CSE - 5334 - 001 : Data Mining

Assignment 1 - Exploratory Data Analysis - Weka

Instructor: Dr. Elizabeth D Diaz

Team 18: Urmi Manish Sheth – 1002064934

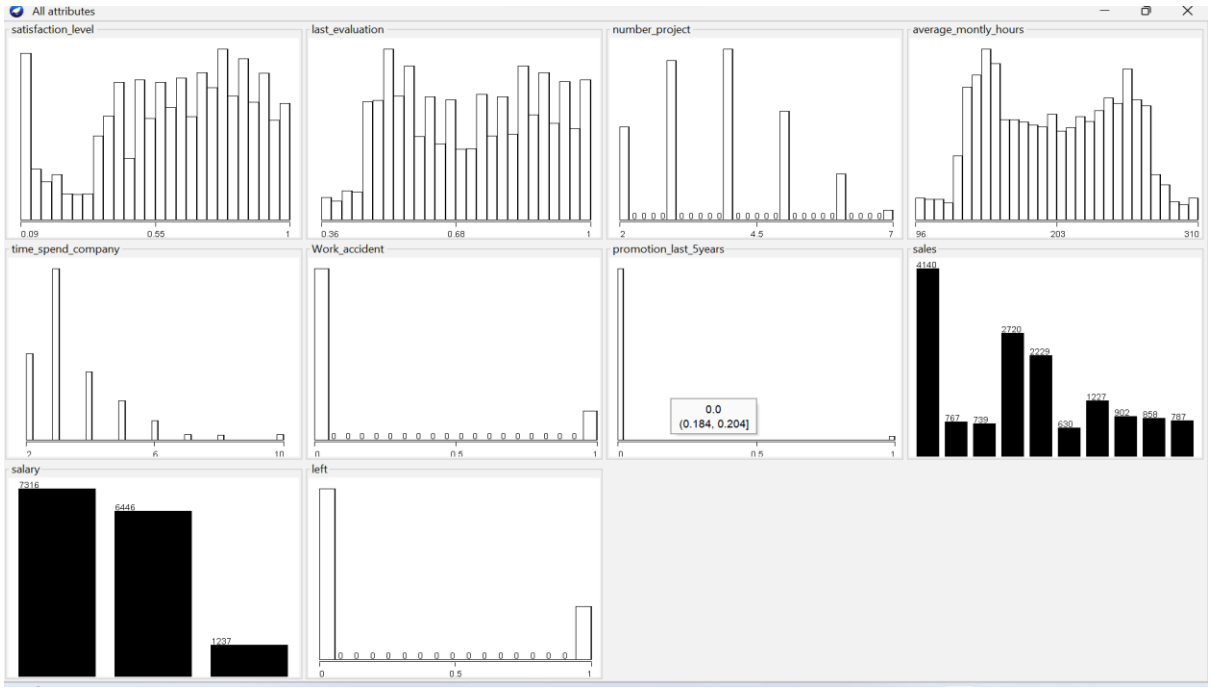
Harsh Navinbhai Shah – 1002057387

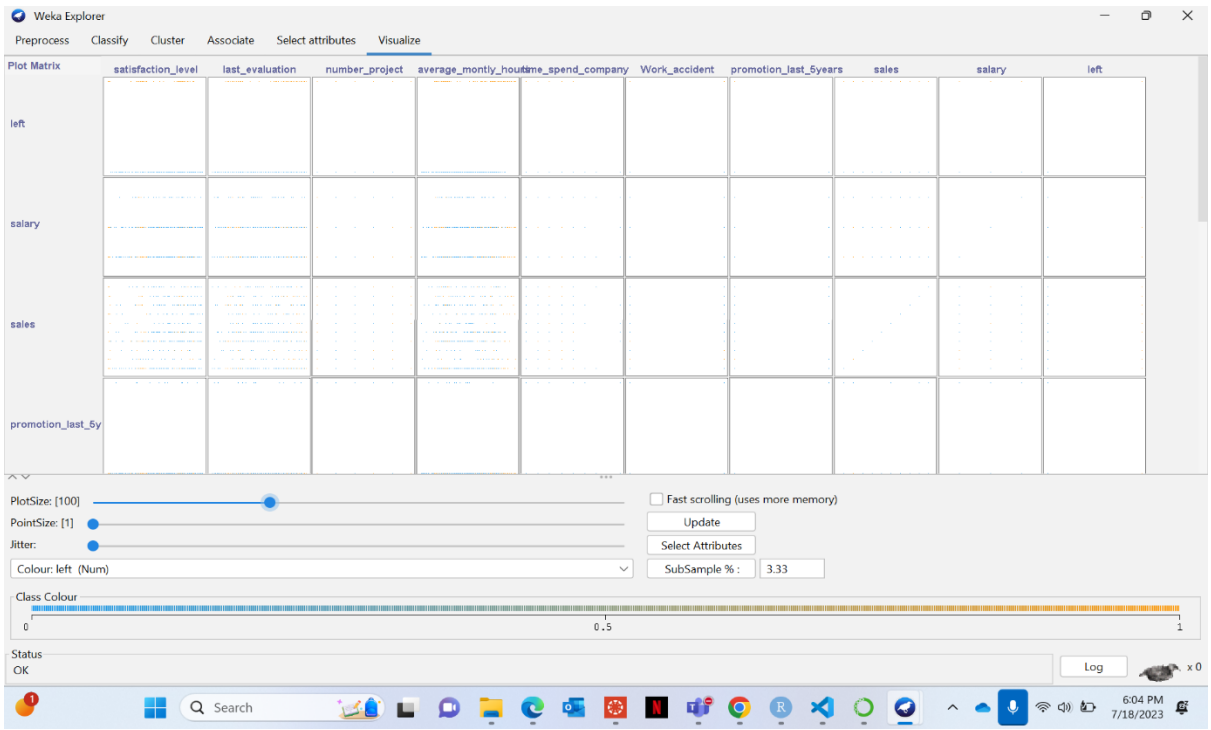
Sai Swetha Tadivaka – 1002112726

Part 3: Weka (20 Points) For this task, use the "Employee_retention.csv" data file in Weka. Perform the following analyses:

Task 3A (3 points): Display a visualization for each column in the dataset.

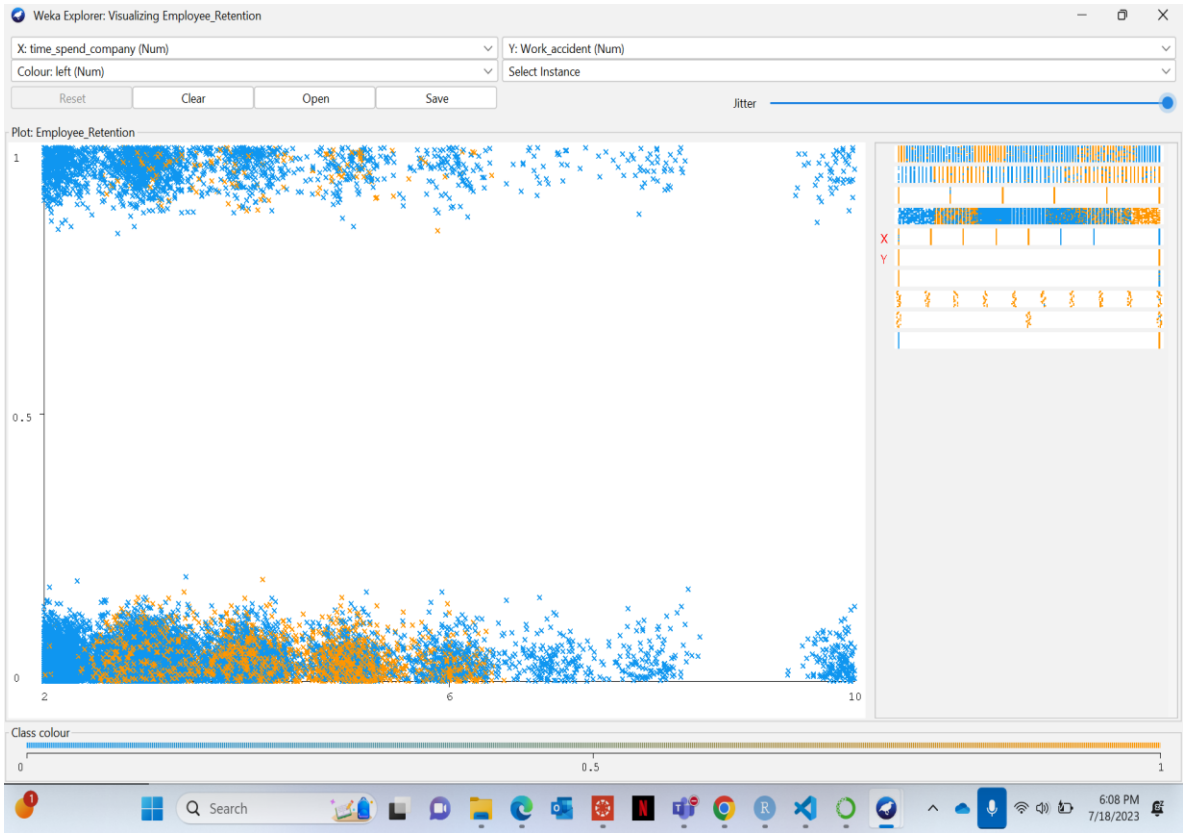
Output





Task 3B (3 points): Display the time spent by employees in the company vs. the occurrence of work accidents. Interpret the graph and provide your observations.

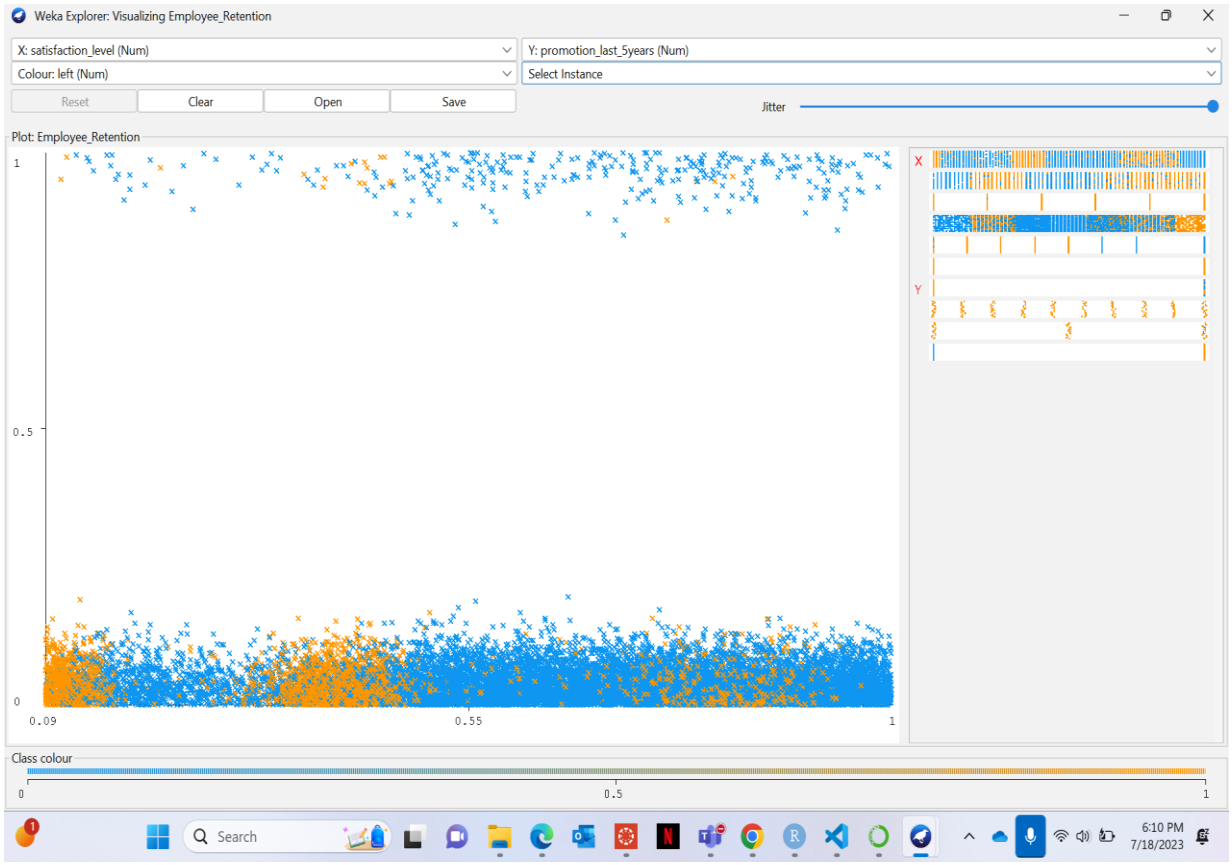
Output



By analyzing, we can observe the relationship between the ratio of workers and work accidents. The frequency of workplace accidents rises as the number of employees grows. It is also possible to find that the two variables have no obvious relationship or even a negative association.

Task 3C (3 points): Display the job satisfaction level of employees vs. whether they received a promotion within the last 5 years. Interpret the graph and provide your observations.

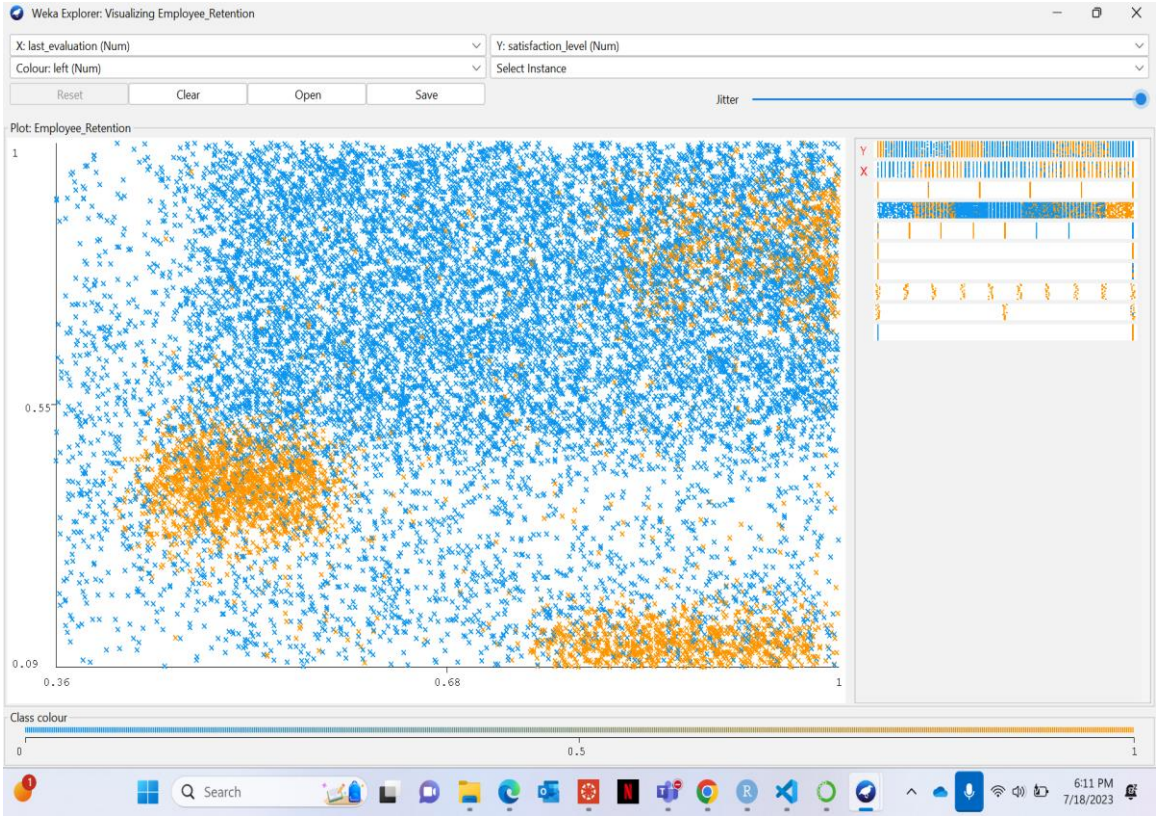
Output



The image above shows the relation between employee's promotion based on their satisfaction level. By examining the above case, we can observe that the employee whose satisfaction level is high are more likely to get promotion early compared to the employee who have a lower satisfaction level.

Task 3D (3 points): Display the last evaluation score of employees vs. their job satisfaction level. Interpret the graph and provide your observations.

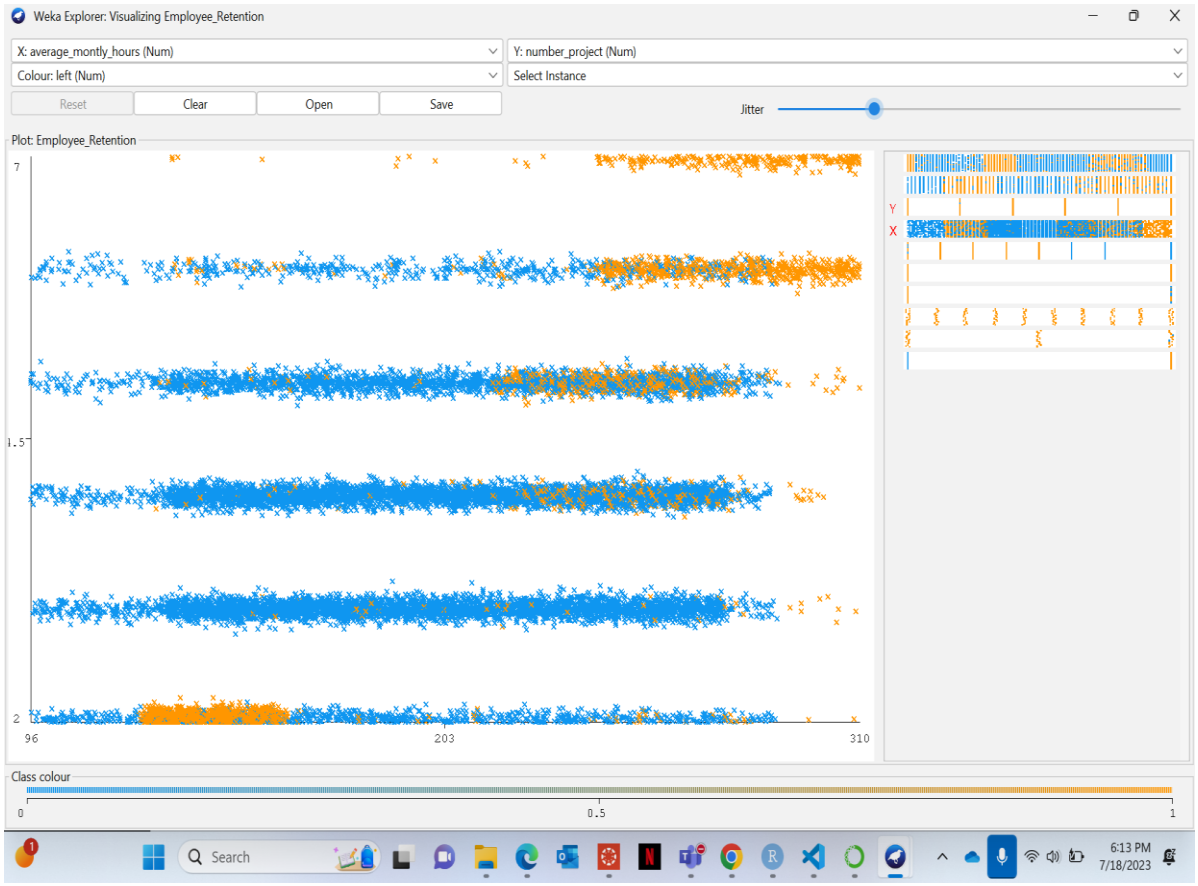
Output



We can see the relationship between the most recent evaluation and the work satisfaction level by looking at the graphic. When a worker's most recent evaluation score was higher, they tended to be more content with their positions. The two variables don't correlate or that there are cutoff points at which levels of work satisfaction fluctuate in connection to evaluation results.

Task 3E (3 points): Display the average monthly working hours for employees vs. the number of projects they are working on. Interpret the graph and provide your observations.

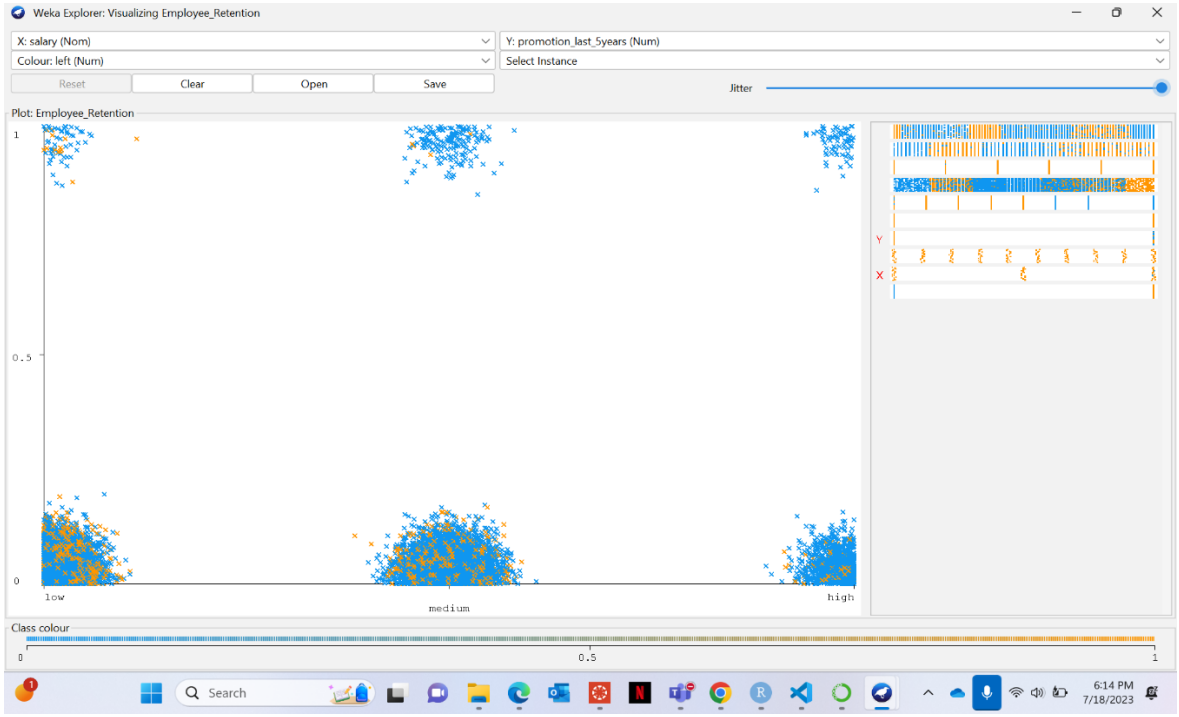
Output

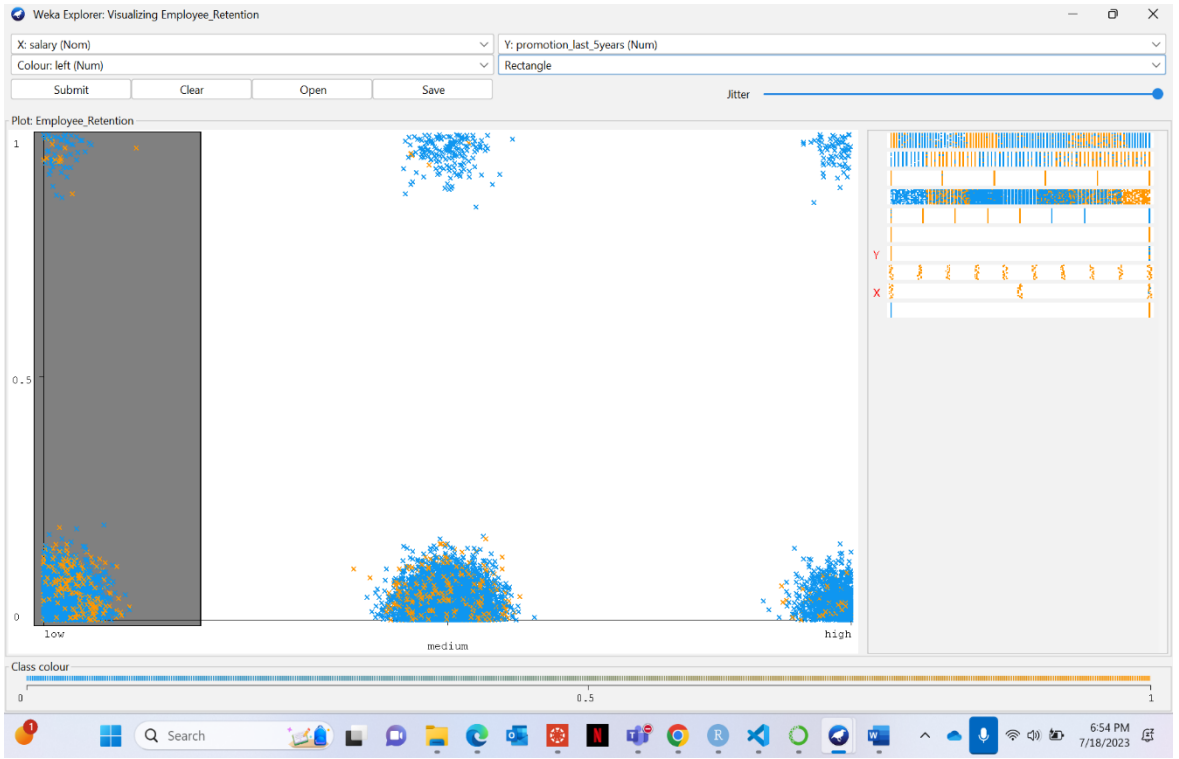


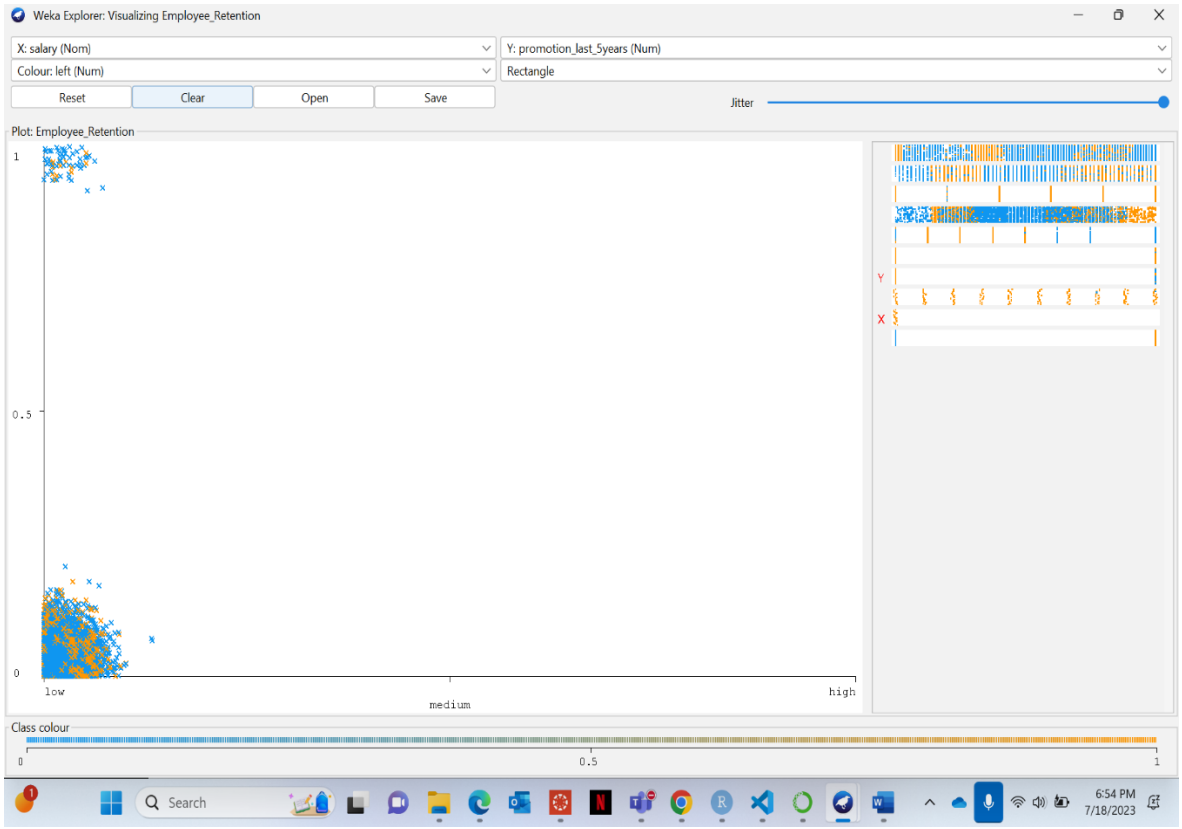
We can observe how the number of projects and the average monthly working hours are related by looking at the visualization. For example, we may notice that workers spend more hours per month on average when they are working on a bigger number of projects. There is no relevant relationship between the two variables as average monthly working hours vary in proportion to the number of projects.

Task 3F (5 points): Display the correlation between low salary levels and the occurrence of promotions within the last 5 years. Interpret the graph and provide your observations.

Output







We can see how the low salary category is related to the occurrence of promotions by inspecting the graphic. Employees with low salary may be less likely than those with higher salaries to have gotten promotions in the prior five years.

References:

Microsoft Word - WEKA Explorer Tutorial.doc (sabanciuniv.edu)