

Data Management Assignment 1

Name : Kanika Kushwaha

Student ID : 40366629

Word Count : 2103

Index

1.0 Introduction	2
2.0 Database Documentation	2
2.1 Database development in Microsoft Access	2
2.2 Data Quality Analysis	6
2.2.1 Validity	6
2.2.2 Accuracy	6
2.2.3 Completeness	7
2.2.4 Uniformity	7
3.0 Results And Findings	7
3.1 Gender Analysis	7
3.2 Age Analysis	8
3.3 Customer Segment Analysis	8
3.4 Location Analysis	8
3.5 Preferred Communication Channel	9
4.0 Bibliography	10
5.0 Appendices	11
5.1 Appendix 1 – SQL Code	11
5.2 Appendix 2 – R Code	27

1.0 Introduction

An insurance company has been sparsely storing data files thus making the analysis difficult. Initially, the analyst working in the firm used SQL to perform the data analysis, but a major shift is happening toward R. Therefore, it wants to establish an analytics firm and as an initial project wants to figure out the similarity in characteristics of the customers that use various insurance policies offered by the company as well as the marketing channels that are being preferred the most by the customer. The challenge is to create an appropriate database that fulfills the established criteria and helps answer the differences between customer characteristics and the preferred communication channels.

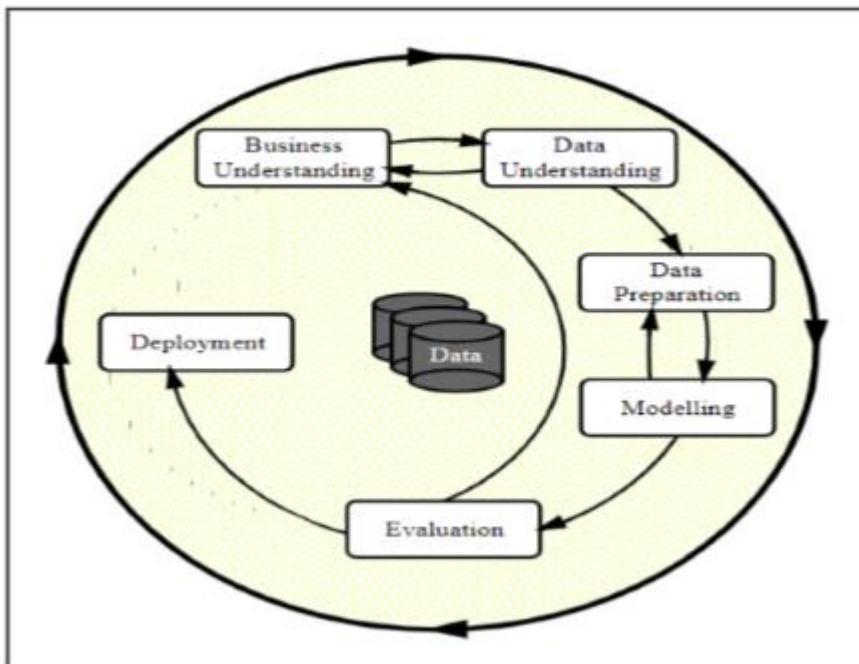
This report aims to solve the challenges faced by the insurance company by creating an appropriate database that helps answer the differences between customer characteristics and the preferred communication channels of various insurance policies offered by the company and presents meaningful information to increase the conversion rate using various findings from the analysis that is performed using SQL.

2.0 Database Documentation

2.1 Database development in Microsoft Access

The CRISP-DM project, which stands for Cross Industry Standard Process for Data Mining, presents a systematic process framework for working on data mining projects (Wirth & Hipp, 2000). Following the CRISP-DM technique, an emphasis on business understanding was undertaken. The business challenge attempts to discover differences in customer attributes across insurance policies and their preferred mode of communication. Creating an analytical base table would be very helpful in formulating information for this problem. The customers who have purchased various insurance policies are the subject of the analytical base table in this business challenge.

The objective is to develop an analytical base table with all the customer data and the insurance policies that the customers have purchased from the insurance firm. This is performed in the stage of data understanding. The data is then examined in the data preparation phase. The data consists of tables that contain information about the customers, motor insurance policies, travel insurance policies and health insurance policies provided by the insurance firm.



Sourced from: (Rüdiger Wirth, Jochen Hipp, *Phases of the Current CRISP-DM Process Model for Data Mining*)

Further, the associations between the variables were identified using the data dictionaries. Different ways of integrating the tables were discovered after evaluating the relationships. The best solution was to perform LEFT and FULL Outer joins on the tables since all the customer data is needed. This contributes to the CRISP-DM Model's Data Modeling stage.

Therefore, three new tables with information about consumers who have purchased the respective insurance policies were created. These tables were then integrated to make an analytical base table (ABT). One subject per row is required to form an analytical base table. Therefore, in the ABT, the subject is the client, and its attributes are numerous consumer traits together with the preferred communication channel, in addition to the insurance policies the customers have chosen. In the following section, the methods involved in the creation of the database are discussed.

The database and analytical basis table were created using the following steps. The database consists of 10 tables. The primary keys and foreign keys of the customer, motor_policies, travel_policies and the health_policies tables were identified. LEFT JOIN on the MotorID column was performed for customer and motor policies table to obtain the Customer_MotorPolicies table that contains information regarding customers who bought motor insurance and also performed LEFT Join using TravelID column for customer and travel policies to obtain the Customer_Travel polices that contains information regarding customers who bought travel insurance. Similarly, the Customer_Healthpolicies table was created using LEFT JOIN on HealthID column for customer and health_policies table to form a table that contains information regarding customers who have bought health insurance. Next, a FULL Outer Join of Customer_MotorPolicies and Customer_TravelPolicies on CustomerID was performed to form the CustMotorTravel table which contains information on all the customers who bought the motor and travel insurance policies. The FULL Outer Join was also performed using CustomerID on the CustMotorTravel and

Customer_HealthPolicies to form the CustMotorTravelHealth table. The CustMotorTravelHealth table consists of all attributes of the customer and the insurance policies assigned to it. Lastly, the redundant columns were removed to form the analytical base table called ABT_Customer_InsuranceTable. The various data quality issues encountered while creating the tables are discussed in the following section.

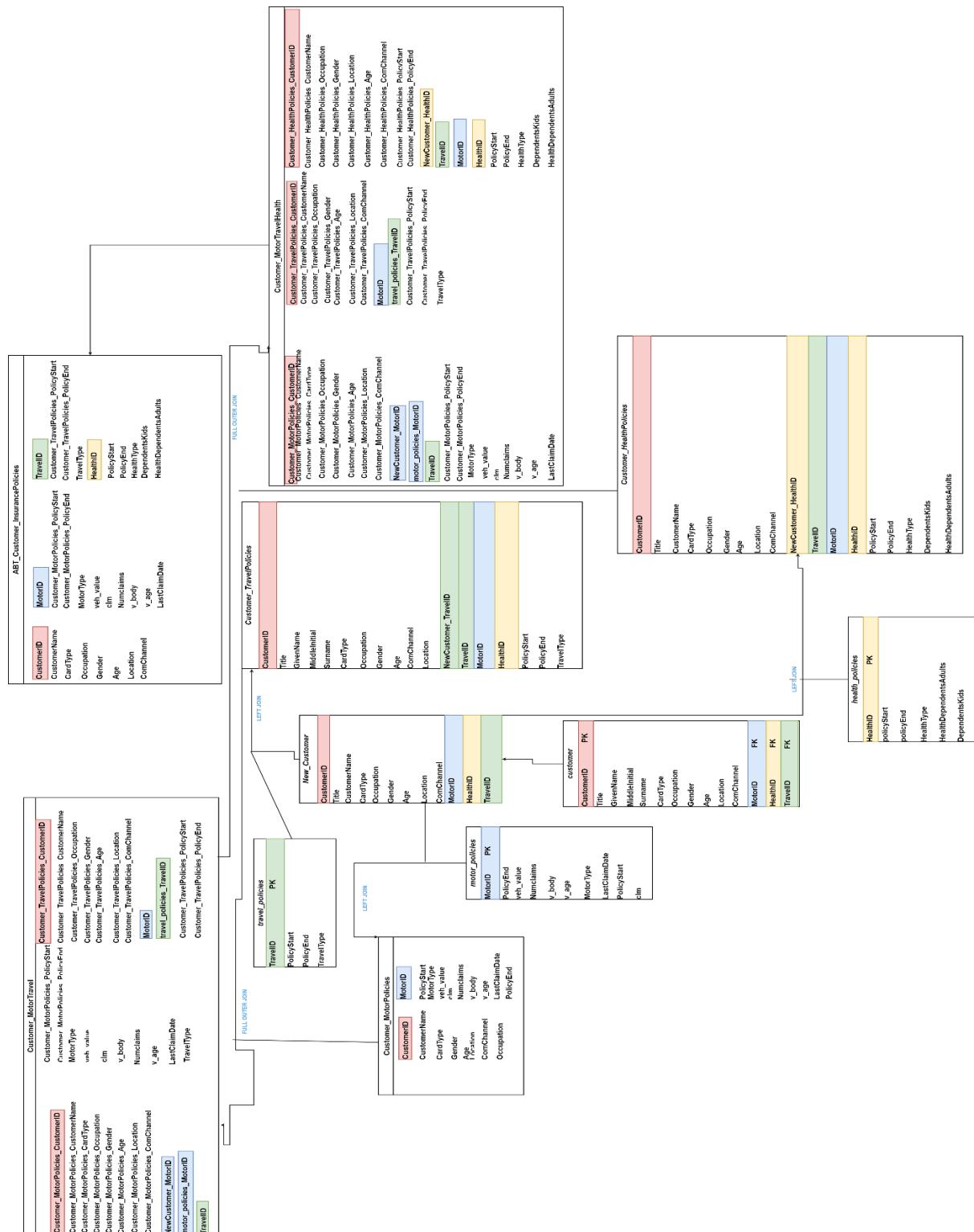


Figure 1: Structure of the Database

Name of Created Table	Tables Used for Creation	Linked using Keys	Operations Performed	Significance of table
Customer_MotorPolicies	Customer, Motor_policies	MotorID	Left Join on MotorID	Information regarding the customers who have taken the respective motor insurance policies offered by the insurance company for automobiles.
Customer_TravelPolicies	Customer, travel_policies	TravelID	Left Join on TravelID	Information regarding the customers who have taken the respective travel insurance policies offered by the insurance company for automobiles.
Customer_HealthPolicies	Customer, health_policies	HealthID	Left Join on HealthID	Information regarding the customers who took the respective motor insurance policies and the customers who have taken travel insurance offered by the insurance company for automobiles.
CustMotorTravel	Customer_MotorPolicies, Customer_TravelPolicies	CustomerID	Full Outer Join on CustomerID	Information regarding the customers who took the respective motor insurance policies and the customers who have taken travel insurance offered by the insurance company for automobiles
CustMotorTravelHealth	CustMotorTravel, Customer_HealthPolicies	CustomerID	Full Outer Join on CustomerID	Consists of all the information about the customer and the insurance policies it has taken offered by the insurance company.
ABT_Customer_InsurancePolicies	CustMotorTravelHealth			Modified version of the CustMotorTravelHealth Table and it consists of useful information pertaining to the

				customers and the insurance policies they have taken.
--	--	--	--	---

Table 1: Detailed Information on Database Structure

2.2 Data Quality Analysis

Adkins, M.C. (no date) emphasised various data quality issues that occur while performing data analysis. These are discussed below along with the summary of the data quality issues that occurred while analysing the ABT_Customer_InsuranceTable. According to Adkins, M.C. (no date), data quality can be classified into numerous categories, which are as follows :

2.2.1 Validity

Validity refers to whether or not a measurement is legitimate or invalid. It does not emphasize how closely it approaches a genuine value. A number or a special character, for example, cannot be a value for a column in a table corresponding to a person's name.

2.2.2 Accuracy

In order to convey accuracy, psychologists frequently use the words validity or proximity of a value to a known standard or real value. The age field in the ABT_Customer InsurancePolicies contains a negative number, -44, and some values like 180 and 240 that do not appear valid as a person cannot live this long. The presence of these outliers was confirmed using a boxplot of the age attribute that shows the presence of outliers between 150 – 200 and below -50. (Figure 2). These outliers were then identified and saved in the outliers variable using \$out. The na_if() function is then used to convert these outliers to null values. After converting these outliers to null, they are replaced by mean values of the age attribute using the replace_na() function. These were replaced with the mean values of the customers' ages, thus, removing the ambiguity associated with the customers' ages relative to this collection of null values.

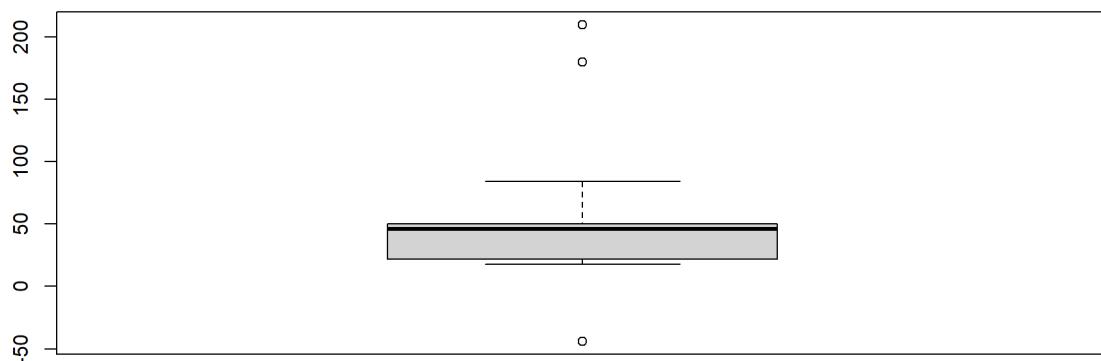


Figure 2 : Boxplot of Age attribute

2.2.3 Completeness

Completeness refers to the degree to which all required measures are known. This is the extent to which all of the values of a specific attribute are known. If the data does not contain null values, the dataset is considered to be complete. The ABT_Customer_Insurance table consists of 22 variables/columns with null values in it. These null values can be found by using the summarise_all function which is later discussed in appendix 2.

2.2.4 Uniformity

This refers to whether or not each piece of data in the table has the same measure or unit. In the gender column of the ABT, there are certain values like f and m that represents the females and males respectively. They are not in uniformity with other values like males and females in the column. Similarly, the ComChannel column, which addresses various customers' modes of communication, contains values such as S for SMS, P for Phone, and E for Email, which are inconsistent with the values of Email, Phone, and SMS. As a result, there lies inconsistency. To address this, the attribute values are converted to factors, the wrong values are replaced with the proper ones, and the null values are removed using the droplevel() function. Appendix 2 delves into the code in greater detail.

3.0 Results & Findings

While carrying out the initial descriptive analysis of the ABT, the following results were found.

3.1 Gender Analysis

The count of several customers using the motor, health and travel policies was analysed about the gender of the customers. It was found that the number of females who have taken any type of policy i.e motor, health or travel policies is slightly greater than the number of males who have taken the policies. This implies that females are much more aware of the policies offered by insurance companies and are successfully using them much more than males. This is evident in motor insurance as the number of females using motor policies is around 1700 and the total number of customers using motor insurance policies is around 3360. Therefore, 50.6% of females are using motor insurance policies and around 49.4% of the males are using motor insurance.

Similarly for health insurance, the number of females using health insurance is around 1296 and the total number of customers using health policies is around 2538. Therefore, 51.06% of females are using health insurance policies and around 48.9% of males are using health insurance.

Lastly, for travel insurance, the number of females using travel insurance is around 1069 and the total number of customers using health policies is around 2105. Therefore, 50.7% of

females are using health insurance policies and around 49.2% of males are using health insurance. Therefore, a marketing strategy specifically geared towards women can bring in more business due to the greater interest in insurance policies for females, especially in the health sector. These results can be seen in Appendix 1 along with the SQL code.

3.2 Age Analysis

An analysis of the age of customers and the three insurance policies was conducted. The following findings were found. The number of customers at the age of 48 is the highest set of customers for the motor policies. Similar is the case with the health policy with 239 customers aged 48. Thus, people that lie within the age groups of 40 -55 should be targeted more as they are more likely to convert and buy health and motor insurance policies.

On the other hand, the highest number of customers for the travel policy is 199 at the age of 18. This community of devoted customers I.e., those between the ages of 18 and 20 are younger and more likely to spend more on travel insurance premiums because they travel more frequently. As travel insurance policies are more popular among youngsters, youngsters can be targeted more for buying travel insurance. These results can be seen in Appendix 1 along with the SQL code.

3.3 Customer Segment Analysis

While analysing the clients who have purchased various insurance plans, it was discovered that some customers have purchased numerous insurance policies. Analysis reveals that there are substantially more customers—roughly 975—who have purchased all three insurance plans, which is very profitable for the business. Additionally, there are only about 207 customers who have purchased both travel and health insurance policies. Therefore, the insurance company can look into reasons why there are fewer customers opting for both travel and health insurance and further can formulate strategies to enhance their customer base in these areas by running suitable marketing strategies like targeting the major customer base for travel insurance i.e the younger generation and influencing them to opt for suitable health insurance. With 1149 customers, those who have purchased both auto and health insurance make up the majority. Finally, there are about 732 clients who have purchased both travel and auto insurance coverage. As a result, even though catering to all of the customers is crucial, greater attention should be given to those who have purchased all three insurance policies, followed by those who have health and auto insurance, as well as travel and auto insurance. Along with the SQL code, these results may be seen in Appendix 1.

3.4 Location Analysis:

To gain insights into the relationships between the location and the insurance policies, the ABT table was analysed. It was found that the number of customers for the urban location is more when compared with the number of customers for the rural location across the insurance policies. Therefore, buying insurance policies in urban areas is more popular than the rural areas. These results can be seen in Appendix 1 along with the SQL code.

3.5 Preferred communication channel

The preferred communication channel is email followed by telephone for auto and travel insurance policies. For health insurers, the preferred communication channel is the phone, followed by email. Therefore, overall, email and phone should be the marketing channels for motor and travel insurance. For health insurance, the marketing channel should be phone followed by email.

To further analyze customers' preferred communication channels, various components were further examined to uncover some useful insights. An analysis was carried out to find out the connection between the age of the customers and the preferred means of communication for different insurance companies. It was found that customers in their late 20s and late 30s are more likely to prefer email and SMS as a means of communication for all policies. Also, customers in their 50s are likely to prefer phones across policies. It should be noted that health insurance customers prefer the telephone as the primary means of communication. It should be noted that health insurance customers prefer the telephone as the primary means of communication. The telephone should therefore be more successful than others when it comes to promoting health insurance. Along with the SQL code, these results may be seen in Appendix 1.

In addition, an analysis of the gender of the customer and the preferred mode of communication for the insurance policies was performed. The results are as follows. Emails and SMS are preferred by females more than males. On the other hand, phones are more likely preferred by males than females. Taking into account the genders involved, customers can therefore be targeted to increase the conversion rate. These results can be seen in Appendix 1 along with the SQL code.

Additionally, a comparison between the clients' preferred communication method and their location was made. It was discovered that urban locations use emails as a form of communication more frequently than rural ones. However, compared to metropolitan areas, cellular phone use is more prevalent in rural areas. Last but not least, urban areas are where SMS are more common than rural places. This holds true for all of the company's insurance coverage. Along with the SQL code, these results may be seen in Appendix 1.

4.0 Bibliography

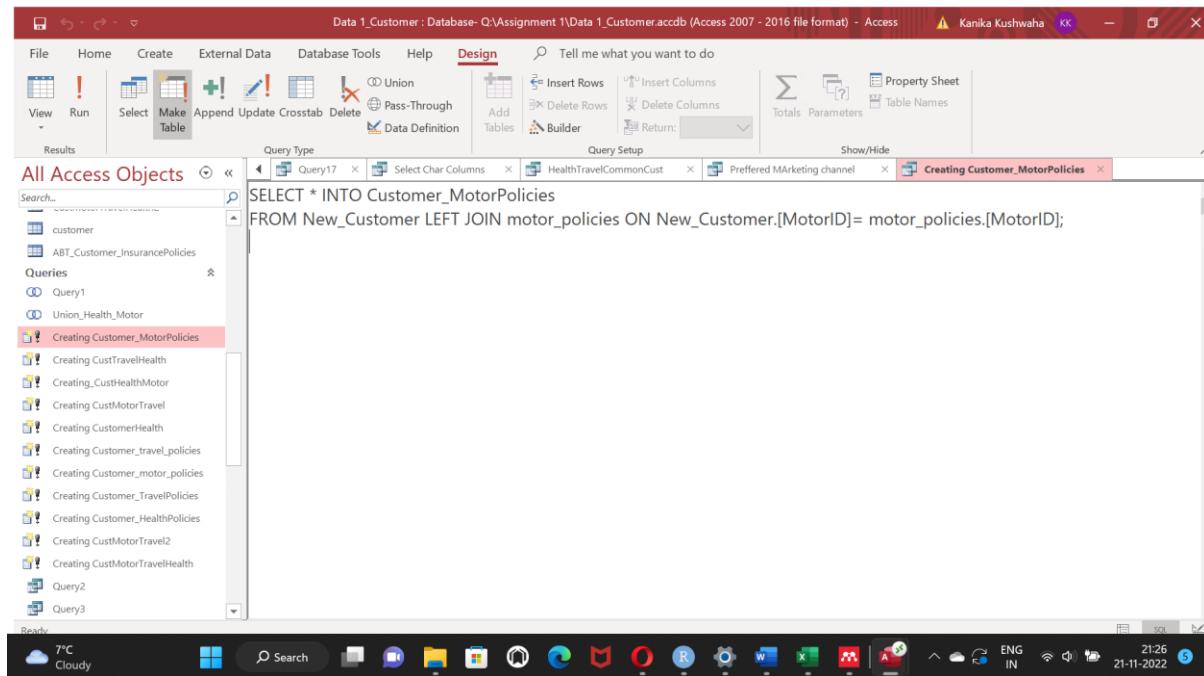
- Adkins, M.C. (no date) *Better Data Cleaning using R and the Tidyverse Section1 : Things to look for, Better data cleaning using R and the Tidyverse*. Available at: https://standard-deviator.com/workshop/cpa_2020_data_cleaning_tidyverse/slides/section_1#29 (Accessed: November 18, 2022).
- Phases of the Current CRISP-DM Process Model for Data Mining* (no date). Available at: <http://www.cs.unibo.it/~danilo.montesi/CBD/Beatriz/10.1.1.198.5133.pdf>.
- Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 1, 29–39.

5.0 Appendices

5.0.1 Appendix 1: SQL Code

Creating customer_motorpolicies table

```
SELECT * INTO Customer_MotorPolicies  
FROM New_Customer LEFT JOIN motor_policies ON New_Customer.[MotorID]=  
motor_policies.[MotorID];
```



The screenshot shows a Microsoft Access interface with a query results grid. The grid displays data from a table with columns: CustomerID, CustomerName, Cardtype, Occupation, Gender, Age, Location, ComChanne, New_Custon, HealthID, TravelID, and motor_polic. The data includes various customer details like names, card types, and locations. The interface includes a navigation pane on the left, a ribbon at the top, and a taskbar at the bottom.

CustomerID	CustomerName	Cardtype	Occupation	Gender	Age	Location	ComChanne	New_Custon	HealthID	TravelID	motor_polic
1 Mrs.MacyABoyle	Mastercard	Clinical laborat	female		23	Urban	SMS		6255		
2 Ms.TheaLMcIntosh	Mastercard		female		44	Urban	Phone		7154		63
4 Ms.MurronPMiller	Mastercard	Sheriff	female		19	Urban	SMS	6391	4975	2178	88
5 Mr.KaiAHenderson	Visa	Automotive pa	male		47	Rural	Phone	8876	4975		88
11 Mrs.KaylaABrown	Visa	Risk manager	female		54	Rural	Email	1850	3888		18
13 Mr.MuhammedKing	Mastercard	Building cleanir	male		49	Rural	E	8617	2953	3617	86
14 Mr.LucasCGilbert	Visa	Farm and homi	male		19	Rural	Email	8333		9005	83
16 Mr.ShayPSinclair	Mastercard	Recreation sup	male		68	Rural	Phone	8960	1907		89
17 Ms.EveCSutherland	Visa	Computer systf			20	Urban	Email	4961	3584		49
18 Mr.MuhamadNMcGregor	Mastercard	Fire inspector	male		49	Urban	Email	9453	3020		94
19 Mr.JamieCHunter	0	Farm and homi	male		49	Rural	Phone	9369	2995	8662	93
20 Mr.HarveyETaylor	0	Electronics rep	male		68	Rural	Phone	3787	9211		37
21 Ms.IsabellaRiley	Mastercard	f			47	Urban	Email	7764	1284		77
22 Mr.CallumMMcDonald	Mastercard		male		46	Urban	SMS	1173			11
23 Mr.HazelCSutherland	Mastercard	Letterpress set	female		42	Urban	Email	4055	9363	9601	40
24 Ms.ImogenCooper	Visa		female		61	Rural	Phone	7317	4558		73
26 Ms.MaisieCGill	Mastercard		female		51	Urban	Email	8311			83
27 Mr.MasonNicholls	Mastercard	Manager	male		19	Rural	SMS	4354			43
28 Mr.MichaelADickson	Visa	Environmental	male		18	Urban	Email	1774		9378	17
29 Mr.NoahGMCCarthy	Mastercard	Personnel anal	male		50	Rural	Phone	2948	2922		29
30 Mr.NairnDonaldson	Visa		male		44	Rural	SMS	1798	6430		17
31 Ms.AmiDWilliamson	Mastercard	Occupational h	female		49	Urban	Email	9959		4393	99

On a similar note, the customer_travelpolicies and customer_healthpolicies tables were created using TravelID and HealthID as the primary key respectively.

*SELECT * INTO Customer_TravelPolicies*

FROM New_Customer LEFT JOIN travel_policies ON New_Customer.[TravelID]=travel_policies.[TravelID];

The screenshot shows a Microsoft Access interface with a query editor. The query is:

```
SELECT * INTO Customer_TravelPolicies
FROM New_Customer LEFT JOIN travel_policies ON New_Customer.[TravelID]=travel_policies.[TravelID];
```

The interface includes a ribbon at the top, a navigation pane on the left, and a taskbar at the bottom.

The screenshot shows a Microsoft Access window with the title "Data 1_Customer : Database- Q:\Assignment 1\Assignment 1\Customer.accdb (Access 2007 - 2016 file format) - Access". The ribbon is visible at the top with tabs like File, Home, Create, External Data, Database Tools, and Help. The Home tab is selected. Below the ribbon is a toolbar with various icons for operations like Cut, Copy, Paste, Filter, Sort & Filter, Refresh, Save, Spelling, Find, and Text Formatting. A navigation pane on the left lists several queries: "Query17", "Select Char Columns", "HealthTravelCommonCust", "Preferred Marketing channel", "Creating Customer_MotorPolicies", and "Creating Customer_TravelPolicies". The main area displays a grid of data from a query named "Query17". The columns include CustomerID, CustomerName, Cardtype, Occupation, Gender, Age, Location, ComChannel, MotorID, HealthID, New_Custon, travel_polic, and policyStat. The data consists of approximately 40 rows of customer information. At the bottom of the screen, the Windows taskbar shows the date as 21-11-2022 and the time as 21:28.

*SELECT * INTO Customer_HealthPolicies*

*FROM New_Customer LEFT JOIN health_policies ON New_Customer.[HealthID]=
health_policies.[HealthID];*

The screenshot shows the Microsoft Access design view for a query named "Creating Customer_HealthPolicies". The ribbon is at the top with the "Design" tab selected. The left pane shows the "All Access Objects" tree, with "Creating Customer_HealthPolicies" highlighted. The main pane displays the SQL code:

```

SELECT * INTO Customer_HealthPolicies
FROM New_Customer LEFT JOIN health_policies ON New_Customer.[HealthID]= health_policies.[HealthID];

```

The status bar at the bottom indicates the date as 21-11-2022 and the time as 21:28.

The screenshot shows a Microsoft Access application window with the title "Data 1_Customer : Database- Q:\Assignment 1\Assignment 1\Customer.accdb (Access 2007 - 2016 file format) - Access". The ribbon menu is visible at the top. A vertical "Navigation Pane" is on the left, showing several table names: "HealthTravelCommonCust", "Preferred Marketing channel", "Creating Customer_MotorPolicies", "Creating Customer_TravelPolicies", and "Creating Customer_HealthPolicies". The main area displays a grid of data from a joined query. The columns include CustomerID, CustomerName, Cardtype, Occupation, Gender, Age, Location, ComChannel, MotorID, New_Custon, TravellID, health_polic, and policyStat. The data consists of approximately 40 rows of customer information. The status bar at the bottom shows the date as 21-11-2022 and the time as 21:29.

The customer_motorpolcies and the customer_travelpolicies are joined together using Full Outer Join to form CustMotorTravel Travel.

```

SELECT * INTO CustMotorTravel2
FROM (SELECT *
      FROM Customer_MotorPolicies
      LEFT JOIN Customer_TravelPolicies
        ON Customer_MotorPolicies.CustomerID = Customer_TravelPolicies.CustomerID
      UNION
      SELECT *
      FROM Customer_MotorPolicies
      RIGHT JOIN Customer_TravelPolicies
        ON Customer_MotorPolicies.CustomerID = Customer_TravelPolicies.CustomerID);
    
```

Screenshot of Microsoft Access 2016 showing the SQL View ribbon tab selected. The query pane displays the following SQL code:

```

SELECT * INTO CustMotorTravel2
FROM (SELECT *
      FROM Customer_MotorPolicies
      LEFT JOIN Customer_TravelPolicies
      ON Customer_MotorPolicies.CustomerID = Customer_TravelPolicies.CustomerID
      UNION
      SELECT *
      FROM Customer_MotorPolicies
      RIGHT JOIN Customer_TravelPolicies
      ON Customer_MotorPolicies.CustomerID = Customer_TravelPolicies.CustomerID) AS [%$##@_Alias];

```

Screenshot of Microsoft Access 2016 showing the Datasheet View ribbon tab selected. The navigation pane on the left shows the table 'Creating CustMotorTravel2' is selected. The main area displays the data from the joined tables:

Customer_ID	Customer_Name	Customer_Type	Customer_Gender	Customer_Age	Customer_City	Customer_Phone	Customer_Email	Customer_Status	Traveler_ID	motor_policy	Customer_HP
1	Mrs.MacyABoy	Mastercard	Clinical labor	female	23	Urban	SMS		6255		
2	Ms.TheaMcIn	Mastercard		female	44	Urban	Phone				6391 01/05/
4	Ms.MurronPM	Mastercard	Sheriff	female	19	Urban	SMS	6391			8876 07/10/
5	Mr.KaiAHender	Visa	Automotive pa	male	47	Rural	Phone	8876	4975	2178	8876 06/05/
11	Mrs.KaylaABro	Visa	Risk manager	female	54	Rural	Email	1850	3888		1850 18/08/
13	Mr.Muhammed	Mastercard	Building cleanir	male	49	Rural	E	8617	2953	3617	8617 29/05/
14	Mr.LucasCGilbe	Visa	Farm and hom	male	19	Rural	Email	8333		9005	8333 21/08/
16	Mr.ShayPSincla	Mastercard	Recreation sup	male	68	Rural	Phone	8960	1907		4961 25/04/
17	Ms.EveCSuther	Visa	Computer syst f		20	Urban	Email	4961	3584		9453 24/12/
18	Mr.Muhammad	Mastercard	Fire inspector	male	49	Urban	Email	9453	3020		9369 10/12/
19	Mr.JamieCHun	O	Farm and hom	male	49	Rural	Phone	9369	2995	8662	3787 14/12/
20	Mr.HarveyETay	O	Electronics rep	male	68	Rural	Phone	3787	9211		7764 04/09/
21	Ms.IsabellaKRii	Mastercard		f	47	Urban	Email	7764	1284		1173 04/09/
22	Mr.CallumMM	Mastercard		male	46	Urban	SMS			9601	4055 31/05/
23	Mr.HazelCSuth	Mastercard	Letterpress sett	female	42	Urban	Email	4055	9363		7317 13/12/
24	Ms.ImogenCo	Visa		female	61	Rural	Phone	7317	4558		8311 30/05/
26	Ms.MaisieCGill	Mastercard		female	51	Urban	Email			9378	1774 04/10/
27	Mr.MasonLnici	Mastercard	Manager	male	19	Rural	SMS	4354			2948 30/05/
28	Mr.MichaelADi	Visa	Environmental	male	18	Urban	Email	1774		1798	13/06/
29	Mr.NoahGMc	Mastercard	Personnel anal	male	50	Rural	Phone	2948	2922		4354 11/01/
30	Mr.NairnPDonz	Visa		male	44	Rural	SMS	1798	6430		1798 13/06/
31	Ms.AmiDWillia	Mastercard	Occupational h	female	49	Urban	Email	9959		4393	9959 23/11/

All the three tables will be joined when CustMotorTravel and Customer_HealthPolicies are joined together using Full Outer join.

```

SELECT * INTO CustMotorTravelHealth
FROM (SELECT *
      FROM CustMotorTravel2
      LEFT JOIN Customer_HealthPolicies
      ON CustMotorTravel2.CustomerID = Customer_HealthPolicies.CustomerID)

```

*ON CustMotorTravel.Customer_MotorPolicies_CustomerID =
Customer_HealthPolicies.CustomerID*

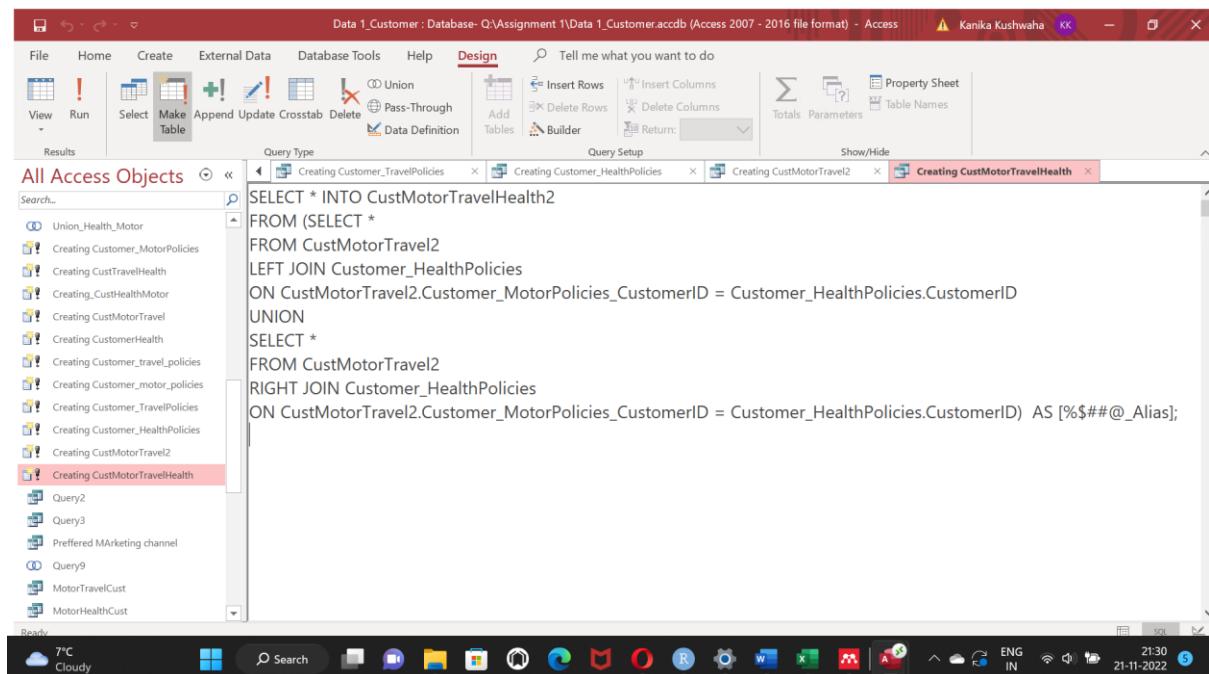
UNION

*SELECT **

FROM CustMotorTravel

RIGHT JOIN Customer_HealthPolicies

*ON CustMotorTravel.Customer_MotorPolicies_CustomerID =
Customer_HealthPolicies.CustomerID);*



The screenshot shows a Microsoft Access application window with the title "Data 1_Customer : Database- Q:\Assignment 1\Assignment 1\Customer.accdb (Access 2007 - 2016 file format) - Access". The ribbon menu is visible at the top. The main area displays five tables in tabs: "Creating Customer_MotorPolicies", "Creating Customer_TravelPolicies", "Creating Customer_HealthPolicies", "Creating CustMotorTravel2", and "Creating CustMotorTravelHealth". The "Creating CustMotorTravelHealth" tab is active, showing a grid of data. The data includes columns such as Customer_ID, Customer_Name, CardType, Occupation, Gender, Age, Location, ComChannel, motor_policies_MotorID (as MotorID), MotorType, Customer_MotorPolicies_policyStart, Customer_MotorPolicies_policyEnd, veh_value, Exposure, clm, Numclaims, v_body, v_age, LastClaimDate, travel_policies_TravelID (as TravelID), TravelType, Customer_TravelPolicies_policyStart, Customer_TravelPolicies_policyEnd, Customer_MotorPolicies_HealthID (as HealthID), HealthType, policyStart (as Customer_HealthPolicies_policyStart), policyEnd (as Customer_HealthPolicies_policyEnd), HealthDependentsAdults, and DependentsKids. The status bar at the bottom shows the date (21-11-2022), time (21:31), and system information (ENG IN). The navigation pane on the left lists the tables: Customer_MotorPolicies, Customer_TravelPolicies, Customer_HealthPolicies, CustMotorTravel2, and CustMotorTravelHealth.

Further, irrelevant columns that are of no use are removed from the CustMotorTravelHealth table to form ABT_Customer_InsurancePolicies.

```
SELECT * INTO ABT_Customer_InsurancePolicies
```

```
FROM (SELECT CustomerID, CustomerName, CardType, Occupation, Gender, Age,
Location, ComChannel, motor_policies_MotorID AS MotorID, MotorType,
Customer_MotorPolicies_policyStart, Customer_MotorPolicies_policyEnd, veh_value,
Exposure, clm, Numclaims, v_body, v_age, LastClaimDate, travel_policies_TravelID AS
TravelID, TravelType, Customer_TravelPolicies_policyStart,
Customer_TravelPolicies_policyEnd, Customer_MotorPolicies_HealthID AS HealthID,
HealthType, policyStart AS Customer_HealthPolicies_policyStart, policyEnd AS
Customer_HealthPolicies_policyEnd, HealthDependentsAdults, DependentsKids FROM
CustMotorTravelHealth2);
```

Navigation Pane

```

SELECT * INTO ABT_Customer_InsurancePolicies
FROM (SELECT CustomerID, CustomerName, CardType, Occupation, Gender, Age, Location, ComChannel, motor_policies_MotorID AS MotorID,
MotorType, Customer_MotorPolicies_policyStart, Customer_MotorPolicies_policyEnd, veh_value, Exposure, clm, Numclaims, v_body, v_age,
LastClaimDate, travel_policies_TravelID AS TravelID, TravelType, Customer_TravelPolicies_policyStart, Customer_TravelPolicies_policyEnd,
Customer_MotorPolicies_HealthID AS HealthID, HealthType, policyStart AS Customer_HealthPolicies_policyStart, policyEnd AS
Customer_HealthPolicies_policyEnd, HealthDependentsAdults, DependentsKids FROM CustMotorTravelHealth2) AS [%$##@_Alias];

```

Navigation Pane

CustomerID	CustomerNa	CardType	Occupation	Gender	Age	Location	ComChanne	MotorID	MotorType	Customer_N	Customer_N	veh_val
271	Mr.JacksonJol	Visa	Job developpe male	male	51	Rural	Phone					
273	Ms.SophiaAPa	Visa	Industrial hygie female	female	24	Rural	Email	9840	Bundle	29/12/2019	29/12/2020	
275	Ms.AliceJHoug	Mastercard	Employment se female	female	73	Urban	Email					
276	Mr.FinlayFMille	Visa	Account collect male	male	47	Urban	Email	2597	Single	10/05/2019	10/05/2020	
277	Mr.ZacharyAEI	Mastercard	Structural and i male	male	71	Rural	Email	3063	Single	01/06/2019	01/06/2020	
278	Mr.FinleyFTalb	0		male	47	Rural	Phone	8882	Single	07/04/2019	07/04/2020	
280	Mr.MorganLW	Mastercard		male	22	Urban	SMS	2926	Bundle	09/10/2019	09/10/2020	
281	Mrs.ShonaMke	Visa	Maxillofacial su female	female	50	Rural	Phone	3392	Single	15/02/2019	15/02/2020	
282	Mrs.VanessaM	Mastercard	Bench jeweler female	female	48	Rural	Phone	8463	Single	15/12/2019	15/12/2020	
283	Ms.AlyssaCBro	Mastercard		female	51	Rural	Email	9278	Single	20/10/2019	20/10/2020	
285	Mrs.EmmaESav	Mastercard		female	18	Urban	SMS	8472	Bundle	23/10/2019	23/10/2020	
286	Mr.AlastairEMc	Mastercard	Executive secre male	male	18	Urban	SMS	8283	Bundle	28/08/2019	28/08/2020	
288	Mr.RioEshaw	Visa	Health care int male	male	70	Rural	Phone	4142	Single	28/10/2019	28/10/2020	
289	Mrs.OrlaChon	Mastercard	Photoengraver female	female	48	Rural	Phone	9437	Bundle	02/06/2019	02/06/2020	
290	Mrs.FaithSteve	Mastercard	Claims represe female	female	44	Urban	Email	2301	Bundle	10/04/2019	10/04/2020	
291	Mrs.JenniferAS	Mastercard	Freight agent female	female	44	Urban	Email	3747	Single	14/10/2019	14/10/2020	
292	Mrs.RhiannonCVisa		Paratransit driv female	female	20	Rural	SMS					
293	Ms.MayKing	Visa	Purchasing cler female	female	20	Urban	Email	6955	Bundle	31/01/2019	31/01/2020	
264	Mrs.RhiannaES	Visa	Direct care wor female	female	47	Rural	Email	9196	Single	30/12/2019	30/12/2020	
266	Mr.RobertSha	0	Real estate app male	male	19	Rural	Email					
267	Mr.FinlayLNort	Visa	Human service male	male	46	Rural	Phone	2802	Bundle	16/12/2019	16/12/2020	
268	Mrs.BellaRAllar	Visa	Electrical and e female	female	22	Rural	SMS	7604	Bundle	21/08/2019	21/08/2020	

Results & Findings - Analysis queries

Gender Analysis

*SELECT Customer_Char.Gender, COUNT(MotorID) AS Customer_Count_Motor,
COUNT(HealthID) AS Customer_Count_Health, COUNT(TravelID) AS
Customer_Count_Travel*

FROM ABT_Customer_InsurancePolicies AS Customer_Char

GROUP BY Customer_Char.Gender;

The screenshot shows the Microsoft Access interface with the title bar "Data 1_Customer : Database- Q:\Assignment 1\Data 1_Customer.accdb (Access 2007 - 2016 file format)". The ribbon is visible with tabs like File, Home, Create, External Data, Database Tools, and Help. The Home tab is selected. A search bar says "Tell me what you want to do". Below the ribbon is a toolbar with various icons for operations like Cut, Copy, Paste, Filter, Sort & Filter, Refresh, Find, and More. The main area displays a table titled "GenderWise_PolicyCount" with four columns: Gender, Customer_Count_Motor, Customer_Count_Health, and Customer_Count_Travel. The data shows:

Gender	Customer_Count_Motor	Customer_Count_Health	Customer_Count_Travel
female	10	7	8
m	1687	1289	1061
male	8	9	7
male	1652	1233	1029

To the left of the table is a sidebar titled "All Access Objects" listing various queries and objects.

Location wise Analysis

```
SELECT Customer_Char.Location, COUNT(MotorID) AS Customer_Count_MotorPolicies,
COUNT(HealthID) AS Customer_Count_HealthPolicies, COUNT(TravelID) AS
Customer_Count_TravelPolicies
```

FROM ABT_Customer_InsurancePolicies AS Customer_Char

GROUP BY Customer_Char.Location;

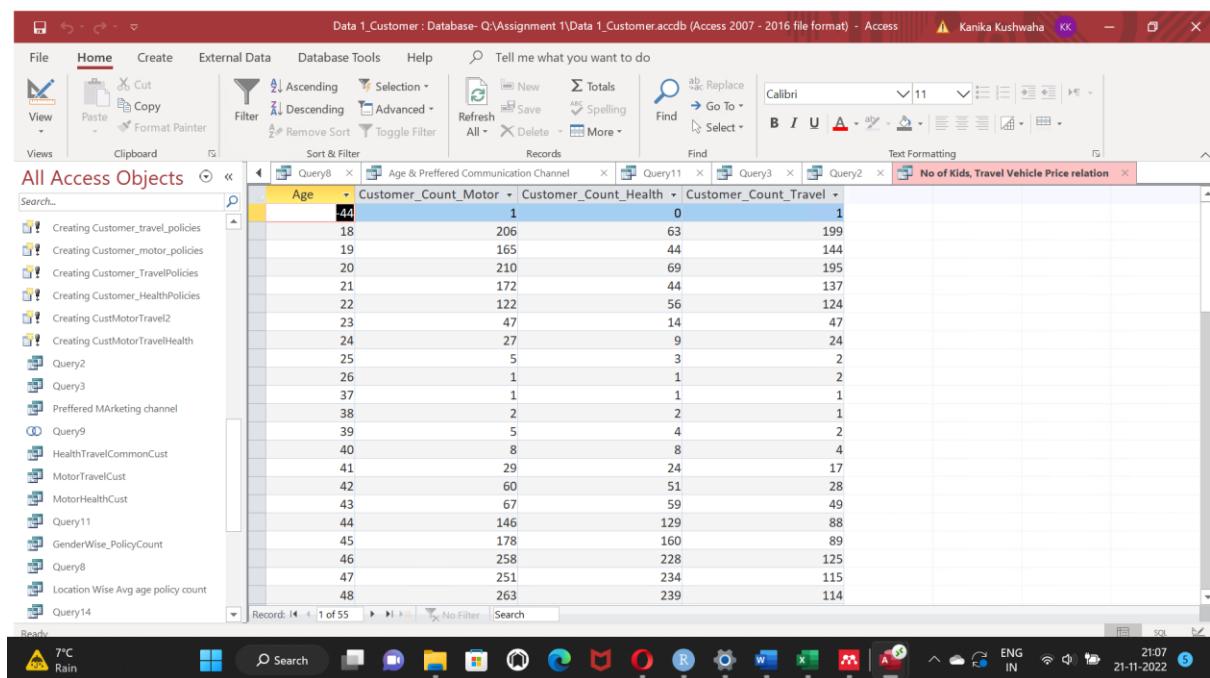
The screenshot shows the Microsoft Access interface with the title bar "Data 1_Customer : Database- Q:\Assignment 1\Data 1_Customer.accdb (Access 2007 - 2016 file format) - Access". The ribbon and toolbar are similar to the previous screenshot. The main area displays a table titled "Location Wise Avg age policy count" with four columns: Location, Customer_Count_MotorPolicies, Customer_Count_HealthPolicies, and Customer_Count_TravelPolicies. The data shows:

Location	Customer_Count_MotorPolicies	Customer_Count_HealthPolicies	Customer_Count_TravelPolicies
Rural	1449	1199	799
Urban	1908	1339	1306

To the left of the table is a sidebar titled "All Access Objects" listing various queries and objects, with "Location Wise Avg age policy count" highlighted.

Age Analysis

```
SELECT Customer_Char.Age, COUNT(Customer_Char.MotorID) AS  
Customer_Count_Motor, COUNT(Customer_Char.HealthID) AS Customer_Count_Health,  
COUNT(Customer_Char.TravelID) AS Customer_Count_Travel  
  
FROM ABT_Customer_InsurancePolicies AS Customer_Char  
  
GROUP BY Customer_Char.Age  
  
HAVING COUNT(Customer_Char.MotorID) OR COUNT(Customer_Char.HealthID) OR  
COUNT(Customer_Char.TravelID) IS NOT NULL  
  
ORDER BY Customer_Char.Age;
```



The screenshot shows the Microsoft Access application interface with a query results grid. The grid displays data from a query named 'Age & Preferred Communication Channel'. The columns are labeled 'Age', 'Customer_Count_Motor', 'Customer_Count_Health', and 'Customer_Count_Travel'. The data shows counts for various ages, with age 44 having the highest count of 239. The Access ribbon is visible at the top, and the taskbar at the bottom shows weather information (7°C Rain), system icons, and the date/time (21-11-2022).

Age	Customer_Count_Motor	Customer_Count_Health	Customer_Count_Travel
44	239	0	1
18	206	63	199
19	165	44	144
20	210	69	195
21	172	44	137
22	122	56	124
23	47	14	47
24	27	9	24
25	5	3	2
26	1	1	2
37	1	1	1
38	2	2	1
39	5	4	2
40	8	8	4
41	29	24	17
42	60	51	28
43	67	59	49
44	146	129	88
45	178	160	89
46	258	228	125
47	251	234	115
48	263	239	114

Customer Segment Analysis

Common Policies

```
SELECT COUNT(MotorID) AS Customer_Count_MotorPolicies, COUNT(HealthID) AS  
Customer_Count_HealthPolicies, COUNT(TravelID) AS Customer_Count_TravelPolicies  
  
FROM ABT_Customer_InsurancePolicies  
  
WHERE MotorID IS NOT NULL AND HealthID IS NOT NULL AND TravelID IS NOT  
NULL;
```

The screenshot shows the Microsoft Access interface with the title bar "Data 1_Customer : Database- Q:\Assignment 1\Assignment 1\Customer.accdb (Access 2007 - 2016 file format) - Access" and the user "Kanika Kushwaha KK". The ribbon is visible with tabs like File, Home, Create, External Data, Database Tools, and Help. The Home tab is selected. The main area displays a grid of data with three columns, each containing the value "975". The grid has a header row with column names: "Customer_Count_MotorPolicies", "Customer_Count_HealthPolicies", and "Customer_Count_TravelPolicies". The entire grid is highlighted with a pink border. On the left, the "All Access Objects" pane is open, showing a list of queries, with "Common Policies Cust" also highlighted with a pink border. The status bar at the bottom shows the date "21-11-2022" and time "21:10".

Common Health and Travel Customers

```
SELECT COUNT(MotorID) AS Customer_Count_MotorPolicies, COUNT(HealthID) AS
Customer_Count_HealthPolicies, COUNT(TravelID) AS Customer_Count_TravelPolicies
FROM ABT_Customer_InsurancePolicies
WHERE MotorID IS NULL AND HealthID IS NOT NULL AND TravelID IS NOT NULL;
```

The screenshot shows the Microsoft Access interface with the title bar "Data 1_Customer : Database- Q:\Assignment 1\Assignment 1\Customer.accdb (Access 2007 - 2016 file format) - Access" and the user "Kanika Kushwaha KK". The ribbon is visible with tabs like File, Home, Create, External Data, Database Tools, and Help. The Home tab is selected. The main area displays a grid of data with two columns, both containing the value "207". The grid has a header row with column names: "Customer_C", "Customer_C", and "Customer_C". The entire grid is highlighted with a pink border. On the left, the "All Access Objects" pane is open, showing a list of queries, with "Common Policies Cust" also highlighted with a pink border. The status bar at the bottom shows the date "21-11-2022" and time "21:12".

Common Motor and Travel Customers

```

SELECT COUNT(MotorID) AS Customer_Count_MotorPolicies, COUNT(HealthID) AS
Customer_Count_HealthPolicies, COUNT(TravelID) AS Customer_Count_TravelPolicies
FROM ABT_Customer_InsurancePolicies
WHERE MotorID IS NOT NULL AND HealthID IS NULL AND TravelID IS NOT NULL;

```

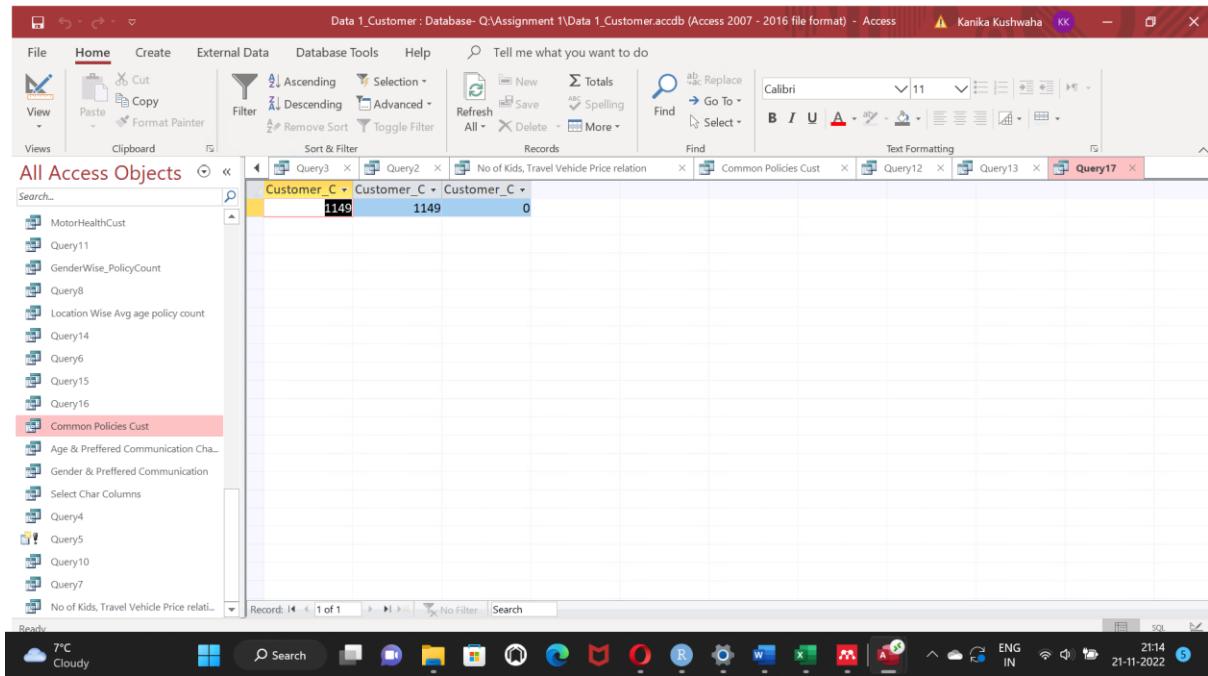
The screenshot shows the Microsoft Access 2016 interface. The ribbon is visible at the top with tabs like File, Home, Create, External Data, Database Tools, and Help. The Home tab is selected. The main area displays a grid of data from a query named 'Common Policies Cust'. The grid has three columns: 'Customer_C' (containing 732), 'Customer_C' (containing 0), and 'Customer_C' (containing 732). The 'Customer_C' column is highlighted with a yellow background. To the left of the grid is the 'All Access Objects' pane, which lists various queries and objects. The 'Common Policies Cust' query is highlighted with a red background. The status bar at the bottom shows the date and time as 21-11-2022 21:13.

Common Motor and Health Customers

```

SELECT COUNT(MotorID) AS Customer_Count_MotorPolicies, COUNT(HealthID) AS
Customer_Count_HealthPolicies, COUNT(TravelID) AS Customer_Count_TravelPolicies
FROM ABT_Customer_InsurancePolicies
WHERE MotorID IS NOT NULL AND HealthID IS NOT NULL AND TravelID IS NULL;

```



Preferred Marketing channel

Overall

```
SELECT ComChannel, COUNT(MotorID) AS Customer_Count_MotorPolicies,  
COUNT(HealthID) AS Customer_Count_HealthPolicies, COUNT(TravelID) AS  
Customer_Count_TravelPolicies  
  
FROM ABT_Customer_InsurancePolicies  
  
GROUP BY ComChannel  
  
ORDER BY COUNT(MotorID), COUNT(TravelID), COUNT(HealthID);
```

Data 1_Customer : Database- Q:\Assignment 1\Data 1_Customer.accdb (Access 2007 - 2016 file format) - Access Kanika

ComChannel	Customer_Count_MotorPolicies	Customer_Count_HealthPolicies	Customer_Count_TravelPolicies
S	1	2	1
E	4	3	4
P	5	4	2
SMS	567	263	458
Phone	1285	1215	637
Email	1495	1051	1003

Gender & Preferred Communication Channel

```

SELECT ComChannel, Gender, COUNT(MotorID) AS Customer_Count_MotorPolicies,
COUNT(HealthID) AS Customer_Count_HealthPolicies, COUNT(TravelID) AS
Customer_Count_TravelPolicies
FROM ABT_Customer_InsurancePolicies
GROUP BY ComChannel, Gender;

```

Data 1_Customer : Database- Q:\Assignment 1\Data 1_Customer.accdb (Access 2007 - 2016 file format) - Access Kanika Kushwaha KK

ComChannel	Gender	Customer_Count_MotorPolicies	Customer_Count_HealthPolicies	Customer_Count_TravelPolicies
E	female	1	2	1
E	male	3	1	3
Email	f	3	2	3
Email	female	753	535	503
Email	m	2	2	2
Email	male	737	512	495
P	female	3	3	1
P	male	2	1	1
Phone	f	3	4	2
Phone	female	628	602	315
Phone	m	4	5	3
Phone	male	650	604	317
S	male	1	2	1
SMS	f	4	1	3
SMS	female	302	147	241
SMS	m	2	2	2
SMS	male	259	113	212

Age & Preferred Communication

```

SELECT ComChannel, ROUND(Avg(Age)), COUNT(MotorID) AS
Customer_Count_MotorPolicies, COUNT(HealthID) AS Customer_Count_HealthPolicies,
COUNT(TravelID) AS Customer_Count_TravelPolicies

FROM ABT_Customer_InsurancePolicies

GROUP BY ComChannel;

```

The screenshot shows the Microsoft Access interface with a query results grid and a list of objects.

Query Results Grid:

ComChannel	Avg_Age_Customer	Customer_Count_MotorPolicies	Customer_Count_HealthPolicies	Customer_Count_TravelPolicies
E	45	4	3	4
Email	38	1495	1051	1003
P	45	5	4	2
Phone	51	1285	1215	637
S	36	1	2	1
SMS	28	567	263	458

Object List:

- Query4
- Query9
- Query7
- Query5
- Query10
- No of Kids, Travel Vehicle Price relat...
- HealthTravelCommonCust
- MotorTravelCust
- MotorHealthCust
- Query11
- GenderWise_PolicyCount
- Query8
- Location Wise Avg age policy count
- Query14
- Query6
- Query15
- Query16
- Common Policies Cust

Location & Preferred Communication

```

SELECT ComChannel, Location AS Location_Customer, COUNT(MotorID) AS
Customer_Count_MotorPolicies, COUNT(TravelID) AS Customer_Count_TravelPolicies,
COUNT(HealthID) AS Customer_Count_HealthPolicies

FROM ABT_Customer_InsurancePolicies

GROUP BY ComChannel, Location;

```

File Home Create External Data Database Tools Help Tell me what you want to do

View Cut Copy Format Painter Filter Ascending Selection Descending Advanced Remove Sort Toggle Filter Sort & Filter

New Save Totals Spelling Refresh All Delete More

Records Find ab Replace Go To Select

Calibri 11 Text Formatting

All Access Objects

Common Policies Cust Age & Preferred Communication Channel Gender & Preferred Communication Query12

ComChanne	Location_Customer	Customer_Count_MotorPolicies	Customer_Count_HealthPolicies	Customer_Count_TravelPolicies
E	Rural	3	3	3
E	Urban	1	0	1
Email	Rural	479	330	303
Email	Urban	1016	721	700
P	Rural	3	3	0
P	Urban	2	1	2
Phone	Rural	766	756	339
Phone	Urban	519	459	298
S	Rural	1	2	1
SMS	Rural	197	105	153
SMS	Urban	370	158	305

Record: 1 of 11 No Filter Search

Common Policies Cust

Ready 8°C Raining now Search

ENG IN 05:25 21-11-2022

5.1.2 Appendix 2: R Code

Please include any R code used. Code should be well commented.

Installing packages and loading libraries

```
## check working directory
getwd()

## set working directory
setwd("C:/Users/kanik/Desktop/Data Mgmt/")

##installing required packages
install.packages("dplyr")
install.packages("tidyverse")
install.packages("tidyrr")

## loading libraries
library(dplyr) #for data manipulation
library(readxl) #for reading in excel files
library(tidyverse)
library(tidyrr)
```

```

1 ## check working directory
2 getwd()
3
4 ## set working directory
5 setwd("C:/Users/kanik/Desktop/Data Mgmt/")
6
7 ##installing required packages
8 install.packages("dplyr")
9 install.packages("tidyverse")
10 install.packages("tidyr")
11
12
13 ## loading libraries
14 library(dplyr) #for data manipulation
15 library(readxl)#for reading in excel files
16 library(tidyverse)
17 library(tidyr)
18
19
20

```

Reading in all the tables and creating new tables using join operations

##reading in all tables

Customer <- read_xlsx("Data 1_Customer.xlsx")

Motor_Policies <- read_xlsx("Data 2_Motor Policies.xlsx")

Travel_Policies <- read_xlsx("Data 4_Travel Policies.xlsx")

Health_Policies <- read_xlsx("Data 3_Health Policies.xlsx")

joining tables using join function

Performing JOIN with Customer and Motor_policies data using MotorID

Customer_MotorPolicies <- left_join(Customer,Motor_Policies, by = "MotorID")

Performing JOIN with Customer and Travel_Policies data using TravelID

Customer_TravelPolicies <- left_join(Customer, Travel_Policies, by = "TravelID")

Performing LEFT JOIN with Customer and Health_Policies data using HealthID

Customer_HealthPolicies <- left_join(Customer, Health_Policies, by = "HealthID")

```

## Performing FULL OUTER JOIN with Customer_MotorPolicies and
Customer_TravelPolicies

##data using CustomerID

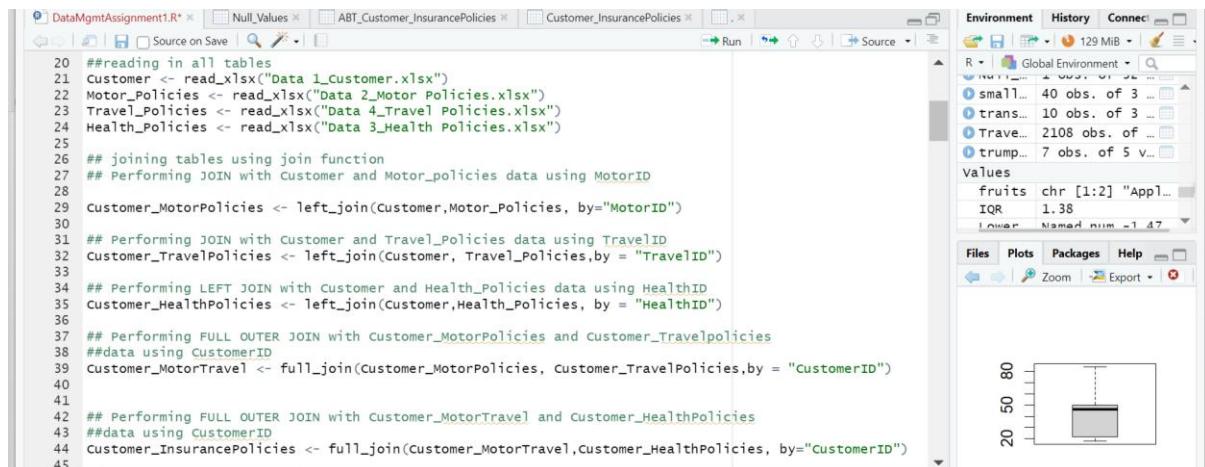
Customer_MotorTravel <- full_join(Customer_MotorPolicies, Customer_TravelPolicies, by
= "CustomerID")

## Performing FULL OUTER JOIN with Customer_MotorTravel and
Customer_HealthPolicies

##data using CustomerID

Customer_InsurancePolicies <- full_join(Customer_MotorTravel, Customer_HealthPolicies,
by="CustomerID")

```



Formation of Analytical Base Table

```

## Forming an ABT

## Saving only useful columns in the ABT

```

```

select(-Customer_InsurancePolicies>Title.x,-Customer_InsurancePolicies$GivenName.x,
      -Customer_InsurancePolicies$MiddleInitial.x,-Customer_InsurancePolicies$Surname.x,
      -Customer_InsurancePolicies$CardType.x,-Customer_InsurancePolicies$Occupation.x,
      -Customer_InsurancePolicies$Gender.x, -Customer_InsurancePolicies$Age.x, -
      Customer_InsurancePolicies$Location.x,

```

```
-Customer_InsurancePolicies$ComChannel.x, -Customer_InsurancePolicies$MotorID.x,  
-Customer_InsurancePolicies$HealthID.x,  
-Customer_InsurancePolicies$TravelID.x,-  
Customer_InsurancePolicies$title.y,Customer_InsurancePolicies$GivenName.y,  
-Customer_InsurancePolicies$MiddleInitial.y,-Customer_InsurancePolicies$Surname.y,  
-Customer_InsurancePolicies$CardType.y,  
-Customer_InsurancePolicies$Occupation.y, -Customer_InsurancePolicies$Gender.y, -  
Customer_InsurancePolicies$Age.y,  
-Customer_InsurancePolicies$Location.y, -Customer_InsurancePolicies$ComChannel.y,  
-Customer_InsurancePolicies$MotorID.y,  
-Customer_InsurancePolicies$HealthID.y, -Customer_InsurancePolicies$TravelID.y)
```

```
## subsetting columns
```

```
ABT_Customer_InsurancePolicies <- Customer_InsurancePolicies[,-c(2:13,14,25:37)]
```

```
## Rearranging the columns
```

```
ABT_Customer_InsurancePolicies <- select(Customer_InsurancePolicies, CustomerID,  
Title, GivenName,  
MiddleInitial, Surname, CardType, Occupation, Gender, Age, Location,  
ComChannel, MotorID,  
MotorType, PolicyStart, PolicyEnd.x, veh_value, Exposure, clm, Numclaims,  
v_body, v_age,  
LastClaimDate, TravelID, policyStart.x, PolicyEnd.y, TravelType, HealthID,  
policyStart.y,  
policyEnd, HealthType, HealthDependentsAdults, DependentsKids)
```

```
## Renaming column names
```

```
ABT_Customer_InsurancePolicies <- ABT_Customer_InsurancePolicies %>%  
rename("customer_motor_policyend" = PolicyEnd.x)%>%  
rename("customer_motor_policystart" = PolicyStart)%>%  
rename("customer_travel_policystart" = policyStart.x)%>%  
rename("customer_travel_policyend" = PolicyEnd.y)%>%
```

```
rename("customer_health_policystart" = policyStart.y)%>%
```

```
rename("customer_health_policyend" = policyEnd)
```

The screenshot shows the RStudio interface with the code editor open. The file is named 'DataMgmtAssignment1.R'. The code performs several operations on datasets: it joins 'Customer_MotorTravel' and 'Customer_HealthPolicies' on 'CustomerID', selects useful columns from the joined dataset, subsets the columns to create 'ABT_Customer_InsurancePolicies', and rearranges the columns. It also includes a section for checking null values.

```
44 Customer_InsurancePolicies <- full_join(Customer_MotorTravel,Customer_HealthPolicies, by="CustomerID")
45
46 ## Forming an ABT
47 ## Saving only useful columns in the ABT
48
49 select(-Customer_InsurancePolicies>Title.x,-Customer_InsurancePolicies$GivenName.x,
50       -Customer_InsurancePolicies$MiddleInitial.x,-Customer_InsurancePolicies$Surname.x,
51       -Customer_InsurancePolicies$CardType.x,-Customer_InsurancePolicies$Occupation.x,
52       -Customer_InsurancePolicies$Gender.x, -Customer_InsurancePolicies$Age.x, -Customer_InsurancePolicies$Location.x,
53       -Customer_InsurancePolicies$ComChannel.x, -Customer_InsurancePolicies$MotorID.x, -Customer_InsurancePolicies$HealthID.x,
54       -Customer_InsurancePolicies$TravelID.x,-Customer_InsurancePolicies>Title.y,Customer_InsurancePolicies$GivenName.y,
55       -Customer_InsurancePolicies$MiddleInitial.y,-Customer_InsurancePolicies$Surname.y, -Customer_InsurancePolicies$CardType.y,
56       -Customer_InsurancePolicies$Occupation.y, -Customer_InsurancePolicies$Gender.y, -Customer_InsurancePolicies$Age.y,
57       -Customer_InsurancePolicies$Location.y, -Customer_InsurancePolicies$ComChannel.y, -Customer_InsurancePolicies$MotorID.y,
58       -Customer_InsurancePolicies$HealthID.y, -Customer_InsurancePolicies$TravelID.y)
59
60 ## subsetting columns
61 ABT_Customer_InsurancePolicies <- Customer_InsurancePolicies[,-c(2:13)]
62 ABT_Customer_InsurancePolicies <- Customer_InsurancePolicies[,-c(2:13,14,25:37)]
63
64 ## Rearranging the columns
65
66 ABT_Customer_InsurancePolicies <- select(Customer_InsurancePolicies, CustomerID, Title, GivenName,
67                                         MiddleInitial, Surname, CardType, Occupation, Gender, Age, Location, ComChannel, MotorID,
68                                         MotorType, PolicyStart, PolicyEnd.x, veh_value, Exposure, clm, NumClaims, v_body, v_age,
69                                         LastClaimDate, TravelID, policystart.x, PolicyEnd.y, TravelType, HealthID, policyStart.y,
70                                         policyEnd, HealthType, HealthDependentsAdults, DependentsKids)
```

Checking for null values in the table

```
## Checking for null values in table
```

```
ABT_Customer_InsurancePolicies %>%
```

```
select(everything()) %>%
```

```
summarise_all(list(~sum(is.na(.))))%>%
```

```
View()
```

The screenshot shows the RStudio interface with the following code in the script editor:

```
82
83
84 View(ABT_Customer_InsurancePolicies)
85
86 ## Data quality analysis of the abt
87
88
89
90 ## summarising the data
91
92 summary(ABT_Customer_InsurancePolicies)
93
94 ## Checking for null values in table
95
96 ABT_Customer_InsurancePolicies %>%
97   select(everything()) %>%
98   summarise_all(list(~sum(is.na(.)))) %>%
99   View()
100
101 ## Found errors in Age : Max is 210
102 ## Vehicle Value; Max : 16
```

Checking for negative values in Age attribute of ABT_CustomerInsuranceTable

```
## Checking for negative values in Age attribute
ABT_Customer_InsurancePolicies %>%
  filter(Age <= 0)%>%
  View()
```

The screenshot shows the RStudio interface with the following code in the script editor:

```
129 #rename_with(.fn = snakecase::to_snake_case)
130
131 boxplot(ABT_Customer_InsurancePolicies$Age)
132
133 ## Checking for negative values in Age attribute
134 ABT_Customer_InsurancePolicies %>%
135   filter(Age <= 0)%>%
136   View()
137
138
139 <--
```

The console output shows:

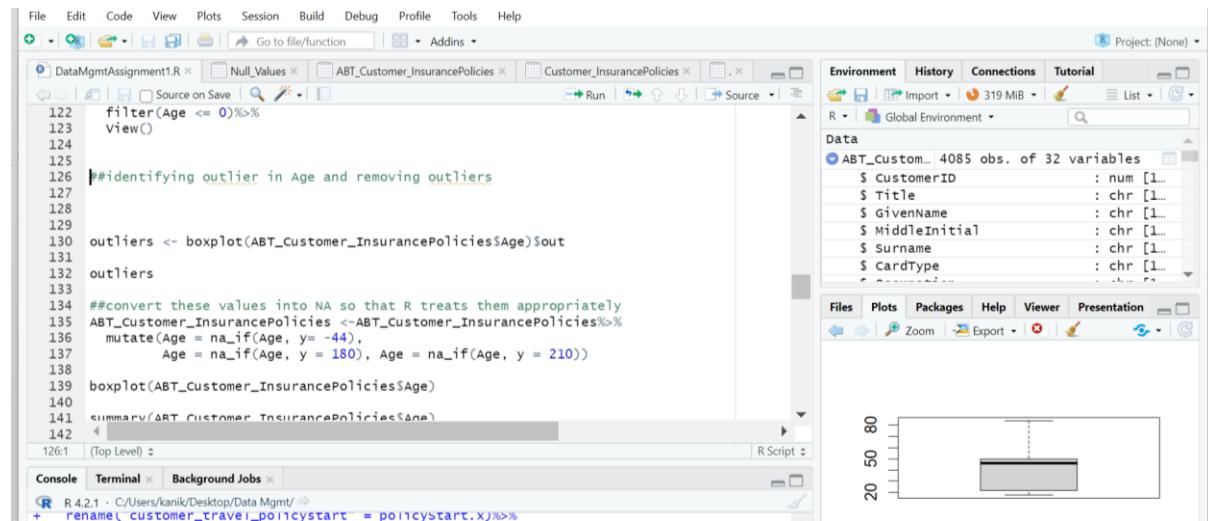
```
R 4.2.1 · C:/Users/kanik/Desktop/Data Mgmt/ ·
[997] "Visa"           "Mastercard" "0"
[1000] "0"
[ reached getOption("max.print") -- omitted 3085 entries ]
> boxplot(ABT_Customer_InsurancePolicies$Age)
> ABT_Customer_InsurancePolicies %>%
+   filter(Age <= 0)%>%
+   View()
```

Finding outliers in Age Attribute

```
##identifying outliers in Age and removing outliers
```

```
outliers <- boxplot(ABT_Customer_InsurancePolicies$Age)$out
```

outliers



Converting outliers to null values

```
##convert these values into NA so that R treats them appropriately
```

```
ABT_Customer_InsurancePolicies <- ABT_Customer_InsurancePolicies%>%
```

```
  mutate(Age = na_if(Age, y= -44),
```

```
    Age = na_if(Age, y= 180), Age = na_if(Age, y= 210))
```

```
boxplot(ABT_Customer_InsurancePolicies$Age)
```

```
summary(ABT_Customer_InsurancePolicies$Age)
```

The screenshot shows the RStudio interface. The script editor pane contains R code for outlier detection and a boxplot. The global environment pane shows a data frame named ABT_Cus... with 4085 observations and 32 variables. The console pane displays an error message related to the 'na_if' function.

```

148 outliers
149
150 ABT_Customer_InsurancePolicies <- ABT_Customer_InsurancePolicies %>%
151   mutate(Age = na_if(Age, y = -44),
152         Age = na_if(Age, y = 180), Age = na_if(Age, y = 210))
153
154 boxplot(ABT_Customer_InsurancePolicies$Age)
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
709
710
711
712
713
714
715
716
717
718
719
719
720
721
722
723
724
725
726
727
728
729
729
730
731
732
733
734
735
736
737
738
739
739
740
741
742
743
744
745
746
747
748
749
749
750
751
752
753
754
755
756
757
758
759
759
760
761
762
763
764
765
766
767
768
769
769
770
771
772
773
774
775
776
777
778
779
779
780
781
782
783
784
785
786
787
788
789
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
809
810
811
812
813
814
815
816
817
818
819
819
820
821
822
823
824
825
826
827
828
829
829
830
831
832
833
834
835
836
837
838
839
839
840
841
842
843
844
845
846
847
848
849
849
850
851
852
853
854
855
856
857
858
859
859
860
861
862
863
864
865
866
867
868
869
869
870
871
872
873
874
875
876
877
878
879
879
880
881
882
883
884
885
886
887
888
889
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
909
910
911
912
913
914
915
916
917
918
919
919
920
921
922
923
924
925
926
927
928
929
929
930
931
932
933
934
935
936
937
938
939
939
940
941
942
943
944
945
946
947
948
949
949
950
951
952
953
954
955
956
957
958
959
959
960
961
962
963
964
965
966
967
968
969
969
970
971
972
973
974
975
976
977
978
979
979
980
981
982
983
984
985
986
987
988
989
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1189
1190
1191
1192
1193
1194
1195
1196
1197
1197
1198
1199
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1789
1790
1791
1792
1793
1794
1795
1796
1797
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1889
1890
1891
1892
1893
1894
1895
1896
1897
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2349
2350
2351
2352
2353
2354
2355
23
```

The screenshot shows the RStudio interface. In the top-left pane, there are tabs for 'DataMgmtAssignment1.R*', 'ABT_Customer_InsurancePolicies', and 'Customer_InsurancePolicies'. The code editor (top-left) contains the following R script:

```

159 ## Replacing NAs with Mean
160
161 mean(ABT_Customer_InsurancePolicies$Age)
162
163 ABT_Customer_InsurancePolicies$Age[is.na(ABT_Customer_InsurancePolicies$Age)] <-
164   mean(ABT_Customer_InsurancePolicies$Age,na.rm = TRUE)
165
166 summary(ABT_Customer_InsurancePolicies$Age)
167
168

```

The 'Console' tab (bottom-left) shows the execution of the script:

```

R 4.2.1 · C:/Users/kanik/Desktop/Data Mgmt/
> mean(ABT_Customer_InsurancePolicies$Age)
[1] 41.3295
>
> ABT_Customer_InsurancePolicies$Age[is.na(ABT_Customer_InsurancePolicies$Age)] <-
+   mean(ABT_Customer_InsurancePolicies$Age,na.rm = TRUE)
>
> summary(ABT_Customer_InsurancePolicies$Age)
  Min. 1st Qu. Median 3rd Qu. Max.
18.00 22.00 46.00 41.33 50.00 84.00

```

The 'Environment' tab (right) shows variables: \$ Surname, \$ CardType, \$ Occupation, \$ Gender, \$ Age, \$ Location, \$ ComChannel.

Data Cleaning For Categorical variables

Data cleaning for categorical variables

summary(ABT_Customer_InsurancePolicies\$Gender)

*ABT_Customer_InsurancePolicies\$Gender <-
as.factor(ABT_Customer_InsurancePolicies\$Gender)*

summary(ABT_Customer_InsurancePolicies\$Gender)

Changing values to accurate values

m to male and f to female in gender attribute

ABT_Customer_InsurancePolicies\$Gender[ABT_Customer_InsurancePolicies\$Gender == "f"] <- "female"

ABT_Customer_InsurancePolicies\$Gender[ABT_Customer_InsurancePolicies\$Gender == "m"] <- "male"

summary(ABT_Customer_InsurancePolicies\$Gender)

```
ABT_Customer_InsurancePolicies$Gender<-
droplevels(ABT_Customer_InsurancePolicies$Gender)

summary(ABT_Customer_InsurancePolicies$Gender)

## s to SMS , E to Email and P to Phone in ComChannel attribute

summary(ABT_Customer_InsurancePolicies$ComChannel)

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

summary(ABT_Customer_InsurancePolicies$ComChannel)

ABT_Customer_InsurancePolicies$ComChannel[ABT_Customer_InsurancePolicies$ComCh
annel == "S"] <- "SMS"

ABT_Customer_InsurancePolicies$ComChannel[ABT_Customer_InsurancePolicies$ComCh
annel == "E"] <- "Email"

ABT_Customer_InsurancePolicies$ComChannel[ABT_Customer_InsurancePolicies$ComCh
annel == "P"] <- "Phone"

summary(ABT_Customer_InsurancePolicies$ComChannel)

ABT_Customer_InsurancePolicies$Gender<-
droplevels(ABT_Customer_InsurancePolicies$ComChannel)

summary(ABT_Customer_InsurancePolicies$ComChannel)
```

The screenshot shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The Project pane is labeled '(None)'. The Environment pane shows a global environment with variables like CustomerID, Title, GivenName, MiddleInitial, Surname, and CardType. The Data pane displays a summary of the ABT_Customer_InsurancePolicies dataset with 4085 observations. The Plots pane contains a box plot for a variable named '80'. The bottom console pane shows R code being run:

```
## Data cleaning for categorical variables
summary(ABT_Customer_InsurancePolicies$Gender)
ABT_Customer_InsurancePolicies$Gender <- as.factor(ABT_Customer_InsurancePolicies$Gender)
summary(ABT_Customer_InsurancePolicies$Gender)
## Changing values to accurate values
## m to male and f to female in gender attribute
ABT_Customer_InsurancePolicies$Gender[ABT_Customer_InsurancePolicies$Gender == "f"] <- "female"
ABT_Customer_InsurancePolicies$Gender[ABT_Customer_InsurancePolicies$Gender == "m"] <- "male"
summary(ABT_Customer_InsurancePolicies$Gender)
ABT_Customer_InsurancePolicies$Gender<- droplevels(ABT_Customer_InsurancePolicies$Gender)
summary(ABT_Customer_InsurancePolicies$Gender)
## s to SMS , E to Email and P to Phone in ComChannel attribute
summary(ABT_Customer_InsurancePolicies$ComChannel)
ABT_Customer_InsurancePolicies$ComChannel <- as.factor(ABT_Customer_InsurancePolicies$ComChannel)
summary(ABT_Customer_InsurancePolicies$ComChannel)
ABT_Customer_InsurancePolicies$ComChannel[ABT_Customer_InsurancePolicies$ComChannel == "s"] <- "SMS"
4
174:1671 (Top Level) ↴
```

The terminal pane shows the command prompt and the output of the R session.

