

HR Analytics

Every organization wishes to improve its gains and achieve higher standards in their respective fields. This requires a constant improvement in the infrastructure and the capacity of the people working for the organization. The field of HR analytics involves the analytical process and its implementation to improve the performance of the employees as well as to retain them for a longer period. This method does not focus on the individuals rather it focuses on the processes and helps understand, how to improve a process which will in turn improve the performance.

HR analysis can help understand questions like the following.

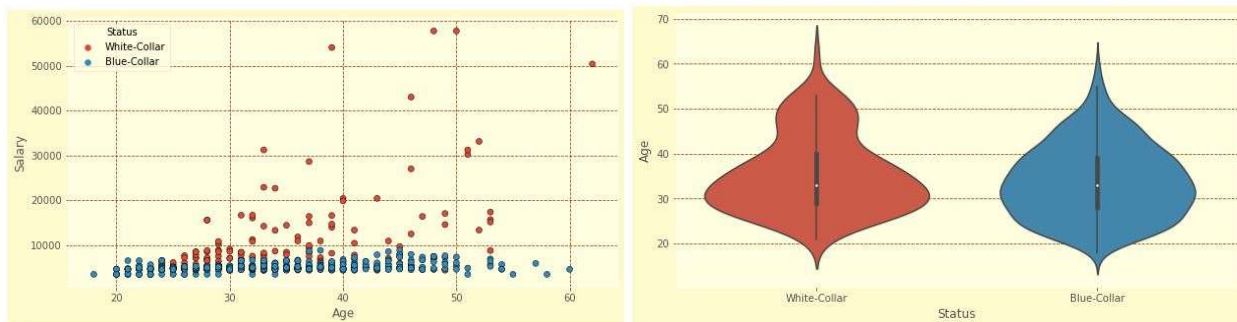
- Employee's turnover rate.
- Which employees are likely to leave within a year?

HR usually gathers a huge amount of data but is unaware on how to use it to get the desirable answers.

This data set is a randomly generated data set gathered from different organizations with a total of 601 entries. First column represents the serial number, second column represents the gender of the employee, third column shows the age, fourth column shows the seniority level of the employee, fifth column represents the region/branch an employee is working at, sixth column shows the department of the employee, seventh columns shows the status of the jobs whether it's a white collar or a blue collar job, eighth column shows the marital status of the employee, ninth columns shows the education level of the employee, tenth column represents the disability status of the employee and eleventh columns mentions their salary.

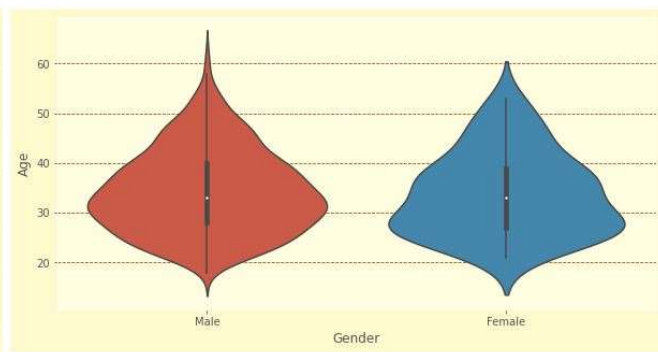
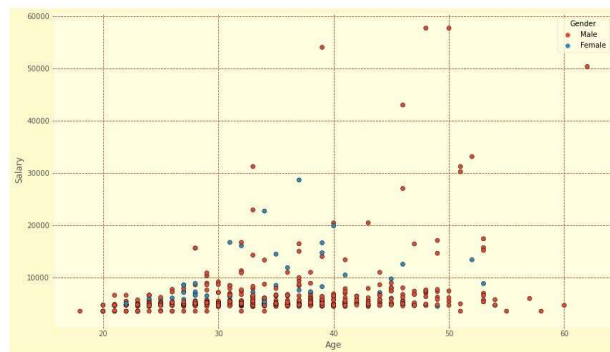
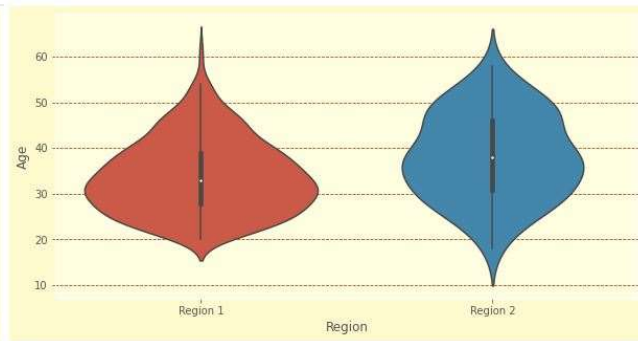
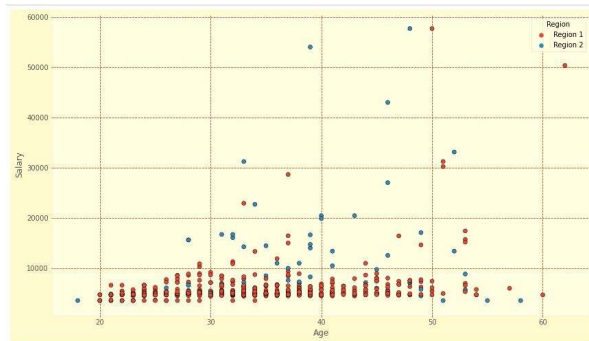
	count	mean	std	min	25%	50%	75%	max	Age Range	Count
Age	601.0	34.229617	8.304247	18.0	28.00	33.0	40.00	62.00	0 18-20	7
Senior	601.0	4.043261	4.423964	0.0	1.00	3.0	5.00	23.00	1 21-35	349
Salary	601.0	6553.212446	5569.475835	3497.1	4663.96	5001.8	6186.18	57722.32	2 36-45	177
									3 46-65	68

These are some descriptive statistics of the data set. There are a total of 601 entries, the minimum age of an employee is 18 and the maximum is 62 with a mean of approx. 34, Senior means the seniority level of the employee with a minimum of 0 and max of 23 with a mean of 4, salary has the minimum value of 3497.1 and goes up to 57722.32 with a mean salary of approx. 6553.

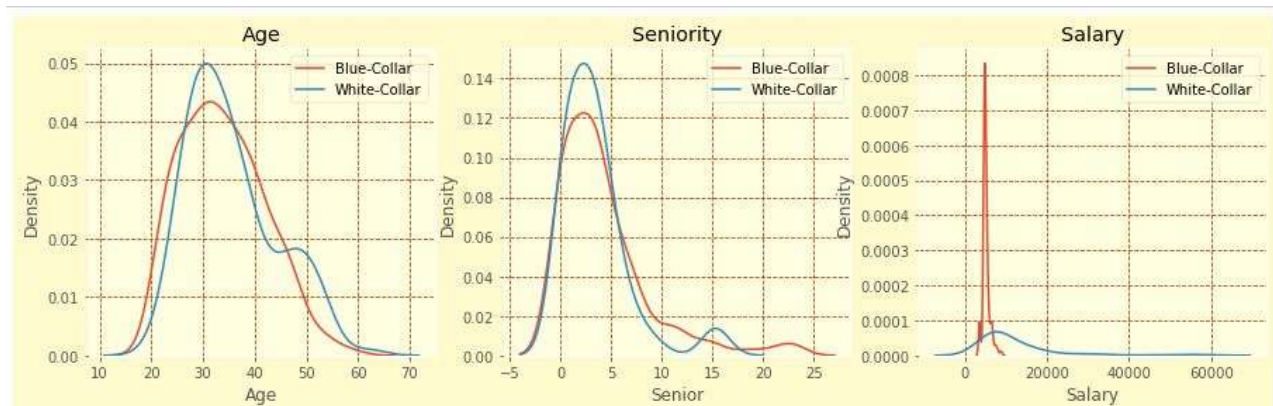


The scatter plot shows the comparison between age and salary for blue and white collard jobs, nearly all of the blue collard jobs are in the lowest salary range whereas for white collard job holders there are a few in higher salary range with some outliers. The smooth distribution curves of blue-collar job holders vs age show that there isn't much gap between the age ranges of the employees whereas for white collar jobs, most employees are between 25 to 40 years old.

Below scatter plot between age and salary shows classified between two regions of the company represents that most employees are below 10k for region 1 and 2 but a significant number lies between 20k to 30k for region 2 as compared to region 1. The distribution shows that for region 2 the employees are somewhat evenly distributed for age range of 35 to 50 but for region 1 most employees are between 25 to 40.



This scatter plot classification for gender shows that nearly the number of males and females is the same for salary under 30k but over this salary limit there are no females.



This kernel density estimation (KDE) plot shows the distribution of observations in this dataset and we can see that the density for age and seniority are nearly equal between blue and white collar but vary hugely for salary.

The bar plot shows that for education level 1 to 4 most employees hold a blue collar job and there are a very few number of employees in this range. But for education level 6 and beyond we can see that most employees have a white collar job.

