Retaining Telecom Customers

Troy Mayrand & Zachary Safir

Questions Asked

Given the data:

- Can we find the most "at risk" customers that we need to focus on?

- If there is a key "at risk" group, what defines this group?
  - Can we model it?

How can we hold on to these customers?

The Data

- 7043 Customers

- 21 Customer Attributes
  - 18 Categorical
  - 3 Numerical

- 11 Missing values

- Churn is turnover

```
customerID          gender        SeniorCitizen     Partner       Dependents    tenure              PhoneService
0002-ORFBO:   1     Female:3488   0:5890            No :3641      No :4933      Min.   : 0.00       No : 682
0003-MKNFE:   1     Male  :3555   1:1142            Yes:3402      Yes:2110      1st Qu.: 9.00       Yes:6361
0004-TLHLJ:   1                                                                Median:29.00
0011-IGKFF:   1                                                                Mean   :32.37
0013-EXCHZ:   1                                                                3rd Qu.:55.00
0013-MHZWF:   1                                                                Max.   :72.00
(Other)   :7037

MultipleLines          InternetService     OnlineSecurity                 OnlineBackup
No          :3390      DSL       :2421     No              :3498          No              :3088
No phone service: 682  Fiber optic:3096    No internet service:1526       No internet service:1526
Yes         :2971      No        :1526     Yes             :2019          Yes             :2429

DeviceProtection               TechSupport                    StreamingTV                    StreamingMovies
No              :3095          No              :3473          No              :2810          No              :2785
No internet service:1526       No internet service:1526       No internet service:1526       No internet service:1526
Yes             :2422          Yes             :2044          Yes             :2707          Yes             :2732

Contract                PaperlessBilling   PaymentMethod                        MonthlyCharges      TotalCharges
Month-to-month:3875     No :2872           Bank transfer (automatic):1544       Min.   : 18.25      Min.   : 18.8
One year      :1473     Yes:4171           Credit card (automatic)  :1522       1st Qu.: 35.50      1st Qu.: 401.4
Two year      :1695                        Electronic check         :2365       Median : 70.35      Median :1397.5
                                           Mailed check             :1612       Mean   : 64.76      Mean   :2283.3
                                                                                3rd Qu.: 89.85      3rd Qu.:3794.7
                                                                                Max.   :118.75      Max.   :8684.8
                                                                                                    NA's   :11

Churn
No :5174
Yes:1869
```
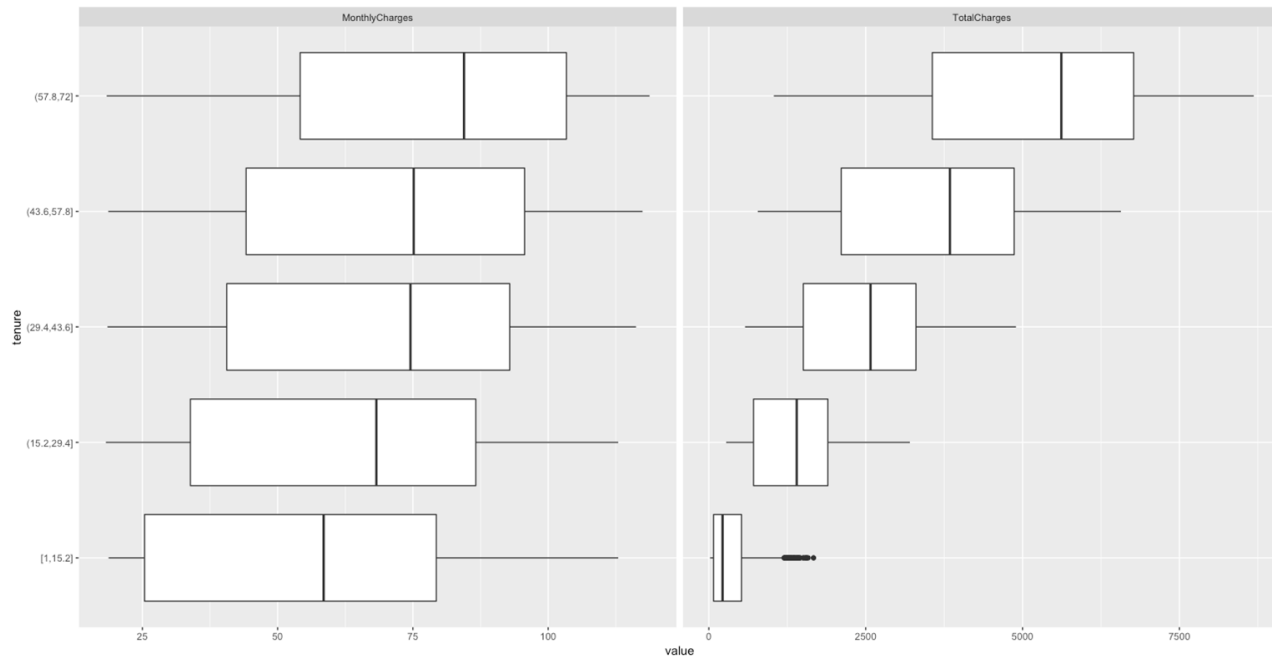
Data Preprocessing I

- Those 11 missing values?
  - Only occur in `totalCharges`
  - `na.omit()`

- Manually convert `SeniorCitizen` to a factor because it was numeric by default.
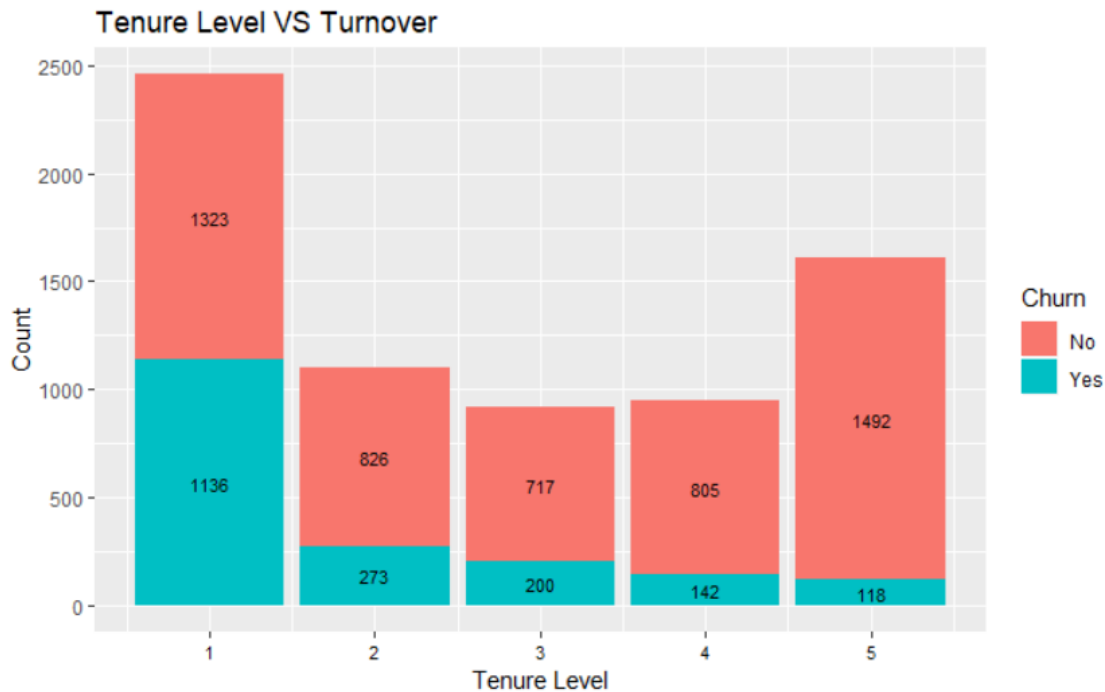
Exploration I

Both `MonthlyCharges` and `TotalCharges` values increase the longer a person has been with the company.

Exploration II

Groups of clients with longer time with the Telecom turn over less.

- Each "level" is about 15 months.

Tenure Level VS Turnover

Identifying the At Risk Group

- With the majority of turnover occuring in the first 15 months.
  We believe the behavior we want to identify primarily exists within this group.

- We created a subset of our data to focus on this window.

R Modeling

Given this identified group:

- We found the three best regression models in R (using glmulti) based on AIC scores

- We can narrow our scope to:
  ```
  SeniorCitizen, MultipleLines, InternetService, OnlineSecurity,
  OnlineBackup, TechSupport, Contract, PaperlessBilling,
  PaymentMethod, tenure, MonthlyCharges, TotalCharges
  ```

Odds Ratios I

From the model, we can look at computed odds for each variable.

We can pick out `Contract` as a key factor.

```
                                        Odds Ratio       2.5 %      97.5 %
(Intercept)                            4.592691e-01 2.029058e-01 1.034991896
SeniorCitizen1                         1.300938e+00 9.973395e-01 1.702429177
MultipleLinesNo phone service          2.897344e+00 1.902196e+00 4.428399995
MultipleLinesYes                       1.359694e+00 1.062584e+00 1.741105491
InternetServiceFiber optic             1.495077e+00 9.643898e-01 2.319625996
InternetServiceNo                      8.046496e-01 4.864395e-01 1.332738988
OnlineSecurityNo internet service               NA           NA          NA
OnlineSecurityYes                      5.507624e-01 4.127503e-01 0.732842994
OnlineBackupNo internet service                 NA           NA          NA
OnlineBackupYes                        7.008759e-01 5.413774e-01 0.906865113
TechSupportNo internet service                  NA           NA          NA
TechSupportYes                         6.955157e-01 5.141790e-01 0.939566728
ContractOne year                       3.515974e-01 1.892778e-01 0.612422942
ContractTwo year                       3.829072e-07 1.007455e-48 0.002460312
PaperlessBillingYes                    1.479569e+00 1.208416e+00 1.811044759
PaymentMethodCredit card (automatic)   6.708471e-01 4.464282e-01 1.005967529
PaymentMethodElectronic check          9.300973e-01 6.751238e-01 1.279404429
PaymentMethodMailed check              7.405657e-01 5.287816e-01 1.038012747
tenure                                 8.476019e-01 7.963772e-01 0.900113990
MonthlyCharges                         1.018981e+00 1.004192e+00 1.034145938
TotalCharges                           1.000881e+00 1.000003e+00 1.001778412
```
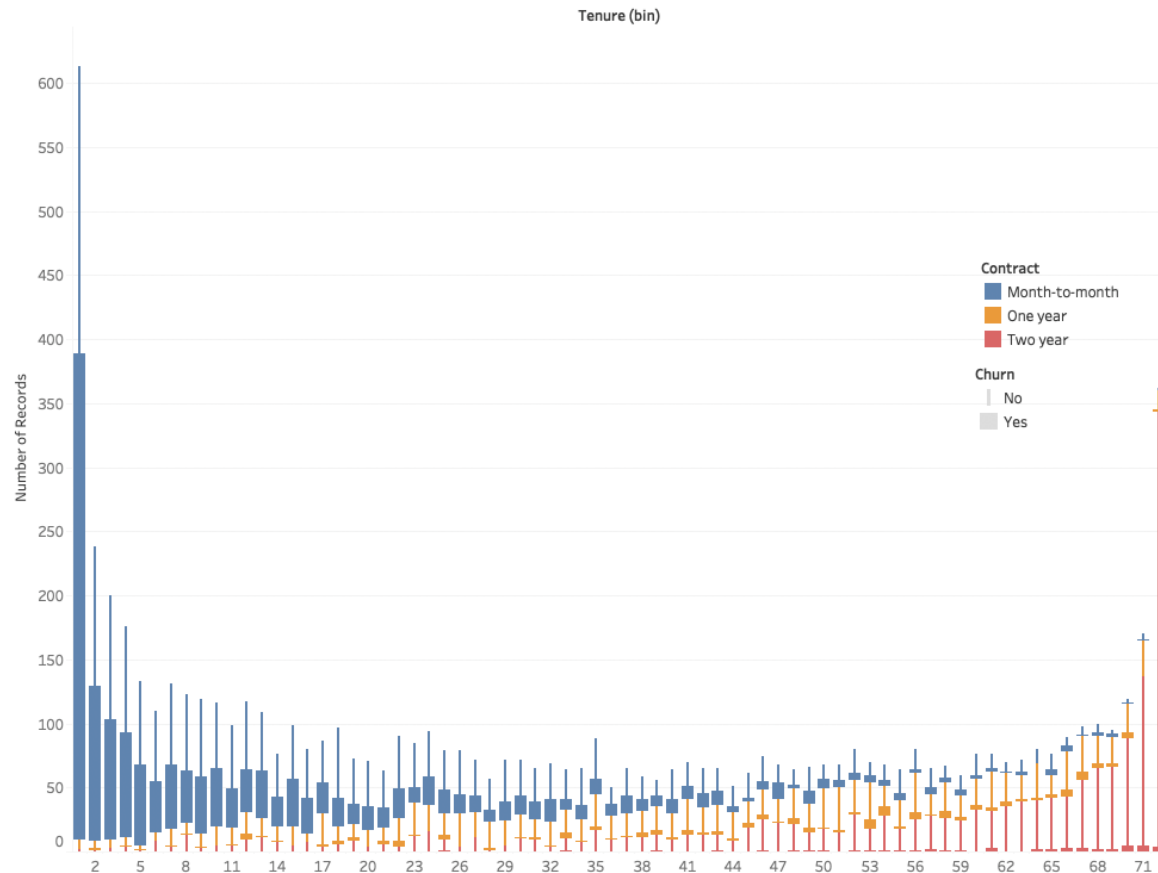
Exploration III

96% of customer turnover occurs on Month-to-month contracts.

60% of documented customer turnover occurs within the first 15 months.
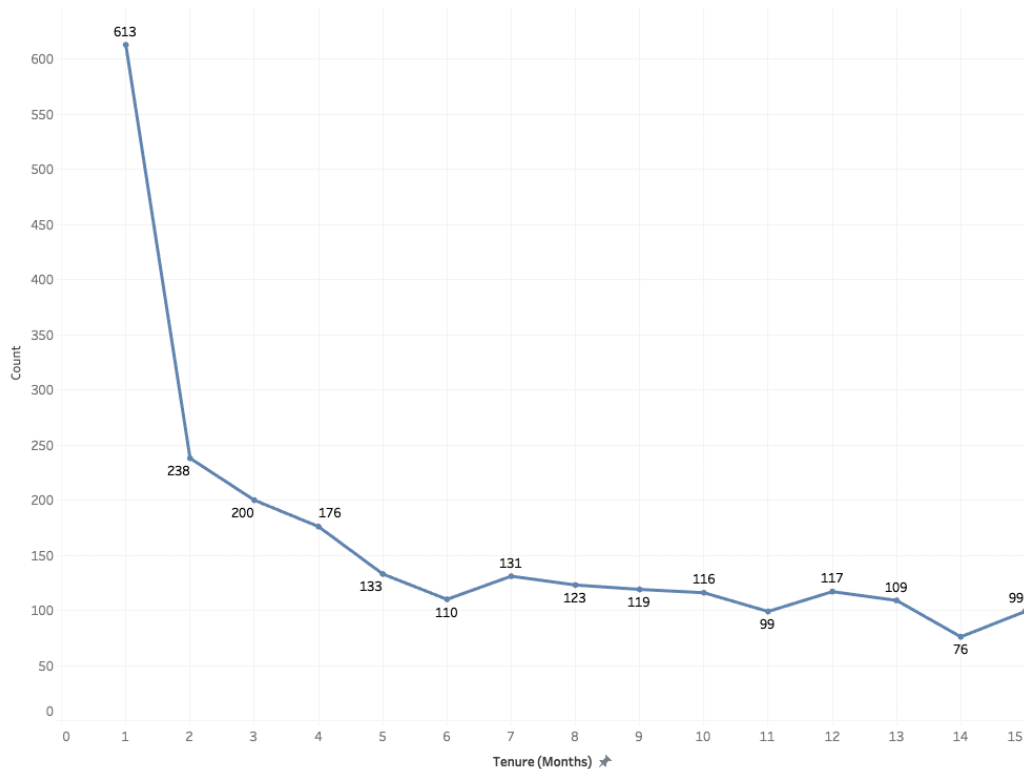

Churn by Contract Type by Tenure

Digging Deeper

Most turnover happens on the first month.

Followed by a sharp decline, and slower decrease.

First Tenure Level Churn

Data Processing II

Given the first month turnover:

- We took another subset focusing on these customers.

- Further refinement of variable scope:
    ```
    SeniorCitizen, PhoneService, OnlineSecurity, TechSupport,
    Contract, PaperlessBilling, TotalCharges
    ```

Odds Ratios II

After the first month the customers most likely to stay all have:

- PhoneService
- TechSupport
- OnlineSecurity

Those that do not are more likely to leave.

	Odds Ratio	2.5 %	97.5 %
(Intercept)	4.074581e-01	0.18655343	8.985406e-01
SeniorCitizen1	2.153155e+00	1.11218613	4.460473e+00
PhoneServiceYes	2.084525e-01	0.08901276	4.763333e-01
OnlineSecurityNo internet service	2.016002e+00	0.90497516	4.531350e+00
OnlineSecurityYes	4.324523e-01	0.20266377	9.145335e-01
TechSupportNo internet service	NA	NA	NA
TechSupportYes	2.899332e-01	0.12916060	6.359916e-01
ContractOne year	2.539274e-07	NA	1.051647e+14
ContractTwo year	3.634430e-07	NA	5.092188e+63
PaperlessBillingYes	1.697654e+00	1.14807985	2.507677e+00
TotalCharges	1.051597e+00	1.03423474	1.070170e+00

Questions Answered

How can Telecom hold on to these customers?

- Avoid Month-to-month contracts.
  - Incentivise One/Two Year options

- If stuck with a Month-to-month option:
  - Push sign ups for phone service
  - Gently remind clients of tech support options
  - Get clients to use/sign up for online security

Reflections

- **Analysis**
  - Working with mostly categorical data was a new experience for us.
  - Retrospectively we isolated the most volatile segment.
    - Additional work required for subtlety of longer contract types.

- **Dataset**
  - Generating models took longer than expected.
  - Subsetting data can be very important