

Visualization Project Proposal

Name: Kunal Kolhe

Group of 1

112748535

Dataset:

I am going to use the same dataset I have been using for the assignments 1 and 2.

It is the Australian credit approval dataset. This dataset has a good mix of Categorical, and numerical attributes. It has 690 data entries, 14 independent variables and 1 dependent variable.

I made a workable mapping based on this source [here](#) between the variables and their labels, since the data has been anonymized before posting it on the UCI Machine Learning Repository.

During the course of the project, I will be using the various tools I have learned over the duration of the course.

The data processing and the back end will be built using Flask.

The styling of the project will be done using Bootstrap and CSS.

The graphs and charts will be made using D3.js.

Approach:

I plan to first find the correlation of each of the independent variables against the target variable.

Based on that, I can get a clearer idea about how important the variables are and how to rank them in term of importance of how much they contribute towards classification of the target variable.

Since there are 14 dimensions, a parallel chart can best be used to visualize all the data points.

First based on the correlations for each variable, I will rank the variables. Then make the parallel chart.

The parallel chart will ideally have brushing on each axis so that the user can select whatever variable and range to view.

Due to this, a deeper understanding of the data can be achieved using a visualization.

Next, I can use multiple charts like barchart, categorical row chart and pie chart to visualize the individual variables to further understand the data in more granularity.

The main insight I aim to draw from this data is how each parameter affects the credit approval for a new credit line.

From my account at Chase, I got the following weightages which contribute to my credit score.

Payment History	40%
Credit History	21%
Credit Usage	20%
Total Balances	11%
Credit Check	5%
Available Credit	3%

Now this dataset is comprising of many more values and these things were asked during my credit approval form for the first time.

Based on this dataset or some new dataset, because this one is from before 1997 (older than me) we can calculate potential credit approval for some person. Banks have leveraged this power of data to insure themselves against high risk.

Different plots:

I will be making use of various plots from d3.js to visualize this dataset extremely intuitively.

I plan to use the following:

- Bar Charts
- Pie Charts
- Parallel Coordinate Chart
- Correlation Scale
- Scatter plot matrix

The final dashboard might look something like this:



The color scheme is chosen in such a way to ensure color blindness does not affect perception of this dashboard.

It is also very eye catching and the entire website is built in such a way that all the colors are cohesive and pleasing to the eye.

The user does not require any scrolling and this does not put load on the user to memorize the anything.

An important part is the titles, since people generally tend to remember what the dataset was conveying rather than the individual data.