



Data
1030



Predicting Churning Customers

Reported by Yingfei Hong
From Data Science Initiative

Date: Oct 14th, 2021

<https://github.com/YingfeiHong01/Data1030-FinalProject>



Introduction



The problem is to predict the "churned customers" for the bank managers



1. Empower CRM and customer experience teams to be creative and proactive in their engagement with the customer
2. Retain existing customers



Dataset comes from LEAPS with 21 columns and 10127 data points



It is a classification problem



Exploratory Data Analysis

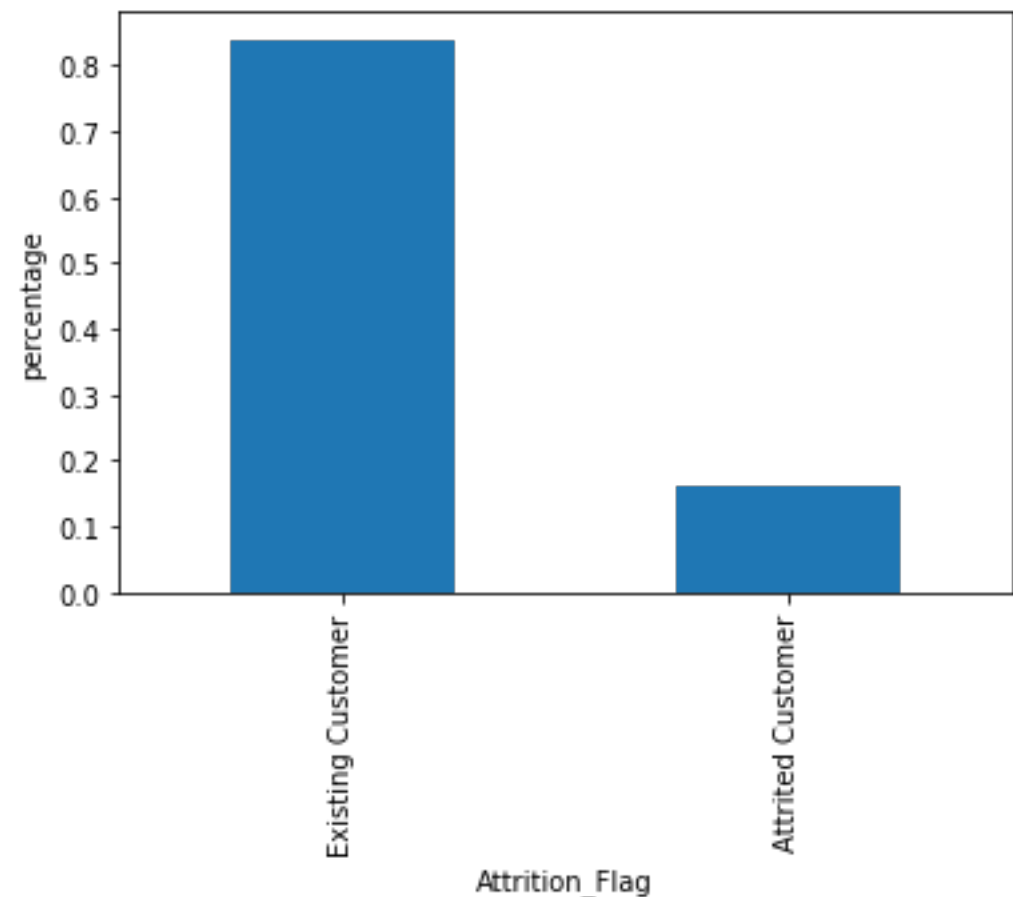


Figure 1 Distribution of Target variable

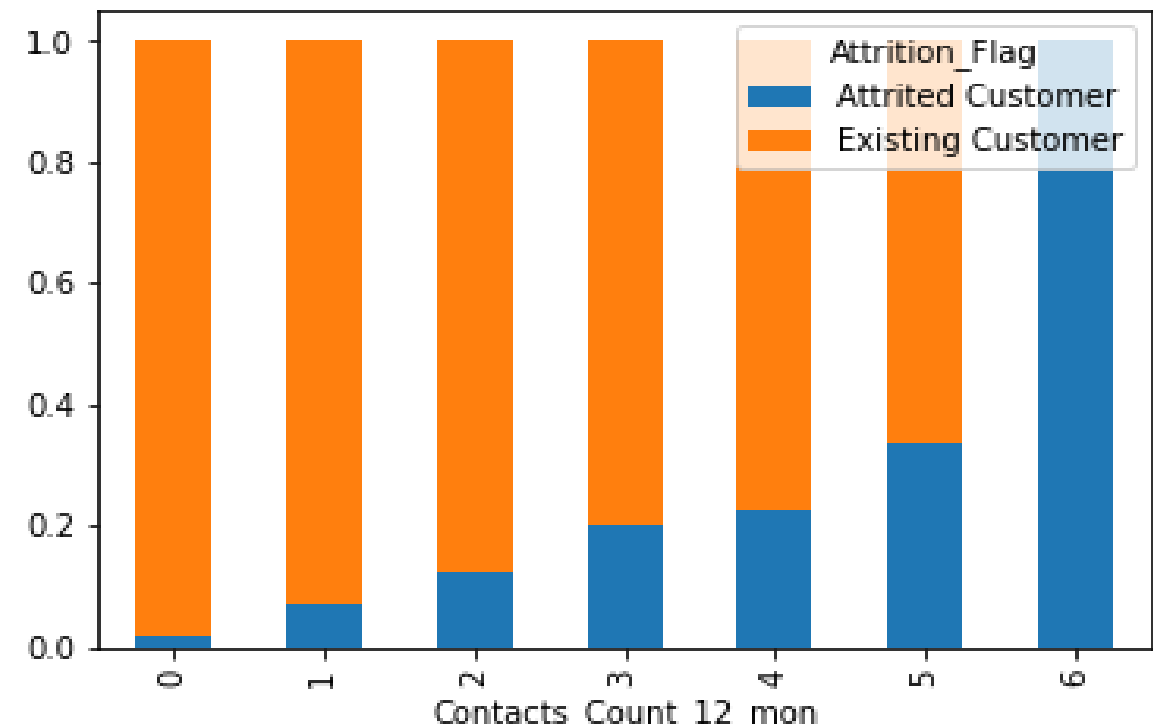


Figure 2 The distribution of the number of contacts across two different customers



Exploratory Data Analysis

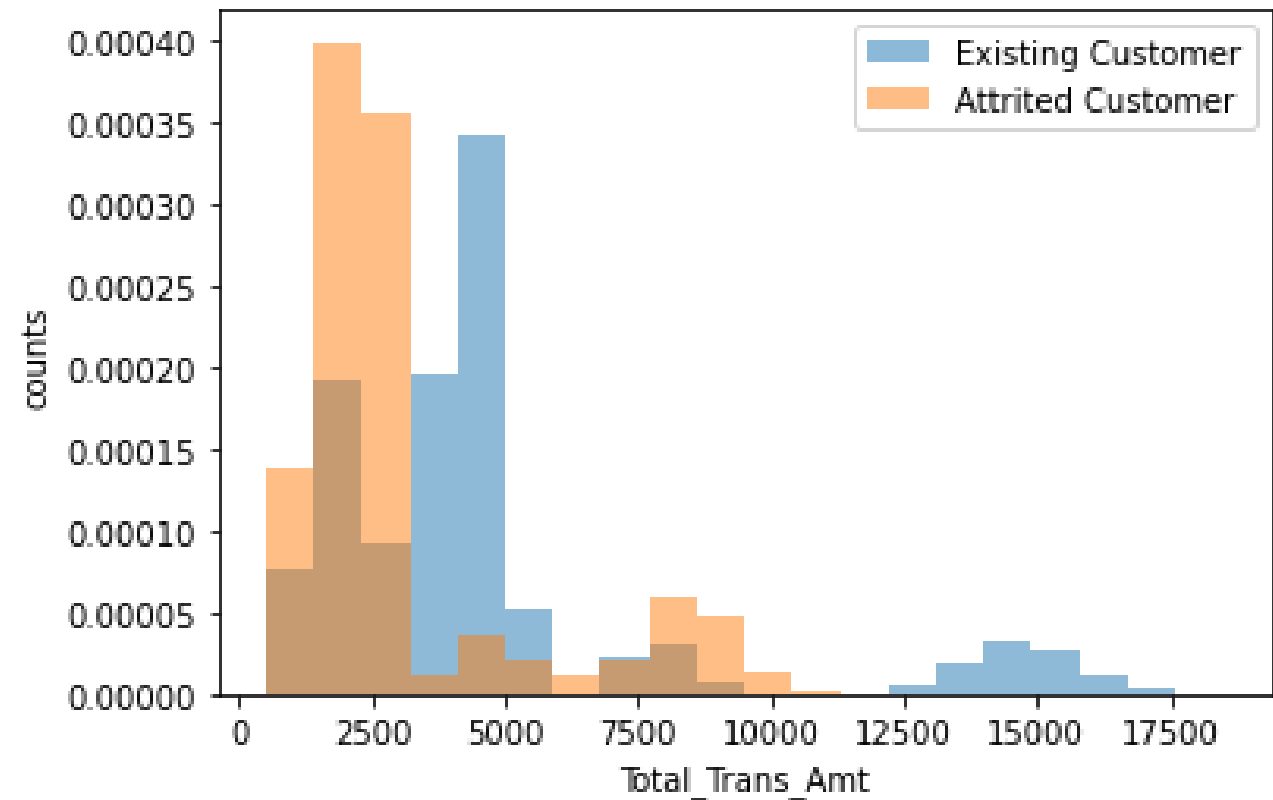


Figure 3 The distribution of total transaction amount across two different customers

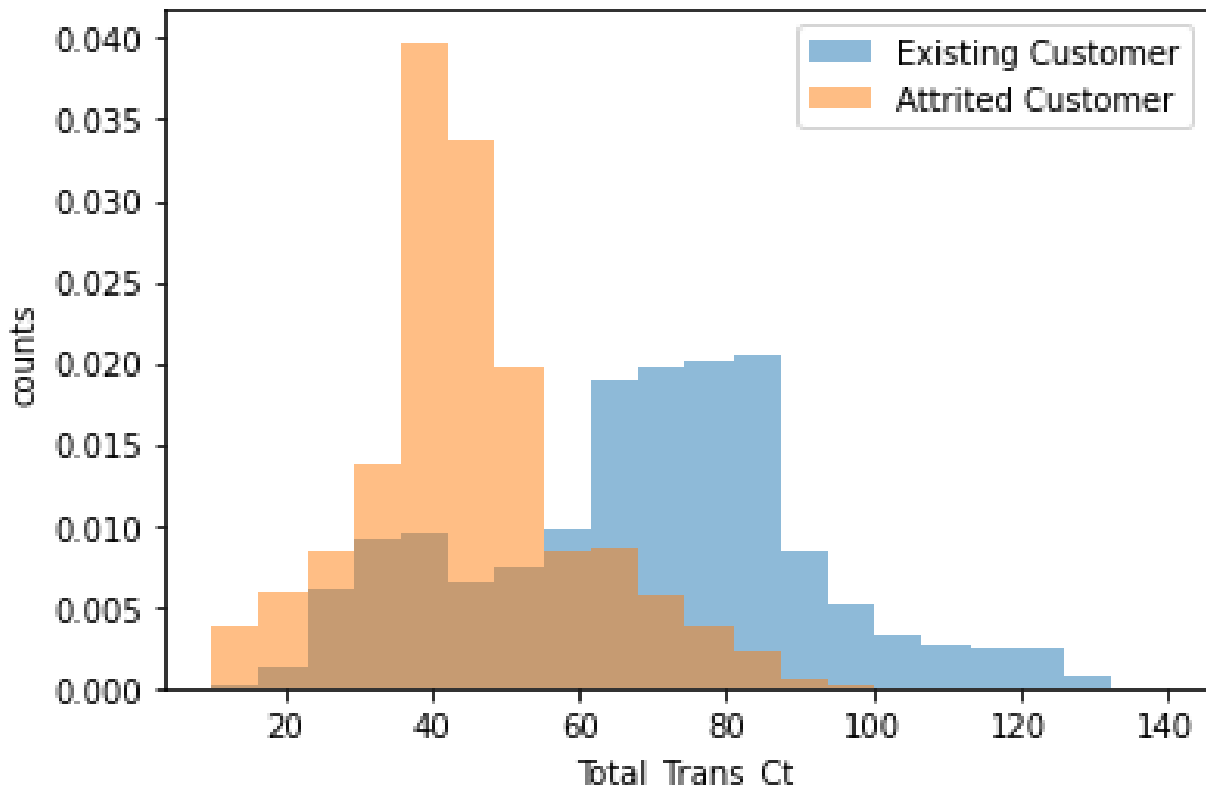


Figure 4 The distribution of total transaction count across two different customers



Splitting and Preprocessing



Stratify method ; train, val, test = 0.6, 0.2, 0.2



OneHotEncoder for category features like gender and marital status

OrdinalEncoder for ordinary features like education level...

StandardScaler for continuous features like total transaction count...



Missing Value: “unknown” in some demographic variables

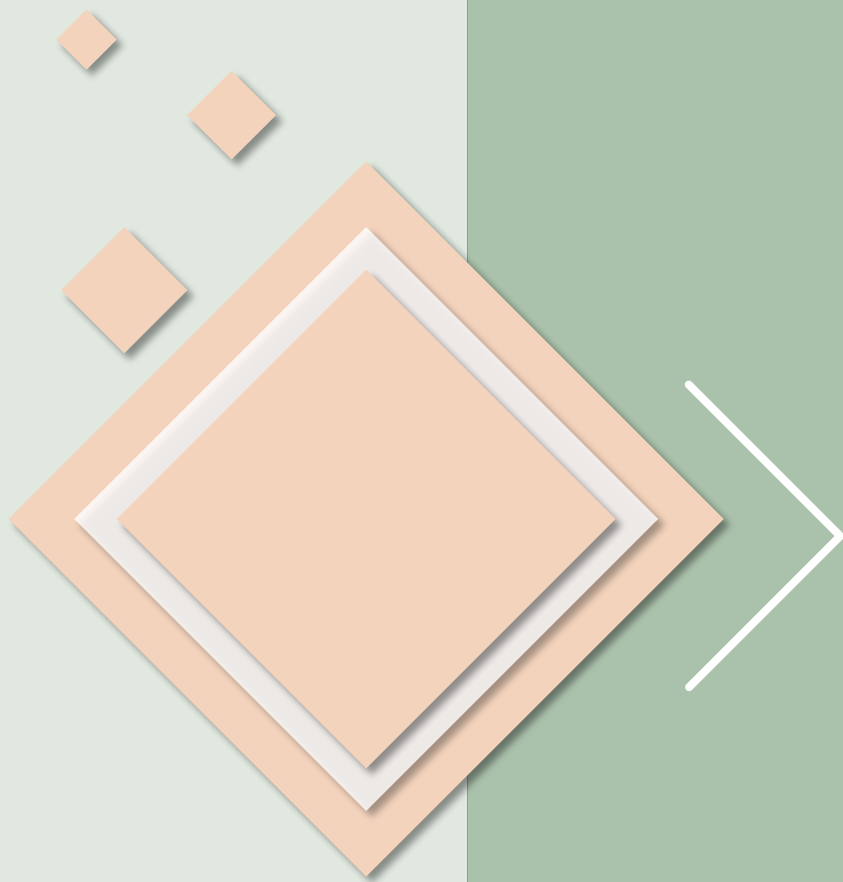
Treat them as a new category



LabelEncoder for target variable



Train size = (8477, 24), validation size = (3649,24), test size = (3644,24)



Thanks for listening

END