2. Using the "Marketing Customer Value Analysis" dataset, complete the following tasks with proper analysis and interpretation:

i. Load the dataset and explore its structure using basic commands.

```
library(ggplot2)
library(knitr)
```

```
marketing_df = read.csv("D:/PYTHON/DATA SCIENCE/DATA//WA_Fn-UseC_-Marketing-Customer-Value-Analysis (1)
```

```
head(marketing_df,5)
```

```
##   Customer      State Customer.Lifetime.Value Response Coverage Education
## 1  BU79786 Washington               2763.519       No    Basic  Bachelor
## 2  QZ44356    Arizona               6979.536       No Extended  Bachelor
## 3  AI49188     Nevada              12887.432       No  Premium  Bachelor
## 4  WW63253 California               7645.862       No    Basic  Bachelor
## 5  HB64268 Washington               2813.693       No    Basic  Bachelor
##   Effective.To.Date EmploymentStatus Gender Income Location.Code Marital.Status
## 1           2/24/11         Employed      F  56274      Suburban        Married
## 2           1/31/11       Unemployed      F      0      Suburban         Single
## 3           2/19/11         Employed      F  48767      Suburban        Married
## 4           1/20/11       Unemployed      M      0      Suburban        Married
## 5            2/3/11         Employed      M  43836         Rural         Single
##   Monthly.Premium.Auto Months.Since.Last.Claim Months.Since.Policy.Inception
## 1                   69                     32                             5
## 2                   94                     13                            42
## 3                  108                     18                            38
## 4                  106                     18                            65
## 5                   73                     12                            44
##   Number.of.Open.Complaints Number.of.Policies    Policy.Type       Policy
## 1                         0                  1 Corporate Auto Corporate L3
## 2                         0                  8  Personal Auto  Personal L3
## 3                         0                  2  Personal Auto  Personal L3
## 4                         0                  7 Corporate Auto Corporate L2
## 5                         0                  1  Personal Auto  Personal L1
##   Renew.Offer.Type Sales.Channel Total.Claim.Amount Vehicle.Class Vehicle.Size
## 1           Offer1         Agent           384.8111  Two-Door Car      Medsize
## 2           Offer3         Agent          1131.4649 Four-Door Car      Medsize
## 3           Offer1         Agent           566.4722  Two-Door Car      Medsize
## 4           Offer1   Call Center           529.8813           SUV      Medsize
## 5           Offer1         Agent           138.1309 Four-Door Car      Medsize
```

```
str(marketing_df)
```

```
## 'data.frame':    9134 obs. of  24 variables:
##  $ Customer                     : chr  "BU79786" "QZ44356" "AI49188" "WW63253" ...
##  $ State                        : chr  "Washington" "Arizona" "Nevada" "California" ...
##  $ Customer.Lifetime.Value      : num  2764 6980 12887 7646 2814 ...
##  $ Response                     : chr  "No" "No" "No" "No" ...
##  $ Coverage                     : chr  "Basic" "Extended" "Premium" "Basic" ...
```

```
##  $ Education                    : chr  "Bachelor" "Bachelor" "Bachelor" "Bachelor" ...
##  $ Effective.To.Date            : chr  "2/24/11" "1/31/11" "2/19/11" "1/20/11" ...
##  $ EmploymentStatus             : chr  "Employed" "Unemployed" "Employed" "Unemployed" ...
##  $ Gender                       : chr  "F" "F" "F" "M" ...
##  $ Income                       : int  56274 0 48767 0 43836 62902 55350 0 14072 28812 ...
##  $ Location.Code                : chr  "Suburban" "Suburban" "Suburban" "Suburban" ...
##  $ Marital.Status               : chr  "Married" "Single" "Married" "Married" ...
##  $ Monthly.Premium.Auto         : int  69 94 108 106 73 69 67 101 71 93 ...
##  $ Months.Since.Last.Claim      : int  32 13 18 18 12 14 0 0 13 17 ...
##  $ Months.Since.Policy.Inception: int  5 42 38 65 44 94 13 68 3 7 ...
##  $ Number.of.Open.Complaints    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Number.of.Policies           : int  1 8 2 7 1 2 9 4 2 8 ...
##  $ Policy.Type                  : chr  "Corporate Auto" "Personal Auto" "Personal Auto" "Corporate Au
##  $ Policy                       : chr  "Corporate L3" "Personal L3" "Personal L3" "Corporate L2" ...
##  $ Renew.Offer.Type             : chr  "Offer1" "Offer3" "Offer1" "Offer1" ...
##  $ Sales.Channel                : chr  "Agent" "Agent" "Agent" "Call Center" ...
##  $ Total.Claim.Amount           : num  385 1131 566 530 138 ...
##  $ Vehicle.Class                : chr  "Two-Door Car" "Four-Door Car" "Two-Door Car" "SUV" ...
##  $ Vehicle.Size                 : chr  "Medsize" "Medsize" "Medsize" "Medsize" ...
```

```
summary(marketing_df)
```

```
##    Customer            State           Customer.Lifetime.Value
##  Length:9134        Length:9134        Min.   : 1898
##  Class :character   Class :character   1st Qu.: 3994
##  Mode  :character   Mode  :character   Median : 5780
##                                        Mean   : 8005
##                                        3rd Qu.: 8962
##                                        Max.   :83325
##    Response           Coverage          Education         Effective.To.Date
##  Length:9134        Length:9134        Length:9134        Length:9134
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  EmploymentStatus      Gender              Income      Location.Code
##  Length:9134        Length:9134        Min.   :    0   Length:9134
##  Class :character   Class :character   1st Qu.:    0   Class :character
##  Mode  :character   Mode  :character   Median :33890   Mode  :character
##                                        Mean   :37657
##                                        3rd Qu.:62320
##                                        Max.   :99981
##  Marital.Status     Monthly.Premium.Auto Months.Since.Last.Claim
##  Length:9134        Min.   : 61.00       Min.   : 0.0
##  Class :character   1st Qu.: 68.00       1st Qu.: 6.0
##  Mode  :character   Median : 83.00       Median :14.0
##                     Mean   : 93.22       Mean   :15.1
##                     3rd Qu.:109.00       3rd Qu.:23.0
##                     Max.   :298.00       Max.   :35.0
##  Months.Since.Policy.Inception Number.of.Open.Complaints Number.of.Policies
##  Min.   : 0.00                 Min.   :0.0000           Min.   :1.000
##  1st Qu.:24.00                 1st Qu.:0.0000           1st Qu.:1.000
##  Median :48.00                 Median :0.0000           Median :2.000
```

```
##  Mean   :48.06                 Mean   :0.3844            Mean   :2.966
##  3rd Qu.:71.00                 3rd Qu.:0.0000            3rd Qu.:4.000
##  Max.   :99.00                 Max.   :5.0000            Max.   :9.000
##  Policy.Type           Policy          Renew.Offer.Type   Sales.Channel
##  Length:9134        Length:9134        Length:9134        Length:9134
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  Total.Claim.Amount Vehicle.Class      Vehicle.Size
##  Min.   :   0.099   Length:9134        Length:9134
##  1st Qu.: 272.258   Class :character   Class :character
##  Median : 383.945   Mode  :character   Mode  :character
##  Mean   : 434.089
##  3rd Qu.: 547.515
##  Max.   :2893.240
```

```
dim(marketing_df)
```

```
## [1] 9134   24
```

The Basic commands provide an overview of the dataset.

ii. Create a new column named "Engaged" by transforming the categorical values in the "Response" variable into numerical values. Why is this transformation important?

```
marketing_df$Engaged = ifelse(marketing_df$Response == "Yes", 1, 0)
head(marketing_df,3)
```

```
##   Customer      State Customer.Lifetime.Value Response Coverage Education
## 1  BU79786 Washington               2763.519       No    Basic  Bachelor
## 2  QZ44356    Arizona               6979.536       No Extended  Bachelor
## 3  AI49188     Nevada              12887.432       No  Premium  Bachelor
##   Effective.To.Date EmploymentStatus Gender Income Location.Code Marital.Status
## 1           2/24/11         Employed      F  56274      Suburban        Married
## 2           1/31/11       Unemployed      F      0      Suburban         Single
## 3           2/19/11         Employed      F  48767      Suburban        Married
##   Monthly.Premium.Auto Months.Since.Last.Claim Months.Since.Policy.Inception
## 1                   69                     32                             5
## 2                   94                     13                            42
## 3                  108                     18                            38
##   Number.of.Open.Complaints Number.of.Policies    Policy.Type       Policy
## 1                         0                 1 Corporate Auto Corporate L3
## 2                         0                 8  Personal Auto  Personal L3
## 3                         0                 2  Personal Auto  Personal L3
##   Renew.Offer.Type Sales.Channel Total.Claim.Amount Vehicle.Class Vehicle.Size
## 1           Offer1         Agent           384.8111  Two-Door Car      Medsize
## 2           Offer3         Agent          1131.4649 Four-Door Car      Medsize
## 3           Offer1         Agent           566.4722  Two-Door Car      Medsize
##   Engaged
## 1       0
## 2       0
## 3       0
```

Interpretation: Converts "Yes/No" responses into 1 and 0 for numerical analysis.

iii. Calculate and interpret the Engagement Rate. How is it computed, and what does it indicate about the customer responses?

```
print(sum(marketing_df$Engaged)/nrow(df)*100)
```

```
## numeric(0)
```

Interpretation: Engagement rate shows the percentage of customers who responded positively.

iv. Analyze engagement rate by "Renew Offer Type" and "Sales Channel":

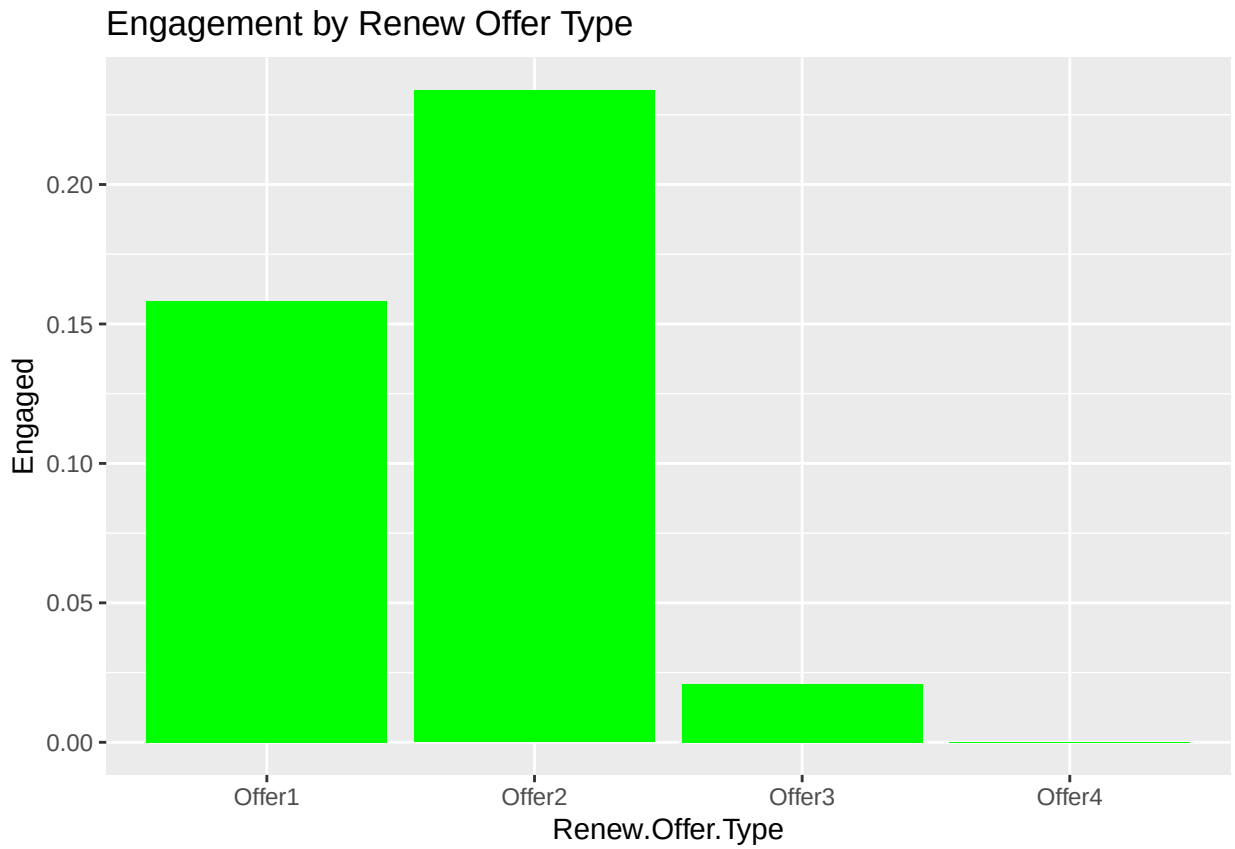
```
aggregate(Engaged ~ Renew.Offer.Type + Sales.Channel, data = marketing_df, mean)
```

```
##    Renew.Offer.Type Sales.Channel    Engaged
## 1            Offer1         Agent 0.19881657
## 2            Offer2         Agent 0.31901840
## 3            Offer3         Agent 0.03474903
## 4            Offer4         Agent 0.00000000
## 5            Offer1        Branch 0.15286624
## 6            Offer2        Branch 0.15989848
## 7            Offer3        Branch 0.00000000
## 8            Offer4        Branch 0.00000000
## 9            Offer1   Call Center 0.07142857
## 10           Offer2   Call Center 0.23183926
## 11           Offer3   Call Center 0.00000000
## 12           Offer4   Call Center 0.00000000
## 13           Offer1           Web 0.12800000
## 14           Offer2           Web 0.18713450
## 15           Offer3           Web 0.05240175
## 16           Offer4           Web 0.00000000
```

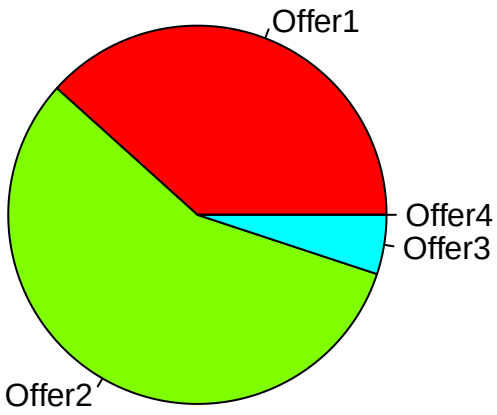
Interpretation: Groups data by "Renew Offer Type" and "Sales Channel".

v. Use a pivot table to summarize engagement by "Renew Offer Type" and visualize the results using both bar and pie charts. Why are these visualizations helpful in understanding customer engagement patterns?

```
#Bar chart
ggplot(marketing_df, aes(x = Renew.Offer.Type, y = Engaged)) +
  stat_summary(fun = mean, geom = "bar", fill = "green") +
  ggtitle("Engagement by Renew Offer Type")
```




```
# Pie chart
engagement_summary = aggregate(Engaged ~ Renew.Offer.Type, data = marketing_df, mean)
pie(engagement_summary$Engaged, labels = engagement_summary$Renew.Offer.Type,
    main = "Engagement by Renew Offer Type", col = rainbow(length(engagement_summary$Engaged)))
```



Interpretation: Bar and pie charts help visualize clear trends.

vi. Explain the purpose of regression analysis in this context. Describe how you would approach regression using (i) continuous variables only, (ii) categorical variables, and (iii) both continuous and categorical variables. How would you interpret the outputs for each approach?

```
# Continuous variables only
model_cont = lm(Engaged ~ Income, data = marketing_df)
summary(model_cont)
```

```
##
## Call:
## lm(formula = Engaged ~ Income, data = marketing_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.1518 -0.1460 -0.1417 -0.1380  0.8620
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.380e-01  5.838e-03   23.64   <2e-16 ***
## Income      1.376e-07  1.207e-07    1.14    0.254
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.3503 on 9132 degrees of freedom
## Multiple R-squared:  0.0001424,  Adjusted R-squared:  3.289e-05
## F-statistic:   1.3 on 1 and 9132 DF,  p-value: 0.2542
```

```
# Categorical variables
marketing_df$Renew.Offer.Type <- as.factor(marketing_df$Renew.Offer.Type)
model_cat <- glm(Engaged ~ Renew.Offer.Type, family = binomial, data = marketing_df)
summary(model_cat)
```

```
##
## Call:
## glm(formula = Engaged ~ Renew.Offer.Type, family = binomial,
##     data = marketing_df)
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)              -1.67081    0.04472 -37.359  < 2e-16 ***
## Renew.Offer.TypeOffer2    0.48365    0.06252   7.736 1.02e-14 ***
## Renew.Offer.TypeOffer3   -2.17364    0.18986 -11.449  < 2e-16 ***
## Renew.Offer.TypeOffer4 -16.89525  203.83246  -0.083    0.934
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 7503.3  on 9133  degrees of freedom
## Residual deviance: 6751.8  on 9130  degrees of freedom
## AIC: 6759.8
##
## Number of Fisher Scoring iterations: 17
```

```
# Both types
both_model <- glm(Engaged ~ Income + Renew.Offer.Type, family = binomial, data = marketing_df)
summary(both_model)
```

```
##
## Call:
## glm(formula = Engaged ~ Income + Renew.Offer.Type, family = binomial,
##     data = marketing_df)
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -1.700e+00  5.522e-02 -30.789  < 2e-16 ***
## Income                   9.426e-07  1.035e-06   0.911    0.362
## Renew.Offer.TypeOffer2   4.712e-01  6.399e-02   7.363 1.79e-13 ***
## Renew.Offer.TypeOffer3 -2.181e+00  1.900e-01 -11.477  < 2e-16 ***
## Renew.Offer.TypeOffer4 -1.691e+01  2.038e+02  -0.083    0.934
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 7503.3  on 9133  degrees of freedom
```

```
## Residual deviance: 6751.0  on 9129  degrees of freedom
## AIC: 6761
##
## Number of Fisher Scoring iterations: 17
```

Purpose of Regression Analysis:

```
Regression analysis helps identify and quantify the relationship between variables, allowing us to p
```

Interpretation: i) Continuous Variables: Measures the direct impact of numerical predictors on the target, with coefficients showing the change in the target for each unit change in predictors.

ii) Categorical Variables: Uses dummy encoding to compare the effect of different groups, where coefficients represent the difference from the reference group.

iii) Both Variables: Combines numeric and categorical data, interpreting continuous variables as direct effects and categorical ones relative to the reference group.