Q1

Gowtham P

1. Load the dataset “WA_Fn-UseC_-Marketing-Customer-Value-Analysis.csv” using pd.read_csv() and perform the following tasks with appropriate interpretation:

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
#### 1. Load Data ####
df <- read.csv("D:/Data Science for Marketing-I& 2/dataset/WA_Fn-UseC_-Marketing-Customer-Value-Analysis.csv")
```

i. Perform basic exploratory data analysis (EDA) such as checking dataset shape and previewing the first few rows. What insights can be drawn from this initial exploration?

```
head(df)
```

```
##   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
## 6  OC83172     Oregon                8256.298      Yes    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
## 6           1/25/11         Employed      F  62902         Rural        Married
##   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
## 6                   69                      14                            94
##   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
## 6                         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
## 4           Offer1   Call Center           529.8813           SUV      Medsize
## 5           Offer1         Agent           138.1309 Four-Door Car      Medsize
## 6           Offer2           Web           159.3830  Two-Door Car      Medsize
```

```
dim(df)
```

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

```
summary(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
```

Interpertation: The dataset has 9134 rows and 24 columns,head() function is used to display the first few rows of a dataset.

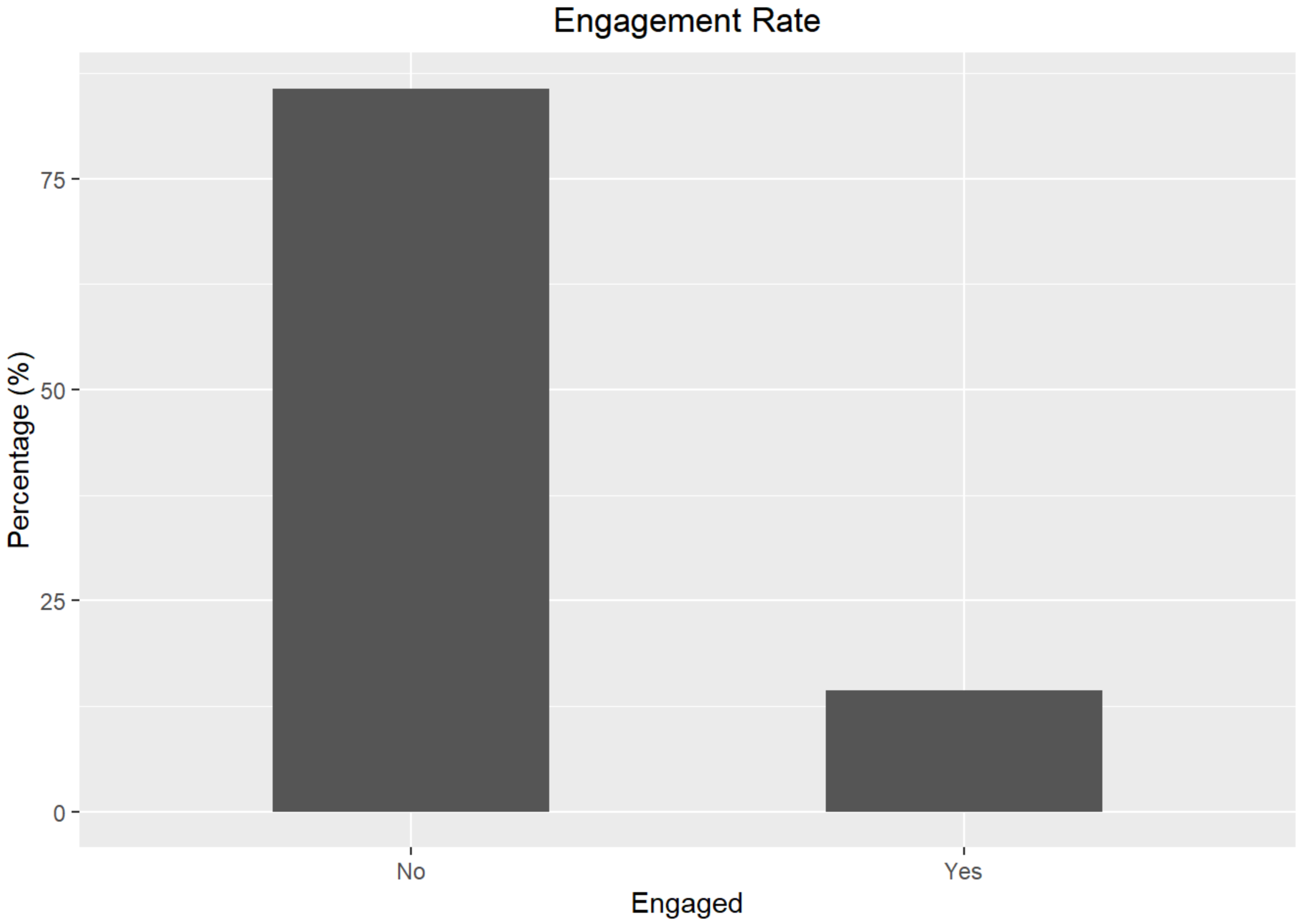
ii. Analyze customer engagement by grouping data based on the Response variable. How does this grouping help in understanding customer behavior?

```
# Encode engaged customers as 0s and 1s
df$Engaged <- rep(0,nrow(df))
df$Engaged[df$Response=='Yes']=1
```

```
## - Overall Engagement Rates ##
engagementRate <- df %>% group_by(Response) %>%
  summarise(Count=n())  %>%
  mutate(EngagementRate=Count/nrow(df)*100.0)
```

iii. Visualize the engagement rate using a bar chart. What is the significance of this visualization, and how does the code achieve it?

```
ggplot(engagementRate, aes(x=Response, y=EngagementRate)) +
  geom_bar(width=0.5, stat="identity") +
  ggtitle('Engagement Rate') +
  xlab("Engaged") +
  ylab("Percentage (%)") +
  theme(plot.title = element_text(hjust = 0.5))
```



Interpertation:

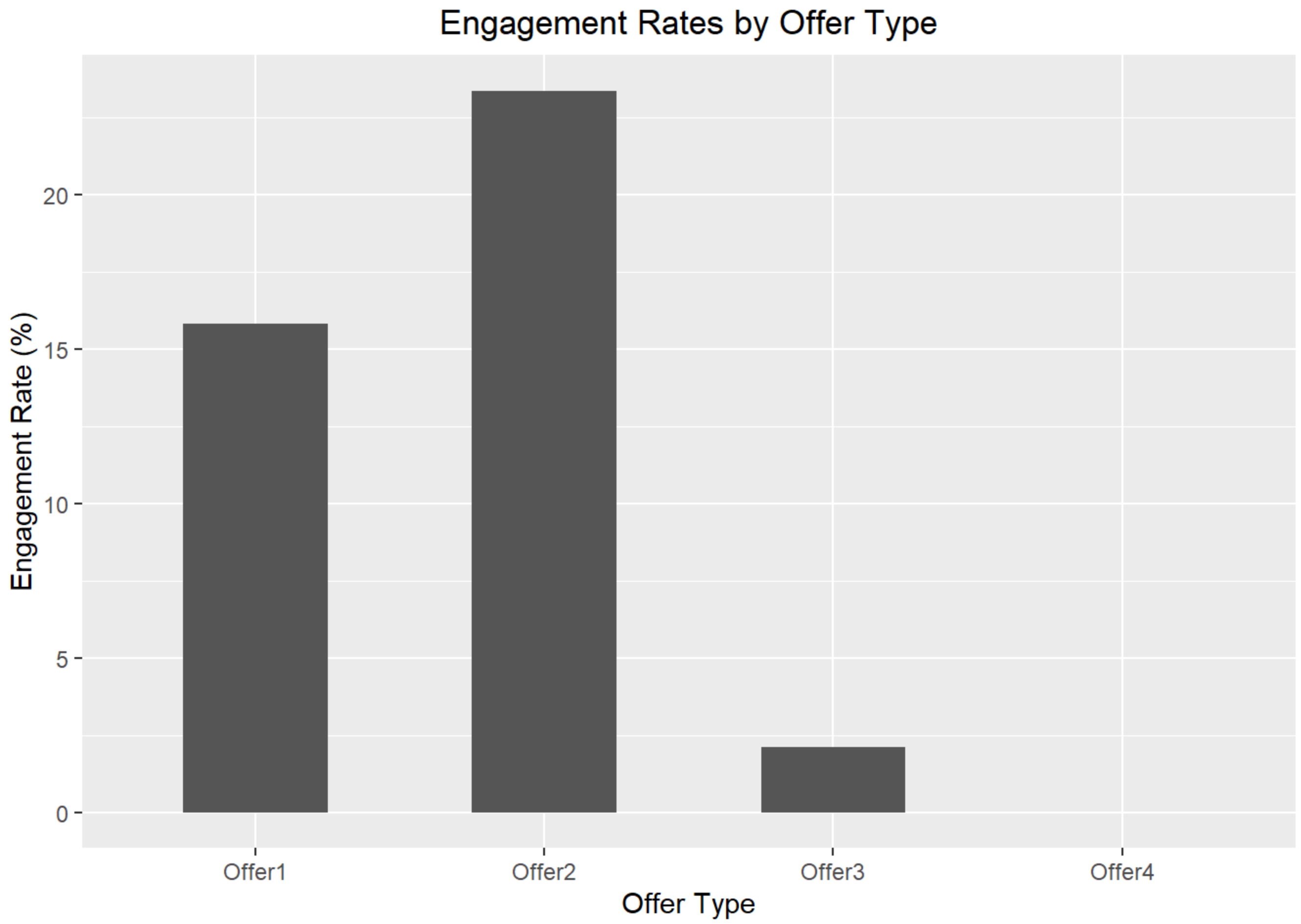
Only 14.3% of customers responded positively, indicating a low engagement rate. This suggests a need for improved marketing strategies to boost response rates.

iv. Calculate the engagement rate for different renewal offer types and interpret the results. Why is this metric useful?

```
## - Engagement Rates by Offer Type ##
engagementRateByOfferType <- df %>%
  group_by(Renew.Offer.Type) %>%
  summarise(Count=n(), NumEngaged=sum(Engaged))  %>%
  mutate(EngagementRate=NumEngaged/Count*100.0)
engagementRateByOfferType
```

```
## # A tibble: 4 × 4
##   Renew.Offer.Type Count NumEngaged EngagementRate
##   <chr>            <int>      <dbl>          <dbl>
## 1 Offer1            3752        594          15.8
## 2 Offer2            2926        684          23.4
## 3 Offer3            1432         30           2.09
## 4 Offer4            1024          0           0
```

```
ggplot(engagementRateByOfferType, aes(x=Renew.Offer.Type, y=EngagementRate)) +
  geom_bar(width=0.5, stat="identity") +
  ggtitle('Engagement Rates by Offer Type') +
  xlab("Offer Type") +
  ylab("Engagement Rate (%)") +
  theme(plot.title = element_text(hjust = 0.5))
```



Interpertation:

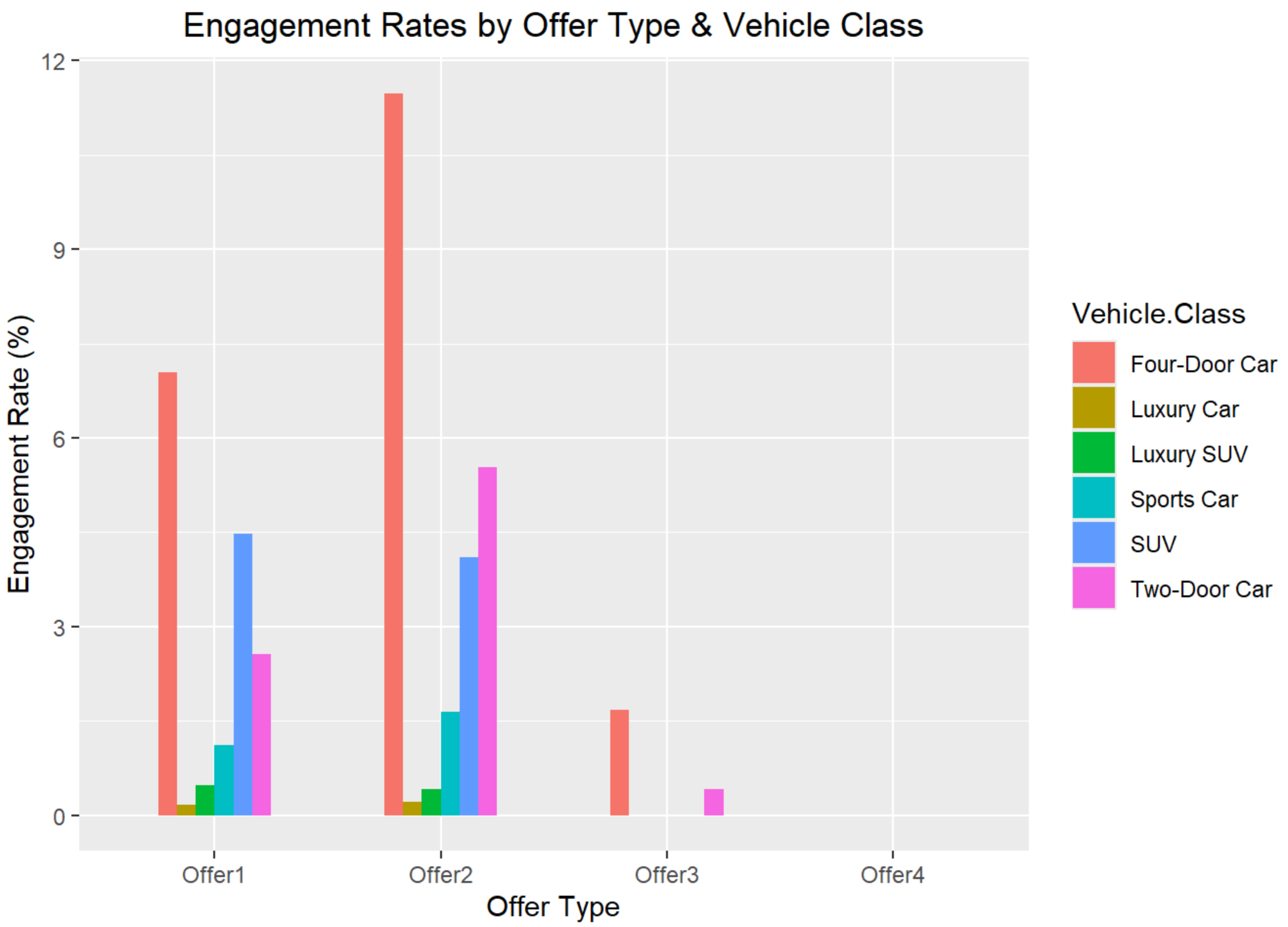
Offer1 and Offer2 have the highest engagement rates (~16%), while Offer4 has the lowest (9.7%). This suggests that some offers are more attractive, guiding future marketing strategies.

v. Extend the analysis by exploring engagement rates segmented by both Renew Offer Type and Vehicle Class. How does this multi-level grouping provide deeper insights?

```
## - Offer Type & Vehicle Class ##
engagementRateByOfferTypeVehicleClass <- df %>%
  group_by(Renew.Offer.Type, Vehicle.Class) %>%
  summarise(NumEngaged=sum(Engaged))  %>%
  left_join(engagementRateByOfferType[,c("Renew.Offer.Type", "Count")], by="Renew.Offer.Type") %
>%
  mutate(EngagementRate=NumEngaged/Count*100.0)
```

```
## `summarise()` has grouped output by 'Renew.Offer.Type'. You can override using
## the `.groups` argument.
```

```
ggplot(engagementRateByOfferTypeVehicleClass, aes(x=Renew.Offer.Type, y=EngagementRate, fill=Veh
icle.Class)) +
  geom_bar(width=0.5, stat="identity", position = "dodge") +
  ggtitle('Engagement Rates by Offer Type & Vehicle Class') +
  xlab("Offer Type") +
  ylab("Engagement Rate (%)") +
  theme(plot.title = element_text(hjust = 0.5))
```



Interpertation:

More customers responded to Offer 2, especially those with Four-Door Cars. Offers 3 and 4 had very few responses.

vi. Perform customer segmentation using the variables ‘Customer Lifetime Value (CLV)’ and ‘Months Since Policy Inception’

```
summary(df$Customer.Lifetime.Value)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##    1898    3994    5780    8005    8962   83325
```

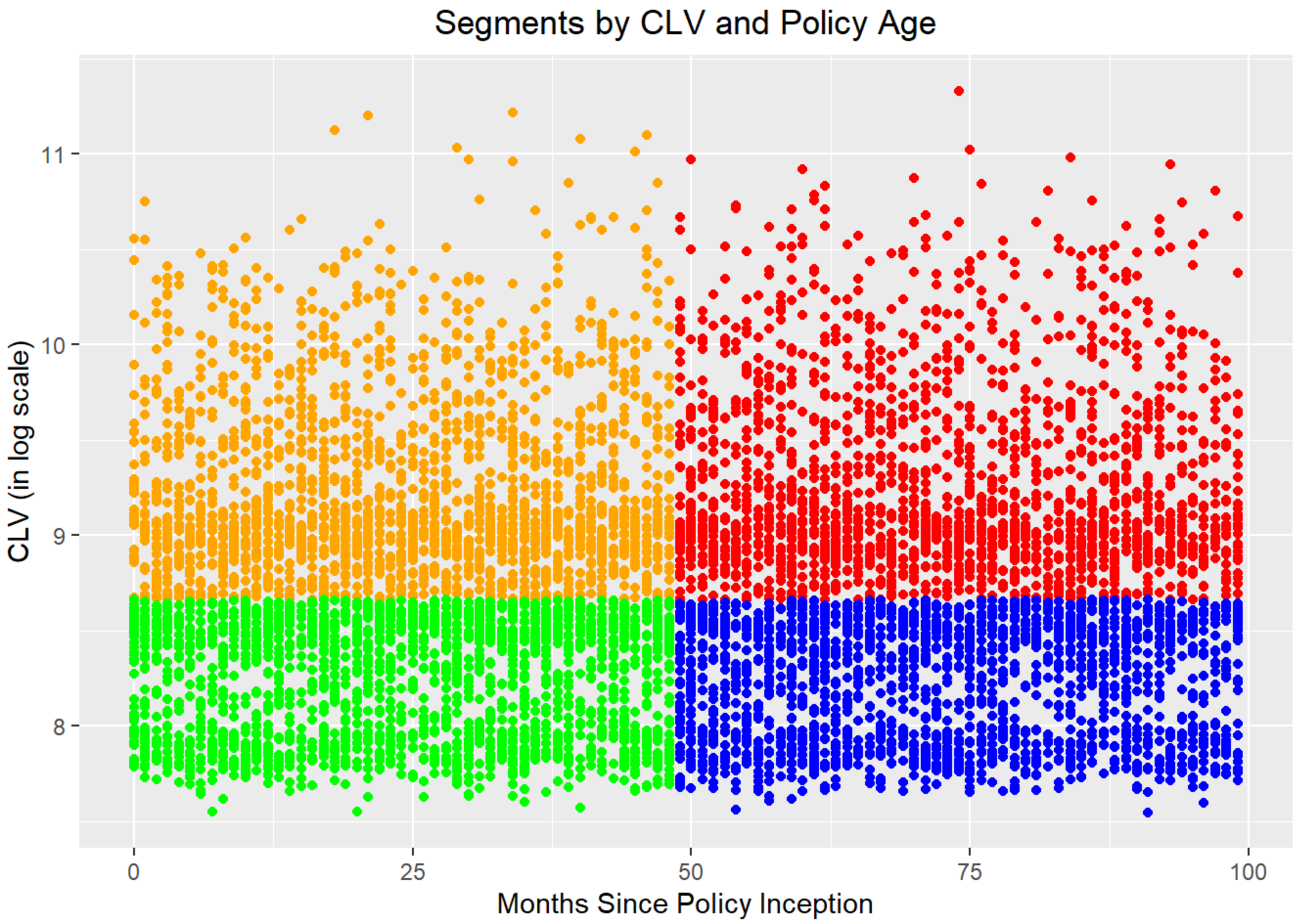
```
summary(df$Months.Since.Policy.Inception)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##    0.00   24.00   48.00   48.06   71.00   99.00
```

```
clv_encode_fn <- function(x) {if(x > median(df$Customer.Lifetime.Value)) "High" else "Low"}
df$CLV.Segment <- sapply(df$Customer.Lifetime.Value, clv_encode_fn)

policy_age_encode_fn <- function(x) {if(x > median(df$Months.Since.Policy.Inception)) "High" else "Low"}
df$Policy.Age.Segment <- sapply(df$Months.Since.Policy.Inception, policy_age_encode_fn)

ggplot(
  df[which(df$CLV.Segment=="High" & df$Policy.Age.Segment=="High"),], 
  aes(x=Months.Since.Policy.Inception, y=log(Customer.Lifetime.Value))
) +
  geom_point(color='red') +
  geom_point(
    data=df[which(df$CLV.Segment=="High" & df$Policy.Age.Segment=="Low"),], 
    color='orange'
  ) +
  geom_point(
    data=df[which(df$CLV.Segment=="Low" & df$Policy.Age.Segment=="Low"),], 
    color='green'
  ) +
  geom_point(
    data=df[which(df$CLV.Segment=="Low" & df$Policy.Age.Segment=="High"),], 
    color='blue'
  ) +
  ggtitle('Segments by CLV and Policy Age') +
  xlab("Months Since Policy Inception") +
  ylab("CLV (in log scale)") +
  theme(plot.title = element_text(hjust = 0.5))
```



Interpertation:

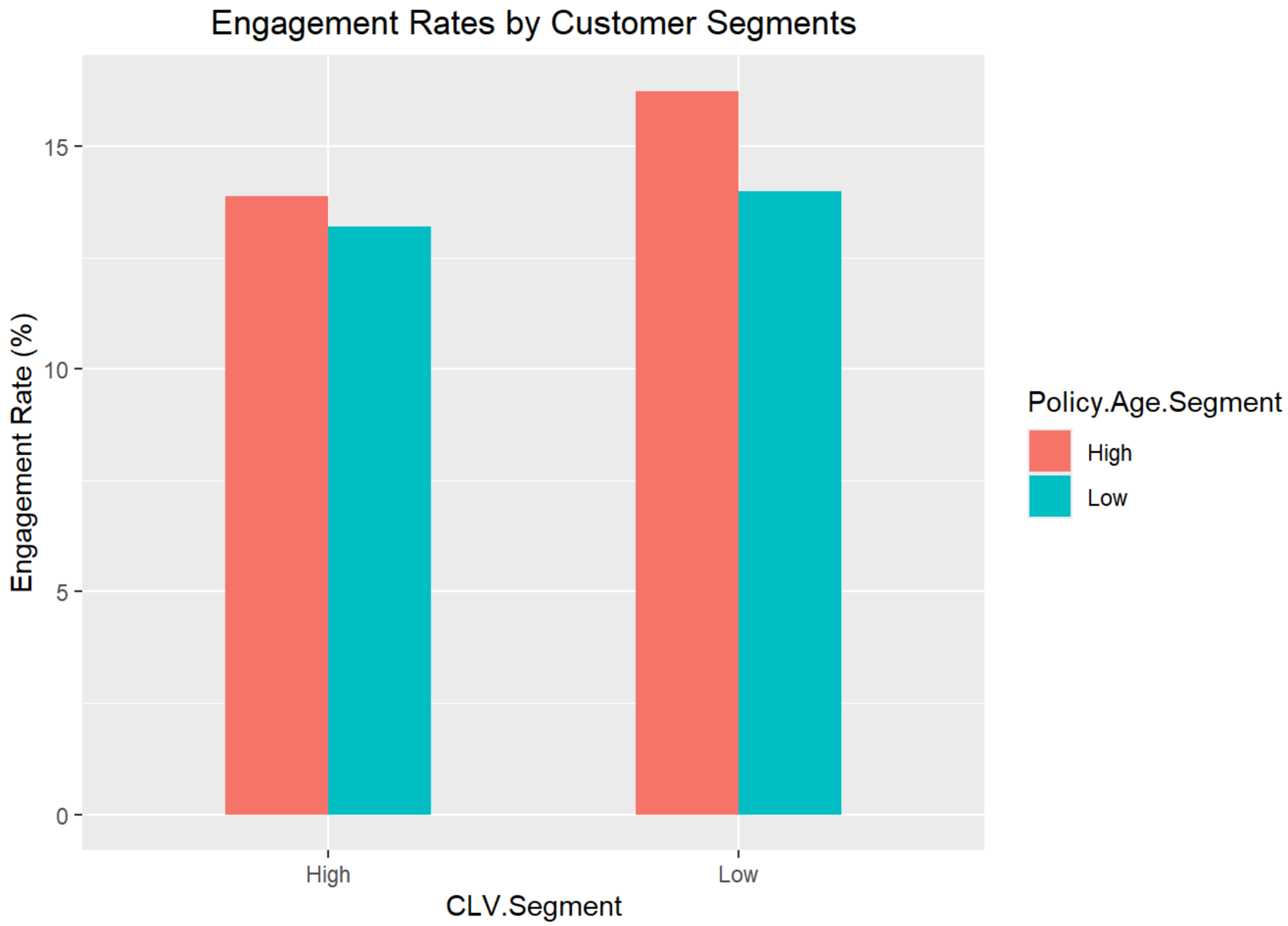
Customers are classified into High/Low CLV and Early/Late Policy Age groups. This segmentation helps in prioritizing high-value customers for retention.

vii. Create a visualization to compare CLV against Months Since Policy Inception

```
engagementRateBySegment <- df %>%
  group_by(CLV.Segment, Policy.Age.Segment) %>%
  summarise(Count=n(), NumEngaged=sum(Engaged))  %>%
  mutate(EngagementRate=NumEngaged/Count*100.0)
```

```
## `summarise()` has grouped output by 'CLV.Segment'. You can override using the
## `.groups` argument.
```

```
ggplot(engagementRateBySegment, aes(x=CLV.Segment, y=EngagementRate, fill=Policy.Age.Segment)) +
  geom_bar(width=0.5, stat="identity", position = "dodge") +
  ggtitle('Engagement Rates by Customer Segments') +
  ylab("Engagement Rate (%)") +
  theme(plot.title = element_text(hjust = 0.5))
```



Customers with high CLV stay longer, so it’s good to keep them happy.
Customers with low CLV might leave early, so they need more attention.