

Electivo Profesional Salud Digital

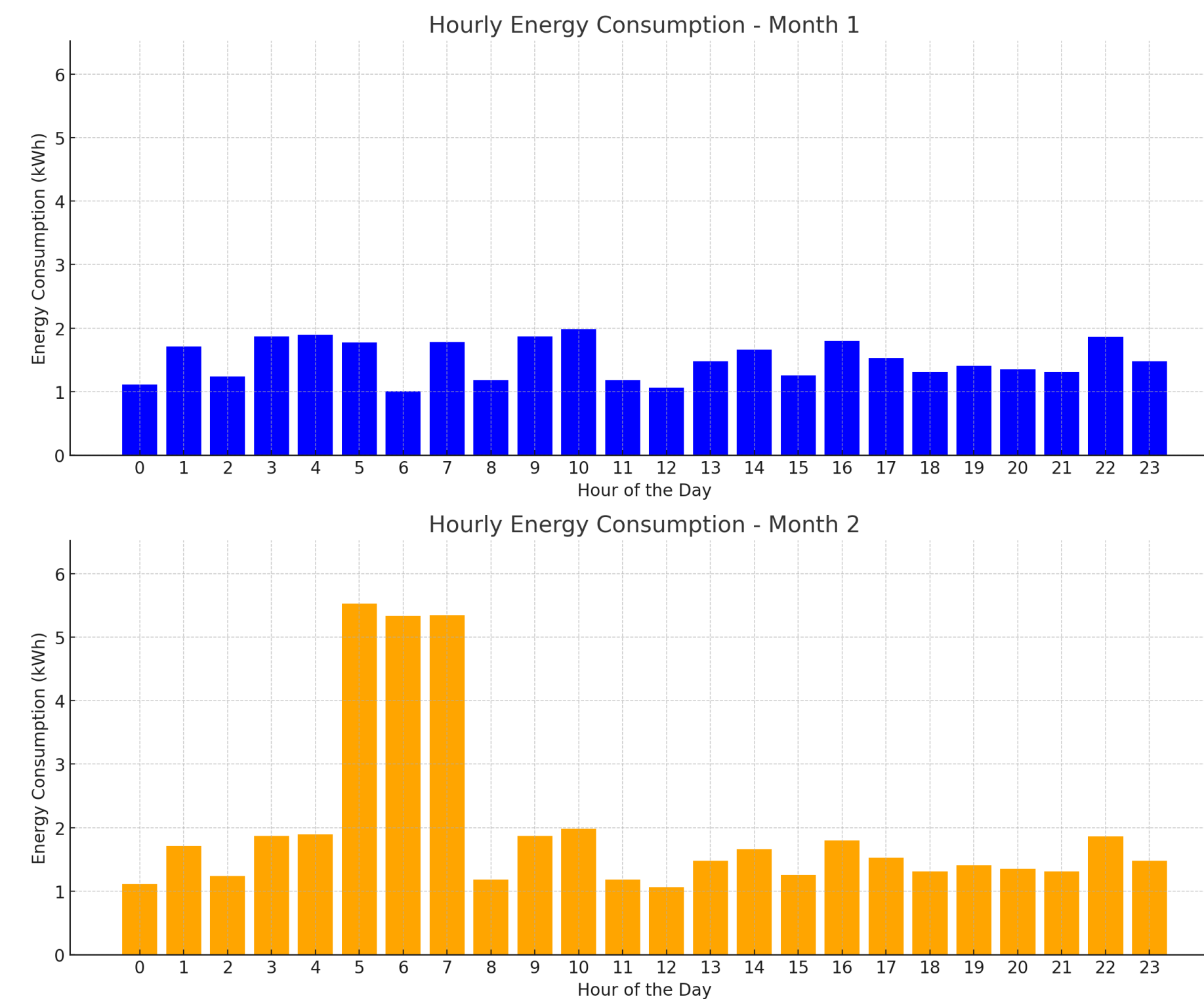
Archivos, DataFrames y Visualización I

**Prof. Mauricio Cerda
Álvaro López Acevedo**

Programa de Biología Integrativa, I.C.B.M.,
Centro de Informática Médica y Telemedicina
Facultad de Medicina, Universidad de Chile

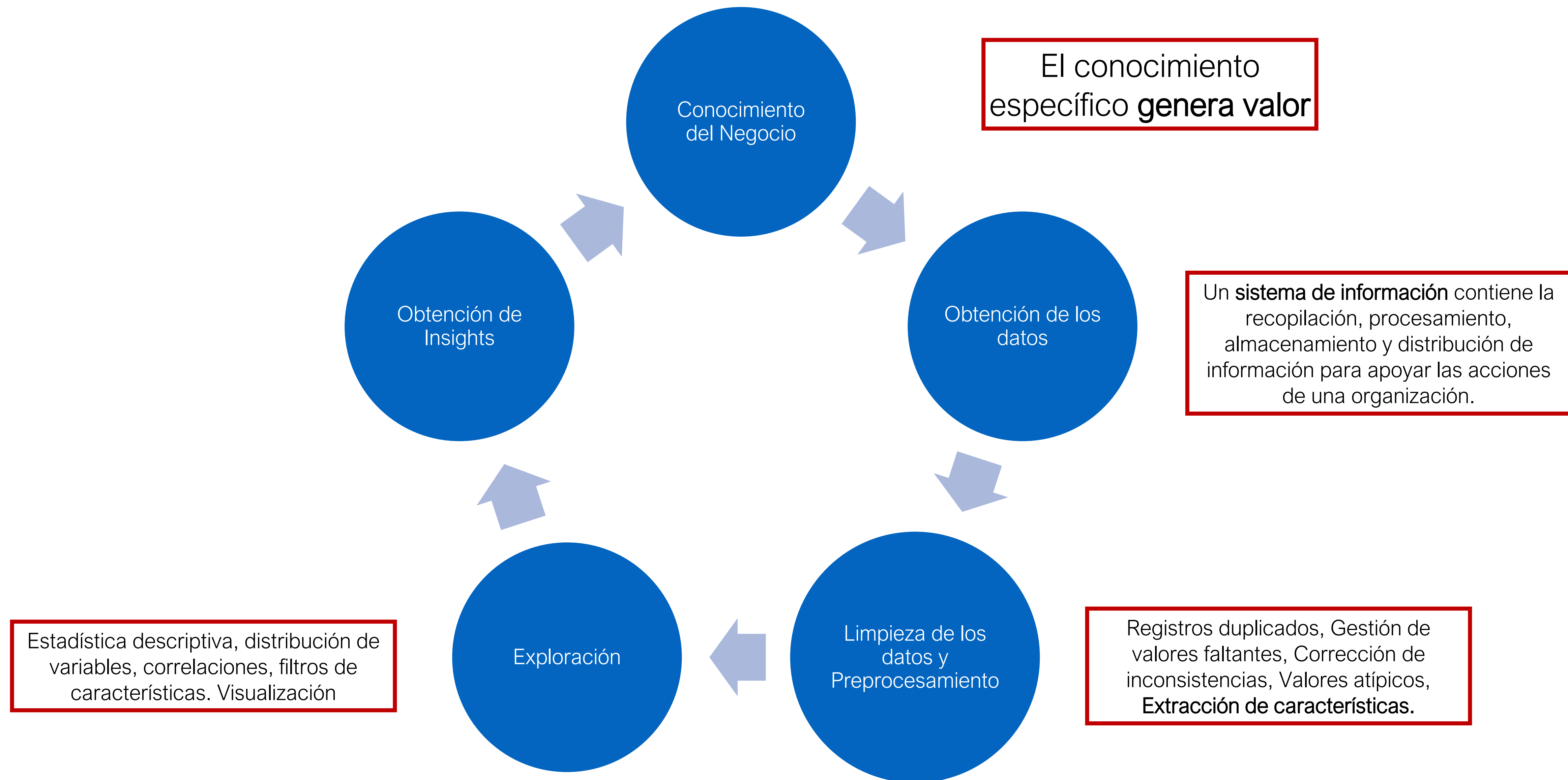
Imagínese que drásticamente la cuenta de luz su casa se duplicó en el último mes. ¿Qué harías?

Exacto. Irías a mirar la cuenta de la luz, a ver si te entrega alguna **información relevante**.



Fuente: Elaboración propia.





Objetivos



- Abrir archivos de datos desde Colab.
- Conocer la librería Pandas, la clase *DataFrame*, los tipos de datos.
- Aprender a utilizar las funciones (*métodos*) propias de la clase.
- Aplicarlas para un set de datos básico y poder responder preguntas atinentes.
- Visualizar los resultados con la librería Matplotlib

- Para abrir archivos de datos desde Colab es necesaria la librería de google.colab importar el módulo drive



```
from google.colab import drive
```

- Utilizamos el método drive.mount para mostrar nuestro drive

```
drive.mount('/content/drive')
```



- Librería diseñada para el análisis de datos
- Introduce el *DataFrame*
- Múltiples métodos
 - Creación
 - Lectura
 - Exploración
 - Manipulación
 - Concatenación
 - Exportación

ID	Col 1	Col 2	Col 3	Col 4
ID1	1	2	3	2
ID2	2	3	4	5
ID3	3	4	5	6
ID4	4	5	6	7
ID5	5	6	7	8
ID6	6	7	8	9

ID	Col 2
ID2	3


```
import pandas as pd

# Crear un DataFrame a partir de un diccionario
data = {
    'Nombre': ['Juan', 'María', 'Pedro', 'Ana'],
    'Edad': [35, 28, 45, 52],
    'Diagnóstico': ['Hipertensión', 'Diabetes Tipo 2', 'Colesterol alto', 'Hipertensión']
}
df = pd.DataFrame(data)

print(df)
```

	Nombre	Edad	Diagnóstico
0	Juan	35	Hipertensión
1	María	28	Diabetes Tipo 2
2	Pedro	45	Colesterol alto
3	Ana	52	Hipertensión

Pandas

```
print(df['Nombre'])
```

	Nombre
0	Juan
1	María
2	Pedro
3	Ana

```
print(df[df['Edad'] > 40])
```

	Nombre	Edad	Diagnóstico
2	Pedro	45	Colesterol alto
3	Ana	52	Hipertensión

```
print(df[(df['Diagnóstico'] == 'Hipertensión') & (df['Edad'] > 45)])
```

	Nombre	Edad	Diagnóstico
3	Ana	52	Hipertensión

```
print(df[(df['Diagnóstico'] == 'Hipertensión') & (df['Edad'] > 45)] ['Nombre'])
```

	Nombre
3	Ana

Pandas

	Nombre	Edad	Diagnóstico		Nombre	Previsión
0	Juan	35	Hipertensión	0	Juan	FONASA
1	María	28	Diabetes Tipo 2	1	María	ISAPRE
2	Pedro	45	Colesterol alto	2	Pedro	FONASA
3	Ana	52	Hipertensión	3	Ana	ISAPRE

```
df_combinado = pd.merge(df, df_prevision, on='Nombre' , how='left')
print(df_combinado)

nuevas_filas = pd.DataFrame([
    {'Nombre': 'Carlos', 'Edad': 40, 'Diagnóstico': 'Migraña', 'Previsión': 'FONASA'},
    {'Nombre': 'Elena', 'Edad': 25, 'Diagnóstico': 'Hipotiroidismo', 'Previsión': 'ISAPRE'}])
df = pd.concat([df_combinado, nuevas_filas], ignore_index=True)
print(df)
```

	Nombre	Edad	Diagnóstico	Previsión
0	Juan	35	Hipertensión	FONASA
1	María	28	Diabetes Tipo 2	ISAPRE
2	Pedro	45	Colesterol alto	FONASA
3	Ana	52	Hipertensión	ISAPRE
4	Carlos	40	Migraña	FONASA
5	Elena	25	Hipotiroidismo	ISAPRE

- Para crear un *DataFrame* a partir de un Excel que tenemos en nuestra Unidad Drive



```
from google.colab import drive  
import pandas as pd
```

```
drive.mount('/content/drive')  
df = pd.read_excel('/content/drive/My Drive/Salud_Digital/SU_data.xlsx')  
print(df.head())
```

Entonces

- La **Ciencia de Datos** se trata de transformar datos en conocimiento para apoyar la toma de decisiones, resolver problemas y crear oportunidades.
- Utilizamos la librería **Pandas** para trabajar con la clase DataFrame y aprendimos a utilizar algunos de sus métodos básicos, tales como la creación de **DataFrames** a partir de tablas Excel, la manipulación de datos, y la extracción de información relevante de ellos.
- En la siguiente cápsula aprenderemos más métodos necesarios para obtener **estadística descriptiva** y aprenderemos técnicas de visualización básicas para expresar los *insights* que obtendremos

