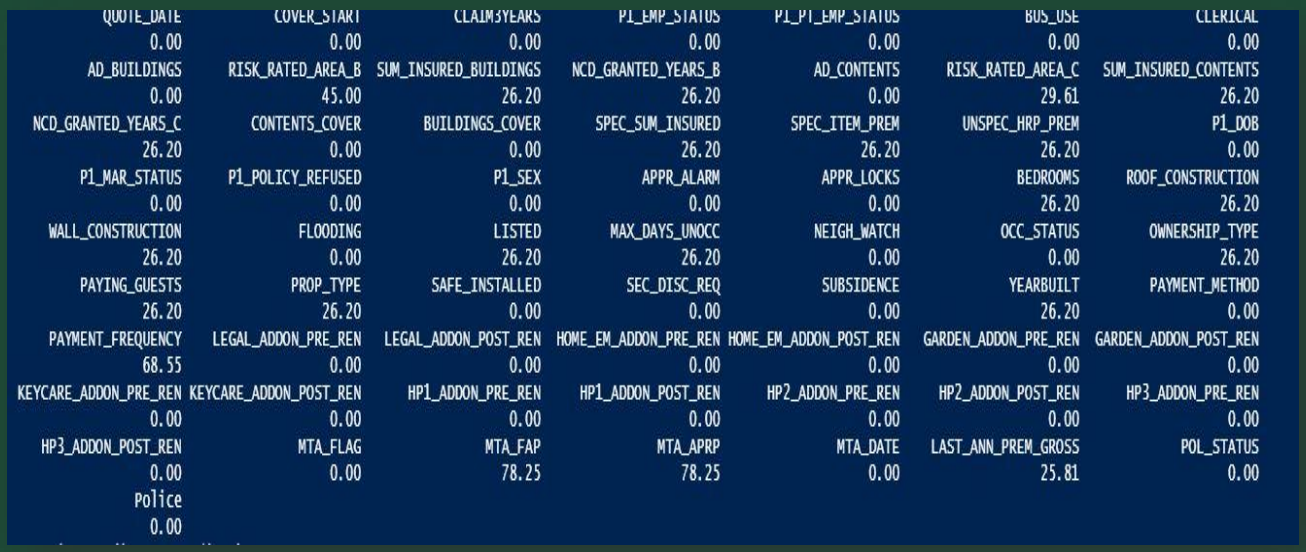
A REPORT ON PREMIUM PRICING ATTRIBUTES USING R

- Krishna Kulkarni

Sr No.	Table Of Contents	Page No.
1.	Initial Findings From The Dataset	2
2.	EDA On The Dataset a) Analysis on the factors affecting the Premium Prices b) Analysis Of Customers Likely To Default	3 9
3.	Conclusion	17

1. Initial Findings From The Dataset :

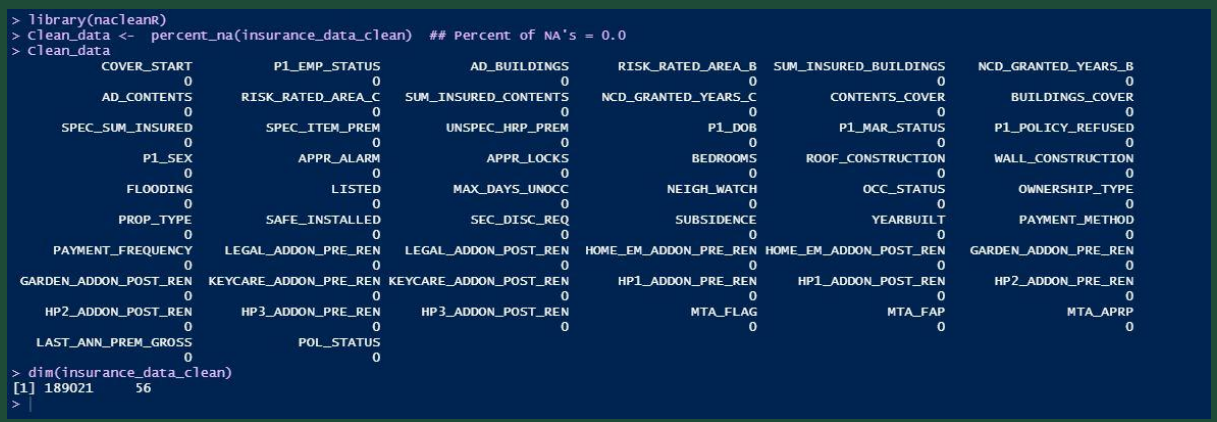
The dataset contained about 2,56,136 rows and 64 columns in total. The dataset contained variables of multiple datatypes. Also the dataset had missing values present in it.



QUOTE_DATE	COVER_START	CLAIM3YEARS	P1_EMP_STATUS	P1_PT_EMP_STATUS	BUS_USE	CLERICAL
0.00	0.00	0.00	0.00	0.00	0.00	0.00
AD_BUILDINGS	RISK_RATED_AREA_B	SUM_INSURED_BUILDINGS	NCD_GRANTED_YEARS_B	AD_CONTENTS	RISK_RATED_AREA_C	SUM_INSURED_CONTENTS
0.00	45.00	26.20	26.20	0.00	29.61	26.20
NCD_GRANTED_YEARS_C	CONTENTS_COVER	BUILDINGS_COVER	SPEC_SUM_INSURED	SPEC_ITEM_PREM	UNSPEC_HRP_PREM	P1_DOB
26.20	0.00	0.00	26.20	26.20	26.20	0.00
P1_MAR_STATUS	P1_POLICY_REFUSED	P1_SEX	APPR_ALARM	APPR_LOCKS	BEDROOMS	ROOF_CONSTRUCTION
0.00	0.00	0.00	0.00	0.00	26.20	26.20
WALL_CONSTRUCTION	FLOODING	LISTED	MAX_DAYS_UNOCC	NEIGH_WATCH	OCC_STATUS	OWNERSHIP_TYPE
26.20	0.00	26.20	26.20	0.00	0.00	26.20
PAYING_GUESTS	PROP_TYPE	SAFE_INSTALLED	SEC_DISC_REQ	SUBSIDENCE	YEARBUILT	PAYMENT_METHOD
26.20	26.20	0.00	0.00	0.00	26.20	0.00
PAYMENT_FREQUENCY	LEGAL_ADDON_PRE_REN	LEGAL_ADDON_POST_REN	HOME_EM_ADDON_PRE_REN	HOME_EM_ADDON_POST_REN	GARDEN_ADDON_PRE_REN	GARDEN_ADDON_POST_REN
68.55	0.00	0.00	0.00	0.00	0.00	0.00
KEYCARE_ADDON_PRE_REN	KEYCARE_ADDON_POST_REN	HP1_ADDON_PRE_REN	HP1_ADDON_POST_REN	HP2_ADDON_PRE_REN	HP2_ADDON_POST_REN	HP3_ADDON_PRE_REN
0.00	0.00	0.00	0.00	0.00	0.00	0.00
HP3_ADDON_POST_REN	MTA_FLAG	MTA_FAP	MTA_APRP	MTA_DATE	LAST_ANN_PREM_GROSS	POL_STATUS
0.00	0.00	78.25	78.25	0.00	25.81	0.00
Police						
0.00						

Missing Values in Percentage

Columns containing NA's were either completely dropped or were replaced with mean/median depending on the percentage of them present in those columns. As shown in fig below missing values have been either dropped or replaced with mean/median. The dataset after removal of missing values has 1,89,021 rows and 54 columns.



```
> library(naclean R)
> clean_data <-  percent_na(insurance_data_clean)  ## Percent of NA's = 0.0
> clean_data
```

COVER_START	P1_EMP_STATUS	AD_BUILDINGS	RISK_RATED_AREA_B	SUM_INSURED_BUILDINGS	NCD_GRANTED_YEARS_B
0	0	0	0	0	0
AD_CONTENTS	RISK_RATED_AREA_C	SUM_INSURED_CONTENTS	NCD_GRANTED_YEARS_C	CONTENTS_COVER	BUILDINGS_COVER
0	0	0	0	0	0
SPEC_SUM_INSURED	SPEC_ITEM_PREM	UNSPEC_HRP_PREM	P1_DOB	P1_MAR_STATUS	P1_POLICY_REFUSED
0	0	0	0	0	0
P1_SEX	APPR_ALARM	APPR_LOCKS	BEDROOMS	ROOF_CONSTRUCTION	WALL_CONSTRUCTION
0	0	0	0	0	0
FLOODING	LISTED	MAX_DAYS_UNOCC	NEIGH_WATCH	OCC_STATUS	OWNERSHIP_TYPE
0	0	0	0	0	0
PROP_TYPE	SAFE_INSTALLED	SEC_DISC_REQ	SUBSIDENCE	YEARBUILT	PAYMENT_METHOD
0	0	0	0	0	0
PAYMENT_FREQUENCY	LEGAL_ADDON_PRE_REN	LEGAL_ADDON_POST_REN	HOME_EM_ADDON_PRE_REN	HOME_EM_ADDON_POST_REN	GARDEN_ADDON_PRE_REN
0	0	0	0	0	0
GARDEN_ADDON_POST_REN	KEYCARE_ADDON_PRE_REN	KEYCARE_ADDON_POST_REN	HP1_ADDON_PRE_REN	HP1_ADDON_POST_REN	HP2_ADDON_PRE_REN
0	0	0	0	0	0
HP2_ADDON_POST_REN	HP3_ADDON_PRE_REN	HP3_ADDON_POST_REN	MTA_FLAG	MTA_FAP	MTA_APRP
0	0	0	0	0	0
LAST_ANN_PREM_GROSS	POL_STATUS				
0	0				

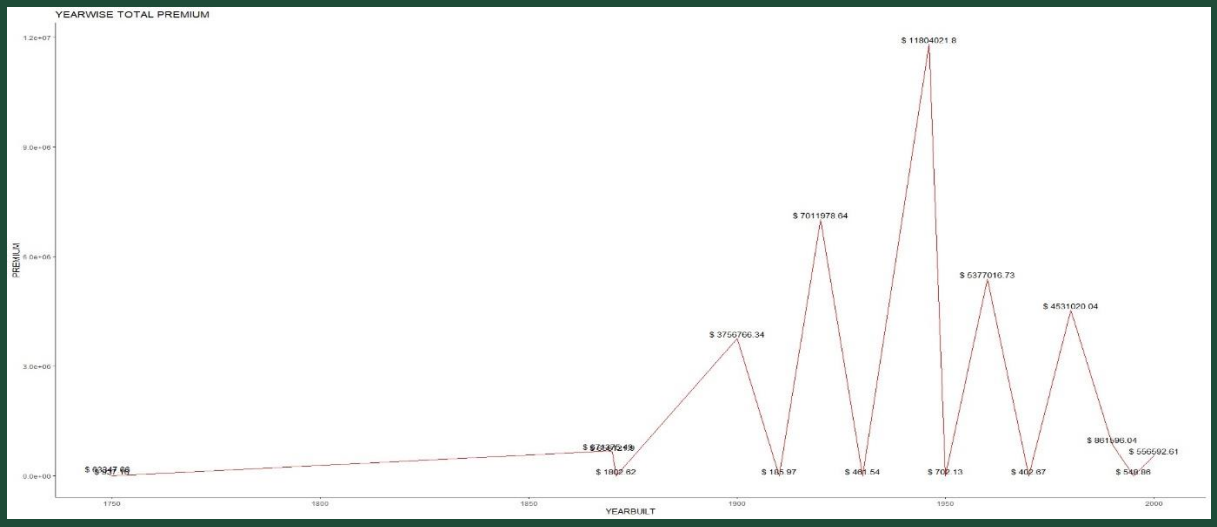
```
> dim(insurance_data_clean)
[1] 189021     56
> |
```

Missing values removed

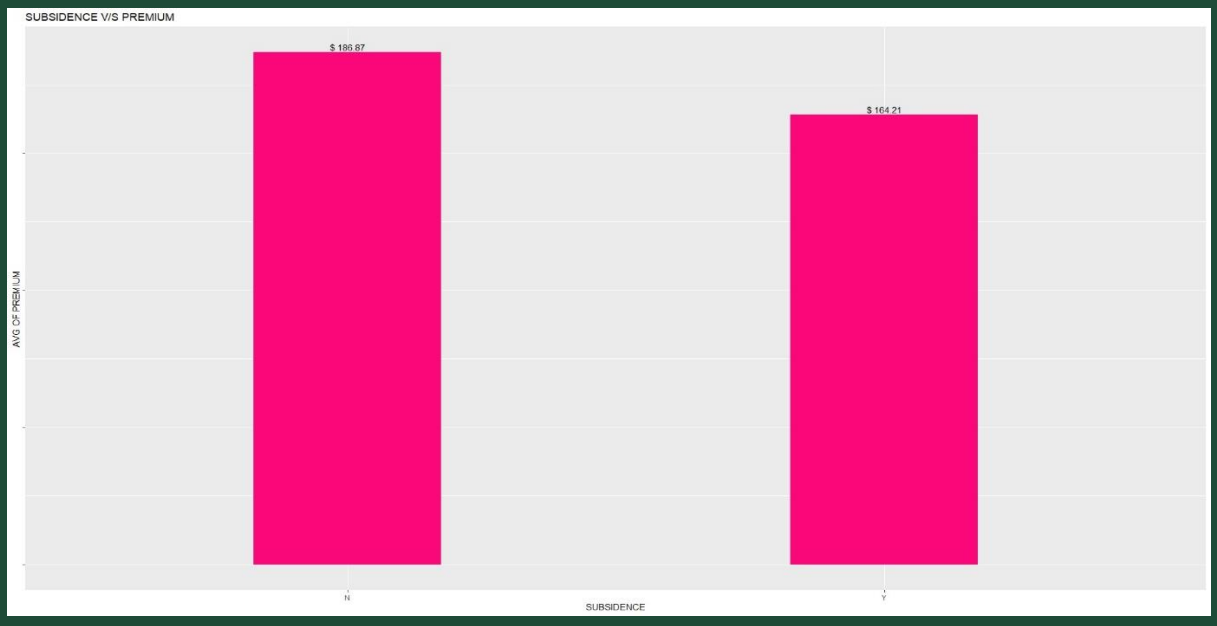
2. EDA on the Dataset :

a) Analysis on the factors affecting the Premium Prices:

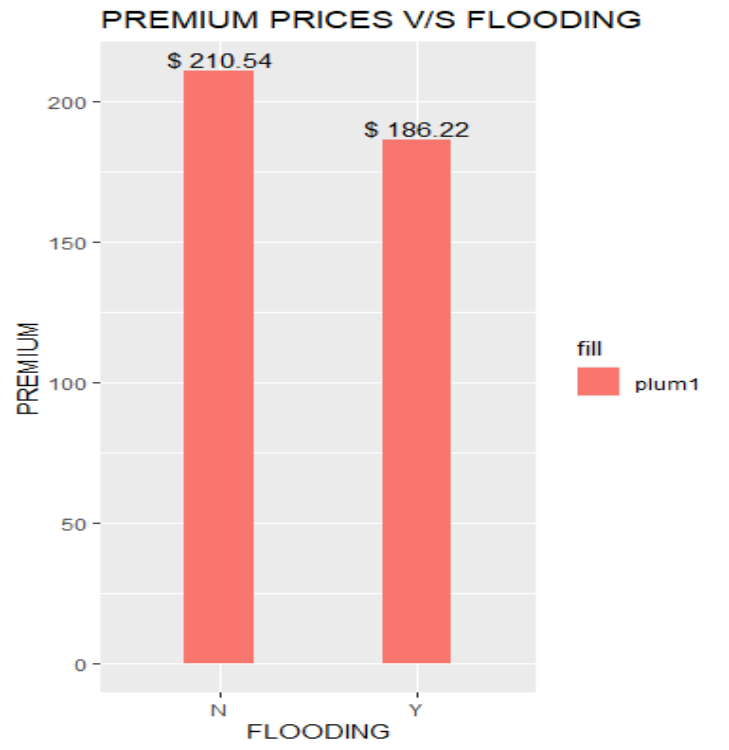
I. YEARBUILT: As shown in the fig. below the premium prices for houses built in between years 1900 – 1950 kept rising till year 1950. In 1950 the premium prices were at an all time high. This may have been influenced by other factors as floods, construction type etc. The buildings constructed in late '90s have significantly lower premium prices than in '50s.



II. SUBSIDENCE : As seen from the fig. presence of subsidence is not proportional to Premium Prices. Premium Prices of houses with subsidence is lower than those which don't have them.

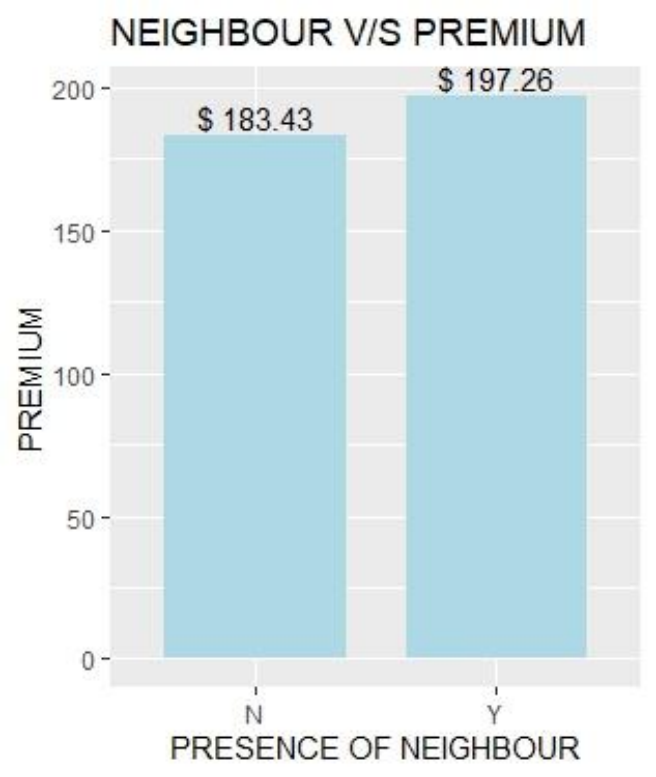


III. FLOODING:



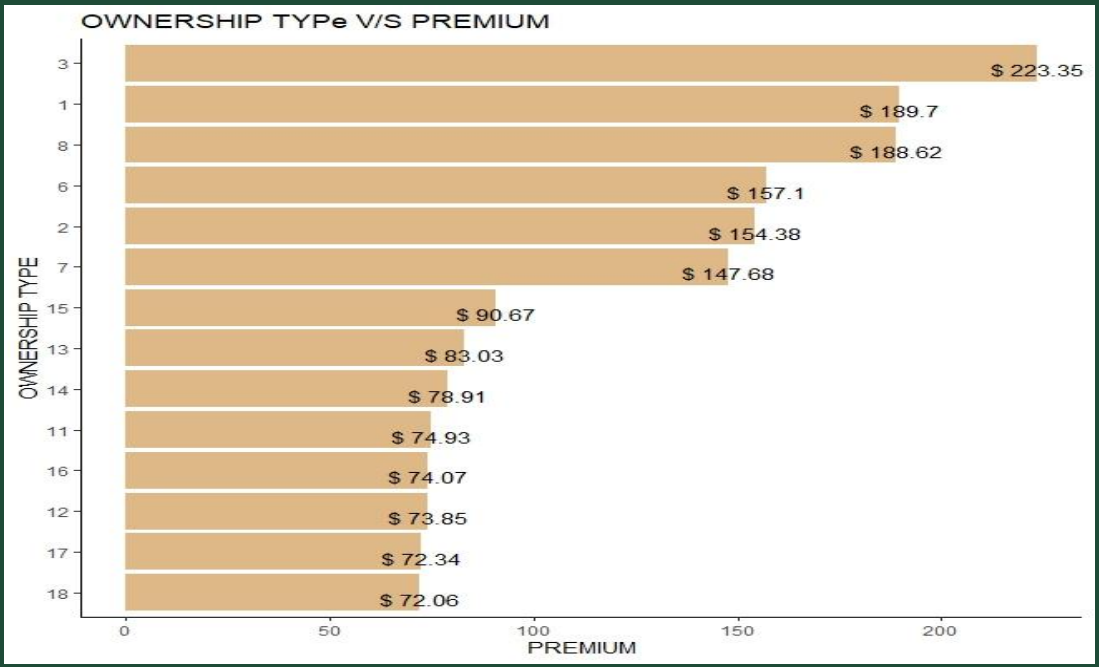
The customers who live in a flood prone area are likely to pay less amount of average premium than those who are not in the flood prone zones.

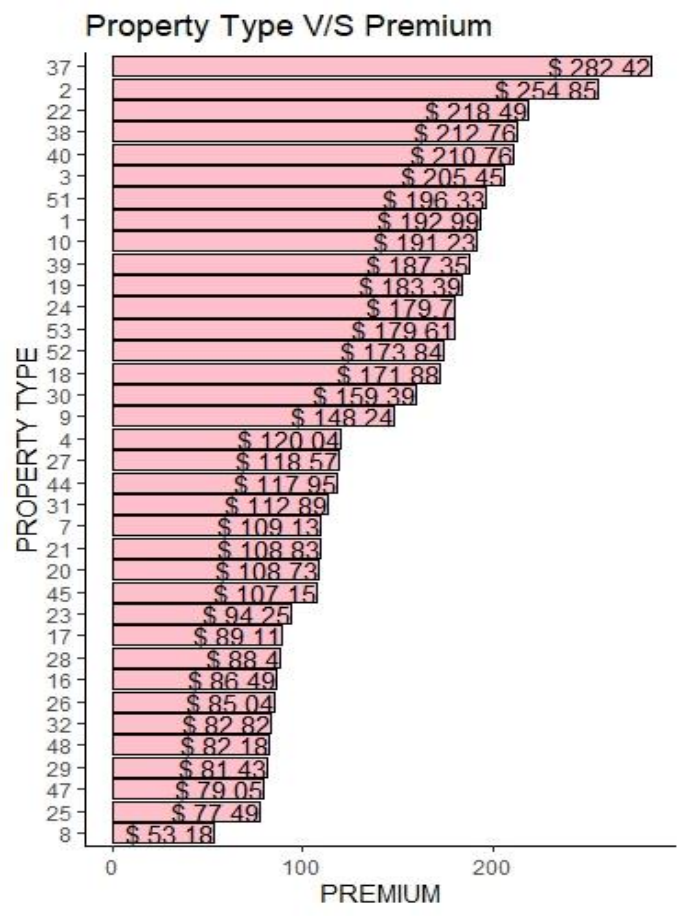
IV. NEIGHBOURHOOD :



Type of neighbourhood has a direct impact on premium prices. Areas with lower standard of living, higher chances of burglary tend to have a significantly lower average premium prices. On the opposite the areas with moderate to high standard of living drive up the premium prices.

V. OWNERSHIP TYPE: As seen on the fig. below the highest average premium price is observed for 'Ownership Type – 3' and lowest for 'Ownership Type – 18'.





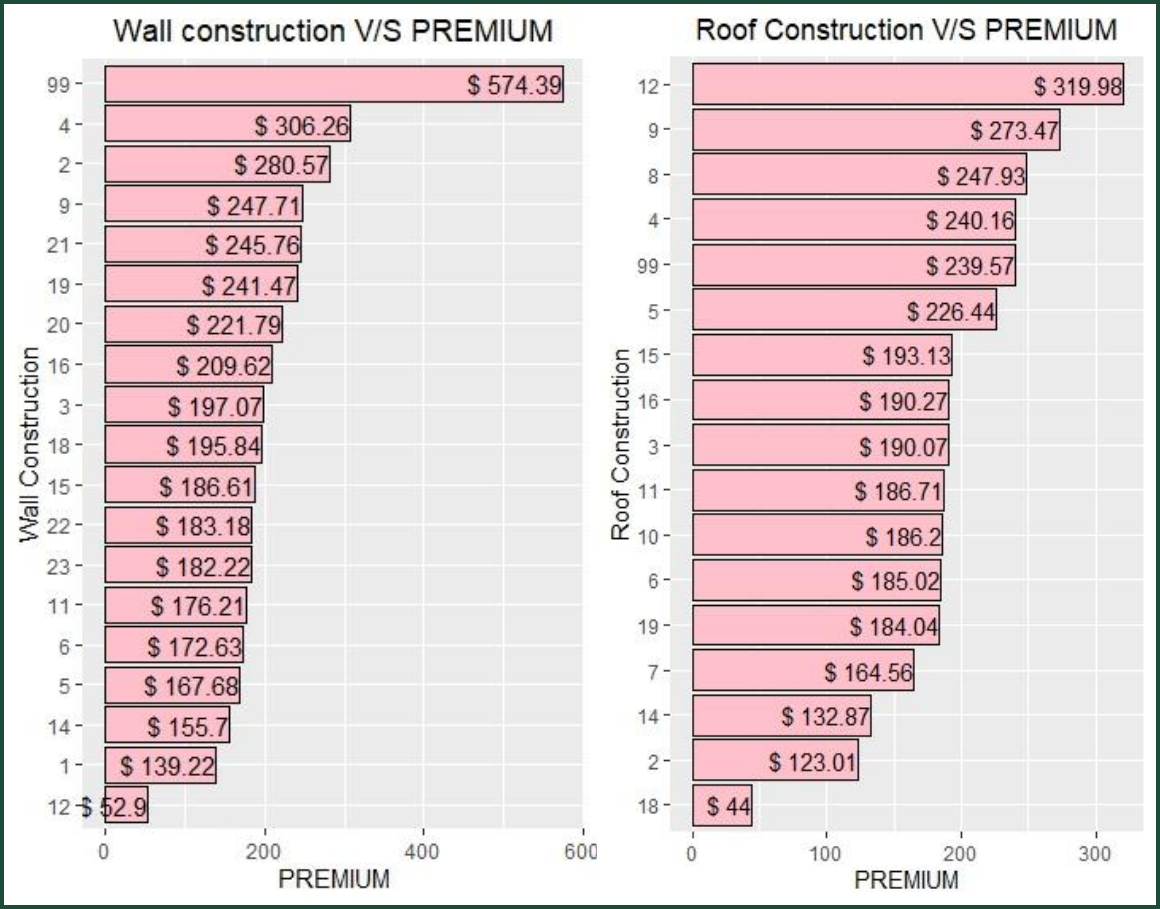
VI . PROPERTY TYPE:

As seen from the fig. 'Property Type – 37' has the highest average premium price and 'Property Type – 8' being the lowest. Meaning Customers will pay the premium prices according to the property types.

VI. BEDROOMS :

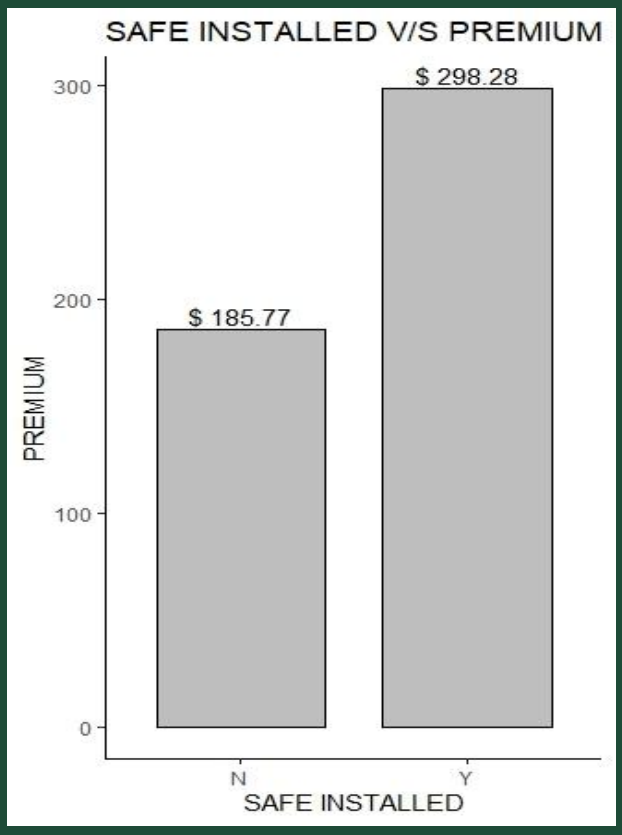
No. of bedrooms in a house has a direct impact on premium prices. Houses with 7 bedrooms have the highest average premium price of $474.69 while the houses with 1 bedrooms have average premium price of $80.79.

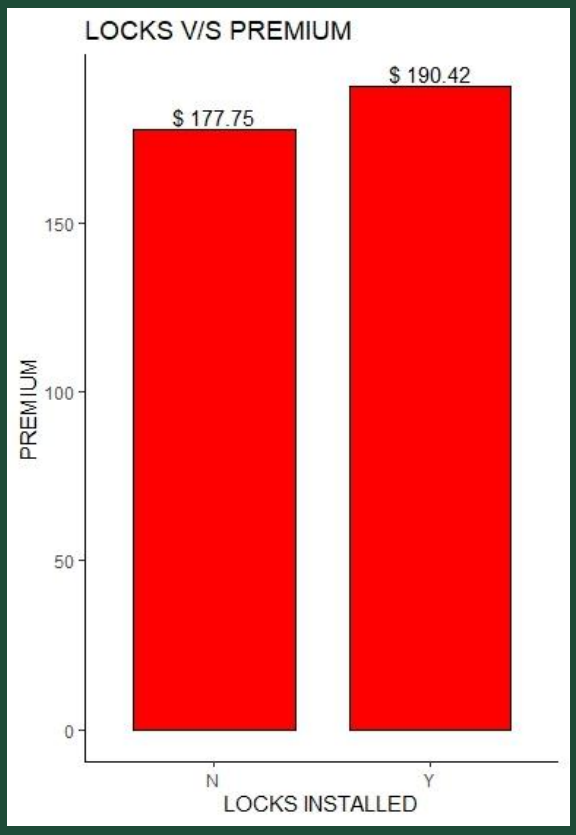
VII. WALL CONSTRUCTION AND ROOF CONSTRUCTION:



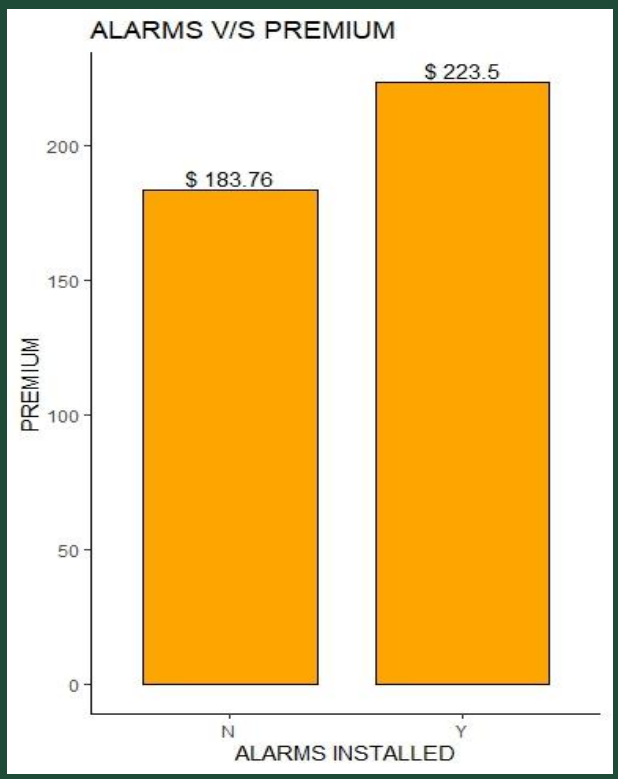
Premium prices are heavily affected by the quality of Roof and Wall Construction. From above fig. it is clear that houses with 'Wall Construction – 99' and 'Roof Construction – 12' have the highest average premium prices. Meaning customers with houses with these characteristics will pay more average premium prices than those who don't.

VIII. HOUSES WITH SAFE'S ,LOCKS AND ALARMS :





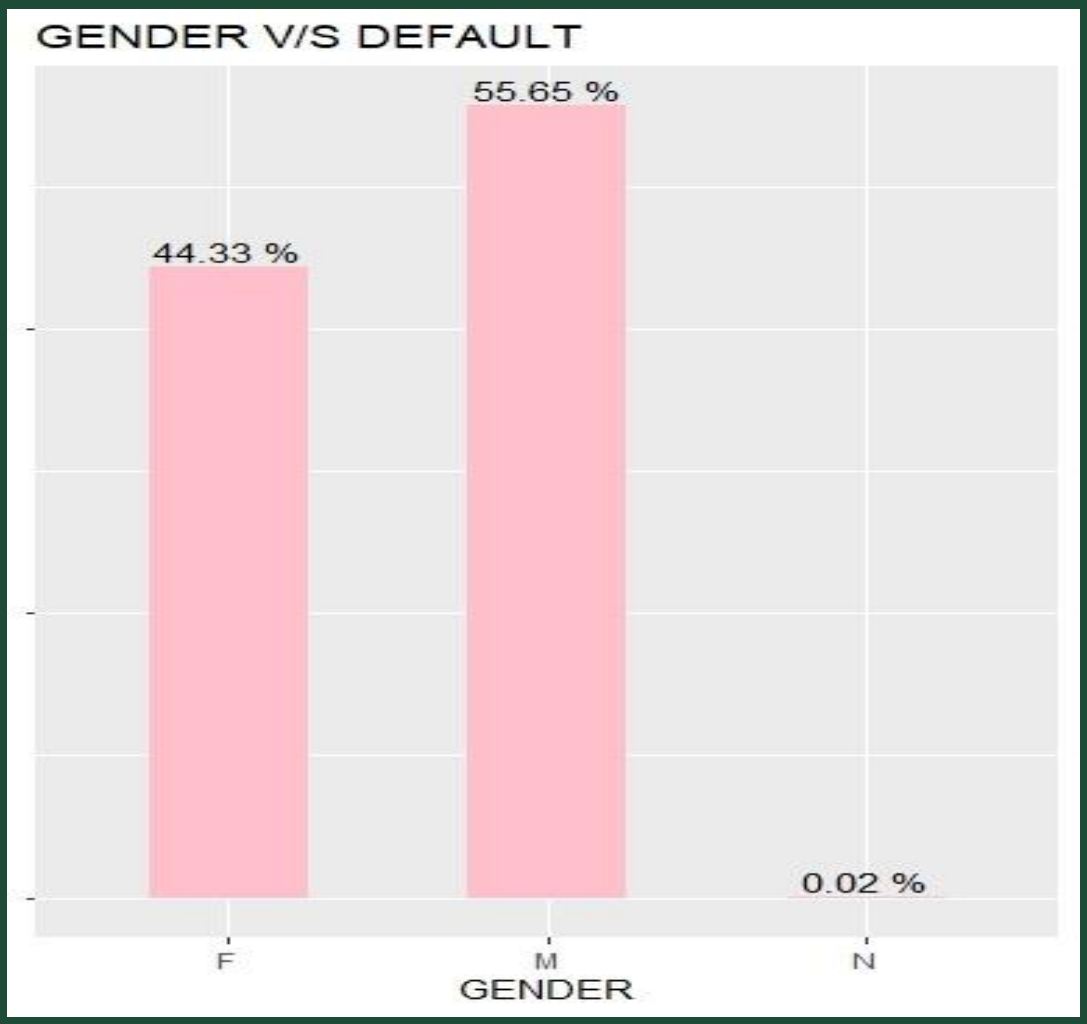
Houses with Safe's, locks and alarms have more avg. premium prices than those which don't have.



b) Analysis Of Customers Likely To Default: For this analysis factors related to customers like gender, employment status , martial status were analysed. Also other key factors like payment methods, payment frequency and policy status were considered.

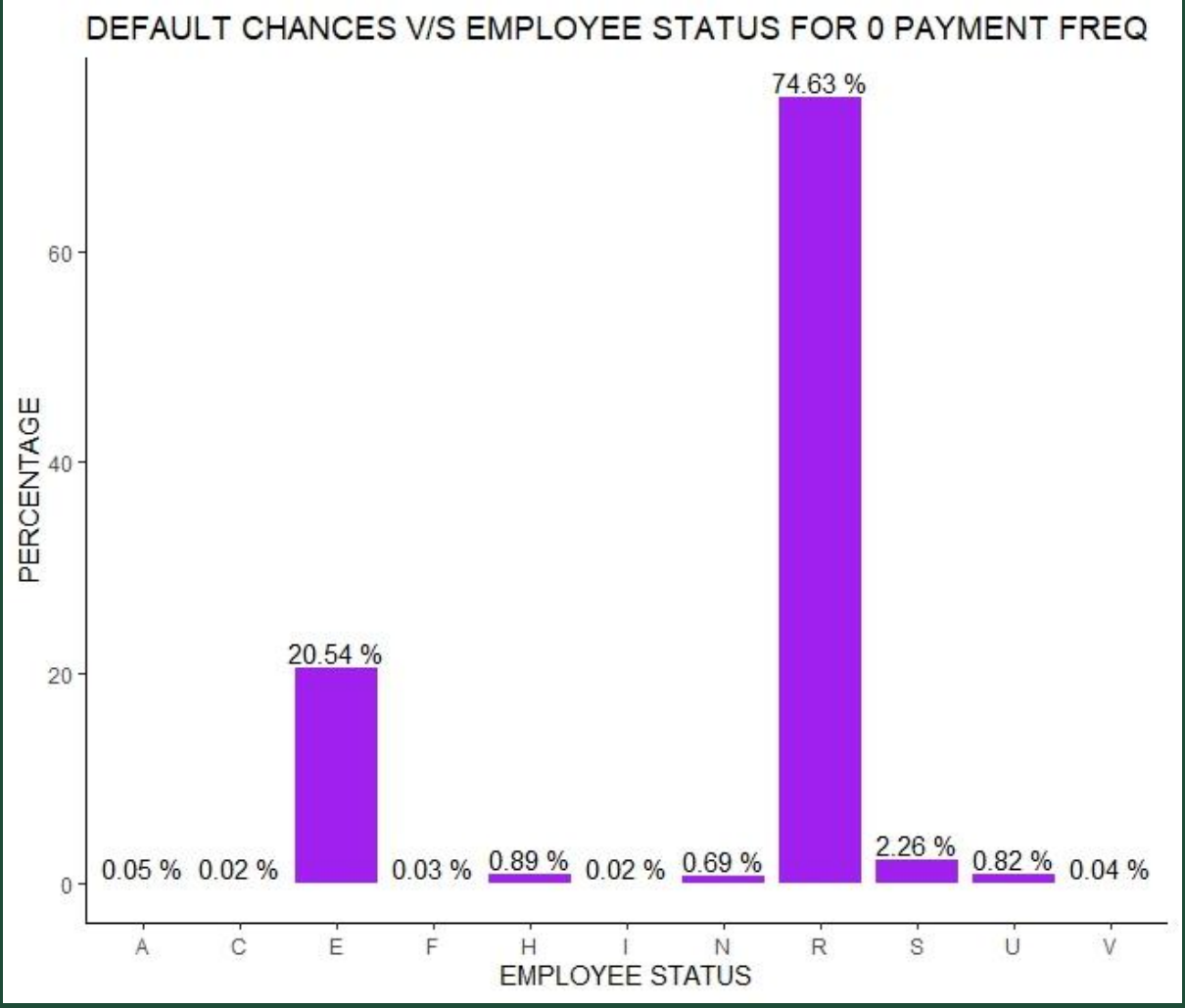
1) Firstly we'll look into default characteristics of customers with policy status : 'Live' and payment frequency : 0 with the help of above mentioned parameters.

i. Gender :



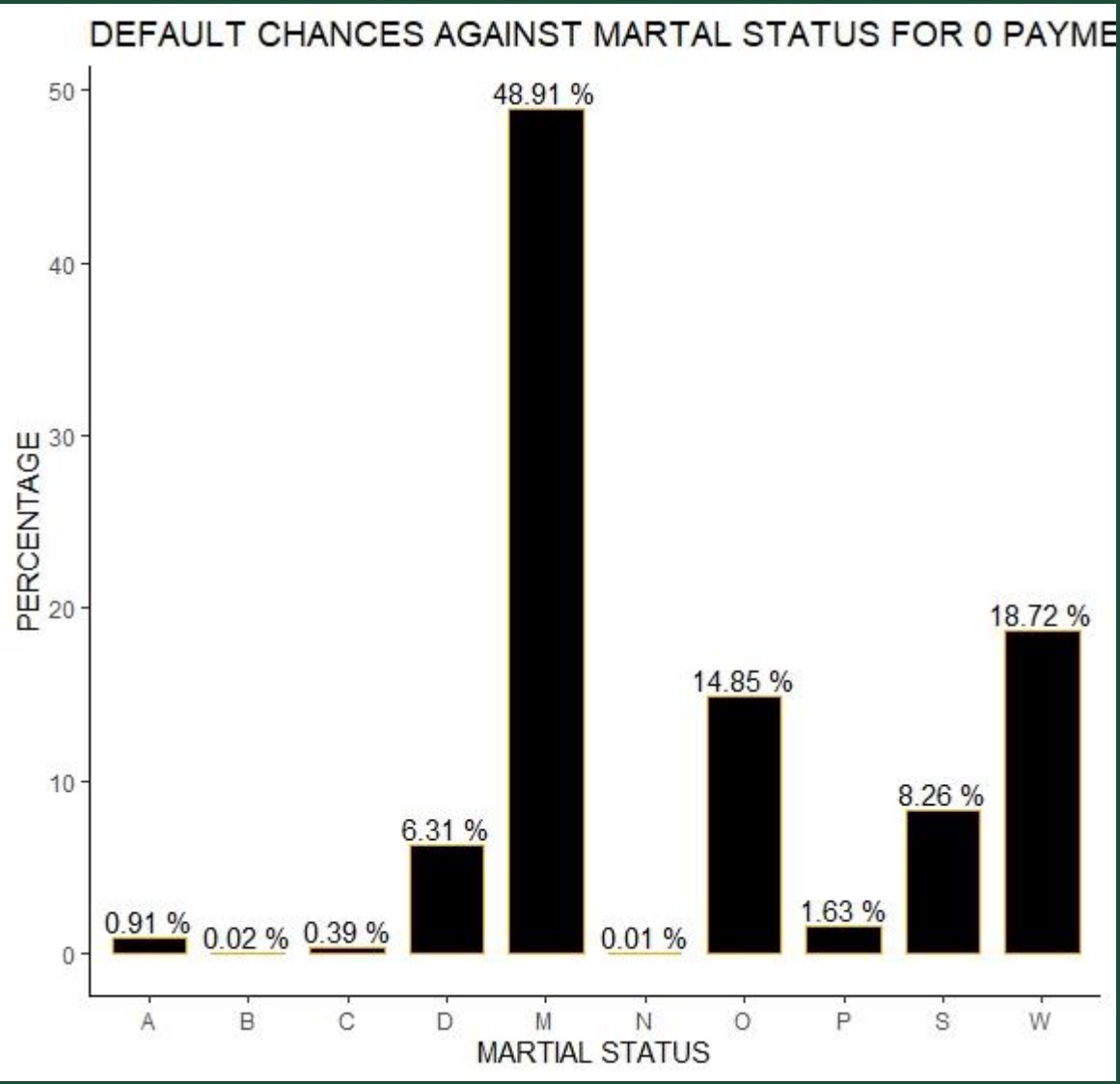
As seen in fig. 55.65% of Male customers couldn't make their payments frequently and 44.33% of Female customers couldn't make frequent payments.

ii. Employee Status:



As seen from the fig the highest no of customers who couldn't make their payments frequently were Retired with a percentage of 74.63%. And 20.54% of them were employed.

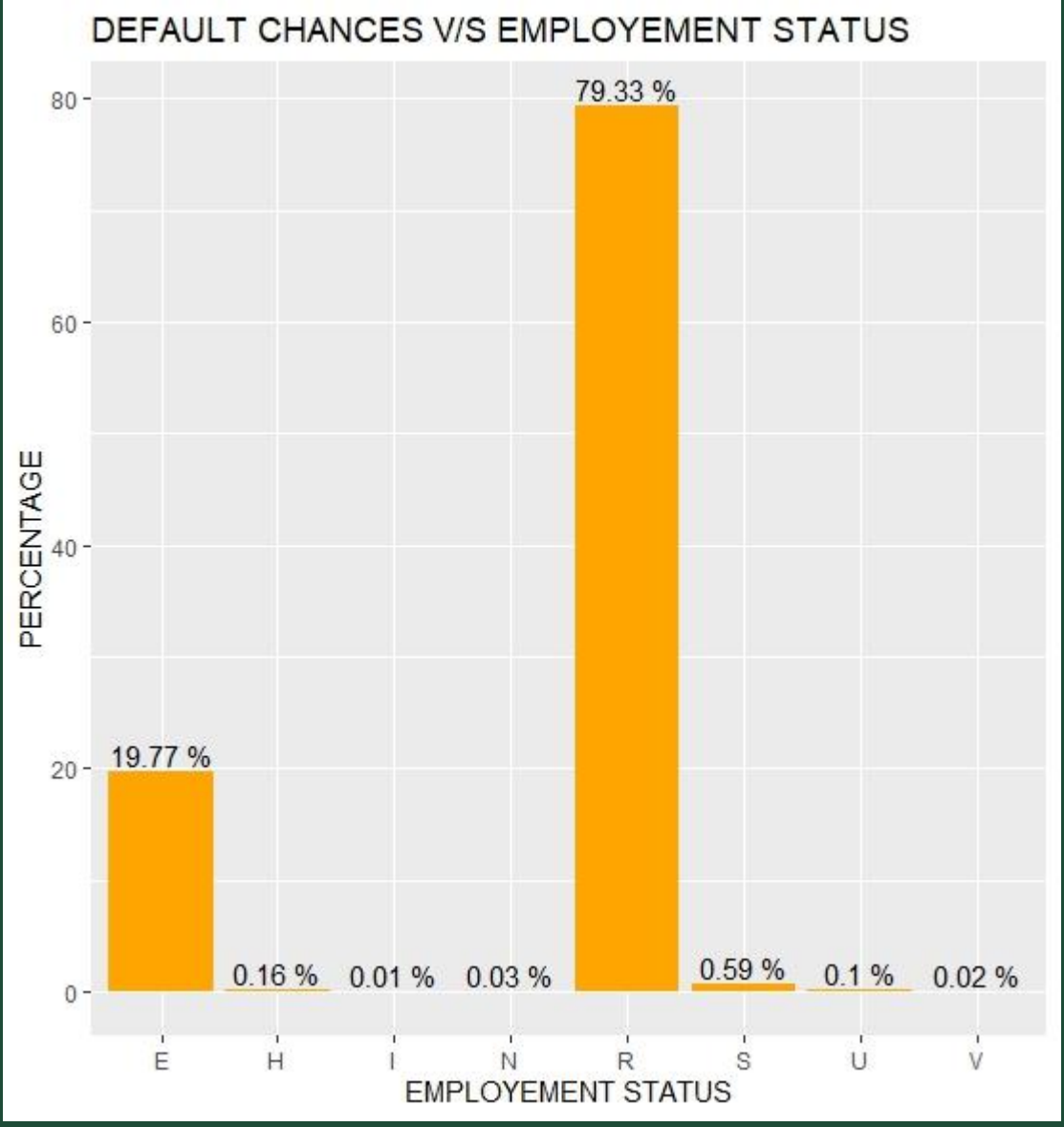
iii. Martial Status :



Of all the customers who failed to make frequent payments, 49% were Married, 18.72% were Widowed , 14.85% were Other and 8.26% were Single.

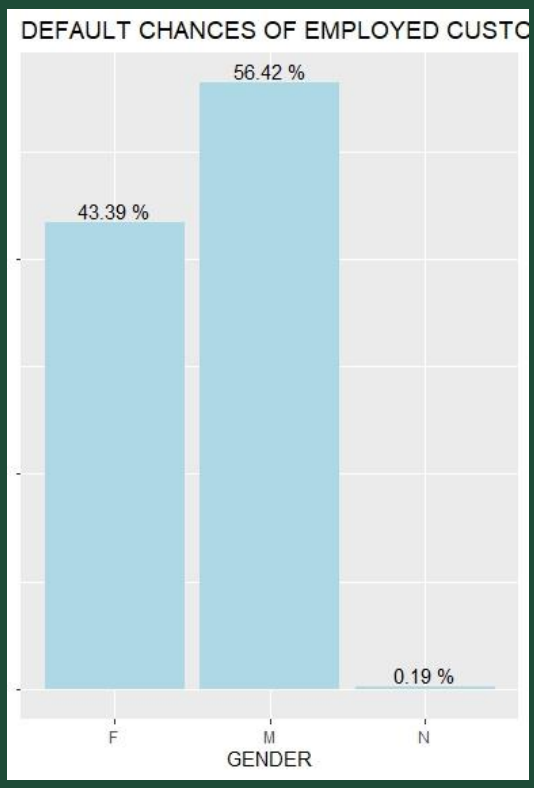
2) Now we'll look into default characteristics of customers with Policy Status : 'Lapsed' & 'Cancelled' and Payment Frequency : 1 in order to analyse

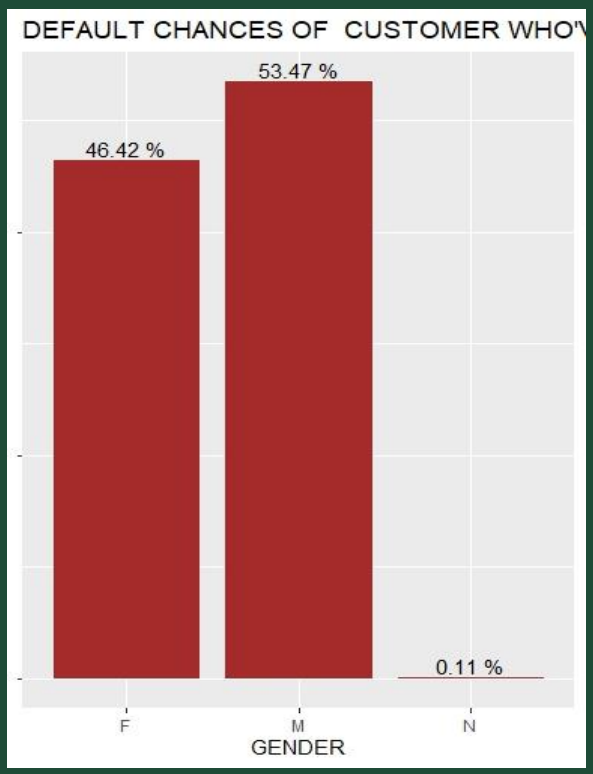
i. Employment Status :



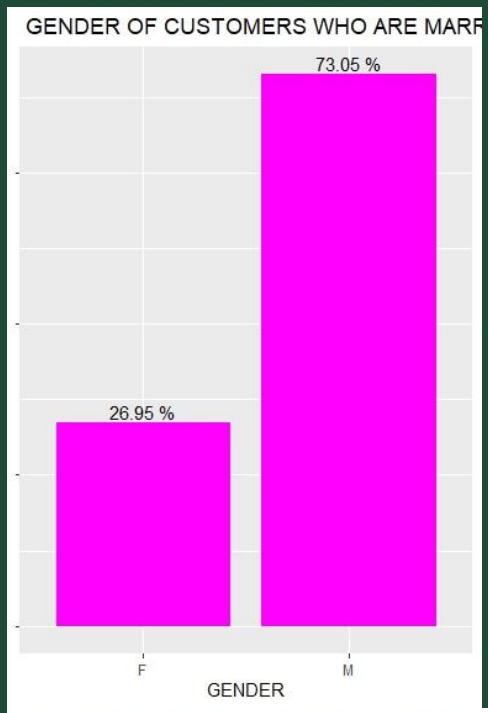
Of all the customers who cancelled their policies about 79%(14,653) were Retired and 19.77%(3651) were Employed

Of those 19.77% employed customers 56.42%(2060) are Males and 43.39%(1584) are Females

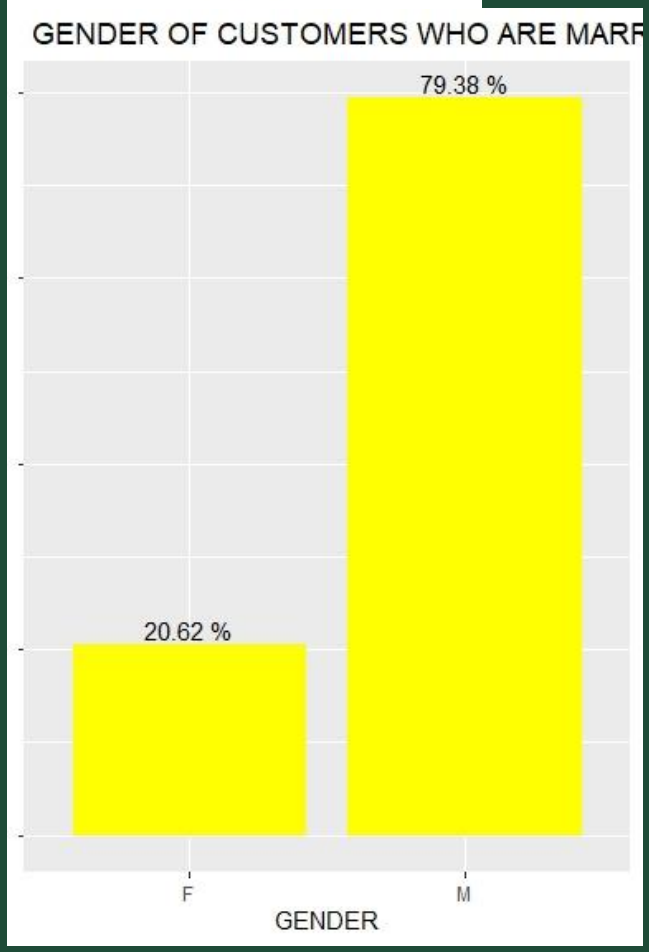




And of the 79% of the Retired customers 53.47%(7835) are Males and 46.42%(6802) are Females.

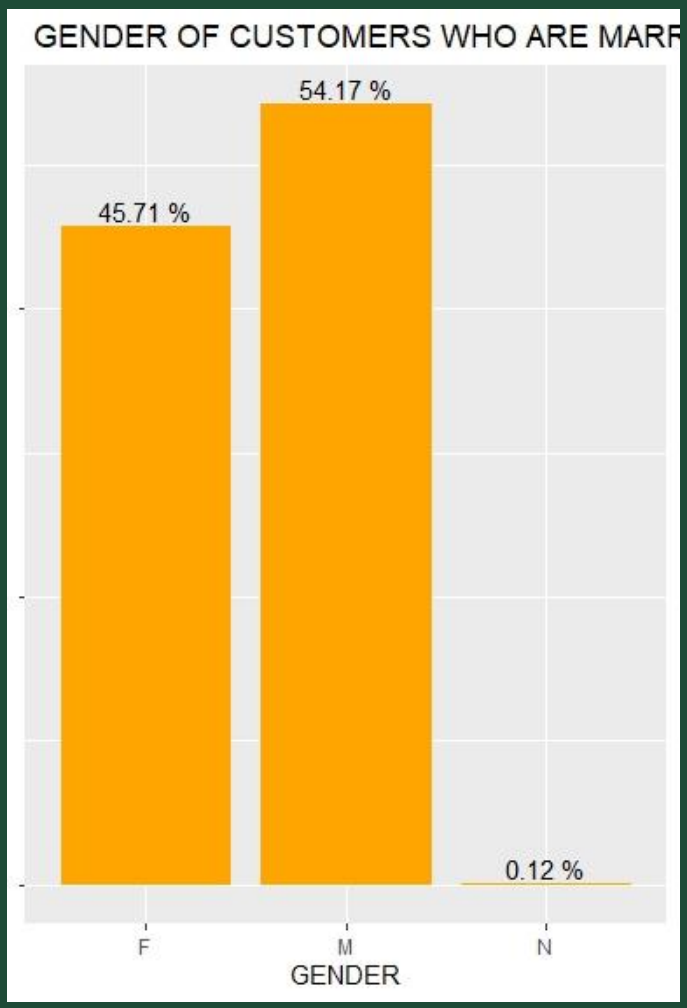


Of customers who are 1.Married and 2. Employed(642): 73.05% (469)of them are Males and 26.95% (173)are Female.



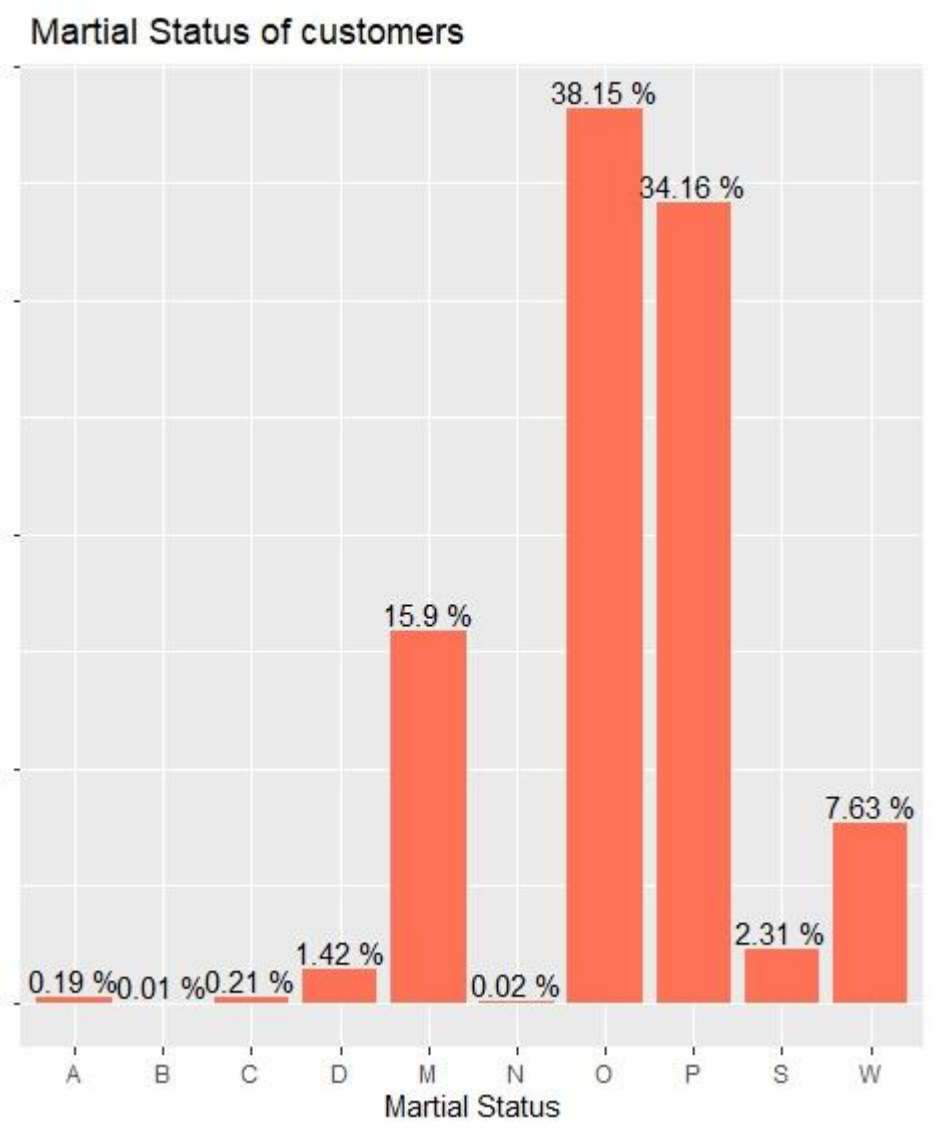
Of customers who are 1. Retired & 2. Married(2192) : 79.40%(1740) are Males and 20.60% are Females

ii. Gender : For the criteria where policy status = 'Lapsed' / 'Cancelled' and Payment Frequency = 1 :



The total no. Of customers are 18471. Of these 54.17%(10005) are Males and 45.71%(8443) are Female.

iii. Martial Status:



For the criteria where policy status = 'Lapsed' / 'Cancelled' and Payment Frequency = 1 :

Out of 18471 customers 38.15%(7046) fell into 'Other' category and 34.16%(6310) fell into 'Domestic Partner' category and 15.9%(2936) were married.

3 . Conclusion :

From this analysis we found out some key factors affecting the premium prices of an insurance policy. Also we found some insights about factors which reveal possibilities of customer defaults.

Flooding , Presence Of Neighbourhood in your vicinity, Construction types(Wall & Roof) , Installation of Safe's , Locks and Alarms, No. of Bedrooms and Subsidence have a clear impact on the premium prices. Except for Flooding and Subsidence all the remaining factors drive up the premium prices.

As for customers who couldn't make payments are : Retired, Married and Male. And the customers who cancelled their policies despite making frequent payments are : Retired , Male and have a martial status of 'Other'.