



Actividad

Las instituciones financieras como cualquier empresa dependen de sus clientes. Por lo que, deben generar la mejor atención y generar confianza, de no hacerlo las personas preferirán otros bancos. Así, el banco debe mantener a los clientes y motivar al público a unirse.

En este trabajo se realizará un análisis descriptivo de un banco con el objetivo de analizar la organización y los procesos que se llevan a cabo. Este análisis servirá para toma de decisiones y de estrategias dentro de la empresa.

Un banco requiere mejorar los tiempos de atención al cliente en ventanilla, para ello ha recolectado esta información anónimamente para cada cajero y transacción realizada.

Se dispone de un archivo libro de Excel **Data_Banco.xlsx**, con tres hojas:

- 1) Tiene llamada "Data" contiene los datos de las transacciones: Sucursal, Cajero, ID_Transaccion, Transaccion, Tiempo_Servicio_seg, Satisfaccion, Monto
- 2) Otra llamada "Data_Sucursal" que indica si en la sucursal se ha puesto o no el nuevo sistema
- 3) Otra hoja llamada "Data_Cajero" que contiene datos del Cajero, Edad, Sexo, Nivel_Formacion, Año Ingreso.

A continuación, se detallan cada una de las variables.

Sucursal: Variable Categórica asociada con el Número con el que se registra las sucursales del Banco

Cajero: Variable Categórica con los códigos de los cajeros

ID_Transaccion: Código que se almacenó cada una de las transacciones.

Tiempo_Servicio_seg: Variable continua con el tiempo que le tomo a cada cajero atender las transacciones.

Satisfacción: Variable Categórica con la calificación que los clientes le han dado a la atención de los cajeros.

Monto: Variable Continua con la cantidad de dinero que los clientes realizaron las transacciones.

NS: Variable Discreta que indica si la sucursal tiene o no el nuevo sistema.

Ubicación: Variable categórica para indicar en donde se encuentra la Sucursal

Edad: Variable continua de la edad de los cajeros.

Sexo: Sexo de los Cajeros.



Nivel de Formación: Nivel de Formación de los Cajeros.

Años de Ingreso: Variable continua con los años de Experiencia que tiene cada cajero.

Comencemos haciendo un análisis exploratorio de los datos

1) Realiza la lectura del archivo asignando a un objeto llamado `data_banco` la hoja llamada `data`. A un objeto llamado `data_sucursal` la hoja llamada `Sucursal` y a un objeto llamado `data_cajero` la hoja llamada `Cajero`.

2) Observa su estructura del objeto `data_banco`

3) Observa la cabecera del objeto `data_sucursal`

4) Escribe el código para observar los nombres de las variables de los tres objetos creados

5) Transforma el objeto `data_banco` a un tibble. Selecciona las variables `Transaccion`, `Tiempo_Servicio_seg`

Tibble es un objeto del paquete `dplyr`, entre las mejoras que da es que no imprime todo el objeto en pantalla, sino un resumen de este.

El operador Pipe `%>%` del paquete `magrittr` (del ecosistema Tidyverse) permiten que el código sea más legible porque:

- *Permite secuencias estructurantes de operaciones de datos de izquierda a derecha (a diferencia de dentro y fuera),

- * utiliza funciones anidadas,

- *Minimiza la necesidad de variables locales y definiciones de funciones

- *Facilita agregar pasos en cualquier lugar de la programación

- *permite programar como si se escribiese “del `data_banco`, selecciona las columnas `Transaccion` y `Tiempo_Servicio_seg`”

Se lee, del `data_banco`, selecciona las columnas `Transaccion` y `Tiempo_Servicio_seg`

6) Selecciona todas las columnas menos cajeros usando el pipe (`%>%`)

Notar que como no se asignó a un objeto, R evalúa la expresión y presenta el resultado.

7) Observa la clase de la variable `Satisfacción`. Conviértela a factor observa sus niveles.

8) Filtrar las filas correspondientes a la sucursal 62.

9) Filtrar las filas correspondientes a la sucursal 62 y hayan durado más de 120 segundos.

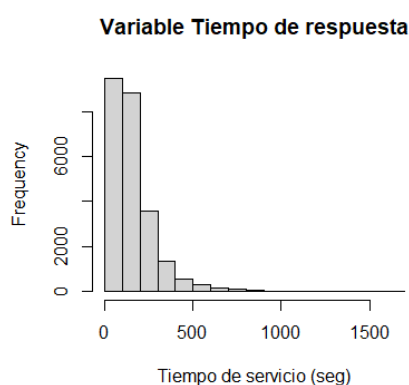
10) Filtrar las filas correspondientes a la sucursal 62, hayan durado más de 120 segundos y la evaluación a la satisfacción sea Muy Buena.

11) Usando la función `arrange()` ordena la variable `satisfacción`.

12) Ordenar cada transacción y dentro de cada transacción de mayor a menor por tiempo de servicio.



- 13) Crear una nueva columna llamada Tiempo_Servicio_Min con el tiempo en minutos
- 14) Observa la variable Monto, la cual tiene una mezcla de “,” y “.”. Vamos a corregirla de manera de reemplazar la coma por el punto.
- 15) Realiza un resumen del conjunto de datos data_banco
- 16) Calcula las medidas de tendencia central de la variable Tiempo_Servicio_seg usando R base y usando la función summary().
- 17) Calcula los deciles y el 5% y 95% de la variable Tiempo_Servicio_seg
- 18) Calcula las medidas de dispersión de la variable Tiempo_Servicio_seg
- 19) Para la Sucursal 62, obtener medidas de tendencia central para el tiempo de servicio para cada transacción y Nivel de Satisfacción. (Creación de resumen para datos agrupados)
summarise() permite aplicar funciones a nuestro data.frame, en R-base se usa tapply(), otra opción es ddply() del paquete plyr.
- 20) Crea una tabla de frecuencias relativas de la variable transacción y de Satisfacción (este tema se profundiza en el próximo encuentro).
- 21) Realiza un histograma de frecuencia de la variable tiempo de Respuesta usando los cortes de Stugers (regla práctica acerca del número de clases que deben considerar al elaborarse un histograma. (Se profundiza en el encuentro de visualización de datos)





- 22) Otra forma de describir variables numéricas es usando la función `describe()` del paquete `prettyR`. Observa cuando lo aplicamos a las variables numéricas del conjunto

```
# install.packages("prettyR")

library(prettyR)

describe(data_banco[, sapply(data_banco, is.numeric)], num.desc = c("mean", "sd",
                                                                    "median", "min", "max", "valid.n"))

class(data_banco$Monto)

data_banco$Monto<-as.numeric(data_banco$Monto, na.rm=TRUE)

class(data_banco$Monto)

describe(data_banco[, sapply(data_banco, is.numeric)], num.desc = c("mean", "sd",
                                                                    "median", "min", "max", "valid.n"))

describe(data_banco, num.desc = c("mean", "sd", "median", "min", "max", "valid.n"))

# calculo el primer cuartil para variable numérica
Q25 <- function(x, na.rm = TRUE) {
  quantile(as.numeric(x), 0.25, na.rm = na.rm, names = FALSE)
}

# calculo tercer cuartil para variable numérica
Q75 <- function(x, na.rm = TRUE) {
  quantile(as.numeric(x), 0.75, na.rm = na.rm, names = FALSE)
}

#describo los datos

describe(data_banco, num.desc = c("mean", "median", "sd", "min", "max", "Q25", "Q75",
                                  "valid.n"))
```

de datos `data_banco`

- 23) Obtener medidas de tendencia central para el tiempo de servicio para cada tipo de transacción.
- 24) Obtener medidas de tendencia central para el tiempo de servicio para cada Sucursal



- 25) Seleccionar la columna Tiempo_Servicio_seg y obtener un boxplot. (se profundiza en el encuentro de visualización de datos)

RESPUESTAS

```
#1)-----  
data_banco <- read_xlsx("Data_Banco.xlsx", col_names = TRUE, sheet = "Data")  
#24299 observaciones y 7 variables  
data_sucursal <- read_xlsx("Data_Banco.xlsx", sheet = "Data_Sucursal")  
observaciones y 3 variables  
data_cajero <- read_xlsx("Data_Banco.xlsx", sheet = "Data_Cajero")  
# 27 observaciones y 5 variables  
#2)-----  
str(data_banco)  
#3)-----  
head(data_sucursal, n = 5)  
#4)-----  
names(data_banco)  
names(data_sucursal)  
names(data_cajero)  
#5)-----  
data_banco <- tbl_df(data_banco)  
data_banco  
select(data_banco, Transaccion, Tiempo_Servicio_seg) # selecciono dos variables  
datos <- data_banco %>% select(Transaccion, Tiempo_Servicio_seg) # selecciono dos  
variables y las asigno al objeto "datos"  
#6)-----  
data_banco %>% select(-Cajero) %>% View  
#7)-----  
class(data_banco$Satisfaccion)  
data_banco$Satisfaccion <- as.factor(data_banco$Satisfaccion)  
levels(data_banco$Satisfaccion)  
#8)-----  
data_banco %>% filter(Sucursal == 62) %>% View
```



```

#9)-----
data_banco %>% filter(Sucursal == 62 & Tiempo_Servicio_seg > 120) %>% View

#10)-----
data_banco %>% filter(Sucursal == 62 & Tiempo_Servicio_seg > 120 & Satisfaccion ==
                      "Muy Bueno") %>% View

datos2 <- data_banco %>% filter(Sucursal == 62 & Tiempo_Servicio_seg > 120 &
                              Satisfaccion == "Muy Bueno")

#11)-----
data_banco %>% arrange(Satisfaccion) %>% View

#12)-----
data_banco %>% arrange(Transaccion, desc(Tiempo_Servicio_seg)) %>% View

#13)-----
data_banco %>% mutate(Tiempo_Servicio_Min = Tiempo_Servicio_seg/60)%>% View

#14)-----
data_banco %>% mutate(Monto = str_replace(Monto, pattern = ",", replacement = "."))

#15)-----
summary(data_banco)

#16)-----
mean(data_banco$Tiempo_Servicio_seg, na.rm = TRUE)
median(data_banco$Tiempo_Servicio_seg, na.rm = TRUE)
summary(data_banco$Tiempo_Servicio_seg, na.rm = TRUE)

#17)-----
quantile(data_banco$Tiempo_Servicio_seg, probs = seq(from = 0.1, to = 1, by = 0.1))
quantile(data_banco$Tiempo_Servicio_seg, probs = c(0.05, 0.95))

#18)-----
var(data_banco$Tiempo_Servicio_seg, na.rm = TRUE)
sd(data_banco$Tiempo_Servicio_seg, na.rm = TRUE)
IQR(data_banco$Tiempo_Servicio_seg, na.rm = TRUE)
range(data_banco$Tiempo_Servicio_seg, na.rm = TRUE)

#19)-----
data_banco %>% filter( Sucursal== 62) %>% group_by (Transaccion, Satisfaccion) %>%
  summarise_at( vars(Tiempo_Servicio_seg), funs ( MEDIA= mean(., na.rm=TRUE),
                                                MEDIA_ACOT= mean(., na.rm = TRUE, trim = 0.05),

```



```

                                CANTIDAD= n() ) )

#20)-----
table(data_banco$Transaccion)
table(data_banco$Satisfaccion)

#21)-----
hist(data_banco$Tiempo_Servicio_seg, breaks = "Sturges",
      main = "Variable Tiempo de respuesta", xlab = "Tiempo de servicio (seg)")

#22)-----

#23)-----
data_banco %>% group_by(Transaccion, Satisfaccion) %>%
summarise_at(vars(Tiempo_Servicio_seg),

  funs(MEDIA = mean(., na.rm = TRUE), MEDIA_ACOT = mean(., na.rm = TRUE, trim =
0.05),

      CANTIDAD = n()))

#24)-----
data_banco %>% group_by(Sucursal) %>%
summarise_at(vars(Tiempo_Servicio_seg),
  funs(MEDIA = mean(., na.rm = TRUE),
      MEDIA_ACOT = mean(., na.rm = TRUE, trim = 0.05),
      CANTIDAD = n()))

data_banco %>%
filter( Sucursal== 85 ) %$% # Operador pipe para seleccion de columnas
cor(Tiempo_Servicio_seg, as.numeric(Monto))

Nota: repetir para cada sucursal

#25)-----
boxplot(data_banco$Tiempo_Servicio_seg) ## Base de R

data_banco %>% select(Tiempo_Servicio_seg) %>% boxplot

boxplot(data_banco$Tiempo_Servicio_seg, main = "Boxplot para Tiempo de Servicio
(seg)",

      ylab = "Tiempo")

```