**CASE STUDY FOR WEEK 4:**

Code for Week4:

==============

#1. Importing necessary library and data

library(dplyr)

library(data.table)

house <- read.csv("C:/Users/I340968/Desktop/ed-code/R programming-elective/Week\_4/housingdata\_v2.0.csv")

View(house)

#2. a) Number of observations - 505

nrow(house)

#2. b) total number of variables - 11

ncol(house)

#3. c) number of continous variable

lapply(house,class)

#3.select and mutate functions

#selecting gender, education and income

house\_subset=select(house,Gender,Education,Income)

head(house\_subset,5)

#selecting coloumn from gender to loan period

house\_subset1 = select(house,Gender:Loan\_Period)

head(house\_subset1,5)

#selecting all coloumn except record and gender

house\_subset2=house[, !names(house) %in% c("Record", "Gender")]

View(house\_subset2)

#============another way==================

house\_subset21=house[,3:ncol(house)]

View(house\_subset21)

#use mutate() new varaiable creation

var1 = mutate(house,g1=PropertyValue/Income)

head(var1,5)

var2 = mutate(house,g2=PropertyValue/Loan\_Period)

head(var2,5)

#4. Filter and Arrange

#4.1.property value based filtering

g3<- filter(house,PropertyValue<80000 | PropertyValue > 150000)

nrow(g3) #to find number of rows

head(g3,5)

#4.2. Property value and income based filtering

g4<- filter(house,PropertyValue>1000000 & Income < 3185)

nrow(g4) #to find number of rows

head(g4,5)

#4.3.Income < 3185 still purchased property

g5<- filter(house,Income < 3185 & Property\_Purchased=='Y')

nrow(g5) #to find number of rows

head(g5,5)

#################arrange()###############

#4.4 bought df - observation when property\_purchased

bought <- filter(house, Property\_Purchased=='Y')

nrow(bought)

#4.5 arrange by income

bought\_income <- arrange(bought,Income)

View(bought\_income)

head(bought\_income,5)

#4.5 arrange bought by geneder

bought\_gender <- arrange(bought,Gender)

View(bought\_gender)

head(bought\_gender,5)

#4.6 group by Gender and education

gpge <- bought%>% group\_by(Gender,Education)

head(gpge,5)

#4.7 Property not purchased and observation.

not\_bought <- filter(house, Property\_Purchased=='N')

nrow(not\_bought)

#4.8 not\_bought arrange by income

not\_bought\_income <- arrange(not\_bought,Income)

View(not\_bought\_income)

head(not\_bought\_income,5)

#4.9 not\_bought arrangement by gender

not\_bought\_Gender <- arrange(not\_bought,Gender)

View(not\_bought\_Gender)

head(not\_bought\_Gender,5)

#4.10 not\_bought group by gender and education

ngpge <- not\_bought%>% group\_by(Gender,Education)

head(gpge,5)

#4.11 arrange house according to gender and desc income

housegi <- arrange(house,Gender,desc(Income))

View(housegi)

head(housegi,5)

#5.Summarise function:

#5.1 MIN and MAX income

summarise(house,min\_income=min(Income),max\_income=max(Income))

#5.2 MIN,MAX,Avg,SD,IQR

summarise(house,min\_income=min(Income),max\_income=max(Income),SD\_inocme=sd(Income),

avg\_income=sum((Income)/nrow(house)),IQR\_income=IQR(Income))

summarise(house,min\_pv=min(PropertyValue),max\_pv=max(PropertyValue),SD\_pv=sd(PropertyValue),

avg\_pv=sum(PropertyValue)/nrow(house),IQR\_pv=IQR(PropertyValue))

summarise(house,min\_lp=min(Loan\_Period),max\_lp=max(Loan\_Period),SD\_lp=sd(Loan\_Period),

avg\_lp=sum(Loan\_Period)/nrow(house),IQR\_lp=IQR(Loan\_Period))

#6.Pipeing function

pip\_1 = house %>% mutate(var1=PropertyValue/Income) %>% filter(var1>50)%>%summarise(var\_mean=mean(var1))

pip\_1

#7.pipeing groupby

grouping <- house %>% group\_by(Education)%>% summarise(avg\_income=sum(Income)/nrow(house),avg\_value=sum(PropertyValue)/nrow(house))%>% arrange(avg\_income,avg\_value)

grouping

Console output:

> #1. Importing necessary library and data

>

> library(dplyr)

> library(data.table)

> house <- read.csv("C:/Users/I340968/Desktop/ed-code/R programming-elective/Week\_4/housingdata\_v2.0.csv")

> View(house)

>

> #2. a) Number of observations - 505

> nrow(house)

[1] 505

>

> #2. b) total number of variables - 11

>

> ncol(house)

[1] 11

>

> #3. c) number of continous variable

>

> lapply(house,class)

$Record

[1] "factor"

$Gender

[1] "factor"

$No\_kids

[1] "integer"

$Education

[1] "factor"

$HasCar

[1] "factor"

$Income

[1] "integer"

$PropertyValue

[1] "integer"

$Loan\_Period

[1] "integer"

$Credit\_Record

[1] "integer"

$Housing\_type

[1] "factor"

$Property\_Purchased

[1] "factor"

>

>

> #3.select and mutate functions

>

> #selecting gender, education and income

> house\_subset=select(house,Gender,Education,Income)

> head(house\_subset,5)

Gender Education Income

1 Female Graduate 710

2 Male Graduate 6516

3 Male Graduate 7040

4 Male Not Graduate 4730

5 Male Graduate 9167

>

> #selecting coloumn from gender to loan period

> house\_subset1 = select(house,Gender:Loan\_Period)

> head(house\_subset1,5)

Gender No\_kids Education HasCar Income PropertyValue Loan\_Period

1 Female 0 Graduate No 710 90400 456

2 Male 0 Graduate No 6516 168800 336

3 Male 0 Graduate Yes 7040 160000 336

4 Male 0 Not Graduate No 4730 155200 336

5 Male 0 Graduate No 9167 149600 336

>

>

> #selecting all coloumn except record and gender

> house\_subset2=house[, !names(house) %in% c("Record", "Gender")]

> View(house\_subset2)

>

> #============another way==================

> house\_subset21=house[,3:ncol(house)]

> View(house\_subset21)

>

> #use mutate() new varaiable creation

>

> var1 = mutate(house,g1=PropertyValue/Income)

> head(var1,5)

Record Gender No\_kids Education HasCar Income PropertyValue

1 Record1 Female 0 Graduate No 710 90400

2 Record8 Male 0 Graduate No 6516 168800

3 Record9 Male 0 Graduate Yes 7040 160000

4 Record10 Male 0 Not Graduate No 4730 155200

5 Record11 Male 0 Graduate No 9167 149600

Loan\_Period Credit\_Record Housing\_type Property\_Purchased g1

1 456 1 Affordable Y 127.32394

2 336 1 Affordable Y 25.90546

3 336 1 Affordable Y 22.72727

4 336 1 Affordable Y 32.81184

5 336 1 Affordable Y 16.31941

>

> var2 = mutate(house,g2=PropertyValue/Loan\_Period)

> head(var2,5)

Record Gender No\_kids Education HasCar Income PropertyValue

1 Record1 Female 0 Graduate No 710 90400

2 Record8 Male 0 Graduate No 6516 168800

3 Record9 Male 0 Graduate Yes 7040 160000

4 Record10 Male 0 Not Graduate No 4730 155200

5 Record11 Male 0 Graduate No 9167 149600

Loan\_Period Credit\_Record Housing\_type Property\_Purchased g2

1 456 1 Affordable Y 198.2456

2 336 1 Affordable Y 502.3810

3 336 1 Affordable Y 476.1905

4 336 1 Affordable Y 461.9048

5 336 1 Affordable Y 445.2381

>

> #4. Filter and Arrange

> #4.1.property value based filtering

>

> g3<- filter(house,PropertyValue<80000 | PropertyValue > 150000)

> nrow(g3) #to find number of rows

[1] 198

> head(g3,5)

Record Gender No\_kids Education HasCar Income PropertyValue

1 Record8 Male 0 Graduate No 6516 168800

2 Record9 Male 0 Graduate Yes 7040 160000

3 Record10 Male 0 Not Graduate No 4730 155200

4 Record76 Male 0 Not Graduate Not Answered 2002 76000

5 Record77 Male 0 Graduate No 3474 71200

Loan\_Period Credit\_Record Housing\_type Property\_Purchased

1 336 1 Affordable Y

2 336 1 Affordable Y

3 336 1 Affordable Y

4 336 1 Affordable Y

5 336 1 Affordable Y

>

> #4.2. Property value and income based filtering

>

> g4<- filter(house,PropertyValue>1000000 & Income < 3185)

> nrow(g4) #to find number of rows

[1] 0

> head(g4,5)

[1] Record Gender No\_kids

[4] Education HasCar Income

[7] PropertyValue Loan\_Period Credit\_Record

[10] Housing\_type Property\_Purchased

<0 rows> (or 0-length row.names)

>

> #4.3.Income < 3185 still purchased property

>

> g5<- filter(house,Income < 3185 & Property\_Purchased=='Y')

> nrow(g5) #to find number of rows

[1] 81

> head(g5,5)

Record Gender No\_kids Education HasCar Income PropertyValue

1 Record1 Female 0 Graduate No 710 90400

2 Record13 Male 0 Graduate No 2888 149600

3 Record25 Male 0 Graduate No 3045 124000

4 Record26 Male 0 Not Graduate Not Answered 3184 124000

5 Record29 Male 0 Graduate Yes 2835 121600

Loan\_Period Credit\_Record Housing\_type Property\_Purchased

1 456 1 Affordable Y

2 336 1 Affordable Y

3 336 1 Affordable Y

4 336 1 Affordable Y

5 336 1 Affordable Y

>

> #################arrange()###############

>

> #4.4 bought df - observation when property\_purchased

>

> bought <- filter(house, Property\_Purchased=='Y')

> nrow(bought)

[1] 328

>

> #4.5 arrange by income

> bought\_income <- arrange(bought,Income)

> View(bought\_income)

> head(bought\_income,5)

Record Gender No\_kids Education HasCar Income PropertyValue

1 Record202 Female 2 Not Graduate Not Answered 231 78400

2 Record1 Female 0 Graduate No 710 90400

3 Record60 Male 0 Graduate No 1128 89600

4 Record313 Male 1 Graduate No 1788 76800

5 Record155 Male 0 Graduate No 1935 104800

Loan\_Period Credit\_Record Housing\_type Property\_Purchased

1 336 1 Mid Range Y

2 456 1 Affordable Y

3 336 1 Affordable Y

4 336 1 Premium Y

5 336 1 Mid Range Y

>

> #4.5 arrange bought by geneder

> bought\_gender <- arrange(bought,Gender)

> View(bought\_gender)

> head(bought\_gender,5)

Record Gender No\_kids Education HasCar Income PropertyValue

1 Record1 Female 0 Graduate No 710 90400

2 Record27 Female 0 Not Graduate No 4785 123200

3 Record38 Female 0 Not Graduate Yes 7857 110400

4 Record41 Female 0 Graduate No 4139 108000

5 Record42 Female 0 Graduate No 5500 105600

Loan\_Period Credit\_Record Housing\_type Property\_Purchased

1 456 1 Affordable Y

2 336 1 Affordable Y

3 336 1 Affordable Y

4 336 1 Affordable Y

5 336 1 Affordable Y

>

> #4.6 group by Gender and education

> gpge <- bought%>% group\_by(Gender,Education)

> head(gpge,5)

# A tibble: 5 x 11

# Groups: Gender, Education [3]

Record Gender No\_kids Education HasCar Income PropertyValue Loan\_Period

*<fct>* *<fct>* *<int>* *<fct>* *<fct>* *<int>* *<int>* *<int>*

1 Recor~ Female 0 Graduate No 710 90400 456

2 Recor~ Male 0 Graduate No 6516 168800 336

3 Recor~ Male 0 Graduate Yes 7040 160000 336

4 Recor~ Male 0 Not Grad~ No 4730 155200 336

5 Recor~ Male 0 Graduate No 9167 149600 336

# ... with 3 more variables: Credit\_Record *<int>*, Housing\_type *<fct>*,

# Property\_Purchased *<fct>*

>

> #4.7 Property not purchased and observation.

> not\_bought <- filter(house, Property\_Purchased=='N')

> nrow(not\_bought)

[1] 177

>

> #4.8 not\_bought arrange by income

> not\_bought\_income <- arrange(not\_bought,Income)

> View(not\_bought\_income)

> head(not\_bought\_income,5)

Record Gender No\_kids Education HasCar Income PropertyValue

1 Record370 Male 0 Graduate No 165 108000

2 Record468 Male 1 Graduate Yes 1100 88000

3 Record349 Male 2 Graduate No 1429 13600

4 Record462 Female 2 Graduate No 1516 133600

5 Record479 Male 0 Not Graduate No 1587 28000

Loan\_Period Credit\_Record Housing\_type Property\_Purchased

1 336 1 Affordable N

2 336 1 Premium N

3 96 1 Premium N

4 336 1 Premium N

5 336 1 Premium N

>

> #4.9 not\_bought arrangement by gender

>

> not\_bought\_Gender <- arrange(not\_bought,Gender)

> View(not\_bought\_Gender)

> head(not\_bought\_Gender,5)

Record Gender No\_kids Education HasCar Income PropertyValue

1 Record330 Female 0 Graduate No 1993 43200

2 Record335 Female 0 Graduate No 3561 24000

3 Record354 Female 0 Graduate No 5500 120800

4 Record361 Female 0 Graduate No 11000 180000

5 Record362 Female 0 Graduate Yes 8186 155200

Loan\_Period Credit\_Record Housing\_type Property\_Purchased

1 336 1 Premium N

2 336 1 Premium N

3 456 1 Affordable N

4 336 1 Affordable N

5 336 1 Affordable N

>

> #4.10 not\_bought group by gender and education

> ngpge <- not\_bought%>% group\_by(Gender,Education)

> head(gpge,5)

# A tibble: 5 x 11

# Groups: Gender, Education [3]

Record Gender No\_kids Education HasCar Income PropertyValue Loan\_Period

*<fct>* *<fct>* *<int>* *<fct>* *<fct>* *<int>* *<int>* *<int>*

1 Recor~ Female 0 Graduate No 710 90400 456

2 Recor~ Male 0 Graduate No 6516 168800 336

3 Recor~ Male 0 Graduate Yes 7040 160000 336

4 Recor~ Male 0 Not Grad~ No 4730 155200 336

5 Recor~ Male 0 Graduate No 9167 149600 336

# ... with 3 more variables: Credit\_Record *<int>*, Housing\_type *<fct>*,

# Property\_Purchased *<fct>*

>

> #4.11 arrange house according to gender and desc income

> housegi <- arrange(house,Gender,desc(Income))

> View(housegi)

> head(housegi,5)

Record Gender No\_kids Education HasCar Income PropertyValue

1 Record96 Female 1 Graduate Yes 21433 480000

2 Record293 Female 0 Not Graduate Yes 19982 100000

3 Record109 Female 0 Not Graduate Yes 18990 180000

4 Record223 Female 0 Graduate Yes 17335 44000

5 Record324 Female 2 Graduate No 16353 56000

Loan\_Period Credit\_Record Housing\_type Property\_Purchased

1 336 1 Mid Range Y

2 336 1 Premium Y

3 336 1 Mid Range Y

4 336 1 Mid Range Y

5 336 1 Premium Y

>

> #5.Summarise function:

> #5.1 MIN and MAX income

> summarise(house,min\_income=min(Income),max\_income=max(Income))

min\_income max\_income

1 165 89100

>

> #5.2 MIN,MAX,Avg,SD,IQR

> summarise(house,min\_income=min(Income),max\_income=max(Income),SD\_inocme=sd(Income),

+ avg\_income=sum((Income)/nrow(house)),IQR\_income=IQR(Income))

min\_income max\_income SD\_inocme avg\_income IQR\_income

1 165 89100 6740.648 5952.921 3215

>

>

> summarise(house,min\_pv=min(PropertyValue),max\_pv=max(PropertyValue),SD\_pv=sd(PropertyValue),

+ avg\_pv=sum(PropertyValue)/nrow(house),IQR\_pv=IQR(PropertyValue))

min\_pv max\_pv SD\_pv avg\_pv IQR\_pv

1 7200 480000 65044.94 115653.1 53600

>

>

> summarise(house,min\_lp=min(Loan\_Period),max\_lp=max(Loan\_Period),SD\_lp=sd(Loan\_Period),

+ avg\_lp=sum(Loan\_Period)/nrow(house),IQR\_lp=IQR(Loan\_Period))

min\_lp max\_lp SD\_lp avg\_lp IQR\_lp

1 12 456 64.90034 317.8693 0

>

> #6.Pipeing function

> pip\_1 = house %>% mutate(var1=PropertyValue/Income) %>% filter(var1>50)%>%summarise(var\_mean=mean(var1))

> pip\_1

var\_mean

1 112.4228

>

> #7.pipeing groupby

> grouping <- house %>% group\_by(Education)%>% summarise(avg\_income=sum(Income)/nrow(house),avg\_value=sum(PropertyValue)/nrow(house))%>% arrange(avg\_income,avg\_value)

> grouping

# A tibble: 2 x 3

Education avg\_income avg\_value

*<fct>* *<dbl>* *<dbl>*

1 Not Graduate 865. 19148.

2 Graduate 5088. 96505.