
Deployment Project

LOAN PREDICTIONS WITH ML MODELS AND PIPELINES

MARCH 3, 2023

Spencer Denton

DATA SCIENCE BOOTCAMP

PROJECT GOALS

- **To predict loan approvals**
 - **To build various machine learning models**
 - **To build pipelines to efficiently deploy machine learning models**
 - **To build an API to connect to the cloud**
 - **To deploy models to cloud**
-

HYPOTHESIS: APPLICANTS MORE LIKELY TO BE APPROVED WHEN...

- **They have higher incomes.**
 - **They have a co-applicant.**
 - **They have credit history.**
 - **Loan amount is lower.**
 - **Property type is urban.**
-

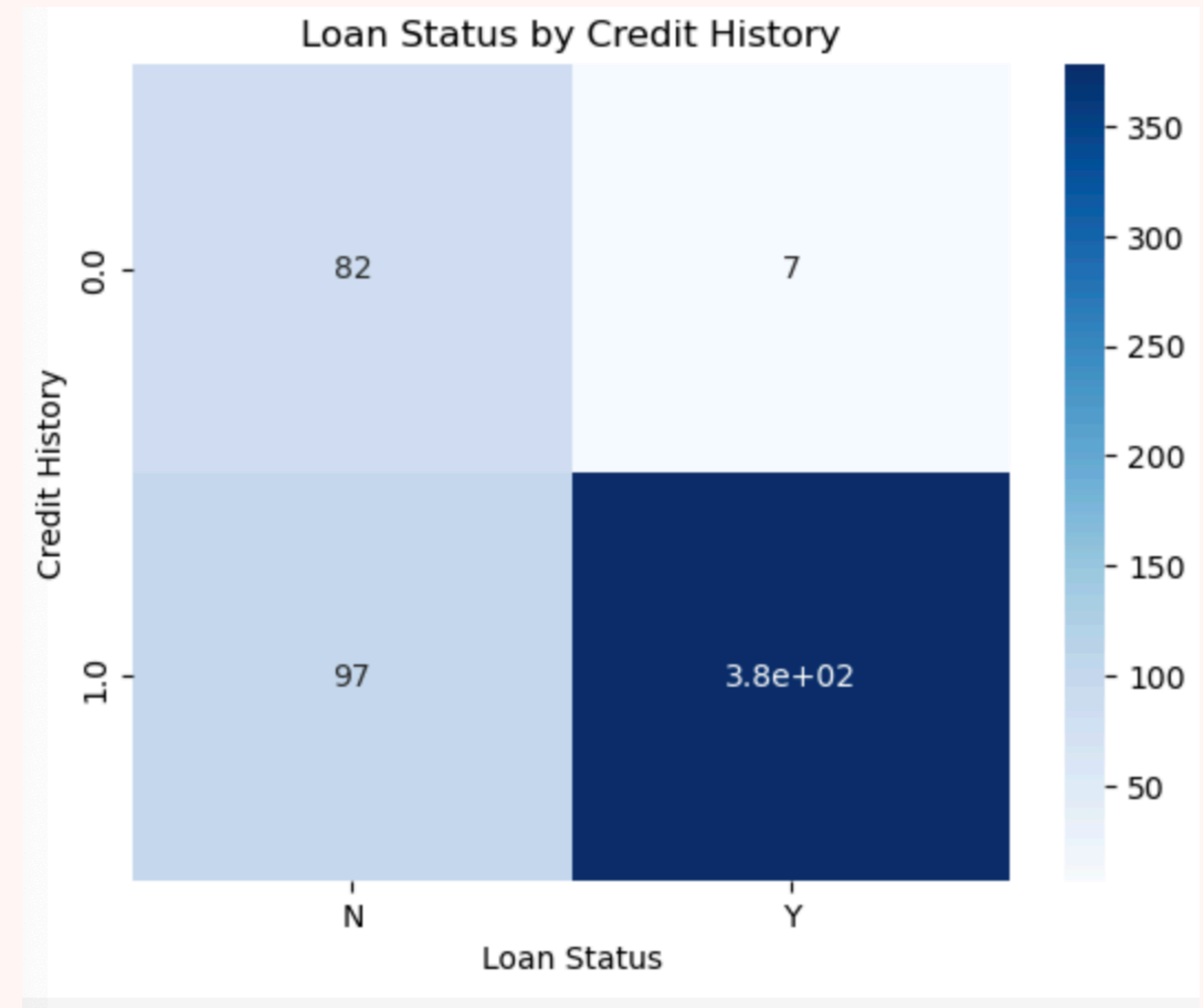
PROCESS

- **Familiarize myself with data**
 - **Perform EDA, Cleaning and Feature Engineering**
 - **Conduct Feature Engineering**
 - **Build Machine Learning Models**
 - **Evaluate Machine Learning Models**
 - **Build API with Flask**
 - **Deploy to Cloud (AWS)**
-

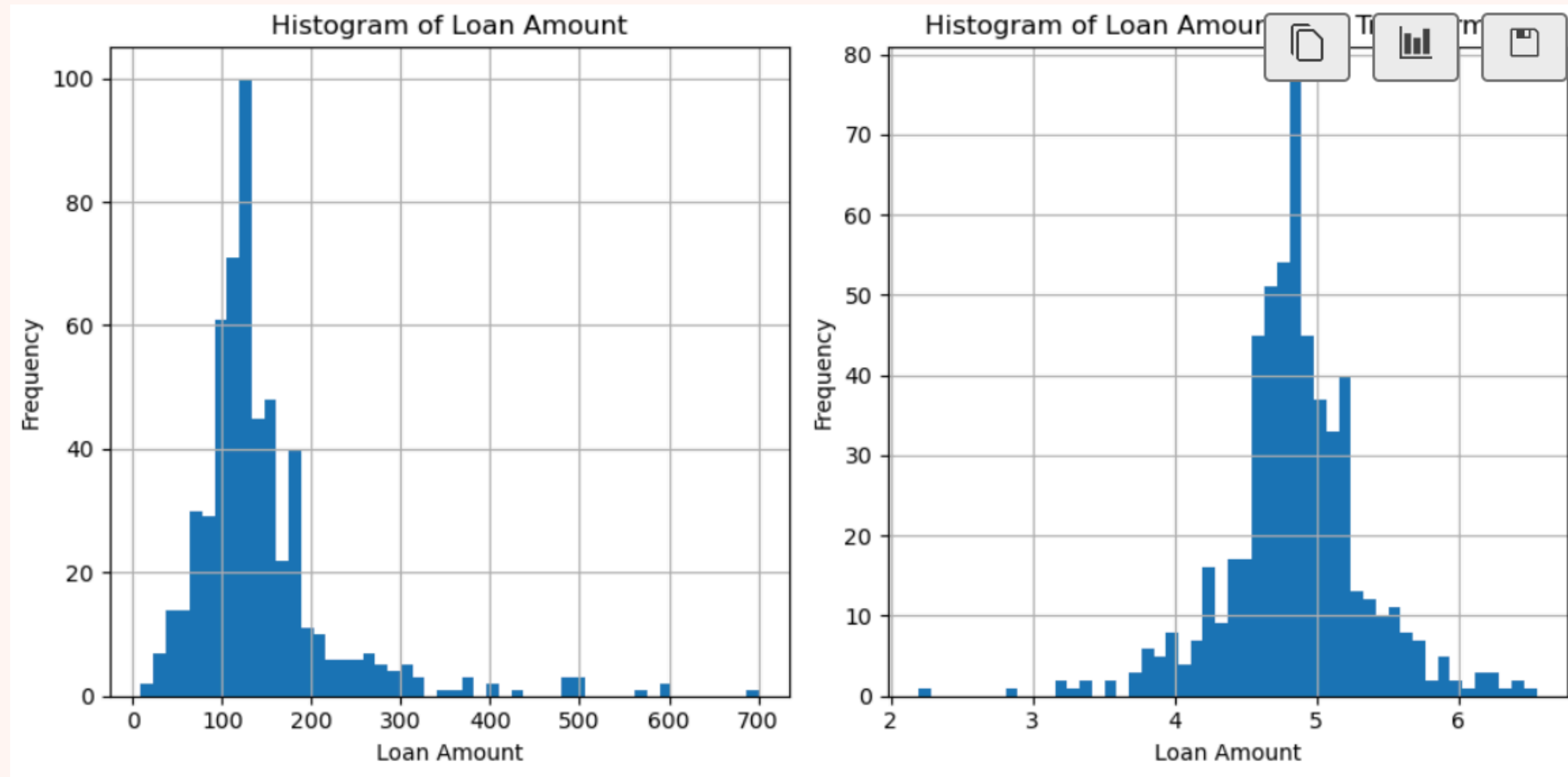
EDA, CLEANING AND FEATURE ENGINEERING

- **Replace NaN's using mode for categorical values and median for numeric (based on the skew of income distribution and loan amounts).**
 - **Dropped all rows missing credit history info as this is integral to deciding on whether or not to grant someone a lone.**
 - **Conducted log transformation on numeric values to make distribution more normal and deal with outliers.**
 - **Looked at scaling and performing PCA and decided against it as PCA made model results worse and scaling seemed to have no affect post log transformation.**
-

EDA - VISUALIZATIONS



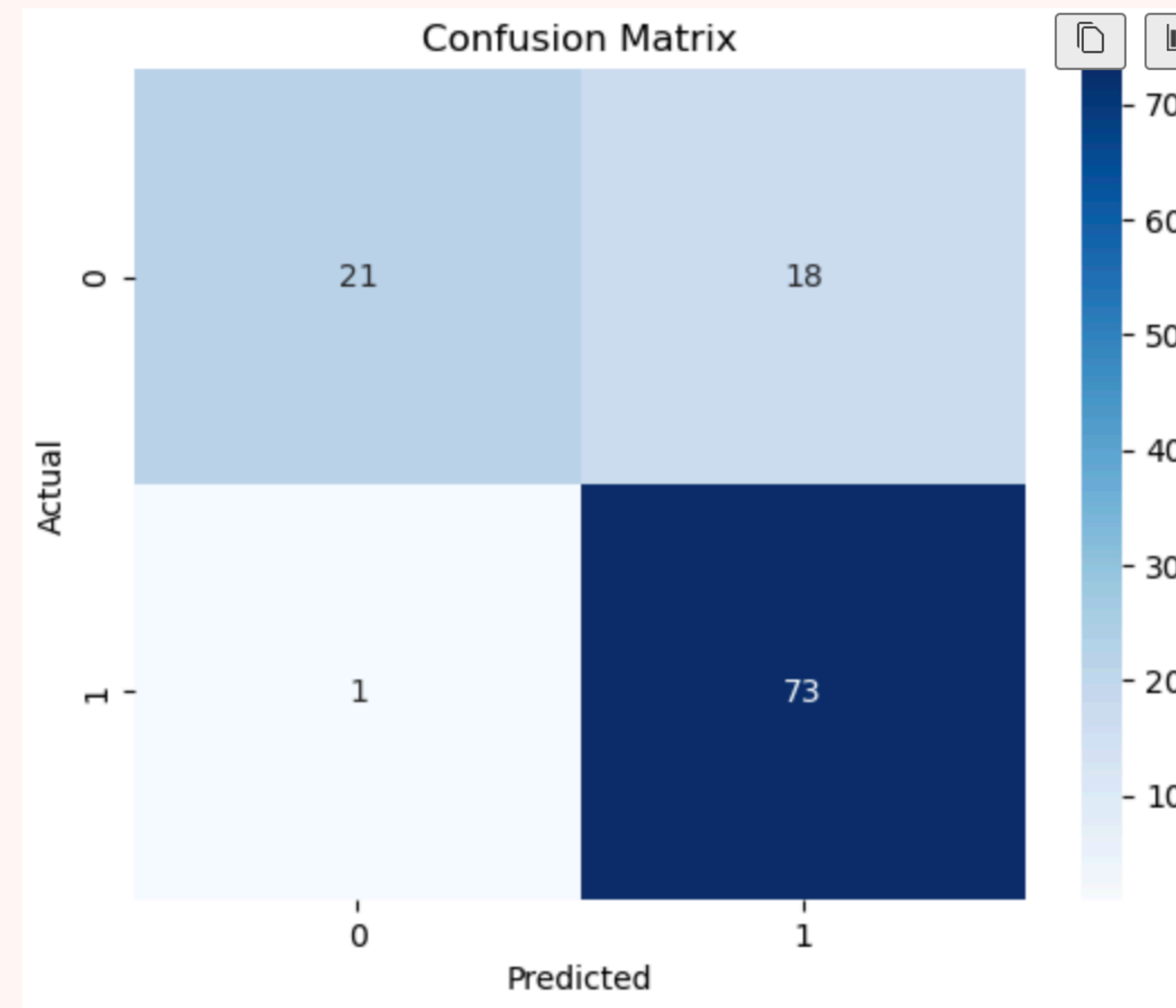
EDA - VISUALIZATIONS



MODEL RESULTS - XGBOOST

➤ Looked at a few different options and seem to get the same results. Went with XGBoost as it produced promising evaluation statistics as well as running much quicker than other models tested.

➤ **Accuracy Score: 0.8319**
Precision Score: 0.8022
Recall Score: 0.9865
F1 Score: 0.8848
ROC-AUC Score: 0.7625



DEPLOYING TO CLOUD & CONNECTING TO API

- **Save as a pickle file to reload the XGBoost model from.**
 - **Initialize Flask Application.**
 - **Send POST request via JSON to make prediction.**
 - **Get returned prediction!**
-

CHALLENGES

- **Issues getting everything put into a pipeline. While things ran independently when trying to combine all of the steps into a single pipeline a lot of debugging was required.**
 - **Time. Only having one day to complete a project this size was a big challenge.**
 - **AWS. Lost key pair location so trying to reconnect caused some issues.**
-

INSIGHTS

- **We can fairly accurately predict loan approvals with Machine Learning Models.**
 - **Machine Learning Models can be leveraged by many different industries for various predictive capabilities.**
 - **APIs are an effective way to interact with cloud deployed Machine Learning Models. It makes sense why this would be done in commercial applications.**
-

FUTURE GOALS

- **Run with more data. Overall dataset for training and testing was quite small. Would prefer more data to hopefully improve the overall accuracy of the model.**
 - **Edit pipelines to make code more efficient.**
 - **Look at a similar but more complex project. This was a fun project that lead me to think about what else might be possible. Could be interesting to see if you could develop a complex ML model to predict if a firm would be acquired by another for example.**
-