



CARRERA: TECNICATURA SUPERIOR DE CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

ESPACIO CURRICULAR: PROYECTO INTEGRADOR I

PROFESORA: SILVIA PEROTTI

PROFESOR: HÉCTOR PRADO

COHORTE: 2022

PROYECTO TECNOLÓGICO INTEGRADOR 2022

PROYECTO GAMA

GRUPO: BSTT4

INTEGRANTES: BUSETTI DIAZ, DANIELA VERONICA

SINDE, JAVIER FERNANDO

TAMINETTO, SERGIO DANIEL

TORRES, TOMAS LAUTARO

OBJETIVO

Realizar un proyecto real de trabajo colaborativo en un entorno ágil.

ALCANCE DEL PROYECTO

El alcance del proyecto consiste en el estudio de la evolución de los precios de combustibles con respecto a la inflación.

PROYECTO

1- HERRAMIENTAS COLABORATIVAS

Utilizamos Trello, Github y Google meet para las reuniones.

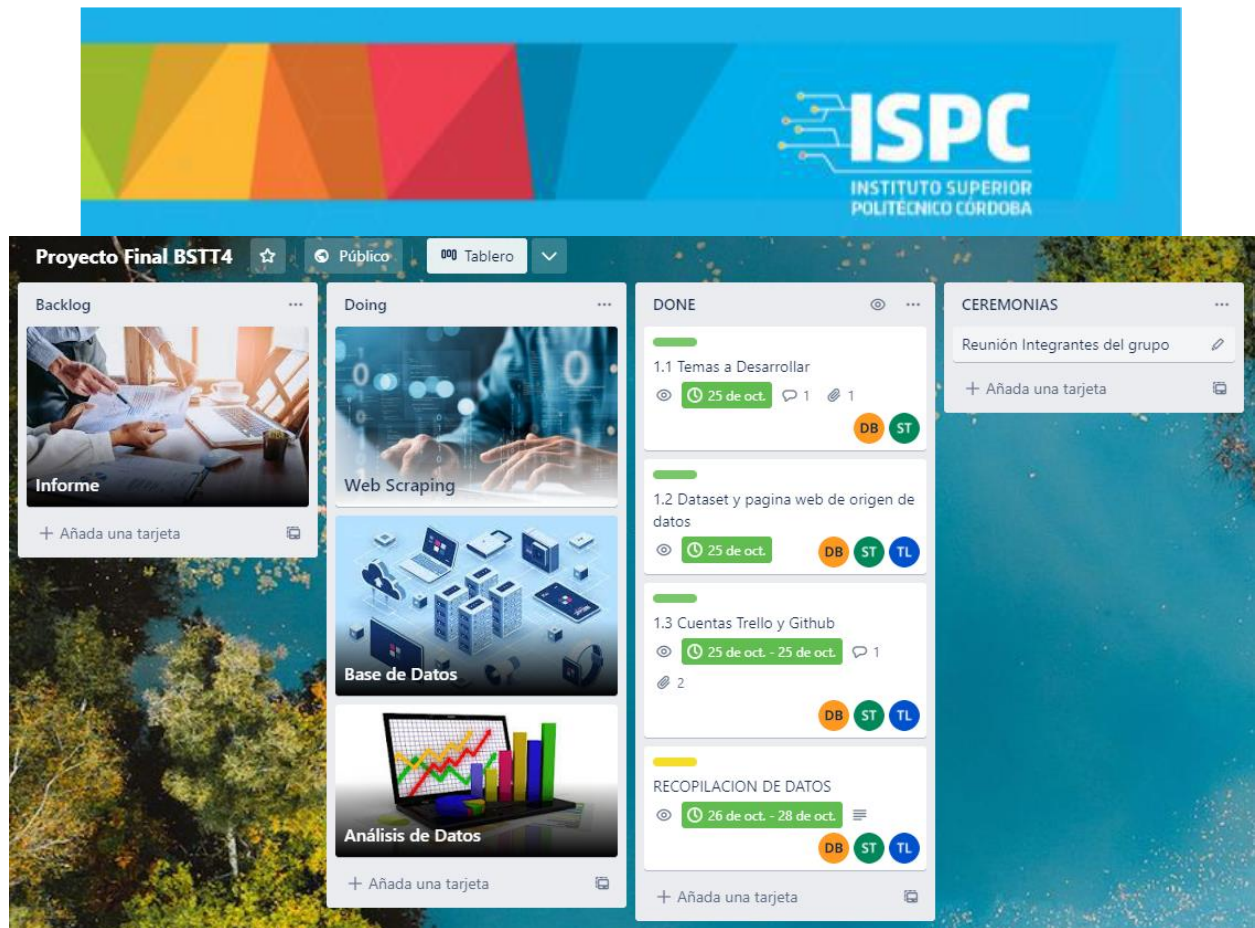
TRELLO: <https://trello.com/b/bGAdA2Ma/proyecto-final-bstt4>

En una primera instancia colocamos las tarjetas correspondientes a:

- Organización de las cuentas de Trello y Github
- Tema a desarrollar: Precios en Surtidor
- Obtención del Dataset: https://datos.gob.ar/dataset/energia-precios-surtidor---resolucion-3142016/archivo/energia_80ac25de-a44a-4445-9215-090cf55cfda5
- Obtención del Dataset de Inflación: <https://datosmacro.expansion.com/ipc-paises/argentina?sector=IPC+General&sc=IPC-IG&anio=2022>

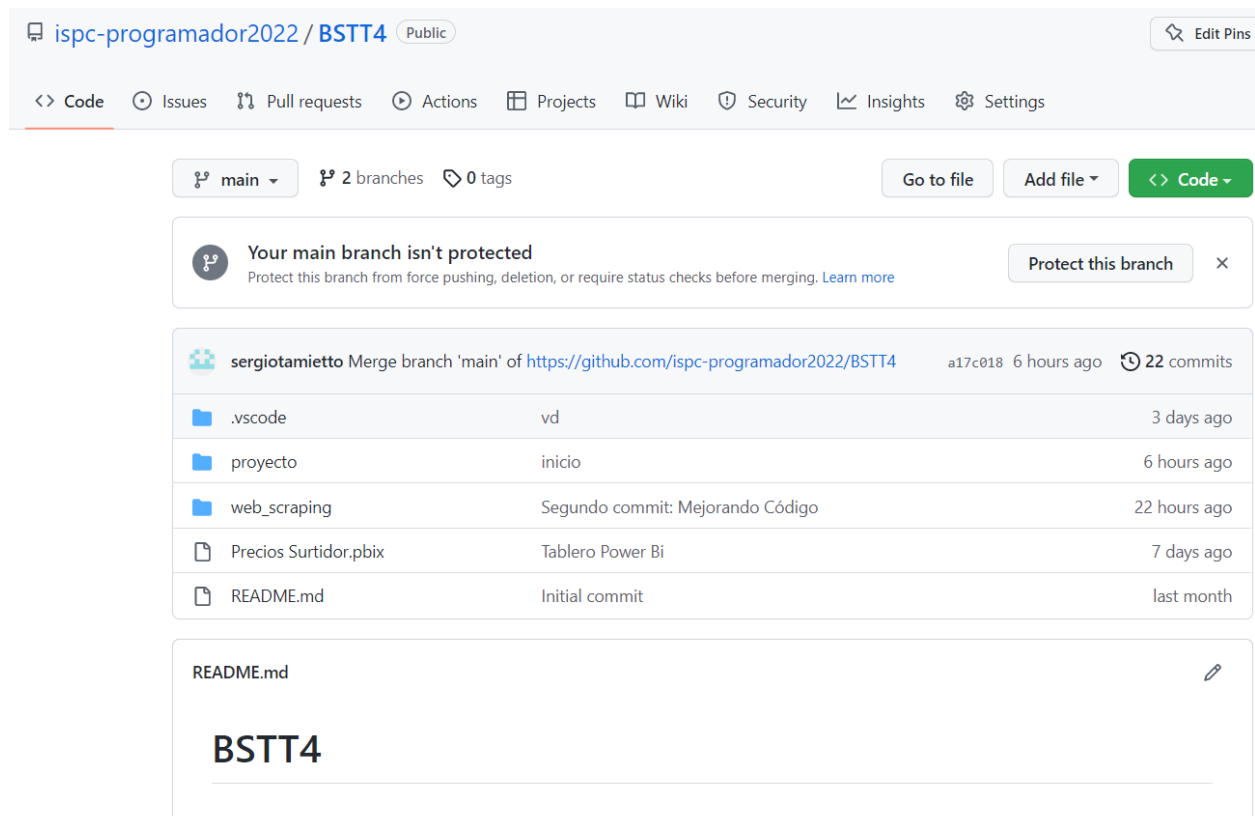
Luego dividimos el proyecto en las siguientes áreas:

- Web Scraping
- Análisis de Datos
- Informe



GITHUB: <https://github.com/ispc-programador2022/BSTT4>

Acá subimos todo el proyecto realizado. Ver las últimas modificaciones.





2- WEB SCRAPING

Página: <https://datosmacro.expansion.com/ipc-paises/argentina?sector=IPC+General&sc=IPC-IG&anio=2022>

Con el lenguaje Python, las bibliotecas BeautifulSoup4, requests, lxml y pandas, y el IDE Pycharm obtuvimos el DataFrame de las tablas de inflación y lo exportamos a un archivo csv.

Primera Versión

```
web_scraping  codigo_tabla_2015-2016.py
Project
  web_scraping
    venv
      Lib
      Scripts
      .gitignore
      pyvenv.cfg
    codigo_mejorado.py
    codigo_tabla_2015-2016.py
    codigo_tabla_2016-2017.py
    codigo_tabla_2017-2018.py
    codigo_tabla_2018-2019.py
    codigo_tabla_2019-2020.py
    codigo_tabla_2020-2021.py
    codigo_tabla_2021-2022.py
    Tabla 2015-2016
    Tabla 2016-2017
    Tabla 2017-2018
    Tabla 2018-2019
    Tabla 2019-2020
    Tabla 2020-2021
    Tabla 2021-2022
  External Libraries
  Scratches and Consoles
Structure
Bookmarks

1  # Importar librerías
2  import requests
3  from bs4 import BeautifulSoup
4  import pandas as pd
5  import lxml
6
7  url_inicial = 'https://datosmacro.expansion.com/ipc-paises/argentina?sector=IPC+General&sc=IPC-IG&anio=2016'
8
9  # Pedido a la página
10 r = requests.get(url_inicial)
11
12 # Armos la sopa
13 soup = BeautifulSoup(r.text, 'html.parser')
14 # print(soup)
15
16 # Tabla
17 rows = soup.find('table', attrs={"id":"tb1_415"}).find('tbody').find_all('tr')
18 # print(rows)
19
20 mes = []
21 interanual = []
22 acum = []
23 var_m = []
24 for row in rows:
25     mes.append(row.find_all('td')[0].get_text())
26     interanual.append(row.find_all('td')[1].get_text())
27     acum.append(row.find_all('td')[3].get_text())
28     var_m.append(row.find_all('td')[5].get_text())
29
30 # print(mes)
31 # print(interanual)
32 # print(acum)
33 # print(var_m)
34
35 # Convertimos datos a un DataFrame
36 df = pd.DataFrame({'mes':mes, 'interanual':interanual, 'acum': acum, 'var_m':var_m})
37 print(df)
38
39 # Exportamos el DataFrame
40 df.to_csv('Tabla 2015-2016')
```

Segunda versión: Código mejorado

Link: <https://github.com/ispc-programador2022/BSTT4/blob/main/proyecto/inicio.py>

Estos datos los guardamos en una Base de Datos utilizando MySQL para realizar consultas, con bibliotecas BeautifulSoup4, requests, pandas, os y mysql connector.

Se unió en un archivo toda la información, con selección, borrado y armado de la Base de Datos.



ispc-programador2022 / BSTT4 Public

<> Code Issues Pull requests Actions Projects Wiki Security Insights Settings

main BSTT4 / proyecto / inicio.py / <> Jump to

sergiotamietto código

1 contributor

159 lines (125 sloc) 5.13 KB

```
1 import pandas as pd
2 import os
3 import mysql.connector
4 import requests
5 from bs4 import BeautifulSoup
6 import time
7
8 os.system('cls')
9
10 lista=list()
11
12
13 def leerArchivo_csv():
14     dirActual=os.getcwd()
```

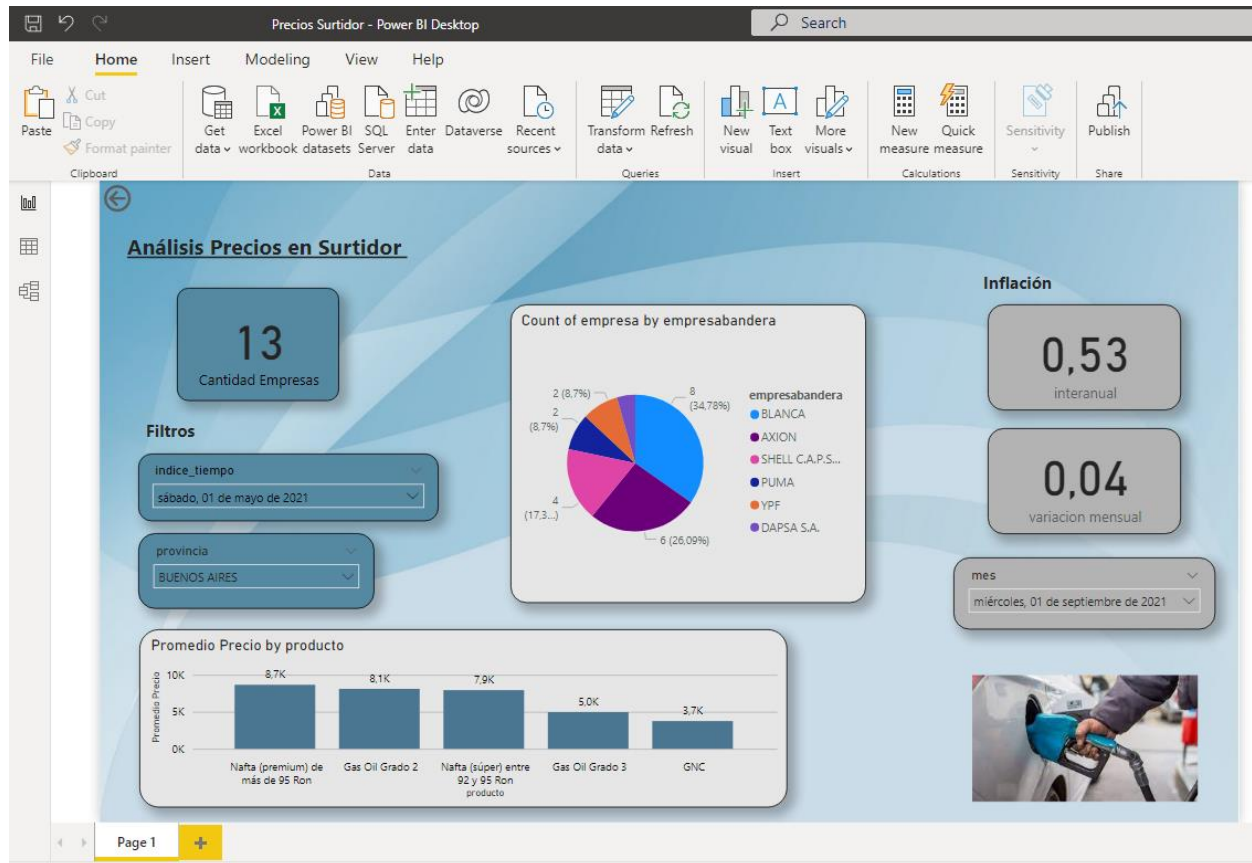
3- ANÁLISIS DE DATOS

Utilizamos el programa Power BI con el que realizamos los cálculos estadísticos y el tablero para la visualización de los datos.

Con los archivos obtenidos ya sea por web scraping o por el csv descargado, lo primero que hicimos fue importar las tablas de datos. En este paso se realiza la limpieza, transformación de datos y elección de las columnas que vamos a utilizar.

El análisis en Power BI lo hicimos a través de expresiones DAX, que es un lenguaje de fórmulas que permiten realizar cálculos matemáticos y estadísticos para trabajar con datos relacionales y agregaciones dinámicas.

- Filtros: Consideramos dos filtros una para elegir la fecha (mes y año) y otro para elegir la provincia.
- KPI o indicador: De la Cantidad de Empresas por provincia, de la Inflación interanual y su variación mensual, vinculado con los filtros.
- Tablas: La primera tabla muestra la cantidad de empresas según fecha y provincia, y la segunda tabla, el precio promedio para cada tipo de combustible, que también se puede filtrar por fecha y provincia.



4- CONCLUSIONES

Por ejemplo, si seleccionamos la provincia de Córdoba para setiembre 2022 vemos que hay 4 empresas banderas correspondientes a Blanca (con 7 empresas), Shell (con 5 empresas), Ypf (con 3 empresas) y Dapsa (con 1 empresa).

Los precios promedios de los combustibles son: Gas oil Gado 2 (22.000), Gas oil Gado 3 (21.000), Nafta Premium (21.000), Nafta Super (19.000) y GNC (7.000).

A su vez, para setiembre del 2022 tenemos una inflación interanual de 0.83% y una variación mensual del 0.06%

