

Adding Virtual Tours to Real Estate listings- Marketing

```
library(tidyverse)
library(data.table)
library(dplyr)
library(lubridate)
library(zoo)
library(gridExtra)
```

Clean & process data

```
Data = read.csv("Data.csv")

# Clean the dataset
Data$List.Date = as.Date(mdy(Data$List.Date))
Data$Sold.Date = as.Date(mdy(Data$Sold.Date))
Data$List.Price <- as.numeric(gsub('$', '', Data$List.Price))
Data$Sold.Price <- as.numeric(gsub('$', '', Data$Sold.Price))
Data$Sold.Price.per.SqFt <- as.numeric(gsub('$', '', Data$Sold.Price.per.SqFt))
Data$TotFlArea = as.numeric(gsub('[,]', '', Data$TotFlArea))
Data$Status = as.factor(Data$Status)
Data$TypeDwel <- as.factor(Data$TypeDwel)
Data$City <- as.factor(Data$City)
Data$Area <- as.factor(Data$Area)
Data$StratMtFee <- as.numeric(gsub('$', '', Data$StratMtFee))

# Creating the virtual column
Data$virtual <- as.numeric(Data$Virtual.Tour.URL!="")
Data$virtual <- factor(Data$virtual)

Data$sold <- as.numeric(Data$Status=="S")
Data$sold <- factor(Data$sold)
```

EDA

```
summary(Data)
```

```

##      PicCount          ML..      Status
## Min.    : 0.00    R2272222:    2    C: 412
## 1st Qu.:10.00    R2279232:    2    S:43245
## Median :16.00    R2282249:    2    T:11516
## Mean    :14.19    R2294222:    2    W:    3
## 3rd Qu.:20.00    F1304705:    1    X: 4824
## Max.    :20.00    F1306404:    1
##
##              (Other) :59990
##
##              Address          City
## 3504 838 W HASTINGS STREET:    7    Vancouver:14215
## 6486 BRANTFORD AVENUE      :    6    Surrey   : 9549
## 677 E 59TH AVENUE          :    6    Richmond  : 5944
## 101 6939 GILLEY AVENUE     :    5    Burnaby   : 5192
## 1206 151 W 2ND STREET      :    5    Langley   : 3068
## 17 7311 MINORU BOULEVARD   :    5    Coquitlam: 3049
## (Other)                    :59966    (Other)   :18983
##
##              Area          S.A          List.Price
## Vancouver West:10277    VVWDT : 1422    Min.    : 52900
## Richmond      : 5947    VRI60 : 1361    1st Qu.: 344000
## Vancouver East: 3926    F26    : 1350    Median : 495000
## Langley       : 3070    VVWYA : 986     Mean   : 610233
## Surrey       : 3049    F63    : 881     3rd Qu.: 699000
## Coquitlam     : 3047    F78    : 782     Max.    :18000000
## (Other)       :30684    (Other):53218
##
##      Price      Sold.Price.per.SqFt    List.Date
## $450,000 : 257    Min.    : 38.54    Min.    :2012-05-11
## $550,000 : 242    1st Qu.: 318.76    1st Qu.:2015-09-28
## $425,000 : 237    Median : 466.02    Median :2016-09-28
## $420,000 : 234    Mean   : 531.10    Mean   :2016-10-10
## $599,000 : 233    3rd Qu.: 686.34    3rd Qu.:2017-10-03
## $475,000 : 232    Max.    :2937.01    Max.    :2018-11-30
## (Other)   :58565    NA's    :16766
##
##      Sold.Date          DOM          Tot.BR          Tot.Baths
## Min.    :2015-01-01    Min.    : 0.00    Min.    :0.000    Min.    :0.000
## 1st Qu.:2015-12-14    1st Qu.: 7.00    1st Qu.:2.000    1st Qu.:1.000
## Median :2016-10-27    Median : 21.00    Median :2.000    Median :2.000
## Mean    :2016-11-08    Mean   : 44.95    Mean   :2.097    Mean   :1.992
## 3rd Qu.:2017-09-20    3rd Qu.: 61.00    3rd Qu.:3.000    3rd Qu.:2.000
## Max.    :2018-12-02    Max.    :1464.00    Max.    :9.000    Max.    :7.000
## NA's    :16755
##
##      TotFlArea          Age          TypeDwel
## Min.    : 0    Min.    : 0.0    1/2 Duplex : 1961
## 1st Qu.: 785    1st Qu.: 4.0    Apartment/Condo:39698
## Median :1008    Median : 11.0    Townhouse   :18341
## Mean    :1130    Mean   : 16.5
## 3rd Qu.:1368    3rd Qu.: 24.0
## Max.    :6833    Max.    :999.0
##
##              NA's :1

```

```

##                                     Commission
## S-3.22% 1ST 100K/1.15% BAL                : 6133
## 3.255% ON 1ST $100,000/1.1625% ON BAL      : 2281
## 3.255% ON 1ST $100,000 & 1.1625% ON BALANCE : 901
## 3.255% ON 1ST 100K/1.1625% ON BAL          : 765
## 3.255% ON THE FIRST $100,000 AND 1.1625% ON THE BALANCE: 687
## S-3.255% 1ST 100K/1.1625% BAL              : 647
## (Other)                                     :48586
##      TotalPrkng      Sold.Price      StratMtFee
## Min.   : 0.000   Min.   : 50000   Min.   : 0.0
## 1st Qu.: 1.000   1st Qu.: 338900   1st Qu.: 220.6
## Median : 1.000   Median : 475000   Median : 278.6
## Mean   : 1.539   Mean   : 558490   Mean   : 302.8
## 3rd Qu.: 2.000   3rd Qu.: 655000   3rd Qu.: 350.0
## Max.   :90.000   Max.   :13250000   Max.   :25123.0
## NA's   :6993     NA's   :16755     NA's   :793
##                                     Virtual.Tour.URL virtual
##                                     :42510 0:42510
## http://www.cotala.com/10854          : 34 1:17490
## http://www.seevirtual360.com/25978   : 18
## http://www.opopmedia.com/virtualtour/ampri-museo/: 16
## http://www.seevirtual360.com/33729   : 14
## http://www.StuBell.com               : 12
## (Other)                             :17396
## sold
## 0:16755
## 1:43245
##
##
##
##
##

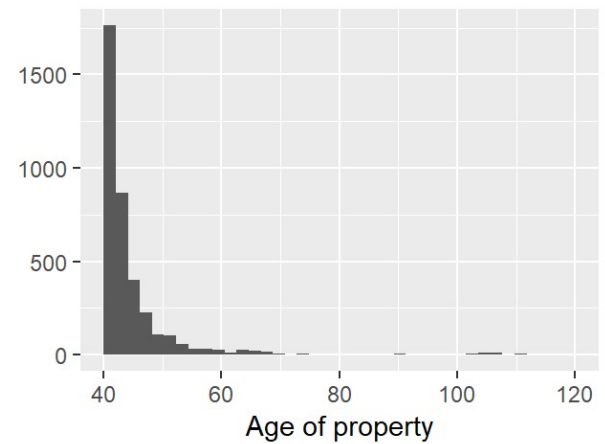
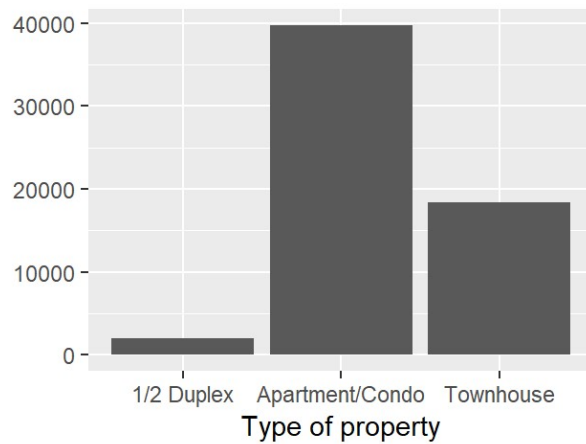
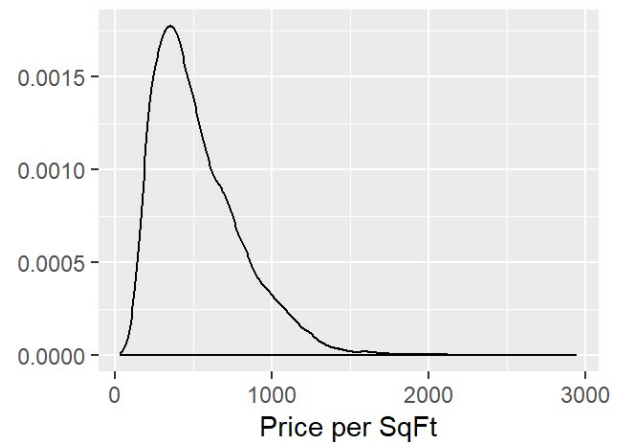
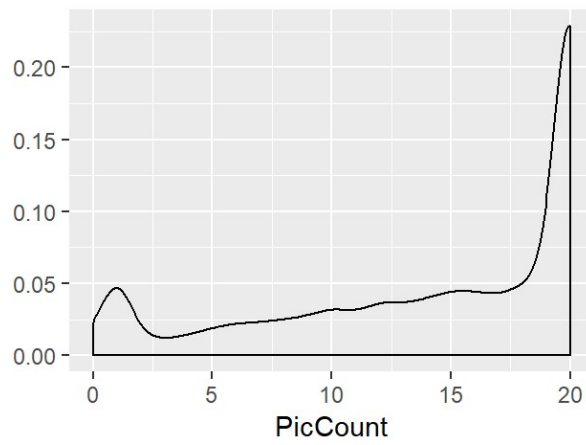
```

```

# plot
plot1 <- qplot(Data$PicCount,geom='density',xlab='PicCount')
plot2 <- qplot(Data$Sold.Price.per.SqFt,geom='density',xlab='Price per SqFt')
plot3 <- qplot(Data$TypeDwel,geom='bar',xlab='Type of property')
plot4 <- qplot(Data$Age,geom='histogram',xlab='Age of property',xlim = c(40,120),bins=
40)

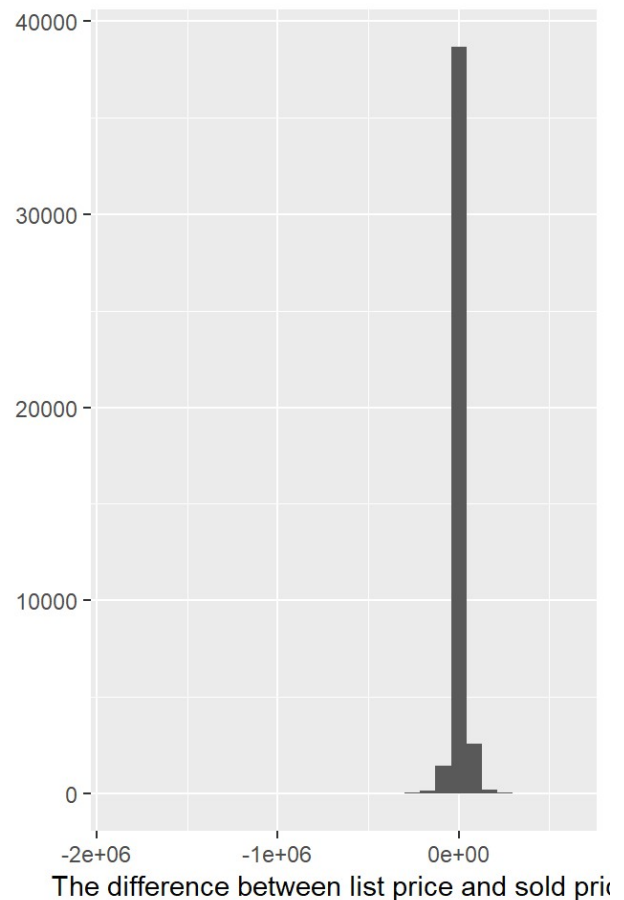
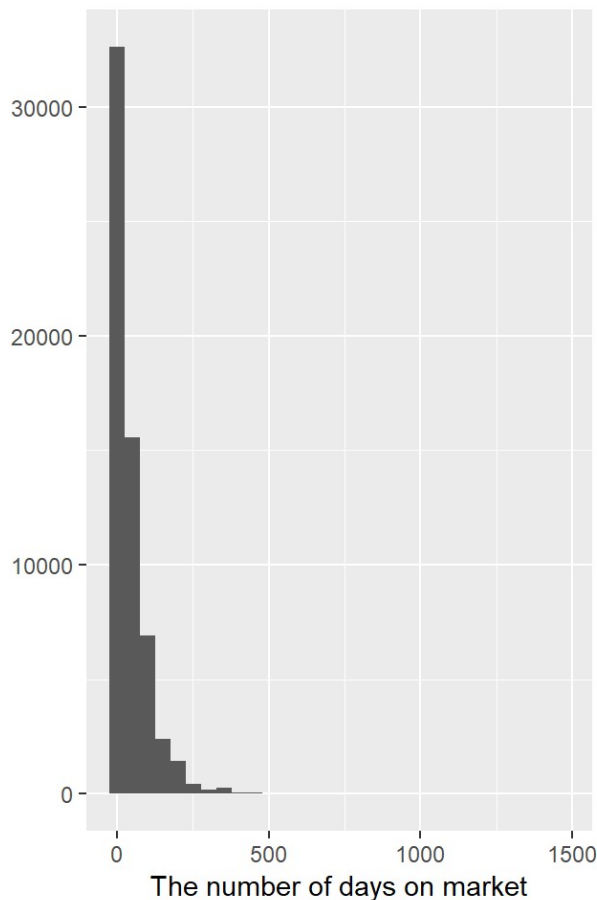
grid.arrange(plot1,plot2,plot3,plot4,ncol=2)

```



```
plot5 <- qplot(Data$DOM,geom='histogram',xlab='The number of days on market')

Data$price.diff <- Data$Sold.Price-Data$List.Price
plot6 <- qplot(Data$price.diff,geom='histogram',xlab='The difference between list price and sold price')
grid.arrange(plot5,plot6,ncol=2)
```



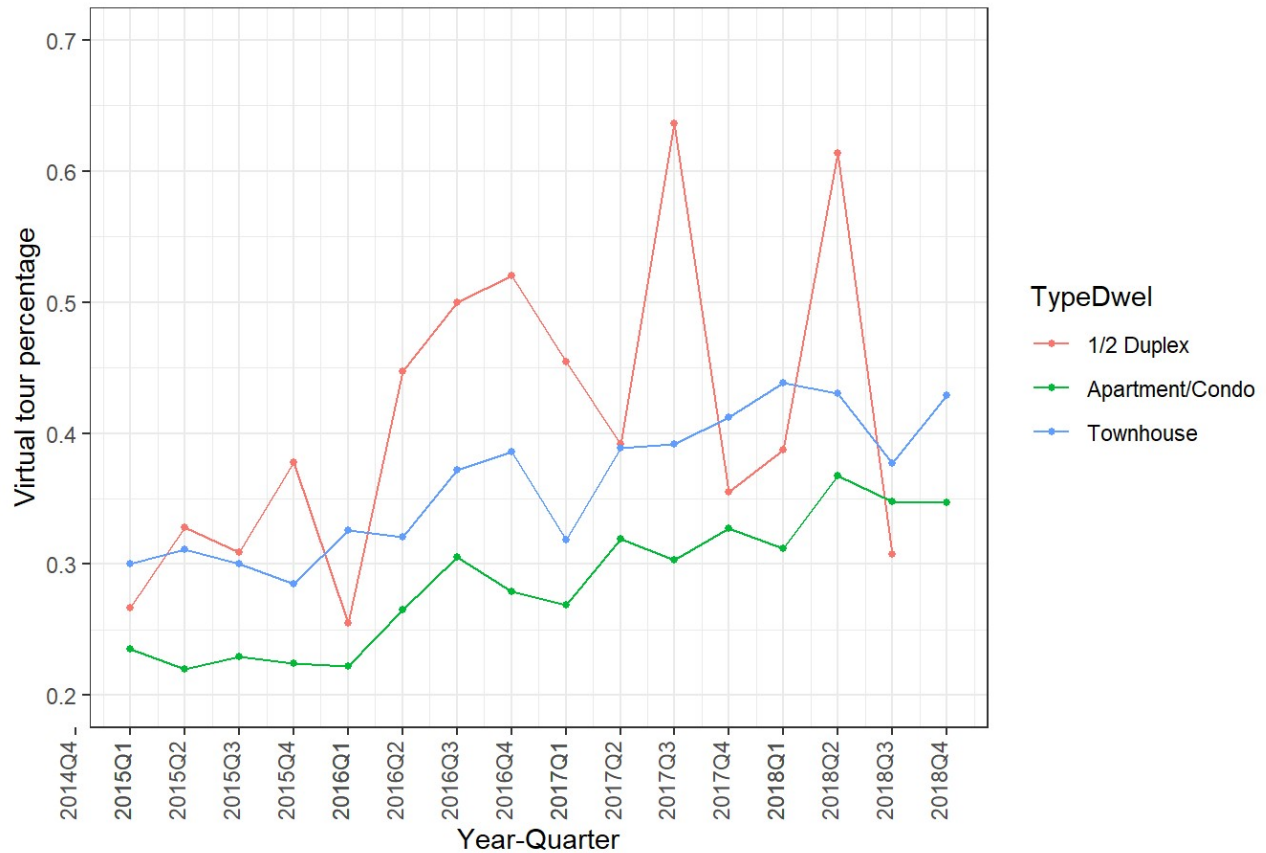
Univariate Analysis

```
# subset the dataset
mask <- Data$City %in% c("Vancouver", "Burnaby", "North Vancouver")
Data.VBV <- subset(Data,mask)

# Plot time trend in term of sold date
Data.VBV$Sold.Date.Year.quarter <- as.yearqtr(Data.VBV$Sold.Date, format = "%Y-%m-%d")
average <- Data.VBV %>% group_by(TypeDwel,Sold.Date.Year.quarter) %>% summarise(average_virtual = mean(as.numeric(I(virtual==1)),na.rm=TRUE))

ggplot(average, aes(Sold.Date.Year.quarter, average_virtual,group=TypeDwel)) +
  geom_point(aes(color=TypeDwel),size=1)+
  geom_line(aes(color=TypeDwel))+
  scale_x_yearqtr(format="%YQ%q",n=16)+
  scale_y_continuous(limits = c(0.2,0.7))+
  xlab("Year-Quarter") + ylab("Virtual tour percentage") +
  ggtitle('Vritual Tour')+
  theme_bw()+
  theme(axis.text.x = element_text(angle = 90, vjust=0, hjust=1))
```

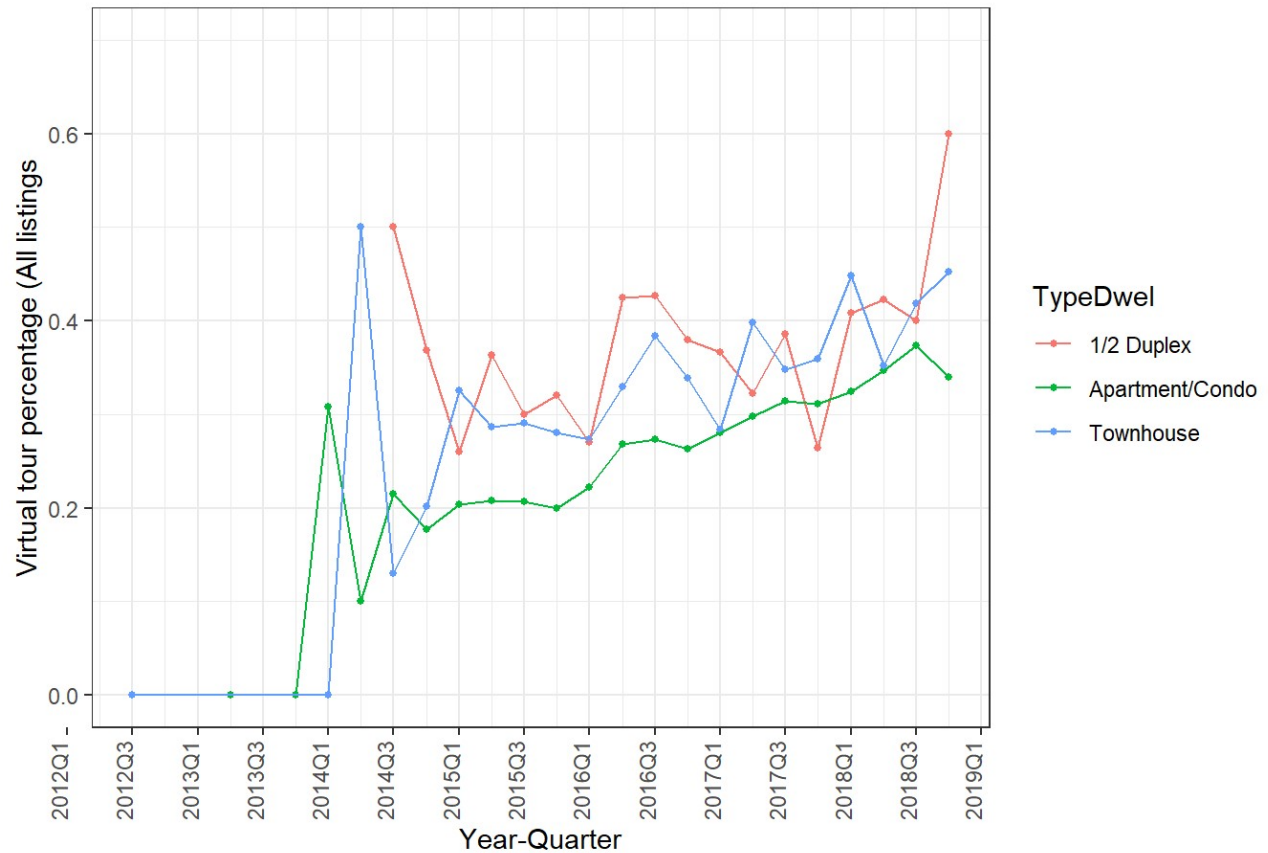
Vritual Tour



```
# Plot time trend in term of list date
Data.VBV$List.Date.Year.quarter <- as.yearqtr(Data.VBV$List.Date, format = "%Y-%m-%d")
average2 <- Data.VBV %>% group_by(TypeDwel,List.Date.Year.quarter) %>% summarise(average_virtual = mean(as.numeric(I(virtual==1)),na.rm=TRUE))

ggplot(average2, aes(List.Date.Year.quarter, average_virtual,group=TypeDwel)) +
  geom_point(aes(color=TypeDwel),size=1)+
  geom_line(aes(color=TypeDwel)) +scale_x_yearqtr(format="%YQ%q",n=16)+scale_y_continuous(limits = c(0,0.7))+
  xlab("Year-Quarter") + ylab("Virtual tour percentage (All listings)") +
  ggtitle('Vritual Tour')+
  theme_bw()+
  theme(axis.text.x = element_text(angle = 90, vjust=0, hjust=1))
```

Vritual Tour



Does type of dwelling have an impact?

```
z.TypeDwel <- glm(virtual~TypeDwel,data=Data.VBV, family = "binomial"(link='logit'))
z.null <- glm(virtual~1,data=Data.VBV, family = "binomial"(link='logit'))
summary(z.TypeDwel)
```

```
##
## Call:
## glm(formula = virtual ~ TypeDwel, family = binomial(link = "logit"),
##      data = Data.VBV)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -0.9437  -0.7909  -0.7909   1.4760   1.6216
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -0.57814    0.06392  -9.044  < 2e-16 ***
## TypeDwelApartment/Condo -0.42386    0.06611  -6.411 1.44e-10 ***
## TypeDwelTownhouse     -0.10092    0.07351  -1.373    0.17
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 26654  on 22362  degrees of freedom
## Residual deviance: 26560  on 22360  degrees of freedom
## AIC: 26566
##
## Number of Fisher Scoring iterations: 4
```

```
anova(z.null,z.TypeDwel,test='Chisq')
```

```
## Analysis of Deviance Table
##
## Model 1: virtual ~ 1
## Model 2: virtual ~ TypeDwel
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      22362      26654
## 2      22360      26560  2   94.027 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Probability of using virtual tour
```

```
z.prob <- sapply(levels(Data.VBV$TypeDwel),function(x){predict(z.TypeDwel,newdata=dat
a.frame(TypeDwel=x),type='response')})
```

```
print(z.prob)
```


##	1/2 Duplex.1 Apartment/Condo.1	Townhouse.1	
##	0.3593603	0.2685475	0.3364706

Effect of Multiple Variates

- DOM & Price Difference between sold price and listing price
- covariates

```
z.DOM<-lm(DOM~PicCount+Sold.Price.per.SqFt+Area+TypeDwel+Age + virtual, data=Data.VBV)
z.difference<-lm(price.diff~PicCount+Sold.Price.per.SqFt+Area+TypeDwel+Age +virtual, data=Data.VBV)

print(summary(z.DOM))
```

```
##
## Call:
## lm(formula = DOM ~ PicCount + Sold.Price.per.SqFt + Area + TypeDwel +
##     Age + virtual, data = Data.VBV)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -40.56 -19.01 -12.72   4.93 644.85
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    44.772697   2.739507  16.343 < 2e-16 ***
## PicCount       -0.511662   0.050927 -10.047 < 2e-16 ***
## Sold.Price.per.SqFt -0.010910  0.001244  -8.771 < 2e-16 ***
## AreaBurnaby North  -0.277980   2.353185  -0.118 0.905967
## AreaBurnaby South   1.601286   2.332772   0.686 0.492452
## AreaLadner        -5.894845  37.090773  -0.159 0.873726
## AreaNorth Vancouver -2.028332   2.313296  -0.877 0.380601
## AreaRichmond     -19.810359  26.276284  -0.754 0.450905
## AreaVancouver East  -2.237173   2.286157  -0.979 0.327806
## AreaVancouver West   3.208156   2.267418   1.415 0.157119
## AreaWest Vancouver  20.743678  37.095262   0.559 0.576033
## TypeDwelApartment/Condo -2.847540   1.569241  -1.815 0.069605 .
## TypeDwelTownhouse    -2.214724   1.707110  -1.297 0.194529
## Age              -0.044245   0.008518  -5.194 2.08e-07 ***
## virtual1          -2.651100   0.683553  -3.878 0.000106 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 37.02 on 15693 degrees of freedom
## (6655 observations deleted due to missingness)
## Multiple R-squared:  0.01919,    Adjusted R-squared:  0.01832
## F-statistic: 21.93 on 14 and 15693 DF,  p-value: < 2.2e-16
```

```
print(summary(z.difference))
```

```
##
## Call:
## lm(formula = price.diff ~ PicCount + Sold.Price.per.SqFt + Area +
##     TypeDwel + Age + virtual, data = Data.VBV)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1793764   -16936    -4170    12876   653528
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6665.174    4048.908    1.646 0.099750 .
## PicCount           430.444      75.268    5.719 1.09e-08 ***
## Sold.Price.per.SqFt    -25.154       1.838   -13.683 < 2e-16 ***
## AreaBurnaby North    1594.988    3477.936    0.459 0.646526
## AreaBurnaby South     693.007    3447.766    0.201 0.840700
## AreaLadner          5638.426   54819.038    0.103 0.918079
## AreaNorth Vancouver  4541.504    3418.982    1.328 0.184092
## AreaRichmond        2118.433   38835.551    0.055 0.956499
## AreaVancouver East   12044.822    3378.870    3.565 0.000365 ***
## AreaVancouver West   4890.492    3351.175    1.459 0.144492
## AreaWest Vancouver  -44014.122   54825.673   -0.803 0.422102
## TypeDwelApartment/Condo -271.864    2319.290   -0.117 0.906688
## TypeDwelTownhouse    -1295.615    2523.057   -0.514 0.607602
## Age                25.590      12.590    2.033 0.042114 *
## virtual1           2610.877    1010.271    2.584 0.009766 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 54720 on 15693 degrees of freedom
## (6655 observations deleted due to missingness)
## Multiple R-squared:  0.0224, Adjusted R-squared:  0.02153
## F-statistic: 25.68 on 14 and 15693 DF,  p-value: < 2.2e-16
```