REPORT – HW1

Dataset

This is a fictional dataset created by IBM data scientists to perform an exploratory analysis on the factors that contribute to the attrition of an employee.
https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset
Categorical variables considered: TrainingTimesLastYear, Department, Education, Attrition, BusinessTravel, Gender, EnvironmentSatisfaction, NumCompaniesWorked
Numerical Variables considered: Age, DailyRate, MonthlyIncome, YearsAtCompany, TotalWorkingYears, PercentSalaryHike, YearsWithCurrManager

Technologies and Languages Used

- D3 js – for generating charts from the data
- Python – for setting up the server to render the homepage
- HTML, CSS, JS – for design of the page
- Bootstrap – for the elegance of the page

Functional requirements

1. Present a menu to allow users to select a variable and update chart.

```
<div class="form-inline custom-centered">
        <label for="exampleFormControlSelect1">Select Variable:   </label>
        <select class="form-control" id="exampleFormControlSelect1">
        <option value="">Select a value</option>
        <option value="Age">Age</option>
```

Used input tag with select list to get the variable under consideration

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

```
bars = ["Department", "Education", "Attrition", "BusinessTravel", "Gender", "
if (bars.indexOf(col) != -1) {
    for (let index = 0; index < data.length; index++) {
        col_data.push(data[index][col]);
    }
    console.log(col_data)
    updateBarChart(col, col_data, 0);
    document.getElementById("bin").hidden = true;
} else {
```

Checked if the select variable is a categorical one and made a call to the bar chart. Else, a histogram is called.

3. Draw a histogram if a numerical variable is selected (bin it into a fixed range (equi-width) of your choice)

```
var histogram = d3.histogram()
    .value(function(d) { return d; })   // I need to give the vector of value
    .domain(x.domain())  // then the domain of the graphic
    .thresholds(x.ticks(nBin)); // then the numbers of bins

// And apply this function to data to get the bins
var bins = histogram(col_data);
```

Generated the bins based on the number of bins selected using the slider. Slider was used as it is more intuitive to a user upon first glance. Mouse drag may not be understood by a normal user without hints.

4. On mouse-over display the value of the bar on top of the bar

```
var tool_tip = d3.tip()
            .attr("class", "d3-tip")
            .offset([-8, 0])
            .html(function(d) { return "Frequency: " + d.length  + ", Range: "+(d.x0)+"-"+(d.x1); });
            d3.selectAll("svg").call(tool_tip);
```

Made a tool tip for displaying the value of the variable and it's category/ range. It is added to the onmouseover event & removed when onmouseleave is called

5. On mouse-over also make the bar wider and higher to focus on it

```
var showTooltip = function(d) {
    d3.select(this).style("fill", "orangered")
    tool_tip.show(d)
}
// A function that change this tooltip when the leaves a point: just need to
var hideTooltip = function(d) {
    tool_tip.hide(d)
    d3.select(this).style("fill", "orange")
}
```

Focus is added by changing the color to a more intense red, from a less intense orange color. It is done as changing the dimensions may give false data comprehension to the user.

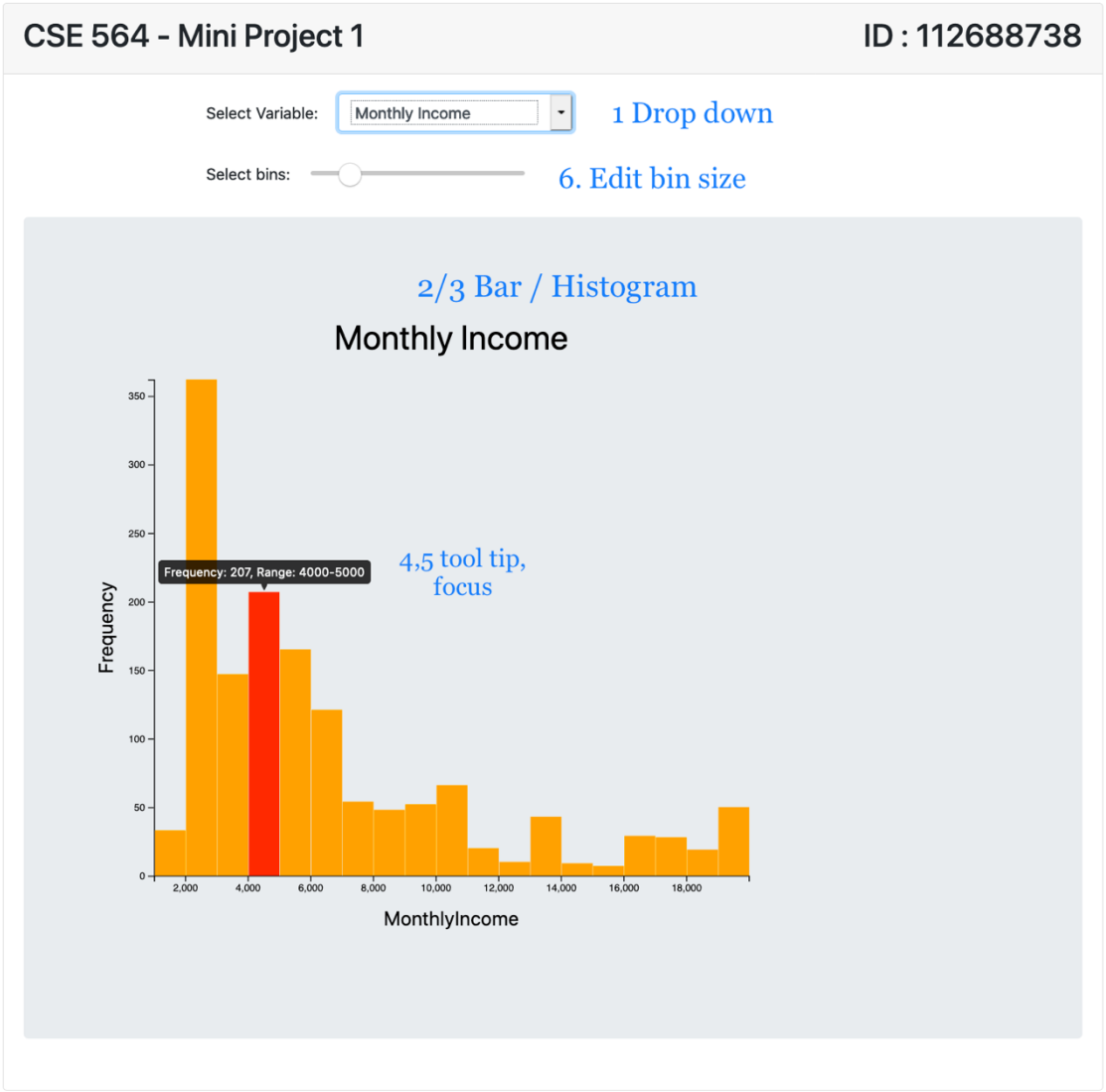
6. Mouse (with left mouse button down) move left (right) should decrease (increase) bin width/size (for numerical variables only)

```
// Listen to the button -> update if user change it
d3.select("#nBin").on("input", function() {
    updateHistogram(col, col_data, +this.value);
});
```

Added a onchange event to the histogram to capture the change and then redraw the histogram. This field is not visible when user selects categorical variable by setting hidden = true for the element.

Glance of features

A detailed overview of the 6 key points and axis labels are shown in the following picture.



Youtube Link

https://youtu.be/3PydE62X0ts