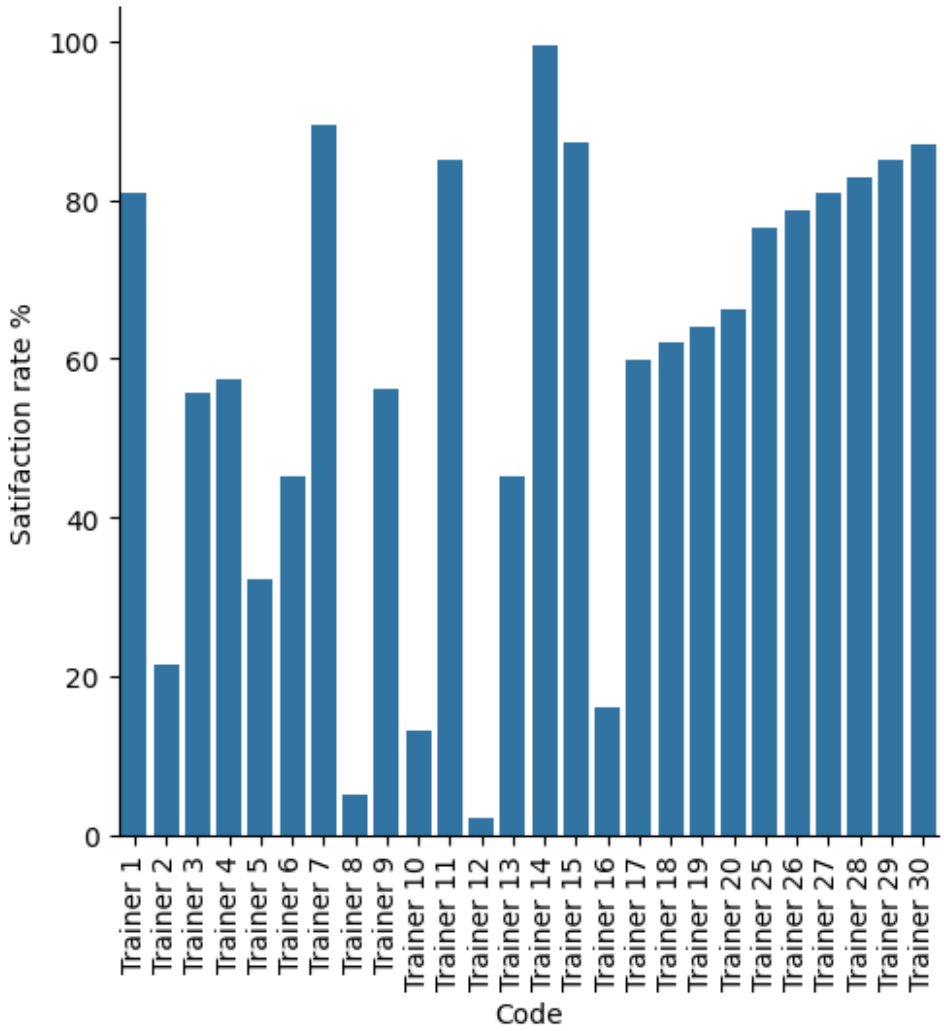
```
df.columns
```

```
Index(['Code', 'Email', 'Phone Number', 'Gender', 'Title', 'Area of
expertise',
       'Years of Experience', 'No. Projects Joined', 'Project',
'Project_year',
       'Satifaction rate %', 'Project_yearly_rate'],
      dtype='object')
```

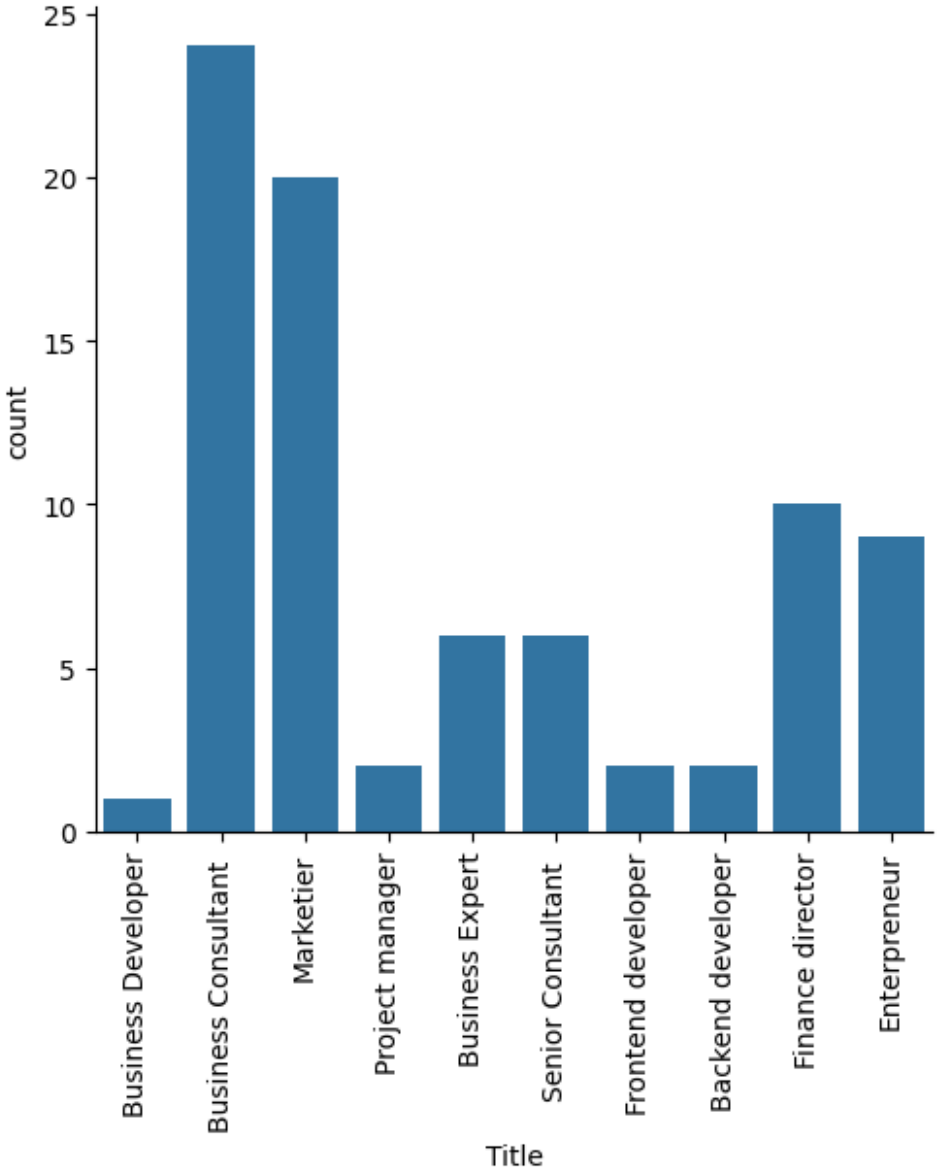
```
df.head()
```

```
        Code                Email Phone Number  Gender                    
Title  \
0  Trainer 1  trainer1@gmail.com   1253334345  Female    Business        
Developer
1  Trainer 2  trainer1@gmail.com   1253334345    Male  Business          
Consultant
2  Trainer 3  trainer1@gmail.com   1253334345  Female                    
Marketier
3  Trainer 4  trainer1@gmail.com   1253334345    Male  Business          
Consultant
4  Trainer 5  trainer1@gmail.com   1253334345  Female       Project      
manager

                         Area of expertise  Years of Experience  \
0                      Business Development                    4
1  (Business Model) Innovation Consultancy                    4
2                         Digital Marketing                    5
3                      Business consultancy                    6
4                         Impact Management                    4

   No. Projects Joined                             Project Project_year
\
0                    1                    Startups of Alex         2023

1                    1                TIEC Pre-incubation          2024

2                    1                     Talent4Startups         2024

3                    1                    Startups of Alex         2023

4                    1  Ladies Startups League Superstar          2023


   Satifaction rate %  Project_yearly_rate
0               81.00                400.0
1               21.52                600.0
2               55.81                400.0
3               57.45                500.0
4               32.19                200.0
```

```
sns.catplot(data=df,x="Code",y="Satifaction rate %",kind="bar")
plt.xticks(rotation=90)
plt.show()
```




```
sns.catplot(data=df,x="Title",kind="count")
plt.xticks(rotation=90)
plt.show()
```


25
20
15
10
5
0
count
Business Developer
Business Consultant
Marketier
Project manager
Business Expert
Senior Consultant
Frontend developer
Backend developer
Finance director
Enterpreneur
Title