**RTI CDS Analytics Exercise 01 – Write Up**

Aditya Anerao

02/02/22

**Overview**

**Methodology & Analysis**

Data Summary

I did not remove missing values from the data. Each missing value was assigned a unique category within each variable.

I split the data into training, validation, and test sets using a 70/20/10 split respectively. I then built a logistic regression model using the training dataset. This model predicts whether a person will make over $50,000 based on the

Used a logistic regression model to classify each person into two categories

**Results**

Age of people in sample ranged from 17 to 90 years old with a mean of 38.6 and median of 37

**Chart, treemap chart

Description automatically generated**

**Figure 1: Confusion Matrix of Predicted & Actual People earning more than $50,000 per year**

**Conclusion**