Credit Card Customers

Predict Churning customers

Abeer Alahmadi

Problem Statement

Problem

Credit card customer churn in the Bank sector, whose customers are most likely to churn and leave the bank.

Why

Knowing in advance customers churn, to maintain them by giving them special offers.

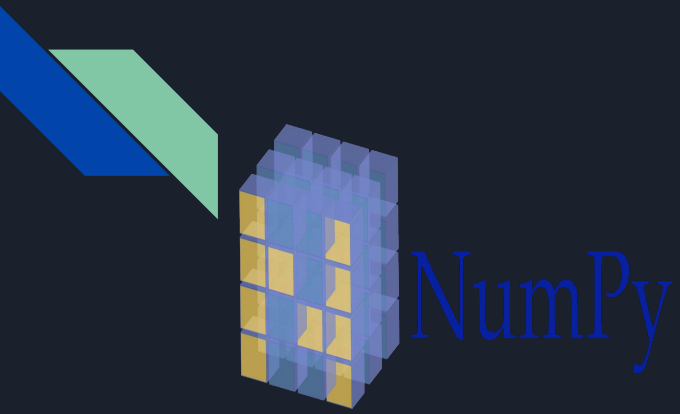
Dateset

- Dataset from Kaggle
- Contains 10127 records and 21 columns

CLIENTNUM	Attrition_Flag	Customer_Age	Gender	Dependent_count	Education_Level	Marital_Status	Income_Category	Card_Category	Months_on_book
802013583	Existing Customer	56	M	3	College	Married	$120K +	Blue	50
711887583	Attrited Customer	47	M	2	Unknown	Married	$80K-$ 80K−120K	Blue	53

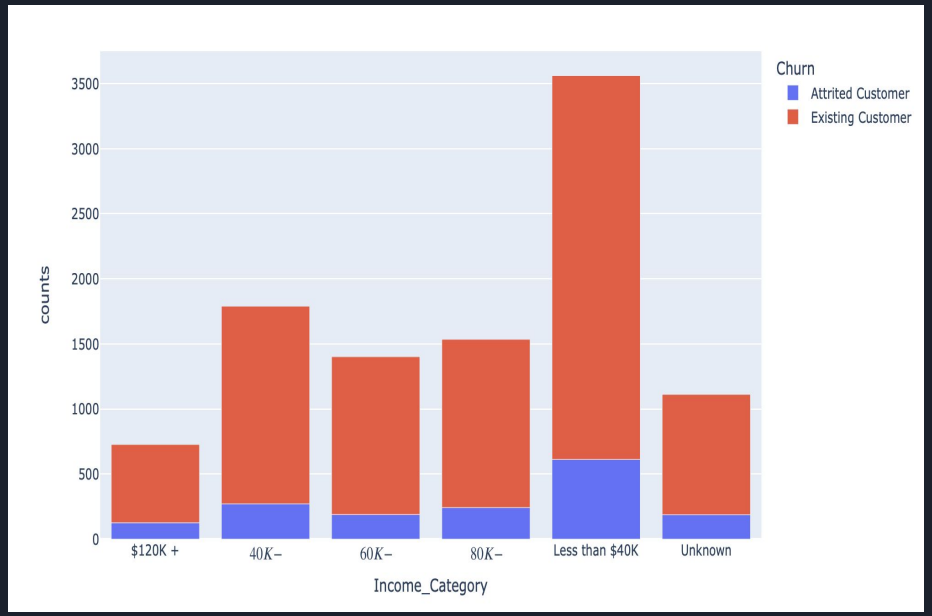
Months_Inactive_12_mon	Contacts_Count_12_mon	Credit_Limit	Total_Revolving_Bal	Avg_Open_To_Buy	Total_Amt_Chng_Q4_Q1	Total_Trans_Amt	Total_Trans_Ct	Total_Ct_Chng_Q4_Q1	Avg_Utilization_Ratio
2	0	9689.0	2250	7439.0	0.576	1158	19	0.727	0.232
3	3	5449.0	1628	3821.0	0.696	836	18	0.385	0.299

Tools

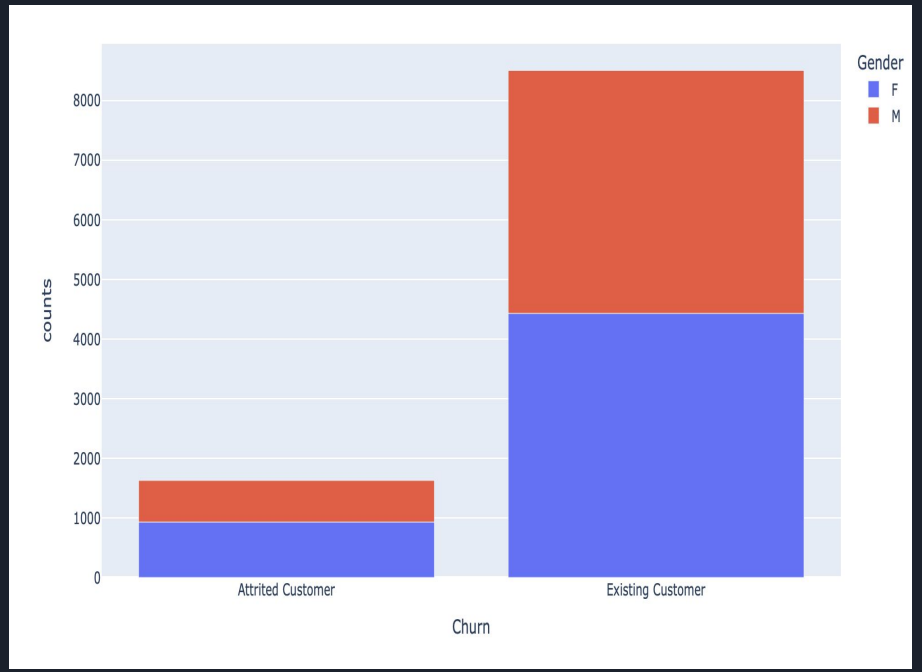


EDA

Churn Based on Income



Churn Based on Gender

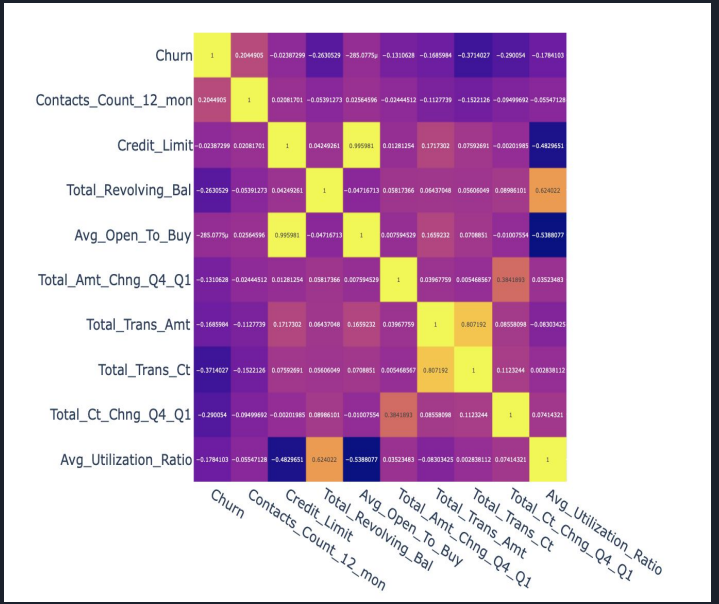


Data Processing

- Rename Columns:

```
bank_df.rename(columns={
   'CLIENTNUM': 'ID',
   'Attrition_Flag':'Churn',
   'Customer_Age':'Age',
   'Dependent_count':'Dependents'
},inplace=True)
```

- Create with dummy variables.
- Measure the correlation between columns.
- Feature Selection using (Recursive Feature Elimination
  -Given an external estimator that assigns weights to features



Over-sampling using SMOTE

```
length of oversampled data is  11330
Number of non-churn in oversampled data 5665
Number of churn 5665
Proportion of non-churn data in oversampled data is  0.5
Proportion of churn data in oversampled data is  0.5
```

Model Training & Comparison

Model	Accuracy	AUC	Recall	Cross validation
Logistic Regression	0.82	0.8698	0.7534	0.84
Naive Bayes	0.80011968	0.8985	0.8323	0.85
KNN	0.7687013	0.7781	0.6528	0.85

Conclusion

To sum up , i recommend in the future adding a new variable with the number of transactions done be the customers using credit cards, it going to help predicting the churn customers.

Thank you