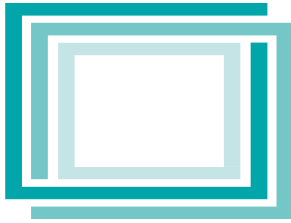


Pre-presentación Proyecto Webscraping



UNIVERSIDAD ADOLFO IBÁÑEZ

Integrantes:

- Rodrigo Bruna Durán
- José Mora González

Contexto del Proyecto



UAI
UNIVERSIDAD ADOLFO IBÁÑEZ

Actualmente Chile es un país sumamente dependiente de otras naciones respecto de sus necesidades alimenticias. Esto genera se traduce en una debilidad para la economía pues frente a situaciones como la guerra en Ucrania o desastres naturales, se generan enormes variaciones de precio debido a la escasez. Frente a esta situación nos hemos enfocado en analizar la situación productiva chilena en alimentos lácteos, con la finalidad de conocer si esta actividad productiva tiene una tendencia positiva o negativa, y establecer un panel de control que permita monitorear estos comportamientos.

Con lo anterior queremos visibilizar esta situación dejando el análisis en un plataforma abierta para que cualquier ciudadano pueda conocer esta información en visualizaciones de rápido entendimiento

Pasos a seguir



UAI
UNIVERSIDAD ADOLFO IBÁÑEZ

- Analizar la web
<https://www.odepa.gob.cl/avance-mensual-y-series-de-tiempo-de-productos-por-planta-o-region-de-la-industria-lactea>



- Crear un scraper que extraiga la la información necesaria.
- Almacenar los datos en archivos csv.
- Generar visualizaciones públicas de actualización mensual usando streamlit.

Pseudocódigo



UAI
UNIVERSIDAD ADOLFO IBÁÑEZ

- 1 Importar librerías pandas, requests y BeautifulSoup
- 2 Crear variable url = "https://aplicativos.odepa.gob.cl/avancemensual.do"
- 3 Crear diccionario headers = {'Cookie': 'JSESSIONID=2E9275A55DAFA54F9AFEE077EA378990'}
- 4 Hacer un request del tipo GET y almacenar el resultado en la variable page
- 5 Crear diccionario campos = {'mesini': 'cboMesIni', 'anioini': 'cboAñoIni', 'mesfin': 'cboMesFin', 'aniofin': 'cboAñoFin', 'producto': 'cboProducto', 'region': 'cboRegion'}
- 6 Crear diccionario parametros vacío
- 7 Parsear el contenido de la variable page y almacenar el resultado en la variable soup
- 8 Para cada key y value en el diccionario campos:
 - 8.1 Encontrar el select con el id igual a value y almacenarlo en la variable temp
 - 8.2 Obtener el listado de valores option almacenados en temp y almacenarlo en la variable valor
 - 8.3 Adicionar el listado en el diccionario parametros

9 Establecer los años a revisar minimo y maximo y almacenar el valor en la variable anios

10 Establecer los productos a revisar almacenarlo en la variable productos

11 Establecer las regiones a revisar y almacenarlo en la variable regiones

12 Para cada producto en la variable productos:

12.1 Para cada region en la variable regiones:

12.1.1 Crear el diccionario payload con los valores payload = {'dataExport': '', 'compressed': 'false', 'fileNameExcel': '', 'decimales': '', 'cboMesIni': '01', 'cboAñoIni': valor mínimo de año almacenado en la variable anios, 'cboMesFin': '06', 'cboAñoFin': valor máximo de año almacenado en la variable anios, 'cboProducto': producto i en productos, 'rdoTipo': 'region', 'cboRegion': region i en regiones, 'rdoFormatoTabla': 'mesaño'}

12.1.2 Hacer el request tipo POST con usando las variables url, headers y payload y almacenar el resultado en la variable page

12.1.3 Crear un dataframe vacio llamado dffinal en el que se almacenarán los resultados

12.1.4 Si el request no devuelve error:

12.1.4.1 Parsear el contenido de la variable page y almacenar el resultado en la variable soup

12.1.4.2 Encontrar la tabla con valores y almacenarla en la variable rows

12.1.4.3 Encontrar los encabezados de la tabla y almacenarla en la variable columns

12.1.4.4 Construir la tabla usando las variables columns y rows

12.1.4.5 Añadir la columna Producto a la tabla, haciendo que su contenido sea igual al producto i

12.1.4.6 Añadir la columna Region a la tabla, haciendo que su contenido sea igual al region i

12.1.4.7 Adicionar los valores de tabla en el dataframe dfffinal