```
import random


from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
import pandas as pd


from datetime import datetime


data = {
    'Name': [
        'Alice', 'Bob', 'Charlie', 'David', 'Eva', 'Frank', 'Grace', 'Helen', 'Ian', 'Jane',
        'Karl', 'Laura', 'Mike', 'Nina', 'Oscar', 'Paula', 'Quentin', 'Rachel', 'Steve', 'Tina',
        'Uma', 'Victor', 'Wendy', 'Xander', 'Yara', 'Zane'
    ],
    'Department': [
        'HR', 'IT', 'Finance', 'IT', 'HR', 'Finance', 'IT', 'Marketing', 'Sales', 'HR',
        'IT', 'Finance', 'Sales', 'Marketing', 'IT', 'HR', 'Finance', 'Sales', 'IT', 'Marketing',
        'HR', 'Finance', 'Sales', 'IT', 'Marketing', 'Finance'
    ],
    'Salary': [
        50000, 60000, 55000, 70000, 48000, 62000, 65000, 52000, 58000, 51000,
        75000, 53000, 60000, 57000, 67000, 49500, 61000, 59000, 72000, 54000,
        47000, 56000, 61000, 68000, 55000, 64000
    ],
    'Join_Date': [
        '2020-05-21', '2019-03-15', '2021-07-10', '2018-11-01', '2022-01-05', '2017-09-12',
        '2020-06-30', '2019-08-20', '2021-02-11', '2020-12-01', '2016-04-25', '2019-11-15',
        '2020-03-03', '2022-04-10', '2018-07-18', '2017-10-22', '2015-05-09', '2020-09-14',
        '2021-01-19', '2016-12-30', '2023-01-11', '2018-02-27', '2019-06-06', '2020-08-08',
        '2021-03-22', '2017-01-17'
    ]
}


df
```

	Name	Department	Salary	Join_Date	Years_with_Company
0	Alice	HR	50000	2020-05-21	4
1	Bob	IT	60000	2019-03-15	5
2	Charlie	Finance	55000	2021-07-10	3
3	David	IT	70000	2018-11-01	6
4	Eva	HR	48000	2022-01-05	3
5	Frank	Finance	62000	2017-09-12	7
6	Grace	IT	65000	2020-06-30	4
7	Helen	Marketing	52000	2019-08-20	5
8	Ian	Sales	58000	2021-02-11	4
9	Jane	HR	51000	2020-12-01	4
10	Karl	IT	75000	2016-04-25	8
11	Laura	Finance	53000	2019-11-15	5
12	Mike	Sales	60000	2020-03-03	4
13	Nina	Marketing	57000	2022-04-10	2
14	Oscar	IT	67000	2018-07-18	6
15	Paula	HR	49500	2017-10-22	7
16	Quentin	Finance	61000	2015-05-09	9
17	Rachel	Sales	59000	2020-09-14	4
18	Steve	IT	72000	2021-01-19	4
19	Tina	Marketing	54000	2016-12-30	8
20	Uma	HR	47000	2023-01-11	2
21	Victor	Finance	56000	2018-02-27	6
22	Wendy	Sales	61000	2019-06-06	5
23	Xander	IT	68000	2020-08-08	4
24	Yara	Marketing	55000	2021-03-22	3
25	Zane	Finance	64000	2017-01-17	8

Next steps: Generate code with df | View recommended plots | New interactive sheet

```
df.describe()
```

	Salary	Join_Date	Years_with_Company
count	26.000000	26	26.000000
mean	58826.923077	2019-08-18 19:23:04.615384576	5.000000
min	47000.000000	2015-05-09 00:00:00	2.000000
25%	53250.000000	2018-04-03 06:00:00	4.000000
50%	58500.000000	2020-01-08 12:00:00	4.500000
75%	63500.000000	2021-01-06 18:00:00	6.000000
max	75000.000000	2023-01-11 00:00:00	9.000000
std	7553.730612	NaN	1.918333

```
df.columns
```

```
Index(['Name', 'Department', 'Salary', 'Join_Date', 'Years_with_Company'], dtype='object')
```

```
df = pd.DataFrame(data)

df['Join_Date'] = pd.to_datetime(df['Join_Date'])
employees_after_2020 = df[df['Join_Date'] > '2020-01-01']['Name']
```

```
print("Employees who joined after January 1, 2020:")
print(employees_after_2020)
print("\n")
```

```
Employees who joined after January 1, 2020:
0       Alice
2     Charlie
4         Eva
6       Grace
8         Ian
9        Jane
12       Mike
13       Nina
17     Rachel
18      Steve
20        Uma
23     Xander
24       Yara
Name: Name, dtype: object
```

```
average_salary_by_department = df.groupby('Department')['Salary'].mean()
print("Average salary for each department:")
print(average_salary_by_department)
print("\n")
```

```
Average salary for each department:
Department
Finance      58500.000000
HR           49100.000000
IT           68142.857143
Marketing    54500.000000
Sales        59500.000000
Name: Salary, dtype: float64
```

```
department_with_highest_average_salary = average_salary_by_department.mean()
print("Department with the highest average salary:")
print(department_with_highest_average_salary)
print("\n")
```

```
Department with the highest average salary:
57948.571428571435
```

```
today = pd.to_datetime('2025-02-12')
df['Years_with_Company'] = (today - df['Join_Date']).dt.days // 365
print("DataFrame with 'Years_with_Company' column:")
print(df)
print("\n")
```

```
DataFrame with 'Years_with_Company' column:
       Name Department  Salary  Join_Date  Years_with_Company
0     Alice         HR   50000 2020-05-21                   4
1       Bob         IT   60000 2019-03-15                   5
2   Charlie    Finance   55000 2021-07-10                   3
3     David         IT   70000 2018-11-01                   6
4       Eva         HR   48000 2022-01-05                   3
5     Frank    Finance   62000 2017-09-12                   7
6     Grace         IT   65000 2020-06-30                   4
7     Helen  Marketing   52000 2019-08-20                   5
8       Ian      Sales   58000 2021-02-11                   4
9      Jane         HR   51000 2020-12-01                   4
10     Karl         IT   75000 2016-04-25                   8
11    Laura    Finance   53000 2019-11-15                   5
12     Mike      Sales   60000 2020-03-03                   4
13     Nina  Marketing   57000 2022-04-10                   2
14    Oscar         IT   67000 2018-07-18                   6
15    Paula         HR   49500 2017-10-22                   7
16  Quentin    Finance   61000 2015-05-09                   9
17   Rachel      Sales   59000 2020-09-14                   4
18    Steve         IT   72000 2021-01-19                   4
19     Tina  Marketing   54000 2016-12-30                   8
20      Uma         HR   47000 2023-01-11                   2
21   Victor    Finance   56000 2018-02-27                   6
22    Wendy      Sales   61000 2019-06-06                   5
23   Xander         IT   68000 2020-08-08                   4
```

```
24     Yara  Marketing   55000 2021-03-22                3
25     Zane    Finance   64000 2017-01-17                8
```

```
df_sorted = df.sort_values(by='Years_with_Company', ascending=False)
print("DataFrame sorted by 'Years_with_Company' in descending order:")
print(df_sorted)
print("\n")
```

```
DataFrame sorted by 'Years_with_Company' in descending order:
       Name Department  Salary  Join_Date  Years_with_Company
16  Quentin    Finance   61000 2015-05-09                   9
25     Zane    Finance   64000 2017-01-17                   8
19     Tina  Marketing   54000 2016-12-30                   8
10     Karl         IT   75000 2016-04-25                   8
5     Frank    Finance   62000 2017-09-12                   7
15    Paula         HR   49500 2017-10-22                   7
3     David         IT   70000 2018-11-01                   6
21   Victor    Finance   56000 2018-02-27                   6
14    Oscar         IT   67000 2018-07-18                   6
22    Wendy      Sales   61000 2019-06-06                   5
7     Helen  Marketing   52000 2019-08-20                   5
11    Laura    Finance   53000 2019-11-15                   5
1       Bob         IT   60000 2019-03-15                   5
23   Xander         IT   68000 2020-08-08                   4
18    Steve         IT   72000 2021-01-19                   4
17   Rachel      Sales   59000 2020-09-14                   4
0     Alice         HR   50000 2020-05-21                   4
12     Mike      Sales   60000 2020-03-03                   4
9      Jane         HR   51000 2020-12-01                   4
8       Ian      Sales   58000 2021-02-11                   4
6     Grace         IT   65000 2020-06-30                   4
4       Eva         HR   48000 2022-01-05                   3
2   Charlie    Finance   55000 2021-07-10                   3
24     Yara  Marketing   55000 2021-03-22                   3
20      Uma         HR   47000 2023-01-11                   2
13     Nina  Marketing   57000 2022-04-10                   2
```

```
%matplotlib inline
```

```
df_sorted.plot(y='Salary', x='Years_with_Company', kind='bar')
```

```
<Axes: xlabel='Years_with_Company'>
```