MGT7178: Data Management

Customers Across Insurance Products

Assignment 1

Name: Muhammad Muneeb Ullah Ansari
Student Number: 40426685

Word Count: 2156

Contents

Introduction

Insurance companies provide their products and services through insurance contracts designed for risk management. In these contracts, the insurer assumes responsibility for covering costs in the event of an uncertain occurrence, while the policyholder pays a smaller premium to the insurer in return for this protection.

The dataset furnished encompasses details regarding customers (including general characteristics such as Age, Gender, Location, and Card type), Motor policies (covering aspects like Claims, Last claim date, Policy start and end dates, Vehicle value, body, and age), Health policies (encompassing Health type, Policy start and end dates, Dependent adults, etc.), and Travel policies (including Travel type, Policy start and end dates).

The study's objective is to assess the correlation between customer characteristics and various policies. In this pursuit, Section 1 involves a descriptive analysis using the SQL programming language, while Section 2 addresses data quality concerns through diverse methods utilizing the "tidyverse" package, which incorporates "dplyr" and "tidyr" to manipulate data. Section 3 includes a Conclusion, References and Appendix.

Section 1: SQL

Forming an ABT

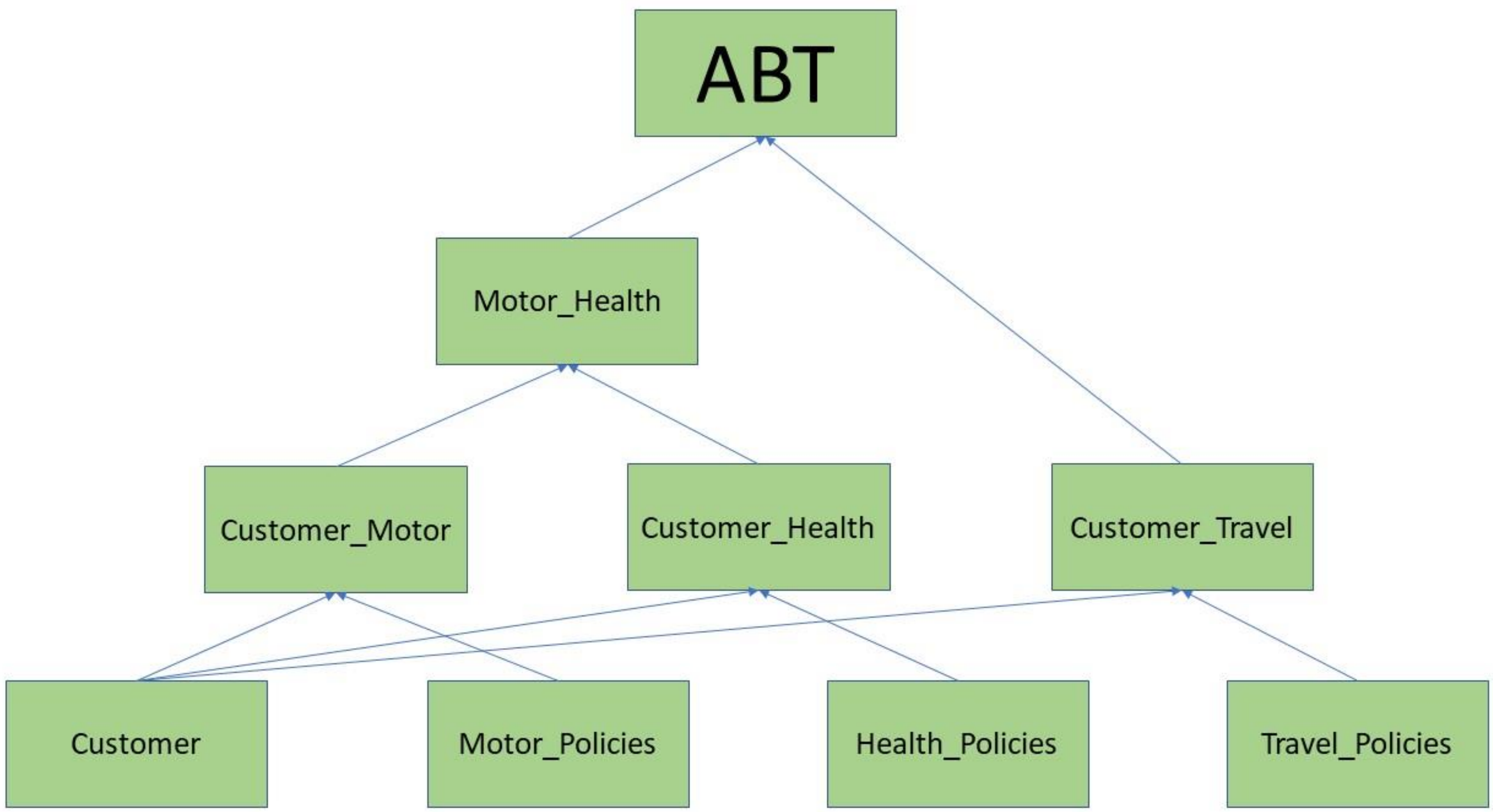
Microsoft Access was used to combined 4 datasets together to form an ABT using left join. We could have also used only the left join.2 datasets were combined and then the resultant was combined with another dataset until all the 4 datasets were completely joined. Some duplicated variables were deleted and the remaining variables were appropriately named.

The Primary Key in the Customer dataset was CustomerID. The MotorID, HealthID, TravelID variables in the dataset Motor_Policies, Health_Policies and Travel_Policies respectively were both Primary and Foreign keys. That is because those IDs uniquely identified entries within those tables but also provided a link between the Customer dataset.

Click Code 1 to view in Appendix.

The following is a representation of how the datasets were combined.

(This space has been intentionally left blank.)


ABT
Motor_Health
Customer_Motor
Customer_Health
Customer_Travel
Customer
Motor_Policies
Health_Policies
Travel_Policies

Data Analysis

Click Code 2 to view in Appendix.

PolicyType	AverageAge
Health	47.2762017336485
Motor	41.7536490914507
Travel	38.762945368171

It was found that the people opting for Health Policies were highest in age while those opting for Travel Policies were relatively younger. This makes sense because at a later age people are at a higher risk of medical complications while younger people tend to travel more.

Click Code 3 to view in Appendix.

PolicyType	Gender	GenderCount
Health		5
Health	f	7
Health	female	1289
Health	m	9
Health	male	1233
Motor		4
Motor	f	10
Motor	female	1687
Motor	m	8
Motor	male	1652
Travel		3
Travel	f	8
Travel	female	1061
Travel	m	7
Travel	male	1029

For all of the policies, the split of gender was almost 50-50. However for each of them, the number of women opting for the policies was slightly higher. This might suggest that communications made could focus more on women as a possible demographic.

Click Code 4 to view in Appendix.

Location	PolicyCount
	12
Rural	1707
Urban	2255

Most of the policies were bought in urban areas. Hence if marketing communications need to be created, they need to be more accustomed to the urban customer.

Click Code 5 to view in Appendix.

ComChannel	PolicyCount
	12
E	5
Email	1712
P	5
Phone	1526
S	2
SMS	712

In absolute terms, most policy holders preferred email for communications and least of them preferred SMS. This could be helpful if there needs to be a decision regarding the mode of communication for a message that needs to be sent to every customer across the board.

Click Code 6 to view in Appendix.

HealthType	AverageDependentsAdults
Level2	0.893939393939394
Level3	0.814992025518341
Level1	0.670694864048338

Customers that opted for a level 2 health policy had the highest number of adult dependents while level 1 had the lowest number. There can be efforts made to enhance level 2 health insurance terms so that it becomes more attractive to customers with dependents.

Click Code 7 to view in Appendix.

ComChannel	MotorPolicy
	4
E	4
Email	1495
P	5
Phone	1285
S	1
SMS	567

Email was most preferred by motor and travel customers. This could suggest that they engage more with the communication methods and want dedicated threads.

Click Code 8 to view in Appendix.

ComChannel	HealthPolicy
	5
E	3
Email	1051
P	4
Phone	1215
S	2
SMS	263

Phone was most preferred by customers opting for health policies. Usually such customers are more elderly and hence might be most comfortable with phone calls. Also unfortunate events that require health policies are also emergencies which require an immediate communication channel.

Click Code 9 to view in Appendix.

ComChannel	TravelPolicy
	3
E	4
Email	1003
P	2
Phone	637
S	1
SMS	458

Click Code 10 to view in Appendix.

Most of the policies were bought in urban areas which is intuitive because urban areas have a higher population density and more vehicles.

Location	MotorPolicyCount
	4
Rural	1449
Urban	1908

Click Code 11 to view in Appendix.

Location	HealthPolicy
	5
Rural	1199
Urban	1339

Click Code 12 to view in Appendix.

Location	TravelPolicy
	3
Rural	799
Urban	1306

Click Code 13 to view in Appendix.

Occupation	MotorPolicyCount
	1261
Broker	5
U.S. deputy marshal	5
Catering manager	4
Construction equipment technician	4
Training manager	4
College counselor	4
Poultry cutter	4
Systems analyst	4
Farm and home management advisor	4
Life Guard	4
Clinical social worker	4
Typesetting machine operator	4
HIV/AIDS nurse	4
Executive recruiter	4
Medical and health services manager	4
Ornamental ironworker	4
Welding machine tender	4
Technical training coordinator	4
Fire fighter	4
Drug Enforcement Administration (DEA	4
ATF agent	4
Railcar repairer	4
Social work planner	4

Motor Policies were most bought by occupations that required more day to day traveling as that results in greater risk like Brokers, U.S deputy marshal, catering manager etc.

Click Code 14 to view in Appendix.

Occupation	HealthPolicy
	966
Physiologist	4
Training manager	4
HIV/AIDS nurse	4
Occupational therapist aide	4
Electronics repairer	4
Welding machine tender	4
Ornamental ironworker	4
Poultry cutter	4
Drug Enforcement Administration (DEA	4
Rancher	4
Payroll representative	3
Hostler	3
Tower controller	3
Tractor-trailer driver	3
College counselor	3
Claims representative	3
Clinical social worker	3
Plant manager	3
Buyer	3
Personnel recruiter	3
Passenger rate clerk	3
Fire inspector	3
Aircraft surfaces assembler	3

Health Policies were most bought by occupations that were at a health risk like training manager, HIV/AIDS nurse, electronics repairer, welding machine tender etc.

Click Code 15 to view in Appendix.

Occupation	TravelPolicy
	802
Social work planner	5
Railroad signal operator	4
HIV/AIDS nurse	4
Farm and home management advisor	4
Training manager	4
Executive recruiter	4
Amusement machine repairer	4
Court municipal and license clerk	4
ATF agent	3
Textile knitting and weaving machine s	3
Engraver set-up operator	3
Marketing research analyst	3
Experimental psychologist	3
Training and development coordinator	3
Physiologist	3
Apartment leasing agent	3
Fence erector	3
Life Guard	3
Tower controller	3
Cost consultant	3
Regional planner	3
Clinical social worker	3
College counselor	3

Travel Policies were most bought by occupations that required more day to day traveling like social work planner, railroad signal operator, farm and home management advisor, executive recruiter etc.

Click Code 16 to view in Appendix.

Most customers used Mastercard and visa to pay for their policies. This can be an opportunity to offer loyalty programs through Mastercard and Visa.

CardType	Motor_Card
	4
0	579
Mastercard	1422
Visa	1356

Click Code 17 to view in Appendix.

CardType	Health_Card
	5
0	469
Mastercard	1064
Visa	1005

Click Code 18 to view in Appendix.

CardType	Travel_Card
	3
0	355
Mastercard	899
Visa	851

A higher percentage of customers opted to pay in cash for travel policies compared to other policies. This is counter intuitive as debit and credit cards make overseas transactions easy which might be more frequently occurring for customers that travel more.

Section 2: R

Forming an ABT

The program used for R coding was RStudio. For forming an ABT, a left join was opted for and all the datasets were combined all together at once in a single code utilizing the DPLYR pipeline.

Click Code 19 to view in Appendix.

Data Cleaning

Click Code 20 to view in Appendix.

For identifying problems with the data, some summary characteristics were drawn out along with the number of missing values in the data. This was done along with identifying unique values for each variable and complimentary graphs.

It was found out that there was inconsistent naming in the title, gender and ComChannel. There were outliers in age, dependents kids, veh_value. Every entry with null CustomerID was entirely null which needed to be removed. Some variables that should have been as factors were either numeric or characters. Since every table had the policy_start and policy_end variable, those variables needed to be renamed so that it was clear which policy_start and policy_end belonged to which policy.

The following were the changes made to clean the data:

- Mr replaced by Mr.
- Male replaced by m
- Female replaced by f
- Data filtered to only give values of age greater than 0 and less than 84
- E was replaced by Email
- P was replaced By Phone
- S was replaced by SMS
- Data was filtered to only give values of DependentsKids less than 10
- Data was filtered to only give values of veh_value less than 8
- Title, Gender, CardType, Location, ComChannel, MotorType, clm, v_body, HealthType, TravelType and v_age was converted to factors
- Data was filtered to remove rows with null customer ids
- policyStart.x was renamed to policyStart_health
- policyEnd was renamed to policyEnd_health
- policyStart was renamed to policyStart_motor
- policyEnd.x was renamed to policyEnd_motor
- polictStart.y was renamed to policyStart_travel
- policyEnd.y was renamed to policyEnd_travel

To rename the policyStart and policyEnd, we needed to look at an observation with only HealthID which would imply that the policyStart and policyEnd for that entry would be for health policy. Then we look for an observation with no TravelID which would suggest that start and end for that observation would either be motor policy or health policy which we can easily identify. Once 2 policies are identified, the remaining policy will be travel policy.

We do need to mention that some missing values in the data are significant and cannot be removed. That is because not every customer might have a TravelID because they might not have opted for a travel policy. Every customer might not have a veh_value since they have not opted for a motor policy and so on. Hence we need to be careful when removing NAs.

Analysis

The ABT formed has a lot of variables that are both general and specific to insurances. For this task since we only need to figure out differences in characteristics between people opting for different insurances and preferred communication channels, we do not need to figure out relationships between insurance specific variables.

Click Code 21 to view in Appendix.

We need to create a variable that denotes 0 or 1 if there is a start date for each respective policy which is also known as dummy coding.

Click Code 22 to view in Appendix.

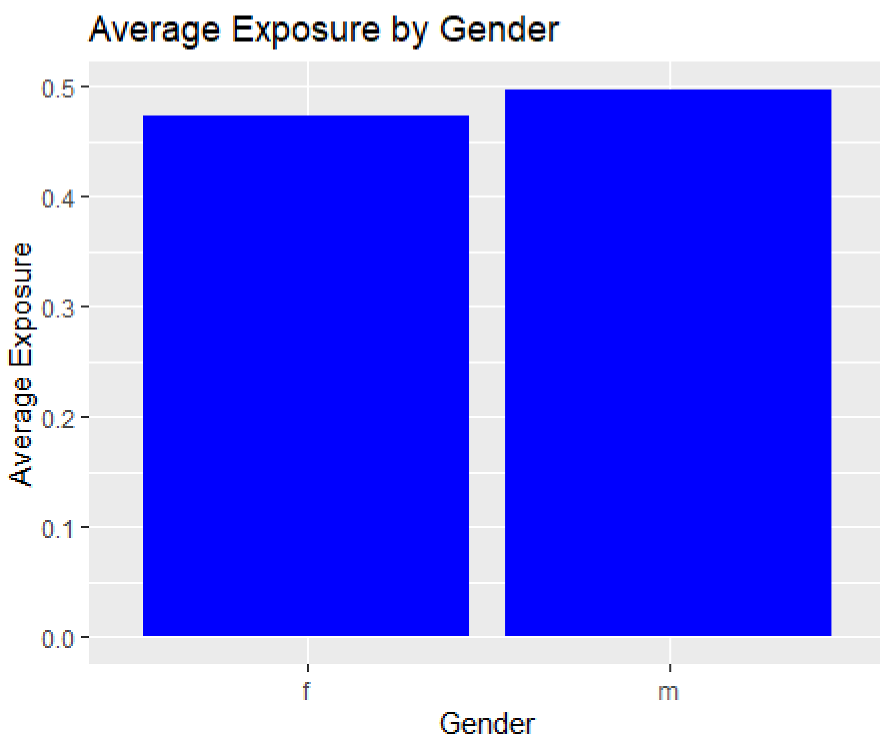
If we try to figure out if there is a relationship between the age of customers and the value of the vehicle that they own, it would be helpful to derive a correlation coefficient between the two variables.

It was found out that there is no relationship between age of customers and the value of the vehicle that they have (Correlation Coefficient of -0.03433799) The same can be said for the relationship between age and the number of claims that the customers make (Correlation Coefficient of 0.01137241)

Click Code 23 to view in Appendix.

If we focus on customer age and vehicle age, we would be able to tell if there is a relationship between how long the customer had a car for and whether they want to buy an insurance for it. But on the flip side, calculating correlation coefficient between customer age and vehicle age indicates no relationship between the variables (Correlation Coefficient of 0.02330586).

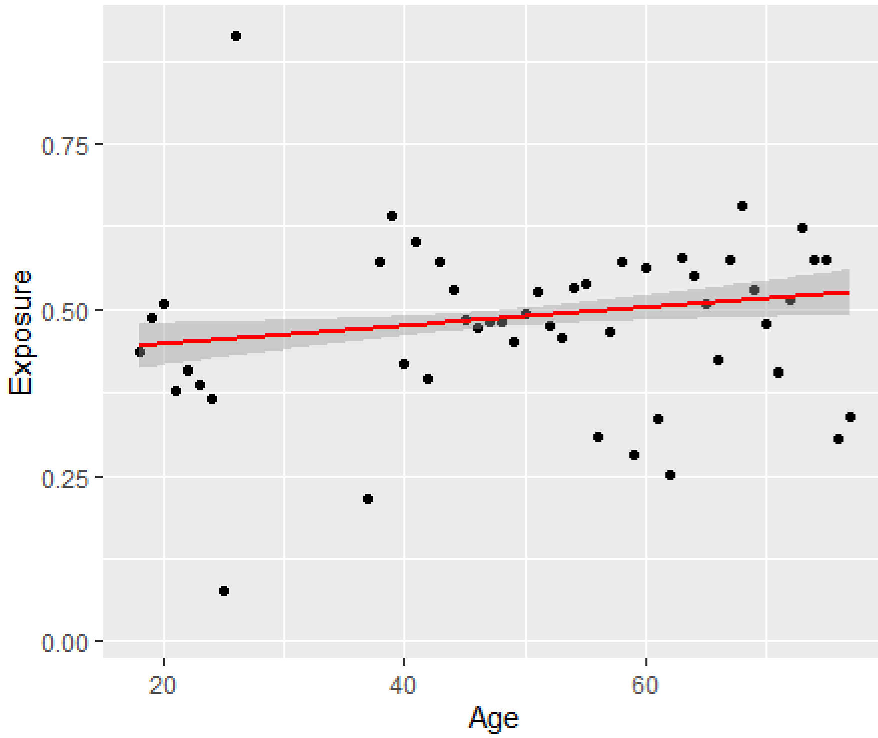
Click Code 24 to view in Appendix.



Exposure is the extent to which a customer would want to have their vehicle insured (FSRAO.1, 2023). The analysis suggests that men tend to have a slightly higher portion of their vehicle insured than women.

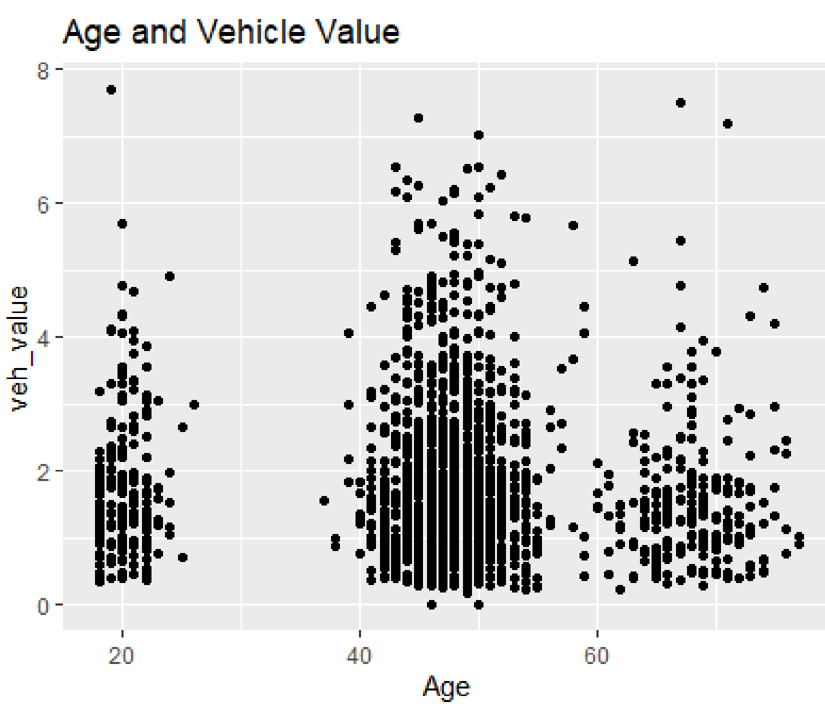
(This space has been intentionally left blank)

Click Code 25 to view in Appendix.

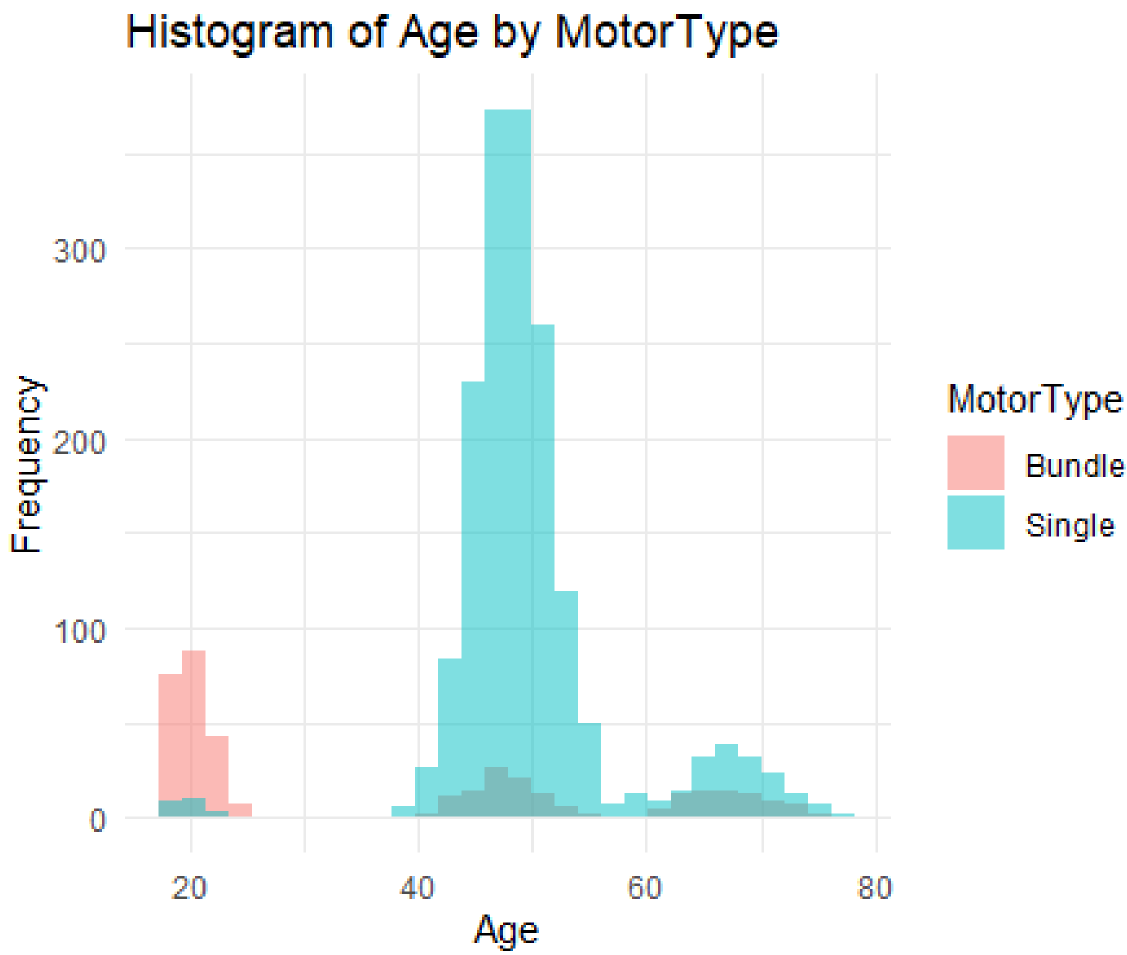


This also has a slight positive relationship with age as people who are older tend to have a higher portion of their vehicle ensured.

Click Code 26 to view in Appendix.

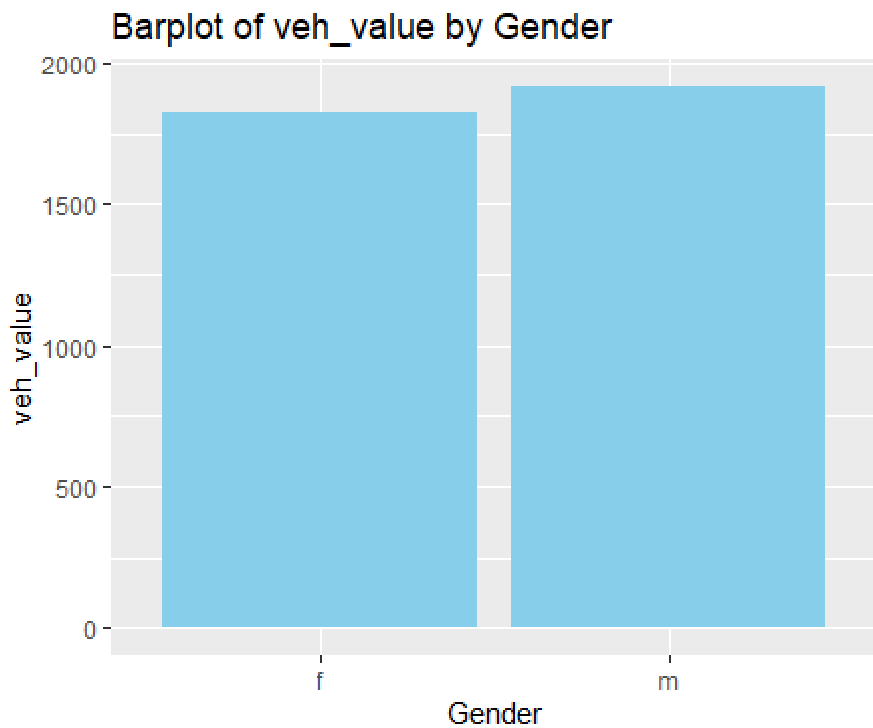


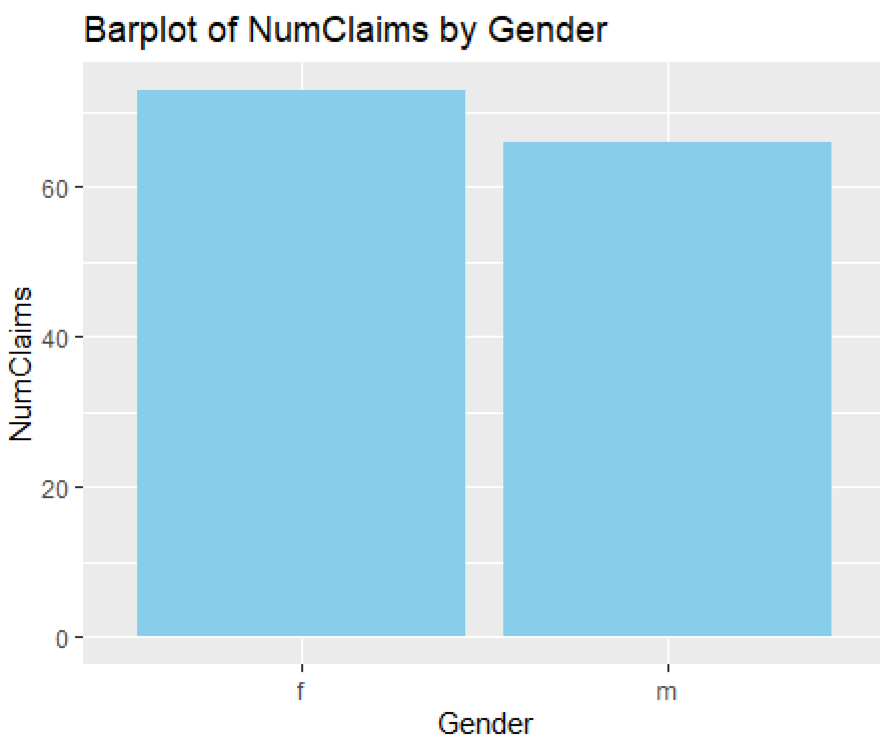
Click Code 27 to view in Appendix.



It is also suggested that for people between the ages of 40 and 60, there is a sharp peak indicating that a huge number of people opt for single motor types. The second highest peak is at the age of 20 for bundle motor type at 90 which goes to show how prevalent is the single motor type for a certain age group.

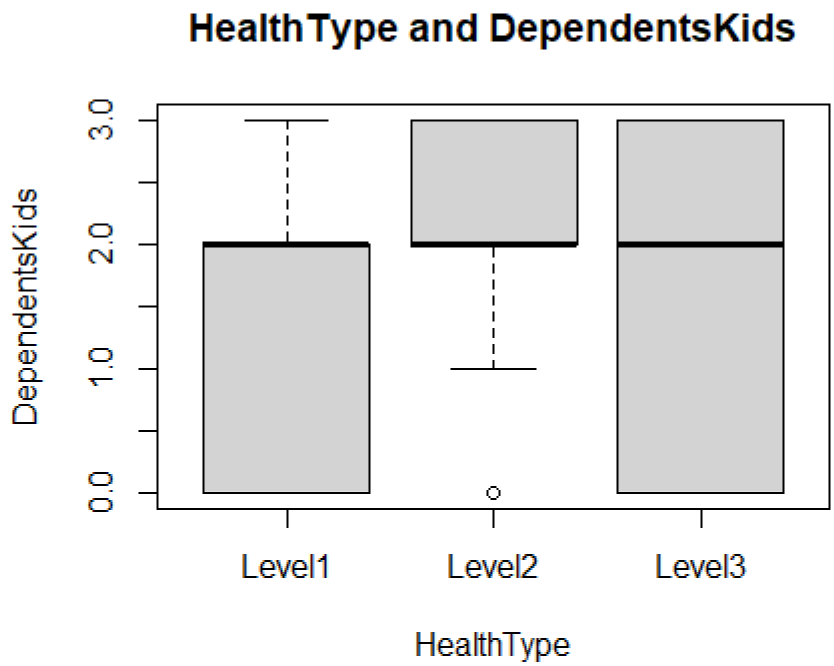
Click Code 28 to view in Appendix.

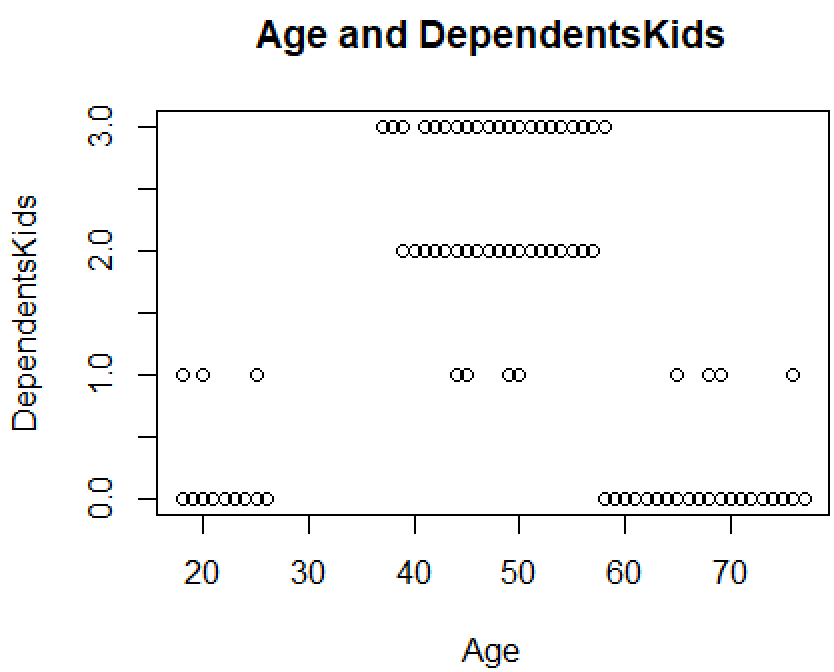


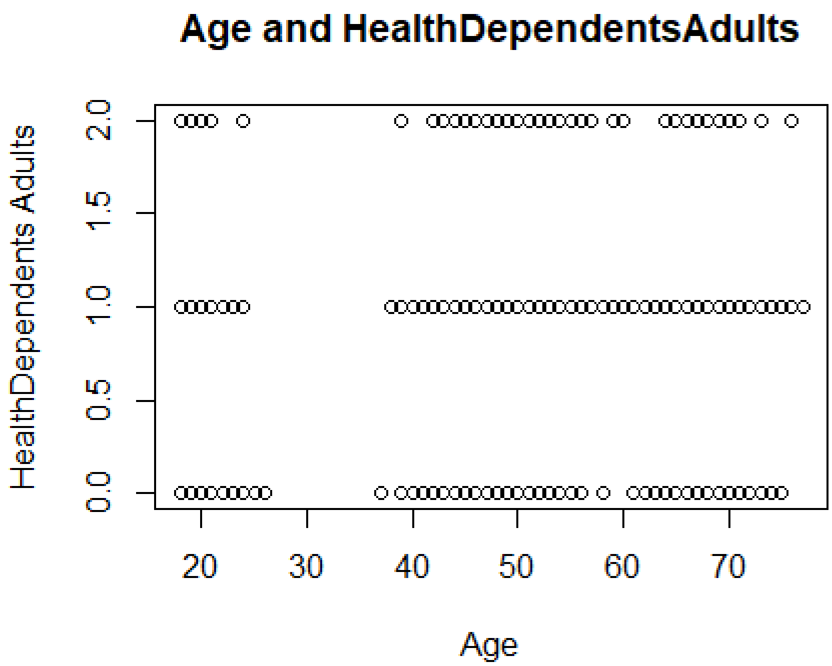


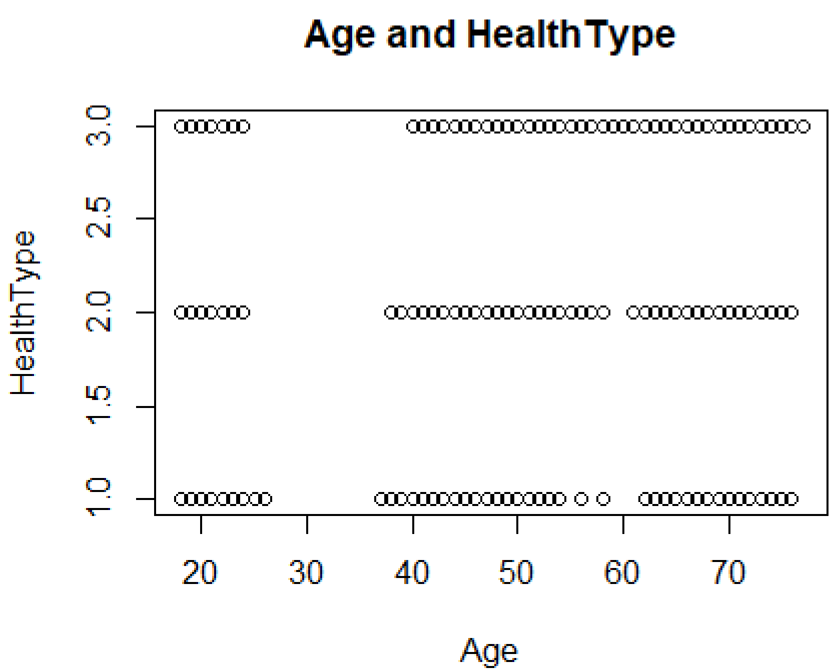
Compared to females, men have vehicles of a higher value and they have claimed less of their insurances. So it would make sense to have marketing communications tailored to men for more premium insurance services as there is less risk and more value associated to them.

Click Code 29 to view in Appendix.



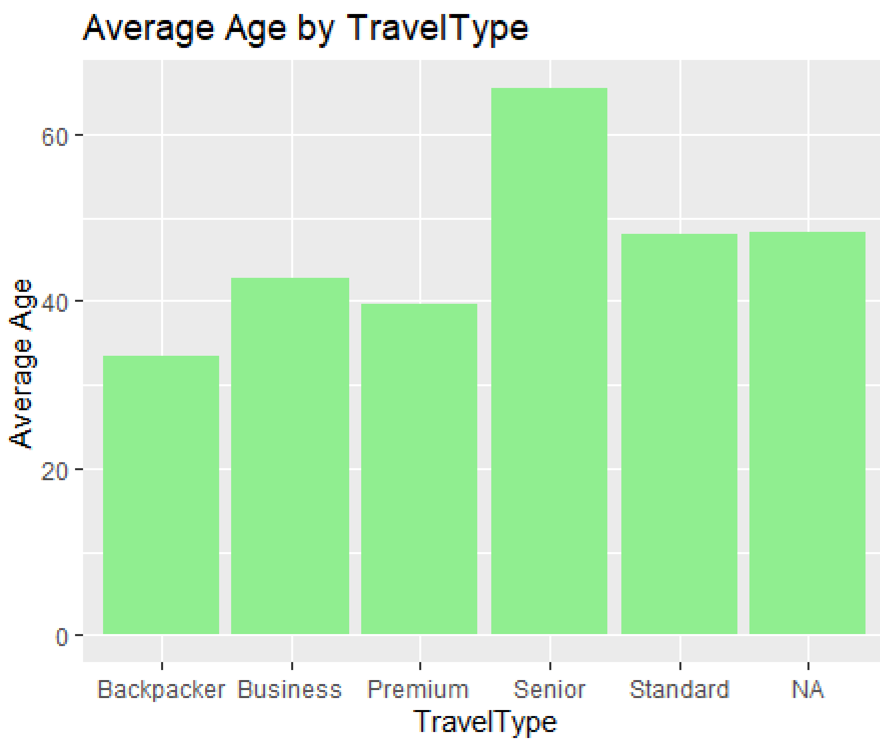






Customers between the ages of 30-60 have mostly 2 to 3 dependent kids. While customers from the age of 0-20 and 60+ have no dependent kids. However that is not the same case with number of adult dependents and age as any age can have any number of adult dependents. Hence marketing communications tailored for customers within the age of 30-60 should focus more on wellbeing of kids. At the same time, there is no preference of customers of different ages for the type of health insurance that they might opt for.

Click Code 30 to view in Appendix.



There is a preference of the type of travel insurance by customers of different ages. Most customers aged 30 will opt for the backpacker policy while business and premium types will be preferred by 40-year-old customers. Standard type will be preferred by customers in there 50s and senior type will be opted by customers who are aged more than 60.

(This space has been intentionally left blank)

Section 3

Conclusion

Knowing the characteristics of customers that opt for certain insurance products can be very helpful in terms of creating highly focused marketing communications for them. For the case of life insurance products, customer satisfaction and trust were the factors that lead to the highest purchase intention (Panigrahi, Azizan and Waris, 2018). Focusing marketing communications on these aspects can help build brand value in return.

Demographic factors like age and gender also have an affect on how likely it is for a vehicle insurance to be claimed which is associated with the risk of an unfortunate event (Bian *et al*., 2018). Hence to reduce the risk of dangerous driving, there has been an attempt to create a new model for motor insurance pricing policy called pay-how-you-drive (Guillen, Nielsen and Pérez-Marín, 2021) which penalizes the customer if they drive recklessly.

Before considering factors that affect consumer intentions to purchase travel insurances, we need to acknowledge the changing behavior of travel post COVID-19 due to increased fear of health hazards and risks at tourist destinations (Petrus, Yahya and Yahya, 2022). If a traveler has a history of traveling very often, they are more likely to purchase travel insurances which also depends on whether those travels were local or international (Dai, 2023).

References

FSRAO.1 (2023). *Key Definitions*. [online] Gisa.ca. Available at: https://www.gisa.ca/KeyDefinitions#:~:text=Exposures%20are%20a%20measure%20 [Accessed 22 Nov. 2023].

Panigrahi, S., Azizan, N.A. and Waris, M., 2018. Investigating the empirical relationship between service quality, trust, satisfaction, and intention of customers purchasing life insurance products. *Indian Journal of Marketing• January*.

Bian, Y., Yang, C., Zhao, J.L. and Liang, L., 2018. Good drivers pay less: A study of usage-based vehicle insurance models. *Transportation research part A: policy and practice*, *107*, pp.20-34.

Guillen, M., Nielsen, J.P. and Pérez-Marín, A.M., 2021. Near-miss telematics in motor insurance. *Journal of Risk and Insurance*, *88*(3), pp.569-589.

Petrus, S., Yahya, A.F. and Yahya, F., 2022, September. Domestic Travelers' Perceptions and Intention to Purchase Travel Insurance. In *Proceedings* (Vol. 82, No. 1, p. 74). MDPI..

Dai, Y., 2023, August. The Factor Drive People Buy Travel Insurance. In *Proceedings of the 2nd International Academic Conference on Blockchain, Information Technology and Smart Finance (ICBIS 2023)* (pp. 217-227). Atlantis Press.

Appendix

Code 1 – ABT

```
SELECT

c.CustomerID, c.Title, c.GivenName, c.MiddleInitial, c.Surname, c.CardType,

c.Occupation, c.Gender, c.Age, c.Location, c.ComChannel, c.MotorID, c.HealthID,

c.TravelID, m.PolicyStart AS Motor_Policy_Start, m.PolicyEnd AS Motor_Policy_End,

m.MotorType, m.veh_value, m.Exposure AS Vehicle_Exposure, m.clm AS Vehicle_Claim,

m.Numclaims AS Vehicle_Number_of_Claims, m.v_body, m.v_age, m.LastClaimDate AS
Vehicle_Last_Claim_Date, h.policyStart AS Health_Policy_Start, h.policyEnd AS Health_Policy_End,
h.Healthtype, h.HealthDependentsAdults, h.DependentsKids, t.policyStart AS Travel_Policy_Start,

t.PolicyEnd AS Travel_Policy_End, t.TravelType

INTO abt

FROM customer AS c

LEFT JOIN motor_policies AS m ON c.MotorID = m.MotorID

LEFT JOIN health_policies AS h ON c.HealthID = h.HealthID

LEFT JOIN travel_policies AS t ON c.TravelID = t.TravelID
```

Code 2

```
SELECT PolicyType, AVG(Age) AS AverageAge

FROM (

  SELECT 'Motor' AS PolicyType, Age FROM abt WHERE Motor_Policy_Start IS NOT NULL

  UNION ALL

  SELECT 'Health' AS PolicyType, Age FROM abt WHERE Health_Policy_Start IS NOT NULL

  UNION ALL

  SELECT 'Travel' AS PolicyType, Age FROM abt WHERE Travel_Policy_Start IS NOT NULL

) AS Policies

GROUP BY PolicyType;
```

Code 3

```
SELECT PolicyType, Gender, COUNT(*) AS GenderCount
```

```
FROM (
  SELECT 'Motor' AS PolicyType, Gender FROM abt WHERE Motor_Policy_Start IS NOT NULL
  UNION ALL
  SELECT 'Health' AS PolicyType, Gender FROM abt WHERE Health_Policy_Start IS NOT NULL
  UNION ALL
  SELECT 'Travel' AS PolicyType, Gender FROM abt WHERE Travel_Policy_Start IS NOT NULL
) AS Policies
GROUP BY PolicyType, Gender;
```

Code 4

```
SELECT Location, COUNT(*) AS PolicyCount
FROM abt
WHERE Motor_Policy_Start IS NOT NULL OR Health_Policy_Start IS NOT NULL OR Travel_Policy_Start IS NOT NULL
GROUP BY Location;
```

Code 5

```
SELECT ComChannel, COUNT(*) AS PolicyCount
FROM abt
WHERE Motor_Policy_Start IS NOT NULL OR Health_Policy_Start IS NOT NULL OR Travel_Policy_Start IS NOT NULL
GROUP BY ComChannel;
```

Code 6

```
SELECT HealthType, AVG(HealthDependentsAdults) AS AverageDependentsAdults
FROM abt
WHERE Health_Policy_Start IS NOT NULL
GROUP BY HealthType
ORDER BY AVG(HealthDependentsAdults) DESC;
```

Code 7

```
SELECT ComChannel, COUNT(*) AS MotorPolicyCount
```

```
FROM abt

WHERE Motor_Policy_Start IS NOT NULL

GROUP BY ComChannel;
```

Code 8

```
SELECT ComChannel, COUNT(*) AS HealthPolicyCount

FROM abt

WHERE Health_Policy_Start IS NOT NULL

GROUP BY ComChannel;
```

Code 9

```
SELECT ComChannel, COUNT(*) AS TravelPolicyCount

FROM abt

WHERE Travel_Policy_Start IS NOT NULL

GROUP BY ComChannel;
```

Code 10

```
SELECT Location, COUNT(*) AS MotorPolicyCount

FROM abt

WHERE Motor_Policy_Start IS NOT NULL

GROUP BY Location;
```

Code 11

```
SELECT Location, COUNT(*) AS HealthPolicyCount

FROM abt

WHERE Health_Policy_Start IS NOT NULL

GROUP BY Location;
```

Code 12

```
SELECT Location, COUNT(*) AS TravelPolicyCount
```

```
FROM abt

WHERE Travel_Policy_Start IS NOT NULL

GROUP BY Location;
```

Code 13

```
SELECT Occupation, COUNT(*) AS MotorPolicyCount

FROM abt

WHERE Motor_Policy_Start IS NOT NULL

GROUP BY Occupation

ORDER BY 2 DESC;
```

Code 14

```
SELECT Occupation, COUNT(*) AS HealthPolicyCount

FROM abt

WHERE Health_Policy_Start IS NOT NULL

GROUP BY Occupation

ORDER BY 2 DESC;
```

Code 15

```
SELECT Occupation, COUNT(*) AS TravelPolicyCount

FROM abt

WHERE Travel_Policy_Start IS NOT NULL

GROUP BY Occupation

ORDER BY 2 DESC;
```

Code 16

```
SELECT CardType, COUNT(*) AS Motor_CardTypeCount

FROM abt

WHERE Motor_Policy_Start IS NOT NULL

GROUP BY CardType;
```

Code 17

```
SELECT CardType, COUNT(*) AS Health_CardTypeCount

FROM abt

WHERE Health_Policy_Start IS NOT NULL

GROUP BY CardType;
```

Code 18

```
SELECT CardType, COUNT(*) AS Travel_CardTypeCount

FROM abt

WHERE Travel_Policy_Start IS NOT NULL

GROUP BY CardType
```

Code 19

```
getwd()

setwd("c:/Users/User/Documents")


library(readxl)

customer <- read_excel("C:/Users/User/Desktop/QUB Classes/Semester 1/Data Management/Assignment 1/Data 1_Customer.xlsx")

motor <- read_excel("C:/Users/User/Desktop/QUB Classes/Semester 1/Data Management/Assignment 1/Data 2_Motor Policies.xlsx")

health <- read_excel("C:/Users/User/Desktop/QUB Classes/Semester 1/Data Management/Assignment 1/Data 3_Health Policies.xlsx")

travel <- read_excel("C:/Users/User/Desktop/QUB Classes/Semester 1/Data Management/Assignment 1/Data 4_Travel Policies.xlsx")


library(tidyverse)

abt <- customer %>%

  left_join(motor, by = "MotorID") %>%

  left_join(health, by = "HealthID") %>%
```

```
  left_join(travel, by = "TravelID")
```

Code 20

```
summary(abt)
str(abt)
sum(is.na(abt))

unique(abt$Title)
unique(abt$Gender)
summary(abt$Age)
plot(abt$Age)
boxplot(abt$Age)
unique(abt$ComChannel)
summary(abt$DependentsKids)
plot(abt$DependentsKids)
null_CustomerID <- abt[is.na(abt$CustomerID), ]
hist(abt$veh_value)
abt$Title[abt$Title == "Mr"] <- "Mr."
abt$Gender[abt$Gender == "male"] <- "m"
abt$Gender[abt$Gender == "female"] <- "f"
abt <- abt[abt$Age >= 0 & abt$Age < 84, ]
abt$ComChannel[abt$ComChannel == "E"] <- "Email"
abt$ComChannel[abt$ComChannel == "P"] <- "Phone"
abt$ComChannel[abt$ComChannel == "S"] <- "SMS"
abt <- abt[abt$DependentsKids <= 10, ]
abt <- abt[abt$veh_value < 8, ]

abt$Title <- as.factor(abt$Title)
abt$Gender <- as.factor(abt$Gender)
```

```
abt$CardType <- as.factor(abt$CardType)

abt$Location <- as.factor(abt$Location)

abt$ComChannel <- as.factor(abt$ComChannel)

abt$MotorType <- as.factor(abt$MotorType)

abt$clm <- as.factor(abt$clm)

abt$v_body <- as.factor(abt$v_body)

abt$HealthType <- as.factor(abt$HealthType)

abt$TravelType <- as.factor(abt$TravelType)

abt$v_age <- as.factor(abt$v_age)

abt <- abt %>%

  rename(policyStart_health = policyStart.x, policyEnd_health = policyEnd, policyStart_motor =
PolicyStart,

        policyEnd_motor = PolicyEnd.x, policyStart_travel = policyStart.y, policyEnd_travel = PolicyEnd.y)

abt <- abt[complete.cases(abt$CustomerID), ]
```

Code 21

```
abt <- abt %>%

  mutate(motor_policy = factor(ifelse(!is.na(policyStart_motor), 1, 0), levels = c(0, 1)))

abt <- abt %>%

  mutate(health_policy = factor(ifelse(!is.na(policyStart_health), 1, 0), levels = c(0, 1)))

abt <- abt %>%

  mutate(travel_policy = factor(ifelse(!is.na(policyStart_travel), 1, 0), levels = c(0, 1)))
```

Code 22

```
cor(abt$Age[complete.cases(abt$Age, abt$veh_value)], abt$veh_value[complete.cases(abt$Age,
abt$veh_value)])


cor(

  abt$Age[complete.cases(abt$Age, abt$Numclaims)],

  abt$Numclaims[complete.cases(abt$Age, abt$Numclaims)]
```

```
)
```

Code 23

```
cor(
  abt$Age[complete.cases(abt$Age, as.numeric(abt$v_age))],
  as.numeric(abt$v_age[complete.cases(abt$Age, as.numeric(abt$v_age))])
)
```

Code 24

```
mean(abt$Exposure, na.rm=TRUE)
ggplot(abt, aes(x = Gender, y = Exposure)) +
  geom_bar(stat = "summary", fun.y = "mean", position = "dodge", fill = "blue") +
  labs(title = "Average Exposure by Gender", x = "Gender", y = "Average Exposure")
```

Code 25

```
ggplot(abt, aes(x = Age, y = Exposure)) +
  geom_point(stat = "summary", position = "dodge", fill = "blue") +
  geom_smooth(method = "lm", color = "red")
  labs(title = "Exposure by Age", x = "Age", y = "Exposure")
```

Code 26

```
as.numeric(abt$Age)
as.numeric(abt$veh_value)
ggplot(abt, aes(x = Age, y = veh_value)) +
  geom_point() +
  labs(title = "Age and Vehicle Value", x = "Age", y = "veh_value")
```

Code 27

```
ggplot(abt, aes(x = Age, fill = MotorType)) +
  geom_histogram(position = "identity", alpha = 0.5, bins = 30) +
  labs(title = "Histogram of Age by MotorType", x = "Age", y = "Frequency") +
```

```
  theme_minimal()
```

Code 28

```
ggplot(abt, aes(x = Gender, y = veh_value)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(title = "Barplot of veh_value by Gender", x = "Gender", y = "veh_value")

ggplot(abt, aes(x = Gender, y = Numclaims)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(title = "Barplot of NumClaims by Gender", x = "Gender", y = "NumClaims")
```

Code 29

```
plot(abt$HealthType, abt$DependentsKids, main = "HealthType and DependentsKids", xlab =
"HealthType", ylab = "DependentsKids")

plot(abt$Age, abt$DependentsKids, main = "Age and DependentsKids", xlab = 'Age', ylab =
'DependentsKids')

plot(abt$Age, abt$HealthDependentsAdults, main = "Age and HealthDependentsAdults", xlab = 'Age',
ylab= 'HealthDependents Adults')

plot(abt$Age, abt$HealthType, main = "Age and HealthType", xlab = "Age", ylab = "HealthType")
```

Code 30

```
ggplot(abt, aes(x = TravelType, y = Age)) +
  geom_bar(position = "stack", stat = "summary",na.rm = TRUE, fill = "lightgreen") +
  labs(title = "Average Age by TravelType", x = "TravelType", y = "Average Age")
```