



Tecnológico
de Monterrey

| Educación
Continua

Plataformas de Aplicación sobre Python

Programas LIVE

SS2 | Preparar para aplicar

Bienvenida e Introducción

- Agenda
- Objetivos de la sesión
- Panorámica de la sesión
- Bienvenida
- Introducción

Agenda

1

Sesión Síncrona 1

Aprender

2

Trabajo asíncrono 1

Profundizar | Ruta de Aprendizaje

3

Sesión Síncrona 2

Preparar para Aplicar

4

Trabajo asíncrono 2

Aplicar en el trabajo | Reto

1

2

3

4



Objetivos particulares de la sesión



- Utilizarás adecuadamente las **estructuras de datos** de **NumPy** y **Pandas** para manipular información

Panorámica de la sesión

1 | Integrar Aprender + Profundizar

2 | Práctica (ciclos)

3 | Reflexión

4 | Plan de aplicación

Cierre de sesión

Bienvenida | Dinámica



 **Mentimeter**

Please enter the code

<https://www.menti.com>

Introducción



- La omisión de las plataformas **NumPy** y **Pandas** en tus proyectos, puede derivar en esfuerzos excesivos de programación, pero si tú las conoces y haces buen uso de ellas, **obtendrás resultados en confiables y en muy poco tiempo**
- Al finalizar la sesión, serás capaz de manipular **grandes cantidades de datos** provenientes de archivos con diversos formatos para **procesarla, resumirla y obtener información valiosa**

Panorámica de la sesión

1 | Integrar Aprender + Profundizar

2 | Práctica (ciclos)

3 | Reflexión

4 | Plan de aplicación

Cierre de sesión

1 | Integrar

Aprender + Profundizar

- 1.1 Aprender + Profundizar
- 1.2 Conexión
- 1.3 Aprendizajes



Integrar
Aprender + Profundizar

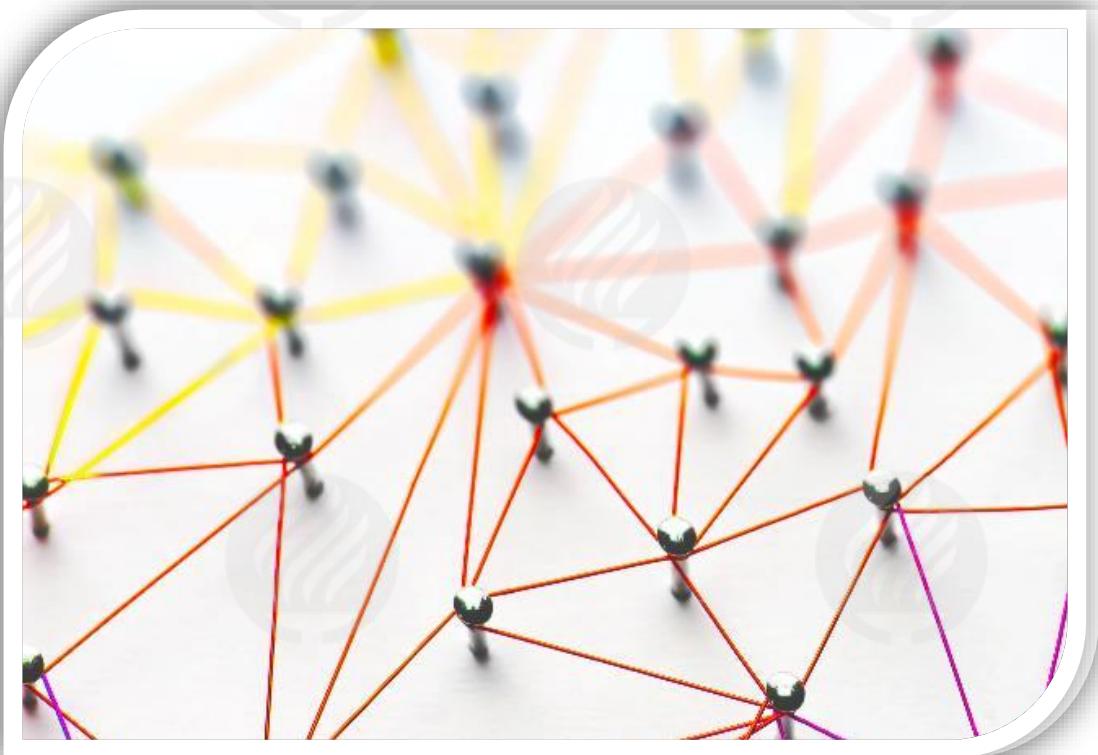
1. Integrar | Aprender + Profundizar



Antes de comenzar nuestras prácticas, pensemos en lo siguiente:

- ¿Notaste la relación entre los conocimientos de la sesión **síncrona y la asíncrona**?
- ¿Qué tan **útil** consideras lo aprendido?
- ¿Cuál podría ser el **primer proyecto** en el que consideres **aplicar estos conocimientos**?

1. Integrar | Conexión



- Ingresa a <https://kahoot.it/> con el **PIN** que te proporcione en el chat

1. Integrar | Aprendizajes

- Instalar/verificar/usrar plataformas en **Python**
- Usar **arreglos** de 1 y 2 dimensiones en **NumPy**
- Emplear **series y dataframes** de **Pandas**
- **Recuperar y almacenar datos** de diferentes tipos de archivos por medio de **Pandas**
- Seleccionar **información importante**
- Crear y eliminar **columnas/filas** en un **dataframe** de **Pandas**
- **Realizar cálculos** con la información de un **dataframe**



Panorámica de la sesión

1 | Integrar Aprender + Profundizar

2 | Práctica (ciclos)

3 | Reflexión

4 | Plan de aplicación

Cierre de sesión



Práctica

2 | Práctica 1

Analizando algunos Tweets de la vacuna de Pfizer

- 2.1 Contexto
- 2.2 Instrucciones
- 2.3 Trabajo en equipo
- 2.4 Integración del trabajo

2.1 Práctica | Contexto



- La creación de la vacuna para combatir el **virus COVID-19** ha generado **diversas opiniones alrededor del mundo**
- Se ha decidido **recolectar tweets** sobre la vacuna de Pfizer y BioNTech usando la plataforma **tweepy** para accesar al **API de Twitter** y con ello **generar un archivo con la información separada por comas (formato csv)**
- Un fragmento de dicho archivo, recuperado del portal **kaggle.com**, será el punto de partida para **explorar la información y realizar diversos análisis de utilidad**

2.2 Práctica | Instrucciones



1. Descarga el archivo [vaccination_tweets.csv](#)
2. Ábrelo con MS Excel para **analizar** su estructura y la **cantidad de registros que contiene**
3. Genera una **libreta en Google Colab** para esta práctica
4. Lee, en un **dataframe**, todos los registros del archivo. Incluye las columnas:
 - ***user_name***
 - ***user_location***
 - ***user_followers***
 - ***user_Friends***
 - ***Source***
 - ***retweets***
 - ***favorites***

2.3 Práctica | Trabajo en equipo



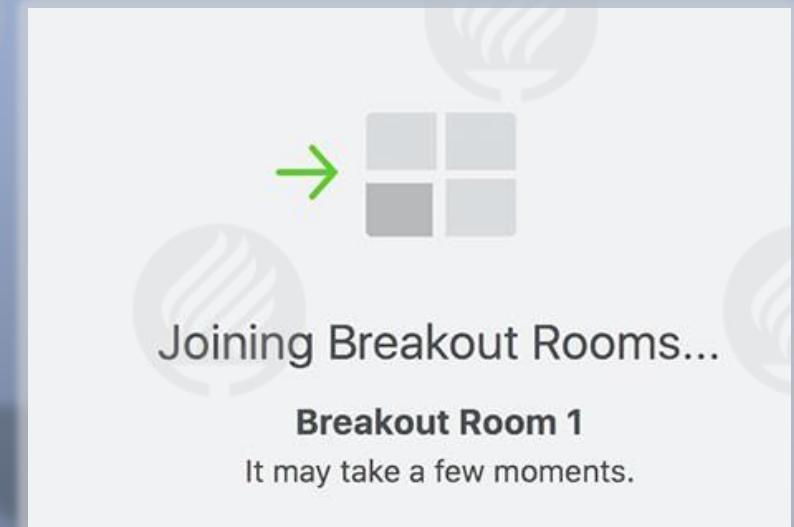
Del *dataframe* anterior, obtén la siguiente información relevante:

1. Los **5 usuarios** con más seguidores
2. La **cantidad promedio** de amigos que tienen los usuarios
3. La cantidad de **usuarios por cada fuente** de donde hicieron el tweet
4. El **promedio de retweets**, sin considerar a aquellos usuarios que no tienen retweets
5. Los **5 primeros usuarios** que tienen más de 100 favoritos

Tiempo: **25 minutos**

Zoom | BreakOut Rooms (BOR)

- 1. Al unirte,
aparecerá un
mensaje**
- 2. Después estarás
con un grupo
pequeño de
compañeros**



2.4 Práctica | Integración del trabajo en equipo



1. ¿Cuál consideras que fue la **dificultad** de la práctica?
2. ¿Cómo **separaron** la información?
3. ¿Qué **funciones** se utilizaron?
4. ¿Qué otra **información relevante** crees poder obtener con los datos que están en el archivo?

2 | Práctica 2

Análisis del progreso de vacunación mundial

- 2.1 Contexto
- 2.2 Instrucciones
- 2.3 Trabajo en equipo
- 2.4 Integración del trabajo en equipo



Práctica

2.1 Práctica | Contexto



Referencias:

Github – www.github.com
Kaggle – www.kaggle.com

- Continuando con el tema del virus **COVID-19**, analizarás información relacionada con la **aplicación de todas las vacunas en el mundo**
- En el repositorio **Our World in Data** de **GitHub**, se almacenan datos en relación a la cantidad de **aplicaciones diarias de las vacunas por país**
- Utilizarás una **muestra** reciente del archivo que los contiene, recuperada del portal **kaggle.com**, para obtener **resúmenes significativos y precisos**

2.2 Práctica | Instrucciones



1. Descarga el archivo [country_vaccinations.csv](#)
2. Ábrelo con MS Excel para **analizar** su estructura y **la cantidad de registros que contiene**
3. Genera una **libreta en Google Colab** para esta práctica
4. Para cada uno de los requerimientos que a continuación se presentan, crea un **dataframe** que contenga únicamente la **información necesaria para dar solución a lo solicitado**

2.3 Práctica | Trabajo en equipo



Obtén la siguiente información relevante:

1. La cantidad de **vacunas aplicadas** de cada **compañía**
2. La cantidad de **vacunas aplicadas** en **todo el mundo**
3. El promedio de **vacunas aplicadas** por **país**
4. La cantidad de **vacunas aplicadas** el día **29/01/21** en todo el mundo

2.3 Práctica | Trabajo en equipo

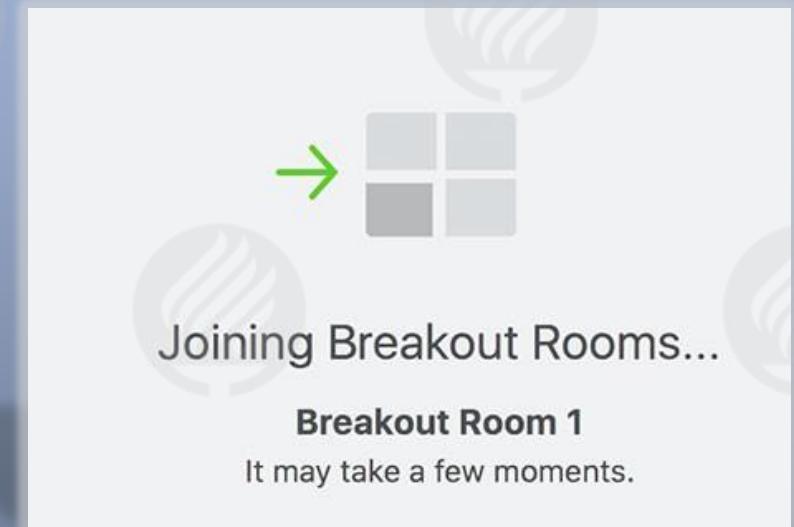


5. Una columna **derivada con las diferencias de aplicación** entre las columnas **daily_vaccionations** y **daily_vaccionations_raw**
6. La **diferencia de vacunación** entre los días 20/01/21 y 31/01/212 para México y EEUU
 - Guardar todos los **dataframes usados** en formato de archivo de Excel

Tiempo: **25 minutos**

Zoom | BreakOut Rooms (BOR)

- 1. Al unirte,
aparecerá un
mensaje**
- 2. Después estarás
con un grupo
pequeño de
compañeros**



2.4 Práctica | Integración del trabajo en equipo



1. ¿Cuál consideras que fue el nivel de **dificultad** de la práctica?
2. ¿Notaste cómo puedes combinar **distintos requerimientos** para obtener un resultado?
3. ¿Te das cuenta de que **guardar el dataframe** te permite seguir explotándolo en otro momento?
4. ¿Qué otra **información relevante** crees poder obtener con los datos que están en el archivo?

Panorámica de la sesión

1 | Integrar Aprender + Profundizar

2 | Práctica (ciclos)

3 | Reflexión

4 | Plan de aplicación

Cierre de sesión

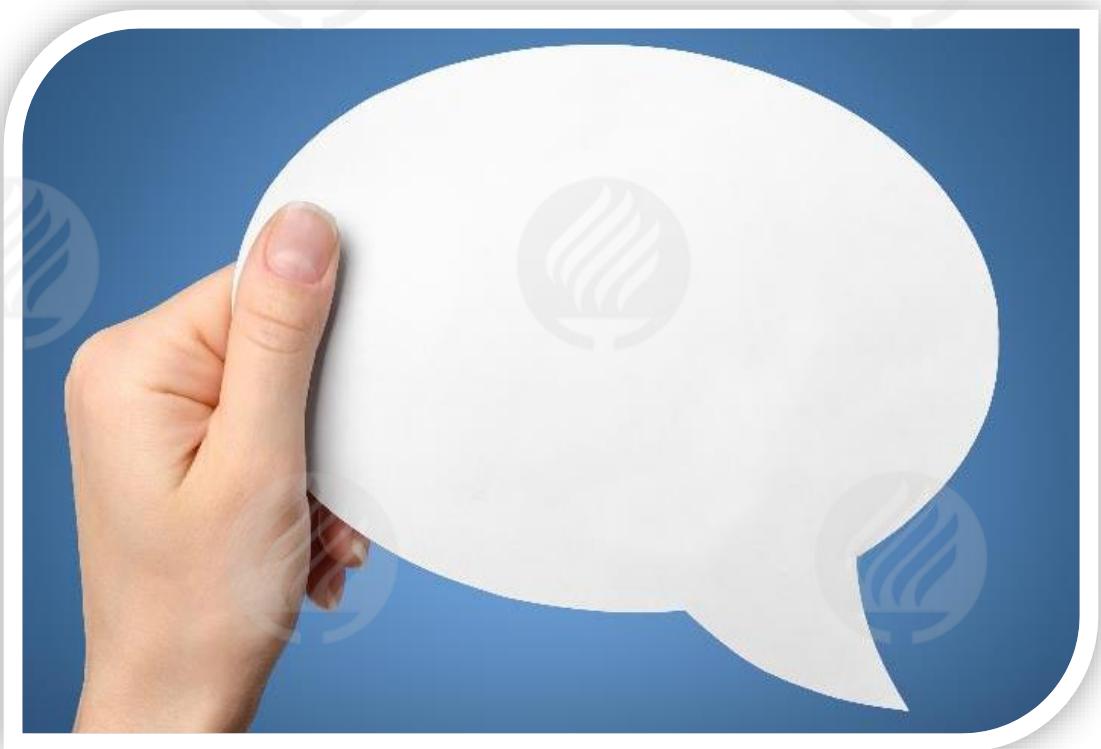
3 | Reflexión General Ciclo de prácticas

- 3.1 Tips y reflexiones
- 3.2 Resultados y hallazgos



Reflexión

3.1 Reflexión General: Resultados y Hallazgos obtenidos de las prácticas



En este momento comprobaste que tienes la capacidad de **manipular y procesar** grandes cantidades de datos **para obtener información valiosa**

- a. Comenta sobre la **importancia** de la información obtenida
- b. Menciona si crees que hay **otras formas de obtener los mismos resultados**
- c. ¿Crees que los datos siempre los podrás **encontrar** en un único archivo?
- d. ¿Cómo te gustaría representar **los resultados obtenidos**?

3.2 Reflexión General: Tips y Reflexiones de la práctica



- Siempre utiliza **plataformas** para facilitar el desarrollo y asegurar la **calidad y consistencia de los resultados**
- Cuando sea necesario procesar información numérica, **NumPy** es la **plataforma recomendada y más ampliamente usada**

3.2 Reflexión General: Tips y Reflexiones de la práctica



- Para **recuperar o almacenar** información de archivos con diferentes formatos, **Pandas** es una plataforma **reconocida por su completitud y simpleza**
- Siempre es mejor trabajar con **subconjuntos de datos** que con todos los datos disponibles
- No olvides verificar la **estructura de tu dataframe** para ver qué operaciones puedes realizar **con los datos que contiene**

Panorámica de la sesión

1 | Integrar Aprender + Profundizar

2 | Práctica (ciclos)

3 | Reflexión

4 | Plan de aplicación

Cierre de sesión



4 | Plan de aplicación

Análisis de deserción de empleados

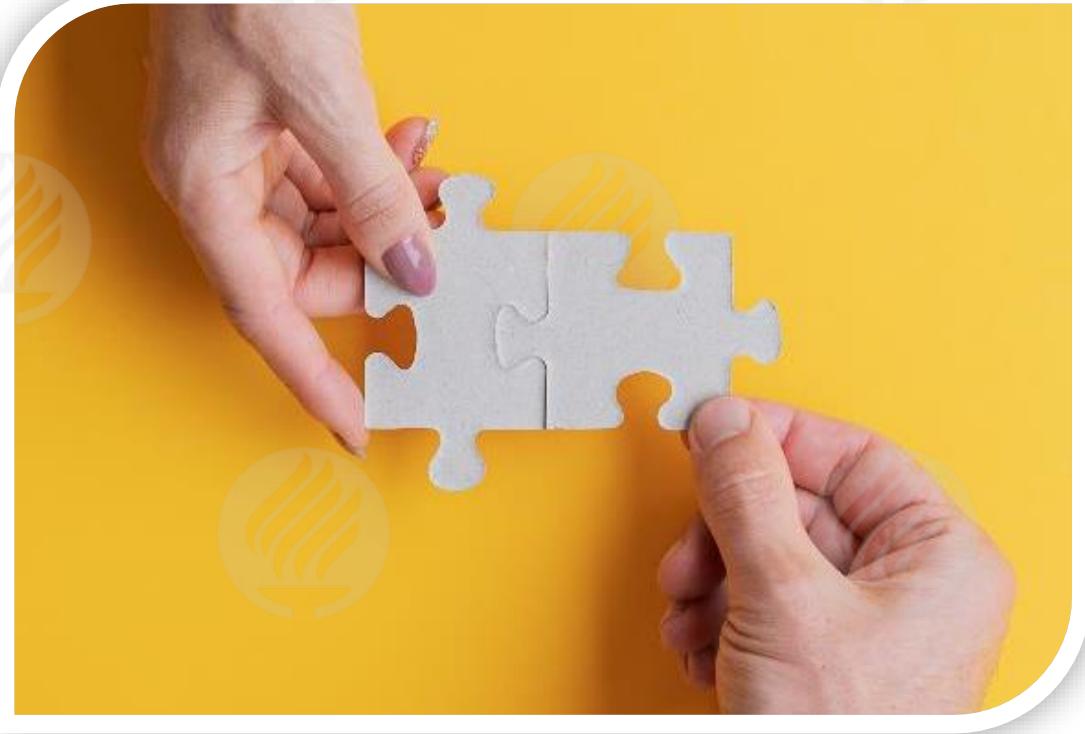
1 Sesión Síncrona 1
Aprender

2 Trabajo asíncrono 1
Profundizar | Ruta de Aprendizaje

3 Sesión Síncrona 2
Preparar para Aplicar

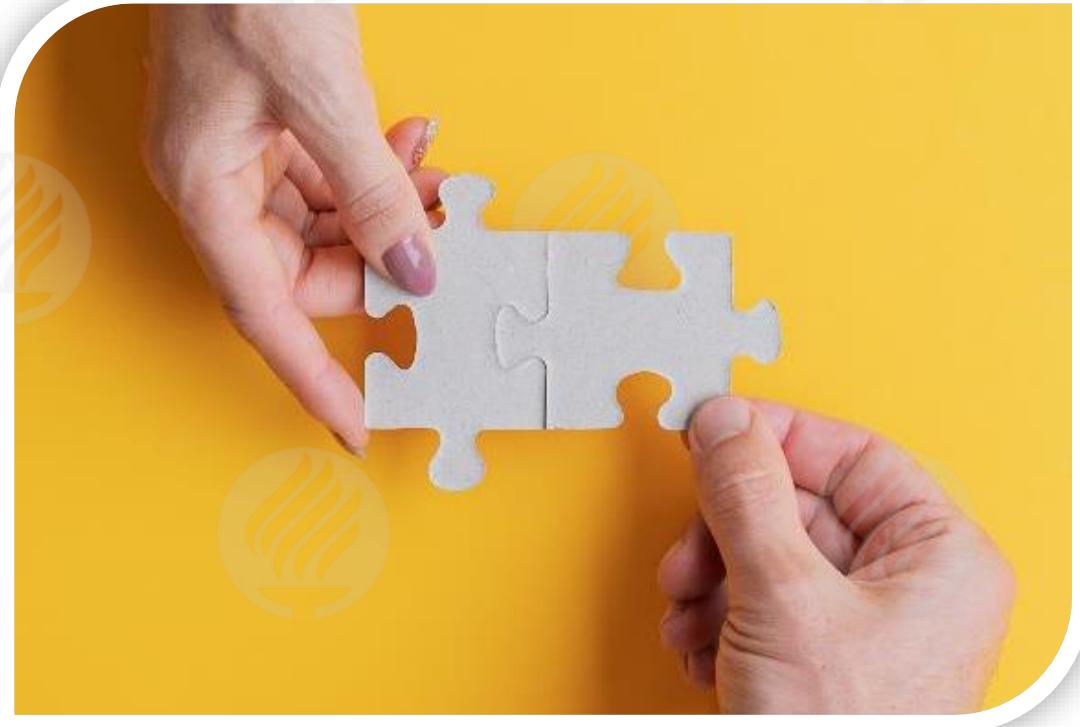
4 Trabajo asíncrono 2
Aplicar en el trabajo | Reto

Plan de aplicación: Reto



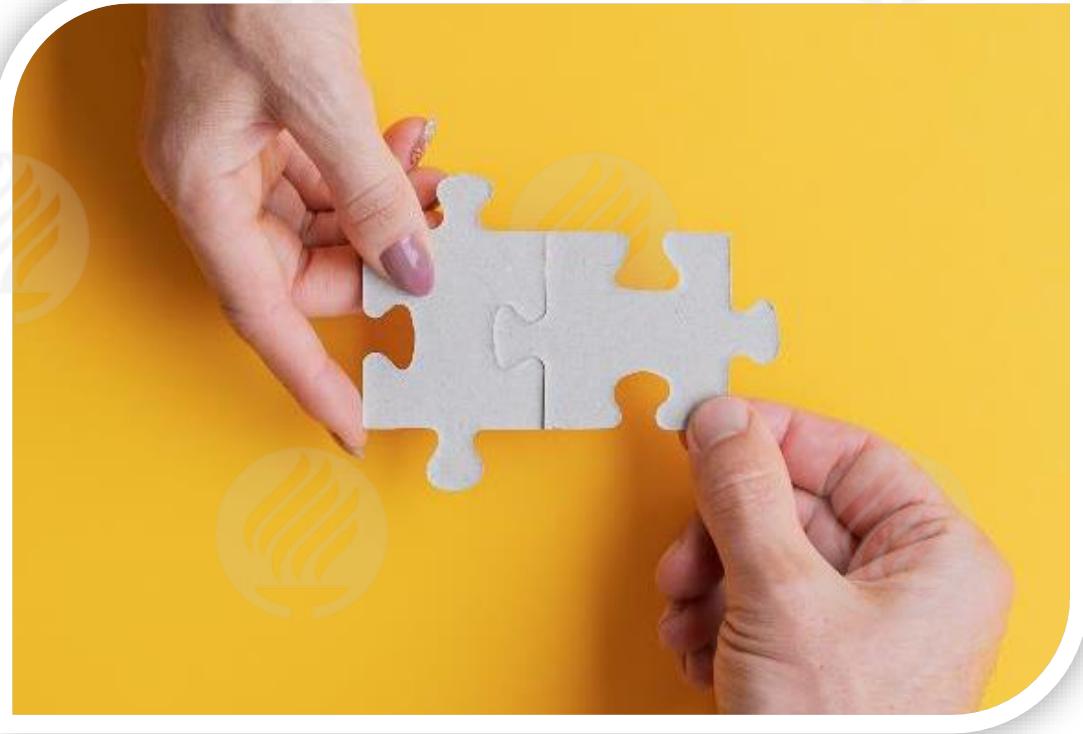
- Para el reto utilizarás uno de los archivos provistos en el **Hackathon HackerEarth 2020** que contiene información diversa de los empleados de una compañía
- Con esta información se busca predecir la **tasa de deserción** para que las empresas puedan **aplicar un plan correctivo**
- El archivo contiene las **siguientes columnas: género, edad, educación, estado civil, escala de pago, tiempo de servicio, tiempo de la última promoción, entre otros**

Plan de aplicación: Reto



1. Descarga el archivo: [Employee Attrition.csv](#)
2. Genera una **libreta en Google Colab** para resolver el reto
3. Recupera la **información** del archivo
4. Añade una **sección de texto** para cada requerimiento que **anteceda a la solución**
5. Guarda los **dataframes resultantes** en archivos xls
6. Al finalizar, **comparte la libreta con tu profesor**

Plan de aplicación: Reto | Rúbrica



Criterio	Puntaje
Libreta en Google Colab con secciones documentadas y compartirlo con el profesor	5
Recupera la información	5
Obtén la información relevante requerida para contestar las preguntas	20
Guardar los dataframes en XLSX	5
Total de puntos	35

Panorámica de la sesión

1 | Integrar Aprender + Profundizar

2 | Práctica (ciclos)

3 | Reflexión

4 | Plan de aplicación

Cierre de sesión

Cierre Sesión Sincrónica 2

- Integración final de aprendizajes
- Reflexión final
- Concepto clave | TakeAway



Cierre de sesión

Cierre | Sesión Sincrónica 2



- En esta sesión realizaste dos prácticas que permitieron mejorar tu **entendimiento y habilidad** para:
 - ✓ Recuperar y almacenar datos
 - ✓ Seleccionar datos relevantes
 - ✓ Crear **columnas derivadas**
 - ✓ Procesar datos para obtener información resumida
- Ahora piensa en la **información** que puedes obtener a partir de los datos en tu trabajo, **aplicando estas habilidades y utilizando las plataformas estudiadas**

Cierre: Concepto clave

Contar con el conocimiento y habilidades para recuperar, seleccionar, procesar y almacenar información de manera sencilla, rápida y confiable es fundamental por los grandes volúmenes de información que se generan y se manipulan en la actualidad.



Tecnológico de Monterrey | 2021

Prohibida la reproducción total o parcial de esta
obra sin expresa autorización del Tecnológico
de Monterrey

Gracias | Programas LIVE