

Trabajo final Estadística Avanzada

true

true

true

true

octubre 28, 2020

Abstract

Este documento analiza los impactos de las variables macroeconómicas en los costos y gastos de una empresa en un determinado sector económico

Contents

Capítulo 1 Lectura de variables de empresa	1
Capítulo 2 Lectura y consolidación de variables económicas	6
Capítulo 3 Consolidación de la base	10

{newpage}

Capítulo 1 Lectura de variables de empresa

Actividad de evaluación de la asignatura Métodos Estadísticos Avanzados Profesor: Juan David Ospina Arango Estudiantes: Cindy Guerra, Diana Benjumea, Carlos Murillo, Luz Florez

Métodos Estadísticos Avanzados Maestría en Ciencia de los Datos Universidad EAFIT

Objetivo Caracterizar las relaciones entre algunos indicadores macroeconómicos y los costos y gastos de ventas de las empresas colombianas vigiladas por la SuperSociedades.

Lineamientos: 1. Con ayuda de un modelo lineal modele cree un modelo o varios modelos que permitan caracterizar la relación entre las variables PIB, Inflación, Desempleo, Tasa de Cambio, Balance Fiscal, Balance en Cuenta Corriente, Tasa de intervención, TRM y los costos y gastos de ventas.

2. Se debe escoger mínimo un tipo de empresas (Clasificación Industrial Internacional Uniforme) que tenga más de 20 empresas y tomar al menos los últimos tres años de información disponible.
3. Se debe evaluar el ajuste y la capacidad predictiva.
4. Se deben explicar todas las transformaciones de variables requeridas por el modelo.
5. Se deben explicar todos los pasos para la construcción de la base de datos: descarga de información, concatenación, etc.

6. Se debe incluir un análisis descriptivo.
7. Se debe incluir un análisis de la razonabilidad de las cifras.
8. Se debe redactar un reporte técnico documentando lo anterior. La sugerencia es utilizar un formato que permita la inclusión de gráficos basados en html o JavaScript (por ejemplo hmtl a partir de Rmarkdown). El código se debe subir a un repositorio Git y referenciarlo en el reporte. El reporte debe incluir una estimación del esfuerzo de las actividades de 1) consolidación de información, 3) transformación de variables y análisis descriptivo, 4) ajuste y validación de modelos y 5) redacción del reporte.
9. El trabajo se debe subir al canal del curso en Teams y se debe notificar por correo a la dirección judaosp@bancolombia.com.co.
10. La fecha de entrega es el viernes 30 de octubre y el trabajo se puede presentar en equipos de máximo cinco estudiantes.

Para acceder a los datos de costos y gastos de ventas: • Entrar a <http://pie.supersociedades.gov.co>
> MENÚ > Descarga Masiva de Información Descargar la información de los años 2016 a 2019

Primera iteración: Código CIIU seleccionado: G4711 Macrosector: Comercio Descripción: Comercio al por menor en establecimientos no especializados con surtido compuesto principalmente por alimentos, bebidas o tabaco.

Esta clase incluye: • Los establecimientos no especializados de comercio al por menor de productos cuyo surtido está compuesto principalmente de alimentos (víveres en general), bebidas o tabaco. No obstante, expenden otras mercancías para consumo de los hogares tales como vestuario, electrodomésticos, muebles, artículos de ferretería, cosméticos, entre otros. Suelen realizar este tipo de actividad los denominados supermercados, cooperativas de consumidores, comisariatos y otros establecimientos similares. También se incluyen las tiendas, los graneros, entre otros, que se encuentran en los pueblos o en barrios tradicionales.

Esta clase excluye: • El expendio de comidas preparadas en restaurantes, cafeterías y por autoservicio.

Al realizar los cargues iniciales de información nos dimos cuenta de que nos cruzaban muy pocas empresas por lo que el conjunto de datos seleccionado no era suficiente.

Iteración 2: Código CIIU seleccionado: B0722

```
library(tidyverse)
library("readxl")
library("dplyr")
```

1. Cargamos los datos básicos de las empresas

```
#Revisamos como son nuestros datos para saber si tenemos que realizar algún ajuste a la carga
file.show("./data/datosBasicosComplete.xlsx")

#Como el archivo no tiene forma de tabla al principio, debemos realizar la carga, ignorando las

#Cargar los archivos a un dataframe
pd_datos_basicos <- read_excel("./data/datosBasicosComplete.xlsx", sheet = "Reporte", skip=8, col_types =
"text", "text", "text", "text","text","text","text","text", "text","text","text","text","text",
```

```

"text","date","text","date", "text", "text"))

pd_datos_basicos %>% mutate(`Órgano Societario` = as.factor(`Órgano Societario`),
                             `Etapla Situación` = as.factor(`Etapla Situación`)) -> pd_datos_basico

head(pd_datos_basicos)

## # A tibble: 6 x 23
##   NIT   `Razón social` `Código CIIU` `Tipo Societari` `Objeto Social`
##   <chr> <chr>         <chr>         <chr>         <chr>
## 1 1001~ NOREÑA MANRI~ 0          PERSONA NATURAL <NA>
## 2 1001~ PEÑA RAMIREZ ~ H5229      PERSONA NATURAL <NA>
## 3 1002~ GONZALEZ SANC~ G4731      PERSONA NATURAL <NA>
## 4 1002~ RODRIGO JAVIE~ L6810      PERSONA NATURAL <NA>
## 5 1002~ BUITRAGO GONZ~ H4923      PERSONA NATURAL <NA>
## 6 1005~ KAREN JULIETH~ M7500      PERSONA NATURAL <NA>
## # ... with 18 more variables: `Dirección Notificación Judicial` <chr>, `Ciudad
## #   Notificación Judicial` <chr>, `Departamento Notificación Judicial` <chr>,
## #   `Teléfono Notificación Judicial` <chr>, `Dirección Domicilio` <chr>,
## #   `Ciudad Domicilio` <chr>, `Departamento Domicilio` <chr>, `Apartado
## #   Domicilio` <chr>, `E-Mail` <chr>, Web <chr>, Estado <chr>, `Fecha
## #   Estado` <dtm>, Situación <chr>, `Fecha Situación` <dtm>, `Etapla
## #   Situación` <fct>, `Fecha Etapla` <dtm>, `Nombre Representante Legal` <chr>,
## #   `Órgano Societario` <fct>

library(dplyr)

pd_datos_basicos_flt <- pd_datos_basicos[,c("NIT","Razón social","Código CIIU","Ciudad Domicilio",
names (pd_datos_basicos_flt) = c("NIT","razon_social","CIIU","ciudad","departamento", "estado",
                                "etapla_situacion")

pd_datos_basicos_flt <- filter(pd_datos_basicos_flt, CIIU == "B0722" & situacion == "ACTIVA")

head(pd_datos_basicos_flt)

## # A tibble: 6 x 9
##   NIT   razon_social CIIU   ciudad departamento estado situacion organo_societar~
##   <chr> <chr>         <chr> <chr>   <chr>         <chr> <chr>         <fct>
## 1 8002~ GRUPO DE BU~ B0722 MEDEL~ ANTIOQUIA   INSPE~ ACTIVA   ACTIVIDAD ECONO~
## 2 8110~ MINERA CROE~ B0722 MEDEL~ ANTIOQUIA   INSPE~ ACTIVA   ACTIVIDAD ECONO~
## 3 8110~ NUEVA CALIF~ B0722 MEDEL~ ANTIOQUIA   INSPE~ ACTIVA   ACTIVIDAD ECONO~
## 4 8110~ COLOMBIA GO~ B0722 MEDEL~ ANTIOQUIA   INSPE~ ACTIVA   ACTIVIDAD ECONO~
## 5 8110~ NEGOCIOS MI~ B0722 MEDEL~ ANTIOQUIA   INSPE~ ACTIVA   ACTIVIDAD ECONO~
## 6 8300~ ECO ORO MIN~ B0722 BUCAR~ SANTANDER   INSPE~ ACTIVA   ACTIVIDAD ECONO~
## # ... with 1 more variable: etapla_situacion <fct>

```

3. Cargamos los datos financieros

```
pd_datos_fin_2017 <- read_excel("./data/Plenas - Individuales2017.xlsx", sheet = "Estado de Resu
pd_datos_fin_2017 <- pd_datos_fin_2017[,c("Nit", "Periodo", "Costo de ventas", "Costos de distr
names (pd_datos_fin_2017) = c("NIT", "Periodo", "costo_ventas", "costo_distribucion", "gastos_a
datos_completos_fin <- merge (pd_datos_basicos_flt, pd_datos_fin_2017, by.x="NIT", by.y="NIT")
```

Para efectos del ejercicio, no tomaremos el archivo de 2017, ya que el archivo 2018 tiene los datos de 2017 con la nueva norma.

```
pd_datos_fin_2018 <- read_excel("./data/Plenas - Individuales2018.xlsx", sheet = "ERI" )
pd_datos_fin_2018 <- pd_datos_fin_2018[,c("Nit", "Periodo", "Costo de ventas", "Gastos de ventas
names (pd_datos_fin_2018) = c("NIT", "Periodo", "costo_ventas", "gastos_ventas", "gastos_adminis
datos_completos_2018 <- merge (pd_datos_basicos_flt, pd_datos_fin_2018, by.x="NIT", by.y="NIT")

#Le damos formato a los periodos

datos_completos_2018$Periodo[datos_completos_2018$Periodo == "Periodo Anterior"] <- "2017"
datos_completos_2018$Periodo[datos_completos_2018$Periodo == "Periodo Actual"] <- "2018"
```

```
pd_datos_fin_2019 <- read_excel("./data/Plenas - Individuales2019.xlsx", sheet = "ERI" )

#Revisar Costos de distribución
pd_datos_fin_2019 <- pd_datos_fin_2019[,c("Nit", "Periodo", "Costo de ventas", "Gastos de admini
names (pd_datos_fin_2019) = c("NIT", "Periodo", "costo_ventas", "gastos_administracion", "otros_
datos_completos <- merge (pd_datos_basicos_flt, pd_datos_fin_2019, by.x="NIT", by.y="NIT")

datos_completos$Periodo[datos_completos$Periodo == "Periodo Actual"] <- "2019"

datos_completos <- filter(datos_completos, Periodo == "2019")
```

```
#Eliminamos variable diferente a 2019
datos_completos_2018 <- select(datos_completos_2018, -gastos_ventas)

#UNimos los 2 dataframes
datos_completos = rbind(datos_completos, datos_completos_2018)
datos_completos %>% mutate( razon_social = as.factor(razon_social), CIIU = as.factor(CIIU), ciudad
                        departamento = as.factor(departamento), estado =as.factor(estado), Per
                        situacion = as.factor(situacion)) -> datos_completos
```

```
library(ggplot2)
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 4.0.3
```

```
## Loading required package: Matrix
```

```
##
```

```
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
```

```
##
```

```
##      expand, pack, unpack
```

```
summary(datos_completos)
```

```
##      NIT                                razon_social      CIIU
## Length:82          ANGLOGOLD ASHANTI COLOMBIA S.A.          : 3      B0722:82
## Class :character    CALDAS GOLD MARMATO S.A.S.              : 3
## Mode :character     CONTINENTAL GOLD LIMITED SUCURSAL COLOMBIA: 3
##                      ECO ORO MINERALS CORP                  : 3
##                      EXPLORACIONES CHAPARRAL COLOMBIA SAS    : 3
##                      EXPLORACIONES NORTHERN COLOMBIA S.A.S   : 3
##                      (Other)                                :64
##      ciudad      departamento      estado      situacion
## BOGOTÁ, D.C.:26  ANTIOQUIA      :44  INSPECCION:41  ACTIVA:82
## BUCARAMANGA :12  BOGOTÁ, D. C.:26  VIGILANCIA:41
## ENVIGADO      : 2  SANTANDER      :12
## MEDELLÍN      :42
##
##
##
##      organo_societario
## ACTIVIDAD ECONOMICA DIFERENTE:82
## ACTIVIDAD DE VIVIENDA          : 0
## BOLSA PROD. AGROPECUARIO       : 0
## CLUBES DEPORTIVOS              : 0
## COMERCIANTE                    : 0
## EMPRESA MULTINACIONAL ANDINA   : 0
## (Other)                        : 0
##
##      etapa_situacion Periodo      costo_ventas
## ACTIVA                                :82      2017:28  Min.      :      0
## A LIQUIDACION POR ADJUDICACION      : 0      2018:28  1st Qu.:   59760
## ACUERDO CONFIRMADO                  : 0      2019:26  Median : 10986154
## ACUERDO DE REORGANIZACION EN EJECUCION: 0      Mean   : 47651275
## ADMISION                            : 0      3rd Qu.: 39093635
## CANCELADA                           : 0      Max.   :408474390
## (Other)                             : 0      NA's   :42
## gastos_administracion otros_gastos      costos_financieros
## Min.      :      1963      Min.      :      2      Min.      :      0
## 1st Qu.:   256194      1st Qu.:   16748      1st Qu.:   87145
```

```
## Median : 1099196      Median : 436728      Median : 1288761
## Mean   : 15247178     Mean   : 3893153     Mean   : 4145977
## 3rd Qu.: 7201307      3rd Qu.: 2322578     3rd Qu.: 3418959
## Max.   :287863704     Max.   :46689684     Max.   :52689630
## NA's   :16           NA's   :13          NA's   :23
## gasto_impuestos      ingresos_actividades_ordinarias otros_ingresos
## Min.    : -8974634    Min.    :      0      Min.    :      0
## 1st Qu.:      499     1st Qu.:      0      1st Qu.: 14600
## Median  :     36788    Median  :      0      Median  : 106427
## Mean    : 6355794     Mean    :100173105     Mean    : 1106485
## 3rd Qu.: 813045      3rd Qu.: 42815317     3rd Qu.: 1551194
## Max.    :131684824    Max.    :954650443     Max.    :11404344
## NA's    :24          NA's    :54          NA's    :20
## ingresos_financieros
## Min.    :      0
## 1st Qu.: 15028
## Median  : 251425
## Mean    : 2067049
## 3rd Qu.: 1794058
## Max.    :26100695
## NA's    :31
```

```
#p1=ggplot(datos, aes(y=cd4,x=week,group=patid,colour=trtarm))
#p1+geom_line()
```

REFERENCIAS:

https://www.dian.gov.co/ciiu/Documents/Resolucion_000139_21_Nov_2012.pdf

<https://linea.ccb.org.co/descripcionciiu/>

<https://siis.ia.supersociedades.gov.co/>

https://www.supersociedades.gov.co/delegatura_aec/Paginas/Base-completa-EF-2019.aspx

Capítulo 2 Lectura y consolidación de variables económicas

A continuación se presenta el proceso que se ejecutó para generar un dataframe con las variables de PIB, Inflación, Desempleo, Balance Fiscal, Balance en Cuenta Corriente, Tasa de intervención, TRM

Para el PIB: Es un indicador económico que refleja el valor monetario de todos los bienes y servicios finales producidos por un país o región en un determinado periodo de tiempo, normalmente un año. Se utiliza para medir la riqueza que genera un país.

```
library(dplyr)
#Los datos son tomados de https://datosmacro.expansion.com/pcb/colombia
# vectores
anyo <- c("2016", "2017", "2018", "2019")
PIB_M.E. <- c(289.239, 280.249, 275.999, 255.416)
Var.PIB <- c(3.3, 2.5, 1.4, 2.1)
```

```
#Crear dataframe de vectores
PIB <- data.frame(anyo, PIB_M.E., Var.PIB)
head(PIB)
```

```
##   anyo PIB_M.E. Var.PIB
## 1 2016  289.239    3.3
## 2 2017  280.249    2.5
## 3 2018  275.999    1.4
## 4 2019  255.416    2.1
```

Para la inflación: La inflación es un fenómeno que se observa en la economía de un país y está relacionado con el aumento desordenado de los precios de la mayor parte de los bienes y servicios que se comercian en sus mercados, por un periodo de tiempo prolongado.

```
#Los datos son tomados de https://es.wikipedia.org/wiki/Anexo:Variaci%C3%B3n_de_la_inflaci%C3%B3n
# vectores
anyo <- c("2016", "2017", "2018", "2019")
Inflacion <- c(5.75, 4.09, 3.18, 3.80)
#Crear dataframe de vectores
Inflacion <- data.frame(anyo, Inflacion)
head(Inflacion)
```

```
##   anyo Inflacion
## 1 2016    5.75
## 2 2017    4.09
## 3 2018    3.18
## 4 2019    3.80
```

Para el desempleo: Es otra de las variables mas importantes de la macroeconomía, porque afecta directamente el bienestar de las personas. El desempleo es el porcentaje de la fuerza de trabajo que está buscando trabajo activamente y que actualmente se encuentra desempleada.

```
#Los datos son tomados de https://www.dane.gov.co/index.php/estadisticas-por-tema/mercado-laboral
# vectores
anyo <- c("2016", "2017", "2018", "2019")
Desempleo <- c(9.2, 9.4, 9.7, 10.5)
Var.Desempleo <- c(3.36, 1.99, 3.19, 8.25)
#Crear dataframe de vectores
Desempleo <- data.frame(anyo, Desempleo, Var.Desempleo)
head(Desempleo)
```

```
##   anyo Desempleo Var.Desempleo
## 1 2016    9.2    3.36
## 2 2017    9.4    1.99
## 3 2018    9.7    3.19
## 4 2019   10.5    8.25
```

Para el balance fiscal: Es la diferencia entre ingresos y gastos públicos en un determinado territorio.

```
#Los datos son tomados de http://www.urf.gov.co/webcenter/portal/EntidadesFinancieras/pages_EntidadFinanciera.jspx
# vectores
```

```

anyo <- c("2016", "2017", "2018", "2019")
GNC <- c(-4, -3.6, -3.1, -2.5)
#Crear dataframe de vectores
GNC <- data.frame(anyo,GNC)
head(GNC)

```

```

##   anyo  GNC
## 1 2016 -4.0
## 2 2017 -3.6
## 3 2018 -3.1
## 4 2019 -2.5

```

Para el balance en cuenta corriente: Es el conjunto de transacciones de intercambio de bienes y servicios, rentas y transferencias (tanto corrientes como de capital), su saldo determina la capacidad o necesidad de financiación de un país.

```

# vectores
anyo <- c("2016", "2017", "2018", "2019")
Balance_Cuenta_Corriente <- c(-13747.75, -13117.66, -10240.88, -12036.18)
#Crear dataframe de vectores
Balance_Cuenta_Corriente <- data.frame(anyo, Balance_Cuenta_Corriente)
head(Balance_Cuenta_Corriente)

```

```

##   anyo Balance_Cuenta_Corriente
## 1 2016                -13747.75
## 2 2017                -13117.66
## 3 2018                -10240.88
## 4 2019                -12036.18

```

Para la tasa de intervención: Corresponde a la tasa de interés mínima que le cobra el Banco de la República a las entidades financieras por los préstamos que les concede generalmente a un día y, además, sirve como referencia para establecer la tasa de interés máxima que les paga por recibirles dinero que tengan como excedente.

#Los datos son tomados de <https://www.banrep.gov.co/es/estadisticas/tasas-interes-politica-monet>

```

# vectores
anyo <- c("2016", "2017", "2018", "2019")
TIM_promedio<- c(7.10, 6.13, 4.35, 4.25)
#Crear dataframe de vectores
TIM <- data.frame(anyo, TIM_promedio)
head(TIM)

```

```

##   anyo TIM_promedio
## 1 2016           7.10
## 2 2017           6.13
## 3 2018           4.35
## 4 2019           4.25

```

Para la TRM: La tasa de cambio representativa del mercado (TRM) es la cantidad de pesos colombianos por un dólar de los Estados Unidos. La TRM se calcula con base en las operaciones de

compra y venta de divisas entre intermediarios financieros que transan en el mercado cambiario colombiano, con cumplimiento el mismo día cuando se realiza la negociación de las divisas.

Actualmente la Superintendencia Financiera de Colombia es la que calcula y certifica diariamente la TRM con base en las operaciones registradas el día hábil inmediatamente anterior.

```
#Se leen los datos - estos datos son tomados de https://www.dolar-colombia.com/historico
dataset = read.csv('./data/TRM.csv', check.names = FALSE, encoding = "UTF-8", blank.lines.skip =

#se conservan unicamente las columnas de año y TRM
df = dataset[1]
df['TRM'] = dataset[3]
df$TRM <- as.numeric(as.character(df$TRM))

#Se agrupa bajo la media
media = df
media = media %>%
  group_by(media[1]) %>%
  summarise(across(.cols = everything(), .fns = mean))

## `summarise()` ungrouping output (override with `.groups` argument)

#para la mediana
mediana = df
mediana = mediana %>%
  group_by(mediana[1]) %>%
  summarise(across(.cols = everything(), .fns = median))

## `summarise()` ungrouping output (override with `.groups` argument)

#Se genera un dataframe con los datos obtenidos
df = media
colnames(df)[2] <- 'TRM_media'
df['TRM_mediana'] <- mediana[2]
df

## # A tibble: 4 x 3
##   Anyo TRM_media TRM_mediana
##   <int>   <dbl>       <dbl>
## 1  2016    3051.        3003.
## 2  2017    2951.        2942.
## 3  2018    2956.        2898.
## 4  2019    3281.        3277.
```

Se unen los datos en un solo dataframe

```
df['PIB_M.E.'] = PIB[2]
df['Var.PIB'] = PIB[3]
df['Inflacion'] = Inflacion[2]
df['Desempleo'] = Desempleo[2]
df['var.Desempleo'] = Desempleo[3]
df['GNC'] = GNC[2]
```

```
df['Balance_Cuenta_Corriente'] = Balance_Cuenta_Corriente[2]
df['TIM_promedio'] = TIM[2]
head(df)
```

```
## # A tibble: 4 x 11
##   Anyo TRM_media TRM_mediana PIB_M.E. Var.PIB Inflacion Desempleo var.Desempleo
##   <int>   <dbl>     <dbl>   <dbl>   <dbl>   <dbl>   <dbl>       <dbl>
## 1  2016    3051.     3003.    289.    3.3     5.75    9.2         3.36
## 2  2017    2951.     2942.    280.    2.5     4.09    9.4         1.99
## 3  2018    2956.     2898.    276.    1.4     3.18    9.7         3.19
## 4  2019    3281.     3277.    255.    2.1     3.8     10.5        8.25
## # ... with 3 more variables: GNC <dbl>, Balance_Cuenta_Corriente <dbl>,
## #   TIM_promedio <dbl>
```

Capítulo 3 Consolidación de la base

En esta sección se unen las dos bases generadas en las fases anteriores en una sola base

```
datos_completos %>% mutate(Periodo=as.numeric(as.character(Periodo))) %>% inner_join(df, by=c("Per
datos_completos %>% mutate(costos_gastos_totales = gastos_administracion + otros_gastos +gasto_i
                          ingresos_totales = ingresos_actividades_ordinarias + ingresos_financi
                          #select(gastos_administracion, -otros_gastos, -gasto_impuestos, -costo
base_modelado <- droplevels(base_modelado)
summary(base_modelado)
```

```
##      NIT                                razon_social    CIIU
## Length:82          ANGLOGOLD ASHANTI COLOMBIA S.A.      : 3    B0722:82
## Class :character    CALDAS GOLD MARMATO S.A.S.          : 3
## Mode  :character    CONTINENTAL GOLD LIMITED SUCURSAL COLOMBIA: 3
##                                ECO ORO MINERALS CORP      : 3
##                                EXPLORACIONES CHAPARRAL COLOMBIA SAS : 3
##                                EXPLORACIONES NORTHERN COLOMBIA S.A.S : 3
##                                (Other)                     :64
##      ciudad      departamento      estado      situacion
## BOGOTÁ, D.C.:26  ANTIOQUIA      :44    INSPECCION:41    ACTIVA:82
## BUCARAMANGA :12  BOGOTÁ, D. C.:26    VIGILANCIA:41
## ENVIGADO      : 2  SANTANDER      :12
## MEDELLÍN      :42
##
##
##
##      organo_societario etapa_situacion      Periodo
## ACTIVIDAD ECONOMICA DIFERENTE:82      ACTIVA:82      Min.      :2017
##                                           1st Qu.:2017
##                                           Median :2018
##                                           Mean   :2018
```

```

##                                     3rd Qu.:2019
##                                     Max.    :2019
##
## costo_ventas      gastos_administracion  otros_gastos
## Min.    :      0   Min.    :      0      Min.    :      0
## 1st Qu.:      0   1st Qu.:   15366      1st Qu.:    848
## Median :      0   Median :   560711     Median :   120317
## Mean    : 23244524   Mean    : 12272119     Mean    : 3275946
## 3rd Qu.: 6666040   3rd Qu.:   5672411     3rd Qu.: 1736150
## Max.    :408474390   Max.    :287863704     Max.    :46689684
##
## costos_financieros gasto_impuestos      ingresos_actividades_ordinarias
## Min.    :      0   Min.    : -8974634     Min.    :      0
## 1st Qu.:      0   1st Qu.:      0      1st Qu.:      0
## Median : 168546   Median :    2546     Median :      0
## Mean    : 2983081   Mean    : 4495562     Mean    : 34205451
## 3rd Qu.: 2383609   3rd Qu.:   141459     3rd Qu.:      0
## Max.    :52689630   Max.    :131684824     Max.    :954650443
##
## otros_ingresos      ingresos_financieros   TRM_media      TRM_mediana
## Min.    :      0   Min.    :      0      Min.    :2951     Min.    :2898
## 1st Qu.:      0   1st Qu.:      0      1st Qu.:2951     1st Qu.:2898
## Median : 36549   Median :    3942     Median :2956     Median :2942
## Mean    : 836610   Mean    : 1285604     Mean    :3058     Mean    :3033
## 3rd Qu.: 601104   3rd Qu.:   850632     3rd Qu.:3281     3rd Qu.:3277
## Max.    :11404344   Max.    :26100695     Max.    :3281     Max.    :3277
##
## PIB_M.E.      Var.PIB      Inflacion      Desempleo
## Min.    :255.4   Min.    :1.400     Min.    :3.180     Min.    : 9.400
## 1st Qu.:255.4   1st Qu.:1.400     1st Qu.:3.180     1st Qu.: 9.400
## Median :276.0   Median :2.100     Median :3.800     Median : 9.700
## Mean    :270.9   Mean    :1.998     Mean    :3.687     Mean    : 9.851
## 3rd Qu.:280.2   3rd Qu.:2.500     3rd Qu.:4.090     3rd Qu.:10.500
## Max.    :280.2   Max.    :2.500     Max.    :4.090     Max.    :10.500
##
## var.Desempleo      GNC      Balance_Cuenta_Corriente  TIM_promedio
## Min.    :1.990   Min.    : -3.60     Min.    : -13118      Min.    :4.250
## 1st Qu.:1.990   1st Qu.: -3.60     1st Qu.: -13118      1st Qu.:4.250
## Median :3.190   Median : -3.10     Median : -12036      Median :4.350
## Mean    :4.385   Mean    : -3.08     Mean    : -11792      Mean    :4.926
## 3rd Qu.:8.250   3rd Qu.: -2.50     3rd Qu.: -10241      3rd Qu.:6.130
## Max.    :8.250   Max.    : -2.50     Max.    : -10241      Max.    :6.130
##
## costos_gastos_totales ingresos_totales
## Min.    :    2869   Min.    :      0
## 1st Qu.:   528039   1st Qu.:   24364
## Median :   4175588   Median :   587346
## Mean    : 43288150   Mean    : 36327665

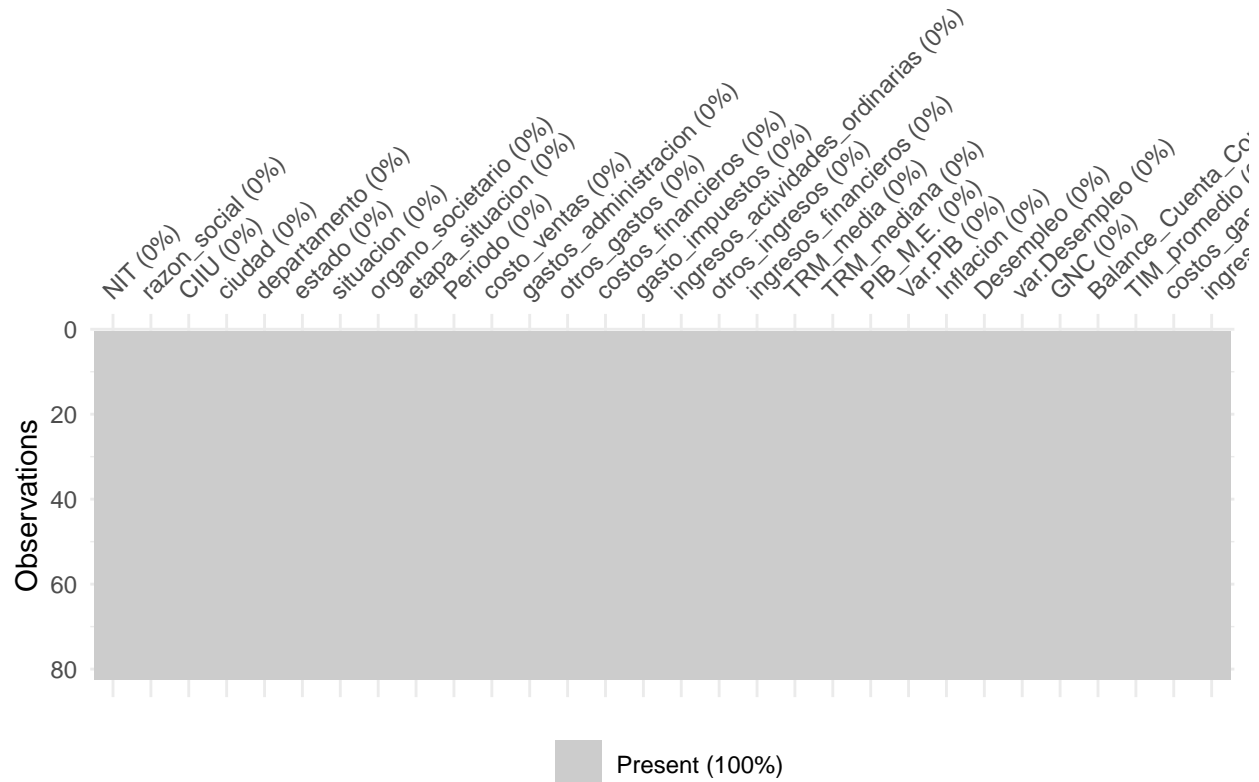
```

```
## 3rd Qu.: 26972121      3rd Qu.: 3698197
## Max.      :700560590    Max.      :967119965
##
```

```
library(visdat)
```

```
## Warning: package 'visdat' was built under R version 4.0.3
```

```
vis_miss(base_modelado)
```



```
# Capítulo 4: análisis descriptivo
```

```
library(ggplot2)
```

```
library(tidyr)
```

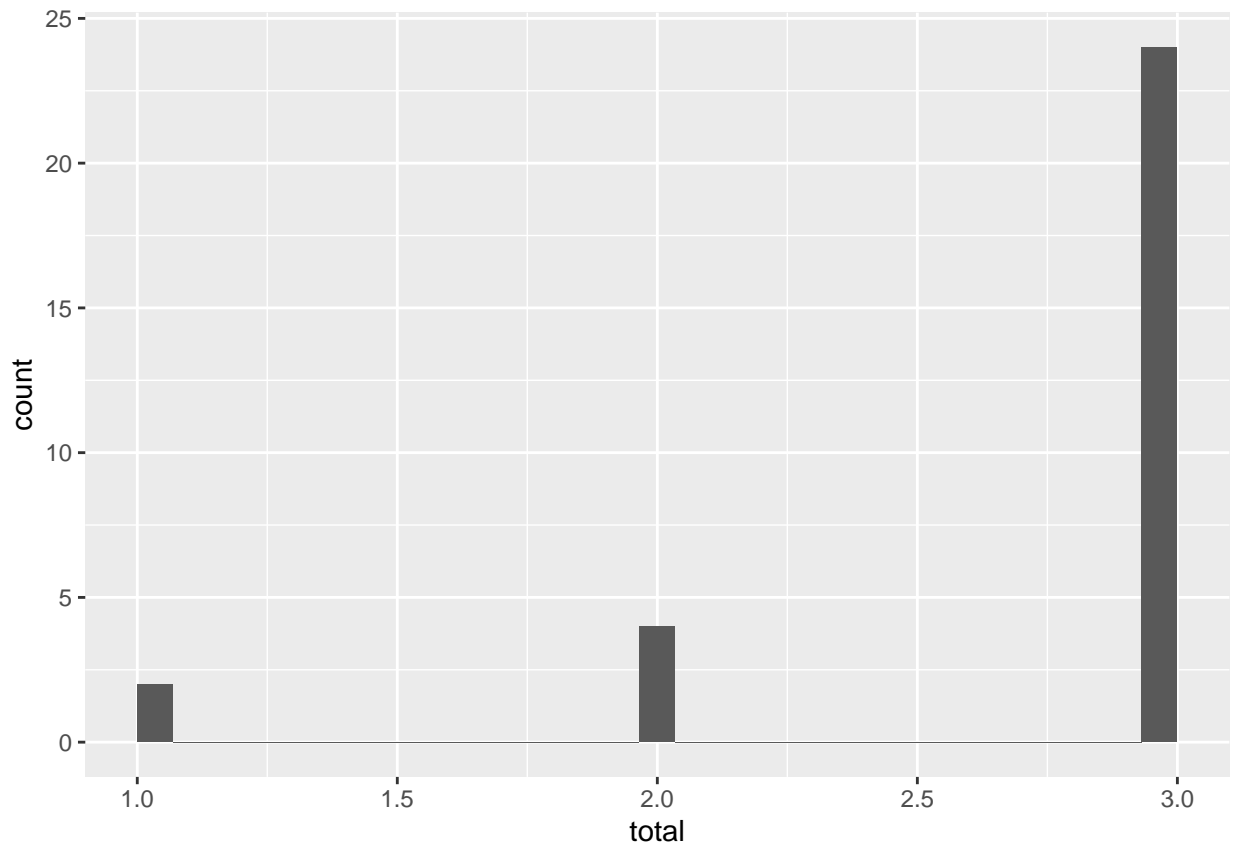
```
datosValidar <- base_modelado %>% group_by(NIT, razon_social) %>% summarise(total=n())
```

```
## `summarise()` regrouping output by 'NIT' (override with `.groups` argument)
```

```
#crear histograma para ver la cantidad de periodos por empresa
```

```
ggplot(datosValidar, aes(x=total)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
datosValidarDepartamento <- base_modelado %>% group_by(departamento, Periodo) %>% summarise(cost
ingre
```

```
## `summarise()` regrouping output by 'departamento' (override with `.groups` argument)
```

```
ggplot(datosValidarDepartamento, aes(x= Periodo))+
  geom_line(aes(y = costo_gasto_total_dep, color="darkred", linetype="twodash"))+
  geom_label(aes(y = costo_gasto_total_dep, label=costo_gasto_total_dep)) +
  geom_line(aes(y = ingresos_totales_dep, label="Ingresos"), color = "steelblue")+
  geom_label(aes(y = ingresos_totales_dep, label=ingresos_totales_dep)) +
  facet_wrap(~departamento, scales = "free_y")
```

```
## Warning: Ignoring unknown aesthetics: label
```

