### employee-data

#### August 23, 2023

```
[29]: import pandas as pd
      import numpy as np
      import matplotlib.pyplot as plt
      import seaborn as sns
      import scipy.stats as stats
      import plotly.express as px
      import statsmodels.api as sm
 [5]: data = pd.read_csv('C:\\Users\\96653\\Desktop\\Test jupyer\\Employee DATA.csv ')
      data.head()
 [6]:
         Age Attrition
                            BusinessTravel DailyRate
                                                                     Department
      0
          41
                   Yes
                             Travel_Rarely
                                                  1102
                                                                          Sales
          49
      1
                    No
                         Travel_Frequently
                                                   279
                                                        Research & Development
      2
          37
                             Travel_Rarely
                                                        Research & Development
                   Yes
                                                  1373
      3
          33
                    No
                         Travel_Frequently
                                                  1392
                                                        Research & Development
      4
          27
                             Travel_Rarely
                                                   591
                                                        Research & Development
                    No
         DistanceFromHome
                            Education EducationField
                                                       EmployeeCount
                                                                       EmployeeNumber
      0
                                    2 Life Sciences
                         1
                         8
                                    1 Life Sciences
                                                                                    2
      1
                                                                    1
                         2
      2
                                    2
                                                Other
                                                                    1
                                                                                    4
      3
                         3
                                      Life Sciences
                                                                    1
                                                                                    5
      4
                         2
                                                                                    7
                                    1
                                             Medical
            RelationshipSatisfaction StandardHours
                                                      StockOptionLevel
      0
                                    4
                                                  80
                                                                      1
      1
                                    2
                                                                      0
      2
                                                  80
                                    3
      3
                                                  80
                                                                      0
                                                  80
                             TrainingTimesLastYear WorkLifeBalance
                                                                     YearsAtCompany
         TotalWorkingYears
      0
                          8
                                                  0
                                                                   1
                                                                                   6
      1
                         10
                                                  3
                                                                   3
                                                                                  10
      2
                          7
                                                  3
                                                                   3
                                                                                   0
```

3	8	3	3	8
4	6	3	3	2

	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
0	4	0	5
1	7	1	7
2	0	0	0
3	7	3	0
4	2	2	2

[5 rows x 35 columns]

#### [7]: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):

#	Column	Non-Null Count	Dtype
0	Age	1470 non-null	int64
1	Attrition	1470 non-null	object
2	BusinessTravel	1470 non-null	object
3	DailyRate	1470 non-null	int64
4	Department	1470 non-null	object
5	DistanceFromHome	1470 non-null	int64
6	Education	1470 non-null	int64
7	EducationField	1470 non-null	object
8	EmployeeCount	1470 non-null	int64
9	EmployeeNumber	1470 non-null	int64
10	${\tt EnvironmentSatisfaction}$	1470 non-null	int64
11	Gender	1470 non-null	object
12	HourlyRate	1470 non-null	int64
13	JobInvolvement	1470 non-null	int64
14	JobLevel	1470 non-null	int64
15	JobRole	1470 non-null	object
16	${\tt JobSatisfaction}$	1470 non-null	int64
17	MaritalStatus	1470 non-null	object
18	MonthlyIncome	1470 non-null	int64
19	MonthlyRate	1470 non-null	int64
20	NumCompaniesWorked	1470 non-null	int64
21	Over18	1470 non-null	object
22	OverTime	1470 non-null	object
23	${\tt PercentSalaryHike}$	1470 non-null	int64
24	PerformanceRating	1470 non-null	int64
25	${\tt RelationshipSatisfaction}$	1470 non-null	int64
26	StandardHours	1470 non-null	int64
27	${\tt StockOptionLevel}$	1470 non-null	int64

${ t TotalWorking Years}$	1470	non-null	int64
${\tt Training Times Last Year}$	1470	non-null	int64
WorkLifeBalance	1470	non-null	int64
YearsAtCompany	1470	non-null	int64
YearsInCurrentRole	1470	non-null	int64
${\tt YearsSinceLastPromotion}$	1470	non-null	int64
YearsWithCurrManager	1470	non-null	int64
	TrainingTimesLastYear WorkLifeBalance YearsAtCompany YearsInCurrentRole YearsSinceLastPromotion	TrainingTimesLastYear 1470 WorkLifeBalance 1470 YearsAtCompany 1470 YearsInCurrentRole 1470 YearsSinceLastPromotion 1470	TrainingTimesLastYear 1470 non-null WorkLifeBalance 1470 non-null YearsAtCompany 1470 non-null YearsInCurrentRole 1470 non-null YearsSinceLastPromotion 1470 non-null

dtypes: int64(26), object(9)
memory usage: 402.1+ KB

#### [8]: data.describe()

[8]:		Age	Da	ilyRate	DistanceFr	omHom	e Educatio	n Er	mployeeCoun	t \
[0]	count	1470.000000		0.000000		00000			1470.	
	mean	36.923810		2.485714		19251			1.	
	std	9.135373		3.509100		10686			0.	
	min	18.000000		2.000000		00000			1.	
	25%	30.000000		5.000000		00000			1.	
	50%	36.000000		2.000000		00000			1.	
	75%	43.000000		.000000		00000			1.	
	max	60.000000		0.000000		00000			1.	0
		EmployeeNumb	er E	nvironme	entSatisfact	ion	HourlyRate	JobIı	nvolvement	\
	count	1470.0000	00		1470.000	000	1470.000000	14	470.000000	
	mean	1024.8653	06		2.721	769	65.891156		2.729932	
	std	602.0243	35		1.093	082	20.329428		0.711561	
	min	1.0000	00		1.000	000	30.000000		1.000000	
	25%	491.2500	00		2.000	000	48.000000		2.000000	
	50%	1020.5000	00		3.000	000	66.000000		3.000000	
	75%	1555.7500	00		4.000	000	83.750000		3.000000	
	max	2068.0000	00		4.000	000	100.000000		4.000000	
		JobLevel	R	Relations	hipSatisfac		StandardHour			
	count	1470.000000	•••		1470.00		1470.			
	mean	2.063946	•••			2245	80.			
	std	1.106940	•••			1209	0.			
	min	1.000000	•••			0000	80.			
	25%	1.000000	•••		2.00		80.			
	50%	2.000000	•••		3.00		80.			
	75%	3.000000	•••		4.00		80.			
	max	5.000000	•••		4.00	0000	80.	0		
		StockOptionL	0110]	To+olWo	rkingYears	Troi	${ t ningTimesLast}$	Voor	\	
	count	1470.00			.470.000000	ııaı.	1470.00		`	
	mean	0.79		1	11.279592			9320		
	std	0.79			7.780782			9320		
	min	0.00			0.000000			0000		
	штп	0.00	0000		0.000000		0.00	,0000		

25%	0.000000	6.000	000	2.	000000	
50%	1.000000	10.000	000	3.	000000	
75%	1.000000	15.000			000000	
max	3.000000	40.000			000000	
	WorkLifeBalance	YearsAtCompany	YearsInC	urrentRole	\	
count	1470.000000	1470.000000	14	470.000000		
mean	2.761224	7.008163		4.229252		
std	0.706476	6.126525		3.623137		
min	1.000000	0.000000		0.000000		
25%	2.000000	3.000000		2.000000		
50%	3.000000	5.000000		3.000000		
75%	3.000000	9.000000		7.000000		
max	4.000000	40.000000		18.000000		
	YearsSinceLastPro		thCurrMan	-		
count		.000000	1470.000			
mean		. 187755	4.123			
std		. 222430	3.568	3136		
min		.000000	0.000	0000		
25%	0	.000000	2.000	0000		
50%	1	.000000	3.000	0000		
75%	3	.000000	7.000	0000		
max	15	.000000	17.000	0000		
[8 row	s x 26 columns]					
	<pre>ion_rate = data[', attrition_rate)</pre>	Attrition'].valu	e_counts(	) / len(dat	(a) * 100	
No	83.877551					
Yes	16.122449					
Name: A	Attrition, dtype:	float64				
, ich ac	tisfaction_by_role	o - doto manushr	(  Johnolo		iafoation	11 maan()
-	job_satisfaction_l	0	( Jobitole	)[ JODSati	ISTACTION	].mean()
Primo		J_1010)				
JobRole	9					
	care Representativ	e 2.786260				
	Resources	2.557692				
	cory Technician	2.691120				
Manager	•	2.705882				
•	cturing Director	2.682759				
	ch Director	2.700000				
-	THE DITECTOR	2.100000				

[9]

[10]

Research Scientist

Sales Representative

Name: JobSatisfaction, dtype: float64

Sales Executive

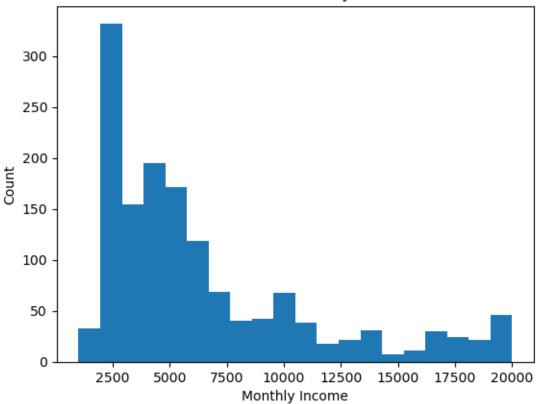
2.773973

2.754601

2.734940

```
[11]: plt.hist(data['MonthlyIncome'], bins=20)
    plt.xlabel('Monthly Income')
    plt.ylabel('Count')
    plt.title('Distribution of Monthly Incomes')
    plt.show()
```

#### Distribution of Monthly Incomes



```
[12]: average_age = data['Age'].mean()
print(average_age)
```

#### 36.923809523809524

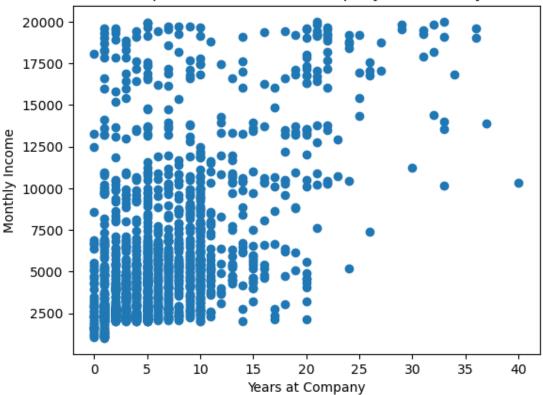
```
[13]: job_levels = data['JobLevel'].value_counts()
print(job_levels)
```

- 1 543
- 2 534
- 3 218
- 4 106
- 5 69

Name: JobLevel, dtype: int64

```
[15]: plt.scatter(data['YearsAtCompany'], data['MonthlyIncome'])
    plt.xlabel('Years at Company')
    plt.ylabel('Monthly Income')
    plt.title('Relationship between Years at Company and Monthly Income')
    plt.show()
```

#### Relationship between Years at Company and Monthly Income

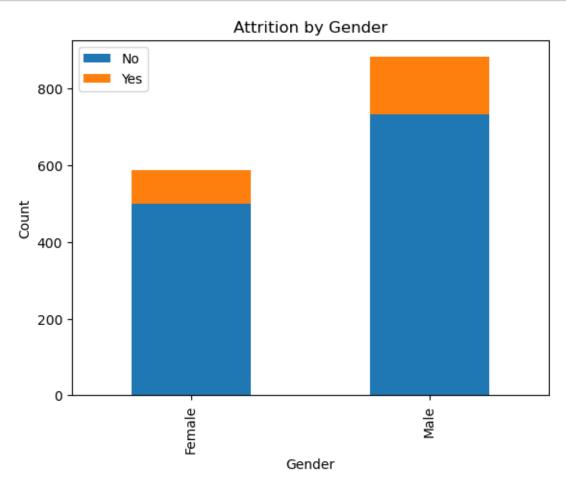


```
[16]: department_income = data.groupby('Department')['MonthlyIncome'].mean()
print(department_income)
```

Department

Human Resources 6654.507937
Research & Development 6281.252862
Sales 6959.172646
Name: MonthlyIncome, dtype: float64

```
plt.title('Attrition by Gender')
plt.legend()
plt.show()
```



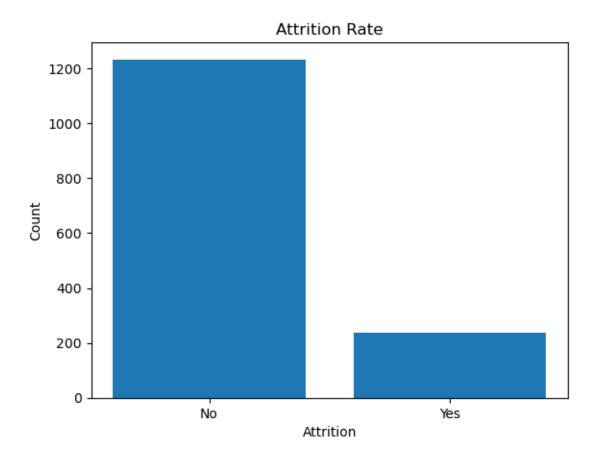
# [18]: distance\_home\_by\_role = data.groupby('JobRole')['DistanceFromHome'].mean() print(distance\_home\_by\_role)

JobRole Healthcare Representative 9.786260 Human Resources 8.173077 Laboratory Technician 9.409266 Manager 8.029412 Manufacturing Director 9.482759 Research Director 8.437500 Research Scientist 9.013699 Sales Executive 9.659509 Sales Representative 8.662651 Name: DistanceFromHome, dtype: float64

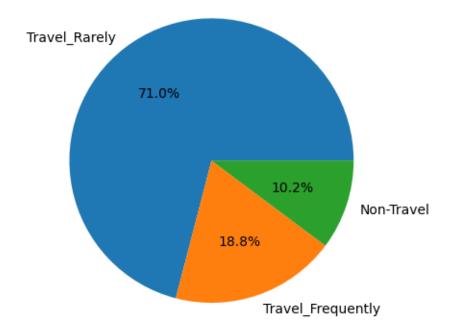
```
[19]: plt.hist(data['Age'], bins=20)
   plt.xlabel('Age')
   plt.ylabel('Count')
   plt.title('Distribution of Age')
   plt.show()
```

### Distribution of Age Count Age

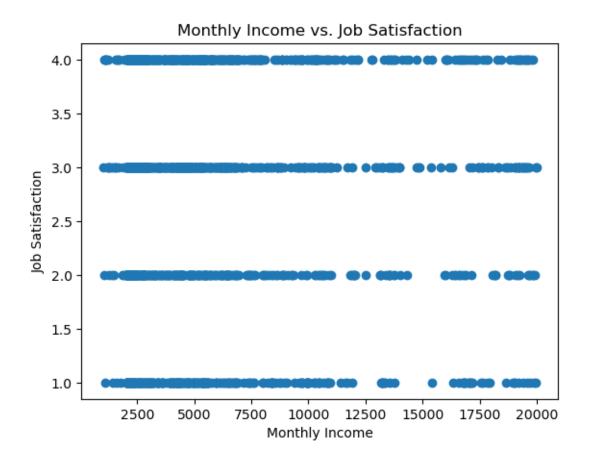
```
[20]: attrition_counts = data['Attrition'].value_counts()
    plt.bar(attrition_counts.index, attrition_counts.values)
    plt.xlabel('Attrition')
    plt.ylabel('Count')
    plt.title('Attrition Rate')
    plt.show()
```

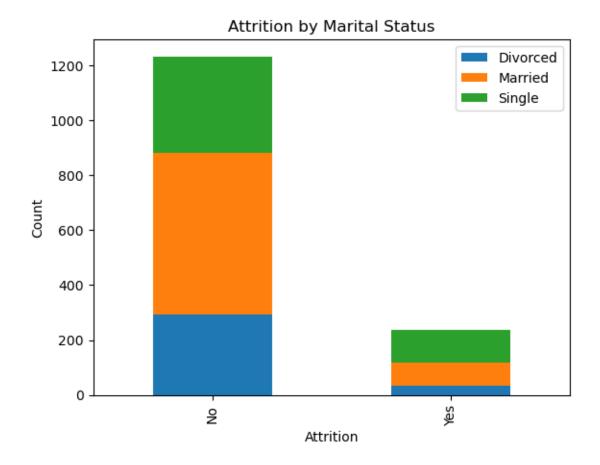


#### **Business Travel Distribution**

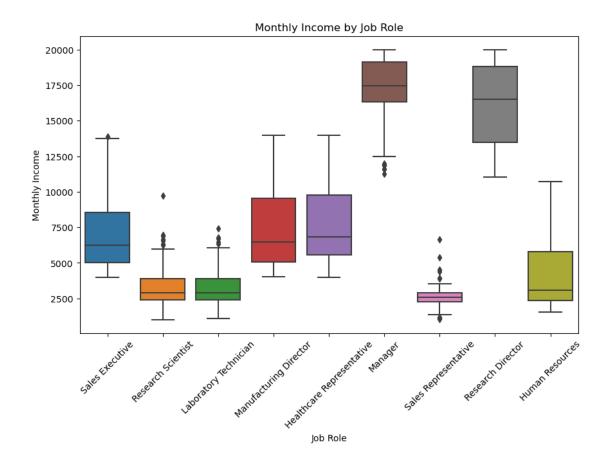


```
[22]: plt.scatter(data['MonthlyIncome'], data['JobSatisfaction'])
   plt.xlabel('Monthly Income')
   plt.ylabel('Job Satisfaction')
   plt.title('Monthly Income vs. Job Satisfaction')
   plt.show()
```





```
[25]: plt.figure(figsize=(10, 6))
    sns.boxplot(x='JobRole', y='MonthlyIncome', data=data)
    plt.xlabel('Job Role')
    plt.ylabel('Monthly Income')
    plt.title('Monthly Income by Job Role')
    plt.xticks(rotation=45)
    plt.show()
```



```
[27]: numeric_cols = data.select_dtypes(include=np.number).columns
    correlation = data[numeric_cols].corr()

plt.figure(figsize=(10, 8))
    sns.heatmap(correlation, annot=True, cmap='coolwarm')
    plt.title('Correlation Matrix')
    plt.show()
```

```
Correlation Matrix
                                                                 ).0D.01.024.0 <mark>0.51</mark>.004<mark>0.5</mark>.02 <mark>0.6</mark> 0@36601.95
                                                                                                                                   .030.650.62.020.3 D.2 D.220.2
                               Age - 10.00.000.21
                      DailyRate -.0111-0.0-0501
                                                                 . O 15 110 D80 2030 4060 003003, D 0-07 70 19 20 308002030 0.497)
                                                                                                                                    040201.100250-880340-0-90-8302
       DistanceFromHome -000.700510.02
                                                                 0308000000.D00800050003.00070327.0209004.00.700
                                                                                                                                    034.50 0-660-587002 07 059.59 159.000.01
                      Education 0.20.00.0211
                                                                 042202.70 D70420.10.0 D1040502061-6.0 D1042.50
                                                                                                                                   010.10.02509.869.06.0540
             EmployeeCount
                                                                                                                                                                                          - 0.8
          EmployeeNumber -0.00.06.1030304
                                                                 10.0 D80-0 50 006.99-D90-4060 D50-0.30 0-0.30 1-03.0-20.0
                                                                                                                                    0.60200140261.001.001.10038.4600.90
                                                                 01<mark>81-</mark>0.40500803010406050060308040303020300
EnvironmentSatisfaction -0.00.0180-060
                                                                                                                                    0036902.700902.8000.5018040600
                                                                 035.05 10,040304080-010-060059020-0910200
                    HourlyRate -.024023030101
                                                                                                                                    .0050-00203008050-406902.0-2040.207.0
             Jobinvolvement -0.03.04.608.84
                                                                 00639038694<mark>31-0</mark>.0-D30-2010-D500a604350-007020903
                                                                                                                                    90,20,905,50-1050-105002,0,0-8,70,12/40,2
                        JobLevel -0.50.00.0056.:
                                                                 .001.9040.20-2080.1<mark>310</mark>.0040.9950.040.149.049.502000
                                                                                                                                    01<mark>4.76</mark>.00£03<mark>0.5£</mark>0.3£0.350.38
                                                                                                                                                                                          - 0.6
              JobSatisfaction -0004991093.01
                                                                  994.6965.89-101.902.10.01.190.000.00.00.64.506.00.20.902.30
                                                                                                                                    0 1·0·020 9·5·8/1.9·03/292.B·D80 2
                                                                 .01.506.39-1060 1<mark>05.905</mark>00 <mark>7 2 0.0 30.1 9.0 20 70 10 20</mark>
                                                                                                                                   005470.00203<mark>0.510.360.340.34</mark>
              MonthlyIncome -0.5 000.000
                                                                 0 D30 308 0-D50 10600. 40 006043 <mark>5 10</mark> . 0 0 80 00 604 09 320
                                                                                                                                    08402.6001.9080-234001.600.603
                  MonthlyRate -.0208062020702
                                                                 001.6 D30202016.14.05<mark>0.16</mark>.01<mark>81-</mark>0.61.0 D405
  NumCompaniesWorked -0.3.0380291
                                                                                                                                    .0.<mark>0.2</mark>9.00.60841020910307.1
                                                                                                                                                                                          - 0.4
          PercentSalaryHike -003.626.04.01
                                                                 .0-D3938.1069.10-1070305.012.002.100-10401_1_0.730.0
                                                                                                                                   0 0 0 5902 (0 00 502903 (B/G) (6) 901 (5) 20 20 1
                                                                  .0-20--0306-02-20-20902-01-02-3901-10-69-80-10-4771-0.0:
                                                                                                                                    @350-6701.6@2©03.4350 D802
        PerformanceRating -00.09 0040 207.02
                                                                  .070 @70701.B 340 212 0 D 232 60 004 0 5 6 . 604 0 3
                                                                                                                                    0 4 60 2. 4 0 2250 2. 0 109 0 10 500 8 (3 0 )
 RelationshipSatisfaction -.06.407090660
               StandardHours -
           StockOptionLevel -.0380402040501
                                                                  065.20 0 304005. 0 2020 104001. 10 0+15 40 304 0030 00705006 50 4
                                                                                                                                   1 0.00L00LD004.D D505010 D402
                                                                                                                                                                                          - 0.2
                                                                 .61.46270230 <mark>557 8</mark>).0<mark>2.70</mark>.020.24.02005702
02461.9685-0550.8655822005566652160
          TotalWorkingYears 4.68.01.50.0515
                                                                                                                                    01<mark>1-0</mark>.09600<mark>0.63</mark>0.460.40.46
                                                                                                                                    0<del>1</del>0103<mark>610</mark>.02.89036050702310
   TrainingTimesLastYear -0.0209250-8702
            WorkLifeBalance -.0-2010-880200
                                                                  00.02.20.04.600.50.308.00.90.3010.0.20.00840.3330.2060
                                                                                                                                    04.D0102<mark>810</mark>.012.05008902
                                                                                                                                    01<mark>0.63</mark>003.61</mark>21 0.70.620.
            YearsAtCompany 0.30.08809.56
                                                                  01.010405602.02<mark>0.50</mark>300<mark>0.50</mark>.0204.102.038603.491
                                                                 00841802400<mark>0.33</mark>000<mark>0.36</mark>.0-0309.001.63501
                                                                                                                                    05<mark>0.45</mark>0057050.76 1 0.550.71
                                                                                                                                                                                           0.0
         YearsInCurrentRole 0.21009919.0
YearsSinceLastPromotion +0.20.03B00.05
                                                                  0090<del>1</del>060-20702<mark>0.31</mark>5.01<mark>0.34</mark>00-060-207022010803
                                                                                                                                    01 0.4 0 0 200 0.6 20.5 5 1 0.5 1
   YearsWithCurrManager -0.20.0060 D406
                                                                 0 699.12995.012 0 2 <mark>0 . 38</mark>. 0 2<mark>0 . 34</mark> . 0 307. 10. 0 0 2012 6 0 0
                                                                                                                                    020.45004002870.710.5
                                                                               Jobinvolvement -
JobLevel -
                                                                                                                                        TotalWorkingYears
                                                                                         JobSatisfaction
                                                                                               MonthlyIncome
                                                                                                                                            TrainingTimesLastYear
                                                                     EnvironmentSatisfaction
                                                                                                   MonthlyRate
                                                                                                              PercentSalaryHike  
                                                                                                                    PerformanceRating
                                                                                                                                  StockOptionLevel
                                                                                                                                                 WorkLifeBalance
                                                                                                                                                      YearsAtCompany
                                                                                                                                                            YearsInCurrentRole
                                                  DistanceFromHome
                                                      Education
                                                           EmployeeCount
                                                                EmployeeNumber
                                                                                                         NumCompaniesWorked
                                                                                                                        RelationshipSatisfaction
                                                                                                                              StandardHours
                                                                                                                                                                      YearsWithCurrManager
```

## # Print the model summary print(results.summary())

## OLS Regression Results

Method: Least Squares F-statistic: Date: Wed, 23 Aug 2023 Prob (F-statistic): Time: 09:05:33 Log-Likelihood: No. Observations: 1470 AIC: Df Residuals: 1468 BIC: Df Model: 1 Covariance Type: nonrobust		0.00		
Date: Wed, 23 Aug 2023	Adj. R-squared:			
Time: 09:05:33 Log-Likelihood: No. Observations: 1470 AIC: Df Residuals: 1468 BIC: Df Model: 1 Covariance Type: nonrobust				
No. Observations: 1470 AIC:  Df Residuals: 1468 BIC:  Df Model: 1  Covariance Type: nonrobust		0.78		
Df Residuals: 1468 BIC:  Df Model: 1  Covariance Type: nonrobust		-2229.		
Df Model: 1 Covariance Type: nonrobust		4462		
Covariance Type: nonrobust		4473		
= coef std err t P> t   0.975]				
const 2.7395 0.049 55.820 0.000 2.836  MonthlyIncome -1.676e-06 6.11e-06 -0.274 0.784 -1 1.03e-05				
0.975]				
const 2.7395 0.049 55.820 0.000 2.836  MonthlyIncome -1.676e-06 6.11e-06 -0.274 0.784 -1.1 1.03e-05  Comnibus: 1411.411 Durbin-Watson: Prob(Omnibus): 0.000 Jarque-Bera (JB): Skew: -0.329 Prob(JB): Kurtosis: 1.778 Cond. No.  Notes: [1] Standard Errors assume that the covariance matrix of the errors expecified. [2] The condition number is large, 1.37e+04. This might indicate strong multicollinearity or other numerical problems.	[0.025			
const 2.7395 0.049 55.820 0.000 2.836  MonthlyIncome -1.676e-06 6.11e-06 -0.274 0.784 -1.3 1.03e-05				
2.836  MonthlyIncome -1.676e-06 6.11e-06 -0.274 0.784 -1.3  1.03e-05				
MonthlyIncome -1.676e-06 6.11e-06 -0.274 0.784 -1.3  1.03e-05	2.643			
1.03e-05  ===================================				
1.03e-05  ===================================	1.37e-05			
Omnibus: 1411.411 Durbin-Watson: Prob(Omnibus): 0.000 Jarque-Bera (JB): Skew: -0.329 Prob(JB): Kurtosis: 1.778 Cond. No.				
Prob(Omnibus):  Skew:  -0.329 Prob(JB):  Kurtosis:  1.778 Cond. No.		2.02		
Skew: -0.329 Prob(JB): Kurtosis: 1.778 Cond. No.		118.02		
Notes:  [1] Standard Errors assume that the covariance matrix of the error specified.  [2] The condition number is large, 1.37e+04. This might indicate strong multicollinearity or other numerical problems.		2.35e-2		
Notes:  [1] Standard Errors assume that the covariance matrix of the errors as a sum of the error		1.37e+0		
	ic ondo	onere a		