TelecomChurnCapstone

Nirmal Patel

October 5, 2018

Read in the Telecom Churn Data

```
setwd("C:/Users/NP/Desktop/SPRINGBOARD/Caspstone telecom")
telecom <- read.csv("WA_Fn-UseC_-Telco-Customer-Churn.csv", header=T)
```

The telecom churn csv file has been read into R and renamed into telecom.

```
library("ggplot2")
library("readxl")
library("rmarkdown")
```

This data looks clean with well defined variable names.

Structure

```
str(telecom)
```

```
## 'data.frame':    7043 obs. of  21 variables:
##  $ customerID      : Factor w/ 7043 levels "0002-ORFBO","0003-MKNFE",..:
5376 3963 2565 5536 6512 6552 1003 4771 5605 4535 ...
##  $ gender          : Factor w/ 2 levels "Female","Male": 1 2 2 2 1 1 2 1 1
2 ...
##  $ SeniorCitizen   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Partner         : Factor w/ 2 levels "No","Yes": 2 1 1 1 1 1 1 1 2 1
...
##  $ Dependents      : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 2 1 1 2
...
##  $ tenure          : int  1 34 2 45 2 8 22 10 28 62 ...
##  $ PhoneService    : Factor w/ 2 levels "No","Yes": 1 2 2 1 2 2 2 1 2 2
...
##  $ MultipleLines   : Factor w/ 3 levels "No","No phone service",..: 2 1 1
2 1 3 3 2 3 1 ...
##  $ InternetService : Factor w/ 3 levels "DSL","Fiber optic",..: 1 1 1 1 2
2 2 1 2 1 ...
##  $ OnlineSecurity  : Factor w/ 3 levels "No","No internet service",..: 1 3
3 3 1 1 1 3 1 3 ...
##  $ OnlineBackup    : Factor w/ 3 levels "No","No internet service",..: 3 1
3 1 1 1 3 1 1 3 ...
##  $ DeviceProtection: Factor w/ 3 levels "No","No internet service",..: 1 3
1 3 1 3 1 1 3 1 ...
##  $ TechSupport     : Factor w/ 3 levels "No","No internet service",..: 1 1
```

```
1 3 1 1 1 1 3 1 ...
##  $ StreamingTV     : Factor w/ 3 levels "No","No internet service",..: 1 1
1 1 1 3 3 1 3 1 ...
##  $ StreamingMovies : Factor w/ 3 levels "No","No internet service",..: 1 1
1 1 1 3 1 1 3 1 ...
##  $ Contract        : Factor w/ 3 levels "Month-to-month",..: 1 2 1 2 1 1 1
1 1 2 ...
##  $ PaperlessBilling: Factor w/ 2 levels "No","Yes": 2 1 2 1 2 2 2 1 2 1
...
##  $ PaymentMethod   : Factor w/ 4 levels "Bank transfer (automatic)",..: 3
4 4 1 3 3 2 4 3 1 ...
##  $ MonthlyCharges  : num  29.9 57 53.9 42.3 70.7 ...
##  $ TotalCharges    : num  29.9 1889.5 108.2 1840.8 151.7 ...
##  $ Churn           : Factor w/ 2 levels "No","Yes": 1 1 2 1 2 2 1 1 2 1
...
```

The sturcture of the telecom company depicts three variables that are integers or numerical type. It would be benificaial if we look into these and see if there are any NA or blank variables. We can also see that most of the data is well organized and has 2 to 4 factors (options for the data)

Summary of Telecom Churn Data

```
summary(telecom)
```

```
##       customerID      gender     SeniorCitizen    Partner    Dependents
##  0002-ORFBO:   1   Female:3488   Min.   :0.0000   No :3641   No :4933
##  0003-MKNFE:   1   Male  :3555   1st Qu.:0.0000   Yes:3402   Yes:2110
##  0004-TLHLJ:   1                 Median :0.0000
##  0011-IGKFF:   1                 Mean   :0.1621
##  0013-EXCHZ:   1                 3rd Qu.:0.0000
##  0013-MHZWF:   1                 Max.   :1.0000
##  (Other)   :7037
##      tenure      PhoneService          MultipleLines     InternetService
##  Min.   : 0.00   No : 682     No              :3390   DSL        :2421
##  1st Qu.: 9.00   Yes:6361     No phone service: 682   Fiber optic:3096
##  Median :29.00                Yes             :2971   No         :1526
##  Mean   :32.37
##  3rd Qu.:55.00
##  Max.   :72.00
##
##              OnlineSecurity              OnlineBackup
##  No                  :3498   No                  :3088
##  No internet service:1526   No internet service:1526
##  Yes                 :2019   Yes                 :2429
##
##
##
##
##             DeviceProtection              TechSupport
##  No                  :3095    No                  :3473
```

```
##  No internet service:1526    No internet service:1526
##  Yes                :2422    Yes                :2044
##
##
##
##
##              StreamingTV              StreamingMovies
##  No                 :2810   No                 :2785
##  No internet service:1526   No internet service:1526
##  Yes                :2707   Yes                :2732
##
##
##
##
##           Contract    PaperlessBilling                   PaymentMethod
##  Month-to-month:3875   No :2872         Bank transfer (automatic):1544
##  One year      :1473   Yes:4171         Credit card (automatic)  :1522
##  Two year      :1695                    Electronic check         :2365
##                                         Mailed check             :1612
##
##
##
##  MonthlyCharges    TotalCharges    Churn
##  Min.   : 18.25   Min.   :  18.8   No :5174
##  1st Qu.: 35.50   1st Qu.: 401.4   Yes:1869
##  Median : 70.35   Median :1397.5
##  Mean   : 64.76   Mean   :2283.3
##  3rd Qu.: 89.85   3rd Qu.:3794.7
##  Max.   :118.75   Max.   :8684.8
##                   NA's   :11
```

The summary function gives us a further break down of the variables inclusing the mean(average), minimum(smallest value), median (middle value), maximum(largest value), and if the variable includes a blank value (NA) this function will let us know. If the variable is a factor, the summary method will give us the total of each factor.

Summary of Telecom Tenure

```
summary(telecom$tenure)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    9.00   29.00   32.37   55.00   72.00
```

We check this for any outliers or NA entries. Since there are no outliers nor NA points, we can move on to the next numerical variable

Summary of Telecom Tenure

```
summary(telecom$MonthlyCharges)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.25   35.50   70.35   64.76   89.85  118.75
```

The monthly charges range from $18.25 to $118.75. There are no blanks (NA's).

Summary of Telecom Tenure

```
summary(telecom$TotalCharges)
```

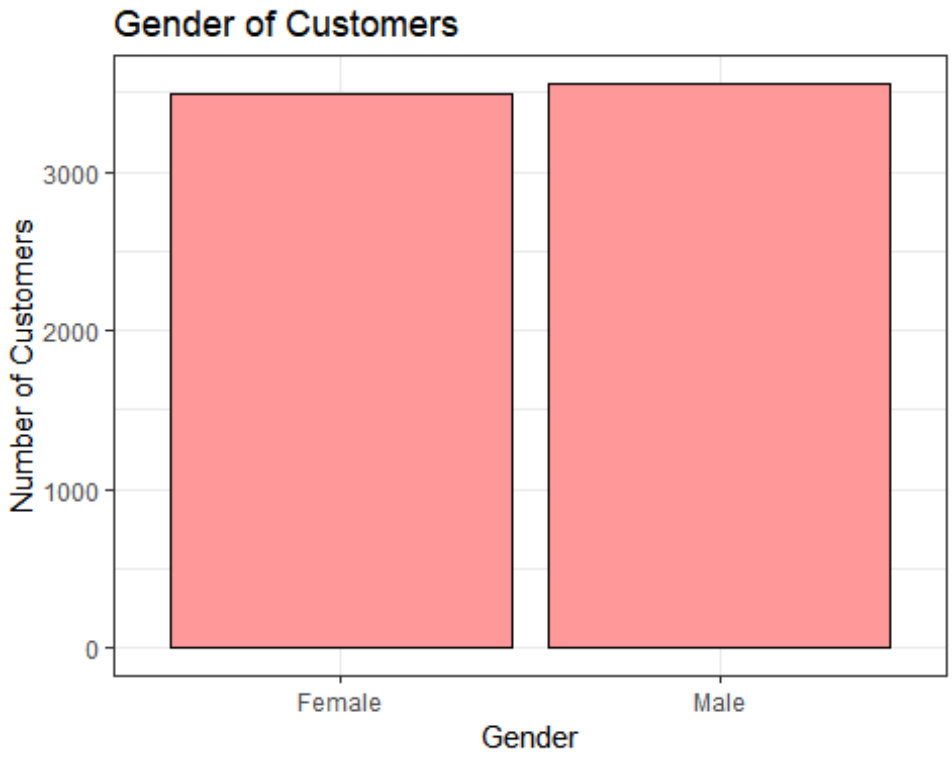
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    18.8   401.4  1397.5  2283.3  3794.7  8684.8      11
```

The total charges range from $18.8 to $8684.8, the max seems vary high it could possibly be an outlier. This variable also has 11 blank points. therfore we will have to determine if removing the rows with blanks would be better than keeping them. I decided to keep them in my data for now. but would use the omit function if needed to remove them.

gender

Is this data bias toward men? Are men more likely to have have coverage?

```
ggplot(telecom,aes(gender))+geom_bar( fill="#FF9999", colour="black")+
  theme_bw()+
  labs(x="Gender",y="Number of Customers",title="Gender of Customers")
```




```
prop.table(table(telecom$gender))
```

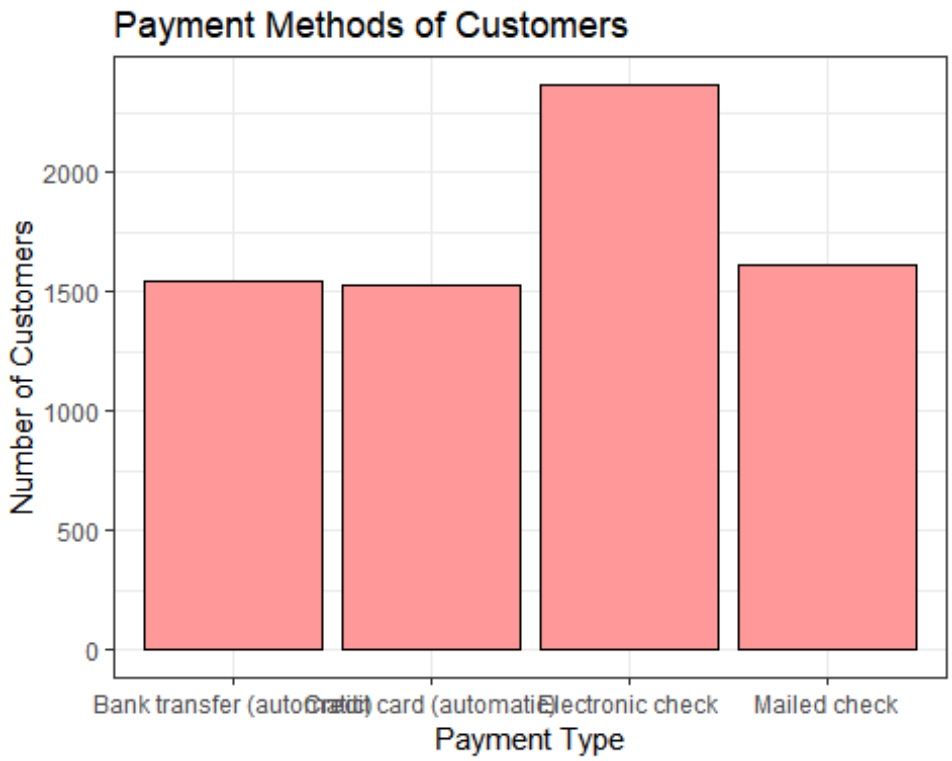
```
##
##    Female      Male
## 0.4952435 0.5047565
```

We have 48.5% Female and 50.5% Male in our data.Therefore the data seems normally distributed and large enough to be unbias.

Payment

What is the most common form of payment?

```
ggplot(telecom,aes(PaymentMethod))+geom_bar( fill="#FF9999", colour="black")+
  theme_bw()+
  labs(x="Payment Type",y="Number of Customers",title="Payment Methods of
Customers")
```




```
prop.table(table(telecom$PaymentMethod))

##
## Bank transfer (automatic)   Credit card (automatic)
##                 0.2192248                 0.2161011
##          Electronic check              Mailed check
##                 0.3357944                 0.2288797
```

The Most common form of payment was Electronic check accounting for 33.5%. Second most common was Mailed check at 22.9%. Third was Bank transfer (automatic) at 22%. Least common was Credit card (automatic) at 21.6%.

Monthly charges

What is the range of monthly charges?

```
summary(telecom$MonthlyCharges)
```

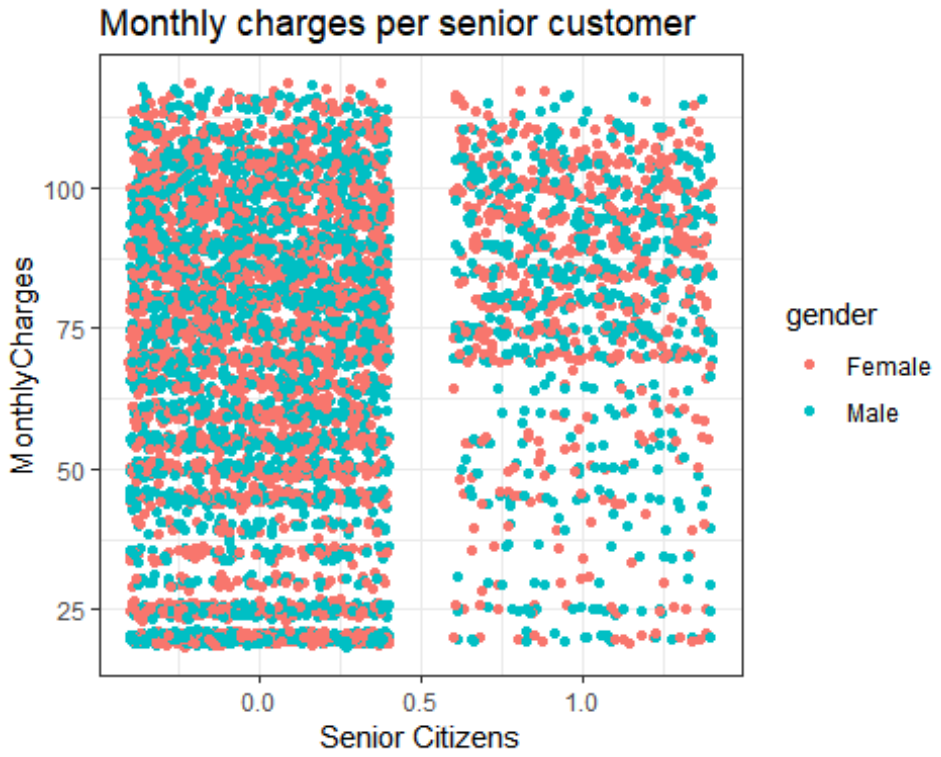
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.25   35.50   70.35   64.76   89.85  118.75
```

The monthly charges range from $18.25 to $118.75 per month. The mean/average monthly bill is $64.76. While the median bill is $70.35.

Do Senior Citizen's pay less in comparison to everyone else?

Do senior citizens get a discount? Are there more senior customers?

```
ggplot(telecom,aes(x=SeniorCitizen,y=MonthlyCharges, color=gender))+
  theme_bw()+
  geom_jitter()+
  labs(x="Senior Citizens",y="MonthlyCharges",title="Monthly charges per
senior customer")
```



The telecom company seems to have about the same minimum and mazimum as for the Senior citizens as the regular non senior customers; However, the majority of Senior Citizens seem to be paying above ~$70. Therfore it seems that being a senior does not give

a bonus to all customers. We would have to do a bar plot to see the difference between number of senior and non senior customers.

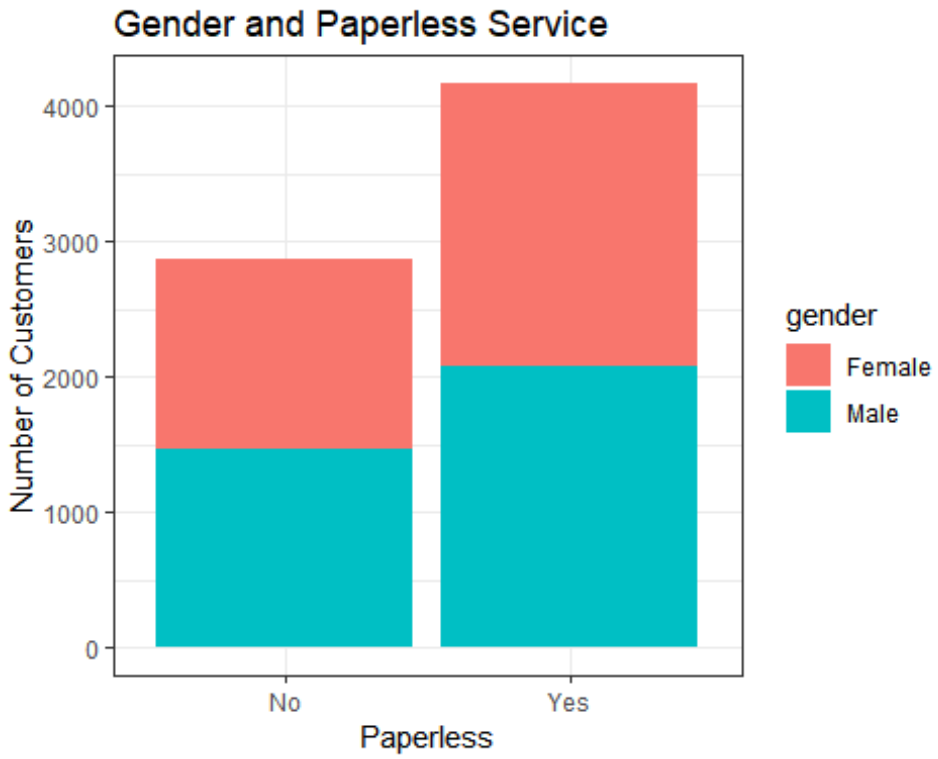
```
ggplot(telecom,aes(x=SeniorCitizen, fill=gender))+ geom_bar()
```

There are very few Senior Customers in comparison to non senior customers.

Ecofriendly initative by going paperless

How eco friendly is the brand? Does it have higher percentage of paperless billing?

```
ggplot(telecom,aes(x=PaperlessBilling,fill=gender))+
  theme_bw()+
  geom_bar () +
  labs(x="Paperless",y="Number of Customers",title="Gender and Paperless
Service")
```



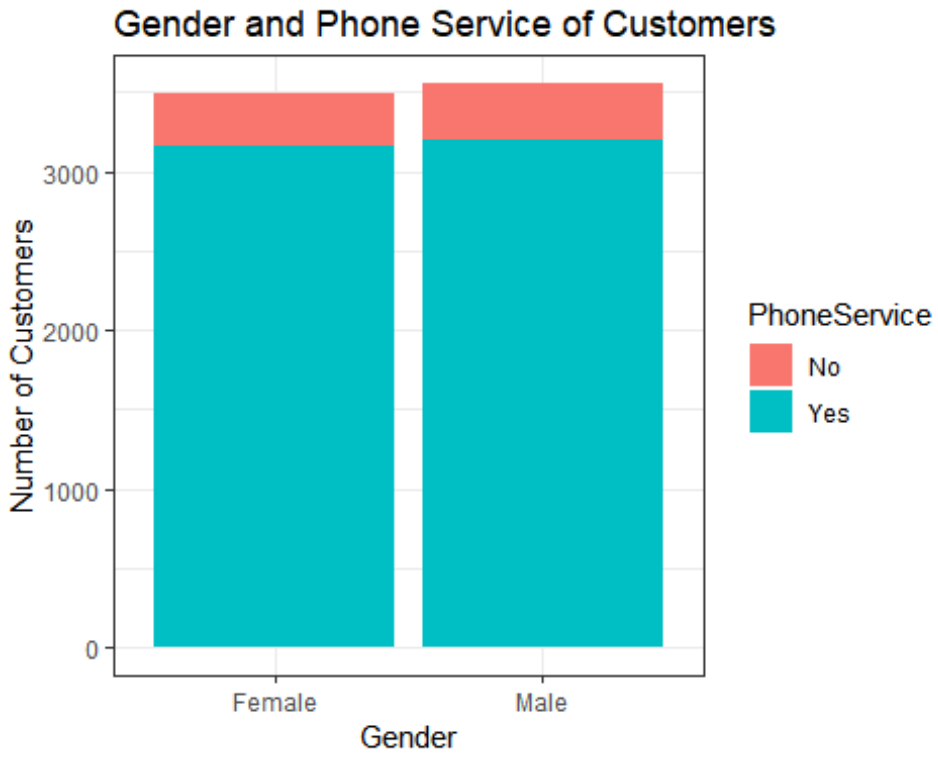

```
prop.table(table(telecom$PaperlessBilling))

##
##        No       Yes
## 0.4077808 0.5922192
```

About 59% of customers choose the paperless route, while 41% still want a paper copy of the bill. Though this can be inproved by giving an incentive to go paperless, which would save the company on stationary supplies.

gender and phoneservice

```
ggplot(telecom,aes(x=gender,fill=PhoneService))+
  theme_bw()+
  geom_bar () +
  labs(x="Gender",y="Number of Customers",title="Gender and Phone Service of
Customers")
```

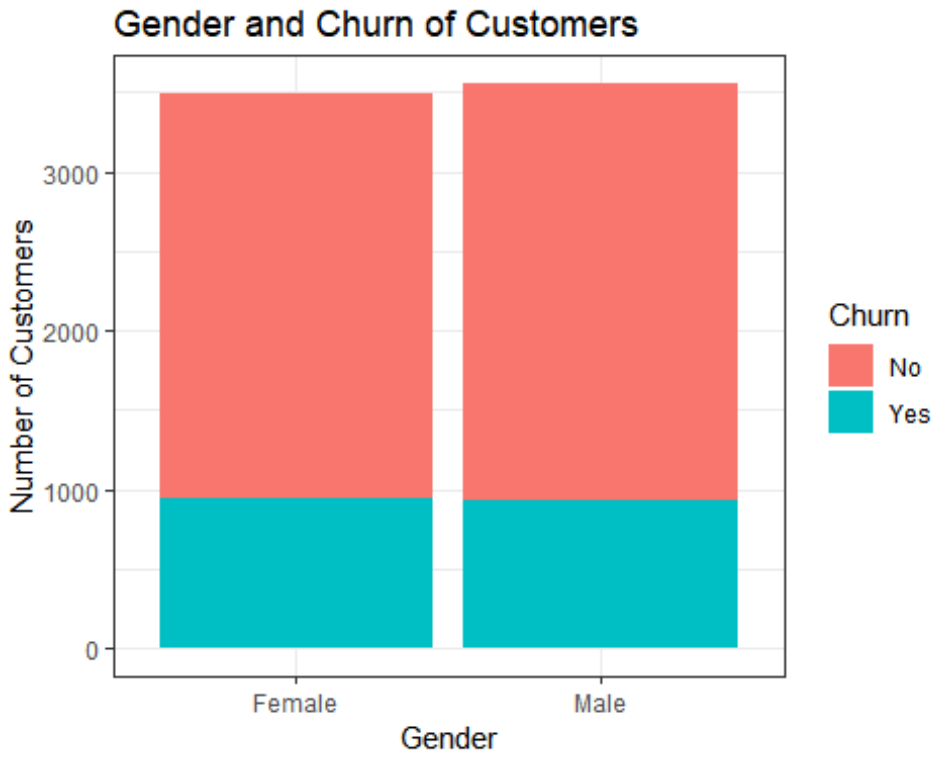



```
prop.table(table(telecom$gender, telecom$PhoneService))
```

```
##
##                   No        Yes
##   Female 0.04699702 0.44824649
##   Male   0.04983672 0.45491978
```

44% women have a phone service with our company while 4% women do not. 45% male customers have the phone service, while 5% male do not have the phone service.

gender and Churn




```
prop.table(table(telecom$gender, telecom$Churn))
```

```
##
##                  No       Yes
##   Female 0.3619196 0.1333239
##   Male   0.3727105 0.1320460
```

About 26.5% of customers churned of these 13.3% where Female while 13.2% were male. While 73.5% stayed with out telecom company. Of these 36.2% where female while 37.3% where male.