Customer Churn Prediction for a Retail Chain

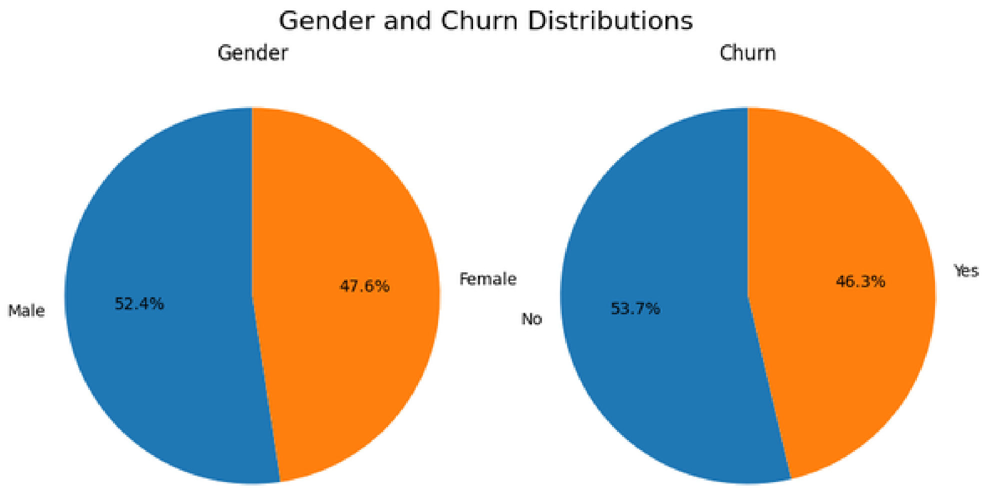
Presented By

AKSH SAINI

Data Analysis:

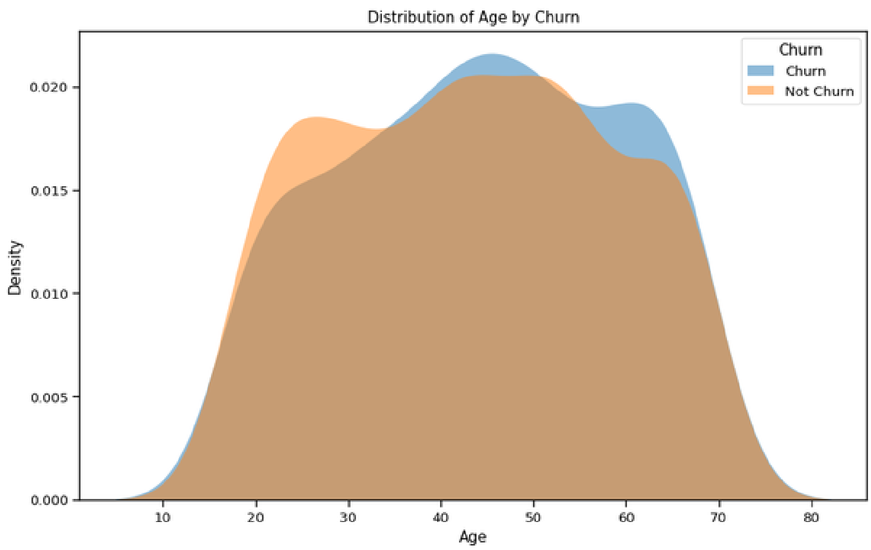
In my exploratory data analysis, I delved into the distribution of features and their connection to the target variable, churn. Through visualizations like Kernel Density Estimate (KDE) plots, pie charts etc , I observed distinct patterns in feature distributions and potential correlations with churn, offering insights into significant predictors of customer churn behavior.

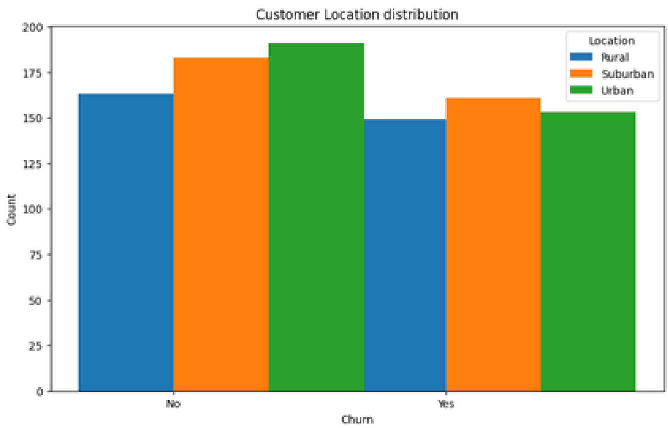
Gender Distribution:



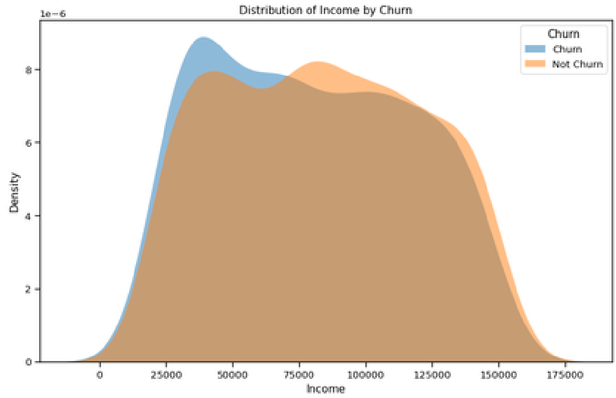
Consider almost equal split between male and female, this data may or may not help with model prediction

- People of age in the range 40 - 50 and 60 - 70 have a higher churning number
- While people with young age ie between 20 - 30 tend to stay .

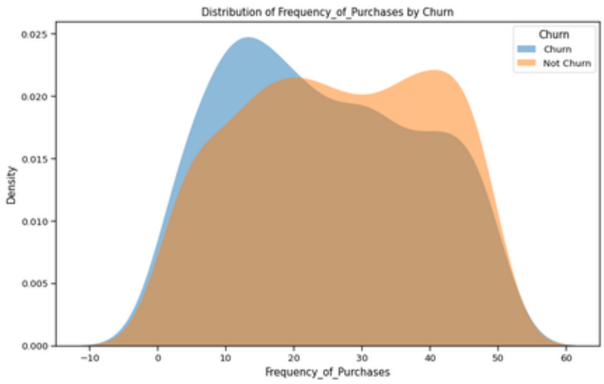




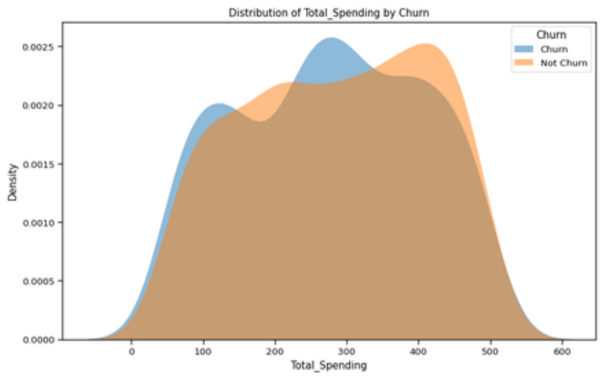
- People living in suburban area have a slight more number in terms of churing



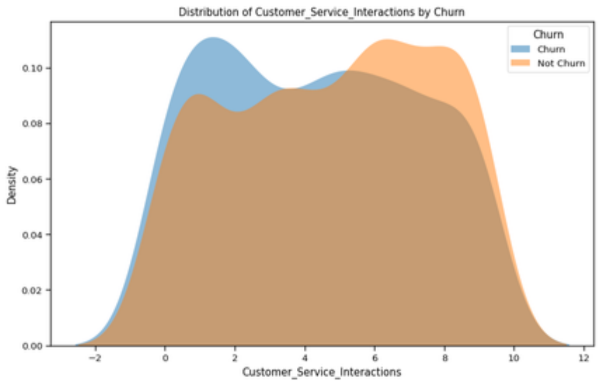
- People with income in the range of around Rs 0 - 50000 are more likely to churn



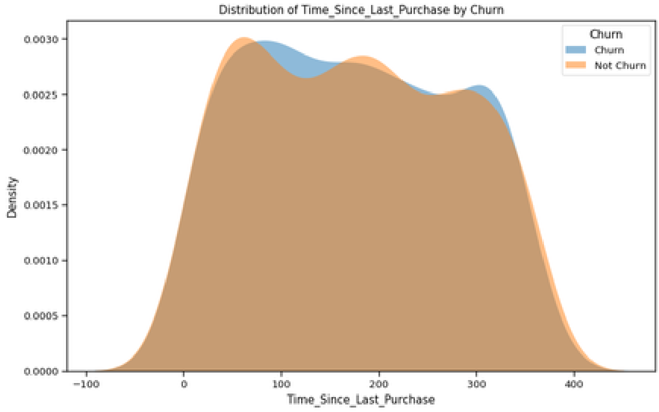
- Customers who are maybe new and have less frequency of purchase have more churning number than those who are old customers which may give information of not satisfactory performance in recent years



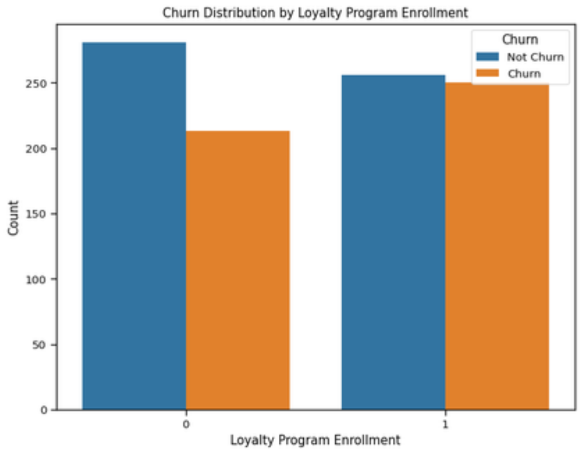
- Customers around 200 - 400 spending are have a higher churning number



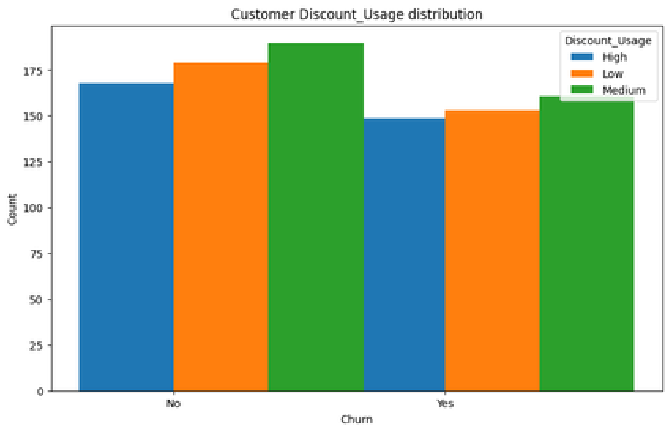
- Customers with less interaction are likely to churn more which implies by increasing customer services interaction, number of people churning may decrease



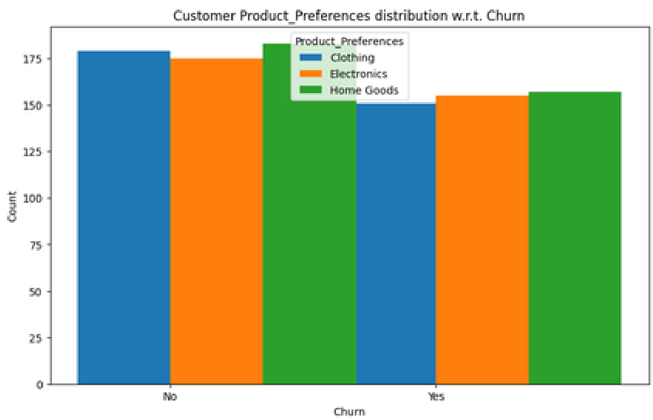
- Almost equal in terms of the Last Purchase
- Doesn’t affect much for predicting



- Customers who have the loyalty program tend to churn more indicating the poor performance Retail Chain’s Loyalty Program



- Customers with medium discount usage are more in both the cases



- Almost equal preference

Model Selection Rationale, Evaluation Metrics:

Model Used for training the dataset:

- AdaBoost
- K - Nearest Neighbour
- Logistic Regression
- Decision Trees
- Random Forests
- Gradient Boosting

Here are results

KNN:
Accuracy: 0.5100
Precision: 0.4706
Recall: 0.4301
F1 Score: 0.4494

AdaBoost:
Accuracy: 0.4750
Precision: 0.4268
Recall: 0.3763
F1 Score: 0.4000

Random Forest:
Accuracy: 0.5000
Precision: 0.4507
Recall: 0.3441
F1 Score: 0.3902

Gradient Boosting:
Accuracy: 0.4800
Precision: 0.4337
Recall: 0.3871
F1 Score: 0.4091

Logistic Regression:
Accuracy: 0.4900
Precision: 0.4400
Recall: 0.3548
F1 Score: 0.3929

Decision Tree:
Accuracy: 0.4600
Precision: 0.4157
Recall: 0.3978
F1 Score: 0.4066

Insights

- As per the results most of the classifiers are giving a accuracy of around 50 %
- This is because most of the data (as shown above in the graphs) does not show a very significant difference. For example:
  - Time since last purchase
  - Location Distribution
  - Product Preference
  - Discount usage etc

Data for the above subpoints does not vary much which makes it difficult for our model to predict with a better accuracy in this specific case.

- Among all the classifiers KNN (K - nearest neighbour performed best with better F 1 score amongst all

Deployment Recommendations.

- Real-time Integration: Integrate the model into existing systems for prompt churn prediction, enabling proactive customer retention measures.
- API Deployment: Deploy the model as an API for seamless integration with various systems and applications.
- Scalability Considerations: Ensure scalability by utilizing cloud-based solutions to handle increasing data volumes.
- Monitoring and Maintenance: Implement continuous monitoring and maintenance to sustain model effectiveness over time.
- UI Integration: Integrate model predictions into the user interface for easy access by stakeholders.
- Feedback Loop: Establish a feedback loop to improve the model iteratively based on intervention effectiveness.
-