Desafio11

```
reticulate::py_install("polars")
```

```
Using virtual environment "C:/Users/rseit/OneDrive/Documentos/.virtualenvs/r-reticulate" ...

+ "C:/Users/rseit/OneDrive/Documentos/.virtualenvs/r-reticulate/Scripts/python.exe" -m pip instal
```

```
reticulate::py_install("pyarrow")
```

```
Using virtual environment "C:/Users/rseit/OneDrive/Documentos/.virtualenvs/r-reticulate" ...

+ "C:/Users/rseit/OneDrive/Documentos/.virtualenvs/r-reticulate/Scripts/python.exe" -m pip instal
```

```
# Registro da data e hora de compilação
cat("Arquivo compilado em:", format(Sys.time(), "%d/%m/%Y às %H:%M:%S"), "\n")
```

```
Arquivo compilado em: 07/10/2025 às 10:27:27
```

```
library(reticulate)
```

```
Warning: pacote 'reticulate' foi compilado no R versão 4.4.3
```

```
pl <- import("polars")
```

```
# Nomes das colunas
colunas <- c(
  "age", "workclass", "fnlwgt", "education", "education_num",
  "marital_status", "occupation", "relationship", "race", "sex",
  "capital_gain", "capital_loss", "hours_per_week", "native_country", "income"
)

# Ler CSV sem cabeçalho, com valores faltantes representados por "?"
renda <- pl$read_csv(
  "renda_adulta.csv",
  has_header = FALSE,
  new_columns = colunas,
  null_values = "?"
)
```

```
renda$head()
```

```
shape: (5, 15)
┌─────┬─────┬─────┬─────┬─────┬─────┬─────┬─────┬─────┬─────┬─────┬─────┬─────┬─────┬─────┐
│ age ┆ wor ┆ fnl ┆ edu ┆ edu ┆ mar ┆ occ ┆ rel ┆ rac ┆ sex ┆ cap ┆ capita ┆ hours ┆ nativ ┆ inco │
│ --- ┆ kcl ┆ wgt ┆ cat ┆ cat ┆ ita ┆ upa ┆ ati ┆ e   ┆ --- ┆ ita ┆ l_loss ┆ _per_ ┆ e_cou ┆ e    │
│ i64 ┆ ass ┆ --- ┆ ion ┆ ion ┆ l_s ┆ tio ┆ ons ┆ --- ┆ str ┆ l_g ┆ ---    ┆ week  ┆ ntry  ┆ ---  │
│     ┆ --- ┆ i64 ┆ --- ┆ _nu ┆ tat ┆ n   ┆ hip ┆ str ┆     ┆ ain ┆ i64    ┆ ---   ┆ ---   ┆ str  │
│     ┆ str ┆     ┆ str ┆ m   ┆ us  ┆ --- ┆ --- ┆     ┆     ┆ --- ┆        ┆ i64   ┆ str   ┆      │
│     ┆     ┆     ┆     ┆ --- ┆ --- ┆ str ┆ str ┆     ┆     ┆ i64 ┆        ┆       ┆       ┆      │
│     ┆     ┆     ┆     ┆ i64 ┆ str ┆     ┆     ┆     ┆     ┆     ┆        ┆       ┆       ┆      │
╞═════╪═════╪═════╪═════╪═════╪═════╪═════╪═════╪═════╪═════╪═════╪════════╪═══════╪═══════╪══════╡
│ 39  ┆ Sta ┆ 775 ┆ Bac ┆ 13  ┆ Nev ┆ Adm ┆ Not ┆ Whi ┆ Mal ┆ 217 ┆ 0      ┆ 40    ┆ Unite ┆ <=50 │
│     ┆ te- ┆ 16  ┆ hel ┆     ┆ er- ┆ -cl ┆ -in ┆ te  ┆ e   ┆ 4   ┆        ┆       ┆ d-Sta ┆ K    │
│     ┆ gov ┆     ┆ ors ┆     ┆ mar ┆ eri ┆ -fa ┆     ┆     ┆     ┆        ┆       ┆ tes   ┆      │
│     ┆     ┆     ┆     ┆     ┆ rie ┆ cal ┆ mil ┆     ┆     ┆     ┆        ┆       ┆       ┆      │
│     ┆     ┆     ┆     ┆     ┆ d   ┆     ┆ y   ┆     ┆     ┆     ┆        ┆       ┆       ┆      │
│ 50  ┆ Sel ┆ 833 ┆ Bac ┆ 13  ┆ Mar ┆ Exe ┆ Hus ┆ Whi ┆ Mal ┆ 0   ┆ 0      ┆ 13    ┆ Unite ┆ <=50 │
│     ┆ f-a ┆ 11  ┆ hel ┆     ┆ rie ┆ c-m ┆ ban ┆ te  ┆ e   ┆     ┆        ┆       ┆ d-Sta ┆ K    │
│     ┆ mp- ┆     ┆ ors ┆     ┆ d-C ┆ ana ┆ d   ┆     ┆     ┆     ┆        ┆       ┆ tes   ┆      │
│     ┆ not ┆     ┆     ┆     ┆ iv- ┆ ger ┆     ┆     ┆     ┆     ┆        ┆       ┆       ┆      │
│     ┆ -in ┆     ┆     ┆     ┆ spo ┆ ial ┆     ┆     ┆     ┆     ┆        ┆       ┆       ┆      │
│     ┆ c   ┆     ┆     ┆     ┆ use ┆     ┆     ┆     ┆     ┆     ┆        ┆       ┆       ┆      │
│ 38  ┆ Pri ┆ 215 ┆ HS- ┆ 9   ┆ Div ┆ Han ┆ Not ┆ Whi ┆ Mal ┆ 0   ┆ 0      ┆ 40    ┆ Unite ┆ <=50 │
│     ┆ vat ┆ 646 ┆ gra ┆     ┆ orc ┆ dle ┆ -in ┆ te  ┆ e   ┆     ┆        ┆       ┆ d-Sta ┆ K    │
│     ┆ e   ┆     ┆ d   ┆     ┆ ed  ┆ rs- ┆ -fa ┆     ┆     ┆     ┆        ┆       ┆ tes   ┆      │
│     ┆     ┆     ┆     ┆     ┆     ┆ cle ┆ mil ┆     ┆     ┆     ┆        ┆       ┆       ┆      │
│     ┆     ┆     ┆     ┆     ┆     ┆ ane ┆ y   ┆     ┆     ┆     ┆        ┆       ┆       ┆      │
│     ┆     ┆     ┆     ┆     ┆     ┆ rs  ┆     ┆     ┆     ┆     ┆        ┆       ┆       ┆      │
│ 53  ┆ Pri ┆ 234 ┆ 11t ┆ 7   ┆ Mar ┆ Han ┆ Hus ┆ Bla ┆ Mal ┆ 0   ┆ 0      ┆ 40    ┆ Unite ┆ <=50 │
│     ┆ vat ┆ 721 ┆ h   ┆     ┆ rie ┆ dle ┆ ban ┆ ck  ┆ e   ┆     ┆        ┆       ┆ d-Sta ┆ K    │
│     ┆ e   ┆     ┆     ┆     ┆ d-C ┆ rs- ┆ d   ┆     ┆     ┆     ┆        ┆       ┆ tes   ┆      │
│     ┆     ┆     ┆     ┆     ┆ iv- ┆ cle ┆     ┆     ┆     ┆     ┆        ┆       ┆       ┆      │
│     ┆     ┆     ┆     ┆     ┆ spo ┆ ane ┆     ┆     ┆     ┆     ┆        ┆       ┆       ┆      │
│     ┆     ┆     ┆     ┆     ┆ use ┆ rs  ┆     ┆     ┆     ┆     ┆        ┆       ┆       ┆      │
│ 28  ┆ Pri ┆ 338 ┆ Bac ┆ 13  ┆ Mar ┆ Pro ┆ Wif ┆ Bla ┆ Fem ┆ 0   ┆ 0      ┆ 40    ┆ Cuba  ┆ <=50 │
│     ┆ vat ┆ 409 ┆ hel ┆     ┆ rie ┆ f-s ┆ e   ┆ ck  ┆ ale ┆     ┆        ┆       ┆       ┆ K    │
│     ┆ e   ┆     ┆ ors ┆     ┆ d-C ┆ pec ┆     ┆     ┆     ┆     ┆        ┆       ┆       ┆      │
│     ┆     ┆     ┆     ┆     ┆ iv- ┆ ial ┆     ┆     ┆     ┆     ┆        ┆       ┆       ┆      │
│     ┆     ┆     ┆     ┆     ┆ spo ┆ ty  ┆     ┆     ┆     ┆     ┆        ┆       ┆       ┆      │
│     ┆     ┆     ┆     ┆     ┆ use ┆     ┆     ┆     ┆     ┆     ┆        ┆       ┆       ┆      │
└─────┴─────┴─────┴─────┴─────┴─────┴─────┴─────┴─────┴─────┴─────┴────────┴───────┴───────┴──────┘
```

```
renda$dtypes
```

```
[[1]]
Int64

[[2]]
String

[[3]]
Int64

[[4]]
String

[[5]]
Int64

[[6]]
String

[[7]]
String

[[8]]
String

[[9]]
String

[[10]]
String

[[11]]
Int64

[[12]]
Int64

[[13]]
Int64

[[14]]
String

[[15]]
String
```

```
renda$shape
```

```
[[1]]
[1] 32561

[[2]]
[1] 15
```

```
renda$group_by("income")$count()
```

```
shape: (2, 2)
┌────────┬───────┐
│ income ┆ count │
│ ---    ┆ ---   │
│ str    ┆ u32   │
╞════════╪═══════╡
│ <=50K  ┆ 24720 │
│ >50K   ┆ 7841  │
└────────┴───────┘
```

```
# Converter para pandas primeiro
renda_df <- renda$select(
  pl$col("capital_gain"), pl$col("capital_loss")
)$to_pandas()

# Usar tidyr no R para pivotar
library(tidyr)
renda_longo <- renda_df |>
  pivot_longer(
    cols = everything(),
    names_to = "tipo",
    values_to = "valor"
  )

head(renda_longo)
```

```
# A tibble: 6 × 2
  tipo         valor
  <chr>        <dbl>
1 capital_gain  2174
2 capital_loss     0
3 capital_gain     0
4 capital_loss     0
5 capital_gain     0
6 capital_loss     0
```

```
renda$group_by("income")$agg(
  pl$col("hours_per_week")$mean()$alias("media_horas")
)
```

```
shape: (2, 2)
┌────────┬─────────────┐
│ income ┆ media_horas │
│ ---    ┆ ---         │
│ str    ┆ f64         │
╞════════╪═════════════╡
│ <=50K  ┆ 38.84021    │
│ >50K   ┆ 45.473026   │
└────────┴─────────────┘
```

```
renda$group_by("occupation")$count()
```

```
shape: (15, 2)
┌─────────────────┬───────┐
│ occupation      ┆ count │
│ ---             ┆ ---   │
│ str             ┆ u32   │
╞═════════════════╪═══════╡
│ Farming-fishing ┆ 994   │
│ Other-service   ┆ 3295  │
│ Armed-Forces    ┆ 9     │
│ Priv-house-serv ┆ 149   │
│ null            ┆ 1843  │
│ ?               ┆ 1843  │
│ Protective-serv ┆ 649   │
│ Tech-support    ┆ 928   │
│ Adm-clerical    ┆ 3770  │
│ Prof-specialty  ┆ 4140  │
│ Transport-moving┆ 1597  │
└─────────────────┴───────┘
```

```
library(ggplot2)

media_horas <- renda$group_by("income")$agg(
  pl$col("hours_per_week")$mean()$alias("media_horas")
)$to_pandas()

ggplot(media_horas, aes(x = income, y = media_horas, fill = income)) +
  geom_col() +
  labs(
    title = "Média de Horas Trabalhadas por Classe Salarial",
    x = "Classe Salarial",
    y = "Media de Horas Semanais"
  ) +
  theme_minimal()
```


```
renda_genero <- renda$group_by("sex", "income")$count()$to_pandas()

renda_genero <- renda_genero %>%
  dplyr::group_by(sex) %>%
  dplyr::mutate(proporcao = count / sum(count))

ggplot(renda_genero, aes(x = sex, y = proporcao, fill = income)) +
  geom_col(position = "dodge") +
  labs(
    title = "Proporção de Renda (>50k e <50k) por Gênero",
    x = "Sexo",
    y = "Proporção"
  ) +
  theme_minimal()
```