

```
</style><meta name="keywords" content="CMF, valores, seguros, banco"
```

```
ar __cid = '515';  
ar __iid = '613';
```

```
</script><script type="text/javascript" src="channels-515_  
<meta http-equiv="X-UA-Compatible" content="IE=edge,chrome=1">
```

```
-- Fonts CSS -->
```

```
<link rel="stylesheet" href="https://cdnjs.cloudflare.com/ajax/
```

```
<link rel="preconnect" href="https://fonts.gstatic.com">
```

```
<link href="https://fonts.googleapis.com/css2?family=Open+Sans:1
```

```
-- CSS iconos Font Awesome y Bootstrap -->
```

```
<link rel="stylesheet" href="https://cdnjs.cloudflare.com/ajax/
```

```
<link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/boo
```

```
<link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/bootst
```

```
-- animated css -->
```

```
<link rel="stylesheet" href="channels-515_css_animated_4_1_1.css
```

```
-- Bootstrap CSS -->
```

```
--link rel="stylesheet" href="https://css-ext.newtenberg.com/c
```

```
<link rel="stylesheet" href="channels-515_css_cmf_styles.css"
```

```
-- Iconos CMF ver en: https://css-ext.newtenberg.com/cmf_2020/
```

```
<link rel="stylesheet" href="https://www.cmfchile.cl/portal/cmf
```

```
<link rel="stylesheet" href="https://www.cmfchile.cl/portal/cmf
```

```
<link rel="stylesheet" href="https://www.cmfchile.cl/portal/cmf
```

```
<link rel="stylesheet" href="https://www.cmfchile.cl/portal/cmf
```

Proyecto Webscrapping: Consulta valor cuota diaria de Fondos Mutuos en CMF

Gabriel Valenzuela

Francisco Letelier

Agenda

- Sitio a Scrapear
- Objetivo
- Pseudocódigo
- Tecnología a usar
- Tamaño de la data
- Información de los datos
- Flujo de Trabajo



Sitio a scrapear: www.cmfchile.cl

- Sitio a scrapear
- Objetivo
 - Se extraerán los valores cuotas de los distintos FFMM registrados diariamente desde 01 Enero 2021 a la fecha.
- Pseudocódigo
- Tecnología a usar
- Tamaño de la data
- Features
- Flujo de Trabajo



Sitio a scrapear: www.cmfchile.cl

- Sitio a scrapear
- Objetivo
- Pseudocódigo
 - for cada día:
 - Buscar: fecha, fondo, RUN, Serie, Moneda, Valor Cuota
 - Agregar a fila de Dataframe
 - Guardar en .csv
 - Accionar cada día al cierre de las operaciones
- Tecnología a usar
- Tamaño de la data (al 6 de Septiembre 2022)
- Información de los datos
- Flujo de Trabajo

```

</style><meta name="keywords" content="CMF,valores,seguros,bancos, tarjetas de crédito, créditos, svcs, sbif">
var __cid = '515';
var __iid = '613';

--></script><script type="text/javascript" src="channels-515_js_main.js"></script><script type="text/javascript">
<meta http-equiv="X-UA-Compatible" content="IE=edge,chrome=1">
<!-- Fonts CSS -->
<link rel="stylesheet" href="https://cdnjs.cloudflare.com/ajax/libs/font-awesome/4.7.0/css/font-awesome.min.css">
<link rel="preconnect" href="https://fonts.gstatic.com">
<link href="https://fonts.googleapis.com/css2?family=Open+Sans:ital,wght@0,300;0,400;0,600;0,700;1,300;1,400;1,600;1,700;1,800;1,900&display=block" rel="stylesheet">
<!-- CSS iconos Font Awesome y Bootstrap -->
<link rel="stylesheet" href="https://cdnjs.cloudflare.com/ajax/libs/font-awesome/4.7.0/css/font-awesome.min.css">
<link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/bootstrap-icons@1.5.0/font/bootstrap-icons.css">
<link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/bootstrap-icons@1.8.1/font/bootstrap-icons.css">
<!-- animated css -->
<link rel="stylesheet" href="channels-515_css_animated_4_1_1.css">
<!-- Bootstrap CSS -->
<!--link rel="stylesheet" href="https://css-ext.newtenberg.com/cmf_2020/css/styles-cmf.css"-->
<link rel="stylesheet" href="channels-515_css_cmf_styles.css">
<!-- Iconos CMF ver en: https://css-ext.newtenberg.com/cmf_2020/cmf-icons.html -->
<link rel="stylesheet" href="https://www.cmfchile.cl/portal/cmf_fonts_2021/css/animation.css">
<link rel="stylesheet" href="https://www.cmfchile.cl/portal/cmf_fonts_2021/css/cmf-icons.css">
<link rel="stylesheet" href="https://www.cmfchile.cl/portal/cmf_fonts_2021/css/cmf-icons-codes.css">
<link rel="stylesheet" href="https://www.cmfchile.cl/portal/cmf_fonts_2021/css/cmf-icons-embedded.css">

```

Sitio a scrapear: www.cmfchile.cl

- Sitio a scrapear
- Objetivo
- Pseudocódigo
- Tecnología a usar
 - **Selenium:** Completar formulario de descarga de valor cuota, patrimonio diario de cada fondo mutuo
 - **XLRD:** La página entrega información en planilla Excel 97 (.xls) con codificación Español Latinoamericano (latin1), lo que genera problemas con pandas. Se usa xlrd para leer archivo y se crea dataframe con pandas
 - **OS:** Para guardar archivo csv y tratar la descarga de las planillas Excel se usa os.
 - **Otras:** La página tiene un robot que impide descargas en masa. Luego, se utiliza Time para crear delays entre que se completa el formulario y se presiona el botón "Descargar a Excel" (5 segundos), y mientras se descarga el archivo (2 segundos). Para el ordenamiento y tratamiento de datos, se usa Pandas

```

} </style><meta name="keywords" content="CMF,valores,seguros,bancos, tarjetas de crédito, créditos

var __cid = '515';
var __iid = '613';

--></script><script type="text/javascript" src="channels-515_js_main.js"></script><script type="t
<meta http-equiv="X-UA-Compatible" content="IE=edge,chrome=1">
<!-- Fonts CSS -->
<link rel="stylesheet" href="https://cdnjs.cloudflare.com/ajax/libs/font-awesome/4.7.0/css/font-a
<link rel="preconnect" href="https://fonts.gstatic.com">
<link href="https://fonts.googleapis.com/css2?family=Open+Sans:ital,wght@0,300;0,400;0,600;0,700;
<!-- CSS iconos Font Awesome y Bootstrap -->
<link rel="stylesheet" href="https://cdnjs.cloudflare.com/ajax/libs/font-awesome/4.7.0/css/font-a
<link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/bootstrap-icons@1.5.0/font/bootstrap-ic
<link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/bootstrap-icons@1.8.1/font/bootstrap-ic
<!-- animated css -->
<link rel="stylesheet" href="channels-515_css_animated_4_1_1.css">
<!-- Bootstrap CSS -->
<!-- link rel="stylesheet" href="https://css-ext.newtenberg.com/cmf_2020/css/styles-cmf.css"-->
<link rel="stylesheet" href="channels-515_css_cmf_styles.css">
<!-- Iconos CMF ver en: https://css-ext.newtenberg.com/cmf_2020/cmf-icons.html -->
<link rel="stylesheet" href="https://www.cmfchile.cl/portal/cmf_fonts_2021/css/animation.css">
<link rel="stylesheet" href="https://www.cmfchile.cl/portal/cmf_fonts_2021/css/cmf-icons.css">
<link rel="stylesheet" href="https://www.cmfchile.cl/portal/cmf_fonts_2021/css/cmf-icons-codes.css">
<link rel="stylesheet" href="https://www.cmfchile.cl/portal/cmf_fonts_2021/css/cmf-icons-embedded

```


Tamaño de la Data

- Sitio a scrapear
- Objetivo
- Pseudocódigo
- Tecnología a usar
- Tamaño de la data (al 6 de Septiembre 2022)
 - Vigentes: 517 FFMM
 - Periodo: 1 Enero 2021 al 30 de Agosto 2022
 - Días: 404 (días laborales)
 - Data: 224.378 filas
- Información de los datos
- Flujo de Trabajo

```

1 Fecha;Fondo;RUN;Administradora;Serie;Moneda en que contabiliza el fondo;Valor cuota
2 30-08-2022;SECURITY PLUS;8253-8;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;A;PESOS;1.776,35
3 30-08-2022;SECURITY PLUS;8253-8;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;B;PESOS;1.146,21
4 30-08-2022;SECURITY PLUS;8253-8;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;C;PESOS;2.822,92
5 30-08-2022;SECURITY PLUS;8253-8;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;D;PESOS;1.335,26
6 30-08-2022;SECURITY PLUS;8253-8;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;E;PESOS;1.206,81
7 30-08-2022;SECURITY PLUS;8253-8;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;G;PESOS;1.010,61
8 30-08-2022;SECURITY PLUS;8253-8;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;I;PESOS;1.175,73
9 30-08-2022;SECURITY PLUS;8253-8;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;S;PESOS;1.208,37
10 30-08-2022;DOLAR MONEY MARKET;8384-4;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;A;DOLAR;1.164,28
11 30-08-2022;DOLAR MONEY MARKET;8384-4;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;B;DOLAR;1.060,37
12 30-08-2022;DOLAR MONEY MARKET;8384-4;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;C;DOLAR;1.098,21
13 30-08-2022;DOLAR MONEY MARKET;8384-4;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;S;DOLAR;1.090,07

```

Información de los datos

- Sitio a scrapear
- Objetivo
- Pseudocódigo
- Tecnología a usar
- Tamaño de la data (al 6 de Septiembre 2022)
- Información de los Datos
 - Fecha: (date) “dd-mm-yyyy”
 - Fondo: (str) nombre del fondo
 - RUN: (str) rol único tributario de la administradora
 - Administradora: (str) razón social de la administradora de fondos.
 - Serie: (str) forma con la que el FFMM clasifica las diferentes cuotas dentro del mismo fondo.
 - Moneda (str) Pesos CLP o USD
 - Valor_Cuota: (float) es la división del valor contables del patrimonio del fondo, por el numero de cuotas suscritas y pagadas al momento del cálculo

```

1 Fecha;Fondo;RUN;Administradora;Serie;Moneda en que contabiliza el fondo;Valor cuota
2 30-08-2022;SECURITY PLUS;8253-8;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;A;PESOS;1.776,35
3 30-08-2022;SECURITY PLUS;8253-8;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;B;PESOS;1.146,21
4 30-08-2022;SECURITY PLUS;8253-8;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;C;PESOS;2.822,92
5 30-08-2022;SECURITY PLUS;8253-8;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;D;PESOS;1.335,26
6 30-08-2022;SECURITY PLUS;8253-8;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;E;PESOS;1.206,81
7 30-08-2022;SECURITY PLUS;8253-8;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;G;PESOS;1.010,61
8 30-08-2022;SECURITY PLUS;8253-8;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;I;PESOS;1.175,73
9 30-08-2022;SECURITY PLUS;8253-8;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;S;PESOS;1.208,37
10 30-08-2022;DOLAR MONEY MARKET;8384-4;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;A;DOLAR;1.164,28
11 30-08-2022;DOLAR MONEY MARKET;8384-4;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;B;DOLAR;1.060,37
12 30-08-2022;DOLAR MONEY MARKET;8384-4;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;C;DOLAR;1.098,21
13 30-08-2022;DOLAR MONEY MARKET;8384-4;ADMINISTRADORA GENERAL DE FONDOS SECURITY S.A.;S;DOLAR;1.090,07

```

Flujo de trabajo

- Sitio a scrapear
- Objetivo
- Pseudocódigo
- Tecnología a usar
- Tamaño de la data (al 6 de Septiembre 2022)
- Información de los datos
- Flujo de Trabajo
 1. Script Python
 2. Github Actions
 3. Scrap web CMF
 4. Descarga xls a csv
 5. Acumulación data en una base de datos
 6. Uso de la data para investigación y analítica de negocios.

