Lab 1

Christina Hartnett

The dataset I decided to work with was found on Kaggle.com. It is IBM HR analytics employee attrition and performance. The dataset includes 32 different attributes sampled from over 1,400 employees. Some of these attributes are factual, such as Daily Rate, Age, and Distance from Home, while others are ratings that the employees gave based on questions. These are all anonymous employees therefore there are no privacy issues with this dataset. This dataset had no missing values therefore there was no need for any imputation methods.

I thought this data would be interesting to visualize because I wanted to see if there was any information about how salary, overtime, or distance from work affecting things like rating of Job satisfaction, work life balance rating or attrition. I also wanted to see the percentage people who are satisfied at their current job. This data has a lot of interesting attributes that can help us understand reasons why employees may quit their job.

I was very surprised that so many employees rated Work Life Balance “Bad”. This was very apparent in both the bar chart and pie charts, as they were both very noticeably large. I was also interested to see that there is no obvious linear correlation between “Monthly Income” and “Job Satisfaction”. I included two other scatterplots for personal interest. The extra scatterplot uses job roles and gender to help differentiate if job role or gender affected different attributes such as monthly income or satisfaction. The second extra scatterplot is used to help determine the reasons why employees may or may not have quit.

Tools used:

Python

Plotly Dash

CSS stylesheet (https://codepen.io/chriddyp/pen/bWLwgP.css)

Kaggle Data (https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset)

Variables:
Age
Attrition describes if the employee quit or not.
Business Travel is how frequently an employee travels for work.
Daily Rate: The rate of their pay per day.
Department is the working sector.
Distance from Home
Education: 1 'Below College'; 2 'College'; 3 'Bachelor'; 4 'Master'; 5 'Doctor'
Education Field is the major in college.
Environment Satisfaction: 1 'Low'; 2 'Medium'; 3 'High'; 4 'Very High'
Gender is Male or Female
Hourly Rate is the hourly pay rate.
Job Involvement: 1 'Low'; 2 'Medium'; 'High'; 4 'Very High’
Job Satisfaction: 1 'Low'; 2 'Medium'; 3 'High'; 4 'Very High'
Monthly income (Income might be a determinant factor for employees to leave the organization)

Visualization Lab 1

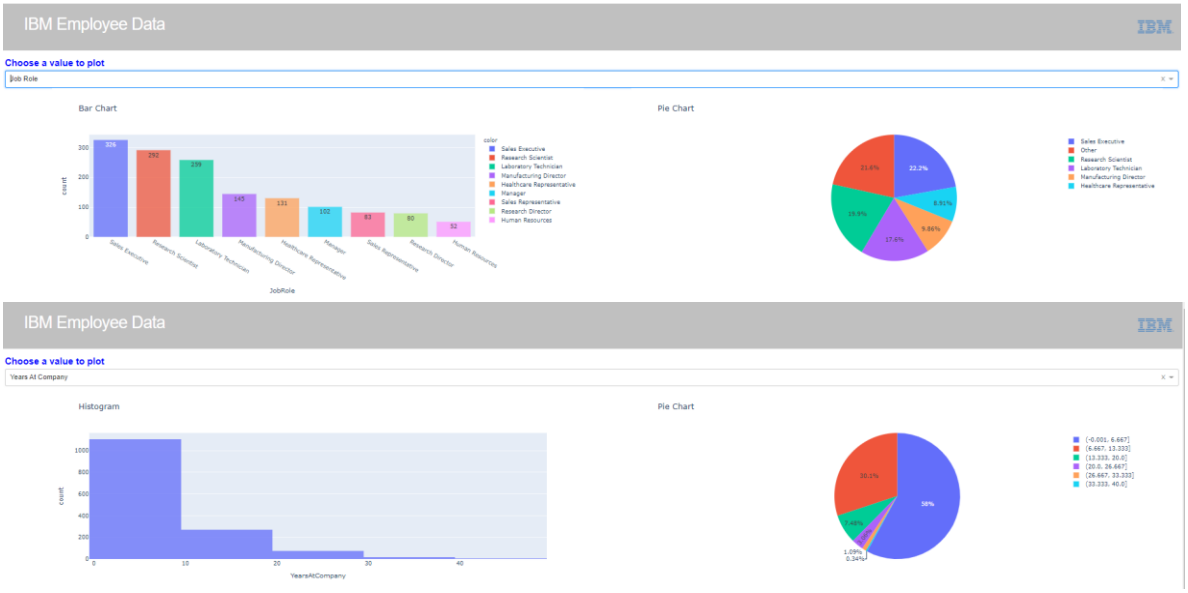
Job Role is the role of the employee.
Job Level is the level of the employee in the corporation.
Overtime is a yes or no that tells us if the employee works overtime.
Percent Salary Hike
Performance Rating: 1 'Low'; 2 'Good; 3 'Excellent'; 4 'Outstanding'
Relationship Satisfaction: 1 'Low'; 2 'Medium'; 3 'High'; 4 'Very High'
Standard hours: 80 hours for all
Stock Option Level: levels of the stock options for the employee
Total Working Years is the number of years the employee has worked overall
Training Times Last Year is how many times the employee was trained in the year
Work Life Balance: 1 'Bad'; 2 'Good'; 3 'Better'; 4 'Best'
Years at Company is the amount of time the employee has been at the company.
Years Since Last Promotion
Years with Current Manager

<u>Tasks:</u>
1. present a menu to allow users to select a variable and update chart.

2. draw a bar chart if a categorical variable is selected.

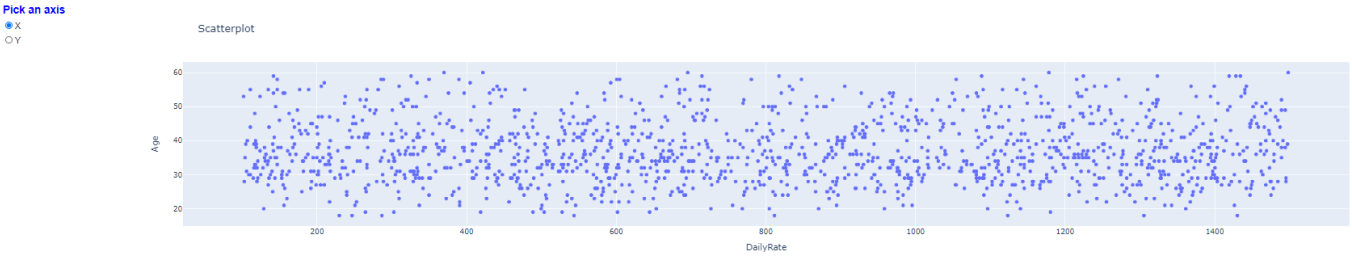
3. draw a histogram if a numerical variable is selected.

4. produce a pie chart of the selected variable (make sure you have no more than 5-6 slices, sum the smaller ones into a single ‘others’ slice)



Visualization Lab 1

5. produce a scatterplot of two selected variables (use a radio button to determine which of the two variable axes is to be loaded)



Extra scatterplots for personal interest: