```python
In [1]:  import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         %matplotlib inline
```

Matplotlib is building the font cache; this may take a moment.

```python
In [4]:  #importing file
         df=pd.read_csv("C:\\Users\\ravie\\Downloads\\Final dataset Attrition (1).csv",parse_dates=["Date_of_Hire"])
         df.head()
```

C:\Users\ravie\AppData\Local\Temp\ipykernel_1072\3758639998.py:2: UserWarning: Parsing dates in DD/MM/YYYY form
at when dayfirst=False (the default) was specified. This may lead to inconsistently parsed dates! Specify a for
mat to ensure consistent parsing.
  df=pd.read_csv("C:\\Users\\ravie\\Downloads\\Final dataset Attrition (1).csv",parse_dates=["Date_of_Hire"])

Out[4]:

	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Gender	JobInvolvement	JobLevel	JobRole	JobSatisfaction	...	Da
0	37	Yes	Travel_Rarely	Research & Development	2	Male	2	1	Laboratory Technician	3	...	
1	21	No	Travel_Rarely	Research & Development	15	Male	3	1	Research Scientist	4	...	
2	45	No	Travel_Rarely	Research & Development	6	Male	3	3	Research Director	1	...	
3	23	No	Travel_Rarely	Sales	2	Male	3	1	Sales Representative	1	...	
4	22	No	Travel_Rarely	Research & Development	15	Female	3	1	Laboratory Technician	4	...	

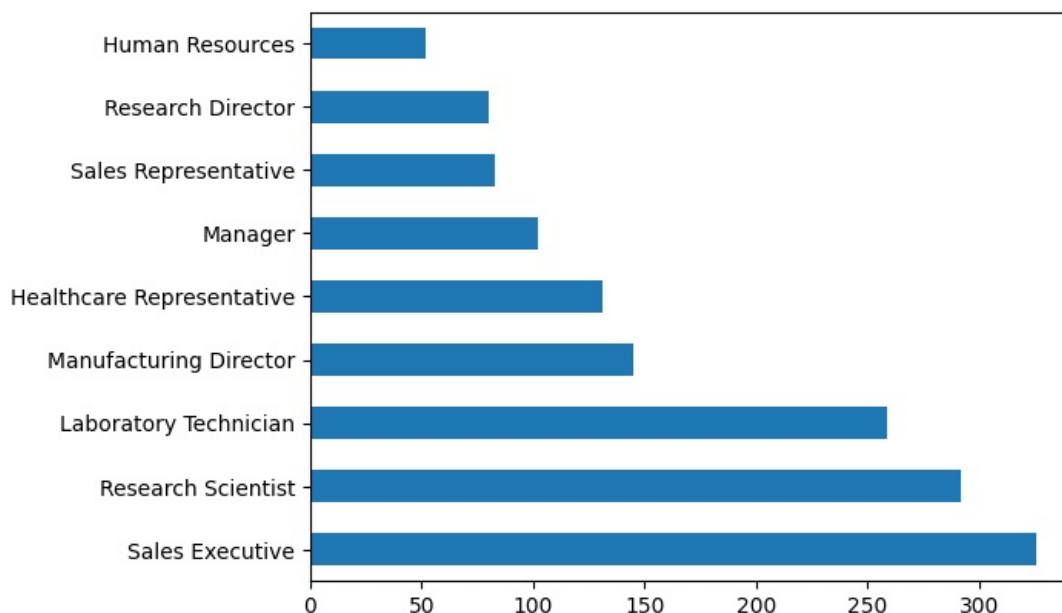5 rows × 33 columns

```python
In [5]:  #count of job roles
         a=df["JobRole"].value_counts()
         a
```

```
Out[5]:  Sales Executive             326
         Research Scientist          292
         Laboratory Technician       259
         Manufacturing Director      145
         Healthcare Representative   131
         Manager                     102
         Sales Representative         83
         Research Director            80
         Human Resources              52
         Name: JobRole, dtype: int64
```

```python
In [6]:  a.plot.barh()
```

Out[6]: <AxesSubplot: >


```python
In [7]:  #mean age
         df.Age.mean()
```
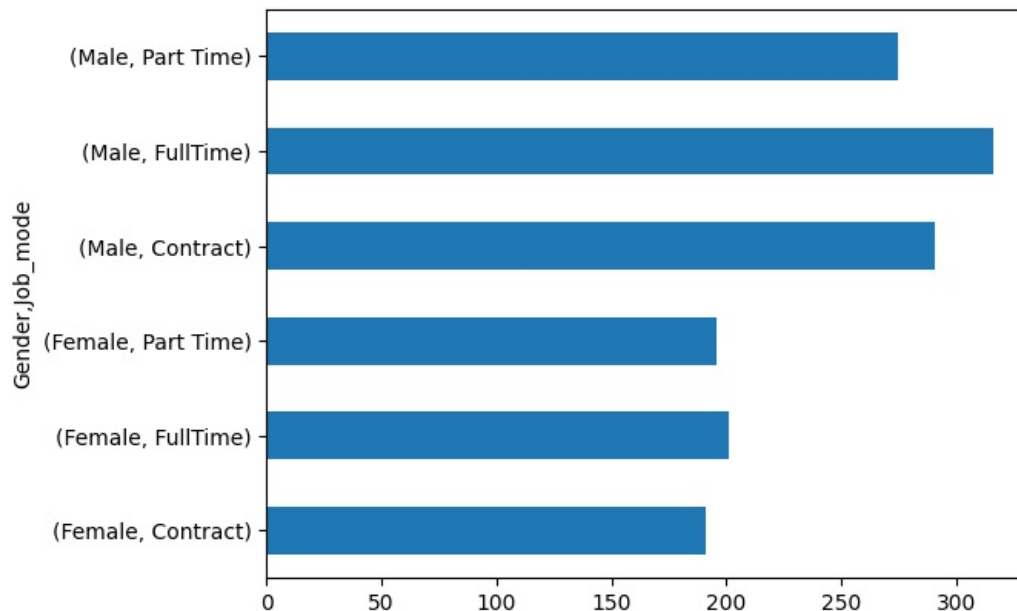
Out[7]: 36.923809523809524

```
In [8]:  #job mode count
         df.head()
         a1=df.groupby(["Gender","Job_mode"])["Gender"].count()
         a1
```

```
Out[8]:  Gender  Job_mode
         Female  Contract     191
                 FullTime     201
                 Part Time    196
         Male    Contract     291
                 FullTime     316
                 Part Time    275
         Name: Gender, dtype: int64
```

```
In [9]:  a1.plot.barh()
```

```
Out[9]:  <AxesSubplot: ylabel='Gender,Job_mode'>
```



```
In [10]:  #count higher education
          df.head()
          b=df.Higher_Education.value_counts()
          b
```
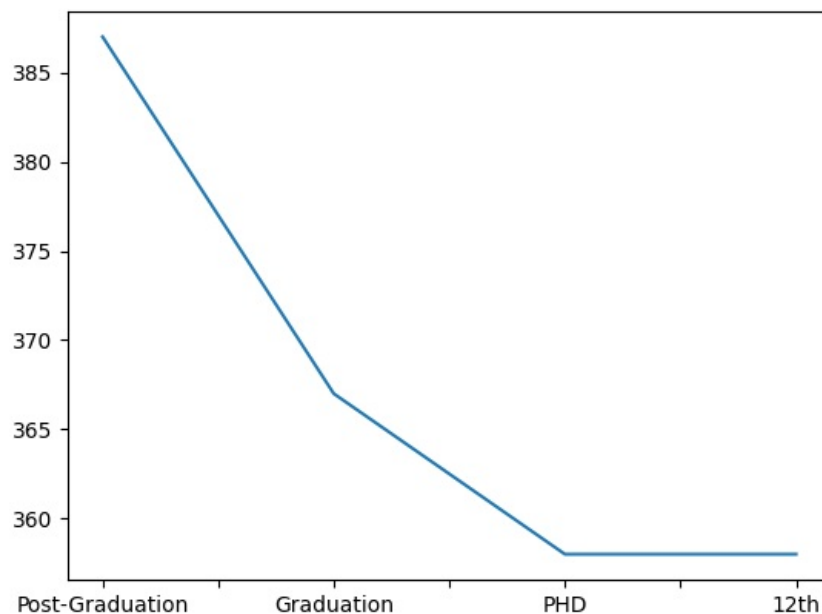
```
Out[10]:  Post-Graduation    387
          Graduation         367
          PHD                358
          12th               358
          Name: Higher_Education, dtype: int64
```

```
In [11]:  b.plot()
```

```
Out[11]:  <AxesSubplot: >
```



```
In [12]:  #number of hires for each month
```

```python
df.head()
c=df.groupby(df["Date_of_Hire"].dt.month_name())["Gender"].value_counts()
c
```

Out[12]:
```
Date_of_Hire  Gender
April         Male      137
              Female     67
August        Male       31
              Female     13
December      Male       36
              Female     31
February      Male      113
              Female     82
January       Male      102
              Female     74
July          Female     26
              Male       21
June          Male      107
              Female     78
March         Male      134
              Female     88
May           Male      133
              Female     77
November      Male       30
              Female     14
October       Female     19
              Male       18
September     Male       20
              Female     19
Name: Gender, dtype: int64
```

In [13]:
```python
c.plot.barh()
```
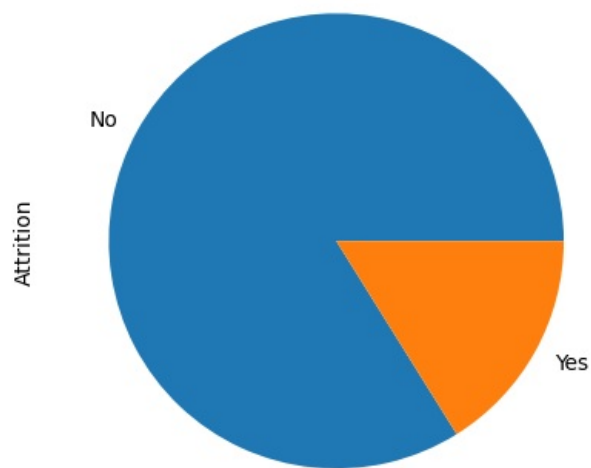
Out[13]: `<AxesSubplot: ylabel='Date_of_Hire,Gender'>`
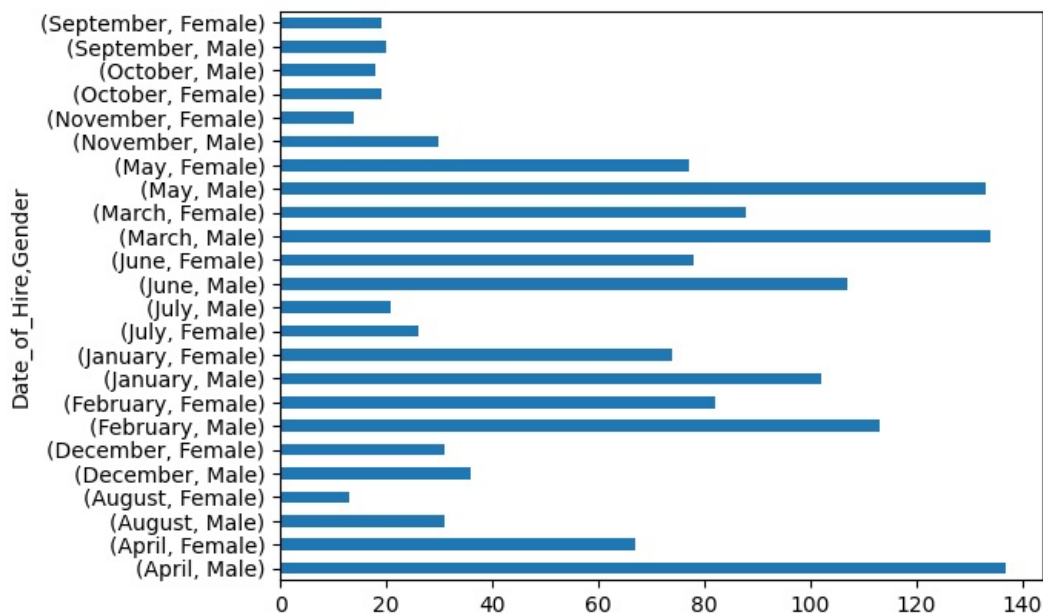
In [15]:
```python
#percentage of attrition
df.head()
d=df.Attrition.value_counts()
```

In [16]:
```python
d.plot.pie()
```

Out[16]: `<AxesSubplot: ylabel='Attrition'>`

```python
#Different job roles with ppl having different education
df.head()
df.groupby("JobRole")["Higher_Education"].value_counts()
```

```
JobRole                      Higher_Education
Healthcare Representative    12th               49
                             Post-Graduation    30
                             Graduation         27
                             PHD                25
Human Resources              Graduation         16
                             PHD                13
                             12th               12
                             Post-Graduation    11
Laboratory Technician        Graduation         67
                             Post-Graduation    67
                             PHD                65
                             12th               60
Manager                      Post-Graduation    31
                             Graduation         28
                             PHD                24
                             12th               19
Manufacturing Director       Graduation         41
                             Post-Graduation    41
                             12th               32
                             PHD                31
Research Director            12th               24
                             Graduation         20
                             PHD                18
                             Post-Graduation    18
Research Scientist           Graduation         89
                             Post-Graduation    75
                             PHD                65
                             12th               63
Sales Executive              Post-Graduation    93
                             PHD                90
                             12th               75
                             Graduation         68
Sales Representative         PHD                27
                             12th               24
                             Post-Graduation    21
                             Graduation         11
Name: Higher_Education, dtype: int64
```

In [18]: df.Department.value_counts()

```
Out[18]: Research & Development    961
         Sales                     446
         Human Resources            63
         Name: Department, dtype: int64
```

```
In [19]: df.head()
         df.groupby(df["Date_of_Hire"].dt.weekday)["Higher_Education"].value_counts()
```

```
Out[19]: Date_of_Hire   Higher_Education
         0              PHD                54
                        Post-Graduation    52
                        Graduation         46
                        12th               42
         1              Graduation         61
                        12th               60
                        PHD                60
                        Post-Graduation    48
         2              Post-Graduation    59
                        PHD                50
                        12th               48
                        Graduation         43
         3              12th               72
                        Post-Graduation    56
                        Graduation         50
                        PHD                45
         4              Graduation         55
                        Post-Graduation    51
                        12th               50
                        PHD                47
         5              Post-Graduation    65
                        Graduation         54
                        12th               45
                        PHD                45
         6              Graduation         58
                        PHD                57
                        Post-Graduation    56
                        12th               41
         Name: Higher_Education, dtype: int64
```

```
In [20]: df.groupby([df["Date_of_Hire"].dt.month_name()])["Higher_Education"].value_counts()
```

```
Out[20]: Date_of_Hire  Higher_Education
         April         Post-Graduation    60
                       12th               52
                       Graduation         51
                       PHD                41
         August        12th               12
                       Graduation         11
                       PHD                11
                       Post-Graduation    10
         December      PHD                26
                       Graduation         20
                       Post-Graduation    11
                       12th               10
         February      12th               56
                       Post-Graduation    53
                       Graduation         49
                       PHD                37
         January       Post-Graduation    54
                       PHD                43
                       12th               40
                       Graduation         39
         July          Graduation         17
                       Post-Graduation    12
                       PHD                10
                       12th                8
         June          Graduation         48
                       PHD                47
                       Post-Graduation    47
                       12th               43
         March         PHD                60
                       Graduation         57
                       12th               54
                       Post-Graduation    51
         May           Post-Graduation    58
                       PHD                56
                       Graduation         51
                       12th               45
         November      12th               16
                       Post-Graduation    11
                       PHD                 9
                       Graduation          8
         October       12th               12
                       PHD                10
                       Post-Graduation     8
                       Graduation          7
         September     Post-Graduation    12
                       12th               10
                       Graduation          9
                       PHD                 8
         Name: Higher_Education, dtype: int64
```

In [21]: `df["BusinessTravel"].value_counts()`

```
Out[21]: Travel_Rarely        1043
         Travel_Frequently     277
         Non-Travel            150
         Name: BusinessTravel, dtype: int64
```
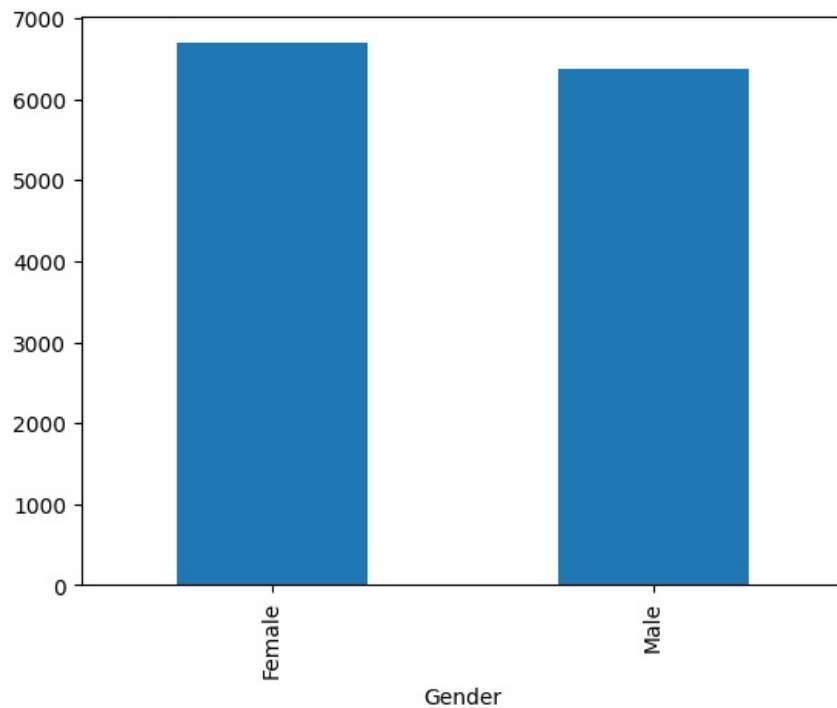
In [22]: `df.columns`

```
Out[22]: Index(['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',
                'Gender', 'JobInvolvement', 'JobLevel', 'JobRole', 'JobSatisfaction',
                'MaritalStatus', 'MonthlyIncome', 'NumCompaniesWorked', 'OverTime',
                'PercentSalaryHike', 'PerformanceRating', 'StockOptionLevel',
                'TotalWorkingYears', 'TrainingTimesLastYear', 'YearsAtCompany',
                'YearsSinceLastPromotion', 'YearsWithCurrManager', 'Higher_Education',
                'Date_of_Hire', 'Date_of_termination', 'Status_of_leaving',
                'Mode_of_work', 'Leaves', 'Absenteeism', 'Work_accident',
                'Source_of_Hire', 'Job_mode', 'Unnamed: 32'],
               dtype='object')
```

In [23]: `f=df.groupby("Gender")["MonthlyIncome"].mean()`

In [24]: `f.plot.bar()`

Out[24]: `<AxesSubplot: xlabel='Gender'>`

`df.head()`

Out[25]:

	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Gender	JobInvolvement	JobLevel	JobRole	JobSatisfaction	...	Da
0	37	Yes	Travel_Rarely	Research & Development	2	Male	2	1	Laboratory Technician	3	...	
1	21	No	Travel_Rarely	Research & Development	15	Male	3	1	Research Scientist	4	...	
2	45	No	Travel_Rarely	Research & Development	6	Male	3	3	Research Director	1	...	
3	23	No	Travel_Rarely	Sales	2	Male	3	1	Sales Representative	1	...	
4	22	No	Travel_Rarely	Research & Development	15	Female	3	1	Laboratory Technician	4	...	

5 rows × 33 columns

In [26]: `df[["Department","MonthlyIncome"]].groupby(["Department"]).max("MonthlyIncome")`

Out[26]:

	MonthlyIncome
Department	
Human Resources	19717
Research & Development	19999
Sales	19847

In [27]: `df["Mode_of_work"].unique()`

Out[27]: `array(['OFFICE', 'WFH'], dtype=object)`

In [ ]:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js