Loan Data

- **Loan_ID** - Removed: not useful for classification models
- **Gender** - Changed to dummies: It was male and female before
- **Married** - Changed to dummies: It was yes and no before
- **Dependents** - Change to dummies: It was 0, 1, 2, 3+
- **Education** - Change to dummies: It was graduate and not graduate before
- **Self_Employed** - Changed to dummies: It was yes and no before
- **Credit_History** - Replaced NA to 0 because we don't have the information if they meet their guidelines, so for now, they haven't met the guidelines. Changed the data type to integer since it was float before
- **Property_Area** - Changed to dummies: It was urban, rural, semiurban before
- **Loan_Status** - Changed to dummies: It was Y and N before
- **Loan_Amount_Term** - Changed all values that are 360 to 1 and other values to 0. Majority of data had 360 months of loan amount term. Then the rest of NA values were dropped.
- **LoanAmount** - Filled NA values with 0 because applicants haven't received their loan yet

Everything was scaled except for the Loan Status

Purchase Dataset

- Stay_In_Current_City_Years had a special character of '+' which was removed
- Stay_In_Current_City_Years column was changed to a float data type
- Mapped Gender column F = 0 and M = 1
- Used pd.get_dummies for all the features except 'User_ID', 'Purchase', and 'Product_ID'
- Dropped 'User_ID', 'Purchase', and 'Product_ID'
- Scaled the data using StandardScaler