Data Visualization Project

Kalmanidis Theodoros
ID:218580

Language Used: Python 3 & Plotly

The dataset

Our dataset is an HR dataset that includes several parameters that were recorded in a company. Those parameters are linked to employees. In specific those parameters are linked to each employee's current position in the company. That is, whether the employee is still working for the company or they have resigned.

The objective

For this assignment we will have to create interactive visualizations using the appropriate visualizations principles and practices the theory tells us in order to reveal additional trends and patterns. In this project we will use interactive output from Python, it will be an html file containing the interactive plotly plots called pythonplots.html

The variables

The HR dataset consists of 15.000 lines and 10 columns. There are no blank values included.

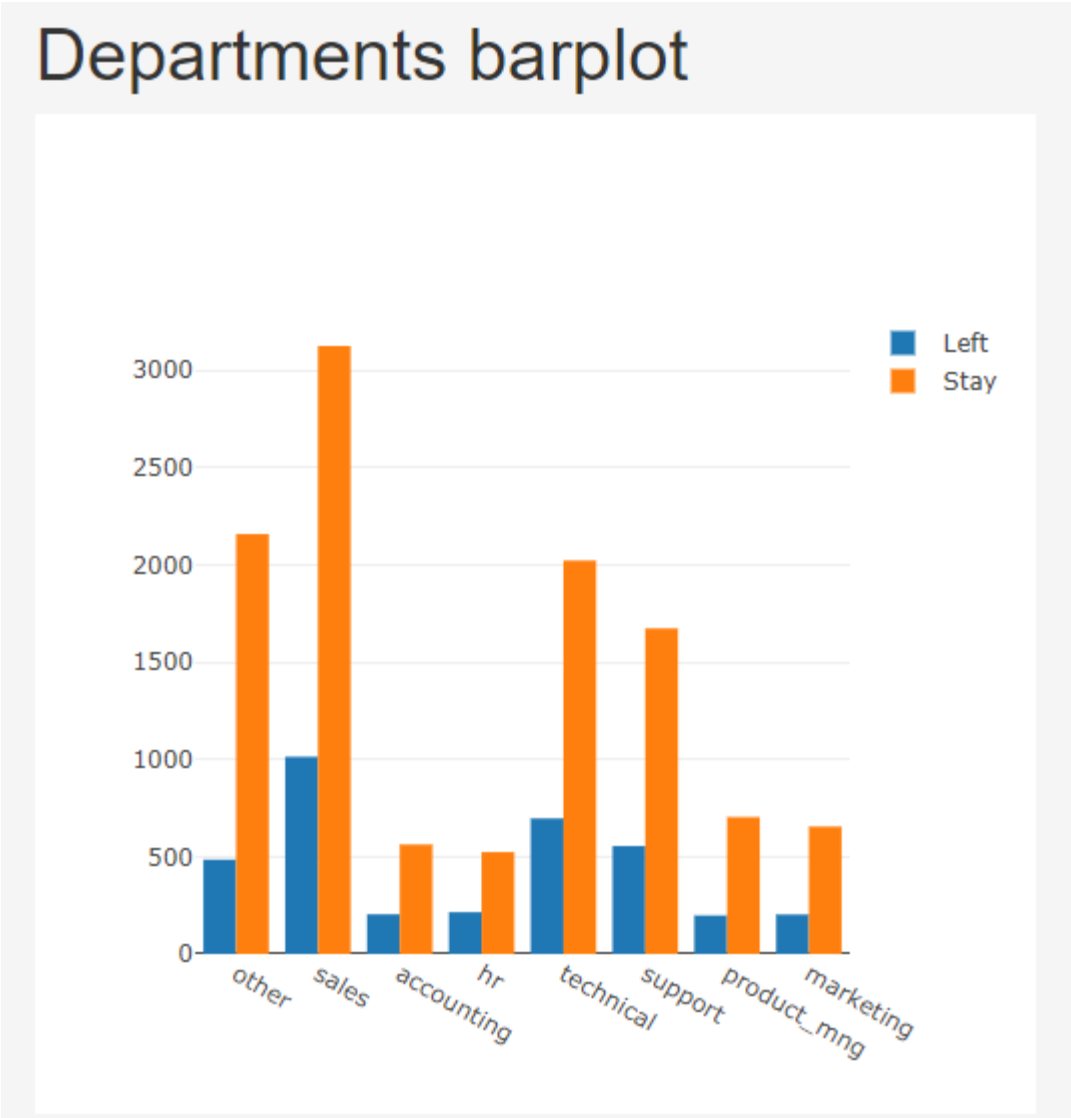
The columns that are included in the dataset are:

- Satisfaction Level
- Last Evaluation
- Number of projects
- Average Monthly Hours
- Time Send Company
- Work Accident
- Left
- Promotion Last 5 Years
- Department
- Salary

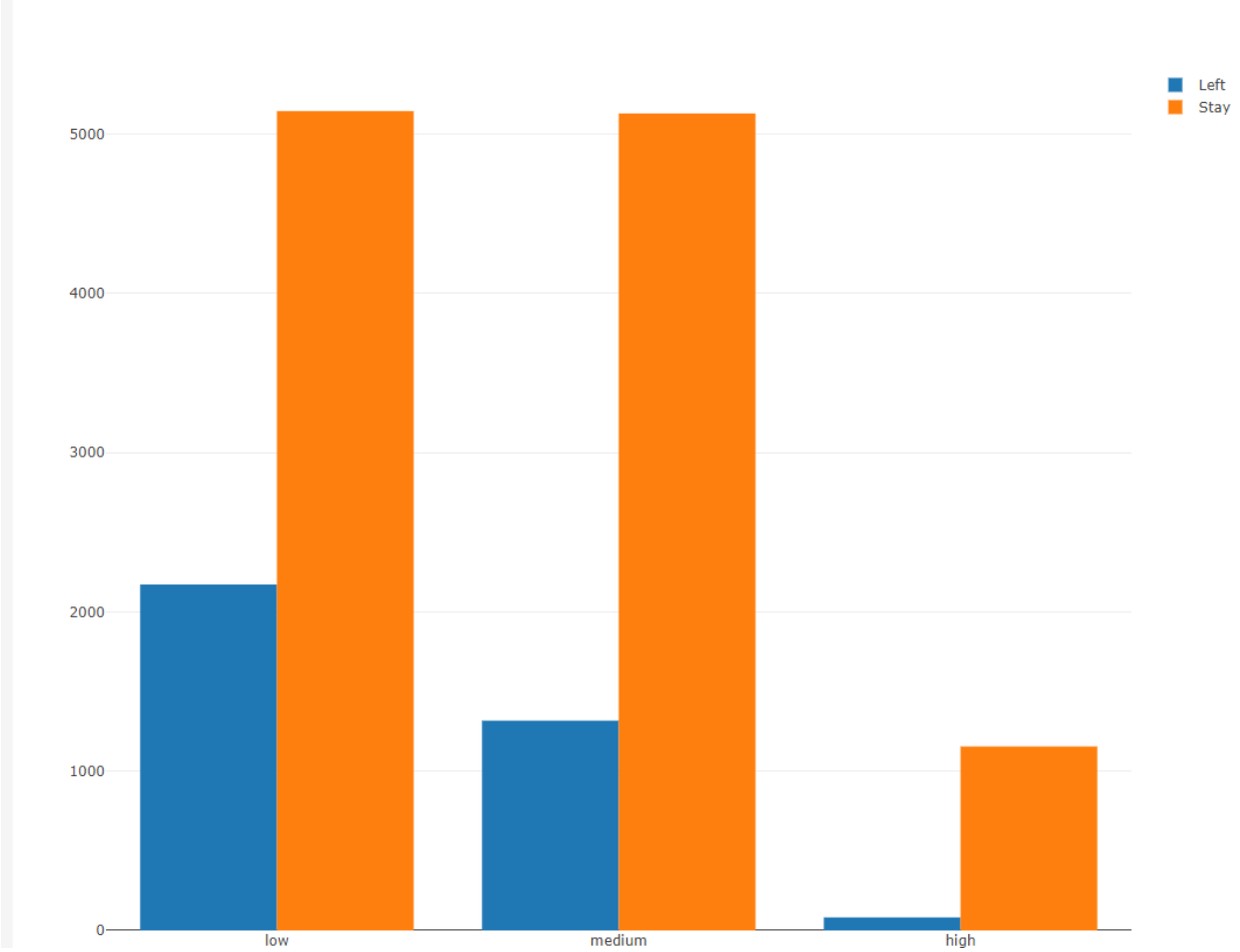
Among those variables, the "work accident", "left" and "promotion last 5 years" are categorical.

The rest variables are continuous.For the project's needs the continuous variables are categorized as well into 3-4 categories depending on the type of data, each one includes.

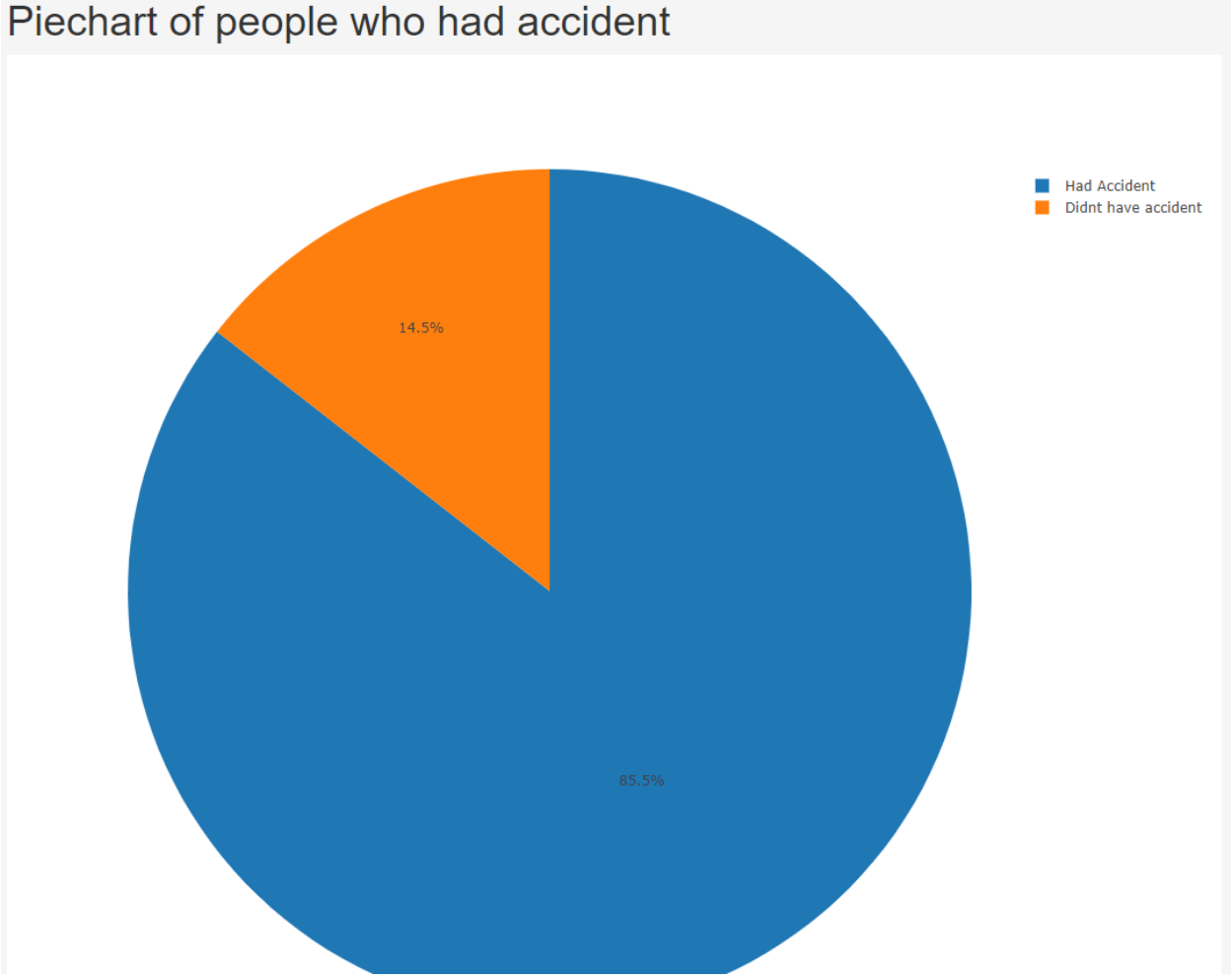
Data Visualizations



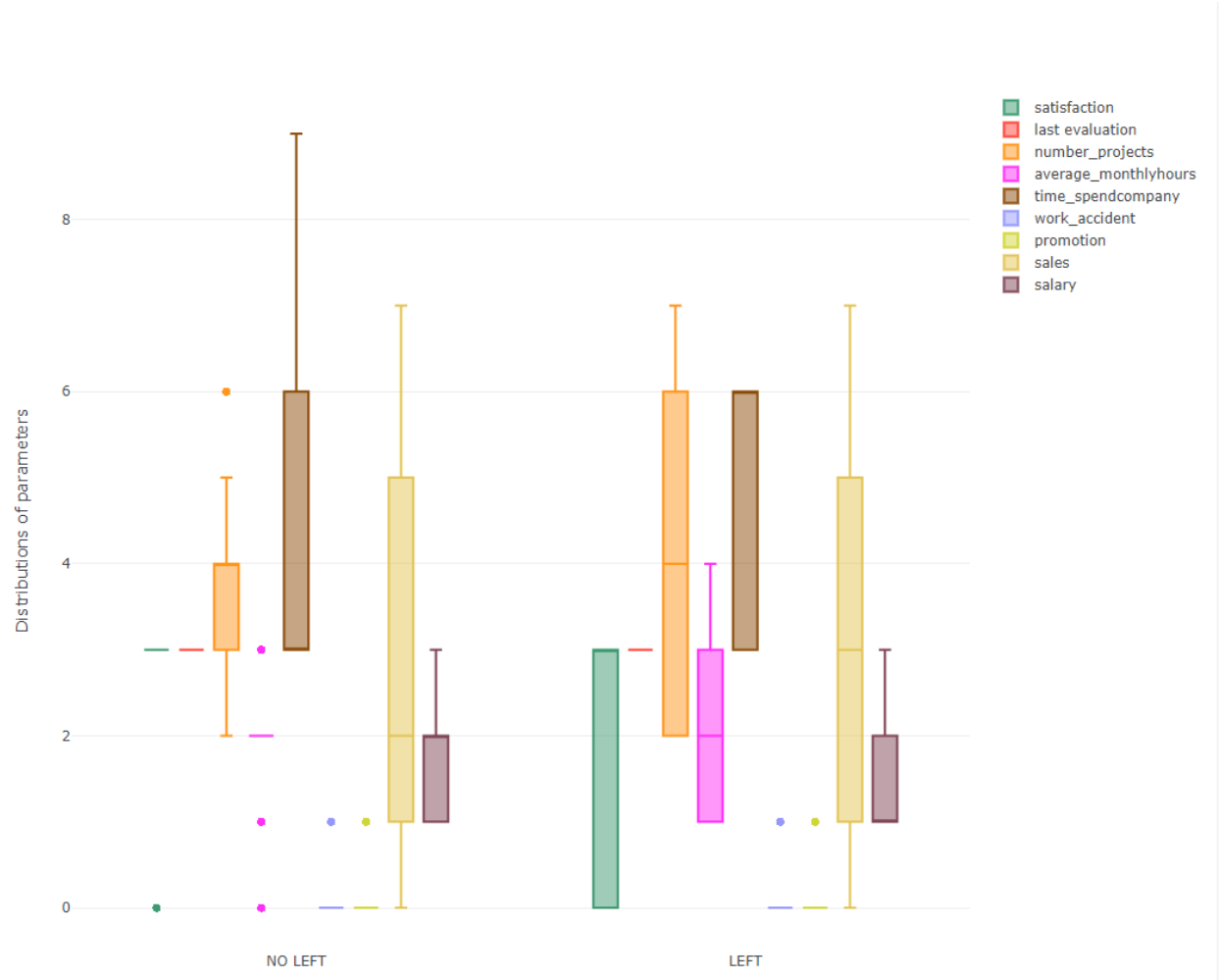
In the barplot above we can see the first interactive plot, it's a grouped barplot for all the departments containing the people who stayed or left from the company. We can see that the most people left from the sales and the technical department.



Below we can see our second grouped barplot it contains the sum of people who quit or stayed based on their salary (low,medium,high). From what we can see we can conclude that the higher the salary the less the probability for someone to leave, the people who low and medium salary who left are way more in count number compared to the high paid employees.



The third interactive plot is a piechart with the percentages of those who had accidents. We can see over 85% of people in dataset had some kind of accident.



The last interactive plot is boxplots for all variables for those who left and those who stayed. We can see that those who left had more projects (orange boxplot) per month and very low salary(light brown boxplot). Also the majority of those who left didn't get promotion the last 5 year and had very low satisfaction (green boxplot).