

Actividad_8_LARFA01363601_Ing_Car

Luis Ángel Ramírez Franco A01363601

21/9/2020

```
library(tidyverse)
```

```
## -- Attaching packages -----
```

```
## v ggplot2 3.3.2      v purrr  0.3.4
## v tibble  3.0.3      v dplyr  1.0.1
## v tidyr   1.1.1      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0
```

```
## -- Conflicts -----
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
datos <- read.csv("E:/The Last Dance/Lab Diseño y Opt/equipo_5_apple_1_datos_limpios.csv")
head(datos)
```

```
##   punto_venta      fecha mes anio num_ventas      sku marca gamma
## 1  5 de mayo zmm 09/06/2018   6 2018         1 N.ISE32GR apple  baja
## 2  5 de mayo zmm 20/12/2018  12 2018         1 N.IP732B apple media
## 3  5 de mayo zmm 01/02/2019   2 2019         1 N.IP732B apple media
## 4  5 de mayo zmm 21/02/2019   2 2019         1 N.ISE32GR apple  baja
## 5      acayucan 15/08/2018   8 2018         1 N.ISE32PTB apple  baja
## 6      acayucan 29/08/2018   8 2018         1 N.ISE32GRB apple  baja
##   costo_promedio      zona      estado ciudad latitud longitud
## 1      4641.083 centro occidente michoacan zamora 19.98131 -102.2833
## 2      8855.867 centro occidente michoacan zamora 19.98131 -102.2833
## 3      8855.867 centro occidente michoacan zamora 19.98131 -102.2833
## 4      4641.083 centro occidente michoacan zamora 19.98131 -102.2833
## 5      4983.012 golfo de mexico veracruz  oluta 17.93999  -94.9106
## 6      4976.971 golfo de mexico veracruz  oluta 17.93999  -94.9106
```

```
dim(datos)
```

```
## [1] 19890    14
```

```
str(datos)
```

```
## 'data.frame': 19890 obs. of 14 variables:
## $ punto_de_venta: chr "5 de mayo zmm" "5 de mayo zmm" "5 de mayo zmm" "5 de mayo zmm" ...
## $ fecha : chr "09/06/2018" "20/12/2018" "01/02/2019" "21/02/2019" ...
## $ mes : int 6 12 2 2 8 8 9 11 11 2 ...
## $ anio : int 2018 2018 2019 2019 2018 2018 2018 2018 2018 2019 ...
## $ num_ventas : int 1 1 1 1 1 1 1 1 1 1 ...
## $ sku : chr "N.ISE32GR" "N.IP732B" "N.IP732B" "N.ISE32GR" ...
## $ marca : chr "apple" "apple" "apple" "apple" ...
## $ gamma : chr "baja" "media" "media" "baja" ...
## $ costo_promedio: num 4641 8856 8856 4641 4983 ...
## $ zona : chr "centro occidente" "centro occidente" "centro occidente" "centro occidente"
## $ estado : chr "michoacan" "michoacan" "michoacan" "michoacan" ...
## $ ciudad : chr "zamora" "zamora" "zamora" "zamora" ...
## $ latitud : num 20 20 20 20 17.9 ...
## $ longitud : num -102.3 -102.3 -102.3 -102.3 -94.9 ...
```

```
summary(datos)
```

```
## punto_de_venta fecha mes anio
## Length:19890 Length:19890 Min. : 1.000 Min. :2018
## Class :character Class :character 1st Qu.: 6.000 1st Qu.:2018
## Mode :character Mode :character Median : 7.000 Median :2018
## Mean : 6.962 Mean :2018
## 3rd Qu.: 9.000 3rd Qu.:2018
## Max. :12.000 Max. :2019
## num_ventas sku marca gamma
## Min. :1 Length:19890 Length:19890 Length:19890
## 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:19890 Length:19890 Length:19890
## 1st Qu.:4977 Class :character Class :character Class :character
## Median :8856 Mode :character Mode :character Mode :character
## Mean :6920
## 3rd Qu.:8856
## Max. :9800
## latitud longitud
## Min. :14.87 Min. : -117.11
## 1st Qu.:19.30 1st Qu.: -101.38
## Median :19.67 Median : -99.27
## Mean :20.97 Mean : -100.07
## 3rd Qu.:21.15 3rd Qu.: -99.03
## Max. :32.67 Max. : -86.81
```

```
datos$punto_de_venta <- as.character(datos$punto_de_venta) #Caracter o factor
datos$fecha <- as.Date(datos$fecha)
datos$mes <- as.numeric(datos$mes) #pueden ser factor / para usos prácticos de este ejercicio conviene
datos$anio <- as.numeric(datos$anio) #pueden ser factor / para usos prácticos de este ejercicio conviene
datos$sku <- as.character(datos$sku) #Caracter o factor
datos$marca <- as.character(datos$marca) #Caracter o factor
datos$sku <- as.character(datos$sku) #Caracter o factor
```

1. Índices - manejo más sencillo de variables cualitativas más importantes

Paso 1. Crear Índices por separado

Índice - Tienda (1 poniente:1, 5 de mayo zmm:2)

```
pdv_id <- datos%>%select(punto_de_venta)%>%unique()%>%arrange()  
head(pdv_id)
```

```
##           punto_de_venta  
## 1           5 de mayo zmm  
## 2           acayucan  
## 3       ace aldama centro  
## 4           ace alhondiga  
## 5       ace benito juarez  
## 6 ace cd guzman santuario
```

```
pdv_id$pdv_id <- as.character(seq.int(nrow(pdv_id)))  
head(pdv_id)
```

```
##           punto_de_venta pdv_id  
## 1           5 de mayo zmm      1  
## 2           acayucan      2  
## 3       ace aldama centro      3  
## 4           ace alhondiga      4  
## 5       ace benito juarez      5  
## 6 ace cd guzman santuario      6
```

Fecha - No porque lo que nos interesa es el pronóstico por mes Mes - ¡CUIDADO! Número distintivo por mes -> Índice distintivo por periodo de registro

```
mes_id <- datos%>%select(mes,anio)%>%unique()%>%arrange(anio,mes)  
mes_id
```

```
##      mes anio  
## 1      6 2018  
## 2      7 2018  
## 3      8 2018  
## 4      9 2018  
## 5     10 2018  
## 6     11 2018  
## 7     12 2018  
## 8      1 2019  
## 9      2 2019  
## 10     3 2019
```

```
mes_id$mes_id <- as.character(seq.int(0,9,1))
mes_id
```

```
##      mes anio mes_id
## 1      6 2018      0
## 2      7 2018      1
## 3      8 2018      2
## 4      9 2018      3
## 5     10 2018      4
## 6     11 2018      5
## 7     12 2018      6
## 8      1 2019      7
## 9      2 2019      8
## 10     3 2019      9
```

SKU

```
sku_id <- datos%>%select(sku)%>%unique()%>%arrange()
head(sku_id)
```

```
##      sku
## 1 N.ISE32GR
## 2 N.IP732B
## 3 N.ISE32PTB
## 4 N.ISE32GRB
## 5 N.ISE32PT
## 6 N.IP732S
```

```
sku_id$sku_id <- as.character(seq.int(nrow(sku_id)))
head(sku_id)
```

```
##      sku sku_id
## 1 N.ISE32GR      1
## 2 N.IP732B      2
## 3 N.ISE32PTB     3
## 4 N.ISE32GRB     4
## 5 N.ISE32PT      5
## 6 N.IP732S      6
```

Marca no - solo hay 1

```
datos%>%select(marca)%>%unique()
```

```
##      marca
## 1 apple
```

Gamma no - solo hay una gamma asignada a cada producto

```
head(datos[,c("gamma", "sku")] %>%
  mutate(contador = 1) %>%
  group_by(gamma, sku) %>%
  summarise(repeticiones = sum(contador)))
```

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

```
## # A tibble: 6 x 3
## # Groups:   gamma [1]
##   gamma sku      repeticiones
##   <chr> <chr>          <dbl>
## 1 baja  N.IPADP932G          1
## 2 baja  N.IPH6S64G           1
## 3 baja  N.IPHSE16D           1
## 4 baja  N.IPHSE16G           4
## 5 baja  N.IPHSE16P           7
## 6 baja  N.IPHSE16R           4
```

Costo promedio no - No es variable cualitativa Extra: El costo promedio esta implícito en la gamma.

costo_promedio <= 5000: "baja"

costo_promedio > 5000 & costo_promedio<=10000: "media"

costo_promedio > 10000 & costo_promedio<=15000: "alta"

costo_promedio > 15000: "premium"

Zona no - El punto de venta tiene implícita esta información Estado no - El punto de venta tiene implícita esta información Ciudad no - El punto de venta tiene implícita esta información Latitud y longitud no - El punto de venta tiene implícita esta información Ventas_totales - No es una variable cualitativa que se pueda indexar Paso 2. En datos agregar nuevas columnas con índices

```
dim(datos)
```

```
## [1] 19890    14
```

```
datos <- left_join(datos, pdv_id, by="punto_de_venta")
head(datos)
```

```
##   punto_de_venta    fecha mes anio num_ventas      sku marca gamma
## 1  5 de mayo zmm 0009-06-20   6 2018         1 N.ISE32GR apple baja
## 2  5 de mayo zmm 0020-12-20  12 2018         1 N.IP732B apple media
## 3  5 de mayo zmm 0001-02-20   2 2019         1 N.IP732B apple media
## 4  5 de mayo zmm 0021-02-20   2 2019         1 N.ISE32GR apple baja
## 5      acayucan 0015-08-20   8 2018         1 N.ISE32PTB apple baja
## 6      acayucan 0029-08-20   8 2018         1 N.ISE32GRB apple baja
##   costo_promedio      zona    estado ciudad latitud longitud pdv_id
## 1      4641.083 centro occidente michoacan zamora 19.98131 -102.2833    1
## 2      8855.867 centro occidente michoacan zamora 19.98131 -102.2833    1
## 3      8855.867 centro occidente michoacan zamora 19.98131 -102.2833    1
## 4      4641.083 centro occidente michoacan zamora 19.98131 -102.2833    1
## 5      4983.012 golfo de mexico veracruz oluta 17.93999 -94.9106     2
## 6      4976.971 golfo de mexico veracruz oluta 17.93999 -94.9106     2
```

```
datos <- left_join(datos, sku_id, by="sku")
datos <- left_join(datos, mes_id, by=c("mes", "anio"))
head(datos)
```

```
## punto_de_venta fecha mes anio num_ventas sku marca gamma
## 1 5 de mayo zmm 0009-06-20 6 2018 1 N.ISE32GR apple baja
## 2 5 de mayo zmm 0020-12-20 12 2018 1 N.IP732B apple media
## 3 5 de mayo zmm 0001-02-20 2 2019 1 N.IP732B apple media
## 4 5 de mayo zmm 0021-02-20 2 2019 1 N.ISE32GR apple baja
## 5 acayucan 0015-08-20 8 2018 1 N.ISE32PTB apple baja
## 6 acayucan 0029-08-20 8 2018 1 N.ISE32GRB apple baja
## costo_promedio zona estado ciudad latitud longitud pdv_id
## 1 4641.083 centro occidente michoacan zamora 19.98131 -102.2833 1
## 2 8855.867 centro occidente michoacan zamora 19.98131 -102.2833 1
## 3 8855.867 centro occidente michoacan zamora 19.98131 -102.2833 1
## 4 4641.083 centro occidente michoacan zamora 19.98131 -102.2833 1
## 5 4983.012 golfo de mexico veracruz oluta 17.93999 -94.9106 2
## 6 4976.971 golfo de mexico veracruz oluta 17.93999 -94.9106 2
## sku_id mes_id
## 1 1 0
## 2 2 6
## 3 2 8
## 4 1 8
## 5 3 2
## 6 4 2
```

Agrupar: Ventas totales

¿Hay más de una venta del mismo producto, en el mismo punto de venta, en la misma fecha (mes)?

Sugerencia:

Quitar variables con información adicional que esta implícita en: punto de venta,

```
datos <- datos %>%
  #quitamos fecha porque vamos a hacer el análisis por mes
  group_by(pdv_id, sku_id, mes_id)%>%
  summarise(ventas_totales = sum(num_ventas))
```

```
## 'summarise()' regrouping output by 'pdv_id', 'sku_id' (override with '.groups' argument)
```

```
head(datos)
```

```
## # A tibble: 6 x 4
## # Groups:   pdv_id, sku_id [4]
## pdv_id sku_id mes_id ventas_totales
## <chr> <chr> <chr> <int>
## 1 1 1 0 1
```

```
## 2 1      1      8      1
## 3 1      2      6      1
## 4 1      2      8      1
## 5 10     1      7      1
## 6 10     11     1      1
```

```
dim(datos)
```

```
## [1] 11777      4
```

```
names(datos)
```

```
## [1] "pdv_id"      "sku_id"      "mes_id"      "ventas_totales"
```

3. Completar Series de Tiempo

Construimos 3 conjuntos nuevos con índices

```
nrow(mes_id)
```

```
## [1] 10
```

```
nrow(sku_id)
```

```
## [1] 41
```

```
nrow(pdv_id)
```

```
## [1] 1293
```

```
#Combinaciones
```

```
nrow(mes_id)*nrow(sku_id)*nrow(pdv_id)
```

```
## [1] 530130
```

```
#Crear dataset con combinaciones
```

```
datos_completos <- merge(mes_id, sku_id)
```

```
datos_completos <- merge(datos_completos, pdv_id)
```

```
head(datos_completos)
```

```
##   mes anio mes_id      sku sku_id punto_de_venta pdv_id
## 1   6 2018      0 N.ISE32GR      1 5 de mayo zmm      1
## 2   7 2018      1 N.ISE32GR      1 5 de mayo zmm      1
## 3   8 2018      2 N.ISE32GR      1 5 de mayo zmm      1
## 4   9 2018      3 N.ISE32GR      1 5 de mayo zmm      1
## 5  10 2018      4 N.ISE32GR      1 5 de mayo zmm      1
## 6  11 2018      5 N.ISE32GR      1 5 de mayo zmm      1
```

```
datos_completos <- datos_completos %>% select(pdv_id, mes_id, sku_id)
head(datos_completos)
```

```
##   pdv_id mes_id sku_id
## 1      1      0      1
## 2      1      1      1
## 3      1      2      1
## 4      1      3      1
## 5      1      4      1
## 6      1      5      1
```

```
datos_completos <- left_join(datos_completos, datos, by=c("pdv_id", "mes_id", "sku_id"))
head(datos_completos)
```

```
##   pdv_id mes_id sku_id ventas_totales
## 1      1      0      1              1
## 2      1      1      1             NA
## 3      1      2      1             NA
## 4      1      3      1             NA
## 5      1      4      1             NA
## 6      1      5      1             NA
```

```
datos_completos[is.na(datos_completos)] <- 0
head(datos_completos)
```

```
##   pdv_id mes_id sku_id ventas_totales
## 1      1      0      1              1
## 2      1      1      1              0
## 3      1      2      1              0
## 4      1      3      1              0
## 5      1      4      1              0
## 6      1      5      1              0
```

4. Variable respuesta - Ventas del siguiente mes

EXTRA (más adelante): Remover el conjunto de datos correspondientes al índice de mes 9

¿Por qué?

Más adelante, se va a hablar de un concepto de Validación Cruzada para Series de Tiempo; sin embargo, en términos sencillos, cuando se aplica un modelo de aprendizaje de máquina, se necesitan 2 conjuntos (entrenamiento y prueba), si mi conjunto de entrenamiento llegara hasta el mes_id:8 y el mes_id:9 correspondiera a mi conjunto de prueba, este no lo puedo utilizar porque necesariamente, para evaluar el desempeño en este conjunto, es necesario contar con la variable y, la cual no tenemos.

```
library(dplyr)
datos_completos <- datos_completos %>%
  group_by(pdv_id, sku_id) %>%
  mutate(y_ventas_siguiente_mes = lead(ventas_totales, n = 1L, default = NA))
head(datos_completos)
```



```
## # A tibble: 6 x 5
## # Groups:   pdv_id, sku_id [1]
##   pdv_id mes_id sku_id ventas_totales y_ventas_siguiente_mes
##   <chr>  <chr>  <chr>          <dbl>          <dbl>
## 1 1      0      1              1              0
## 2 1      1      1              0              0
## 3 1      2      1              0              0
## 4 1      3      1              0              0
## 5 1      4      1              0              0
## 6 1      5      1              0              0
```

```
datos_completos <- datos_completos %>% filter(mes_id <= 8)
head(datos_completos)
```

```
## # A tibble: 6 x 5
## # Groups:   pdv_id, sku_id [1]
##   pdv_id mes_id sku_id ventas_totales y_ventas_siguiente_mes
##   <chr>  <chr>  <chr>          <dbl>          <dbl>
## 1 1      0      1              1              0
## 2 1      1      1              0              0
## 3 1      2      1              0              0
## 4 1      3      1              0              0
## 5 1      4      1              0              0
## 6 1      5      1              0              0
```

#5. Crear nuevas características: Conteos, promedios y rezagos ## Paso 1. Creamos las características de ventas promedio por mes, tienda y producto y ventas totales con las cuáles se crean las características que necesitamos de manera rezagada más adelante.

```
#conteos y promedios por duplas de características
```

```
mes_pdv <- datos_completos %>%
  group_by(mes_id, pdv_id) %>%
  summarise(ventas_totales_en_tienda_de_cada_mes = sum(ventas_totales),
            ventas_promedio_en_tienda_de_cada_mes = mean(ventas_totales))
```

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

```
mes_pdv
```

```
## # A tibble: 11,637 x 4
## # Groups:   mes_id [9]
##   mes_id pdv_id ventas_totales_en_tienda_de_ca~ ventas_promedio_en_tienda_de_c~
##   <chr>  <chr>          <dbl>          <dbl>
## 1 0      1              1              0.0244
## 2 0      10              2              0.0488
## 3 0      100             3              0.0732
## 4 0      1000            1              0.0244
## 5 0      1001            0              0
## 6 0      1002            0              0
## 7 0      1003            8              0.195
## 8 0      1004            0              0
```

```
## 9 0      1005      17      0.415
## 10 0     1006      0      0
## # ... with 11,627 more rows
```

#En el mes ____ y en el punto de venta _____, se tuvieron _____ ventas totales y se obtuvo un promedio de ventas por duplas de características

```
mes_sku <- datos_completos %>%
  group_by(mes_id, sku_id) %>%
  summarise(ventas_totales_en_tienda_de_cada_sku = sum(ventas_totales),
            ventas_promedio_en_tienda_de_cada_sku = mean(ventas_totales))
```

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

```
mes_sku
```

```
## # A tibble: 369 x 4
## # Groups:   mes_id [9]
##   mes_id sku_id ventas_totales_en_tienda_de_cada_sku ventas_promedio_en_tienda_de_cada_sku
##   <chr>  <chr>                <dbl>                <dbl>
## 1 1 0      1                379                0.293
## 2 2 0      10                9                0.00696
## 3 3 0      11                12               0.00928
## 4 4 0      12                15               0.0116
## 5 5 0      13                20               0.0155
## 6 6 0      14                17               0.0131
## 7 7 0      15                3               0.00232
## 8 8 0      16                0                0
## 9 9 0      17                4               0.00309
## 10 10 0     18                8               0.00619
## # ... with 359 more rows
```

#En el mes ____ y en el punto de venta _____, se tuvieron _____ ventas totales y se obtuvo un promedio de ventas por duplas de características

Paso 2. Incluir variables en datos completos

```
datos_completos <- left_join(datos_completos, mes_pdv, by=c("mes_id", "pdv_id"))
datos_completos <- left_join(datos_completos, mes_sku, by=c("mes_id", "sku_id"))
head(datos_completos)
```

```
## # A tibble: 6 x 9
## # Groups:   pdv_id, sku_id [1]
##   pdv_id mes_id sku_id ventas_totales y_ventas_sigue ventas_totales_~
##   <chr>  <chr>  <chr>                <dbl>                <dbl>                <dbl>
## 1 1 1      0      1                1                0                1
## 2 1 1      1      1                0                0                0
## 3 1 2      2      1                0                0                0
## 4 1 3      3      1                0                0                0
## 5 1 4      4      1                0                0                0
```

```
## 6 1      5      1      0      0      0
## # ... with 3 more variables: ventas_promedio_en_tienda_de_cada_mes <dbl>,
## #   ventas_totales_en_tienda_de_cada_sku <dbl>,
## #   ventas_promedio_en_tienda_de_cada_sku <dbl>
```

Paso 3. Crear Rezagos (3 tiempos)

```
datos_completos <- datos_completos %>%
  group_by(pdv_id, sku_id)%>%
  mutate(
    ventas_totales_1_mes_pasado = lag(ventas_totales, n=1),
    ventas_totales_2_meses_pasados = lag(ventas_totales, n=2),
    ventas_totales_3_meses_pasados = lag(ventas_totales, n=3),

    ventas_totales_tienda_y_mes_del_mes_pasado = lag(ventas_totales_en_tienda_de_cada_mes, n=1),
    ventas_totales_tienda_y_mes_2_pasado = lag(ventas_totales_en_tienda_de_cada_mes, n=2),
    ventas_totales_tienda_y_mes_3_pasado = lag(ventas_totales_en_tienda_de_cada_mes, n=3),

    ventas_promedio_tienda_y_mes_del_mes_pasado = lag(ventas_promedio_en_tienda_de_cada_mes, n=1),
    ventas_promedio_tienda_y_mes_2_pasado = lag(ventas_promedio_en_tienda_de_cada_mes, n=2),
    ventas_promedio_tienda_y_mes_3_pasado = lag(ventas_promedio_en_tienda_de_cada_mes, n=3),

    ventas_totales_tienda_y_sku_del_mes_pasado = lag(ventas_totales_en_tienda_de_cada_sku, n=1),
    ventas_totales_tienda_y_sku_2_pasado = lag(ventas_totales_en_tienda_de_cada_sku, n=2),
    ventas_totales_tienda_y_sku_3_pasado = lag(ventas_totales_en_tienda_de_cada_sku, n=3),

    ventas_promedio_tienda_y_sku_del_mes_pasado = lag(ventas_promedio_en_tienda_de_cada_sku, n=1),
    ventas_promedio_tienda_y_sku_2_pasado = lag(ventas_promedio_en_tienda_de_cada_sku, n=2),
    ventas_promedio_tienda_y_sku_3_pasado = lag(ventas_promedio_en_tienda_de_cada_sku, n=3))

head(datos_completos,20)
```

```
## # A tibble: 20 x 24
## # Groups:   pdv_id, sku_id [3]
##   pdv_id mes_id sku_id ventas_totales y_ventas_sigue~ ventas_totales_~
##   <chr> <chr> <chr> <dbl> <dbl> <dbl>
## 1 1      0      1      1      0      1
## 2 1      1      1      0      0      0
## 3 1      2      1      0      0      0
## 4 1      3      1      0      0      0
## 5 1      4      1      0      0      0
## 6 1      5      1      0      0      0
## 7 1      6      1      0      0      1
## 8 1      7      1      0      1      0
## 9 1      8      1      1      0      2
## 10 1      0      2      0      0      1
## 11 1      1      2      0      0      0
## 12 1      2      2      0      0      0
## 13 1      3      2      0      0      0
## 14 1      4      2      0      0      0
## 15 1      5      2      0      1      0
## 16 1      6      2      1      0      1
```

```
## 17 1      7      2      0      1      0
## 18 1      8      2      1      0      2
## 19 1      0      3      0      0      1
## 20 1      1      3      0      0      0
## # ... with 18 more variables: ventas_promedio_en_tienda_de_cada_mes <dbl>,
## #   ventas_totales_en_tienda_de_cada_sku <dbl>,
## #   ventas_promedio_en_tienda_de_cada_sku <dbl>,
## #   ventas_totales_1_mes_pasado <dbl>, ventas_totales_2_meses_pasados <dbl>,
## #   ventas_totales_3_meses_pasados <dbl>,
## #   ventas_totales_tienda_y_mes_del_mes_pasado <dbl>,
## #   ventas_totales_tienda_y_mes_2_pasado <dbl>,
## #   ventas_totales_tienda_y_mes_3_pasado <dbl>,
## #   ventas_promedio_tienda_y_mes_del_mes_pasado <dbl>,
## #   ventas_promedio_tienda_y_mes_2_pasado <dbl>,
## #   ventas_promedio_tienda_y_mes_3_pasado <dbl>,
## #   ventas_totales_tienda_y_sku_del_mes_pasado <dbl>,
## #   ventas_totales_tienda_y_sku_2_pasado <dbl>,
## #   ventas_totales_tienda_y_sku_3_pasado <dbl>,
## #   ventas_promedio_tienda_y_sku_del_mes_pasado <dbl>,
## #   ventas_promedio_tienda_y_sku_2_pasado <dbl>,
## #   ventas_promedio_tienda_y_sku_3_pasado <dbl>
```

Paso 4. NA - Rezagos

```
library(zoo)
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric
```

```
datos_completos<-na.locf(datos_completos, fromLast = TRUE)
head(datos_completos)
```

```
## # A tibble: 6 x 24
## # Groups:   pdv_id, sku_id [1]
##   pdv_id mes_id sku_id ventas_totales y_ventas_sigue~ ventas_totales_~
##   <chr>   <chr>   <chr>         <dbl>         <dbl>         <dbl>
## 1 1      0      1          1          0          1
## 2 1      1      1          0          0          0
## 3 1      2      1          0          0          0
## 4 1      3      1          0          0          0
## 5 1      4      1          0          0          0
## 6 1      5      1          0          0          0
## # ... with 18 more variables: ventas_promedio_en_tienda_de_cada_mes <dbl>,
## #   ventas_totales_en_tienda_de_cada_sku <dbl>,
## #   ventas_promedio_en_tienda_de_cada_sku <dbl>,
## #   ventas_totales_1_mes_pasado <dbl>, ventas_totales_2_meses_pasados <dbl>,
## #   ventas_totales_3_meses_pasados <dbl>,
```

```
## #   ventas_totales_tienda_y_mes_del_mes_pasado <dbl>,  
## #   ventas_totales_tienda_y_mes_2_pasado <dbl>,  
## #   ventas_totales_tienda_y_mes_3_pasado <dbl>,  
## #   ventas_promedio_tienda_y_mes_del_mes_pasado <dbl>,  
## #   ventas_promedio_tienda_y_mes_2_pasado <dbl>,  
## #   ventas_promedio_tienda_y_mes_3_pasado <dbl>,  
## #   ventas_totales_tienda_y_sku_del_mes_pasado <dbl>,  
## #   ventas_totales_tienda_y_sku_2_pasado <dbl>,  
## #   ventas_totales_tienda_y_sku_3_pasado <dbl>,  
## #   ventas_promedio_tienda_y_sku_del_mes_pasado <dbl>,  
## #   ventas_promedio_tienda_y_sku_2_pasado <dbl>,  
## #   ventas_promedio_tienda_y_sku_3_pasado <dbl>
```

```
#write.csv(datos_completos, file="E:/The Last Dance/Lab Diseño y Opt/equipo_5_apple_1_datos_completos.c
```