**Final Project – DSC530 Data Exploration and Analysis**

**Customer Churn Prediction**

**Author : Sathish Manthani**

**Bellevue University**

**Statistical/Hypothetical Question:**

I took Customer Churn dataset in Bank dataset from Kaggle and answered the following questions:

**What variables help in predicting the customer attrition?**

**Do Female customers churn more than the Male customers?**

**Outcome of your EDA:**

I built a couple of models to explore the relationship between target variable and independent variables. Logistic regression with Customer Exit Status variable gave me a better model. So, I concluded Price has relationship between Age and IsActiveMember variables in the dataset.

**What do you feel was missed during the analysis?**

I didn’t explore the multicollinearity, autocorrelation and time-series analysis on the dataset.

**Were there any variables you felt could have helped in the analysis?**

The dataset only contains current balances and statuses of the bank accounts. If I had historical monthly balances and statuses, time series analysis would have given us another model to compare with.

**Were there any assumptions made you felt were incorrect?**

I went with logistic regression since the target variable is binary. There maybe better models fit the dataset but I went with logistic regression due to my limited knowledge.

**What challenges did you face, what did you not fully understand?**

I struggled little bit during Null Hypothesis Testing as I was not sure what variable I need to take to calculate the test statistic. Plotting t-distribution was also a tough task without fully understating the concept. Allen Downey’s book was helpful in understanding the concepts with coding examples. Thank you for referring the right content along with coding exercises.