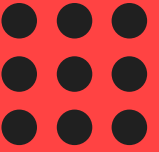


CAR RENTAL DATA

Natasya Septina Arianti - 2043201091





Dataset

Car Rental Data

GASOLINE	4.82
GASOLINE	4.80
GASOLINE	5.00
GASOLINE	5.00
GASOLINE	NA
GASOLINE	5.00
GASOLINE	5.00
	5.00
GASOLINE	5.00

Pada dataset Car Rental Data terdapat missing value dan outlier, sehingga dataset tersebut dapat digunakan



Missing Value

```
      rating
Min.      :1.00
1st Qu.   :4.90
Median    :5.00
Mean      :4.92
3rd Qu.   :5.00
Max.      :5.00
NA's      :501
```

```
GASOLINE
GASOLINE
GASOLINE

GASOLINE
  HYBRID
GASOLINE
```

Pada variabel rating dan fuelType diketahui terdapat missing value sehinggaa akan dilakukan habdling missing value pada variabel tersebut.,

Handling Missing Value

Car Rental Data

```
#handling missing value
data$rating <- with(data, impute(rating,median))
summary(data$rating)
```

```
#mendapatkan modus dari data kategorik
getmode <- function(v) {
  uniqv <- unique(v)
  uniqv[which.max(tabulate((match(v, uniqv))))]
}
```

```
#mendapatkan modus dari variabel fueltype
getmode(data$fuelType)
```

```
#handling missing value
library(dplyr)
data %>%
  mutate(imputed_fuelType = if_else(is.na(fuelType),
                                     getmode(fuelType),
                                     fuelType))
```

Handling missing value pada variabel rating menggunakan mean, sedangkan pada variabel fuelType menggunakan data kategori, karena data tersebut merupakan data kategorik



```
#identifikasi outlier pada data rating
summary(data$rating)
Q1 = quantile(data$rating, probs=0.25)
Q3 = quantile(data$rating, probs = 0.75)
IQR = Q3-Q1
IQR(data$rating)

#threshold/batas bawah dan batas atas
Tmin = Q1 - (1.5*IQR)
Tmax = Q3 + (1.5*IQR)
data$rating[which(data$rating < Tmin | data$rating > Tmax)]
```

Identifikasi outlier dilakukan menggunakan metode interquartile range

```
> data$rating[which(data$rating < Tmin | data$rating > Tmax)]
 [1] 4.42 4.76 4.70 4.50 4.67 4.50 4.72 4.50 4.58 4.73 4.74 4.77 4.51 4.74 4.75 4.77 4.6
[34] 4.71 4.64 4.50 4.69 4.65 4.58 4.70 4.65 4.75 4.73 4.63 4.75 4.69 4.75 4.50 4.75 4.6
[67] 4.75 4.33 4.75 4.38 4.00 4.73 4.67 4.17 4.77 4.00 4.60 4.65 4.37 4.41 4.60 4.55 4.0
[100] 4.60 4.57 4.64 4.71 4.63 4.63 4.77 4.62 4.72 4.59 4.75 4.77 4.50 4.67 4.67 4.75 2.0
[133] 4.71 4.65 4.63 4.64 3.67 4.73 4.77 4.00 4.71 4.50 4.76 4.77 3.67 4.56 4.76 4.75 4.0
[166] 4.71 4.00 4.50 4.76 4.56 4.74 4.75 4.70 4.58 4.75 4.66 4.67 4.69 4.56 4.75 4.74 4.7
[199] 4.57 4.00 4.50 4.00 4.50 4.69 4.72 4.20 4.33 4.57 4.60 4.71 3.67 4.60 4.64 4.77 4.6
[232] 4.76 4.59 3.56 4.76 4.60 4.75 4.67 4.48 4.67 4.48 4.64 4.71 4.58 4.57 4.63 4.75 4.7
[265] 4.75 4.75 4.72 4.77 4.71 4.60 4.00 3.50 3.25 4.00 4.71 4.76 4.67 4.69 4.73 4.71 4.6
[298] 4.50 4.43 4.33 4.67 4.60 4.71 4.50 4.76 4.75 4.69 4.76 4.72 4.75 4.60 4.67 4.77 3.0
[331] 4.60 4.77 4.50 4.63 4.67 4.61 4.63 4.75 4.57 4.75 4.50 4.43 4.70 4.70 4.75 4.65 4.7
[364] 4.71 4.25 4.55 4.71 4.75 4.74 4.75 4.76 4.71 4.67 4.00 4.67 4.77 4.00 4.64 4.77 4.7
[397] 4.20 4.69 4.44 4.67 4.71 4.60 4.73 4.64 4.75 4.71 4.71 4.75 4.70 4.51 4.55 4.68 4.5
[430] 4.62 4.63 4.60 4.63 4.71 4.77 4.73 4.77 4.50 4.50 4.77 3.67 4.64 4.00 3.00 4.60 4.7
[463] 4.67 4.77 4.74 4.47 4.56 4.68 4.71 4.40 4.75 4.50 4.50 4.68 4.75 4.69 4.75 4.60 4.0
[496] 4.75 4.67 4.75 4.68 4.67 4.76 4.64 4.57 4.48 4.73 4.64 4.69 4.60 4.60 4.60 4.73 4.0
[529] 4.50 4.68 4.71 4.63 4.40
```

OUTLIER

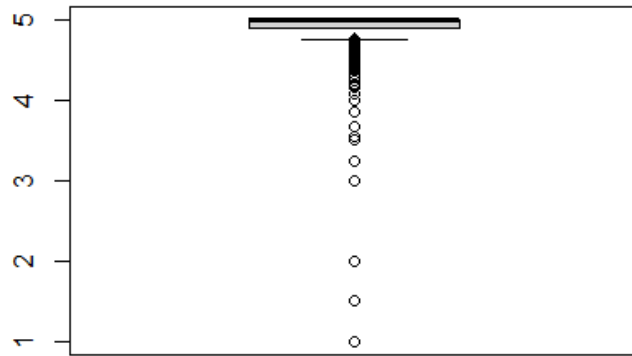
```
#handling outlier
log <- log10(data$rating)
hist(data$rating, col='steelblue', main='Original')
hist(log, col='coral2', main='Log Transformed')

#Removing outliers pada data rating
p75 = quantile(data$rating,0.75, na.rm = TRUE)
p25 = quantile(data$rating,0.25, na.rm = TRUE)
# print(p75)
# print(p25)
iqr = p75-p25
# print(head(is.na(data$rating)))
data$rating = ifelse(data$rating<p25-1.5*iqr,ave(data$rating, FUN = function(x) p25-1.5*iqr),data$rating)
data$rating = ifelse(data$rating>p75+1.5*iqr,ave(data$rating, FUN = function(x) p75+1.5*iqr),data$rating)
boxplot(data$rating, col ="lightblue", main="without outliers") # Attribute without outliers
|
```

Handling outlier dilakukan menggunakan metode transformasi log kemudian data yang outlier dihapus

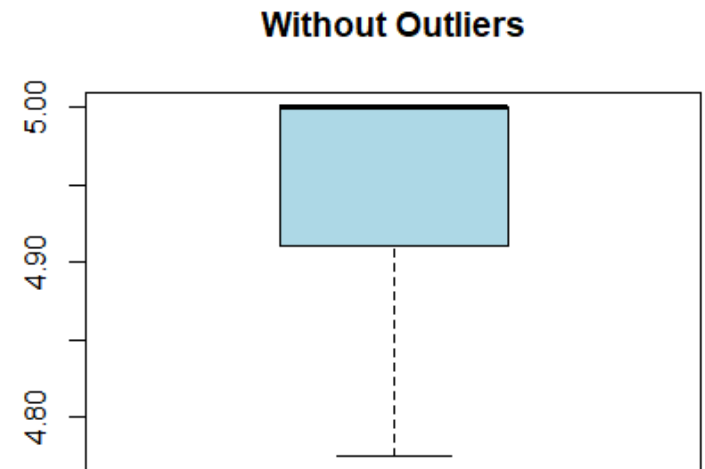
HANDLING OUTLIER

```
[1] 5.000 5.000 4.920 5.000 5.000 5.000 5.000 4.775 4.900 5.000 4.775 4
[28] 5.000 5.000 5.000 5.000 5.000 5.000 5.000 5.000 5.000 5.000 4.775 4
[55] 5.000 5.000 5.000 4.880 4.950 5.000 4.830 4.920 4.880 5.000 5.000 4
[82] 5.000 4.980 5.000 5.000 5.000 5.000 4.940 5.000 5.000 5.000 5.000 5
[109] 4.910 5.000 5.000 5.000 5.000 5.000 5.000 4.820 5.000 4.775 4.960 5
[136] 4.860 5.000 5.000 5.000 5.000 5.000 4.920 5.000 5.000 5.000 5.000 5
[163] 4.900 5.000 5.000 5.000 5.000 5.000 5.000 4.775 4.860 5.000 5.000 4
[190] 4.860 4.970 4.775 5.000 4.800 4.775 4.860 4.950 5.000 5.000 5.000 4
[217] 4.900 4.880 4.830 5.000 4.970 4.870 5.000 5.000 4.800 4.910 4.910 4
[244] 5.000 5.000 4.900 4.775 4.950 4.950 5.000 4.920 4.920 4.800 4.800 4
[271] 4.890 5.000 5.000 5.000 5.000 5.000 4.850 5.000 4.930 4.960 5.000 5
[298] 5.000 5.000 4.990 5.000 5.000 5.000 5.000 5.000 4.820 4.775 4.970 4
[325] 4.950 5.000 4.930 4.930 5.000 5.000 5.000 5.000 5.000 4.840 5.000 5
[352] 5.000 5.000 5.000 5.000 5.000 5.000 5.000 5.000 4.775 5.000 5.000 5
[379] 5.000 5.000 5.000 5.000 5.000 5.000 5.000 5.000 5.000 4.840 5.000 5
[406] 4.790 5.000 4.890 4.780 4.920 5.000 5.000 5.000 5.000 5.000 4.960 5
[433] 4.970 5.000 5.000 5.000 4.775 5.000 5.000 4.830 5.000 5.000 5.000 5
[460] 4.775 5.000 5.000 4.920 5.000 5.000 5.000 5.000 5.000 5.000 4.775 5
[487] 5.000 4.970 5.000 4.880 4.920 5.000 4.940 5.000 5.000 5.000 4.960 5
[514] 5.000 5.000 5.000 5.000 5.000 5.000 5.000 4.910 5.000 5.000 5.000 5
[541] 5.000 4.870 5.000 5.000 4.775 5.000 5.000 5.000 4.775 4.930 4.930 5
[568] 5.000 5.000 4.910 5.000 5.000 5.000 5.000 5.000 4.775 4.860 4.790 4
```



Sebelum dilakukan
handling outlier

DATA VISUALIZA TION



Setelah dilakukan
handling outlier