



Tecnológico
de Monterrey

| Educación
Continua

Manipulación de datos en 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



- Practicarás la **generación de reportes**, procesando más de un archivo de **información**

Panorámica de la sesión

1 | Integrar Aprender + Profundizar

2 | Práctica (ciclos)

3 | Reflexión

4 | Plan de aplicación

Cierre de sesión

Introducción



- Hay muchas maneras de **manipular** la información, algunas **más eficientes que otras**
- Utilizar una herramienta **no adecuada** puede traducirse en **esfuerzos y en tiempo que podría aprovecharse de mejor manera**
- **Pandas** nos ofrece algunas funciones para manipular de manera **eficiente una colección grande de datos**

1 | Integrar Aprender + Profundizar

- 1.1 Aprender + Profundizar
- 1.2 Conexión
- 1.3 Aprendizajes



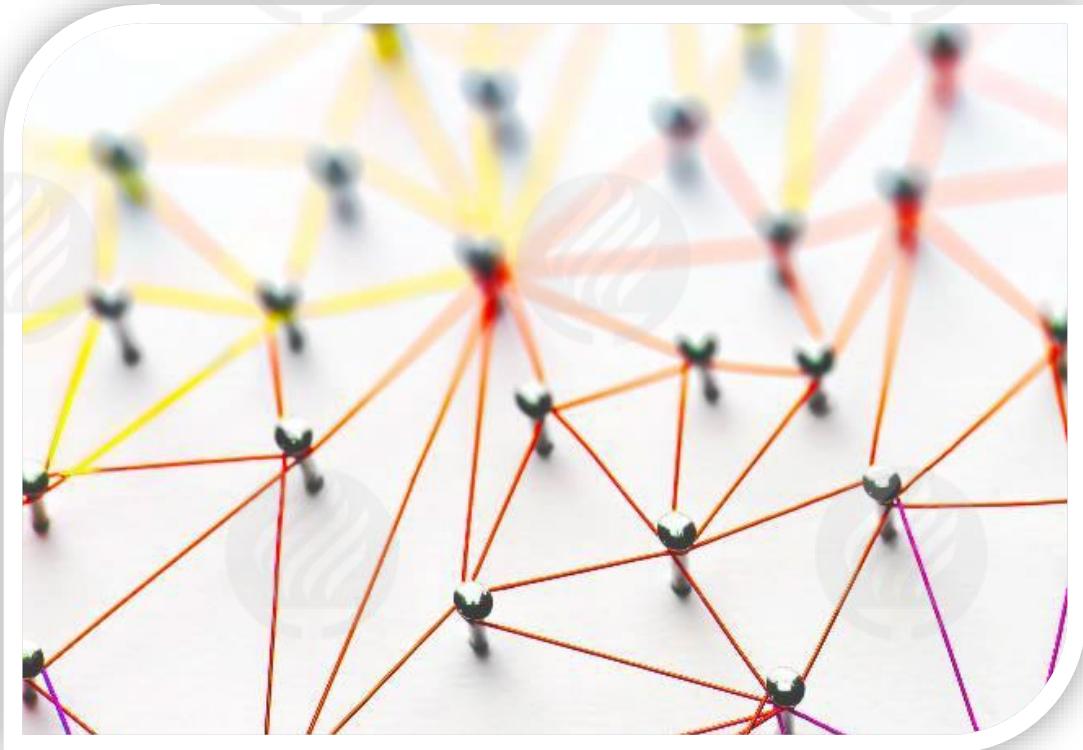
Integrar
Aprender + Profundizar

1. Integrar | Aprender + Profundizar



- Para iniciar, recordemos los **contenidos presentados** en las dos primeras sesiones
- ¿Puedes pensar en un **escenario** en tu lugar de trabajo donde puedas **aplicar estos conceptos**?
- ¿Se te ocurre algo que puedas **hacer para mejorar algún proceso** o toma de decisión ?

1. Integrar | Conexión



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

Tiempo: **5 minutos**

1. Integrar | Aprendizajes



- Principios básicos de las **bases de datos**
- El **álgebra relacional** y sus operaciones
- Conjuntar información desde **diferentes fuentes** para manipularla
- Analogías entre **pandas** y uno de los **lenguajes más utilizados para manipular datos (SQL)**

Panorámica de la sesión

1 | Integrar Aprender + Profundizar

2 | Práctica (ciclos)

3 | Reflexión

4 | Plan de aplicación

Cierre de sesión

2 | Práctica 1

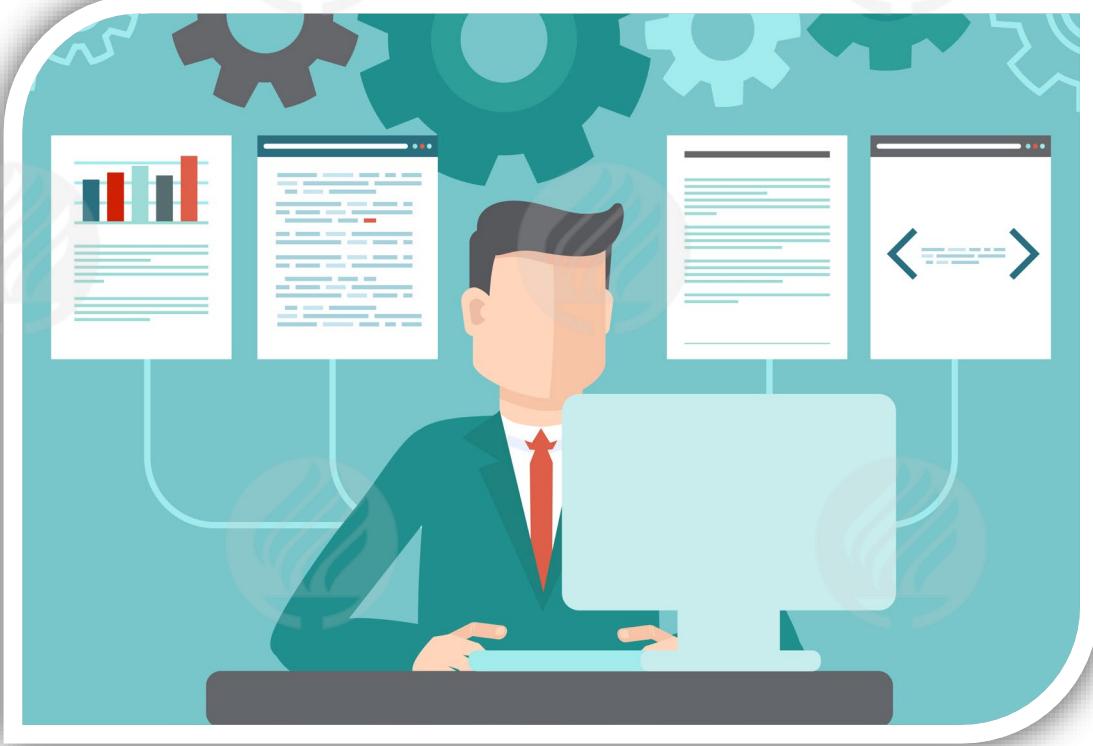
Personajes, especies y estatura (STAR WARS)

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



Práctica

2.1 Práctica | Contexto



- Muchas veces, la **información** que nos interesa se encuentra **organizada en diferentes archivos**
- En esta práctica juntaremos la información de **dos archivos** de forma que **tenga sentido la información conjuntada**
- Trabajaremos con **información recuperada del sitio kaggle.com**, relativa a la tan famosa franquicia de **star wars**, para ver que tanto coinciden las estaturas de los personajes con la estatura promedio de su especie

2.2 Práctica | Instrucciones



- Descarga los archivos **characters.csv** y **species.csv**
- De manera **opcional** puedes abrirlos con **MS Excel** para familiarizarte con el **contenido**, o si lo prefieres puedes ir directo al siguiente paso
- Carga el archivo de personajes en un **DataFrame** y el archivo de especies en otro **DataFrame**

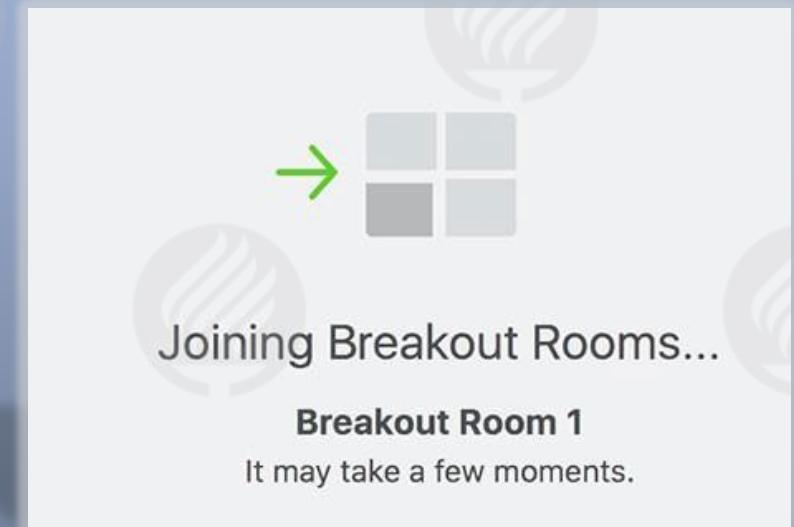
2.3 Práctica | Trabajo en equipo



- Crea un nuevo **DataFrame** con la **información relacionada entre los dos archivos leídos** (probablemente tendrás que utilizar funciones de pandas como **merge, join, concat o append**)
- Muestra una tabla que contenga la **estatura de cada personaje** junto con la **estatura promedio de su especie**
- Tiempo: **20 minutos** para realizar las distintas fases de la práctica

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. ¿Lograste **concluir** la práctica?
2. ¿Cuál consideras que fue la parte **más interesante** de la práctica?
3. ¿Qué **función** utilizaste para **relacionar correctamente la información** en los dos archivos?
4. ¿Qué otro **tipo de consultas** crees que se puedan **realizar con la información relacionada**?

Panorámica de la sesión

1 | Integrar Aprender + Profundizar

2 | Práctica (ciclos)

3 | Reflexión

4 | Plan de aplicación

Cierre de sesión

2 | Práctica 2

Población en el planeta de origen

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



Práctica

2.1 Práctica | Contexto



- Ahora vamos a trabajar con diferentes archivos que tienen **la misma estructura**, podemos pensar en la información de **diferentes sucursales de una tienda, estudiantes de diferentes grupos, etc.**
- También tendremos que resolver algunos **detalles** en el formato de la **información para poder realizar los análisis deseados**
- Seguiremos trabajando con archivos recuperados del sitio **kaggle.com**

2.2 Práctica | Instrucciones



- Descarga los archivos **characters_1.csv**, **characters_2.csv**, **characters_3.csv**, y **planets.csv**
- Carga cada uno de los archivos de personajes descargados en su propio **DataFrame**, y luego **junta la información de los 3 en uno solo**
- Carga la información del archivo de planetas en su propio **DataFrame** y **reúne la información de todos los personajes con la información de los planetas**

2.3 Práctica | Trabajo en equipo

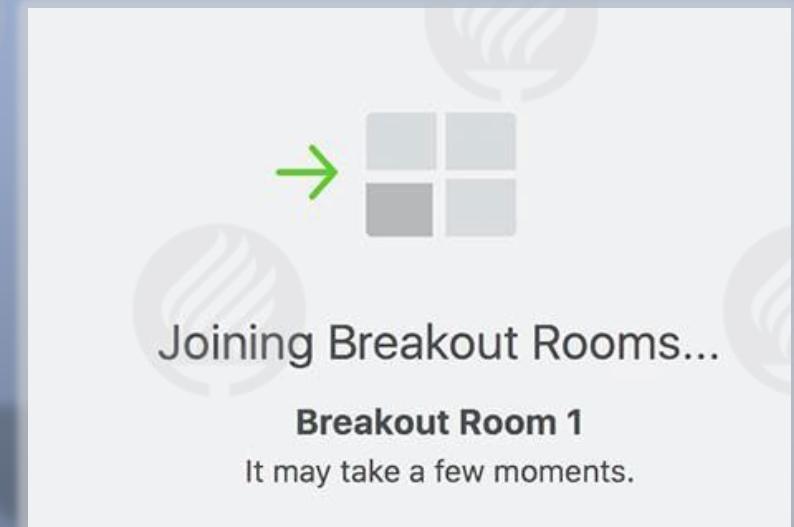


- Muestra una **tabla** que contenga a **cada personaje, su lugar de origen y la población en ese lugar**
- Esta tabla también debe mostrar **todos los planetas** con su población, **aunque ningún personaje sea de ahí**

Tiempo: **20 minutos** para realizar las **distintas fases de la práctica**

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. ¿Lograste **concluir** la práctica?
2. ¿Cuál consideras que fue la parte **más interesante** de la práctica?
3. ¿Qué **función** utilizaste para conjuntar la información de todos los personajes y para **relacionar correctamente la información con los planetas**?
4. ¿Qué otro **tipo de consultas** crees que se puedan **realizar de esta forma**?

Panorámica de la sesión

1 | Integrar Aprender + Profundizar

2 | Práctica (ciclos)

3 | Reflexión

4 | Plan de aplicación

Cierre de sesión

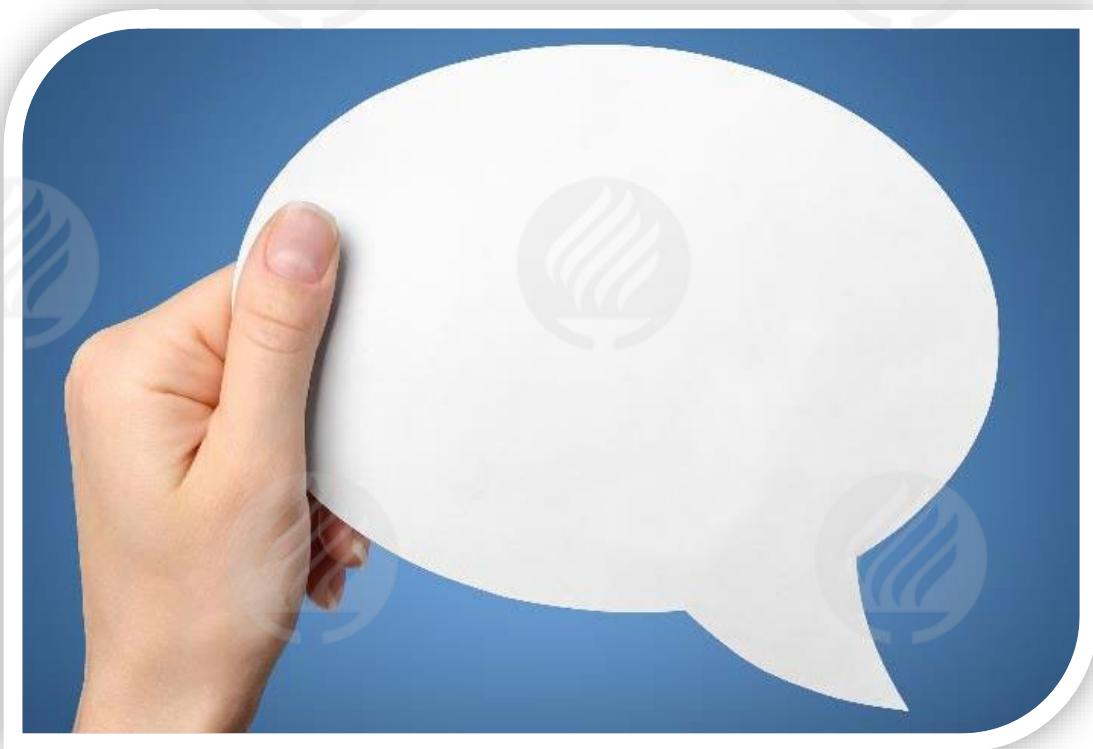
3 | Reflexión General Ciclo de prácticas

- 3.1 Tips y reflexiones
- 3.2 Resultados y hallazgos



Reflexión

3.1 Reflexión General: Resultados y Hallazgos obtenidos de las prácticas



- Con las prácticas realizadas, ahora eres capaz de **conjuntar información** relacionada **proveniente de varios archivos**
- ¿Puedes **diferenciar** cuándo usar **merge, join y concat** ?
- ¿Crees que podamos **mostrar la información obtenida** de alguna manera más "**digerible**" ?

3.2 Reflexión General: Tips y Reflexiones de la práctica



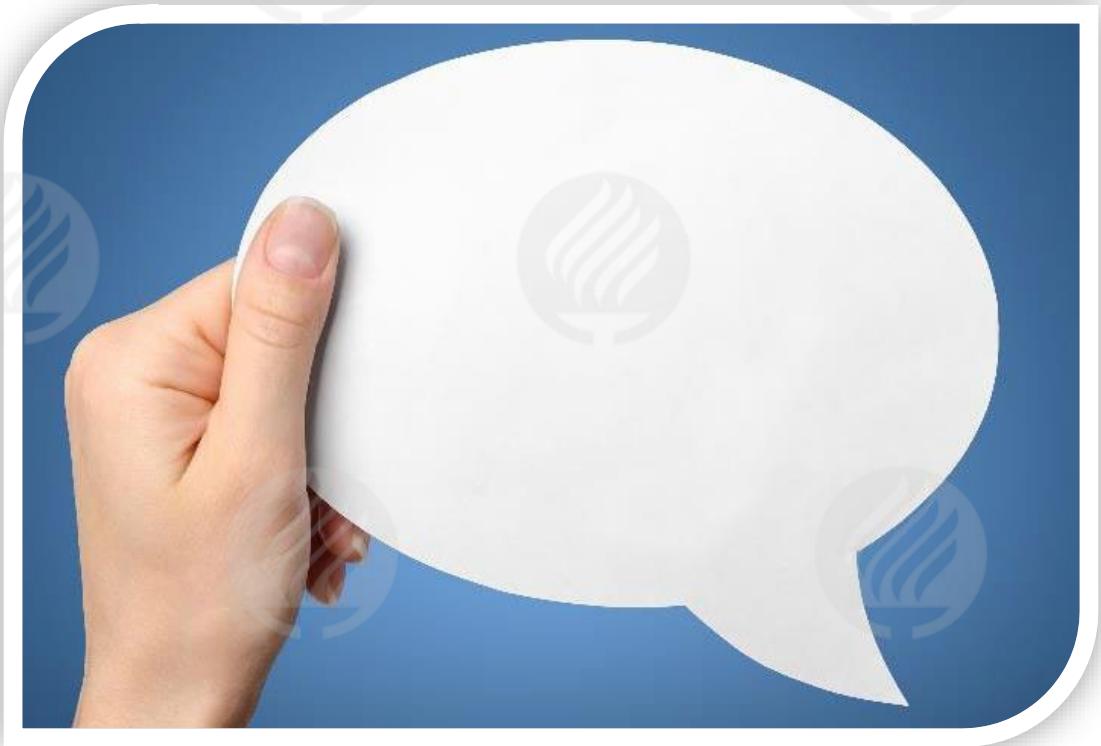
- **Pandas** es una plataforma reconocida por su **eficiencia y simpleza para recuperar, almacenar y manipular información** de diferentes formatos de archivos.
- **Pandas** facilita la **manipulación de la información utilizando DataFrames** y muchas operaciones eficientes ya implementadas

3.1 Reflexión General: Tips y Reflexiones de la práctica



- ***join* y *merge*** son operaciones muy parecidas, ambas **relacionan información de diferentes dataframes**
- ***join*** es más usado cuando **relacionamos dataframes por sus índices** o una columna llave
- ***merge*** es una función más versátil, y más general. Nos permite **especificar otras columnas para relacionar la información**

3.2 Reflexión General: Tips y Reflexiones de la práctica



- **append** y **concat** son operaciones muy parecidas, ambas **agregan información de diferentes dataframes**
- **append** es utilizado para agregar "**filas**" al **dataframe**
- **concat** agrega información de diferentes **dataframes** ofreciéndonos la flexibilidad de agregar ya sea **filas o columnas**

Panorámica de la sesión

1 | Integrar Aprender + Profundizar

2 | Práctica (ciclos)

3 | Reflexión

4 | Plan de aplicación

Cierre de sesión



Plan de aplicación

4

4 | Plan de aplicación

Análisis de deserción de empleados

- 1 Sesión Síncrona 1
Aprender
- 2 Trabajo asincrónico 1
Profundizar | Ruta de Aprendizaje
- 3 Sesión Síncrona 2
Preparar para Aplicar
- 4 Trabajo asincrónico 2
Aplicar en el trabajo | Reto

Plan de aplicación: Reto



- Para el reto seguirás trabajando con uno de los archivos provistos en el **Hackathon HackerEarth 2020** que contiene **información diversa de los empleados de una compañía**, junto con otro archivo creado **específicamente para este reto**
- De manera similar a las actividades realizadas en esta sesión, tendrás que **relacionar información** de diferentes archivos, **agregar otro tipo de información**, y **hacer manipulaciones a esos resultados**

Plan de aplicación: Reto | Rúbrica



Rúbrica de evaluación para el reto

Criterio	Puntaje
Relacionar correctamente información de 2 archivos	10
Agregar información proveniente de otro archivo	10
Ejecutar adecuadamente consultas solicitadas	10
Guardar resultados en hojas de un archivo XLS	5
Total	35 pts

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 | Take Away



Cierre de sesión

Cierre | Sesión Sincrónica 2



¿Dudas?

Cierre | Sesión Sincrónica 2



En esta sesión practicaste:

- **Relacionar información** en diferentes archivos
- **Agregar información** a un *DataFrame* existente
- **Realizar consultas** a la información resultante
- El siguiente paso es **poner en práctica lo revisado en este módulo en tu lugar de trabajo**, tomando una situación en donde puedas aplicar tus **nuevas habilidades**

Cierre: Concepto clave

Procesar **grandes volúmenes de información** provenientes de varias fuentes, es una tarea cada vez más común en nuestros días. Es por eso que contar con **herramientas poderosas, eficientes y confiables**, nos ayuda a mejorar el **manejo de la información**



Tecnológico de Monterrey | 2021

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

Gracias | Programas LIVE