**RETENTION MODELING and Analysis**

**Section 1: Retention modeling**

## Assumption:

The assumption made is be that the each record will be treated as a new customer even though customer may appear more than once in the dataset because every record is associated with a renewed flag which needs to be predicted and the new feature called ‘frequency of customer’ is deduced to capture the information that associates each customer with itself which tell the prediction model that the customer is same as that of previous one if exist.

A new feature called ‘contract days’ is feature engineered to capture the tenure of the customer.

## Training of Retention Prediction Model:

The retention prediction model is trained on below features:

Table

Description automatically generated

## Prediction Modeling:

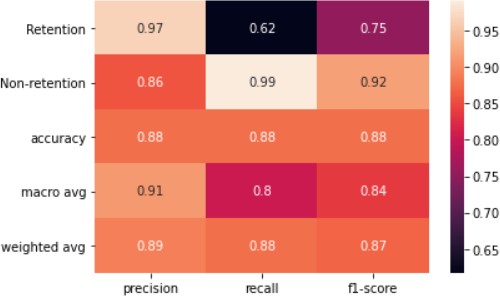
The dataset was spilt into 80-20 ratio for training and testing the model.

Used Random Forest Classifier and Logistic Regression to predict the class in which an employee may fall into.

**Performance of Retention Prediction Model:**

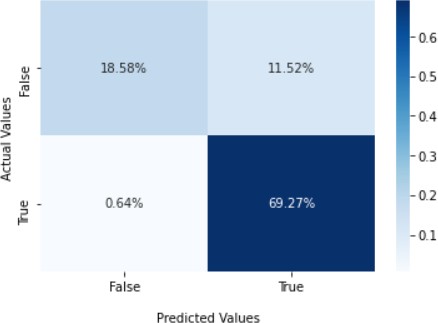
The retention model accuracy turned out to be 87 % which is a way of assessing the performance of a model which implies how well the models can predict the retention of the customers correctly.

|  |  |
| --- | --- |
| **Metrics** | **Score** |
| Accuracy | 0.87 |
| Precision | 0.85 |
| Recall | 0.99 |



## Error Rate in of Retention Prediction Model:

The model prediction shows that Out of total percent of sample, only 0.64% were detected as Type I error (False positive) while 11.52% were detected as Type II error.

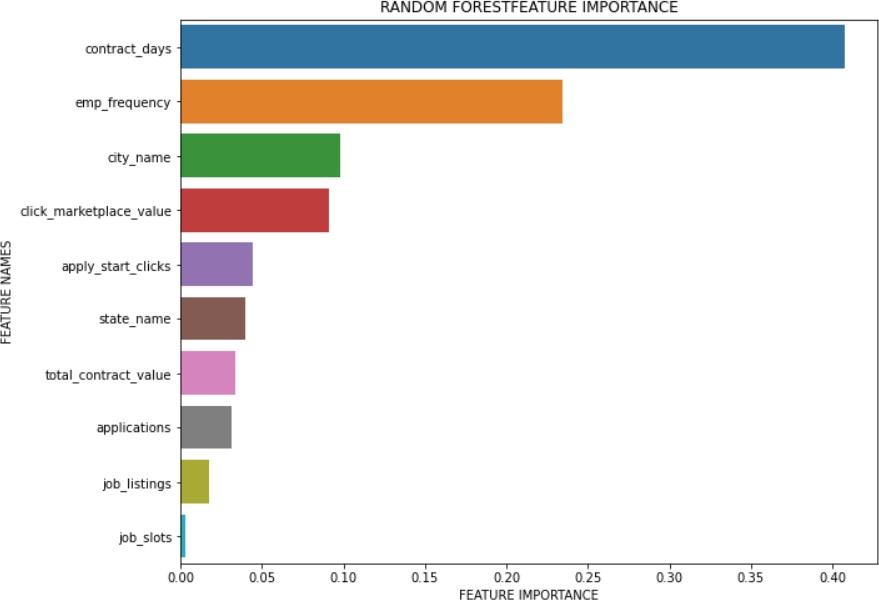


## Factors best predict an employer's likelihood to retain:

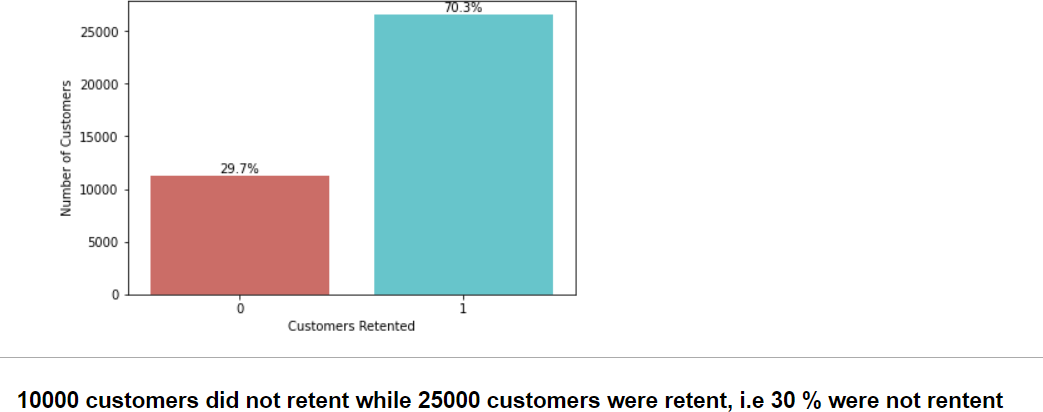
The graph below shows the factors which best predicts an employer’s likelihood to retain with the importance of each feature. The major feature that matters are tenure of the customer’s contract (contract days) followed by feature of how many times customer is repeated in dataset (‘frequency of customer’ ) .

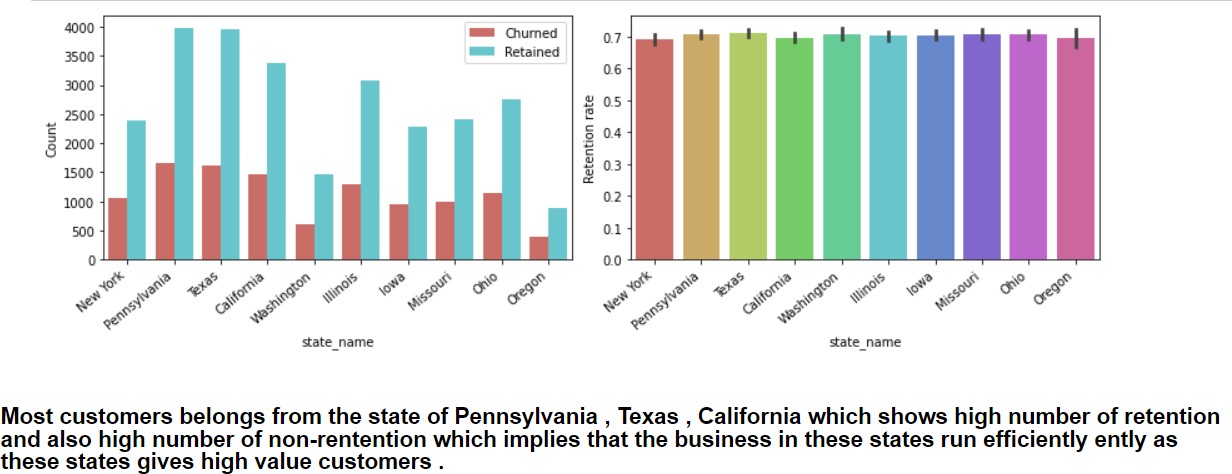
### The feature/factors which matters with their importance score:

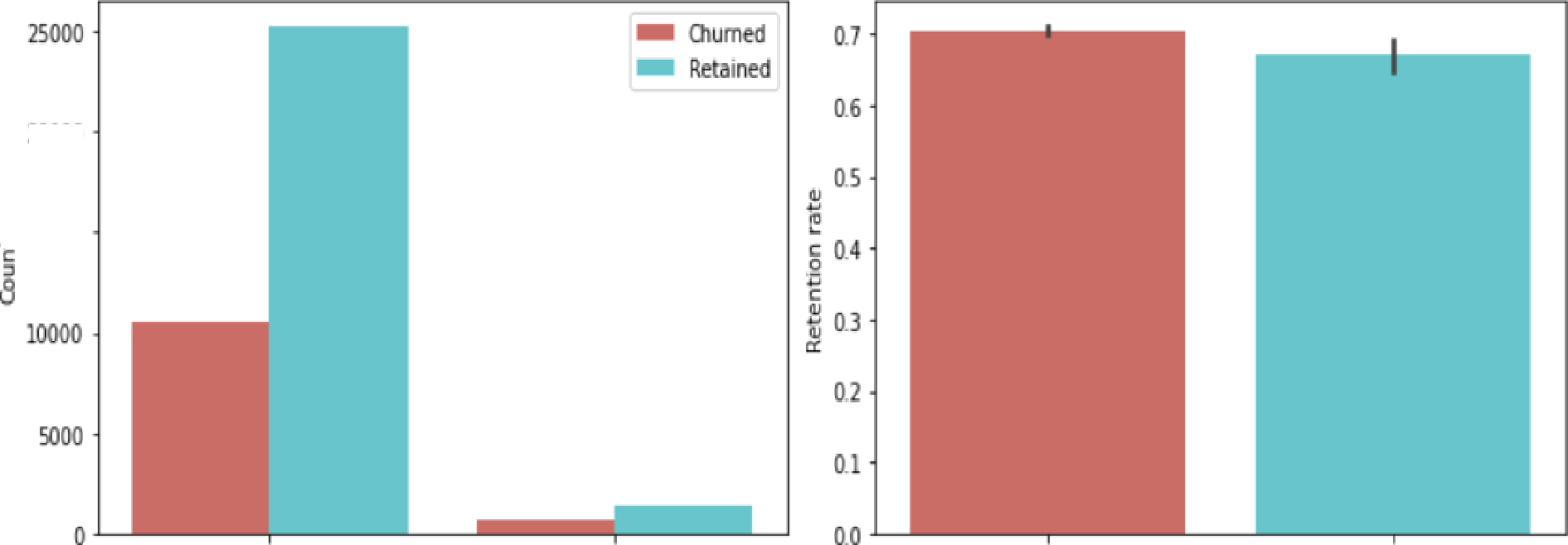
|  |  |
| --- | --- |
| contract\_days | 0.41 |
| emp\_frequency | 0.23 |
| city\_name | 0.10 |
| click\_marketplace\_value | 0.09 |
| apply\_start\_clicks | 0.04 |
| state\_name | 0.04 |
| total\_contract\_value | 0.03 |
| Applications | 0.03 |
| job\_listings | 0.02 |
| job\_slots | 0.00 |



**Section 2 : DATA ANALYSIS & RECOMMENDATION**

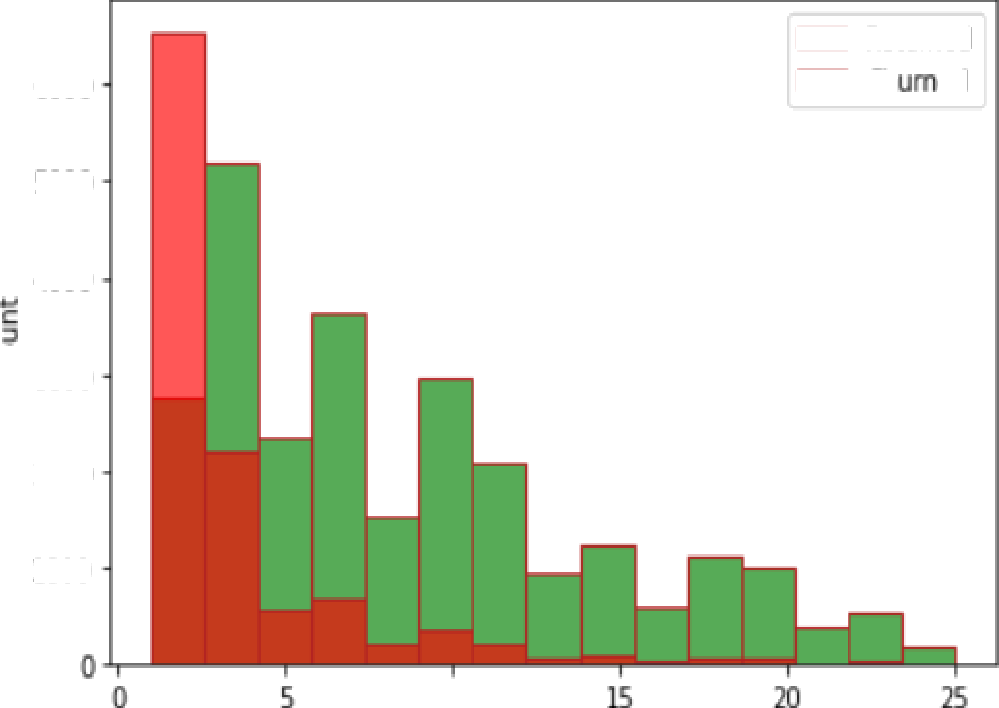






pb slots

The customers having 15 job slots has more probability to retain than the customers having 50 job slots.



6000

Retained Ch ed

5000

4000

3000

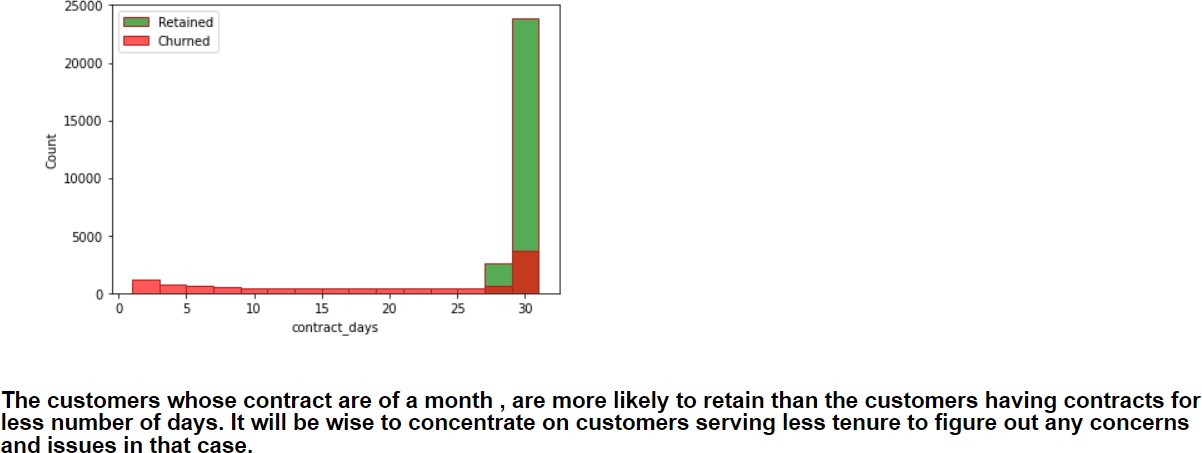
2000

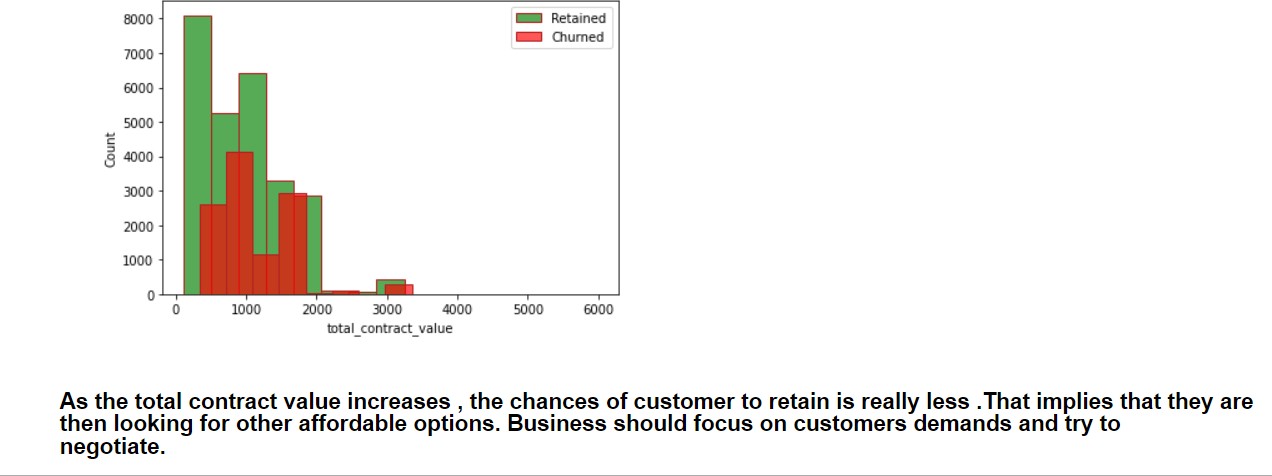
1000

10

same customer across country

The graph implies that If the customers are taking the products across the country they are less likely to retain. Such frequent customers are on red alert and should be paid more attention to .





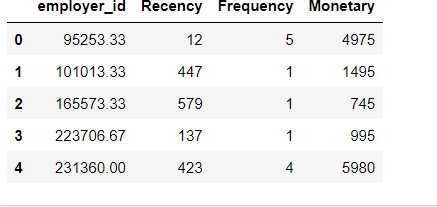
# Section 3: Segmentation & Recommendation:

### The customers should be segmented and behavioral study like below should be done to focus on customers appropriately and target them because every customer is different.

A behavioral segmentation by Recency, Frequency, and Monetary is done to segment the customers to focus appropriately. Based on the data, new features were added, including the following:

Recency: Number of days since the last purchase.

Frequency: Number of transactions made over a given period. Monetary: Amount spent over a given period of time.



These features can be used to calculate RFM score (Recency, Frequency, Monetary) which can segment customers into ‘high value customers’, ‘lost customers’, ‘Medium value customer’ to target better.

The modeling results into three segments of customers shown below:

* Lost customers can be win back by giving incentives and taking surveys.
* High value customers should be Provided with Value-Packed Content That Keeps Customers Engaged and Offer them a Personalized Experience to build relationships , Encouraging Customers to Switch to an Annual Billing Cycle , cross-selling and up scaling to increase revenue .

