

# **Day 15**

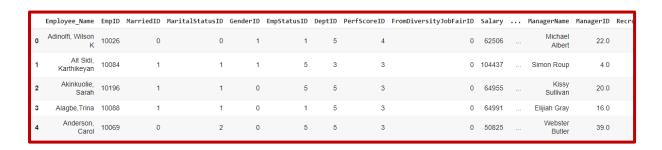
#### DIY

### Q1. Problem Statement: Measures of Dispersion

Load the "HRdataset\_v14.csv" dataset into a DataFrame and group the data based on the "Department" column. From the grouped DataFrame and perform the following tasks:

- 1. Calculate the Mean and Median, of the "Salary" column.
- 2. Find out which Department has the highest number of employees
- 3. Calculate the standard deviation and variance of the "Salary" column
- 4. Find the interquartile range of the "Salary" column
- 5. Find skewness and kurtosis of the original dataset
- 6. Find the z-score of the "Salary" column
- 7. Find the 30th percentile of the "Salary" column
- 8. Find the Quartiles of the "GenderID" column (You can set quartile values of your choice)

#### **Dataset:**



## Sample Output:



1. Calculate the Mean and Median, of the "Salary" column

|   | Department           | Salary        |
|---|----------------------|---------------|
| 0 | Admin Offices        | 71791.888889  |
| 1 | Executive Office     | 250000.000000 |
| 2 | IT/IS                | 97064.640000  |
| 3 | Production           | 59953.545455  |
| 4 | Sales                | 69061.258065  |
| 5 | Software Engineering | 94989.454545  |

| Department |                      | Salary   |
|------------|----------------------|----------|
| 0          | Admin Offices        | 63003.0  |
| 1          | Executive Office     | 250000.0 |
| 2          | IT/IS                | 92328.5  |
| 3          | Production           | 59472.0  |
| 4          | Sales                | 65310.0  |
| 5          | Software Engineering | 95660.0  |

2. Find out which Department has the highest number of employees

This department has highest number of employees 0 Production

3. Calculate the standard deviation and variance of the "Salary" column

Department-wise Standard deviation of salary:\_\_\_\_\_ 73050.21312523098

Department-wise Variance of salary:\_\_\_\_\_ 5336333637.641668



4. Find the interquartile range of the "Salary" column

Interquartile range of salary:\_\_\_\_\_ 31247.375

5. Find skewness and kurtosis of the original dataset.



```
Skewness in Original Dataset:
EmpID
                        0.000000
MarriedID
                       0.415730
MaritalStatusID
                       1.408602
0.267278
GenderID
                       0.631944
EmpStatusID
DeptID
                       -1.536392
PerfScoreID
                        -1.248091
FromDiversityJobFairID 2.811250
                       3.306181
Termd
                       0.705404
PositionID
                       -1.231676
                       4.105494
Zip
ManagerID
                       0.759271
EngagementSurvey
EmpSatisfaction
                       -1.116979
                      -0.222609
SpecialProjectsCount 1.539271
DaysLateLast30
                        3.143468
                        0.029283
Absences
dtype: float64
```

```
Kurtosis in Original Dataset:
                        -1.200000
MarriedID
EmpID
                        -1.839037
MaritalStatusID
                        2.053512
GenderTD
                        -1.941087
EmpStatusID
                       -1.488610
DeptID
                        2.241434
PerfScoreID
                        4.049610
FromDiversityJobFairID
                         5.941296
                      15.452149
Salarv
                       -1.512171
Termd
PositionID
                        0.812346
                      16.187425
Zip
                       1.608422
ManagerID
EngagementSurvey
EmpSatisfaction
                         1.164560
                      -0.762600
SpecialProjectsCount
                        0.641415
DaysLateLast30
                        8.830523
Absences
                        -1.301962
dtype: float64
```

6. Finding the z-score of the "Salary" column

```
array([-0.6192146 , 2.18495176, -0.17945578, -0.67216471, -0.58461933, -0.12949733])
```

7. Finding the 30th percentile of the "Salary" column

```
65th percentile of Salary column is: 65988.0
```

8. Finding Quartiles of the "GenderID" column