```
#########################################################################################################
#                                    LOAD THE REQUIRED LIBRARIERS                                       #
#########################################################################################################
```
##NOTE: Pre-req is to run the EDA_and_DataPreparation_CodeFinal.R file for running the below code**

```
library(DataExplorer)
library(lubridate)

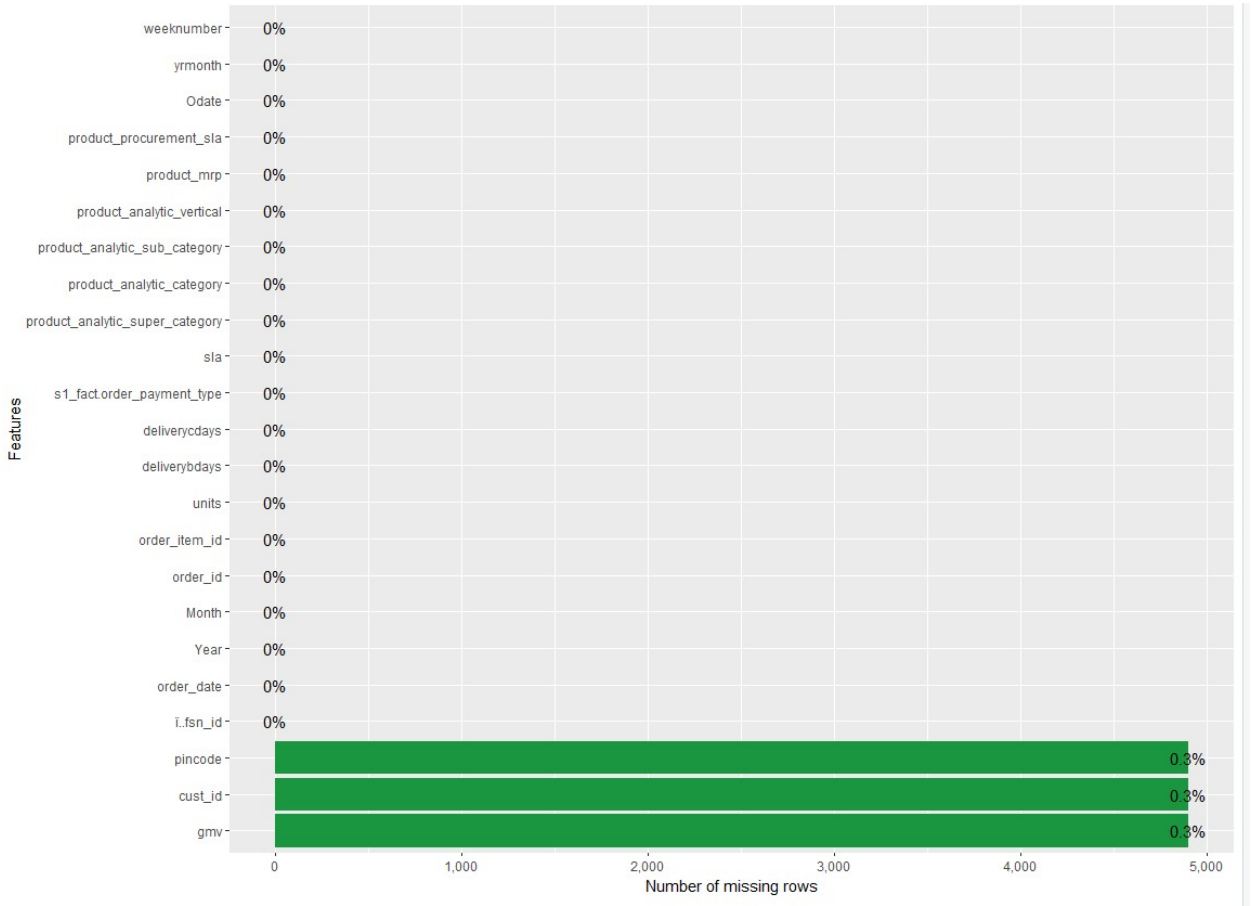
capstonewithoutna_withinperiod_bk <- capstonewithoutna_withinperiod
#str(capstonewithoutna_withinperiod)
analysisCategory <- c('CameraAccessory','HomeAudio','GamingAccessory')
capstonewithoutna_withinperiod_3subcategory <- filter ( capstonewithoutna_withinperiod
,capstonewithoutna_withinperiod$product_analytic_sub_category %in% analysisCategory )

capstonewithoutna_withinperiod1b<-capstonewithoutna_withinperiod_3subcategory
capstonewithoutna_withinperiod <-capstonewithoutna_withinperiod_3subcategory
```

#To Check the min and max of weeknumber after treating records not in our analysis period "July 15 to June 16"

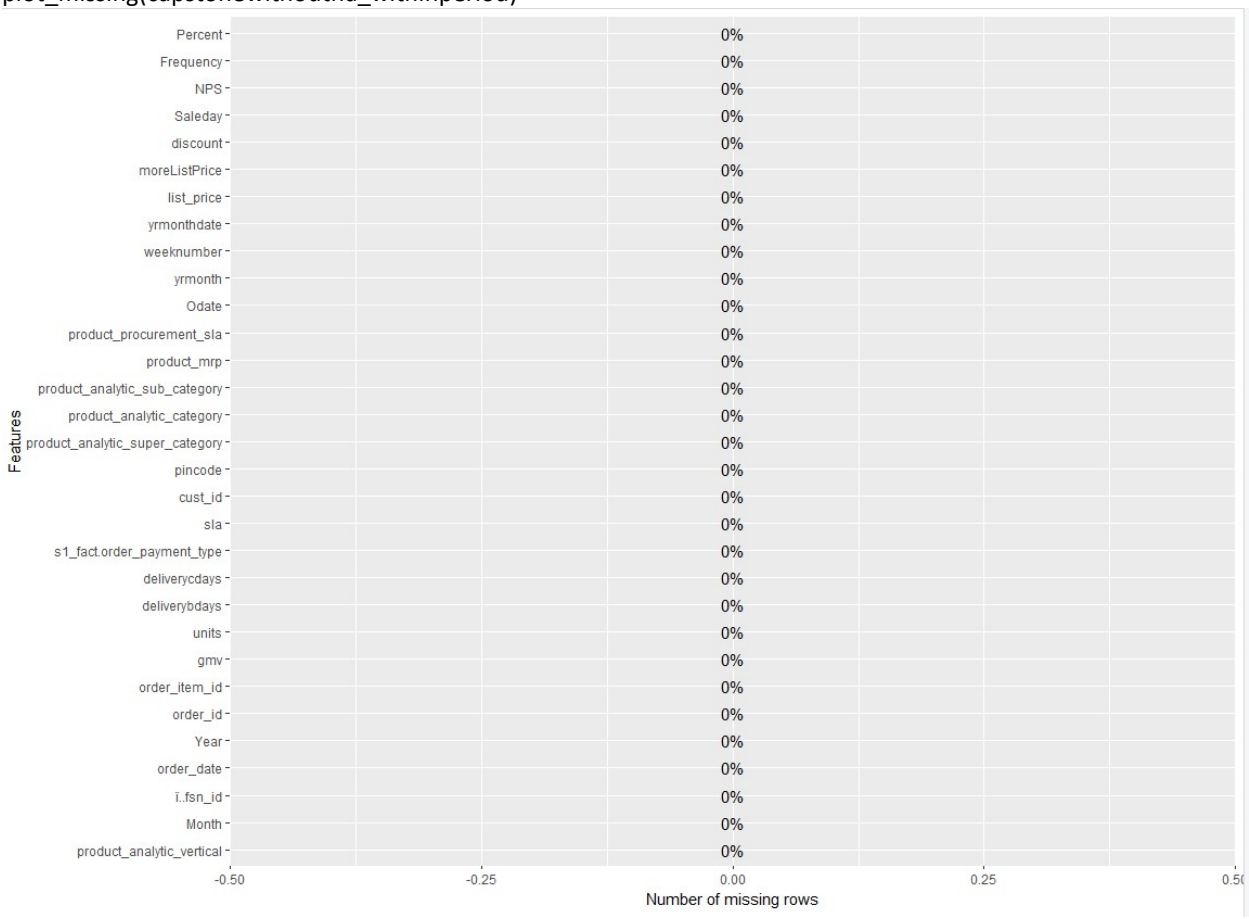
```
min(capstonewithoutna_withinperiod1b$weeknumber)
```  **#01**

```
max(capstonewithoutna_withinperiod1b$weeknumber)
```  **#53**

```
options(repr.plot.width=8, repr.plot.height=3)
# look for missing values using the DataExplorer package
plot_missing(capstone)
```



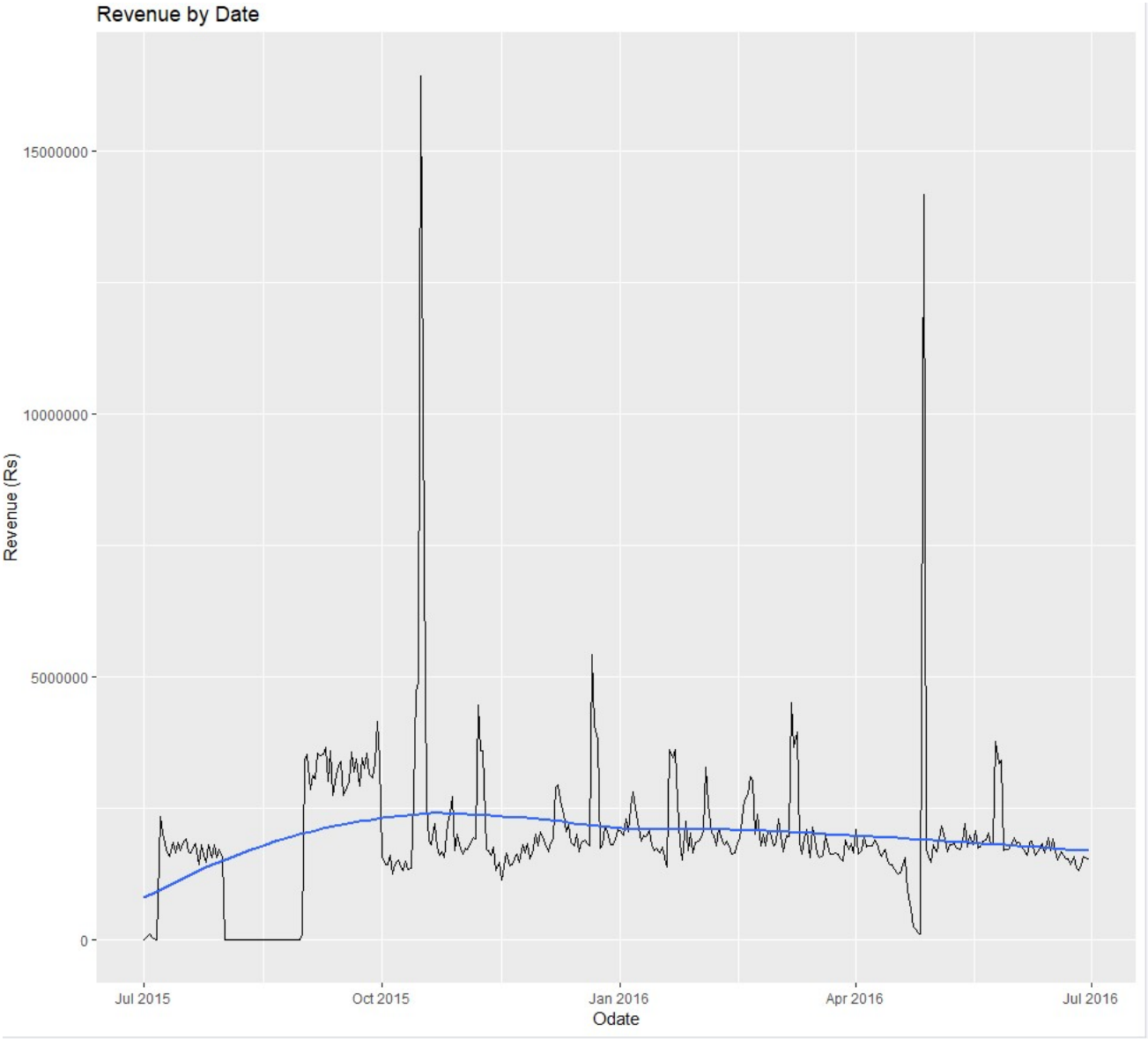
#After treating the N/A

plot_missing(capstonewithoutna_withinperiod)



#Revenue Vs Order Date

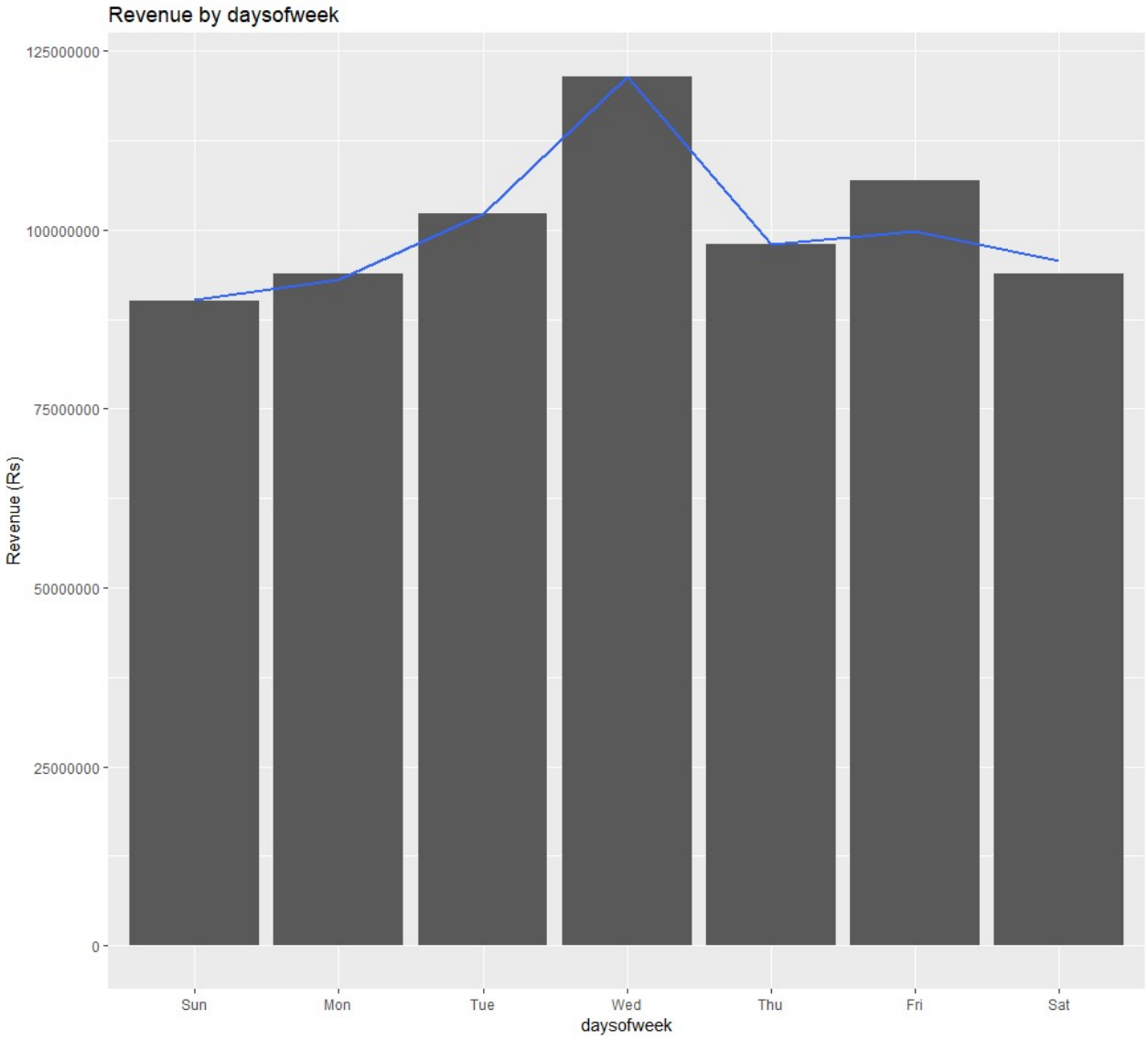
```
capstonewithoutna_withinperiod_DateRevenue <- capstonewithoutna_withinperiod %>% group_by(Odate) %>%
summarise(revenue = sum(gmv))
ggplot(capstonewithoutna_withinperiod_DateRevenue, aes(x = Odate, y = revenue)) + geom_line() +
geom_smooth(method = 'auto', se = FALSE) + labs(x = 'Odate', y = 'Revenue (Rs)', title = 'Revenue by Date')
```



Revenue Vs Order Date

#Revenue Vs DaysOfWeek

capstonewithoutna_withinperiod$daysofweek <- wday(capstonewithoutna_withinperiod$Odate,,label=TRUE)



Revenue Vs Day of Week

capstonewithoutna_withinperiod_weekdaysRevenue <- capstonewithoutna_withinperiod %>% group_by(daysofweek) %>% summarise(revenue = sum(gmv))
ggplot(capstonewithoutna_withinperiod_weekdaysRevenue, aes(x = daysofweek, y = revenue,group = 1)) + geom_col() + geom_smooth(method = 'auto', se = FALSE) + labs(x = 'daysofweek', y = 'Revenue (Rs)', title = 'Revenue by daysofweek')

#TO summarises what is happening on each day,

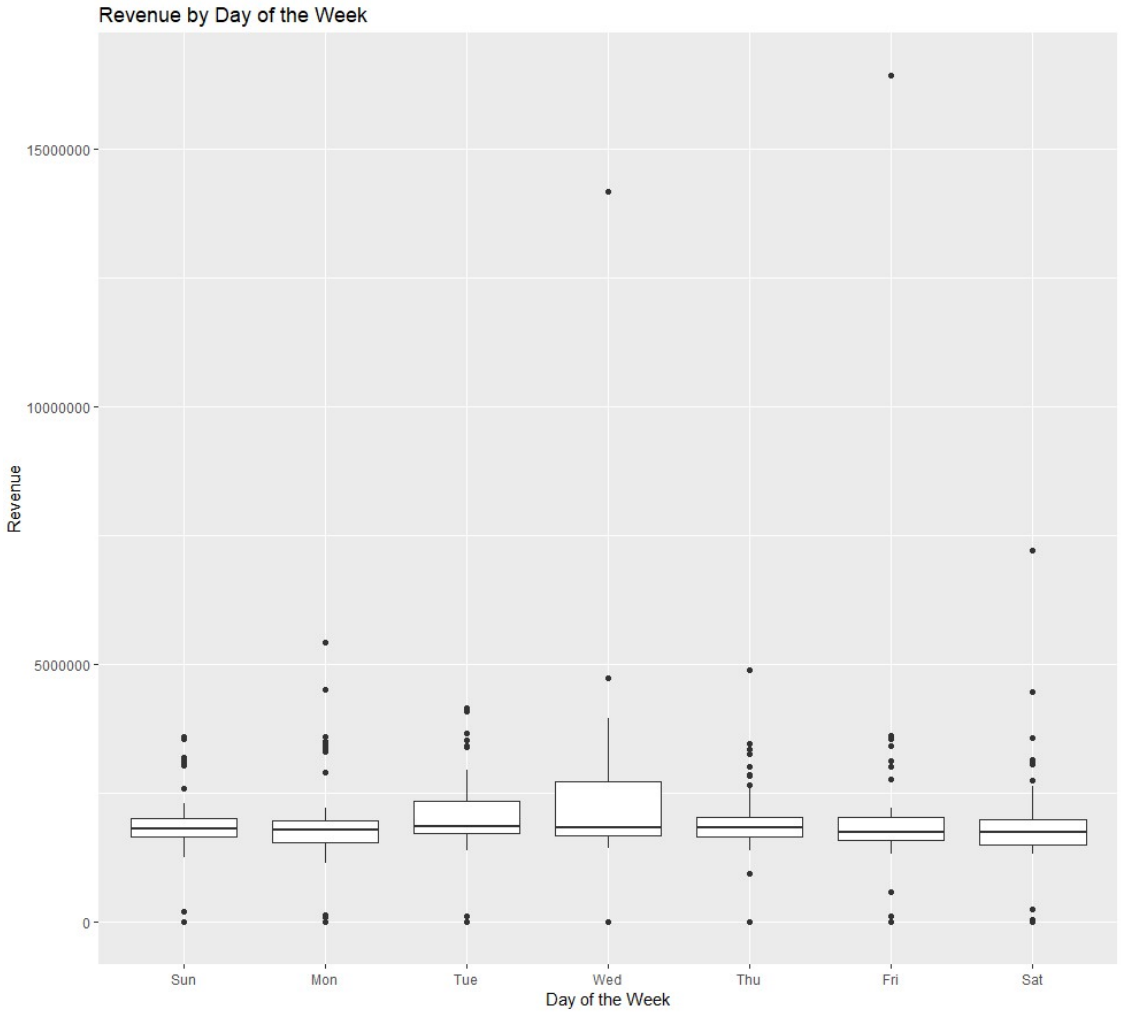
```
#Summarise Revenue, Transaction & AvgOrderValue With respect to dayOfWeek
capstonewithoutna_withinperiod_weekdayssummary <- capstonewithoutna_withinperiod %>%
  group_by(Odate,daysofweek) %>% summarise(revenue = sum(gmv), transactions = (n_distinct(order_id))) %>%
  mutate(AvgOrderValue = (round((revenue / transactions),2))) %>% ungroup()

head(capstonewithoutna_withinperiod_weekdayssummary)
```

	Odate	daysofweek	revenue	transactions	AvgOrderValue
	<date>	<ord>	<dbl>	<int>	<dbl>
1	2015-07-01	Wed	11051	6	1842.
2	2015-07-03	Fri	115347	71	1625.
3	2015-07-04	Sat	40775	16	2548.
4	2015-07-06	Mon	9877	3	3292.
5	2015-07-07	Tue	2346034	1721	1363.
6	2015-07-08	Wed	2044820	1522	1344.

#Plot for Revenue per day of the week

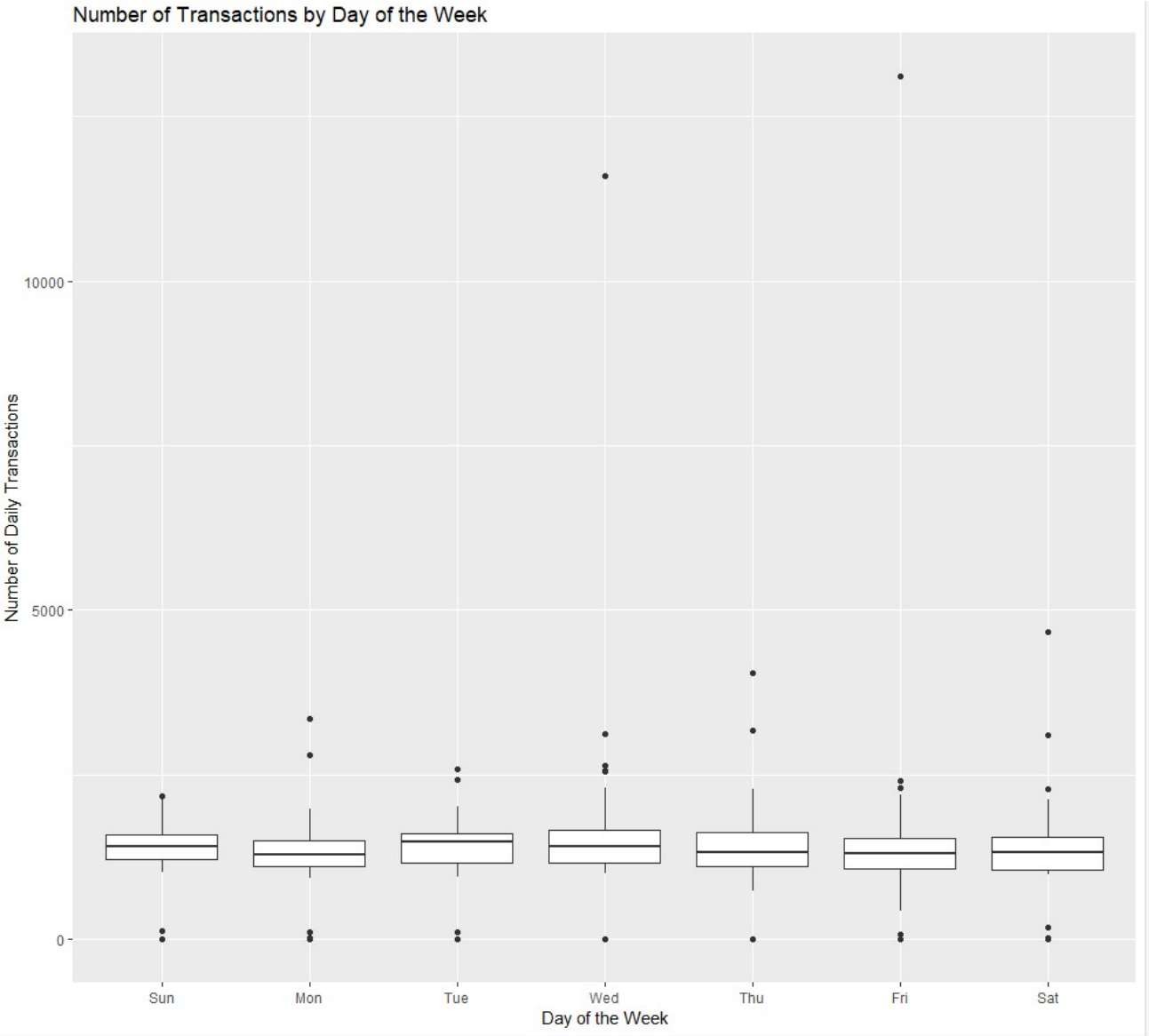
```
ggplot(capstonewithoutna_withinperiod_weekdayssummary, aes(x = daysofweek, y = revenue)) + geom_boxplot()
+ labs(x = 'Day of the Week', y = 'Revenue', title = 'Revenue by Day of the Week')
```



Revenue Per Day Of the Week

#Plot for Transaction per day of the week

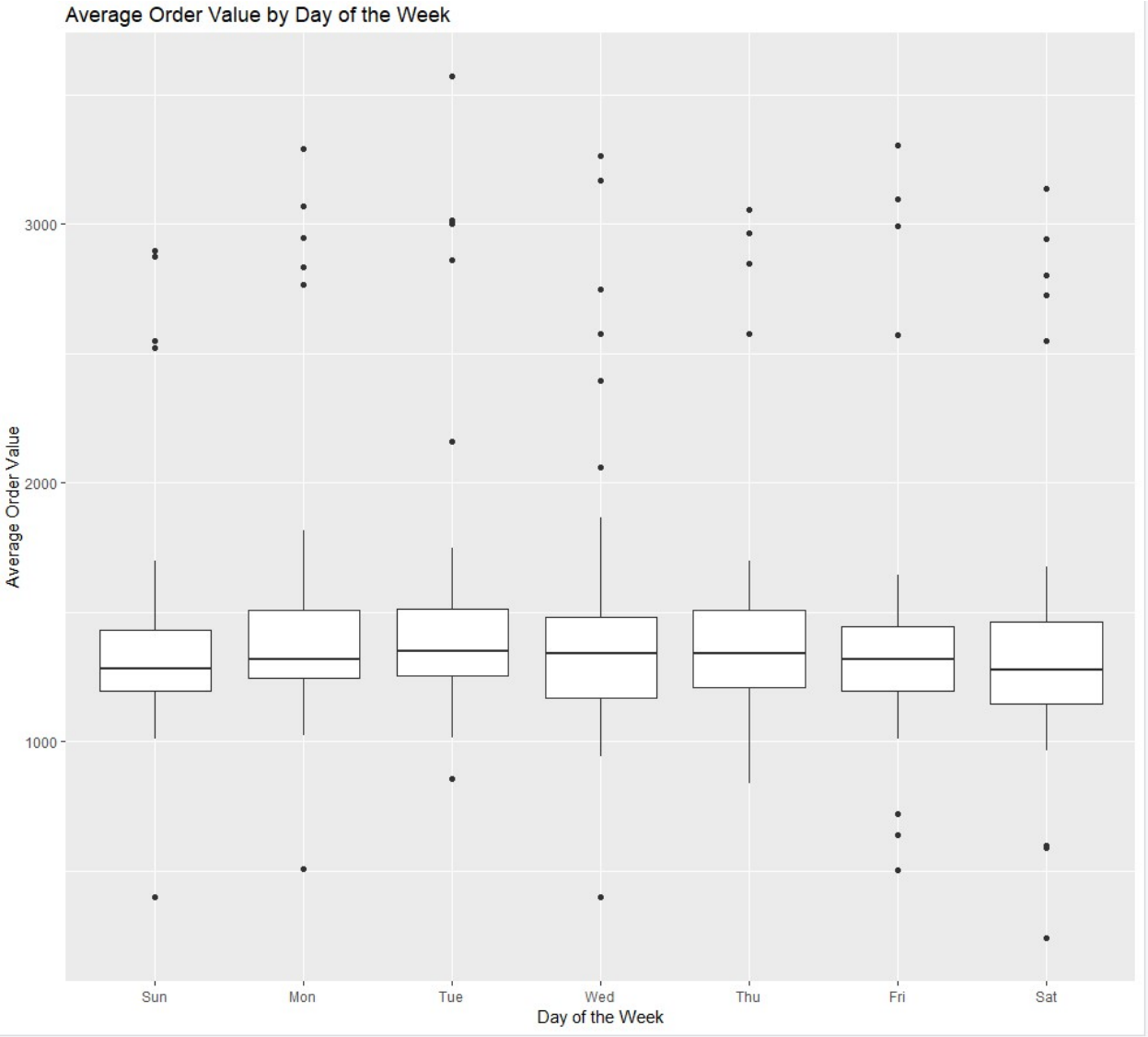
```
ggplot(capstonewithoutna_withinperiod_weekdayssummary, aes(x = daysofweek, y = transactions)) +
geom_boxplot() + labs(x = 'Day of the Week', y = 'Number of Daily Transactions', title = 'Number of Transactions by Day of the Week')
```



Transaction Per Day Of the Week

#Plot AvgOrderValue per day of the week

```
ggplot(capstonewithoutna_withinperiod_weekdayssummary, aes(x = daysofweek, y = AvgOrderValue)) +
geom_boxplot() + labs(x = 'Day of the Week', y = 'Average Order Value', title = 'Average Order Value by Day of the Week')
```

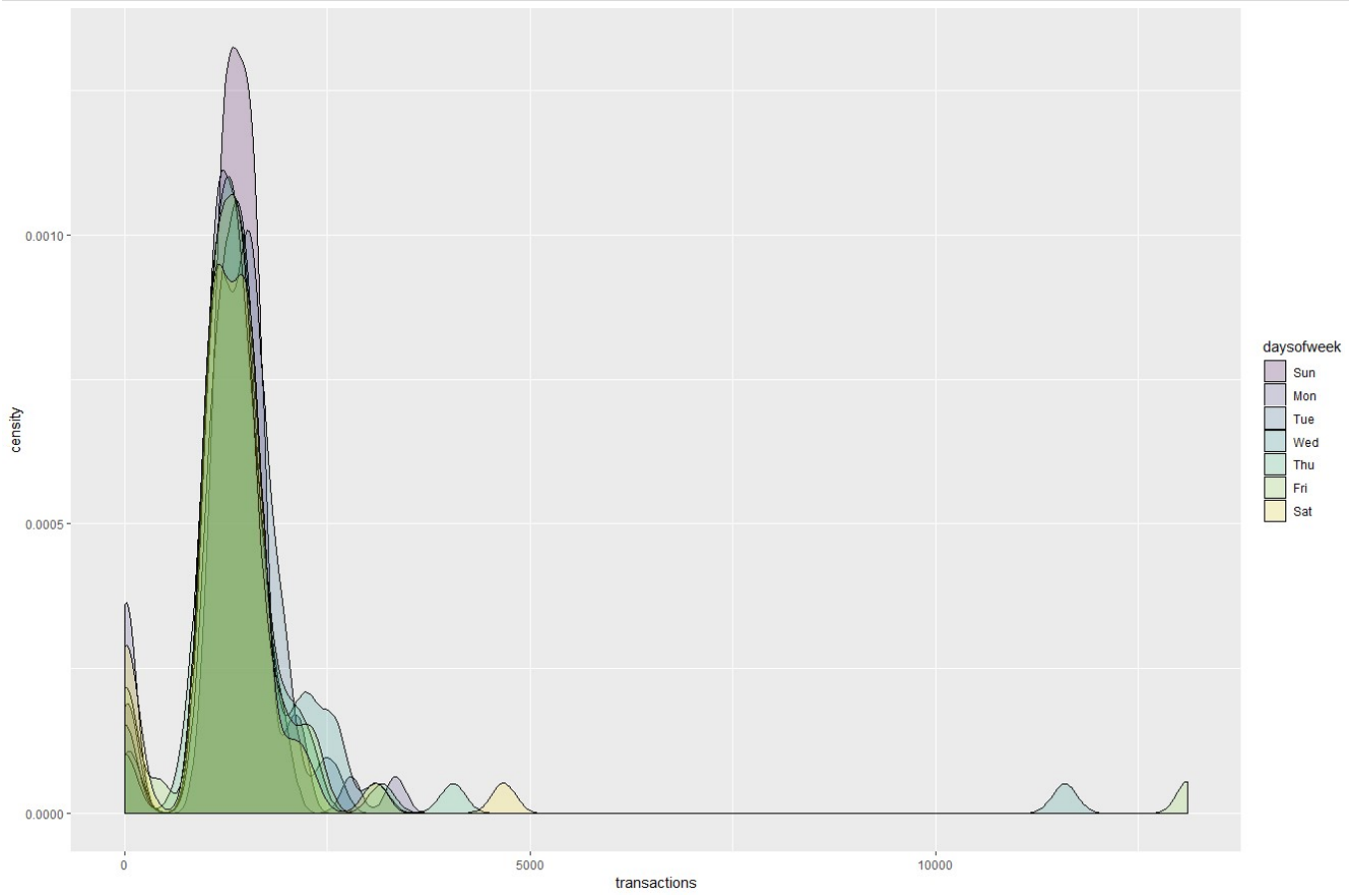


Average Order Value Per Day Of the Week

#To Understand the Differences in the amount of revenue on each day of the week, it is driven by a difference in the number of transactions, rather than the average order value.

#Plot "density plot" to see how the data are distributed.

```
ggplot(capstonewithoutna_withinperiod_weekdayssummary, aes(transactions, fill = daysofweek)) +
geom_density(alpha = 0.2)
```

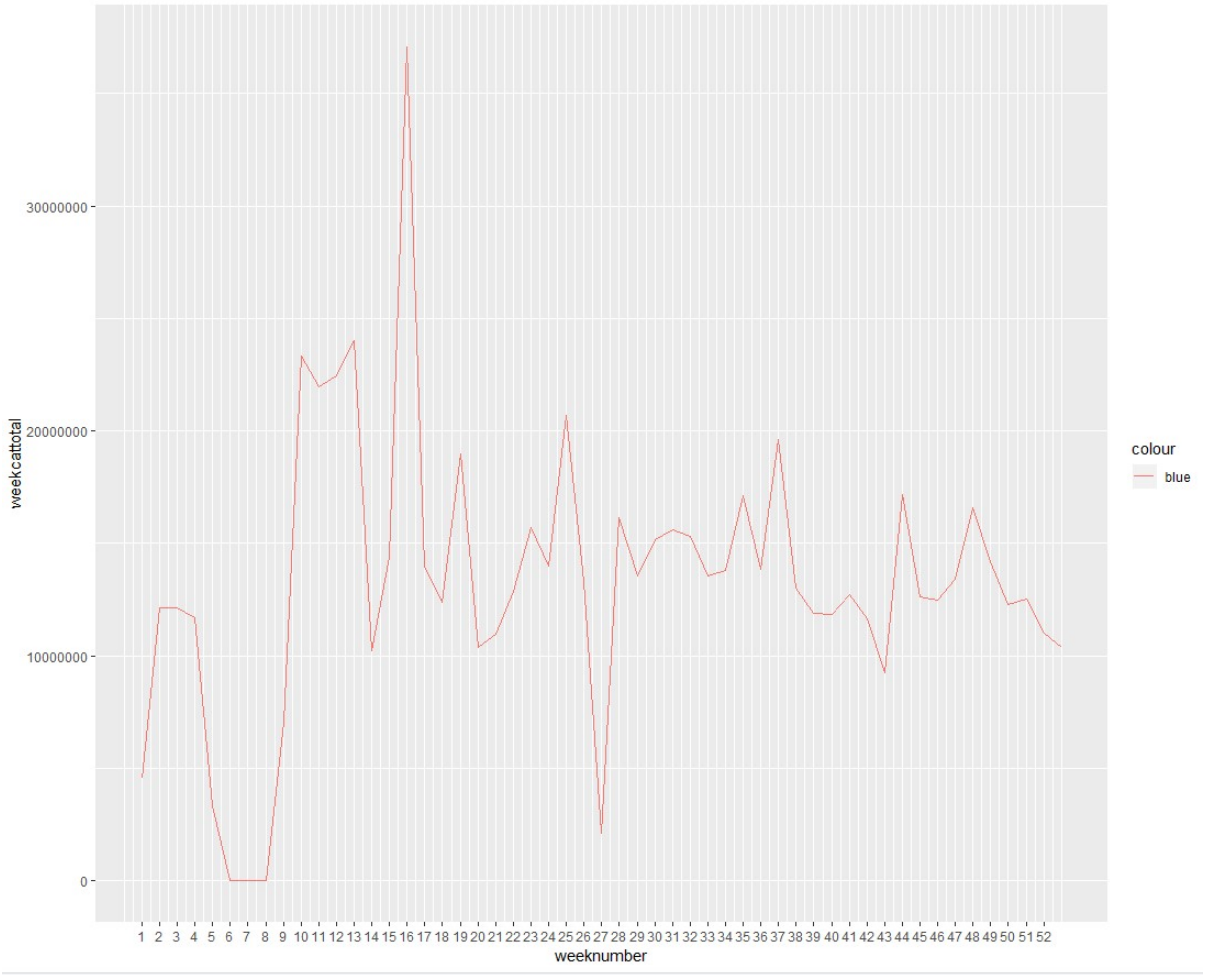


"Density Plot" to see how the data are distributed

capstonewithoutna_withinperiod_3subcategory <- filter (capstonewithoutna_withinperiod ,capstonewithoutna_withinperiod$product_analytic_sub_category %in% analysisCategory)

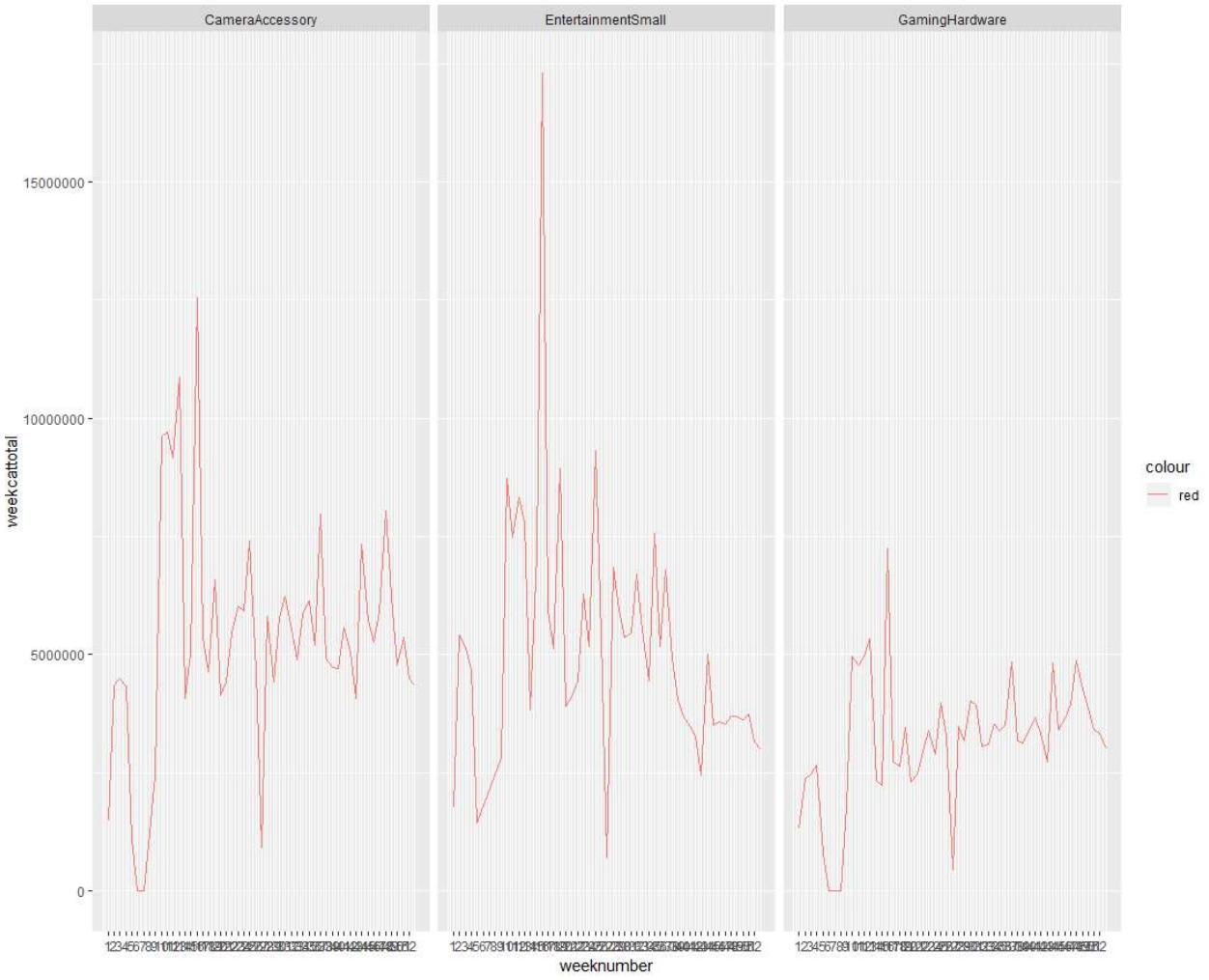
weeklygmv3categorytotals<-capstonewithoutna_withinperiod1b %>% dplyr::group_by(weeknumber,product_analytic_category)%>% dplyr::summarise(weekcattotal=sum(gmv, na.rm = TRUE))
weeklygmvtotals<-capstonewithoutna_withinperiod1b %>% group_by(weeknumber) %>% dplyr::summarise(weekcattotal=sum(gmv, na.rm = TRUE))

#Weekly GMV Total vs WeekNumber
ggplot(weeklygmvtotals,aes(x=weeknumber,y=weekcattotal,color="blue"))+geom_line()+scale_x_continuous(breaks=seq(1,52,1))



#Weekly GMV Total vs WeekNumber with Product Category

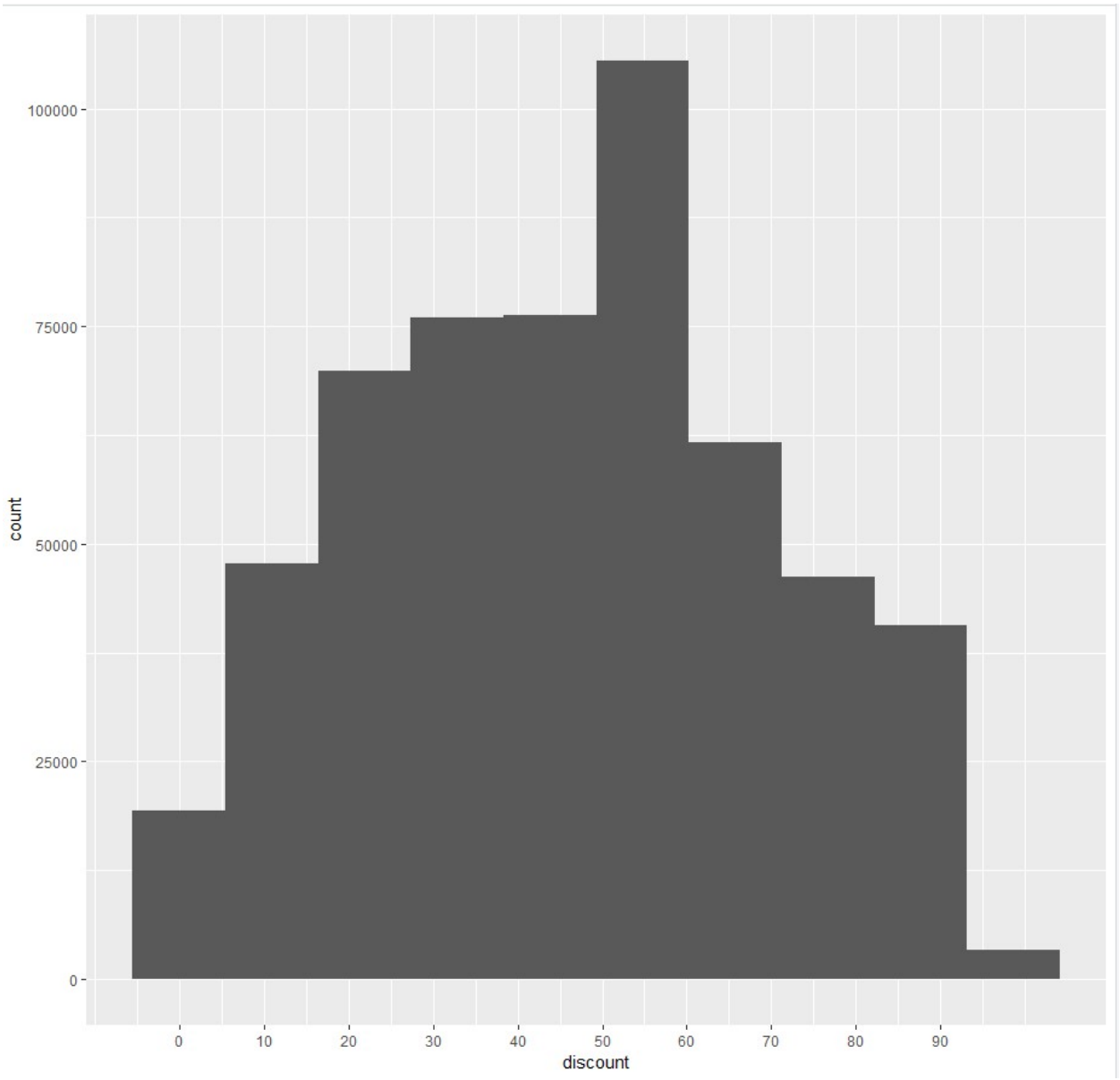
```
ggplot(weeklygmv3categorytotals,aes(x=weeknumber,y=weekcattotal,color="red"))+geom_line()+scale_x_continuous(breaks=seq(1,52,1)) +facet_grid(.~product_analytic_category)
```




```
capstonewithoutna_withinperiod1c<-capstonewithoutna_withinperiod1b
capstonewithoutna_withinperiod1c$discount<-round(capstonewithoutna_withinperiod1c$discount*100,2)
```

#Plot for Discount Vs Counts

```
ggplot(capstonewithoutna_withinperiod1c, aes(x=discount)) +
geom_histogram(bins=10)+scale_x_continuous(breaks=seq(0,99,10))
```



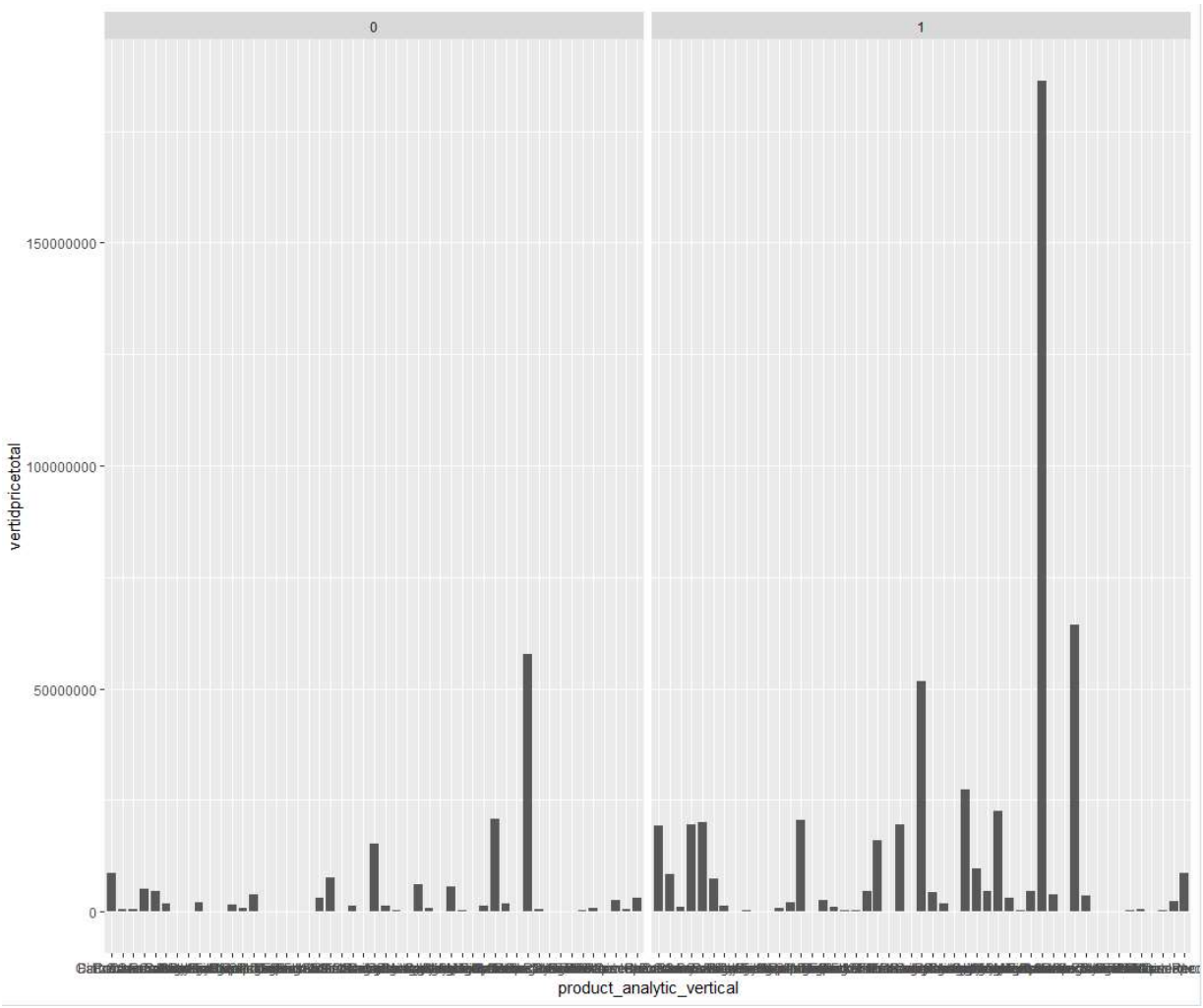

```
capstonewithoutna_withinperiod1c$dprice<-
ifelse(capstonewithoutna_withinperiod1c$product_mrp%%100==0,0,1)
verticaldecoratedprice<-
capstonewithoutna_withinperiod1c%>%dplyr::group_by(product_analytic_vertical,dprice)%>%dplyr::summarise(v
ertidpricetotal=sum(gmv, na.rm = TRUE))
```

#Plot for Decorated price from Order level data

```
verticaldecoratedprice%>%ggplot(aes(x=product_analytic_vertical,y=vertidpricetotal,fill=dprice)) +
geom_bar(stat='identity',width=0.8,position="dodge")
```

#Plot for Decorated price Vs NOT-Decorated price from Order level data

```
verticaldecoratedprice%>%ggplot(aes(x=product_analytic_vertical,y=vertidpricetotal)) +
geom_bar(stat='identity',width=0.8,position="dodge")+facet_grid(.~dprice)
```




```
categorydecoratedprice<-
capstonewithoutna_withinperiod1c%>%dplyr::group_by(product_analytic_category,dprice)%>%dplyr::summarise(
categorydpricetotal=sum(gmv, na.rm = TRUE))
categorydecoratedprice%>%ggplot(aes(x=product_analytic_category,y=categorydpricetotal)) +
geom_bar(stat='identity',width=0.8,position="dodge")+facet_grid(.~dprice)
```

#Plot for Decorated price from Order level data Vs Product Category

```
categorydecoratedprice%>%ggplot(aes(x=product_analytic_category,y=categorydpricetotal,color="blue")) +
geom_bar(stat='identity',width=0.8,position="dodge")+facet_grid(.~dprice)
```