

A decorative graphic on the left side of the slide, consisting of a network of white lines and small circles on a blue gradient background, resembling a circuit board or a neural network.

# PROOF OF CONCEPT

BY PRAKIRTI LAKHOTIYA

# PROBLEM STATEMENT

The data is for company X which is trying to control attrition. There are two sets of data: "Existing employees" and "Employees who have left". Following attributes are available for every employee.

- Satisfaction Level
- Last evaluation
- Number of projects
- Average monthly hours
- Time spent at the company
- Whether they have had a work accident
- Whether they have had a promotion in the last 5 years
- Departments (column sales)
- Salary
- Whether the employee has left

## Objective

- What type of employees are leaving? Determine which employees are prone to leave next. Present your results in the presentation sheet's presentation area.

# BASIC REFORMATTING

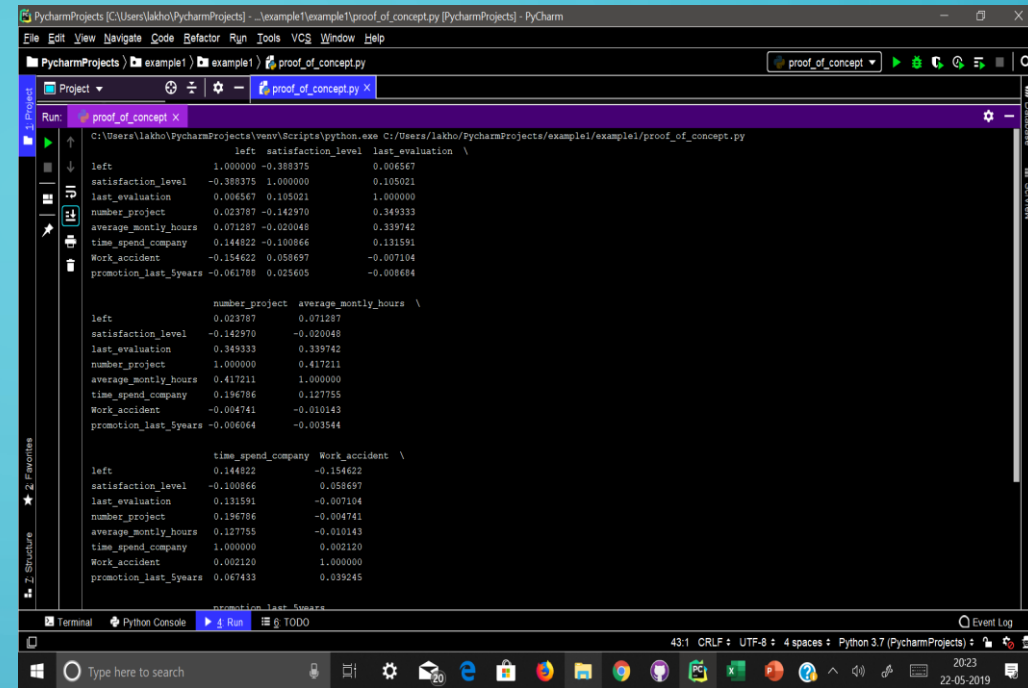
- At first, the data is splitted into two different sheets.
- They have Emp ID as one of the attribute.
- For analysing things, this column is not at all required.
- So we drop the column Emp ID .
- Also, we add a left column to both sheets with left value 1 in 'existing employee sheet' and left value 0 in 'employees who left' sheet.
- After that we merge both the sheets and make them a single dataframe.

# CHECKING NULL VALUES

- Now, we will check to see if any null values are present in the dataset.
- We will use `dataframe.isnull().any()` functions which will return true in front of their column name if null value is present in that column.

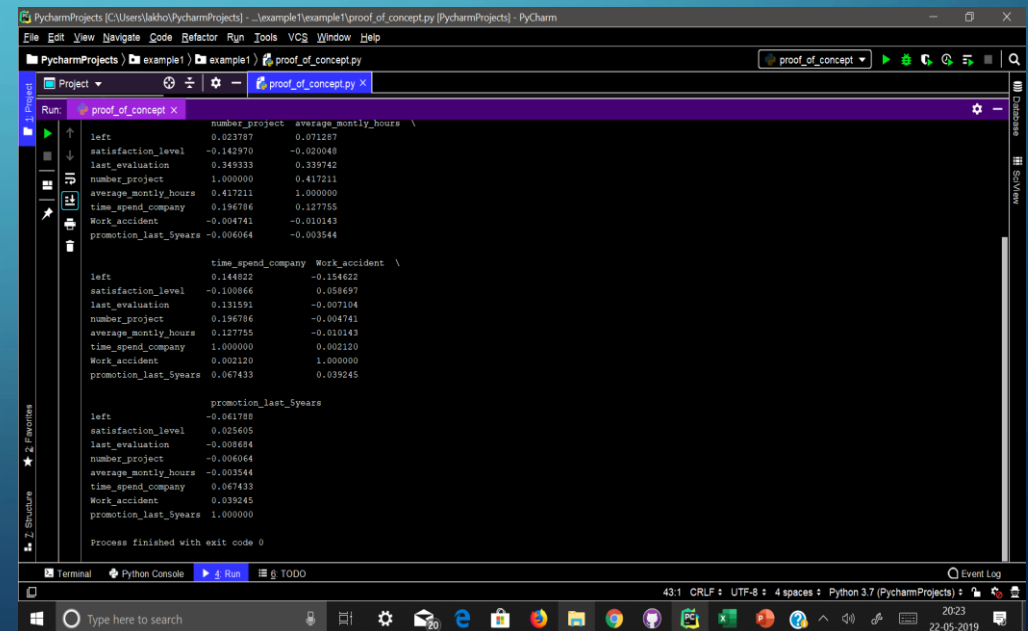
# CORRELATION TABLE

- At first, we will get a correlation matrix using `dataframe.corr()`
- This will give positive correlation between `projectCount`, `averageMonthlyHours`, and `evaluation`. Which could mean that the employees who spent more hours and did more projects were evaluated highly.
- This also gives negative correlation between `turnover` and `satisfaction`



The screenshot shows the PyCharm IDE with a Python script named `proof_of_concept.py` open. The script has executed, and the output is displayed in the Run console. The output is a correlation matrix for the following variables: `left`, `satisfaction_level`, `last_evaluation`, `number_project`, `average_monthly_hours`, `time_spent_company`, `Work_accident`, and `promotion_last_5years`. The matrix is symmetric, with the diagonal elements all equal to 1.000000. The correlations between variables are as follows:

	left	satisfaction_level	last_evaluation	number_project	average_monthly_hours	time_spent_company	Work_accident	promotion_last_5years
left	1.000000	-0.388375	0.005567	0.023787	0.071287	0.144822	-0.154622	-0.061708
satisfaction_level	-0.388375	1.000000	0.105021	-0.142970	-0.020048	-0.100866	0.058697	0.025605
last_evaluation	0.005567	0.105021	1.000000	0.349333	0.339742	0.131591	-0.007104	0.006064
number_project	0.023787	-0.142970	0.349333	1.000000	0.417211	0.196786	-0.004741	-0.006064
average_monthly_hours	0.071287	-0.020048	0.339742	0.417211	1.000000	0.127755	-0.010143	0.039245
time_spent_company	0.144822	-0.100866	0.131591	0.196786	0.127755	1.000000	0.002120	0.067433
Work_accident	-0.154622	0.058697	-0.007104	-0.004741	-0.010143	0.002120	1.000000	0.067433
promotion_last_5years	-0.061708	0.025605	0.006064	-0.006064	0.039245	0.067433	0.067433	1.000000



This screenshot shows the same PyCharm IDE with the same Python script `proof_of_concept.py`. The output in the Run console is the same correlation matrix as in the previous screenshot, but it is formatted differently, with the variables listed in a different order. The correlations are identical to the first screenshot.

	left	satisfaction_level	last_evaluation	number_project	average_monthly_hours	time_spent_company	Work_accident	promotion_last_5years
left	1.000000	-0.388375	0.005567	0.023787	0.071287	0.144822	-0.154622	-0.061708
satisfaction_level	-0.388375	1.000000	0.105021	-0.142970	-0.020048	-0.100866	0.058697	0.025605
last_evaluation	0.005567	0.105021	1.000000	0.349333	0.339742	0.131591	-0.007104	0.006064
number_project	0.023787	-0.142970	0.349333	1.000000	0.417211	0.196786	-0.004741	-0.006064
average_monthly_hours	0.071287	-0.020048	0.339742	0.417211	1.000000	0.127755	-0.010143	0.039245
time_spent_company	0.144822	-0.100866	0.131591	0.196786	0.127755	1.000000	0.002120	0.067433
Work_accident	-0.154622	0.058697	-0.007104	-0.004741	-0.010143	0.002120	1.000000	0.067433
promotion_last_5years	-0.061708	0.025605	0.006064	-0.006064	0.039245	0.067433	0.067433	1.000000

# PLOTTING HEATMAP OF CORRELATION MATRIX

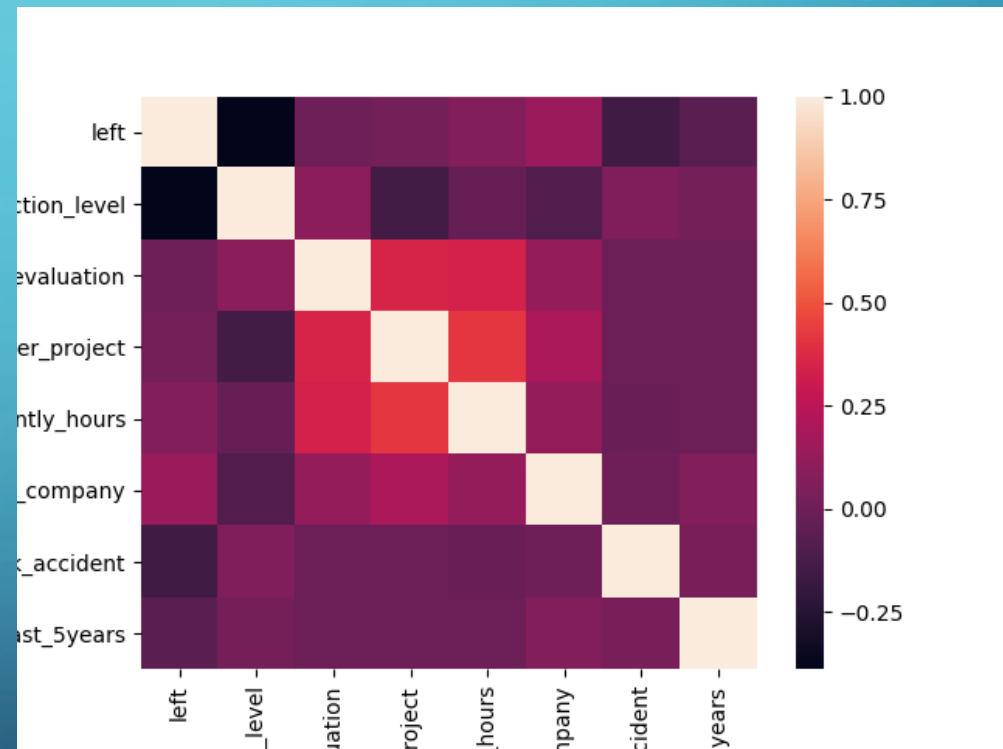
Positive correlation:  
projectCount vs evaluation: 0.349333  
projectCount vs averageMonthlyHours: 0.417211  
averageMonthlyHours vs evaluation: 0.339742

Negative correlation:  
satisfaction vs turnover: -0.388375

The heatmap displays the correlation matrix for eight variables: left, action\_level, evaluation, er\_project, ntly\_hours, \_company, k\_accident, and ast\_5years. The color scale ranges from -0.25 (dark purple) to 1.00 (light yellow). The diagonal elements are all 1.00. The strongest positive correlations are between projectCount and evaluation (0.349333), projectCount and averageMonthlyHours (0.417211), and averageMonthlyHours and evaluation (0.339742). The strongest negative correlation is between satisfaction and turnover (-0.388375).

projectCount vs evaluation: 0.349333  
projectCount vs averageMonthlyHours: 0.417211  
averageMonthlyHours vs evaluation: 0.339742

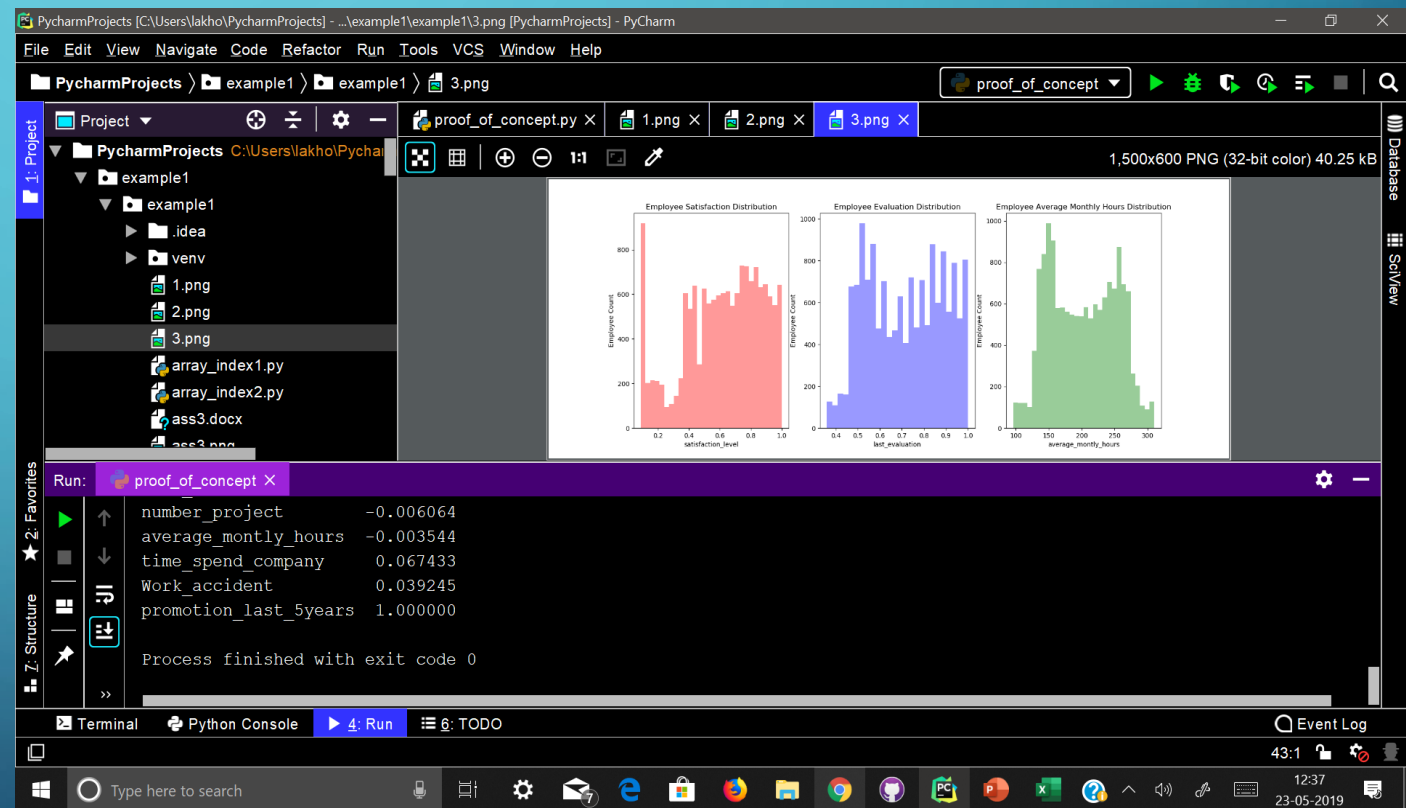
satisfaction vs turnover: -0.388375



# EXAMINING DISTRIBUTION ON SOME FEATURES

## PLOTTING DISTPLOTS

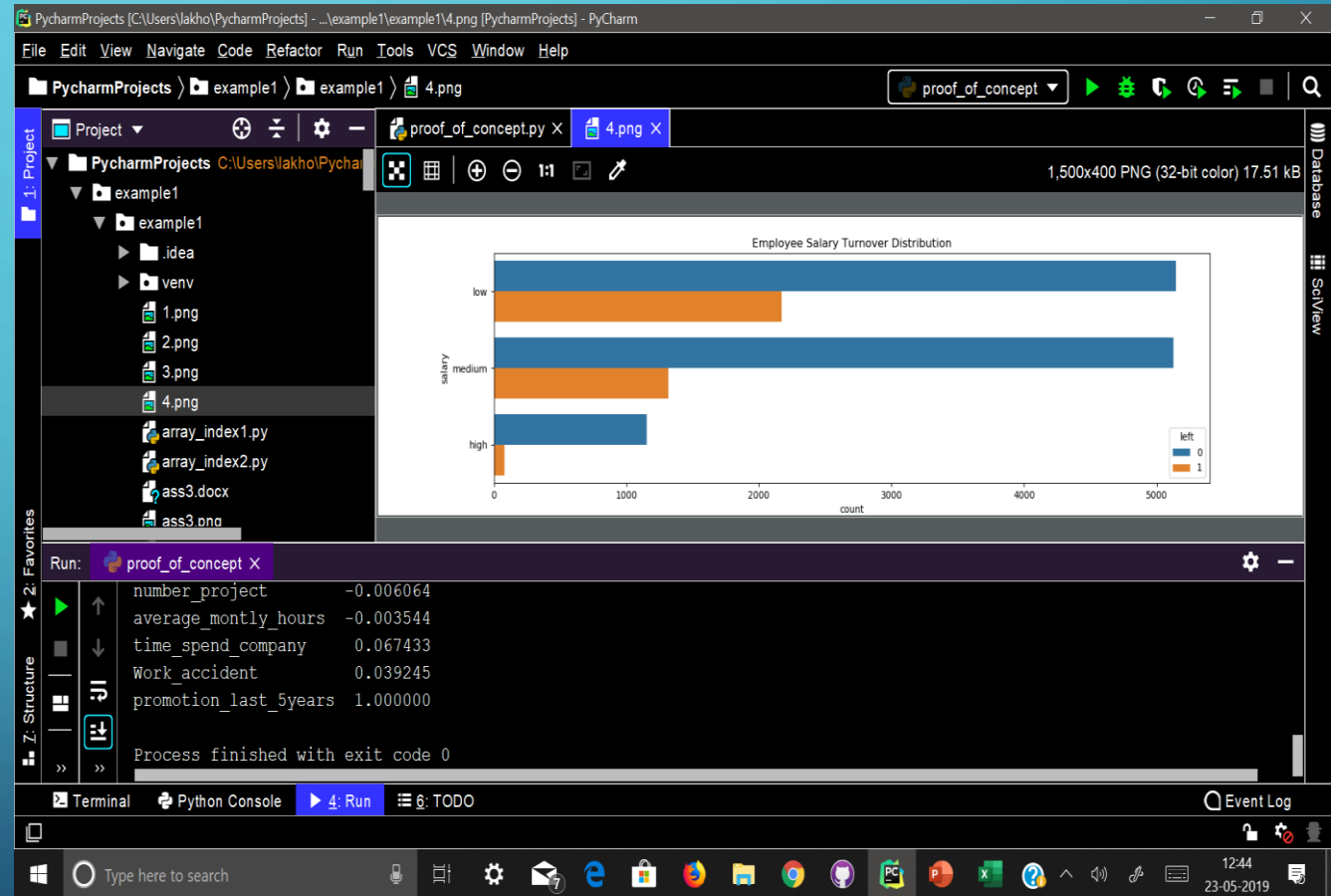
- Satisfaction – A large number of employees are in low and high satisfaction.
- Evaluation - There is a bimodal distribution of employees for low evaluations (less than 0.6) and high evaluations (more than 0.8)
- Average Monthly Hours - There is another bimodal distribution of employees with lower and higher average monthly hours (less than 150 hours & more than 250 hours)
- The evaluation and average monthly hour graphs both share a similar distribution.
- Employees with lower average monthly hours were evaluated less and vice versa.
- If you look back at the correlation matrix, the high correlation between evaluation and average Monthly Hours does support this finding.





# SALARY VS LEFT EMPLOYEES

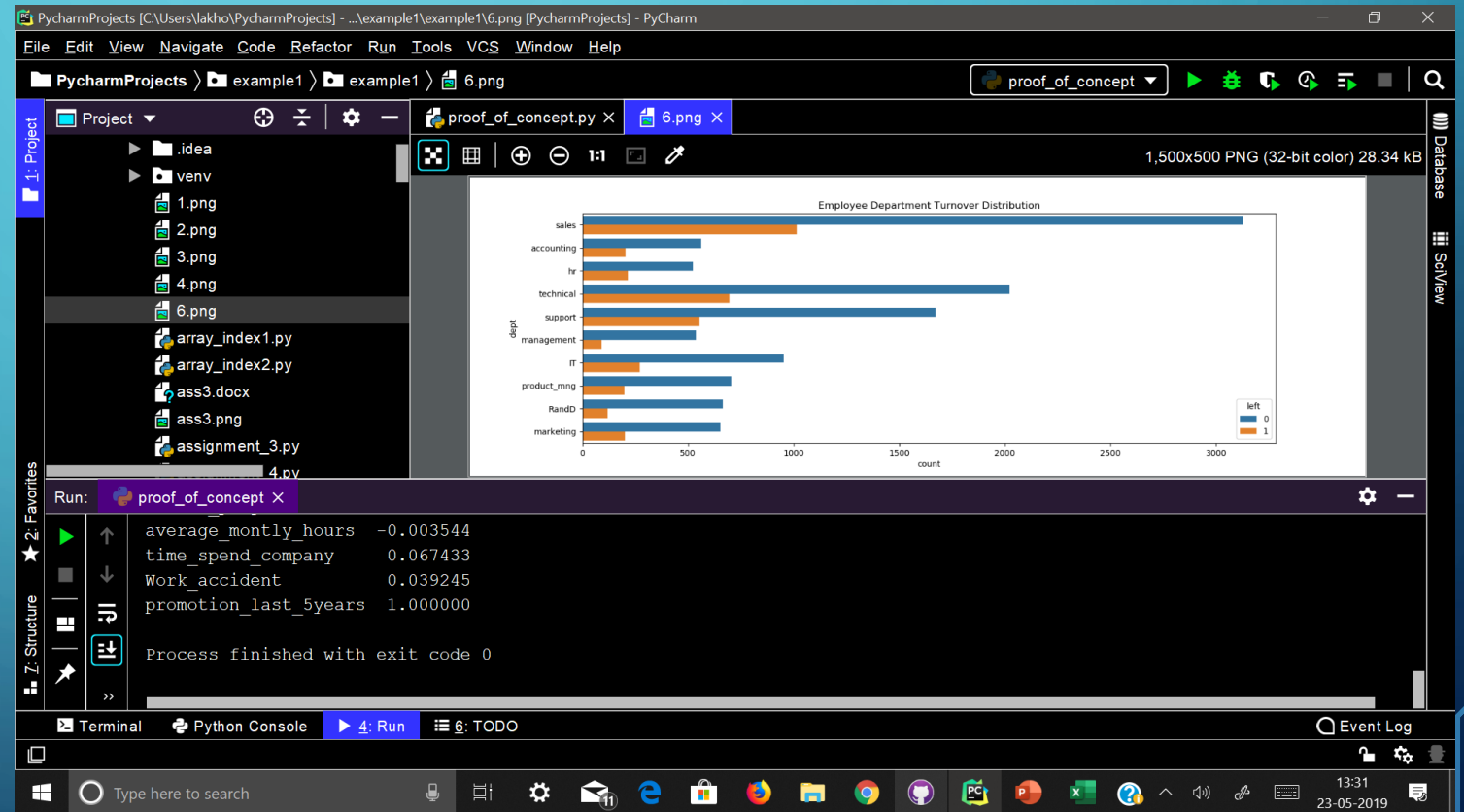
- > Majority of employees who left either had low or medium salary. Barely any employees left with high salary
- > Employees with low to average salaries tend to leave the company





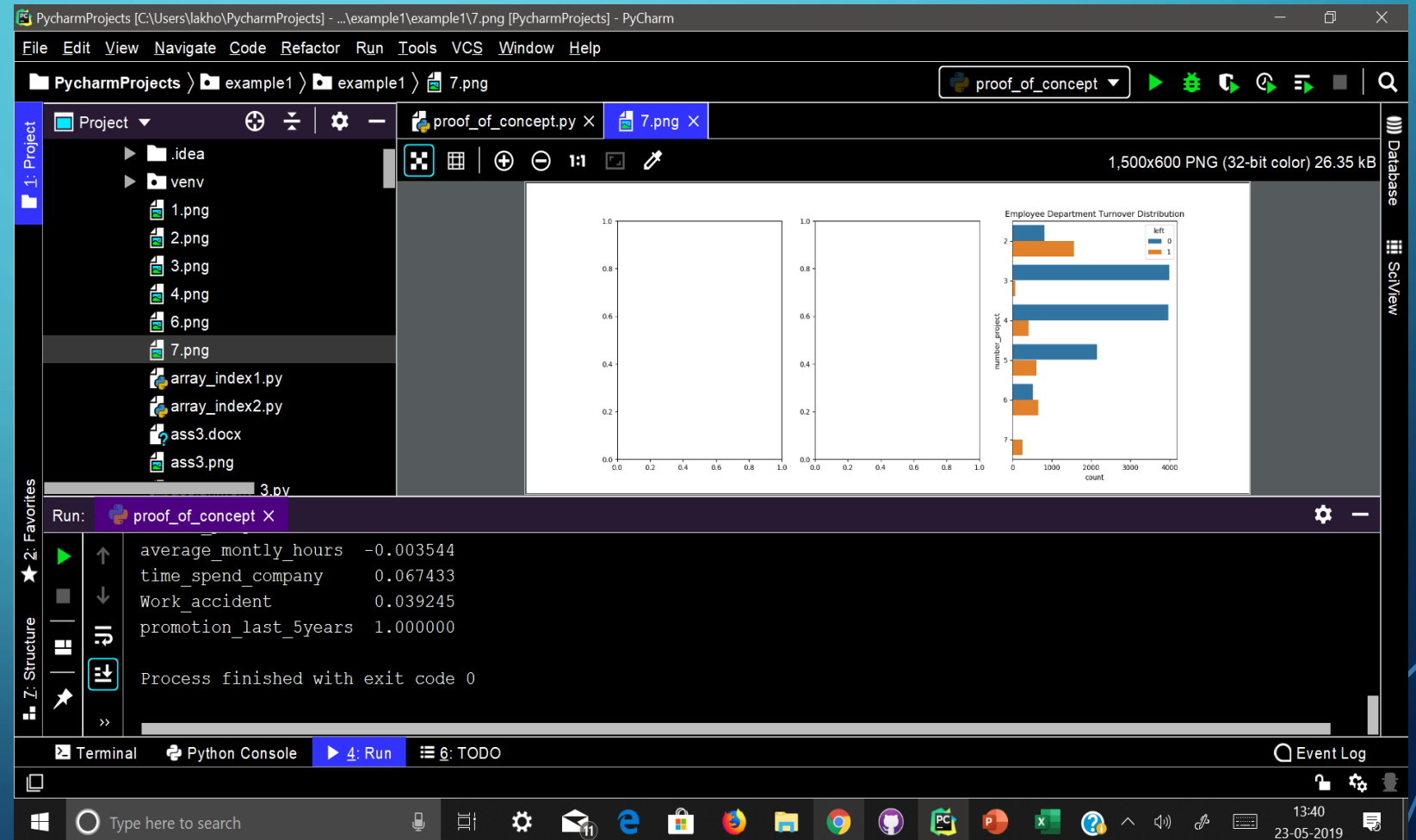
# DEPARTMENT VS LEFT EMPLOYEES

- > The sales, technical, and support department were the top 3 departments to have employee turnover
- > The management department had the smallest amount of turnover



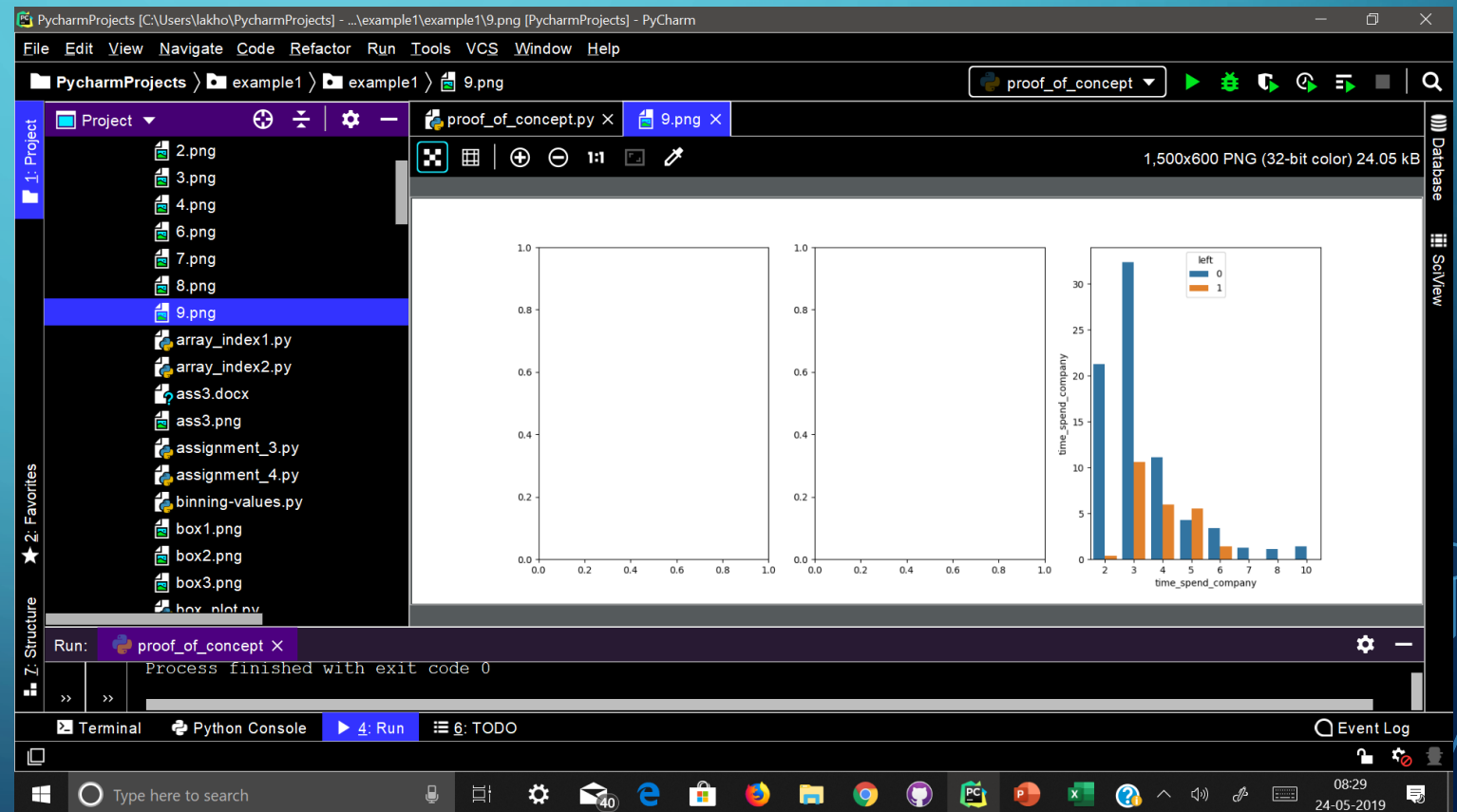
# PROJECT COUNT VS LEFT EMPLOYEES

- > More than half of the employees with 2,6, and 7 projects are leaving the company.
- > Majority of the employees who did not leave the company had 3,4, and 5 projects
- > All of the employees with 7 projects left the company
- > There is an increase in employee turnover rate as project count increases



# WORKING YEARS VS LEFT EMPLOYEE

- > More than half of the employees with 4 and 5 years are leaving the company
- > Employees with 5 years should highly be looked into. They are very likely to leave the company.



# ANALYSING THE PLOTS (CONCLUSION)

- Employees generally leave the company when they are either underworked or overworked.(  $< 6$  hr per day or  $> 10$  hr per day).
- Employees who have either very low or very high evaluations should be taken into consideration for high leaving rate
- Employees having low and medium salary are the bulk of leaving employee.
- Employees that had 2,6, or 7 project count was at risk of leaving the company meaning the employee should not have very less or very high workload.
- Employee leaving is majorly affected by their satisfaction levels.
- Employee that had 4 and 5 years At Company should be taken into consideration for high turnover rate
- Employee satisfaction, years At Company, and evaluation were the three biggest factors in determining number of employees leaving the company.

# EMPLOYEES WHO MAY LEAVE NEXT

- Employees with 5 years of work should highly be looked into.
- Employees with number of projects = 6.
- The major attributes which are responsible for an employee to leave are satisfaction level, evaluation and their working years in the company.
- So if an employee is not satisfied or has less satisfaction level , or not evaluated up to the mark, are more likely to leave the company next.