HR Project

Exploratory Data Analysis (EDA) of data

a. Missing values

As, it can be seen in the figure below, the provided data has no missing values, since the count of all the columns is the same.

	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	promotion_last_5years	department	salary
count	14999	14999	14999	14999	14999	14999	14999	14999	14999
unique	92	65	6	215	8	2	2	10	3
top	0.1	0.55	4	156	3	0	0	sales	low
freq	358	358	4365	153	6443	12830	14680	4140	7316

Stats of data after converting into float type:

	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	promotion_last_5years	department	salar
count	14999.000000	14999.000000	14999.000000	14999.000000	14999.000000	14999.000000	14999.000000	14999.000000	14999.00000
mean	0.612834	0.716102	3.803054	201.050337	3.498233	0.144610	0.021268	3.339823	0.5947(
std	0.248631	0.171169	1.232592	49.943099	1.460136	0.351719	0.144281	2.820837	0.63718
min	0.090000	0.360000	2.000000	96.000000	2.000000	0.000000	0.000000	0.000000	0.00000
25%	0.440000	0.560000	3.000000	156.000000	3.000000	0.000000	0.000000	0.000000	0.00000
50%	0.640000	0.720000	4.000000	200.000000	3.000000	0.000000	0.000000	3.000000	1.00000
75%	0.820000	0.870000	5.000000	245.000000	4.000000	0.000000	0.000000	6.000000	1.00000
max	1.000000	1.000000	7.000000	310.000000	10.000000	1.000000	1.000000	9.000000	2.00000

Stats of employees who stayed:

	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	promotion_last_5years	department	sala
count	11428.000000	11428.000000	11428.000000	11428.000000	11428.000000	11428.000000	11428.000000	11428.000000	11428.0000
mean	0.666810	0.715473	3.786664	199.060203	3.380032	0.175009	0.026251	3.408908	0.6509
std	0.217104	0.162005	0.979884	45.682731	1.562348	0.379991	0.159889	2.853289	0.6552
min	0.120000	0.360000	2.000000	96.000000	2.000000	0.000000	0.000000	0.000000	0.0000
25%	0.540000	0.580000	3.000000	162.000000	2.000000	0.000000	0.000000	0.000000	0.0000
50%	0.690000	0.710000	4.000000	198.000000	3.000000	0.000000	0.000000	3.000000	1.0000
75%	0.840000	0.850000	4.000000	238.000000	4.000000	0.000000	0.000000	6.000000	1.0000
max	1.000000	1.000000	6.000000	287.000000	10.000000	1.000000	1.000000	9.000000	2.0000

Stat of employees who left:

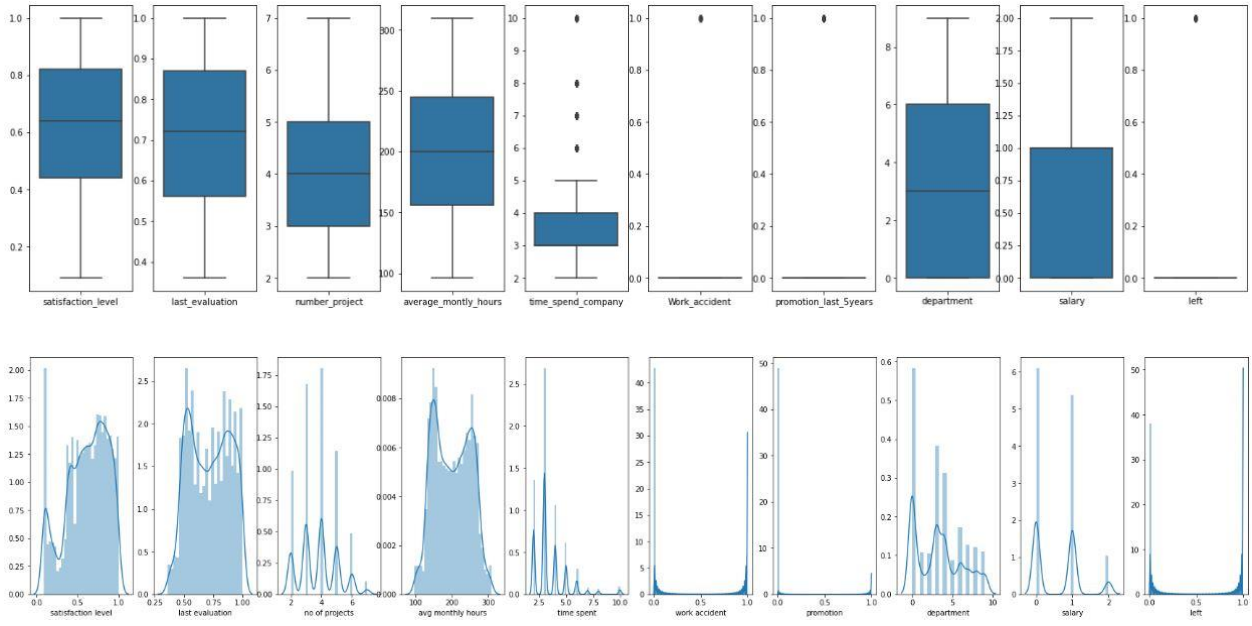
	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	promotion_last_5years	department	salary
count	3571.000000	3571.000000	3571.000000	3571.000000	3571.000000	3571.000000	3571.000000	3571.000000	3571.000000
mean	0.440098	0.718113	3.855503	207.419210	3.876505	0.047326	0.005321	3.118734	0.414730
std	0.263933	0.197673	1.818165	61.202825	0.977698	0.212364	0.072759	2.702922	0.537341
min	0.090000	0.450000	2.000000	126.000000	2.000000	0.000000	0.000000	0.000000	0.000000
25%	0.130000	0.520000	2.000000	146.000000	3.000000	0.000000	0.000000	0.000000	0.000000
50%	0.410000	0.790000	4.000000	224.000000	4.000000	0.000000	0.000000	3.000000	0.000000
75%	0.730000	0.900000	6.000000	262.000000	5.000000	0.000000	0.000000	4.000000	1.000000
max	0.920000	1.000000	7.000000	310.000000	6.000000	1.000000	1.000000	9.000000	2.000000

From the above figures, we can see the **clear difference** between mean of attributes like **salary and satisfaction level**, which are discussed in detail later.

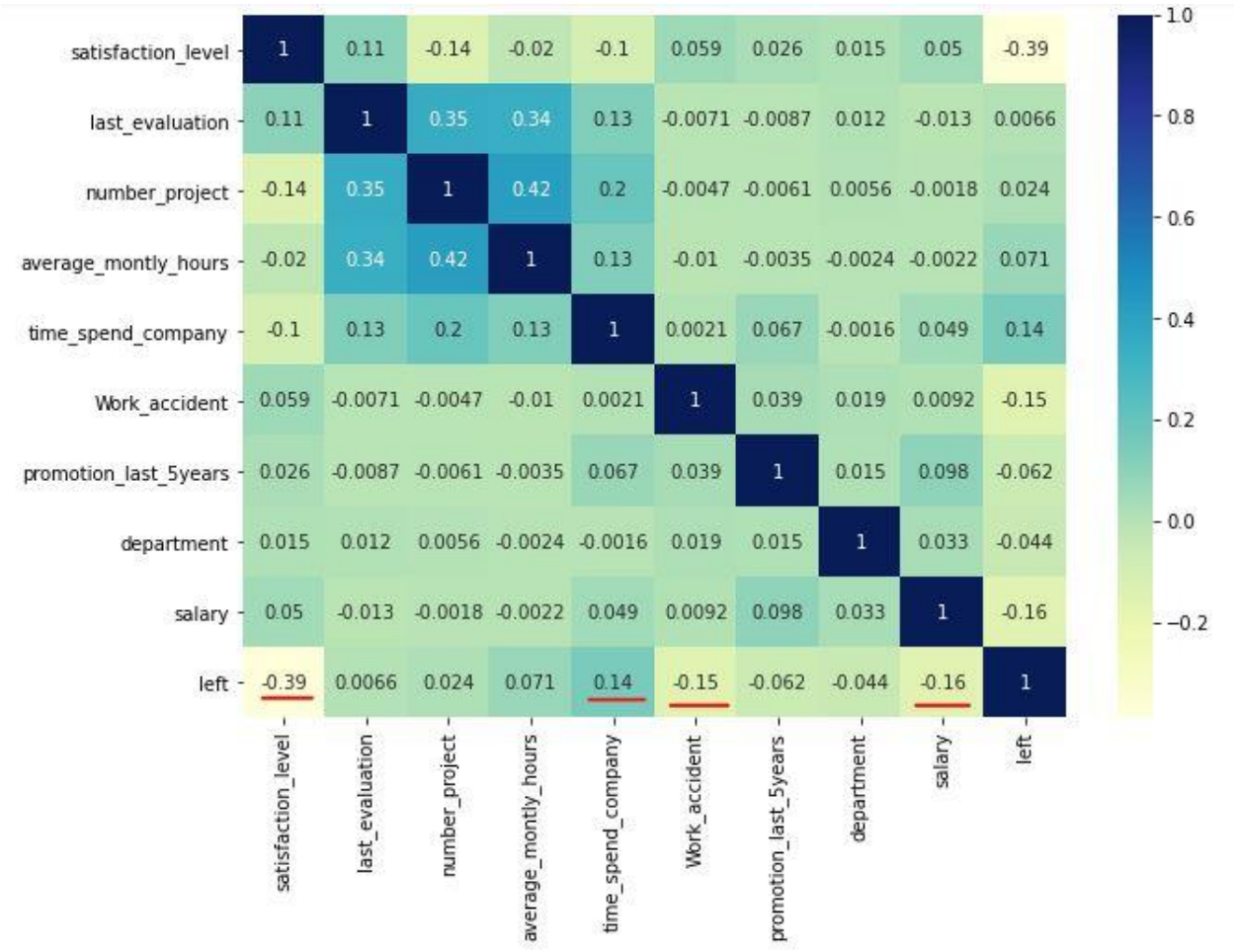
b. Visualisation of statistics and summary

Boxplot and Kernel density estimation (KDE):

From the below graphs, we can visualise the data and its attributes' mean, spread, variance and outliers.
It is visible that the majority of the employees are working from 3 to 4 years and employees with 6 or more years' experience are outliers.
We can also see that very few people had work accidents and very few employees had promotions in the last 5 years.



c. Correlation



The 'left' feature in the data has correlation of -0.39, -0.16, -0.15, +0.14 with satisfaction level, salary, work accident and time spent in company (years) respectively, with **satisfaction level** having **highest correlation**, meaning employee **satisfaction had great role** in employees' decision whether to leave the company or not.

d. Employee Retention

In order to find the attributes with high influence on employee's retention, we first observed logistic regression weights for different attributes for retention class, from which three attributes were shortlisted, which are:

- Satisfaction level (of employee with company)
- Last evaluation (score given by company to employee)
- Average monthly hours

Regression Feature Weights

Retention:

3.63081044	**satisfaction_level**
3.18155641	**last_evaluation**
2.04684737	number_project
2.75860625	**average_montly_hours**
0.98812414	time_spend_company
1.00247202	work_accident
0.1503708	promotion_last_5years
2.16962791	department
1.86434734	salary
5.72812512	bias

On separating the employee data into retained and left groups and finding its statistics, it was noticed that there was a big difference in mean of satisfaction level between both groups, with **retained** employees having **0.666 mean** and employees who **left** having **0.444**, as shown below.

	satisfaction_level
count	3571.000000
mean	0.440098
std	0.263933
min	0.090000
25%	0.130000
50%	0.410000
75%	0.730000
max	0.920000

left

	satisfaction_level
count	11428.000000
mean	0.666810
std	0.217104
min	0.120000
25%	0.540000
50%	0.690000
75%	0.840000
max	1.000000

retained

Mean of ‘avg month hours’ of employees who left, had a difference of 8 hours than who did not left, meaning they worked slightly more as compared to other group. And since stats of attribute ‘last evaluation’ were similar across both groups, we can ignore them.

average_montly_hours	average_montly_hours	last_evaluation	last_evaluation
11428.000000	3571.000000	11428.000000	3571.000000
199.060203	207.419210	0.715473	0.718113
45.682731	61.202825	0.162005	0.197673
96.000000	126.000000	0.360000	0.450000
162.000000	146.000000	0.580000	0.520000
198.000000	224.000000	0.710000	0.790000
238.000000	262.000000	0.850000	0.900000
287.000000	310.000000	1.000000	1.000000
Left	**Retained**	**Left**	**Retained**