

64036_Assignment_1

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   :  2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean   :15.4    Mean   : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```
summary(Online_Retail)
head(Online_Retail) summary(Online_Retail)
X1_test <- Online_Retail$Country summary(X1_test) head(Online_Retail) spec(Online_Retail) sum-
mary(Online_Retail)
colSums(is.na(Online_Retail))
table(grepl("^C", Online_Retail$InvoiceNo))
table(Online_Retail$Country) revenue <- summarise (sales=sum(Online_Retail$Quantity*
Online_Retail$UnitPrice))
revenue <- Online_Retail %>% group_by(Online_Retail$Country) %>% summarise (Country_Count
= n()) %>% mutate(country_percentage = (Country_Count/sum(Country_Count))*100) %>% fil-
ter(country_percentage>1)
str(revenue) str(Online_Retail) revenue %>% select(sales) precoftotal <- revenuesales * sum(revenuesales)
head(precoftotal) Percents <- Online_Retail %>% c(Online_Retail$Country, Online_Retail$Quantity,
Online_Retail$UnitPrice) head(Percents) summary(Percents) Percent_1 <- c(Online_Retail$Country,
Online_Retail$Quantity, Online_Retail$UnitPrice) summary(Percents_1)
percent_2 <- Online_Retail %>% select(Online_Retail$Country, Online_Retail$Quantity, Online_Retail$UnitPrice)
revenue["percents"] <- revenuesales / sum(sales) summary(revenue) sum(revenue$sales)
summary(Online_Retail)
```

```

revenue <- Online_Retail %>% group_by(Online_RetailCountry)table(Online_RetailCountry)

Online_Retail <- Online_Retail %>% mutate(TransactionValue = Quantity * UnitPrice)

Online_Retail %>% group_by(Country) %>% summarise (sum_TransactionValue = sum(TransactionValue))
%>% filter (sum_TransactionValue > 130000)

Online_Retail %>% filter(Country == "Germany")

Germany <- Online_Retail %>% filter(Country == "Germany") hist(Germany$TransactionValue,main =
paste("Histogram of Germany Transactions"))

plot(Germany)

Online_Retail [!is.na(Online_Retail$CustomerID),] %>% group_by(CustomerID) %>% summarise (Cus-
tomer_Count = n(),sum_total= sum(TransactionValue)) %>% arrange (desc(Customer_Count))

Online_Retail [!is.na(Online_Retail$CustomerID),] %>% group_by(CustomerID) %>% summarise (Cus-
tomer_Count = n(),sum_total= sum(TransactionValue)) %>% arrange (desc(sum_total))

colMeans(is.na(Online_Retail)*100)

Online_Retail %>% group_by(Country) %>% summarise(CustomerID_Missing = sum(is.na(CustomerID)))
%>% filter(CustomerID_Missing>0)

Online_Retail_Cancelled <- Online_Retail %>% filter(Country=="France",Quantity<0) %>% count()

Online_Retail_Total <- Online_Retail %>% filter(Country == "France") %>% count()

(Online_Retail_Cancelledn/Online_Retail_Totaln)*100

(149/8557)*100

Online_Retail %>% group_by(StockCode) %>% summarise(sum_transactionvalue = sum(TransactionValue))
%>% arrange(desc(sum_transactionvalue))

length(unique(Online_Retail$CustomerID))

Temp=strptime(Online_Retail$InvoiceDate,format='%m/%d/%Y %H:%M',tz='GMT')

Online_Retail$New_Invoice_Date <- as.Date(Temp)

Online_RetailInvoice_DayWeek = weekdays(Online_RetailNew_Invoice_Date)

Online_Retail$New_Invoice_Hour = as.numeric(format(Temp, "%H"))

Online_Retail$New_Invoice_Month = as.numeric(format(Temp, "%m"))

Online_Retail %>% group_by(Invoice_Day_Week) %>% summarise(number_of_transactions=n()) %>%
mutate(trans_percent = (number_of_transactions/sum(number_of_transactions))*100)

```