

ACTIVIDAD 7

Equipo 1_Samsung

27 de septiembre de 2021

```
library(ggplot2)
library(tidyverse)

## — Attaching packages — tidyverse
1.3.1 —

## ✓ tibble 3.1.3    ✓ dplyr 1.0.7
## ✓ tidyr 1.1.3    ✓ stringr 1.4.0
## ✓ readr 2.0.1    ✓ forcats 0.5.1
## ✓ purrr 0.3.4

## — Conflicts —
tidyverse_conflicts() —
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()

d2<- read.csv("~/Desktop/TEC IIS/8vo/Diseño y optimizacion de
operaciones/proyecto Samsung/equipo_1_samsung_datos_limpios (1).csv")
str(d2)

## 'data.frame': 148575 obs. of 14 variables:
## $ punto_de_venta: chr "1 poniente" "1 poniente" "1 poniente" "1
poniente" ...
## $ fecha : chr "30/06/2018" "30/06/2018" "24/07/2018"
"01/08/2018" ...
## $ mes : int 6 6 7 8 8 8 8 8 8 8 ...
## $ anio : int 2018 2018 2018 2018 2018 2018 2018 2018 2018 2018
...
## $ num_ventas : int 1 1 1 1 1 1 1 1 1 1 ...
## $ sku : chr "N.SJ7PROD" "N.SJ7PRON" "N.SAMGA6PLA" "N.SAMGJ4NG"
...
## $ marca : chr "samsung" "samsung" "samsung" "samsung" ...
## $ gamma : chr "baja" "baja" "media" "baja" ...
## $ costo_promedio: num 4184 4196 5109 1815 2505 ...
## $ zona : chr "centro sur" "centro sur" "centro sur" "centro
sur" ...
## $ estado : chr "puebla" "puebla" "puebla" "puebla" ...
## $ ciudad : chr "tehuacan" "tehuacan" "tehuacan" "tehuacan" ...
## $ latitud : num 18.5 18.5 18.5 18.5 18.5 ...
## $ longitud : num -97.4 -97.4 -97.4 -97.4 -97.4 ...
```

De las variables enlistadas en el cuadro anterior el punto de venta hace referencia al nombre de la sucursal en donde se realizo la venta del dispositivo. Se enlistan otros datos complementarios que permiten saber el lugar y la fecha de transaccion, asi como el modelo con la variable SKU, y la gamma en la cual se cataloga el dispositivo. La precision del lugar en el que se realizo la compra la podemos detallar con las variables de Zona, Estado, Ciudad y mas especificamente con latitud y longitud.

```
#contador de 0
```

```
#año 2019
```

```
contador_0<-  
d2%>%filter(anio=="2019")%>%filter(costo_promedio==0)%>%group_by(gamma)  
count(contador_0,costo_promedio)
```

```
## # A tibble: 1 × 3  
## # Groups:   gamma [1]  
##   gamma costo_promedio      n  
##   <chr>          <dbl> <int>  
## 1 baja              0  1759
```

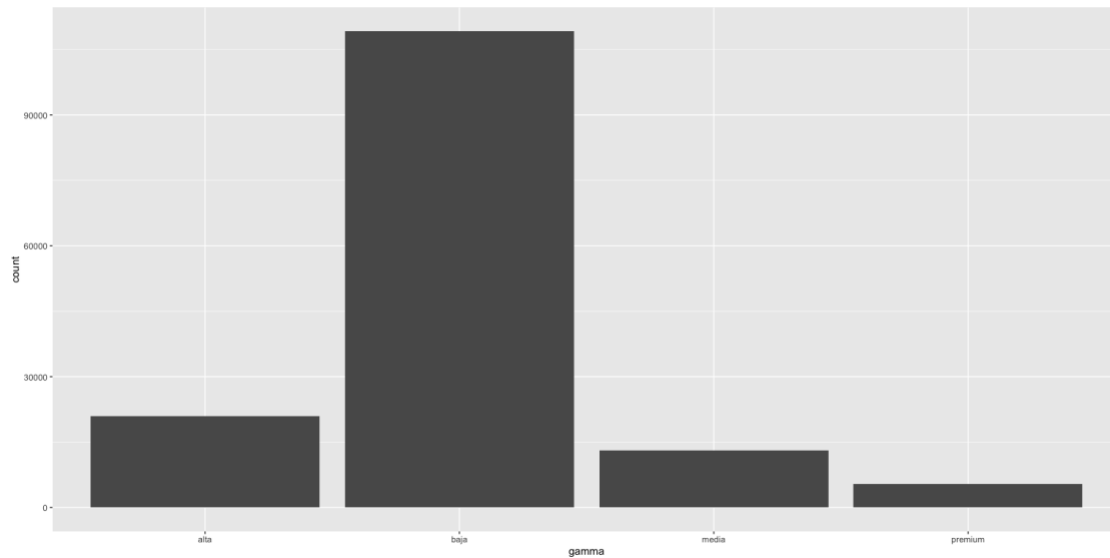
Contamos con un total de 1759 registros con 0 en la gamma baja en la columna de costo promedio esto en el año 2019 lo cual es mayor comparado con el total del 2018 que fue de 113 registros.

```
#contador de 0 2018
```

```
contador2_0<-  
d2%>%filter(anio=="2018")%>%filter(costo_promedio==0)%>%group_by(gamma)  
count(contador2_0,costo_promedio)
```

```
## # A tibble: 1 × 3  
## # Groups:   gamma [1]  
##   gamma costo_promedio      n  
##   <chr>          <dbl> <int>  
## 1 baja              0   113
```

```
ggplot(d2, aes (x=gamma))+  
  geom_bar()
```

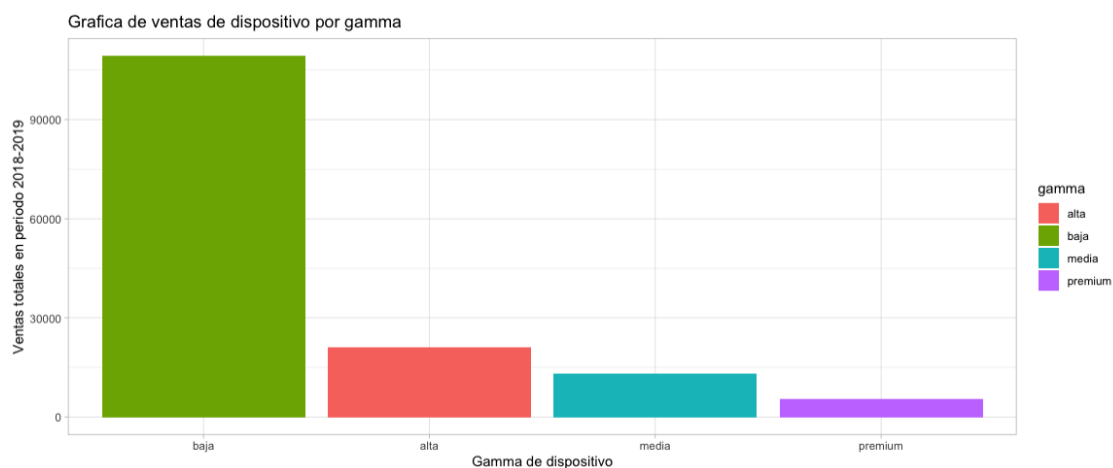


```
count(d2,gamma)
```

```
##      gamma      n
## 1    alta  20980
## 2    baja 109201
## 3    media  13032
## 4 premium   5362
```

La gamma baja es la que tiene un mayor numero de registros

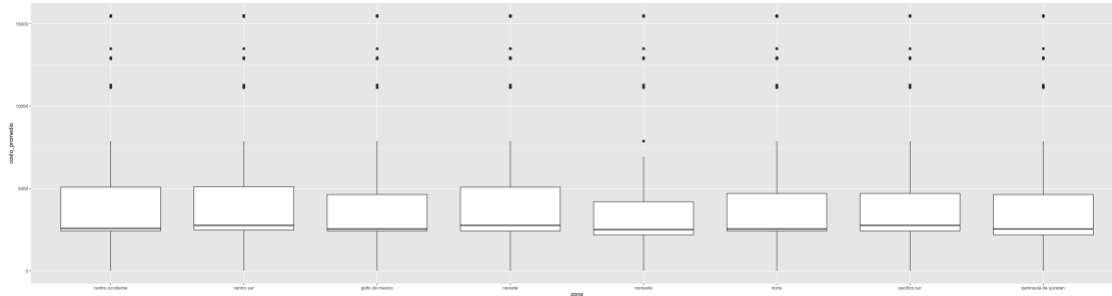
```
ggplot(d2, aes(x=forcats:: fct_infreq(gamma), fill = gamma))+
  geom_bar()+
  theme_light()+
  xlab("Gamma de dispositivo")+
  ylab("Ventas totales en periodo 2018-2019")+
  labs(title= "Grafica de ventas de dispositivo por gamma")
```



La grafica muestra de en los años 2018 y 2019 la gamma de que registro más ventas fue la baja con un margen considerablemente alto frente a la gama alta que fue la segunda enlistada en preferencias, mientras que la gamma media en tercer lugar no mantiene gran

margen frente a la cuarta posición de preferencias en la que se ubican los dispositivos premium.

```
ggplot(d2,aes(x=zona, y=costo_promedio))+  
  geom_boxplot()
```



Gracias

a estas graficas de cajas podemos observar que el comportamiento de las zonas cuenta con datos anormales en cada una de ellas, es decir no se mantiene un comportamiento normal de la preferencia de los equipos.

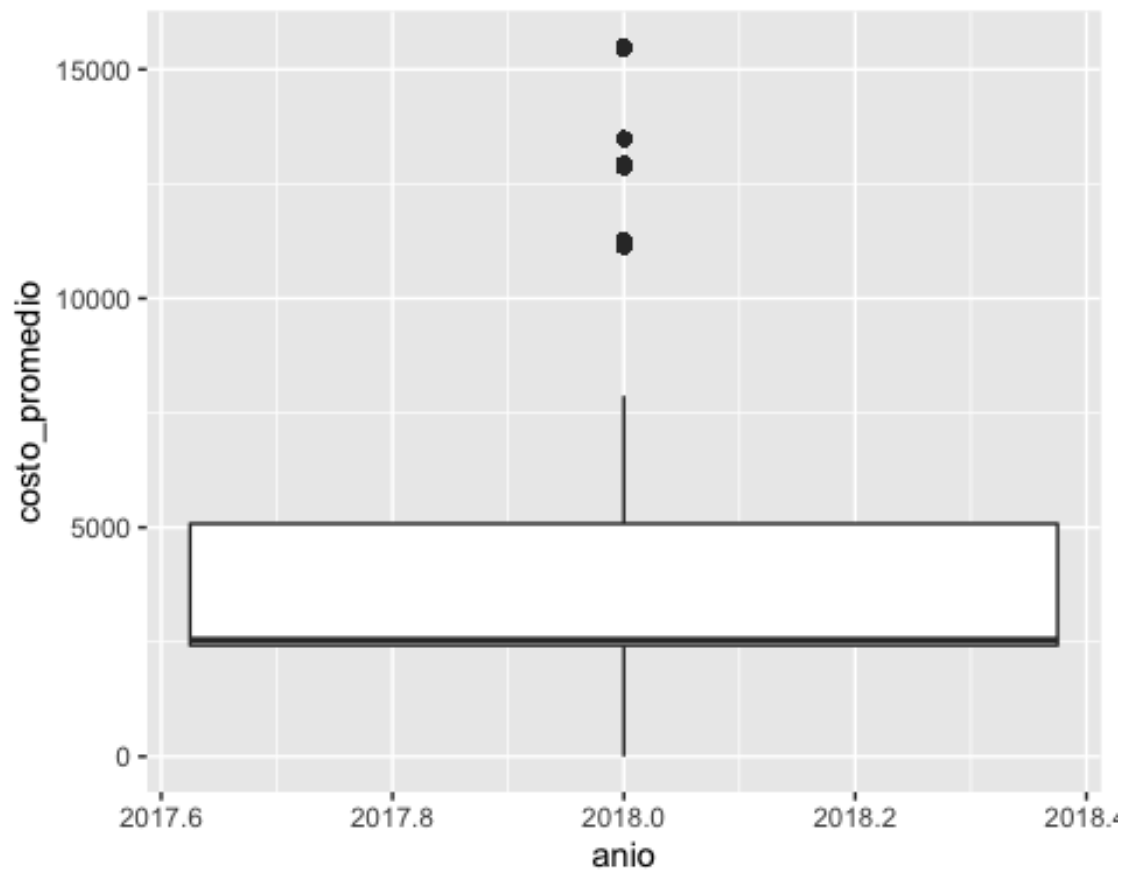
```
count(d2,num_ventas)
```

```
##   num_ventas      n  
## 1           1 148575
```

Numero de ventas totales

```
d3<-d2%>%group_by((anio))
```

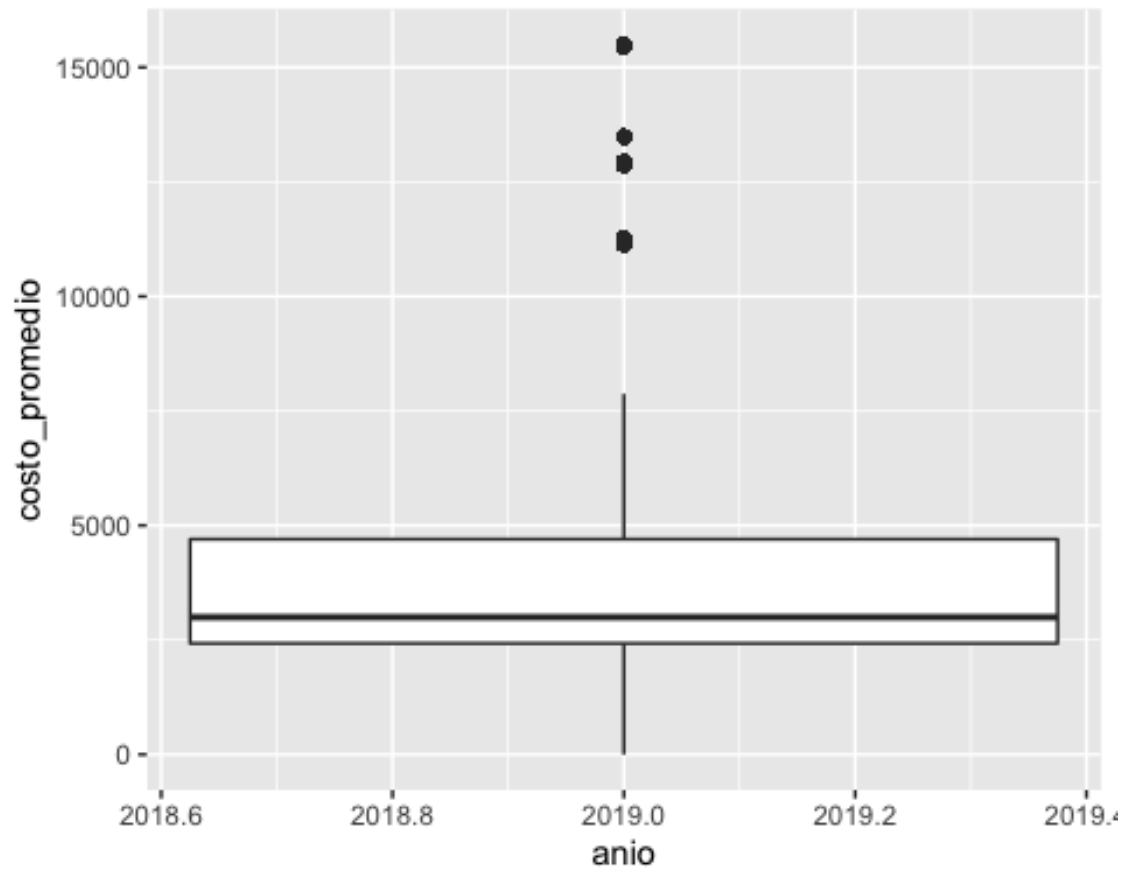
```
d3<-d3%>%filter(anio=="2018")  
ggplot(d3, aes (x=anio,y=costo_promedio))+  
  geom_boxplot()
```



En este diagrama de caja se observan datos atipicos en el año 2018 que estan en un rango de entre 10,000 y 15,000 del costo promedio. La mayoría de los datos esta en 5000, referente al costo de los dispositivos moviles.

```
d4<-d2%>%group_by((fecha))

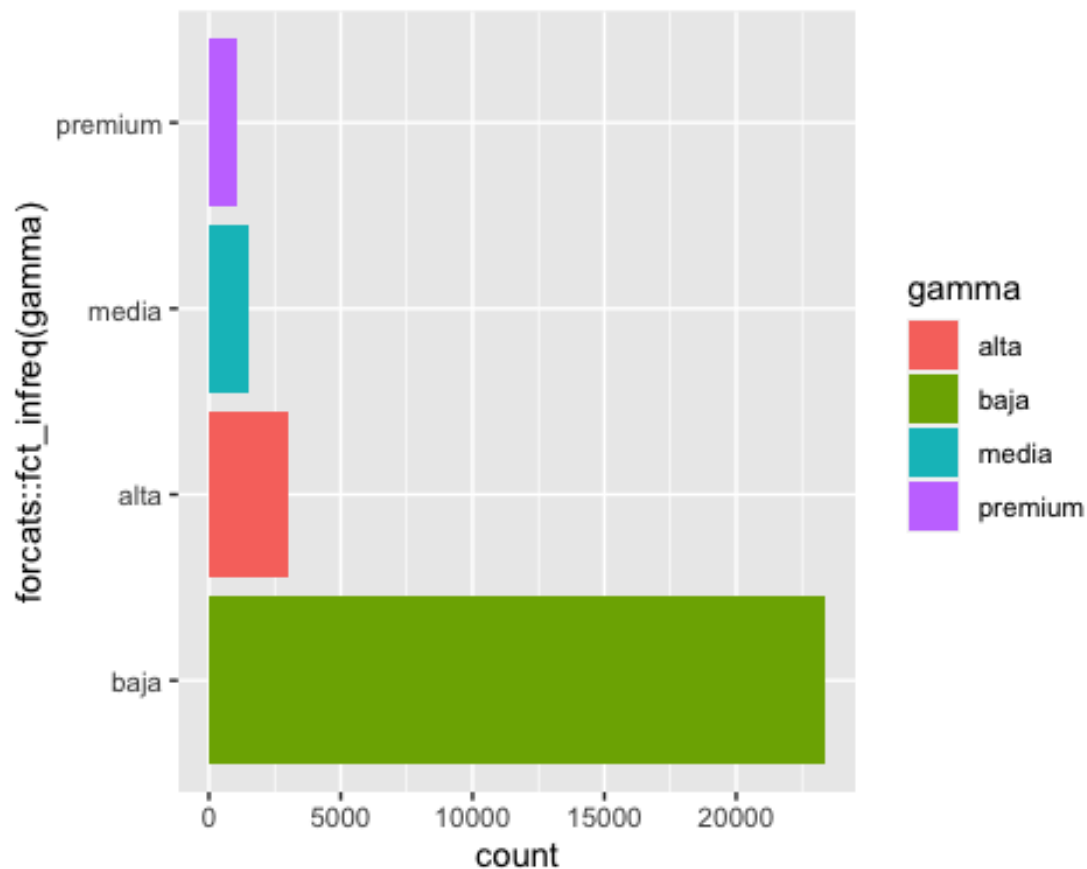
d4<-d4%>%filter(anio=="2019")
ggplot(d4, aes (x=anio,y=costo_promedio))+
  geom_boxplot()
```



En este diagrama de caja se observan datos atipicos en el año 2019 que estan en un rango de entre 11,000 y 15,000 del costo promedio. La mayoría de los datos esta en 5000, referente al costo d elos dispositivos moviles.

#anio 2019

```
ggplot(d4, aes (y=forcats:: fct_infreq(gamma), fill = gamma))+
  geom_bar()
```

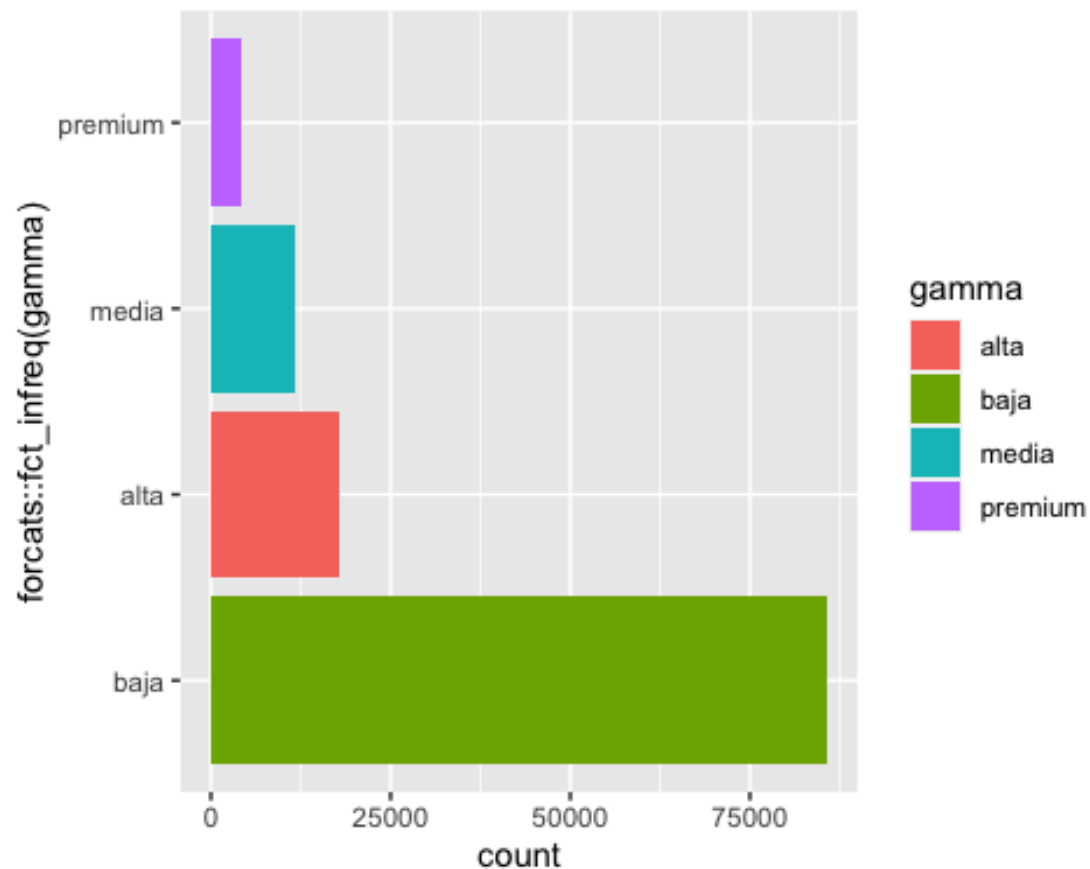


En esta grafica de barras podemos apreciar el analisis de la demanda que se proyecta para los dispositivos de cada gamma durante el año 2019, en donde podemos apreciar que aproximadamente un 80% de la demanda de los dispositivos es de gamma baja, mientras que el resto se reparte en la gamma alta en segundo lugar, seguidos por la gamma media y la menos demandada es la premium.

```
t<-d4%>%group_by((gamma))
count(t,gamma)

## # A tibble: 4 × 3
## # Groups:   (gamma) [4]
##   `(gamma)` gamma      n
##   <chr>    <chr> <int>
## 1 alta      alta    3017
## 2 baja      baja   23391
## 3 media     media    1495
## 4 premium  premium  1073

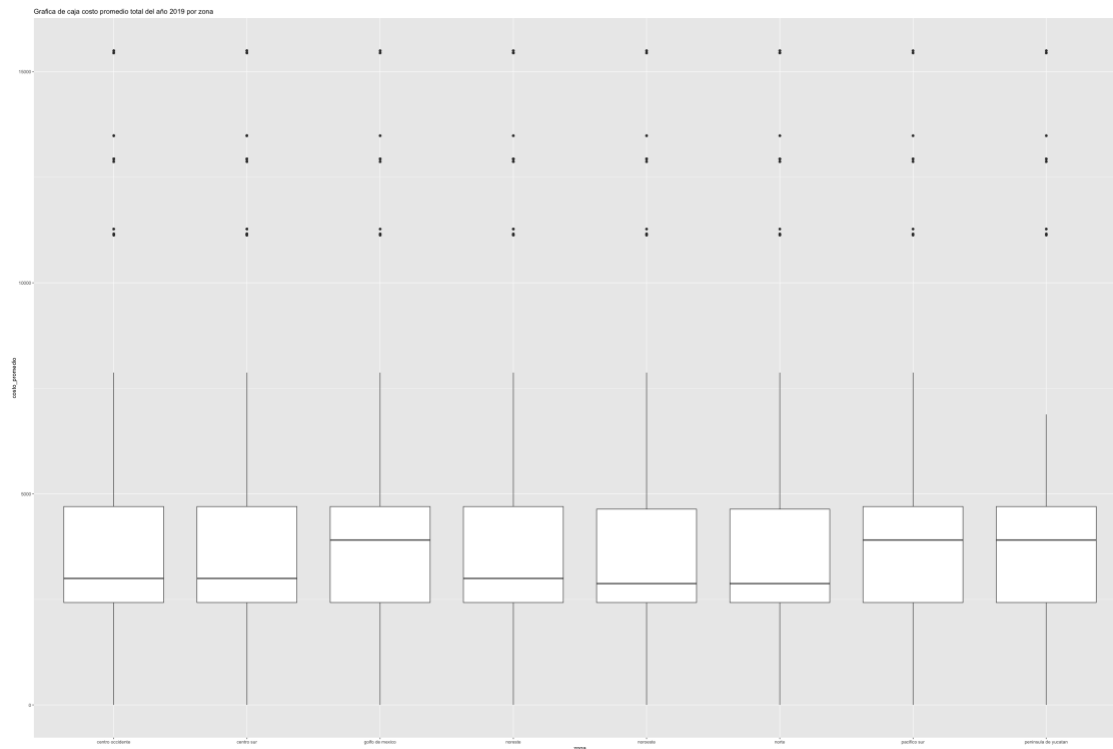
ggplot(d3, aes (y=forcats:: fct_infreq(gamma), fill = gamma))+
  geom_bar()
```



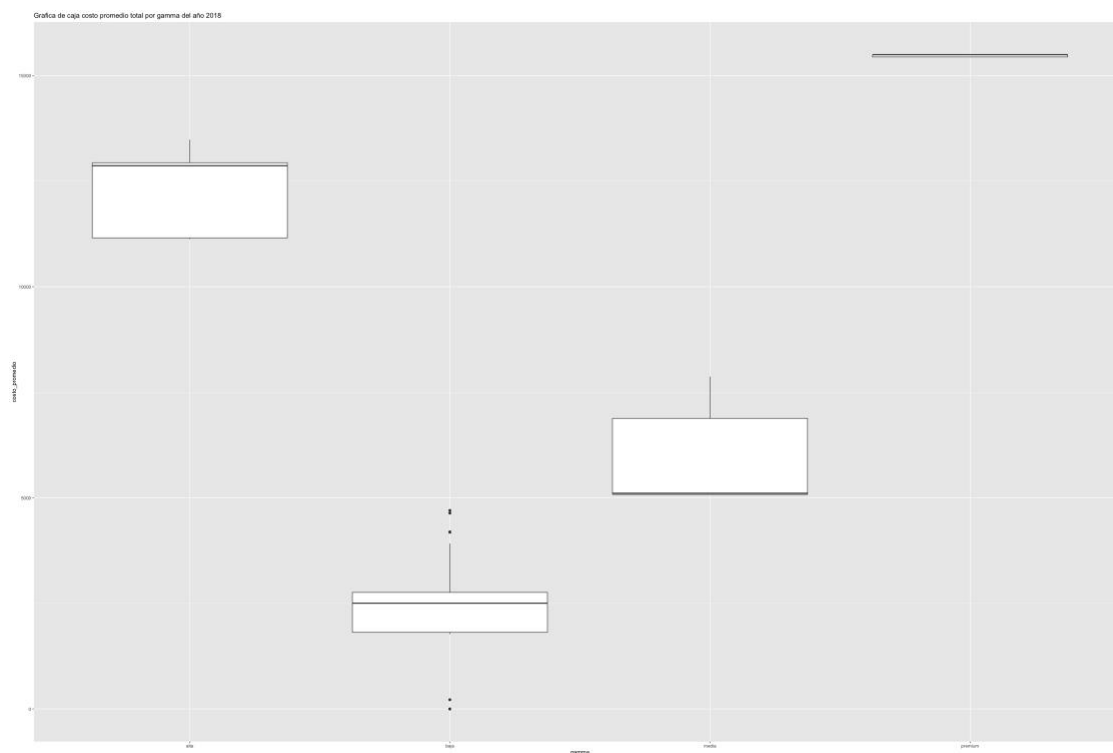
```
t2<-d3%>%group_by((gamma))
count(t2)

## # A tibble: 4 × 2
## # Groups:   (gamma) [4]
##   `(gamma)`      n
##   <chr>      <int>
## 1 alta       17963
## 2 baja       85810
## 3 media      11537
## 4 premium    4289

ggplot(d4,aes(x=zona, y=costo_promedio))+
  geom_boxplot()+
  labs(title= "Grafica de caja costo promedio total del año 2019 por zona")
```

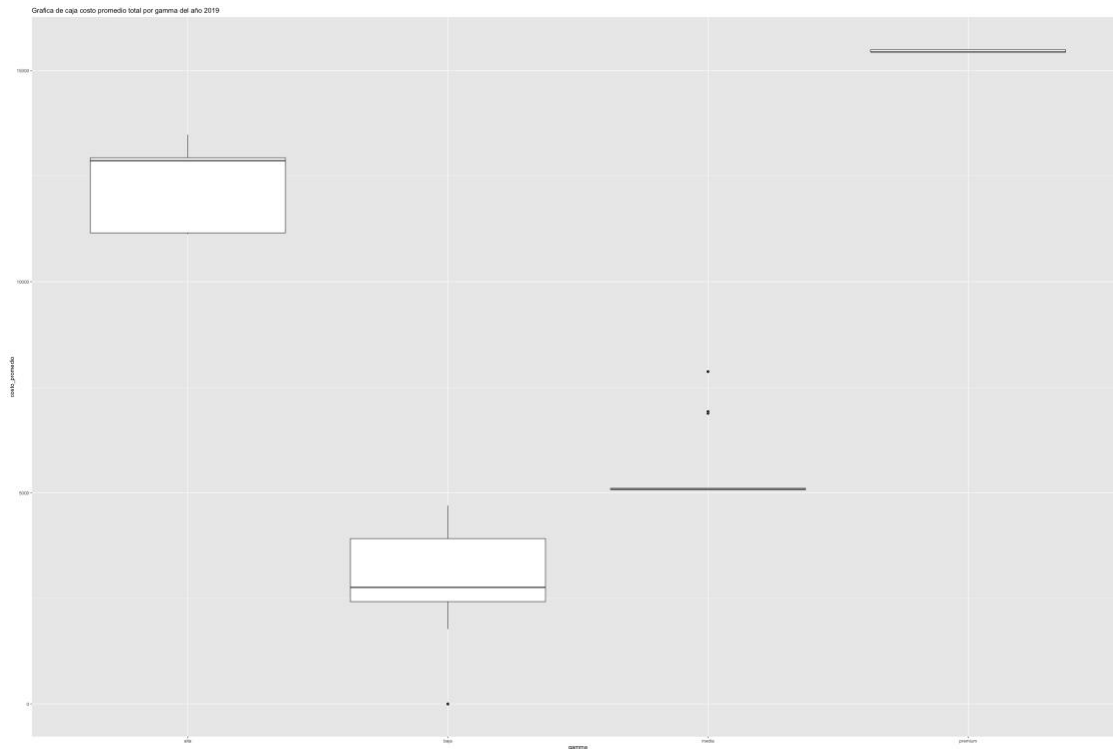
#2018 comportamiento de acuerdo a La gamma
 ggplot(d3,aes(x=gamma, y=costo_promedio))+
 geom_boxplot()+labs(title= "Grafica de caja costo promedio total por gamma
 del año 2018")



Comportamiento inusual en la gamma baja las otras no cuentan con valores atipicos

#2019 comportamiento de acuerdo a La gamma

```
ggplot(d4,aes(x=gamma, y=costo_promedio))+geom_boxplot()+labs(title= "Grafica de caja costo promedio total por gamma del año 2019")
```



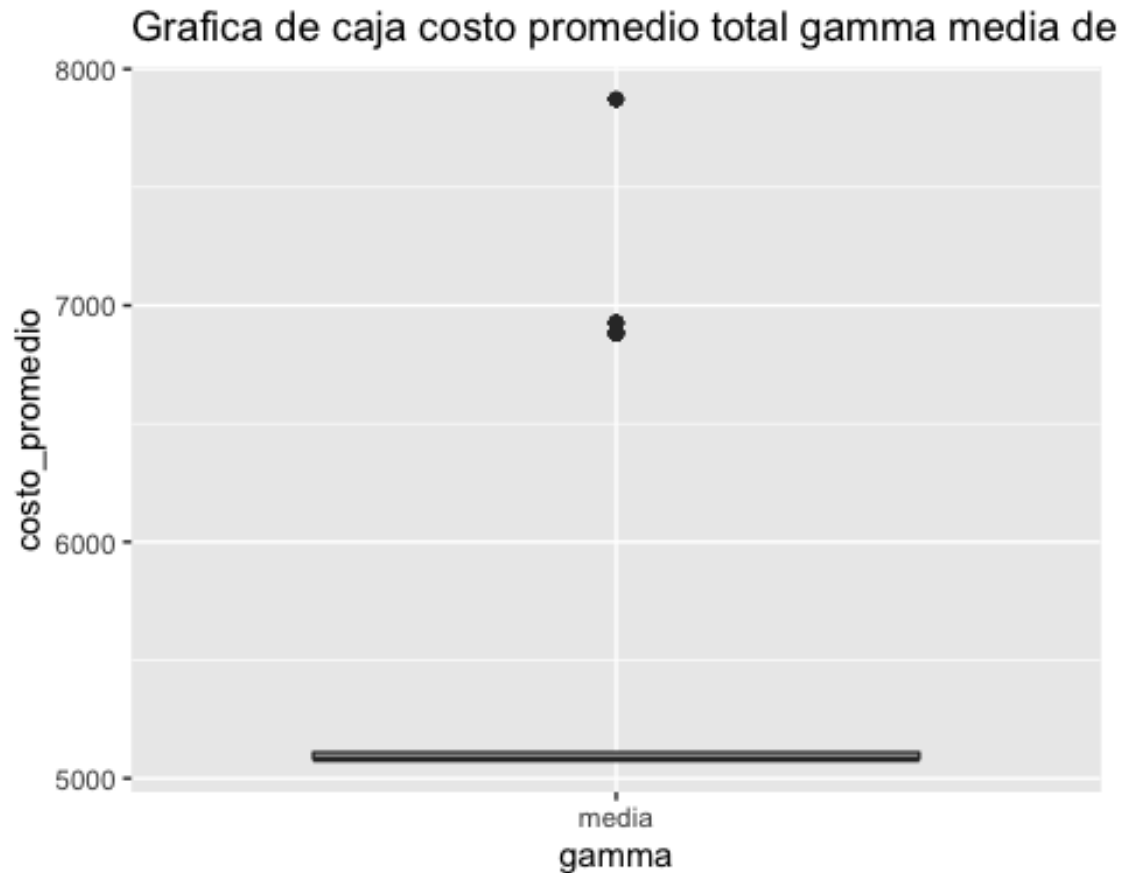
```
geom_boxplot()
```

```
## geom_boxplot: outlier.colour = NULL, outlier.fill = NULL, outlier.shape =  
19, outlier.size = 1.5, outlier.stroke = 0.5, outlier.alpha = NULL, notch =  
FALSE, notchwidth = 0.5, varwidth = FALSE, na.rm = FALSE, orientation = NA  
## stat_boxplot: na.rm = FALSE, orientation = NA  
## position_dodge2
```

A diferencia del año 2018 esta si cuenta con datos inusuales en la gamma baja y media

```
filtergamma<-d4%>%group_by((gamma))
```

```
filtergamma<-filtergamma%>%filter(gamma=="media")  
ggplot(filtergamma, aes (x=gamma,y=costo_promedio))+  
  geom_boxplot()+labs(title= "Grafica de caja costo promedio total  
gamma media del año 2019")
```



En la grafica podemos observar que los valores anormales estan entre 6900 y 8000 para el costo promedio total de los dispositivos clasificados como gamma media en el 2019

```
summary(d2)
```

```
## punto_de_venta      fecha      mes      anio
## Length:148575      Length:148575      Min.   : 1.000      Min.   :2018
## Class :character    Class :character    1st Qu.: 7.000      1st Qu.:2018
## Mode  :character    Mode  :character    Median : 9.000      Median :2018
##                               Mean  : 8.246      Mean   :2018
##                               3rd Qu.:11.000      3rd Qu.:2018
##                               Max.   :12.000      Max.   :2019
## num_ventas      sku      marca      gamma
## Min.   :1      Length:148575      Length:148575      Length:148575
## 1st Qu.:1      Class :character    Class :character    Class :character
## Median :1      Mode  :character    Mode  :character    Mode  :character
## Mean   :1
## 3rd Qu.:1
## Max.   :1
## costo_promedio      zona      estado      ciudad
## Min.   : 0      Length:148575      Length:148575      Length:148575
## 1st Qu.: 2426      Class :character    Class :character    Class :character
## Median : 2761      Mode  :character    Mode  :character    Mode  :character
## Mean   : 4758
```

```

## 3rd Qu.: 5082
## Max. :15498
##      latitud      longitud
## Min. :14.87   Min. : -117.12
## 1st Qu.:19.34   1st Qu.: -101.63
## Median :19.66   Median :  -99.23
## Mean   :21.46   Mean   : -100.43
## 3rd Qu.:21.92   3rd Qu.:  -99.05
## Max.   :32.67   Max.    :  -86.81

#cambio de variable fecha a DATE
d2$fecha <- as.Date(d2$fecha)

str(d2)

## 'data.frame': 148575 obs. of 14 variables:
## $ punto_de_venta: chr "1 poniente" "1 poniente" "1 poniente" "1
poniente" ...
## $ fecha : Date, format: "0030-06-20" "0030-06-20" ...
## $ mes : int 6 6 7 8 8 8 8 8 8 8 ...
## $ anio : int 2018 2018 2018 2018 2018 2018 2018 2018 2018 2018
...
## $ num_ventas : int 1 1 1 1 1 1 1 1 1 1 ...
## $ sku : chr "N.SJ7PROD" "N.SJ7PRON" "N.SAMGA6PLA" "N.SAMGJ4NG"
...
## $ marca : chr "samsung" "samsung" "samsung" "samsung" ...
## $ gamma : chr "baja" "baja" "media" "baja" ...
## $ costo_promedio: num 4184 4196 5109 1815 2505 ...
## $ zona : chr "centro sur" "centro sur" "centro sur" "centro
sur" ...
## $ estado : chr "puebla" "puebla" "puebla" "puebla" ...
## $ ciudad : chr "tehuacan" "tehuacan" "tehuacan" "tehuacan" ...
## $ latitud : num 18.5 18.5 18.5 18.5 18.5 ...
## $ longitud : num -97.4 -97.4 -97.4 -97.4 -97.4 ...

library(dplyr)
filtermes<-d3%>%group_by(mes)%>%summarise(sum_costo_promedio =
sum(costo_promedio))
filtermes

## # A tibble: 7 × 2
## mes sum_costo_promedio
## <int> <dbl>
## 1 6 55651200.
## 2 7 65032616.
## 3 8 75842095.
## 4 9 88116825.
## 5 10 70717178.
## 6 11 91175373.
## 7 12 128906921.

```

```

str(filtermes)

## tibble [7 × 2] (S3: tbl_df/tbl/data.frame)
## $ mes : int [1:7] 6 7 8 9 10 11 12
## $ sum_costo_promedio: num [1:7] 55651200 65032616 75842095 88116825
70717178 ...

filtermes

## # A tibble: 7 × 2
##   mes sum_costo_promedio
##   <int>          <dbl>
## 1     6      55651200.
## 2     7      65032616.
## 3     8      75842095.
## 4     9      88116825.
## 5    10      70717178.
## 6    11      91175373.
## 7    12     128906921.

```

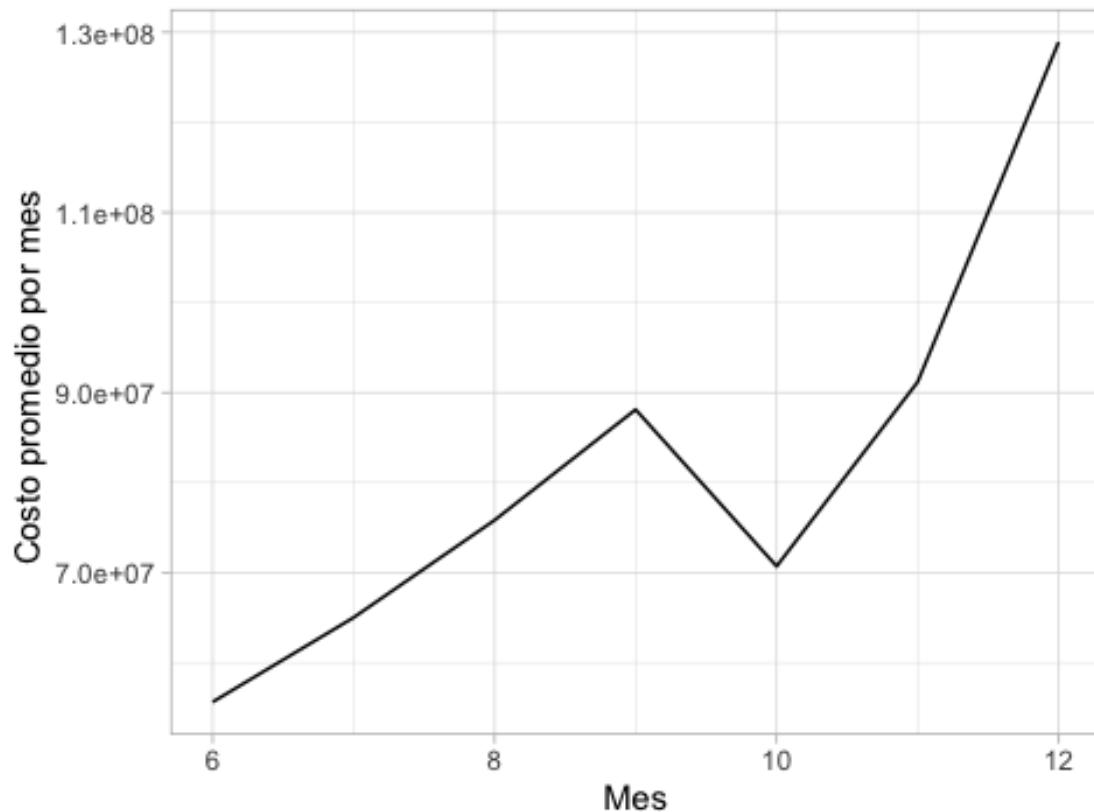
El mejor mes fue diciembre con un total de 128906921 en el año 2018

```

ggplot(filtermes,aes(x = mes, y = sum_costo_promedio ))+geom_line()+
  theme_light()+
  xlab("Mes")+
  ylab("Costo promedio por mes ")+
  labs(title= "Grafica de costo promedio total por mes del año 2018")

```

Grafica de costo promedio total por mes del año 201



Se puede apreciar en la grafica que se obtuvo un incremento considerable en el costo minimo a partir del mes 10 en el año 2018 y que tiende a la alta durante el ultimo periodo de ese año.

```
library(dplyr)
filtermes2<-d4%>%group_by(mes)%>%summarise(sum_costo_promedio =
sum(costo_promedio))
filtermes2

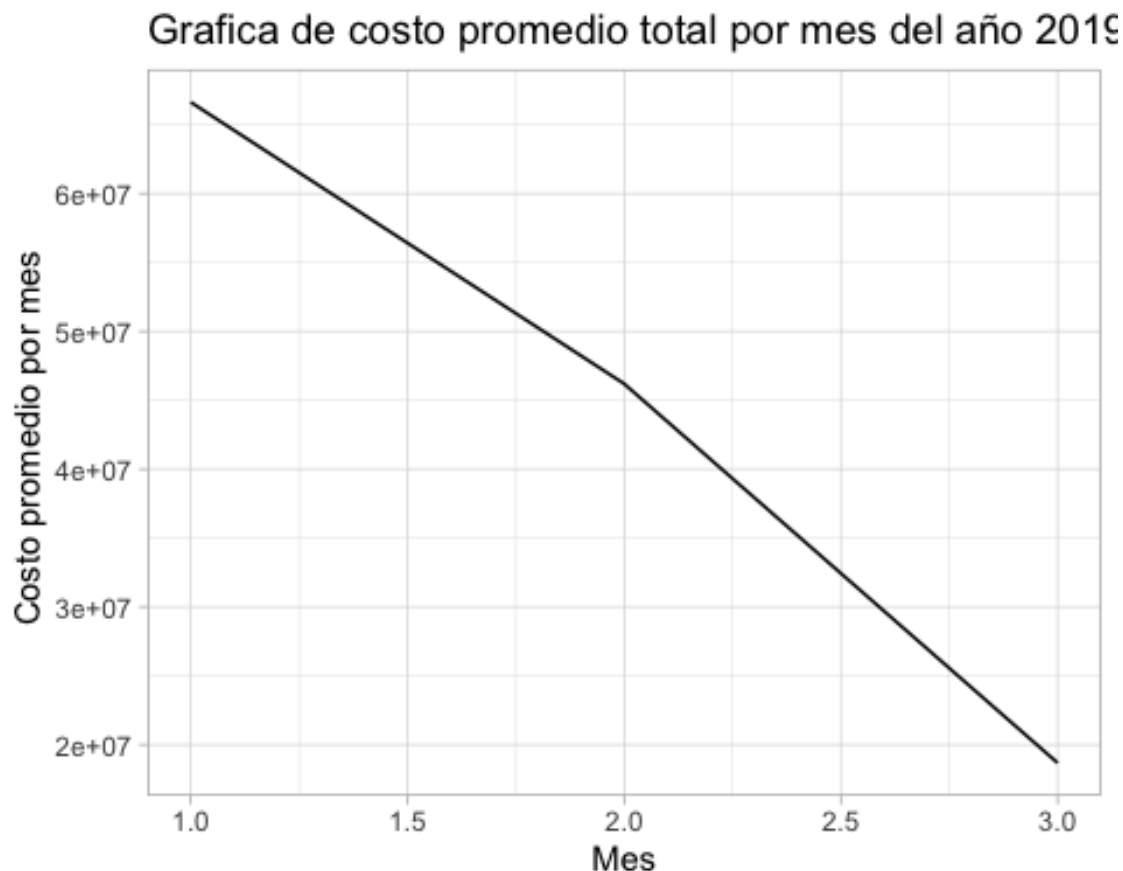
## # A tibble: 3 × 2
##   mes sum_costo_promedio
##   <int>          <dbl>
## 1     1          66629492.
## 2     2          46194490.
## 3     3          18684205.

str(filtermes2)

## tibble [3 × 2] (S3: tbl_df/tbl/data.frame)
## $ mes : int [1:3] 1 2 3
## $ sum_costo_promedio: num [1:3] 66629492 46194490 18684205

ggplot(filtermes2,aes(x = mes, y = sum_costo_promedio ))+geom_line()+
  theme_light()+
```

```
xlab("Mes")+
ylab("Costo promedio por mes ")+
labs(title= "Grafica de costo promedio total por mes del año 2019")
```



Por otra parte en este grafico que hace referencia al costo promedio mensual durante el 2019, y en referencia al 2018, se puede observar un comportamiento decreciente considerable, a partir del mes 3.

```
filter(mes2)

## # A tibble: 3 × 2
##   mes sum_costo_promedio
##   <int>          <dbl>
## 1     1      66629492.
## 2     2      46194490.
## 3     3      18684205.
```

El mes que menos costo promedio total tiene fue el 3 con un total de 18684205

Como podemos observar existe un incremento referente al costo promedio en el año 2018 a partir del mes 10 , a diferencia del 2019 el cual a inicios de año se mantiene una cifra representativa de 66,629,492 y termina decreciendo en el mes marzo con 18,684,205. La informacion anterior es de suma utilidad para considerar realizar un analisis en los mese

que se tiende a la baja y la posible relacion que se tenga en ciertas zonas geograficas o preferencias de ciertos modelos.