

**Universidad Finis Terrae**  
**Segundo Semestre 2023**  
**Asignatura: TOPIC AVANZADO EN DATA SCIENCE**

---

**Profesor: Darío Rojas**



---

# Laboratorio 1

## TÓPIC AVANZADO EN DATA SCIENCE

---



**Autores:**

Alexandra Vargas

Carlos Ramirez

Nicolas Navarro

**Sección:**

30 de Agosto del 2023

## Adquisición de Datos

Dado un sensor o un método para obtener estadísticas de uso del sistema operativo (ej: psutil o top):

Psutil (process and system utilities), es una librería Python que obtiene información de los procesos y utilización del sistema como pueden ser la cpu, memoria, discos, redes o sensores y que funciona bajo Windows, MacOS o Linux.

Para poder utilizarlo en Windows, es necesario instalar el siguiente complemento de python, con la ayuda de la consola de comandos:

```
pip install psutil
```

1. Obtenga una medición del sensor/componente en forma individual. Comprenda su funcionamiento y proponga una evaluación de la medición bajo dos condiciones distintas. No intente ser tan formal en la definición

Se obtendrán mediciones en la red del equipo, específicamente la entrada y salida de datos, en el primer caso se realizarán mediciones con el equipo tratando de tener un mínimo tráfico de datos y para la segunda condición, se medirá el tráfico de datos al estar dentro de una partida del videojuego llamado “Apex Legends”, esto con la finalidad de conocer la diferencia entre el consumo de datos y poder realizar las relaciones correspondientes. A continuación se adjunta el script con el que se toman los datos estipulados.

### Script:

```
import psutil

import time

import openpyxl

# Almacenamiento de datos

Excel = openpyxl.Workbook()

hoja = Excel.active

hoja.title = "Entradas_Salidas Minimo"

hoja.append(["Iteración", "bytes enviados", "bytes recibidos", "Paquetes enviados", "Paquetes recibidos",
"Total errores al recibir",

"Total errores al enviar", "Paquetes recibidos descartados", "Paquetes enviados descartados"])

#Recoleccion de datos
```

```

Tiempo_Ejecucion = 300 #15 Minutos = 900 Segundos

T_Recolector = 15

N_para_For = Tiempo_Ejecucion // T_Recolector #Division Entera

datosRed = [] # Lista de datos

for i in range(N_para_For):

    datos = psutil.net_io_counters()

    datosRed.append(datos)

    hoja.append([i+1, datos.bytes_sent, datos.bytes_recv, datos.packets_sent, datos.packets_recv, datos.errin,
datos.errout, datos.dropin, datos.dropout])

    time.sleep(T_Recolector)

for idx, datos in enumerate(datosRed): #Imprimir en consola

    print(f"Iteración {idx + 1}: {datos}")

# Almacenamiento de datos

hojad = Excel.create_sheet("Entradas_Salidas Maximo")

hojad.append(["Iteración", "bytes enviados", "bytes recibidos", "Paquetes enviados", "Paquetes
recibidos", "Total errores al recibir",

            "Total errores al enviar", "Paquetes recibidos descartados", "Paquetes enviados descartados"])

for id in range(N_para_For):

    datos = psutil.net_io_counters()

    datosRed.append(datos)

    hojad.append([id+1, datos.bytes_sent, datos.bytes_recv, datos.packets_sent, datos.packets_recv,
datos.errin, datos.errout, datos.dropin, datos.dropout])

    time.sleep(T_Recolector)

Nombre_Excel = "Psutil_Red_Recoleccion.xlsx"

Excel.save(Nombre_Excel)

print("Datos Guardados en --> ", Nombre_Excel)

```

**bytes\_sent (número de bytes enviados):** Los bytes enviados por segundo contador de rendimiento notifican el flujo de datos total en bytes desde el servidor integrador de transacciones (TI) al host por segundo. Esto se convierte en un contador útil cuando las transacciones de TI se vuelven grandes y los vínculos de comunicación pueden convertirse en tes los vuelve a combinar el ordenador o el dispositivo que los recibe.

**paquetes\_recv:** número de paquetes recibidos, número de respuestas recibidas a los pings enviados.

**errin:** número total de errores al recibir

**errout:** número total de errores durante el envío

**dropin:** número total de paquetes entrantes que se descartaron

**abandono:** número total de paquetes salientes que se descartaron el cuello de botella en el sistema.

**bytes\_recv (número de bytes recibidos):** Muestra el número total de paquetes de difusión recibidos y procesados por este dispositivo.

**paquetes\_enviados:** un paquete es un pequeño segmento de un mensaje más grande. Los datos enviados a través de redes informáticas\*, como Internet, se dividen en paquetes.

2. Obtenga varias mediciones del sensor/componente en forma periódica y responda:

a. **Defina frecuencia de muestreo. Argumente su decisión**

Se definió una frecuencia de muestreo de 15. Esto significa que estamos tomando mediciones o muestras de los datos del sensor o componente a cada 15 segundos. Este enfoque nos permite capturar de manera efectiva las variaciones significativas en tiempo real. Además, proporciona una precisión de los datos necesarios para aplicaciones en las que es crucial obtener información actualizada y tomar decisiones basadas en esos datos de manera rápida y oportuna.

b. **Defina tiempo de muestreo. Argumente su decisión**

Se definió un tiempo de muestreo de 15 minutos, ya que los datos cambian lentamente por lo que se puede utilizar un periodo de tiempo de muestreo más largo para reducir la cantidad de mediciones y el consumo de recursos.

c. **Defina codificación del muestreo. Argumente su decisión**

No se definió una codificación de muestreo específica.

3. Determinar dos condiciones de la experiencia para realizar la medición (comparables):

- a. Grupo Experimental: las mediciones que son intervenidas. Explique cuál será su intervención, es decir la causa que desea evaluar. Argumente

La intervención seleccionada consistió en el aumento del tránsito de datos al jugar un videojuego, con el objetivo de conocer y analizar la variación entre la cantidad de bytes. En donde se realizaron los siguientes pasos a continuación:

- 1) Se seleccionó el juego de Valorant, debido a sus prominentes elementos en línea y multijugador, ya que este juego tienden a demandar un mayor flujo de datos para garantizar la sincronización entre los jugadores y el servidor.
- 2) Se garantizó una conexión a internet estable y rápida, para poder maximizar el flujo de datos.
- 3) Se empleó un dispositivo equipado con un potente procesador y suficiente memoria para evitar posibles restricciones en el flujo de datos.
- 4) Se incrementó la configuración gráfica del juego para aumentar la carga de datos y observar la variación en la cantidad de bytes transmitidos.

- b. Grupo Control: las mediciones bases sin intervención. Explique por qué es un caso base. Argumente

El grupo control se utiliza como un caso base en un experimento o estudio para proporcionar un punto de comparación con el grupo experimental. Es decir, que es un grupo en el que no se implica ninguna intervención o tratamiento, por lo que las mediciones realizadas en este grupo representan la condición natural en ausencia de cualquier influencia externa.

4. Describa las variables que controlará en su experimento:

Las variables que controlan el experimento son las siguientes:

- **bytes\_sent:** informa el total de bytes enviados desde el host. Los bytes recibidos junto con los bytes enviados representan la transferencia total de datos en los enlaces de comunicaciones entre el servidor y la computadora host.

Los bytes enviados por el segundo contador de rendimiento notifican el flujo de datos total en bytes desde el servidor integrador de transacciones (TI) al host por segundo. Esto se convierte en un contador útil cuando las transacciones de TI se vuelven grandes y los vínculos de comunicación pueden convertirse en el cuello de botella en el sistema. Esto es especialmente importante cuando se usan WAN.

- **bytes\_recv:** esta variable muestra el número total de paquetes de difusión recibidos y procesados por este dispositivo, .
- **paquetes\_recv:** esta variable muestra el número total de paquetes los cuales serán recibidos
- **paquetes\_enviados:** esta variable mostrará el total de paquetes enviados

5. Desarrollar experimento y almacenar mediciones en uno o varios archivos (incluir los archivos en el envío)

Los resultados de la aplicación del experimento se guardan en el excel adjunto llamado “Psutil\_Recoleccion.xlsx”

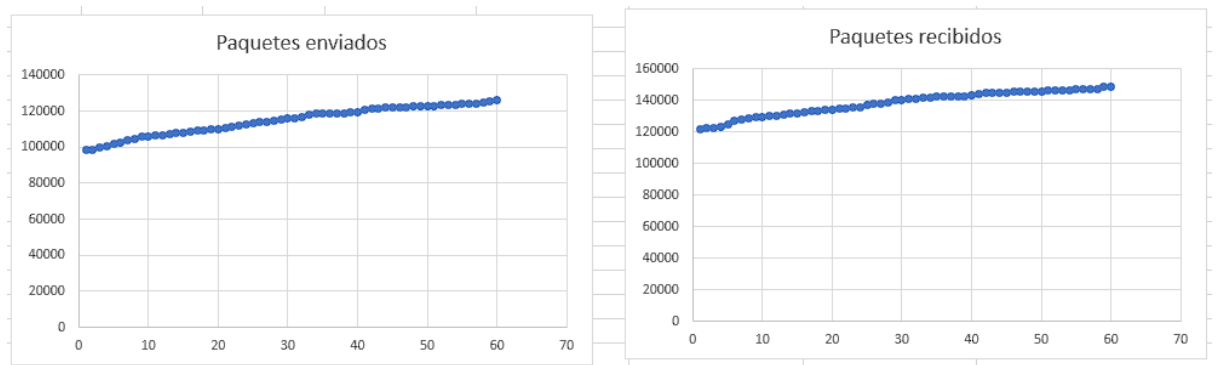
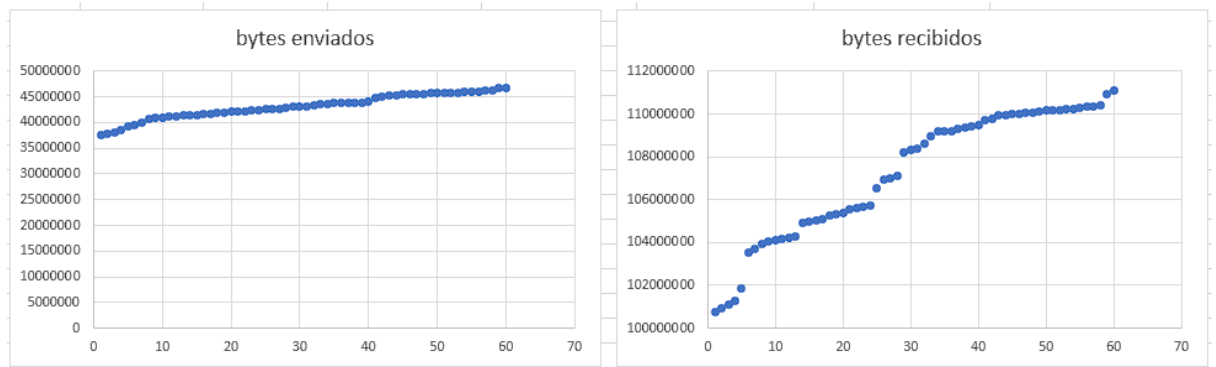
6. Utilizar estadística descriptiva básica para comparar experimentos (media, desviación estándar, máximo, mínimo y gráfico acorde a la lectura). Incluir aquí

Estadísticos ambas condiciones

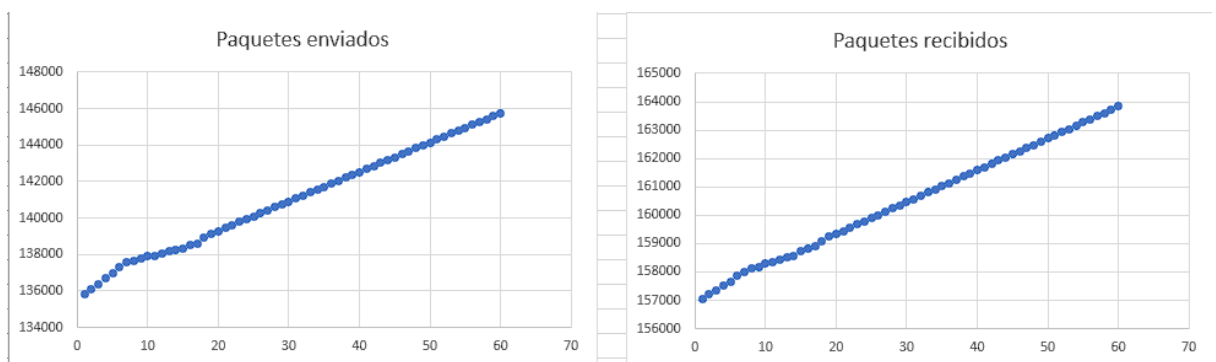
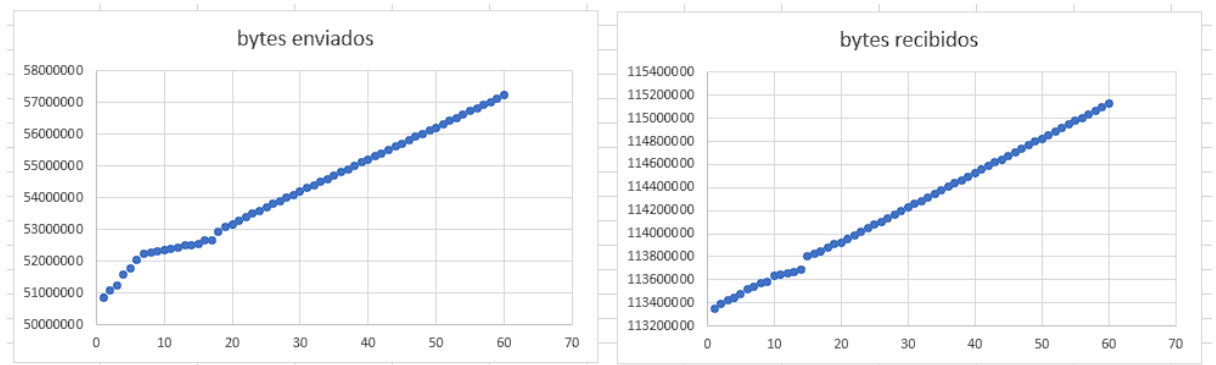
1	E_S Minimo byte Enviado					E_S Maximo byte Enviado	
2	Media	54252953,57				Media	43141682,97
3	Desv. Estandar	1768398,037				Desv. Estandar	2377437,849
4	Maximo	57233715,24				Maximo	46837616
5	Minimo	50868442				Minimo	37695980
6							
7	E_S Minimo byte Recibido					E_S Maximo byte Recibido	
8	Media	114240221,5				Media	107256190,3
9	Desv. Estandar	524619,9255				Desv. Estandar	2986670,181
10	Maximo	115126143,1				Maximo	111100900
11	Minimo	113351415				Minimo	100758407

13	E_S Minimo Paquete Enviado					E_S Maximo Paquete Enviado	
14	Media	140989,4252				Media	114726,4333
15	Desv. Estandar	2824,751533				Desv. Estandar	7915,793386
16	Maximo	145756,7361				Maximo	126150
17	Minimo	135798				Minimo	98281
18							
19	E_S Minimo Paquete Recibido					E_S Maximo Paquete Recibido	
20	Media	160514,0744				Media	138362,5167
21	Desv. Estandar	1965,067474				Desv. Estandar	7693,479351
22	Maximo	163832,4917				Maximo	148944
23	Minimo	157025				Minimo	121851

## Gráficos Máximo



## Graficos Minimo



7. Según estadística descriptiva, verificar si la hipótesis se cumple (no es necesario hacer test de hipótesis). Argumentar

Como se puede observar en los gráficos de la pregunta anterior:

Se calcularon múltiples medidas descriptivas, como la dispersión en la cual se calcula la desviación estándar, la media o también el rango intercuartílico del volumen de datos, generando así los gráficos máximos y los gráficos mínimos. En los gráficos máximos se observa un aumento significativo en el tráfico de datos, por un elevado tránsito de bytes, por lo que la hipótesis se cumple, debido a que se teorizaba que al solicitar una mayor cantidad de internet y banda ancha, el tráfico tenderá al alza en base a los datos obtenidos con el computador casi en modo estático.

8. Responder las siguientes preguntas brevemente:

- a. ¿Es una adquisición experimental o no-experimental? Argumente.

Respuesta: Esta adquisición es del tipo experimental, debido a que se estipula un caso base o genérico, el cual sería que el equipo este con el mínimo de tráfico de datos, mientras que el segundo caso que sería aumentar el tráfico de datos de manera significativa, esto significa que existe una intervención directa, debido a que se transforma el sistema (El equipo con conexión a internet) y se busca encontrar un cambio en los valores analizados.

- b. ¿Su análisis es deductivo o inductivo? Argumente

Respuesta: Este corresponde a un análisis deductivo, debido a que se comienza con un caso general (el equipo con mínimo tráfico de datos) y se le aplica un caso particular, el cual sería analizar el tráfico de datos con un videojuego específico.

- c. ¿Qué aspectos cualitativos y cuantitativos tiene la experiencia? Argumente.

Respuesta: Los aspectos cuantitativos de este análisis son los valores entregados por el método desarrollado, debido a que todos son valores numéricos que se sabe específicamente lo que significan. En esta recolección de datos no se encuentran valores cualitativos, debido a que no se maneja información del tipo cualidades o atributos.

## Referencias

- 1) *PSUTIL Documentation* — *PSUTIL 5.9.6 documentation*. (s. f.). <https://psutil.readthedocs.io/en/latest/>
- 2) *OpenPyxl - a Python library to read/write Excel 2010 xlsx/xlsm files* — *OpenPyxl 3.1.2 documentation*. (s. f.). <https://openpyxl.readthedocs.io/en/stable/>