

Reporte

Juan Cadenas, Sol Astengo, Gabriel Bermudez

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.5.1      v tibble     3.2.1
v lubridate  1.9.3      v tidyr      1.3.1
v purrr      1.0.2
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
library(here)
```

here() starts at C:/Users/asten/cdcR/Proyecto

```
library(readxl)
```

Introducción

En este proyecto, analizamos una base de datos de ingresos y egresos de personas al país. Nuestro objetivo es responder varias preguntas específicas relacionadas con los motivos de ingreso, nacionalidades, y gastos de los individuos. Este reporte incluye un análisis exploratorio inicial de los datos.

Preguntas planteadas

- Distribución de gastos (por persona) según motivo de ingreso al país.
- ¿Con que motivo ingresa al país según nacionalidad del grupo?.
- De qué nacionalidades son las personas que ingresan al país (mapa).
- Gasto promedio diario por persona segun motivo, nivel educativo, transporte por el que ingresa al país.
 - Variable de respuesta: Gasto promedio diario (por persona).
 - Variables explicativas: motivo, nivel educativo, transporte por el que ingresa al país.

Datos

```
receptivo <- read_excel(here::here("receptivo.xlsx"))
diccionario <- read_excel(here::here("Diccionario.xlsx"))

# Mostrar una vista preliminar de los datos
View(receptivo)
View(diccionario)
```

Análisis Exploratorio de Datos

Cantidad de observaciones por hoteles

```
receptivo |>
  group_by(Alojamiento) |>
  summarise(cant. = n()) |>
  arrange(desc(cant.)) |>
  filter(str_starts(string = Alojamiento, pattern = "Hotel")) |>
  print(n = 22)
```

```
# A tibble: 6 x 2
  Alojamiento      cant.
  <chr>          <int>
1 Hotel 3 estrellas 8623
```

2	Hotel sin categorizar	7136
3	Hotel 4 estrellas	4417
4	Hotel 1 y 2 estrellas	1430
5	Hotel/Albergue	1332
6	Hotel 5 estrellas	712

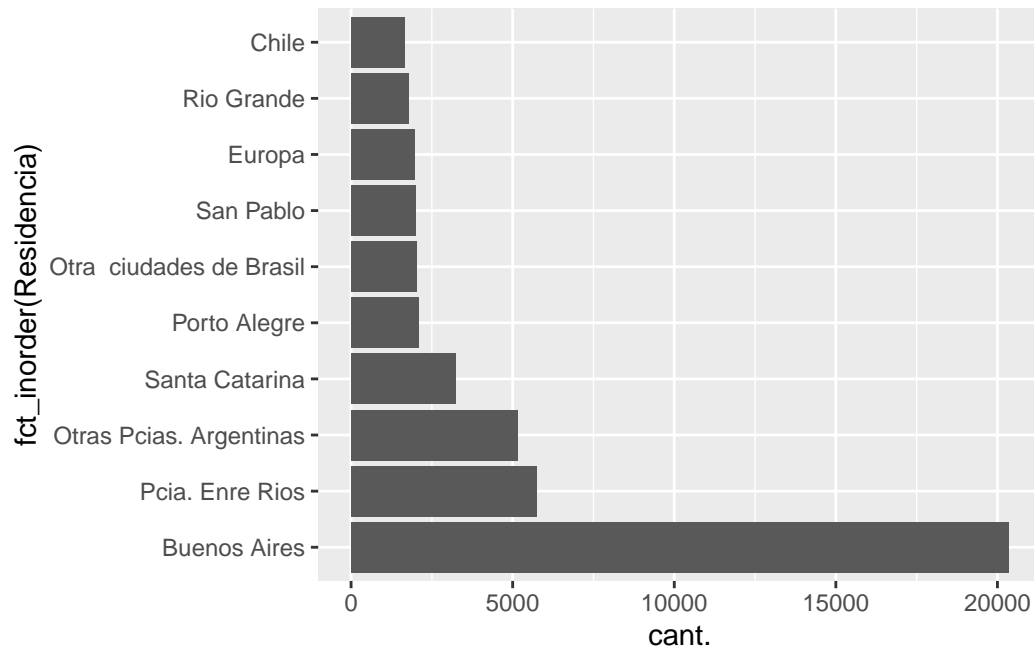
Distribución de la RESIDENCIA de las personas que ingresaron.

```
receptivo |>
  group_by(Residencia) |>
  summarise(cant. = n()) |>
  arrange(desc(cant.))
```

```
# A tibble: 32 x 2
  Residencia      cant.
  <chr>          <int>
1 Buenos Aires  20343
2 Pcia. Enre Rios  5749
3 Otras Pcias. Argentinas  5143
4 Santa Catarina  3225
5 Porto Alegre   2080
6 Otra ciudades de Brasil  2040
7 San Pablo      2008
8 Europa         1970
9 Rio Grande     1788
10 Chile         1662
# i 22 more rows
```

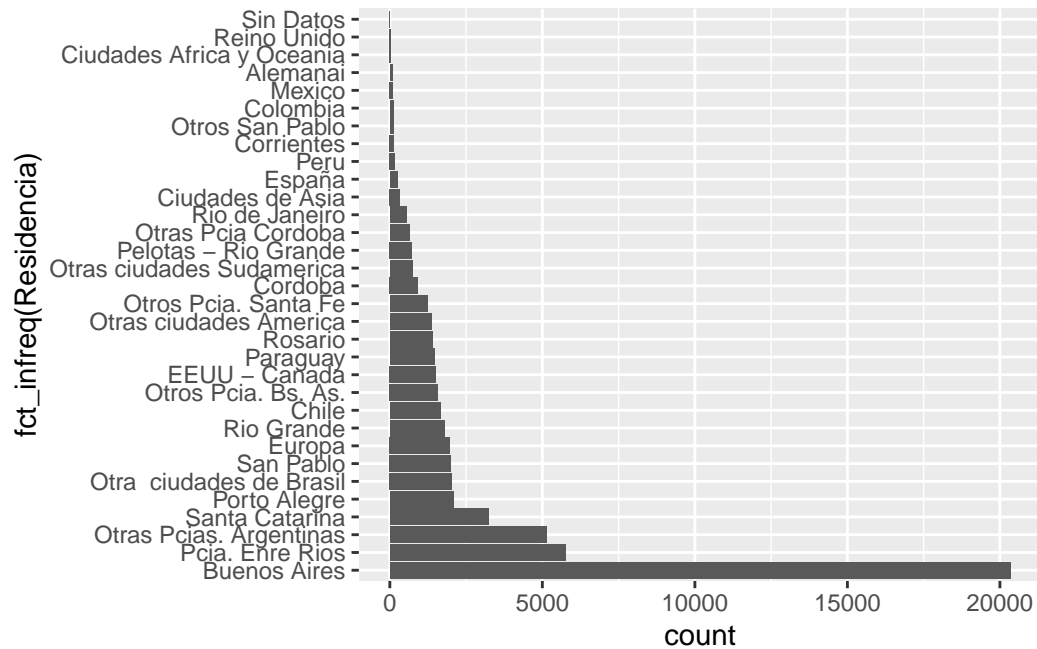
Gráfica de las 10 principales residencias

```
ggplot(
  receptivo |>
  group_by(Residencia) |>
  summarise(cant. = n()) |>
  arrange(desc(cant.)) |>
  slice_max(cant., n=10)
) +
  geom_col(aes(fct_inorder(Residencia), cant.)) +
  coord_flip()
```



Gráfica de la distribución de residencia

```
ggplot(receptivo) +  
  geom_bar(aes(fct_infreq(Residencia))) +  
  coord_flip()
```



Distribución del MOTIVO de ingreso

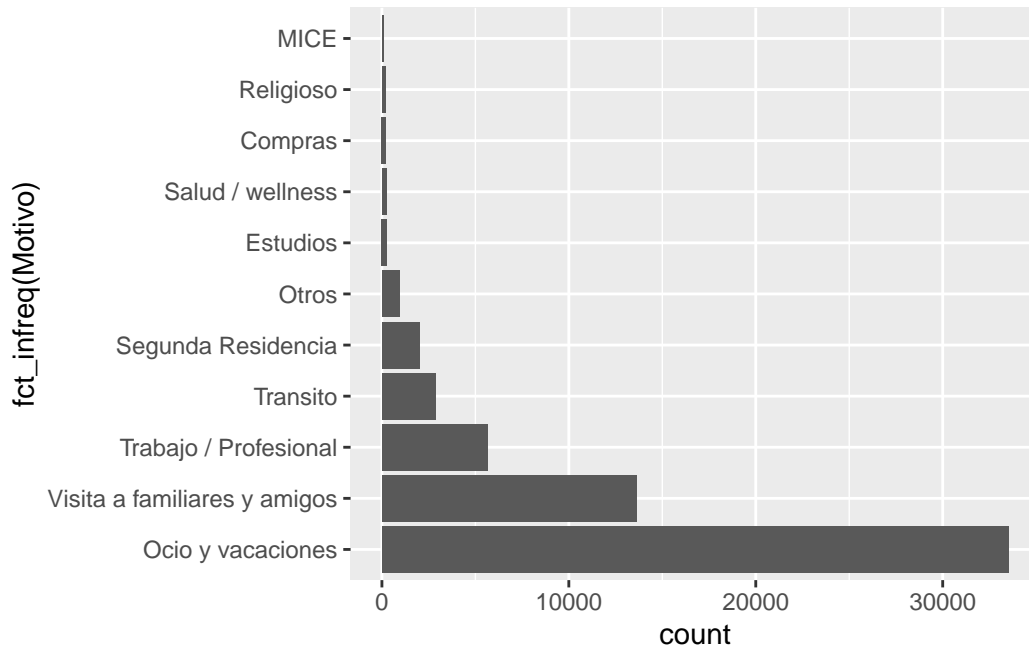
```
receptivo |>
  group_by(Motivo) |>
  summarise(cant. = n()) |>
  arrange(desc(cant.))
```

A tibble: 11 x 2

	Motivo	cant.
	<chr>	<int>
1	Ocio y vacaciones	33498
2	Visita a familiares y amigos	13597
3	Trabajo / Profesional	5631
4	Transito	2876
5	Segunda Residencia	2008
6	Otros	925
7	Estudios	276
8	Salud / wellness	245
9	Compras	223
10	Religioso	176
11	MICE	75

Gráfica de la distribución del motivo de ingreso

```
ggplot(receptivo) +  
  geom_bar(aes(fct_infreq(Motivo))) +  
  coord_flip()
```

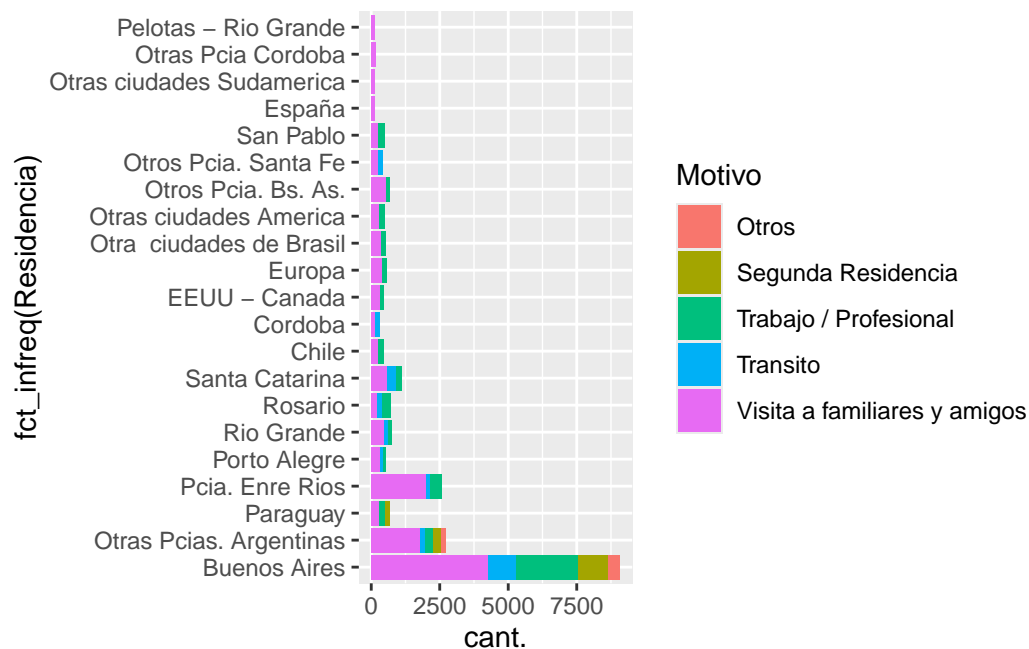


MOTIVO de ingreso según RESIDENCIA

```
receptivo |>  
  select(Motivo, Residencia) |>  
  group_by(Motivo, Residencia) |>  
  summarise(cant. = n()) |>  
  arrange(desc(cant.)) |>  
  filter(cant. >= 100 & Motivo != "Ocio y vacaciones") |>  
  ggplot() +  
  geom_col(aes(  
    x = fct_infreq(Residencia),  
    y = cant.,  
    fill = Motivo  
  )) +
```

```
coord_flip()
```

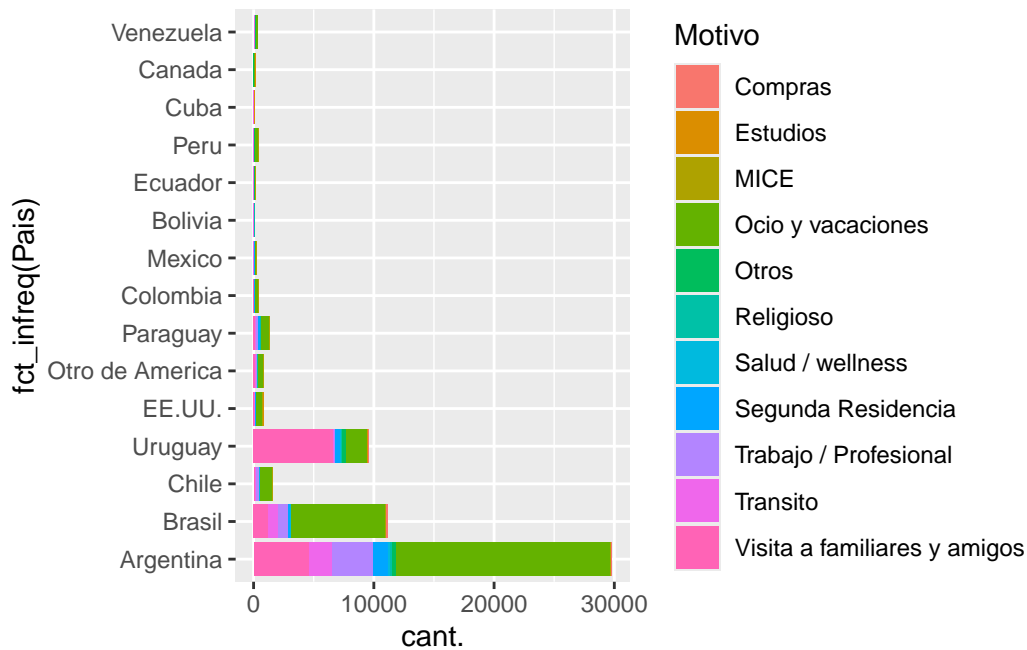
`summarise()` has grouped output by 'Motivo'. You can override using the `.groups` argument.



MOTIVO según NACIONALIDAD, sólo quienes ingresan desde américa

```
receptivo |>
  select(Motivo, Pais, IdNacionalidad) |>
  group_by(Motivo, Pais, IdNacionalidad) |>
  summarise(cant. = n()) |>
  arrange(desc(cant.)) |>
  filter(!(IdNacionalidad %in% c(51:57, 59, 61:63, 69, 79))) |>
  ggplot() +
  geom_col(aes(
    x = fct_infreq(Pais),
    y = cant.,
    fill = Motivo
  )) +
  coord_flip()
```

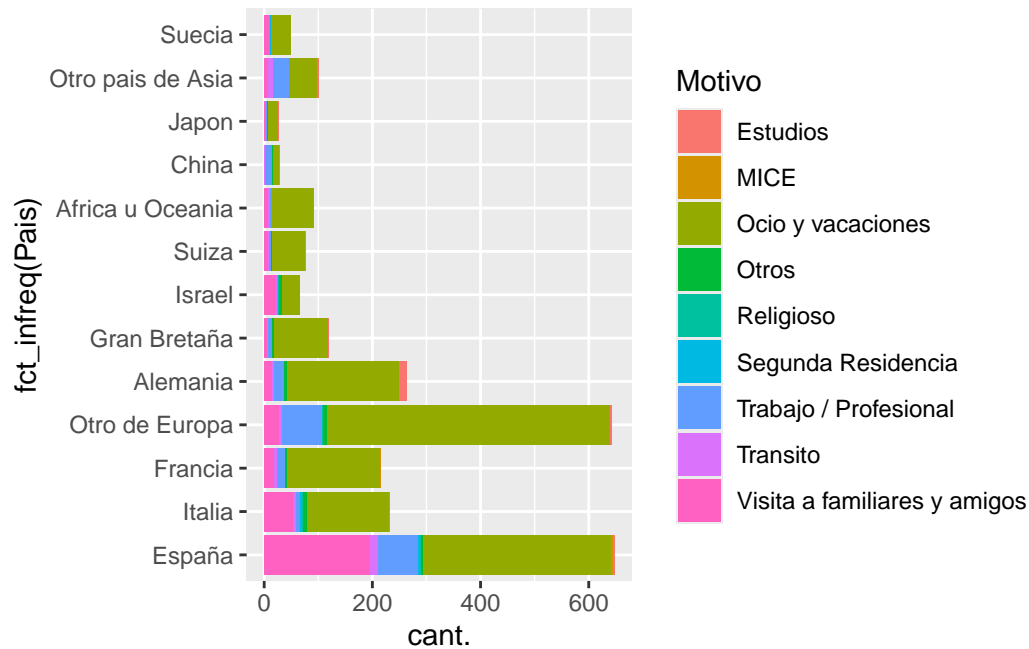
`summarise()` has grouped output by 'Motivo', 'Pais'. You can override using the `.groups` argument.



MOTIVO según NACIONALIDAD, sólo quienes ingresan desde afuera de américa

```
receptivo |>
  select(Motivo, Pais, IdNacionalidad) |>
  group_by(Motivo, Pais, IdNacionalidad) |>
  summarise(cant. = n()) |>
  arrange(desc(cant.)) |>
  filter(IdNacionalidad %in% c(51:57, 59, 61:63, 69, 79)) |>
  ggplot() +
  geom_col(aes(
    x = fct_infreq(Pais),
    y = cant.,
    fill = Motivo
  )) +
  coord_flip()
```

`summarise()` has grouped output by 'Motivo', 'Pais'. You can override using the `.groups` argument.



Nacionalidades de las personas que ingresan al país

```
receptivo |>
  select(IdNacionalidad, Pais) |>
  group_by(IdNacionalidad, Pais) |>
  summarise(cant. = n()) |>
  arrange(desc(cant.))
```

`summarise()` has grouped output by 'IdNacionalidad'. You can override using the `groups` argument.

```
# A tibble: 28 x 3
# Groups:   IdNacionalidad [28]
  IdNacionalidad Pais      cant.
      <dbl> <chr>      <int>
1          10 Argentina 29832
2          19 Brasil   11168
3           1 Uruguay   9522
4          40 Chile    1561
5          30 Paraguay  1310
6          49 Otro de America 818
```

```

7           70 EE.UU.           804
8           51 España           648
9           59 Otro de Europa   642
10          35 Peru             424
# i 18 more rows

```

Nota:

IdNacionalidad, de 50 para abajo es america (sur, central y norte) y 50 para arriba es Europa, Asia, Oceania (resto del mundo).

```

receptivo |>
  select(IdNacionalidad, Pais) |>
  group_by(IdNacionalidad, Pais) |>
  summarise(cant. = n()) |>
  arrange(desc(cant.))

```

`summarise()` has grouped output by 'IdNacionalidad'. You can override using the `.groups` argument.

```

# A tibble: 28 x 3
# Groups:   IdNacionalidad [28]
  IdNacionalidad Pais      cant.
      <dbl> <chr>      <int>
1           10 Argentina  29832
2           19 Brasil    11168
3            1 Uruguay    9522
4           40 Chile     1561
5           30 Paraguay  1310
6           49 Otro de America  818
7           70 EE.UU.     804
8           51 España     648
9           59 Otro de Europa  642
10          35 Peru      424
# i 18 more rows

```

Obs.: Más gente que observaciones

NACIONALIDAD y RESIDENCIA la tomamos como la del grupo, o del que contesta la encuesta?

Número total de personas

```
receptivo |>
  select(Gente) |>
  sum()
```

```
[1] 144425
```

```
receptivo |>
  select(Gente) |>
  count()
```

```
# A tibble: 1 x 1
```

```
  n
```

```
<int>
```

```
1 59530
```

Cambio de fecha de ingreso/egreso por año y mes

```
# Separo FechaIngreso en anio, mes y día:
```

```
receptivo_fecha <-
  receptivo |>
  separate(
    col = FechaIngreso,
    into = c("anio_I", "mes_I", "dia_I"),
    convert = TRUE
  )
```

```
# Separo FechaEgreso en anio, mes y día:
```

```
receptivo_fecha <-
  receptivo_fecha |>
  separate(
    col = FechaEgreso,
    into = c("anio_E", "mes_E", "dia_E"),
    convert = TRUE
  )
```

```
# Creo Variable aniomes:
```

```
receptivo_fecha <-
```

```
receptivo_fecha |>
mutate(
  aniomes_I = anio_I*100 + mes_I,
  aniomes_E = anio_E*100 + mes_E,
  .before = anio_I
)

receptivo_fecha |>
select(aniomes_I:IdFecEgr)
```

A tibble: 59,530 x 10

	aniomes_I	aniomes_E	anio_I	mes_I	dia_I	IdFecIng	anio_E	mes_E	dia_E	IdFecEgr
	<dbl>	<dbl>	<int>	<int>	<int>	<dbl>	<int>	<int>	<int>	<dbl>
1	201702	201703	2017	2	22	13567	2017	3	3	13576
2	201701	201701	2017	1	5	13519	2017	1	9	13523
3	201701	201701	2017	1	4	13518	2017	1	12	13526
4	201701	201701	2017	1	18	13532	2017	1	26	13540
5	201701	201701	2017	1	2	13516	2017	1	17	13531
6	201701	201701	2017	1	9	13523	2017	1	11	13525
7	201701	201701	2017	1	20	13534	2017	1	27	13541
8	201701	201701	2017	1	14	13528	2017	1	16	13530
9	201701	201701	2017	1	20	13534	2017	1	26	13540
10	201701	201701	2017	1	26	13540	2017	1	30	13544

i 59,520 more rows

Nota sobre fechas:

Se identificó un posible error en la fecha de ingreso del 2011-12-27, con una fecha de egreso 8 años después. Se supone que en vez de 2011, debería ser 2018.

Cálculo de los gastos por persona

```
receptivo_fecha <-
receptivo_fecha |>
mutate(across(
  .cols = starts_with("Gasto"),
  .fns = ~ round((.x/Gente), 2),
  .names = "{.col}_porPersona"
)) |>
```

```
relocate(IdIngresos:Gente, GastoTotal, GastoTotal_porPersona, GastoAlojamiento, GastoAlo
```

```
receptivo_fecha |> head()
```

```
# A tibble: 6 x 62
  IdIngresos `Lugar Ingreso` IdTranspIngreso Transporte Internaci~1 aniomes_I
    <dbl> <chr>                <dbl> <chr>                <dbl>
1         3 Aeropuerto de Car~         1 Aereo                201702
2        18 Río Branco                2 Terrestre Auto        201701
3         3 Aeropuerto de Car~         1 Aereo                201701
4         7 Fray Bentos                2 Terrestre Auto        201701
5         7 Fray Bentos                2 Terrestre Auto        201701
6         2 Puerto de monte~         5 Maritimo - Fluvial    201701
# i abbreviated name: 1: `Transporte Internacional de Ingreso`
# i 57 more variables: aniomes_E <dbl>, anio_I <int>, mes_I <int>, dia_I <int>,
#   IdFecIng <dbl>, anio_E <int>, mes_E <int>, dia_E <int>, IdFecEgr <dbl>,
#   IdNacionalidad <dbl>, Pais <chr>, IdResidencia <dbl>, Residencia <chr>,
#   IdMotivo <dbl>, Motivo <chr>, IdOcupacion <dbl>, Ocupacion <chr>,
#   IsEstudio <dbl>, Estudio <chr>, IdDestinoLocalidad <dbl>, Localidad <chr>,
#   IdDepartamentoDestino <dbl>, Departamento <chr>, ...
```