Desafio 3 - ME315

Bruce Trevisan

2025-09-01

```r
# Arquivo JSON

# Payment Card Fraud Detection 2025
# Real-time fraud prevention system for luxury retail transactions analysis 2025

# PACOTES NECESSARIOS PARA FAZER UPLOAD
library(arrow)
```

```
##
## Anexando pacote: 'arrow'
```

```
## O seguinte objeto é mascarado por 'package:utils':
##
##     timestamp
```

```r
library(tibble)
library(jsonlite)

# Ler o JSON
json_data <- fromJSON("luxury_cosmetics_fraud_analysis_2025.json")

str(json_data, max.level = 2) #Com esse comando mostra se é lista, matriz ou um dataframe disfarçado.
```

```
## 'data.frame':    2133 obs. of  16 variables:
##  $ Transaction_ID       : chr  "702bdd9b-9c93-41e3-9dbb-a849b2422080" "2e64c346-36bc-4acf-bc2b-8b0fd
##  $ Customer_ID          : chr  "119dca0b-8554-4b2d-9bec-e964eaf6af97" "299df086-26c4-4708-b6d7-fcaec
##  $ Transaction_Date     : num  1.75e+12 1.74e+12 1.74e+12 1.75e+12 1.74e+12 ...
##  $ Transaction_Time     : chr  "04:04:15" "20:23:23" "12:36:02" "19:09:43" ...
##  $ Customer_Age         : int  56 46 32 60 NA 38 56 36 40 28 ...
##  $ Customer_Loyalty_Tier: chr  "Silver" "Platinum" "Silver" "Bronze" ...
##  $ Location             : chr  "San Francisco" "Zurich" "Milan" "London" ...
##  $ Store_ID             : chr  "FLAGSHIP-LA" "BOUTIQUE-SHANGHAI" "POPUP-TOKYO" "BOUTIQUE-NYC" ...
##  $ Product_SKU          : chr  "NEBULA-SERUM-07" "STELLAR-FOUND-03" "SOLAR-BLUSH-04" "GALAXIA-SET-08
##  $ Product_Category     : chr  "Concealer" "Lipstick" "Mascara" "Serum" ...
##  $ Purchase_Amount      : num  158 86 256 283 206 ...
##  $ Payment_Method       : chr  "Mobile Payment" "Credit Card" "Gift Card" "Gift Card" ...
##  $ Device_Type          : chr  "Desktop" "Tablet" "Desktop" "Mobile" ...
##  $ IP_Address           : chr  "239.249.58.237" "84.49.227.90" "79.207.35.55" "176.194.167.253" ...
##  $ Fraud_Flag           : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Footfall_Count       : int  333 406 96 186 179 244 166 413 481 118 ...
```

```r
# Transformar em tibble
json_tbl <- as_tibble(json_data)
```

```r
# Ver a estrutura
glimpse(json_tbl)
```

```
## Rows: 2,133
## Columns: 16
## $ Transaction_ID        <chr> "702bdd9b-9c93-41e3-9dbb-a849b2422080", "2e64c34~
## $ Customer_ID           <chr> "119dca0b-8554-4b2d-9bec-e964eaf6af97", "299df08~
## $ Transaction_Date      <dbl> 1.753574e+12, 1.741910e+12, 1.740010e+12, 1.7455~
## $ Transaction_Time      <chr> "04:04:15", "20:23:23", "12:36:02", "19:09:43", ~
## $ Customer_Age          <int> 56, 46, 32, 60, NA, 38, 56, 36, 40, 28, 28, 41, ~
## $ Customer_Loyalty_Tier <chr> "Silver", "Platinum", "Silver", "Bronze", "Plati~
## $ Location              <chr> "San Francisco", "Zurich", "Milan", "London", "M~
## $ Store_ID              <chr> "FLAGSHIP-LA", "BOUTIQUE-SHANGHAI", "POPUP-TOKYO~
## $ Product_SKU           <chr> "NEBULA-SERUM-07", "STELLAR-FOUND-03", "SOLAR-BL~
## $ Product_Category      <chr> "Concealer", "Lipstick", "Mascara", "Serum", "Se~
## $ Purchase_Amount       <dbl> 158.24, 86.03, 255.69, 282.76, 205.86, 135.91, 8~
## $ Payment_Method        <chr> "Mobile Payment", "Credit Card", "Gift Card", "G~
## $ Device_Type           <chr> "Desktop", "Tablet", "Desktop", "Mobile", "Mobil~
## $ IP_Address            <chr> "239.249.58.237", "84.49.227.90", "79.207.35.55"~
## $ Fraud_Flag            <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, ~
## $ Footfall_Count        <int> 333, 406, 96, 186, 179, 244, 166, 413, 481, 118,~
```

```r
# Arquivo Parquet

# Serial Killer Data Blog
# Pandas Parquet with the blog posts of a Serial Kerial scrapped

library(arrow)
# Como o arquivo é pequeno não precisa usar SPARK, somente com o Arrow ele funciona perfeitamente

# Carregar o arquivo Parquet
parquet_tbl <- read_parquet("Serial_KIller.parquet")

# Explorar os dados
glimpse(parquet_tbl)
```

```
## Rows: 171
## Columns: 4
## $ post_title <chr> "", "", "I Know What I Know", "", "Life is in the details..~
## $ post_date  <chr> "Saturday, January 31, 2004", "Saturday, January 31, 2004",~
## $ post_hour  <chr> "9:08 PM", "8:53 PM", "12:20 AM", "12:17 AM", "2:44 PM", "1~
## $ post_text  <chr> " \n   ¶ 9:08 PM\n\nFact: Whoever controls the media contro~
```

```r
#Teste de tempo resposta dataset

bench::mark(
  arrow = read_parquet("Serial_KIller.parquet")
)
```

```
## # A tibble: 1 x 6
##   expression      min   median `itr/sec` mem_alloc `gc/sec`
##   <bch:expr> <bch:tm> <bch:tm>     <dbl> <bch:byt>    <dbl>
## 1 arrow        6.27ms   7.58ms      130.    7.96KB     4.20
```