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:

	Employee_Name	EmpID	MarriedID	MaritalStatusID	GenderID	EmpStatusID	DeptID	PerfScoreID	FromDiversityJobFairID	Salary	...	ManagerName	ManagerID	Recr
0	Adinolfi, Wilson K	10026	0	0	1	1	5	4	0	62506	...	Michael Albert	22.0	
1	Ait Sidi, Karthikeyan	10084	1	1	1	5	3	3	0	104437	...	Simon Roup	4.0	
2	Akinkuolie, Sarah	10196	1	1	0	5	5	3	0	64955	...	Kissy Sullivan	20.0	
3	Alagbe,Trina	10088	1	1	0	1	5	3	0	64991	...	Elijiah Gray	16.0	
4	Anderson, Carol	10069	0	2	0	5	5	3	0	50825	...	Webster Butler	39.0	

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
GenderID                   0.267278
EmpStatusID                0.631944
DeptID                    -1.536392
PerfScoreID               -1.248091
FromDiversityJobFairID     2.811250
Salary                     3.306181
Termd                      0.705404
PositionID                -1.231676
Zip                        4.105494
ManagerID                  0.759271
EngagementSurvey          -1.116979
EmpSatisfaction           -0.222609
SpecialProjectsCount       1.539271
DaysLateLast30             3.143468
Absences                   0.029283
dtype: float64
```

```
Kurtosis in Original Dataset: _____
EmpID                     -1.200000
MarriedID                 -1.839037
MaritalStatusID            2.053512
GenderID                  -1.941087
EmpStatusID               -1.488610
DeptID                     2.241434
PerfScoreID                4.049610
FromDiversityJobFairID     5.941296
Salary                    15.452149
Termd                     -1.512171
PositionID                 0.812346
Zip                       16.187425
ManagerID                  1.608422
EngagementSurvey           1.164560
EmpSatisfaction           -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

```
Quartile values of 20, 40 from GenderID column:
[40 40 20 20 20 20 20 40 20 40 20 40 40 40 40 40 20 20 20 40 40 20 20 20
 20 40 20 40 20 20 40 20 20 20 40 40 40 20 40 20 20 20 40 20 40 40 20 20
 40 20 40 40 40 40 40 40 20 40 20 40 20 20 20 40 40 40 20 40 20 40 40 40
 20 20 20 20 40 20 20 20 20 40 40 20 40 20 20 20 20 20 40 40 20 40 20 40
 40 20 40 20 40 20 20 20 20 40 20 20 20 40 20 40 40 20 20 40 20 20 40 20
 40 40 20 40 20 20 40 20 40 20 20 20 20 20 20 20 20 20 40 20 20 20 20 20 40
 20 20 20 20 40 40 20 20 20 40 40 40 40 40 20 40 40 40 20 20 40 40 40 20
 20 40 20 20 20 20 20 40 20 20 20 20 20 20 20 20 20 40 20 20 40 40 40 40 20
 40 20 20 40 40 40 40 40 40 20 20 20 20 20 20 20 20 40 40 40 20 40 40 40 20
 20 20 20 20 20 40 40 20 20 40 20 40 20 40 40 40 20 20 40 40 20 40 20 20
 20 20 40 40 40 20 20 40 40 40 20 20 20 20 40 20 20 40 40 20 20 20 40 20
 20 20 20 40 40 40 20 20 40 20 40 40 20 20 20 20 40 20 40 20 40 40 20 40
 20 40 40 40 40 20 20 40 20 20 40 20 40 40 20 20 40 20 40 20 20 20 20]
```