



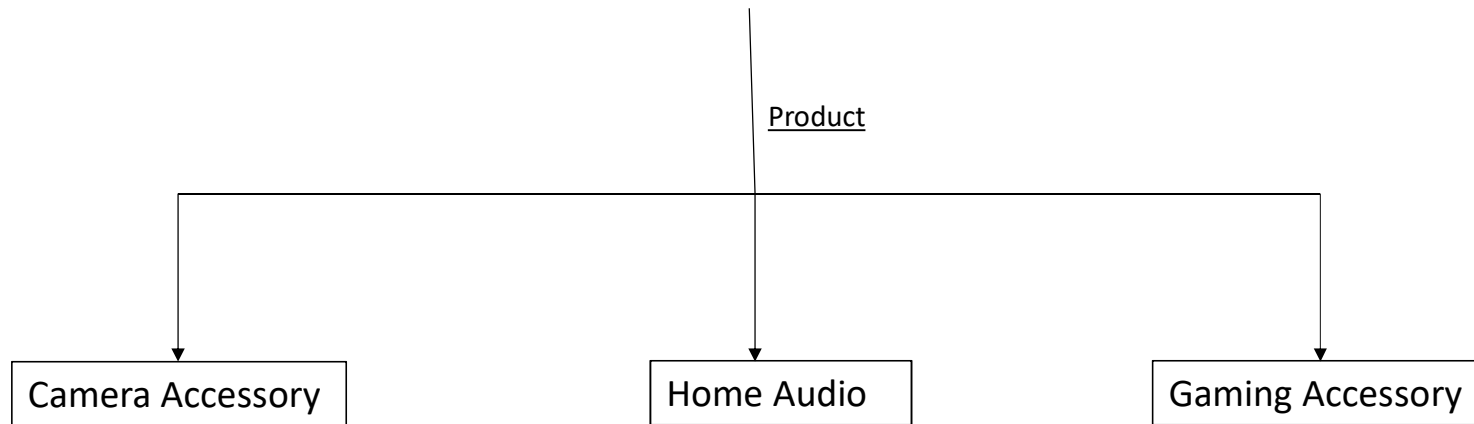
CAPSTONE PROJECT STUDY

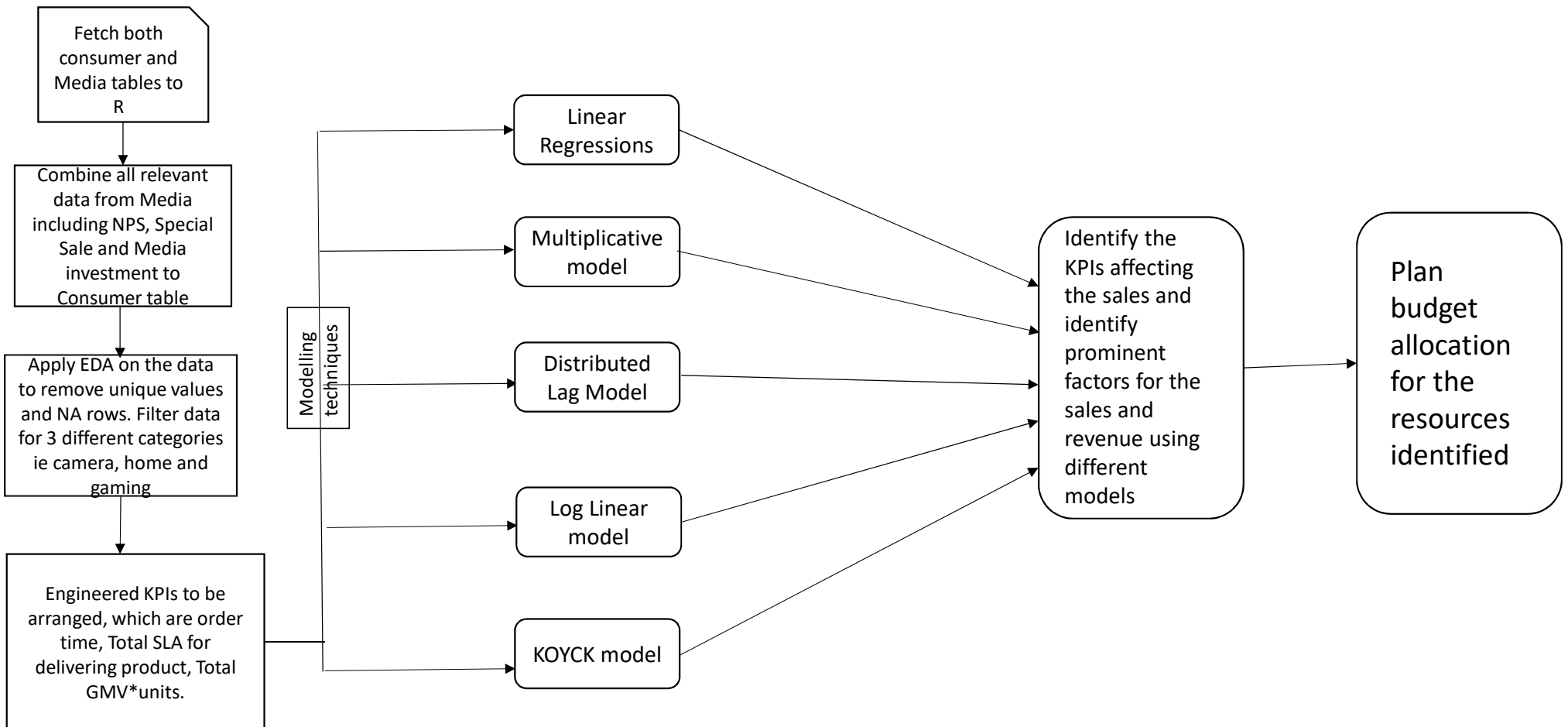
Group Name:

1. Vivek Sachdeva
2. Vikash Sharma
3. Nilaksh Bajpai

Goals

Optimal Budget allocation for various marketing levers based on last 1 year impact







Basic Linear Model

```
install.packages("readxl")
library("readxl")
library(car)
library(MASS)
library(tidyr)

consumerelectro <- read.csv("ConsumerElectronics.csv",header = T,na.strings = c("", " ", "NA"),stringsAsFactors = FALSE)
product_list <- read_excel("Media data and other information.xlsx",sheet = 1,col_names = TRUE)
media_investment <- read_excel("Media data and other information.xlsx",sheet = 2,col_names = TRUE)
Special_sale <- read_excel("Media data and other information.xlsx",sheet = 3,col_names = TRUE)
NPS <- read_excel("Media data and other information.xlsx",sheet = 4,col_names = TRUE)

## Media headers correction

names(media_investment) <- media_investment[1,]
media_investment <- media_investment[-c(1),]
```



Basic Linear Model – Contd.

```
## Changing NPS column name
```

```
names(NPS)[names(NPS) == "July'15"] <- "7"  
names(NPS)[names(NPS) == "Aug'15"] <- "8"  
names(NPS)[names(NPS) == "Sept'15"] <- "9"  
names(NPS)[names(NPS) == "Oct'15"] <- "10"  
names(NPS)[names(NPS) == "Nov'15"] <- "11"  
names(NPS)[names(NPS) == "Dec'15"] <- "12"  
names(NPS)[names(NPS) == "Jan'16"] <- "1"  
names(NPS)[names(NPS) == "Feb'16"] <- "2"  
names(NPS)[names(NPS) == "Mar'16"] <- "3"  
names(NPS)[names(NPS) == "Apr'16"] <- "4"  
names(NPS)[names(NPS) == "May'16"] <- "5"  
names(NPS)[names(NPS) == "June'16"] <- "6"
```

```
## Unique data in consumerelectr and other EDA
```

```
consumerelectro <- unique(consumerelectro)
```



Basic Linear Model – Contd.

```
## removing negative customer id
consumerelectro$cust_id[consumerelectro$cust_id <= 0] <- NA

## Checking and removing all NA entries
sum(is.na(consumerelectro)) ## 531062 entries

row_has_na <- apply(consumerelectro,1,function(x){any(is.na(x))})
sum(row_has_na)
consumerelectro_final <- consumerelectro[!row_has_na,]

sum(is.na(consumerelectro_final)) ## is 0

consumerelectro_final$order_date <- as.POSIXlt(consumerelectro_final$order_date,format="%m/%d/%Y %H:%M")
consumerelectro_final$order_time <- format(consumerelectro_final$order_date,"%H:%M")
```



Basic Linear Model – Contd.

```
for (i in 1:nrow(consumerelectro_final))  
{  
  consumerelectro_final[i,22] <- NPS[,colnames(NPS)==consumerelectro_final[i,4]]  
  consumerelectro_final[i,23] <- media_investment[media_investment$Month==consumerelectro_final[i,4],3]  
  consumerelectro_final[i,24] <- media_investment[media_investment$Month==consumerelectro_final[i,4],4]  
  consumerelectro_final[i,25] <- media_investment[media_investment$Month==consumerelectro_final[i,4],5]  
  consumerelectro_final[i,26] <- media_investment[media_investment$Month==consumerelectro_final[i,4],6]  
  consumerelectro_final[i,27] <- media_investment[media_investment$Month==consumerelectro_final[i,4],7]  
  consumerelectro_final[i,28] <- media_investment[media_investment$Month==consumerelectro_final[i,4],8]  
  consumerelectro_final[i,29] <- media_investment[media_investment$Month==consumerelectro_final[i,4],9]  
  consumerelectro_final[i,30] <- media_investment[media_investment$Month==consumerelectro_final[i,4],10]  
  consumerelectro_final[i,31] <- media_investment[media_investment$Month==consumerelectro_final[i,4],11]  
  consumerelectro_final[i,32] <- media_investment[media_investment$Month==consumerelectro_final[i,4],12]  
}
```



Basic Linear Model – Contd.

```
# naming column as per Naming
names(consumerelectro_final)[22] <- "NPS"
names(consumerelectro_final)[23] <- "Total_Investment"
names(consumerelectro_final)[24] <- "TV"
names(consumerelectro_final)[25] <- "Digital"
names(consumerelectro_final)[26] <- "Sponsorship"
names(consumerelectro_final)[27] <- "Content_Marketing"
names(consumerelectro_final)[28] <- "Online_Marketing"
names(consumerelectro_final)[29] <- "Affiliates"
names(consumerelectro_final)[30] <- "SEM"
names(consumerelectro_final)[31] <- "Radio"
names(consumerelectro_final)[32] <- "Other_Investment"
```




Basic Linear Model – Contd.

```
# For weekly effect dividing NPS by 4
consumerelectro_final$NPS <- (consumerelectro_final$NPS/4)
consumerelectro_final$Total_Sla <- (consumerelectro_final$sla + consumerelectro_final$product_procurement_sla)

# convert factors with 2 levels to numerical variables such as s1_fact.order_payment_type
levels(consumerelectro_final$s1_fact.order_payment_type) <- c(1,0)
consumerelectro_final$s1_fact.order_payment_type <-
as.numeric(levels(consumerelectro_final$s1_fact.order_payment_type))[consumerelectro_final$s1_fact.order_payment_type]

library(lubridate)
consumerelectro_final$week <- week(consumerelectro_final$order_date)

## Engineer KPIs
consumerelectro_final$sales <- (consumerelectro_final$gmrv * consumerelectro_final$units)
```



Basic Linear Model – Contd.

```
## removing few unwanted columns
consumerelectro_final <- consumerelectro_final[-c(1,5:6,9:11,14)]

## Filter by 3 categories to built the mode
camera_accessory <- subset(consumerelectro_final,consumerelectro_final$product_analytic_category=="CameraAccessory"
                           | consumerelectro_final$product_analytic_category=="Camera")

home_audio <- subset(consumerelectro_final,consumerelectro_final$product_analytic_category=="EntertainmentSmall")

gaming_accessory <- subset(consumerelectro_final,consumerelectro_final$product_analytic_category=="GameCDDVD"
                           | consumerelectro_final$product_analytic_category=="GamingHardware")

# set separate training and test data

set.seed(100)
```



Basic Linear Model – Contd.

```
camera_trainindices = sample(1:nrow(camera_accessory),0.7*nrow(camera_accessory))  
Camera_training = camera_accessory[camera_trainindices,]  
Camera_test = camera_accessory[-camera_trainindices,]
```

```
home_trainindices = sample(1:nrow(home_audio),0.7*nrow(home_audio))  
home_training = home_audio[home_trainindices,]  
home_test = home_audio[-home_trainindices,]
```

```
gaming_trainindices = sample(1:nrow(gaming_accessory),0.7*nrow(gaming_trainindices))  
gaming_training = gaming_accessory[gaming_trainindices,]  
gaming_test = gaming_accessory[-gaming_trainindices,]
```



Basic Linear Model – Contd.

```
## checking models for home, camera and gaming
```

```
camera_model_1 <- lm(formula = gmv ~ order_date + units + sla + product_mrp + product_procurement_sla +  
  order_time + NPS + Total_Sla + week + sales + Total_Investment + TV + Digital + Sponsorship  
  + Content_Marketing + Online_Marketing + Affiliates + SEM ,data = Camera_training)
```

```
home_model_1 <- lm(formula = gmv ~ order_date + units + sla + product_mrp + product_procurement_sla +  
  order_time + NPS + Total_Sla + week + sales + Total_Investment + TV + Digital + Sponsorship  
  + Content_Marketing + Online_Marketing + Affiliates + SEM ,data = home_training)
```

```
gaming_model_1 <- lm(formula = gmv ~ order_date + units + sla + product_mrp + product_procurement_sla +  
  order_time + NPS + Total_Sla + week + sales + Total_Investment + TV + Digital + Sponsorship  
  + Content_Marketing + Online_Marketing + Affiliates + SEM ,data = gaming_training)
```

```
summary(camera_model_1)
```

```
summary(home_model_1)
```

```
summary(gaming_model_1)
```

```
vif(camera_model_1)
```

```
vif(home_model_1)
```

```
vif(gaming_model_1)
```