

Introducción a Tidyverse para Manipulación de Datos I

2023-05-13

Seleccionar columnas

`select()`: Se utiliza para seleccionar columnas específicas de un data frame.

```
##Importar la libreria tidyverse y lubridate para manipulacion de fechas  
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --  
## v ggplot2 3.4.0      v purrr  1.0.1  
## v tibble  3.1.8      v dplyr  1.1.0  
## v tidyr   1.3.0      v stringr 1.5.0  
## v readr   2.1.3      v forcats 1.0.0  
## -- Conflicts ----- tidyverse_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
##  
## Attaching package: 'lubridate'  
##  
## The following objects are masked from 'package:base':  
##  
##    date, intersect, setdiff, union
```

```
##Importacion de datos  
letras_bc <- read.table("data/letras_bc_consolidado_clean.csv")
```

```
##Seleccionando Columnas  
letras_bc_selected_cols <- letras_bc |>  
  select(FechadeSubasta, FechaLiquidacion, MontoAdjudicado)  
glimpse(letras_bc)
```

```
## Rows: 536  
## Columns: 6  
## $ FechadeSubasta   <chr> "2007-04-03", "2010-02-17", "2010-02-24", "2023-04-12~  
## $ FechaLiquidacion <chr> "2007-04-04", "2010-02-02", "2010-02-02", "2023-04-04~  
## $ MontoSubastado   <int> 400, 300, 500, 5000, 2500, 2500, 500, 800, 400, 750, ~  
## $ MontoDemandado   <dbl> 2281.80, 1088.00, 575.70, 9955.15, 3635.00, 3823.00, ~  
## $ MontoAdjudicado  <dbl> 400.00, 300.00, 500.00, 7951.91, 1515.00, 2025.00, 50~  
## $ RendimientoPPA  <dbl> 0.09570000, 0.05225600, 0.05243900, 0.12305392, 0.072~
```

Filtering

`filter()`: Permite filtrar filas de un data frame en base a una condición.

```
##Filtrando datasets
letras_monto_subastado_5000 <- letras_bc |>
  select(FechadeSubasta, FechaLiquidacion, MontoAdjudicado) |>
  filter(MontoAdjudicado==5000)

glimpse(letras_monto_subastado_5000)
```

```
## Rows: 1
## Columns: 3
## $ FechadeSubasta   <chr> "2021-06-30"
## $ FechaLiquidacion <chr> "2021-07-02"
## $ MontoAdjudicado  <dbl> 5000
```

Modificación de columnas (mutate)

`mutate()`: Permite agregar nuevas columnas a un data frame o modificar las existentes.

```
letras_monto_subastado_5000 <- letras_bc |>
  select(FechadeSubasta, FechaLiquidacion, MontoAdjudicado, RendimientoPPA) |>
  filter(MontoAdjudicado==5000) %>%
  mutate(RendimientoPPA= RendimientoPPA*100,
         RendimientoPPA= 1-RendimientoPPA
        )

glimpse(letras_monto_subastado_5000)
```

```
## Rows: 1
## Columns: 4
## $ FechadeSubasta   <chr> "2021-06-30"
## $ FechaLiquidacion <chr> "2021-07-02"
## $ MontoAdjudicado  <dbl> 5000
## $ RendimientoPPA   <dbl> -2.924179
```

- `group_by()`: Agrupa las filas de un data frame por una o más columnas, lo que permite realizar operaciones sobre cada grupo por separado
- `summarise()`: Permite resumir los datos de un data frame, calculando estadísticas descriptivas o aplicando funciones personalizadas.

```
letras_bc |>
  select(FechadeSubasta, MontoDemandado, RendimientoPPA) |>
  filter(FechadeSubasta> as.Date('2022-12-31')) |>
  mutate(mesFechaSubasta= month(FechadeSubasta)) |>
  group_by(mesFechaSubasta) |>
  summarise(
    PromedioTasa= mean(RendimientoPPA),
    TotalSubastado= sum(MontoDemandado)) |>
  arrange(mesFechaSubasta)
```

```
## # A tibble: 5 x 3
##   mesFechaSubasta PromedioTasa TotalSubastado
##           <dbl>         <dbl>         <dbl>
## 1             1         0.118         12183.
## 2             2         0.122         29263.
## 3             3         0.121         11985.
## 4             4         0.123         32166.
## 5             5         0.119         10624.
```

`arrange()`: Ordena las filas de un data frame en base a una o más columnas.

```
letras_bc |>
  select(FechadeSubasta, MontoDemandado, RendimientoPPA) |>
  filter(FechadeSubasta > as.Date('2022-12-31')) |>
  mutate(mesFechaSubasta= month(FechadeSubasta)) |>
  group_by(mesFechaSubasta) |>
  summarise(
    PromedioTasa= mean(RendimientoPPA),
    TotalSubastado= sum(MontoDemandado)) |>
  arrange(desc(mesFechaSubasta))
```

```
## # A tibble: 5 x 3
##   mesFechaSubasta PromedioTasa TotalSubastado
##           <dbl>         <dbl>         <dbl>
## 1             5         0.119         10624.
## 2             4         0.123         32166.
## 3             3         0.121         11985.
## 4             2         0.122         29263.
## 5             1         0.118         12183.
```