

ATTRITION IN THE WORKPLACE

KELSEY BARTON
DSC530-T304
MARCH 2, 2024

THE DATASET

IBM HR Analytics Employee Attrition & Performance

- WA_Fn-UseC_-HR-Employee-Attrition
- This dataset explores a few of the factors that could result in employees leaving an organization. The dataset was created by IBM data scientists as a fictional study about attrition.

THE QUESTION

Can certain characteristics about a person predict whether they will choose to remain employed within an organization or leave the organization?

EXAMINED VARIABLES IN THE DATASET

Age

Attrition

Rate of Pay
(DailyRate)

Time it Takes
to get to
Work
(DistanceFromHome)

Education

MEANING OF THE CHOSEN VARIABLES

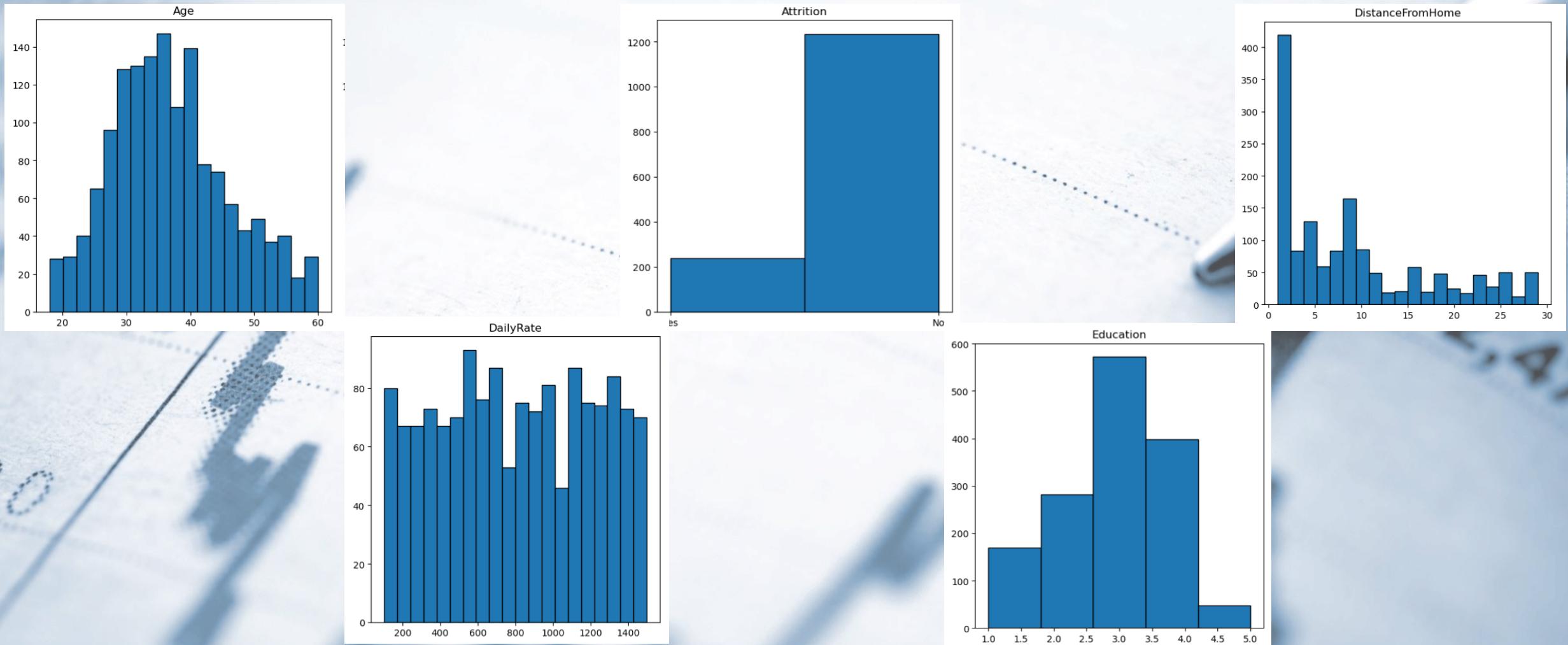
Age: Age of the examined employee

Attrition: Whether the employee has left the company or not

DailyRate: Daily rate of pay for the employee

DistanceFromHome: Distance traveled from the employee's home to the workplace

Education: Level of education attained by the employee



HISTOGRAMS

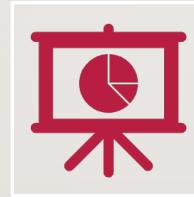
OUTLIERS

Age



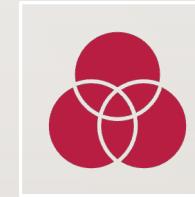
Any data points that fall far from the bulk of the data, such as ages below 20 or above 60. Outliers in age might represent valid data points (e.g., experienced employees) or data entry errors.

Daily Rate



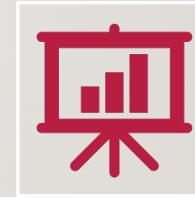
Any extremely high or low daily rates that deviate significantly from the majority of the data. These outliers could be due to genuine differences in salaries or data entry errors.

Distance from Home



Large distances from home that are not representative of the majority of employees.

Education



Unusual education levels compared to employees. These outliers might indicate employees with advanced degrees or employees with less formal education than the majority.

DESCRIPTIVE CHARACTERISTICS

```
EmployeeNumber EnvironmentSatisfaction HourlyRate JobInvolvement
count    1470.000000      1470.000000      1470.000000
mean    1024.865306      2.721769      65.891156      2.729932
std     602.024335      1.093082      20.329428      0.711561
min     1.000000      1.000000      30.000000      1.000000
25%    491.250000      2.000000      48.000000      2.000000
50%    1020.500000      3.000000      66.000000      3.000000
75%    1555.750000      4.000000      83.750000      3.000000
max    2068.000000      4.000000     100.000000      4.000000
```

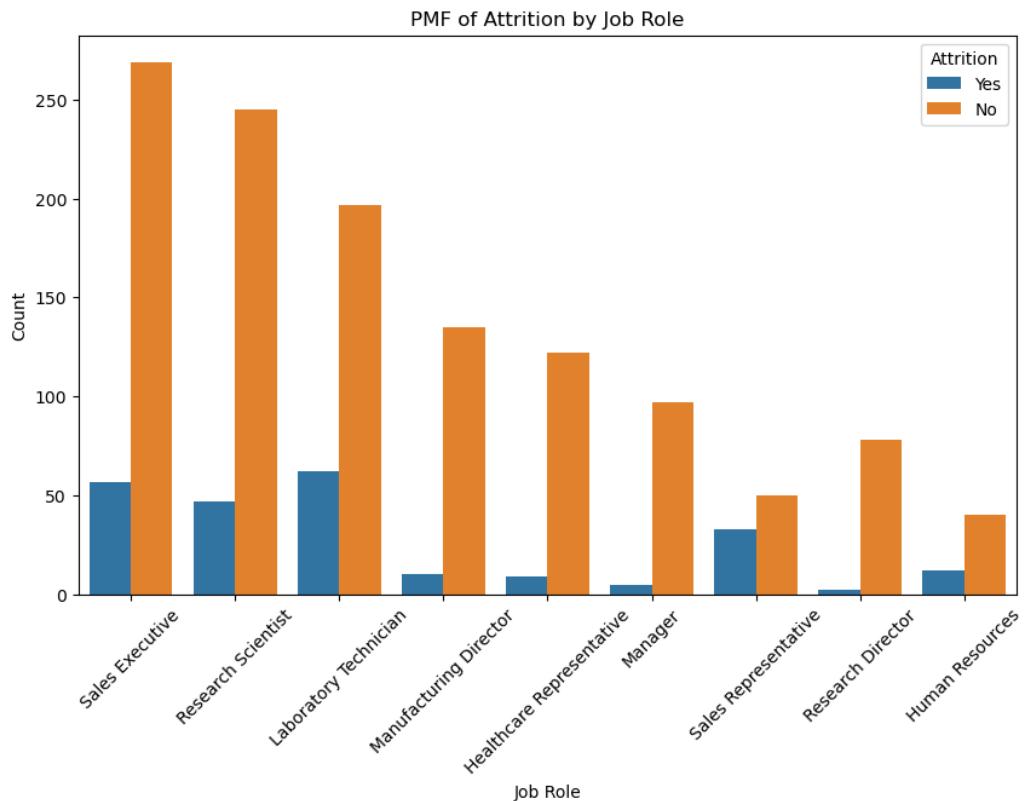
```
JobLevel ... RelationshipSatisfaction StandardHours
count    1470.000000    ...      1470.000000      1470.0
mean    2.063946    ...      2.712245      80.0
std     1.106940    ...      1.081209      0.0
min     1.000000    ...      1.000000      80.0
25%    1.000000    ...      2.000000      80.0
50%    2.000000    ...      3.000000      80.0
75%    3.000000    ...      4.000000      80.0
max    5.000000    ...      4.000000      80.0
```

```
Variable      Mean Mode Spread (Standard Deviation)
0   Age      36.923810 35      9.135373
1   Attrition  NaN  No      NaN
2   DailyRate 802.485714 691    403.509100
3   DistanceFromHome 9.192517 2      8.106864
4   Education  2.912925 3      1.024165
```

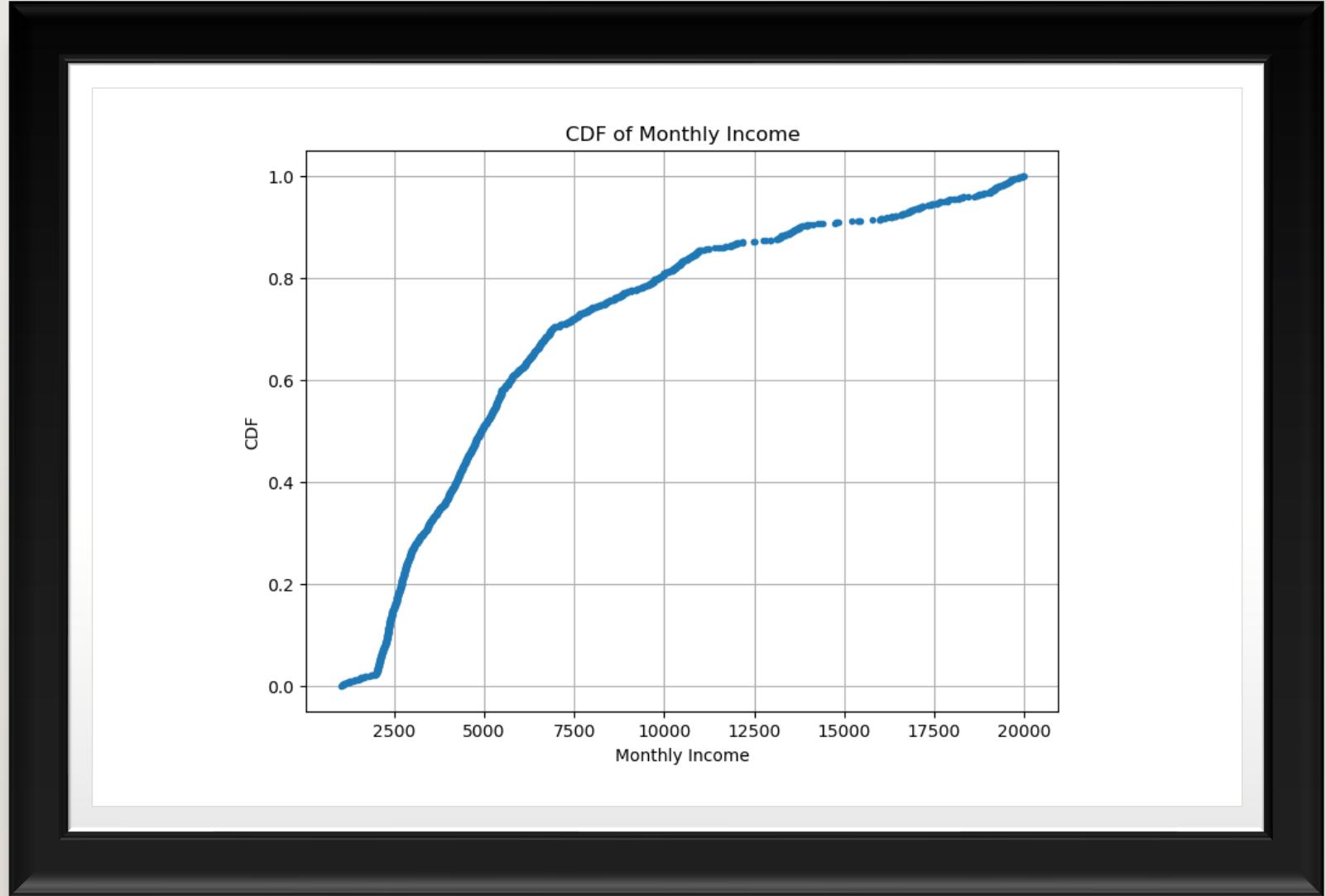
```
Tails
0   Two-tailed
1   None
2   Right-tailed
3   Right-tailed
4   None
```

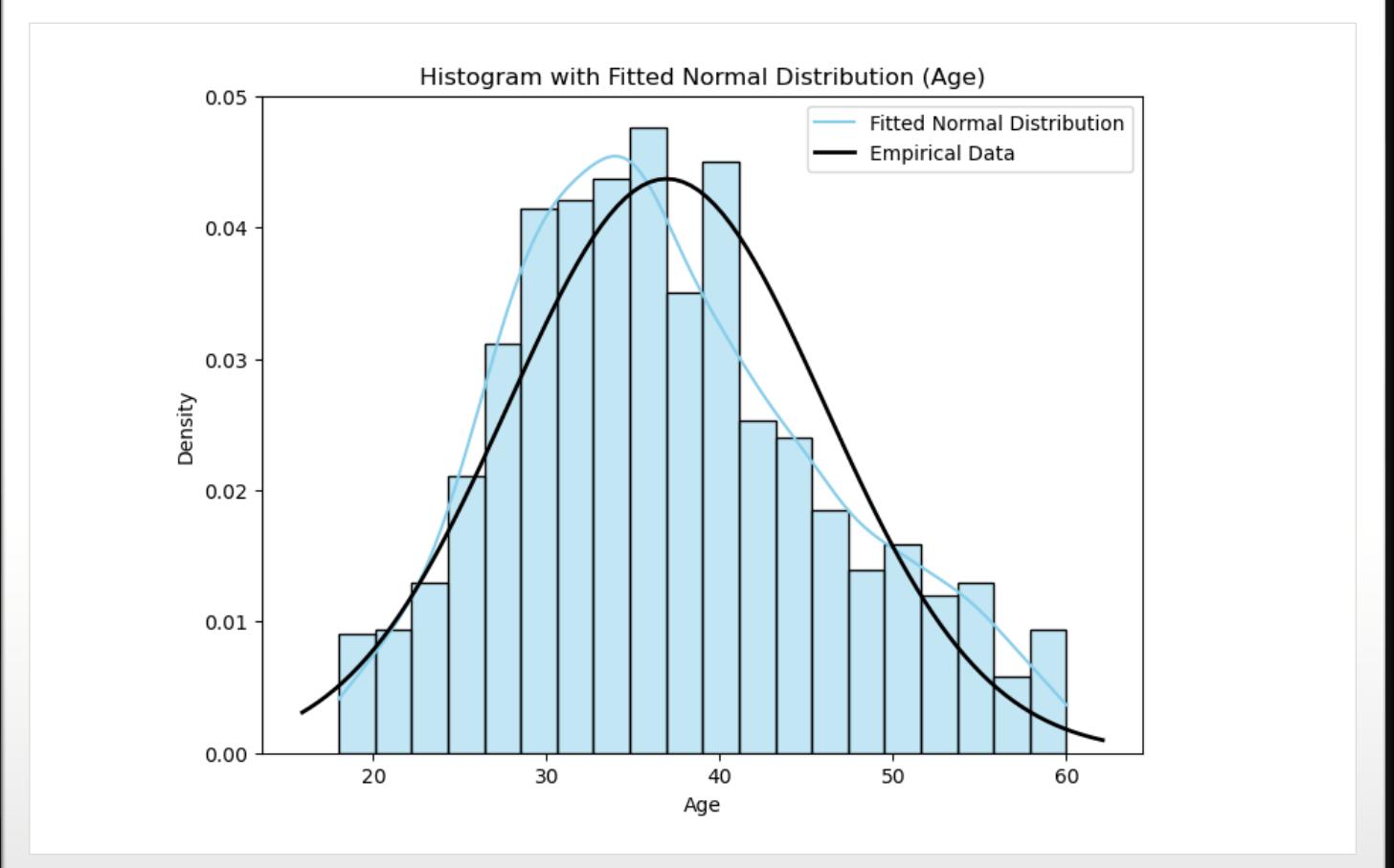
```
Age DailyRate DistanceFromHome Education EmployeeCount
count 1470.000000 1470.000000 1470.000000 1470.000000 1470.0
mean 36.923810 802.485714 9.192517 2.912925 1.0
std 9.135373 403.509100 8.106864 1.024165 0.0
min 18.000000 102.000000 1.000000 1.000000 1.0
25% 30.000000 465.000000 2.000000 2.000000 1.0
50% 36.000000 802.000000 7.000000 3.000000 1.0
75% 43.000000 1157.000000 14.000000 4.000000 1.0
max 60.000000 1499.000000 29.000000 5.000000 1.0
```

PMF COMPARISON

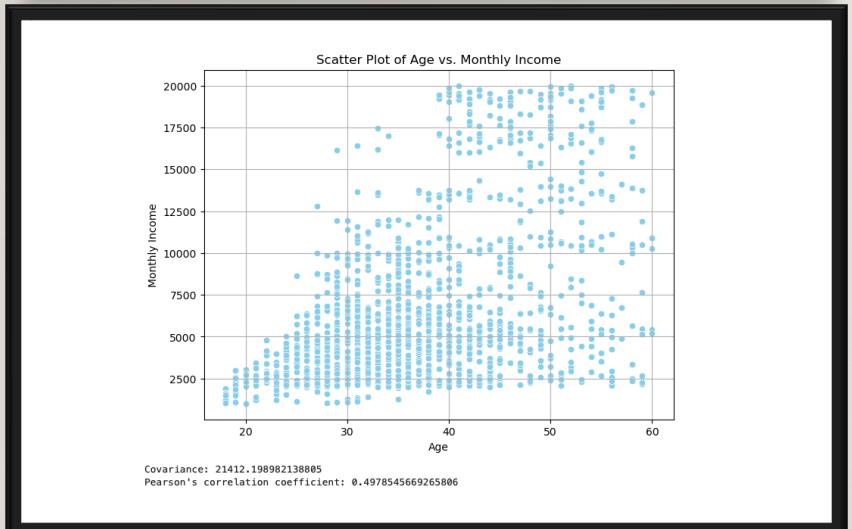
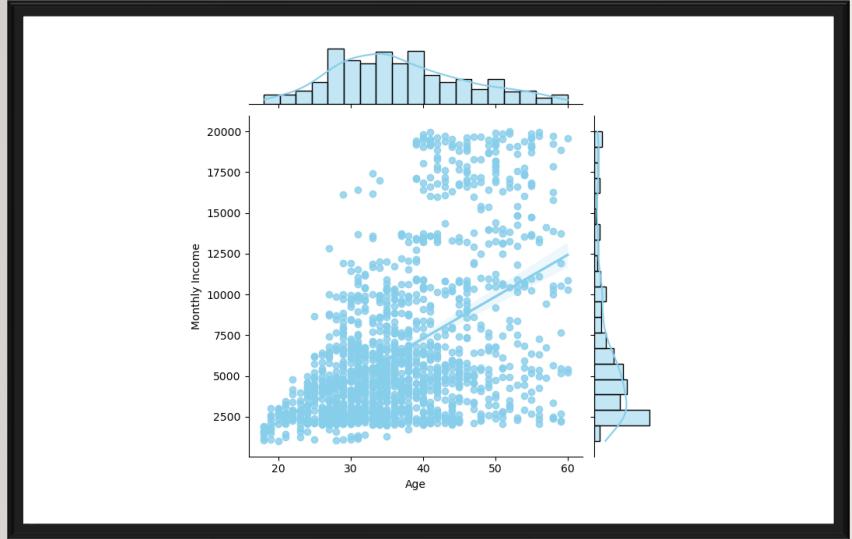


CDF OF MONTHLY INCOME





ANALYTICAL DISTRIBUTION



SCATTER PLOTS OF AGE VERSUS MONTHLY INCOME

TESTING THE HYPOTHESIS

T-Statistic: -6.203935765608938

P-Value: 7.14736398535381e-10

Reject the null hypothesis. There is a significant difference in monthly income between employees who have left the company and those who are still with the company.

REGRESSION ANALYSIS

```
OLS Regression Results
=====
Dep. Variable: MonthlyIncome   R-squared:      0.248
Model:            OLS            Adj. R-squared:  0.247
Method:           Least Squares F-statistic:    483.8
Date:            Mon, 26 Feb 2024 Prob (F-statistic): 6.67e-93
Time:             23:52:07     Log-Likelihood: -14308.
No. Observations: 1470          AIC:            2.862e+04
Df Residuals:    1468          BIC:            2.863e+04
Df Model:         1
Covariance Type: nonrobust
=====
            coef    std err        t    P>|t|    [0.025    0.975]
const     -2970.6712   443.702   -6.695   0.000   -3841.030   -2100.313
Age       256.5716    11.665    21.995   0.000     233.689    279.454
=====
Omnibus:            140.178   Durbin-Watson:    2.069
Prob(Omnibus):      0.000    Jarque-Bera (JB): 182.119
Skew:                0.799    Prob(JB):       2.84e-40
Kurtosis:             3.649    Cond. No.       159.
=====
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

CITATIONS

- Downey, A. B. (n.d.). Think Stats. Retrieved from <https://platform.virdocs.com/read/629508/198/#/4/4/2,/1:0,/1:0>
- Pavansubhash. (2017). IBM HR Analytics Employee Attrition & Performance [Data file and code book]. Retrieved from <https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset>