# Data Exploration

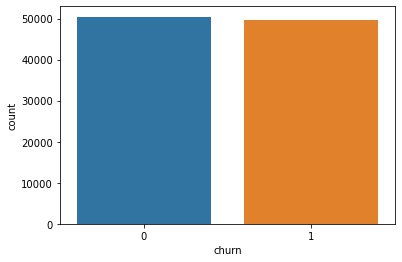
The following document provides a breakdown of the data for telecom churn customers wrt various fields

## Structure of the Data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | A | B | C | D |
| 1 | Data Type | Number of Columns | Number of Rows | % of Null Values |
| 2 | Numerical Columns | 79 | 100,000 |  |
| 3 | Categorical Columns | 21 | 100,000 |  |
| 4 | Total | 100 | 100,000 |  |
| 5 |  |  |  |  |
| 6 |  |  |  |  |

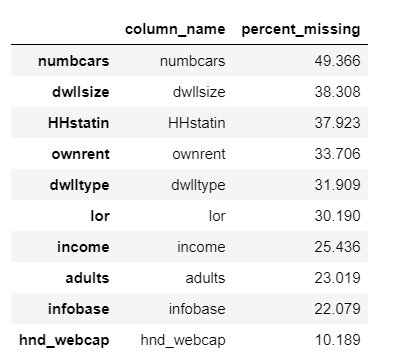
## Distribution of Churn and Not Churned Customers

This is an example of balanced dataset with equal proportion of churn and not churned customers.



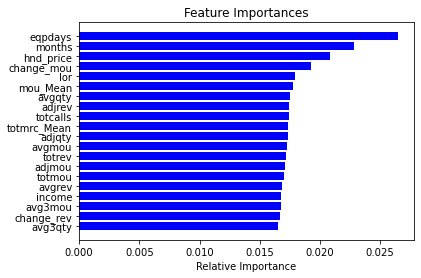
## Missing Values in the columns

Since NumbCars, dwlsize, HHSTatin and ownrent have more than >30% of missing values, we will be removing them from our dataset as imputing them won’t add any extra information and might skew the data

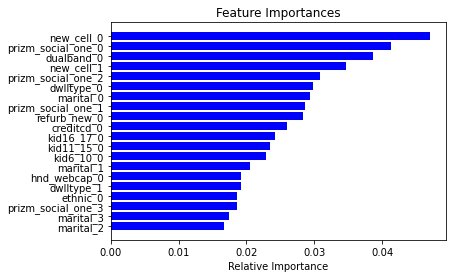


## Important Columns in the Dataset based on Variable importance

Numerical Columns



Categorical Columns



## Correlation of Relevant Numerical Values with Churn

## Correlation of Relevant Categorical Values with Churn