

Test - Data y Analítica

Pronus

Durante la prueba podrá desarrollar los ejercicios haciendo uso de toda la información que tenga a su disposición. El test podrá ser realizado en R o Python. La estructura del código, su eficiencia y simplicidad serán factores que se tomarán en cuenta a la hora de evaluar.

La idea de esta prueba es simular una situación que seguramente se enfrentará cotidianamente dentro del equipo de Pronus. Al finalizar comparta un repositorio en GitHub con su código y un documento con las respuestas. El código debe ser autosuficiente.

Manipulación de Datos

Dentro del material compartido encontrará una carpeta titulada datos_prueba. Allí encontrará información descargada directamente de la Superintendencia Financiera de Colombia (SFC). La primera tarea será utilizar código para unificar la información del respectivo análisis.¹

Preparación de la base

Concéntrese únicamente en la hoja TarjetasXFranquicia. Esta hoja cuenta con 27 columnas. Para este ejercicio solo estamos interesados en las siguientes 5 columnas : (i) Entidad ; (ii) Vigentes a la fecha de corte; (iii) Vigentes durante el mes; (iv) Canceladas; y (v) Bloqueadas temporalmente. Corresponden a la sección de “Totales”². El resultado debe ser similar al siguiente (fíjese que se añadió una columna referente al documento. Será importante para luego estimar la fecha.)

```
## # A tibble: 6 x 6
##   ENTIDAD                Total_vigentes_fecha_c~1 Total_vigentes_duran~2 Canceladas
##   <chr>                  <chr>                  <chr>                  <chr>
## 1 Banco de Bogotá      1183146                37960                27691
## 2 Banco Popular        257470                 6125                 2052
## 3 Itaú                 273995                 6159                 1989
## 4 Bancolombia          2321209                88711                59682
```

¹Fíjese que el nombre del archivo corresponde a la fecha de los datos. Por ejemplo, “0122.xlsx” son datos de enero del 2022.

²Antes de empezar el código analice los datos desde excel

```
## 5 Citibank          11498          316          72
## 6 Banco GNB Sudameris 16099          16          163
## # i abbreviated names: 1: Total_vigentes_fecha_corte,
## #   2: Total_vigentes_durante_mes
## # i 2 more variables: Bloqueada_temporalmente <chr>, ref_doc <chr>
```

Columnas numéricas y Fecha

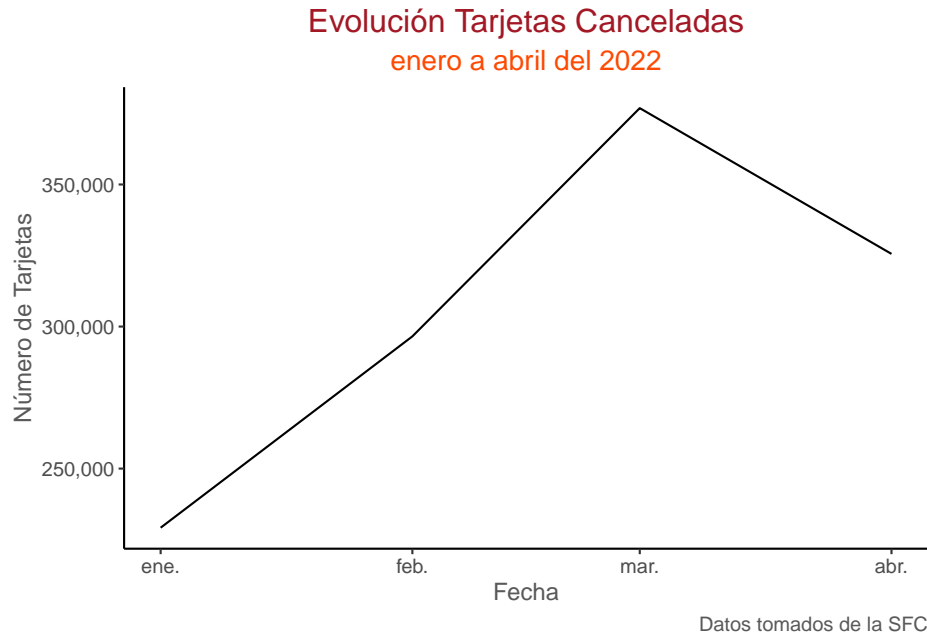
Tal vez, uno de los retos más importantes de un analista de datos es trabajar con fechas. En este caso, la prueba espera la siguiente base resultante (cambie el nombre de las variables para que coincida con lo siguiente):

```
## # A tibble: 6 x 6
##   ENTIDAD          vigentes_corte vigentes_mes canceladas bloqueadas fecha
##   <chr>            <dbl>         <dbl>         <dbl>         <dbl> <date>
## 1 Banco de Bogotá    1183146        37960        27691        200883 2022-01-01
## 2 Banco Popular      257470         6125         2052         3932 2022-01-01
## 3 Itaú               273995         6159         1989         20324 2022-01-01
## 4 Bancolombia        2321209        88711        59682        168169 2022-01-01
## 5 Citibank           11498          316          72         22441 2022-01-01
## 6 Banco GNB Sudame~  16099          16          163          722 2022-01-01
```

Análisis de Datos

Construya la siguiente gráfica, la paleta de colores usada es la siguiente:

- Título = #A21824
- Subtitulo = #FE4902
- Texto dentro de la gráfica = # 595959



Responda las siguientes preguntas, puede contestarlas dentro del código como comentarios:

- ¿La gráfica presentada arriba es congruente con respecto al título, subtítulo, ejes, etc? ¿Por qué no? ¿Qué le añadiría?
- ¿Cuál es crecimiento mensual promedio de las tarjetas canceladas?
- ¿Cuál fue el Establecimiento de Crédito que más tarjetas canceló durante todo el horizonte de tiempo?
- El cliente le pregunta ¿Cuál es la razón de esta tendencia? Detalle los pasos a seguir de su metodología, entre más robusta sea, mejor.