DATA SCIENCE PORTFOLIO

Jombaba.s7@gmail.com

I enjoy working on data and highlighted in this portfolio is one of the projects I have implemented using the python-based tools and programming platform. The entire project is presented in five sub-projects with each of them capturing specifics of the entire work.

Imagine that, a financial institution / bank wish to find a solution to a 'Customer Acquisition and Customer Retention' related problem. As a data Scientist, this is my attempt at providing a wholesome solution and the series of five projects illustrate a plausible approach in resolving the problem.

Interactive Data Visualization Project
Purpose:
This is the third in the series of projects. Our purpose is to examine the 'Credit Card Application' (CAD) dataset and present an interactive dashboard which will be useful for data science teams and more business oriented end-users.

In the first of the projects in this series, we successfully removed all the missing values in the dataset. After that, we featured an investigative look at the dataset where we were looking for unintended discoveries of characteristics in the dataset.

Here, we develop a single-item dashboard which is based on the distribution of two of the sixteen variables of the dataset. The two variables used in the plot were chosen arbitrarily.

Dataset:
The dataset is found at: - https://archive.ics.uci.edu/ml/datasets/credit+approval
The dataset comprises continuous and nominal attributes of small and large values. Originally, there were 67 missing values in the dataset. The missing values were treated in a previous sub-project. Here are some specifics of the dataset:

The number of instances (observations) is 690. The number of attributes is 15 which are labeled A1-A15. There is one class attribute which is labeled A16 where 307 or 44.5% of the classifier is "+" while 383 or 55.5% is "-".

Attribute Label	Value Type	Attribute Label	Value Type	A6	Nominal	A16	Class attribute
A1	Nominal	A11	Continuous (Integer)	A7	Nominal		
A2	continuous	A12	Nominal	A8	continuous		
A3	continuous	A13	Nominal	A9	Nominal		
A4	Nominal	A14	Continuous (Integer)	A10	Nominal		
A5	Nominal	A15	Continuous (Integer)				

An Interactive Data Scatterplot.

This is an interactive scatterplot of variable A2 values on the x axis and variable A8 values on the y axis. A2 and A8 exhibit the highest R^2 value of 15.4.

The user interacts by hovering the mouse over the plot. As the mouse hovers, the values of other variables A1, A2, A8, A9, A10, A11 and A16 for the particular record (observation / row) are displayed.

variable A2 values vs. variable A8 values
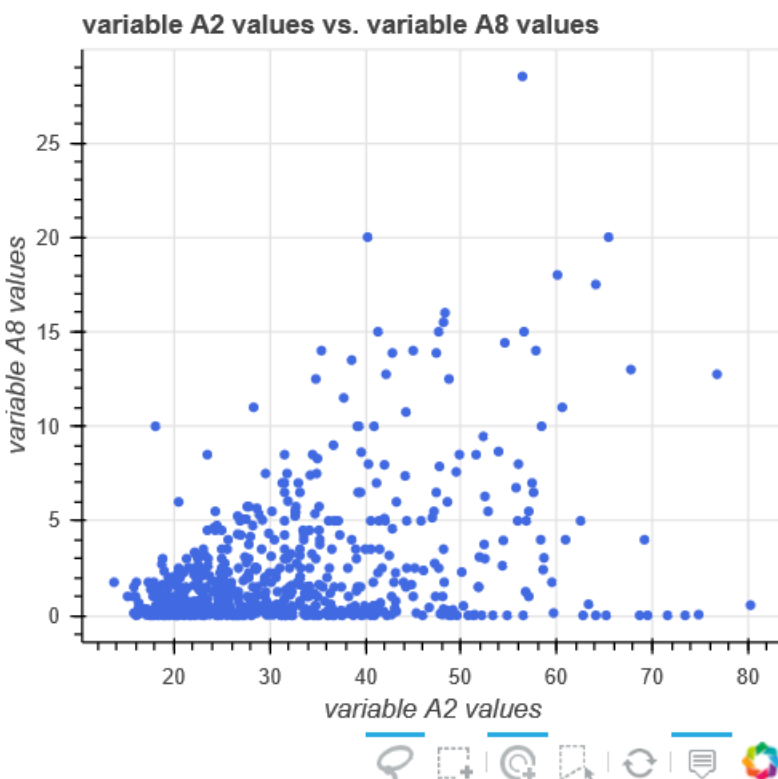
This plot displays the importance and relevance of such dashboard elements in deriving insights from this dataset. Through such dashboards we extend the scope of our diagnostic analysis by inviting all stakeholders to participate in the derivation of the model.

Intuitive questions and or hypotheses, worthy of further examination, might evolve from the data team, board room as well as other stakeholders utilizing the dashboard. This will foster collaborative effort which can aid the derivation of the final model immensely.

Similar plots showing some other features of the dataset can be developed as desired.

tyJ_A