# Data Exploration

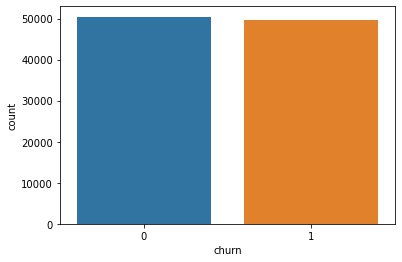
The following document provides a breakdown of the data for telecom churn customers wrt various fields

## Structure of the Data

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

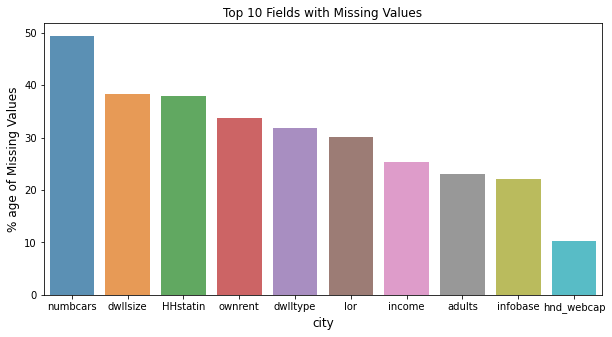
## Distribution of Churn and Not Churned Customers

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

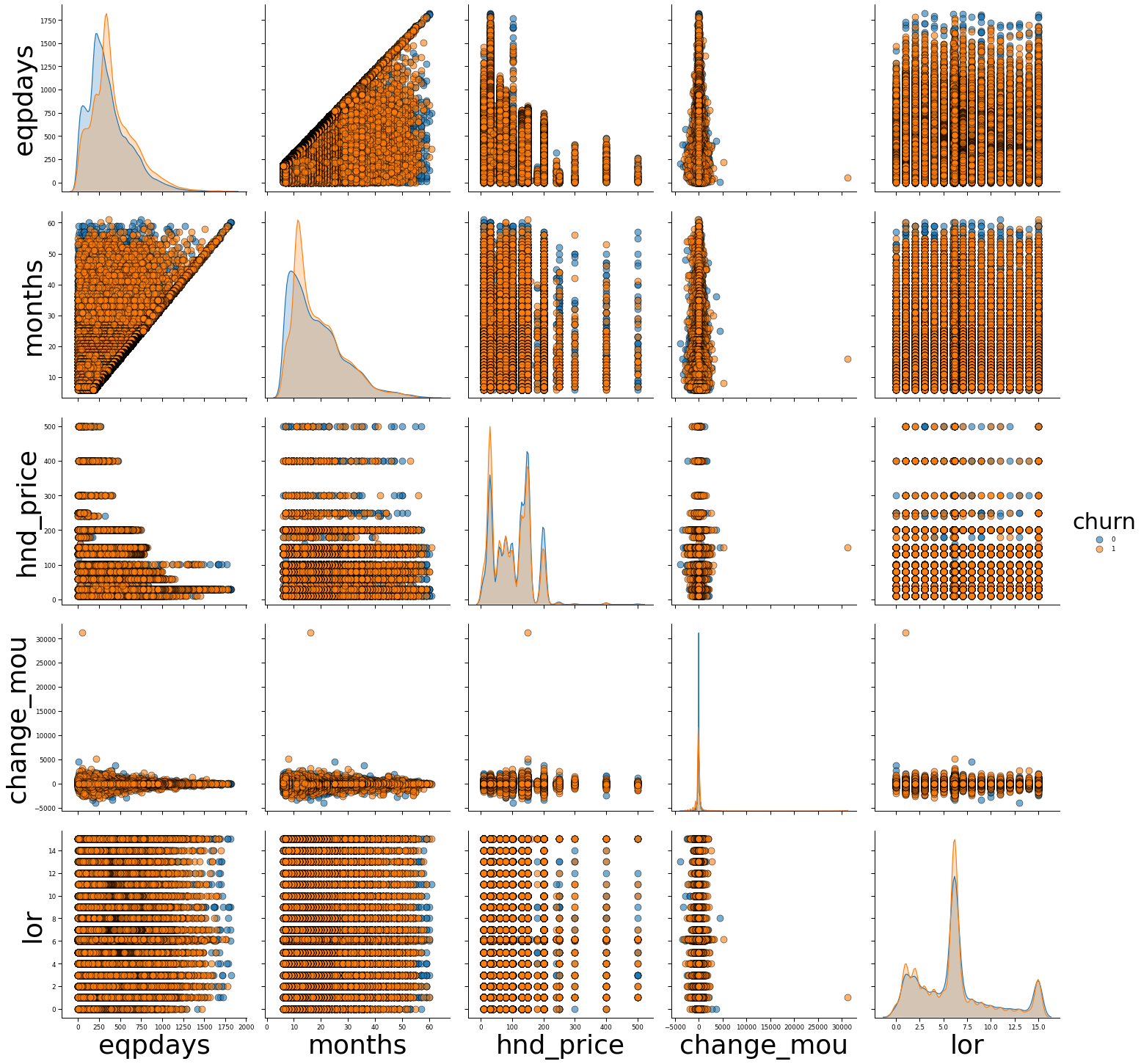


## Missing Values in the columns

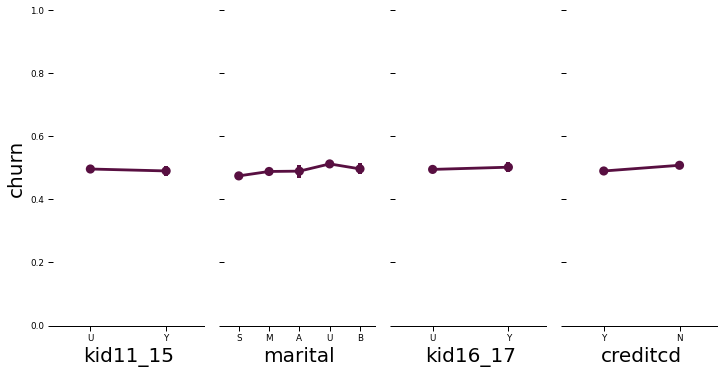
Since NumbCars, dwlsize, HHSTatin and ownrent have more than >30% of missing values, we will be removing them from our dataset as imputing them won’t add any extra information and might skew the data

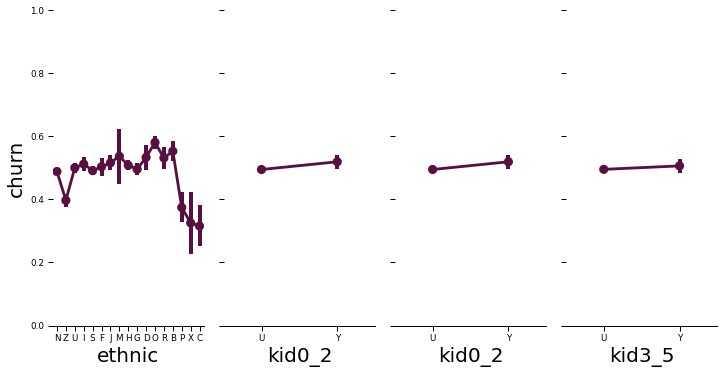


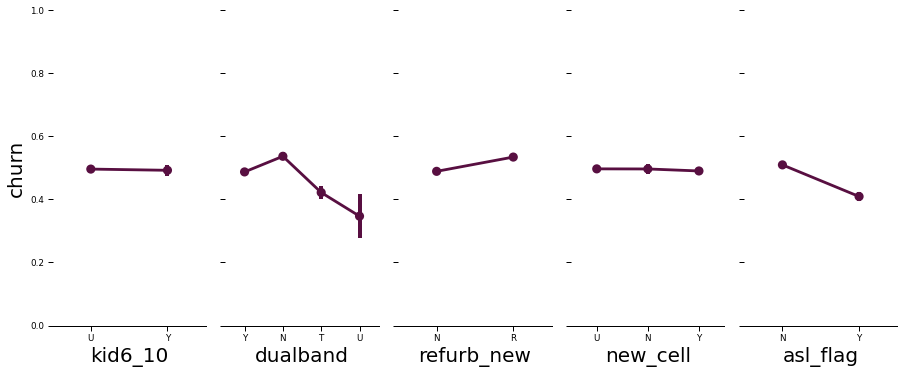
## Correlation of Relevant Numerical Values with Churn



## Relationship of Categorical Values with Churn







## Appendix

**Data Dictionary**

Distribution of Numerical Values



Categorical Variables

|  |  |  |
| --- | --- | --- |
| column\_name | Column Meaning | percent\_missing |
| dwllsize | Dwelling size | 38.308 |
| HHstatin | Premier household status indicator | 37.923 |
| ownrent | Home owner/renter status | 33.706 |
| dwlltype | Dwelling Unit type | 31.909 |
| infobase | InfoBase match | 22.079 |
| hnd\_webcap | Handset web capability | 10.189 |
| prizm\_social\_one | Social group letter only | 7.388 |
| marital | Marital Status | 1.732 |
| ethnic | Ethnicity roll-up code | 1.732 |
| kid16\_17 | Child 16 - 17 years of age in household | 1.732 |
| kid11\_15 | Child 11 - 15 years of age in household | 1.732 |
| kid6\_10 | Child 6 - 10 years of age in household | 1.732 |
| kid3\_5 | Child 3 - 5 years of age in household | 1.732 |
| kid0\_2 | Child 0 - 2 years of age in household | 1.732 |
| creditcd | Credit card indicator | 1.732 |
| area | Geogrpahic area | 0.04 |
| refurb\_new | Handset: refurbished or new | 0.001 |
| dualband | Dualband | 0.001 |
| crclscod | Credit class code | 0 |
| asl\_flag | Account spending limit | 0 |
| new\_cell | New cell phone user | 0 |