



Tecnológico
de Monterrey

| Educación
Continua

Plataformas de visualización

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



- Manipularás **información** contenida en diversos archivos para combinarla y **obtener representaciones relevantes de los datos**
- Usarás las **gráficas adecuadas** de acuerdo al tipo y contenido de las **variables a visualizar**
- Identificarás los **parámetros** a modificar en las funciones de ploteo para **caracterizar las gráficas resultantes**

Panorámica de la sesión

1 | Integrar Aprender + Profundizar

2 | Práctica (ciclos)

3 | Reflexión

4 | Plan de aplicación

Cierre de sesión

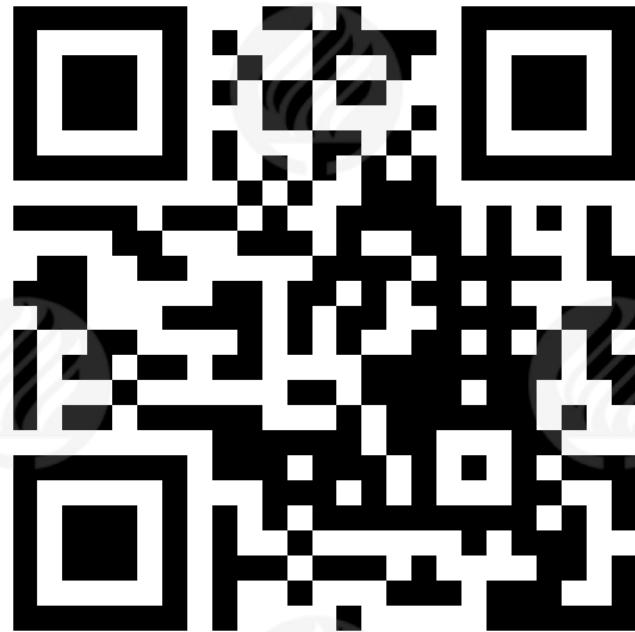
Bienvenida | Dinámica



Please enter the code



- <https://www.menti.com/f6b3zxzzr>



Introducción



- El **desconocimiento** de la estructura y contenido de la información puede impactar **negativamente** en la **precisión de los proyectos de ciencia de datos**
- La etapa de **exploración** evita esos problemas y encuentra tendencias y correlaciones que constituyen un **punto de partida para fases subsecuentes del análisis**
- Al término de esta sesión serás capaz de crear **representaciones relevantes** de los datos de acuerdo a su contenido y composición

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

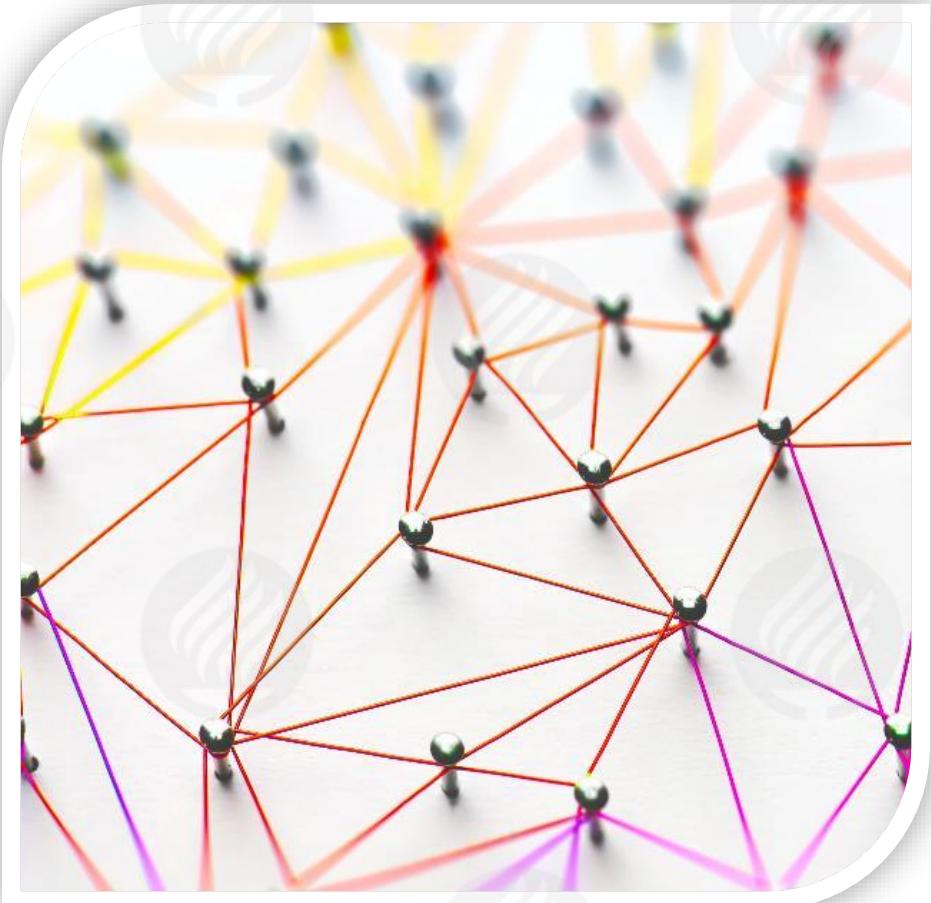
1. Integrar | Aprender + Profundizar



En esta sesión asociarás los temas de las sesiones previas para reflexionar:

- ¿Cómo **integro los conocimientos** adquiridos en mi quehacer diario?
- ¿Conozco con certeza la **información que manipulo** en mi trabajo?
- ¿Serán importantes los **hallazgos** derivados de la exploración para la toma de decisiones en mi empresa?

1. Integrar | Conexión



- Ingresa a app.gosoapbox.com con el **event code** que te proporcione en el chat
- Duración: **5 minutos**

1. Integrar | Aprendizajes



1. Explorar **rangos típicos, categorías, valores faltantes** en la información, **distribuciones y estadísticas** de los datos
2. Trazar y personalizar la estética con tres **plataformas de visualización**
3. Graficar con **ejes compartidos y subplots**
4. Hacer **anotaciones de texto** en las gráficas e incluir **tablas de datos**
5. Emplear los tipos más conocidos de **gráficas para exploración de datos**

Panorámica de la sesión

1 | Integrar Aprender + Profundizar

2 | Práctica (ciclos)

3 | Reflexión

4 | Plan de aplicación

Cierre de sesión

2 | Práctica 1

Explorando el índice de felicidad

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



Práctica

2.1 Práctica | Contexto



- El **informe mundial sobre la felicidad** es una encuesta que clasifica a 156 países por sus niveles de bienestar, tomando en cuenta **6 factores**:
 1. Producción económica
 2. Apoyo social
 3. Esperanza de vida
 4. Libertad
 5. Ausencia de corrupción y
 6. Generosidad
- Se tomará el **estudio del 2019**, recuperado del portal **kaggle.com**, para explorar los datos y puntuación por país o región y obtener gráficas relevantes

2.2 Práctica | Instrucciones



- Descarga el archivo ***happiness_report.csv***
- Genera una libreta en **Google Colab** para esta práctica, cuya estructura esté basada en los análisis solicitados
- Lee, en un **dataframe**, todos los registros del archivo

2.3 Práctica | Trabajo en equipo



Del **dataframe** anterior:

1. Usa los atributos y funciones estudiadas para familiarizarte con la **estructura y contenido** del dataframe (`shape`, `columns`, `head()`, `tail()`, `isna()`)
2. A través de la función **describe()** responde:
 - ¿Cuál es el **máximo valor** de felicidad? ¿El mínimo?
 - ¿El factor con **mayor desviación** estándar? ¿La menor? ¿Qué significa?
 - ¿Cuál es el valor del **3er cuartil** en la percepción de corrupción? ¿Cómo se interpreta?
3. Filtra el dataframe para quedarte con **5 países** (el más feliz, el menos feliz, México y dos más de tu interés) y **visualiza en una misma gráfica los 6 factores**

2.3 Práctica | Trabajo en equipo



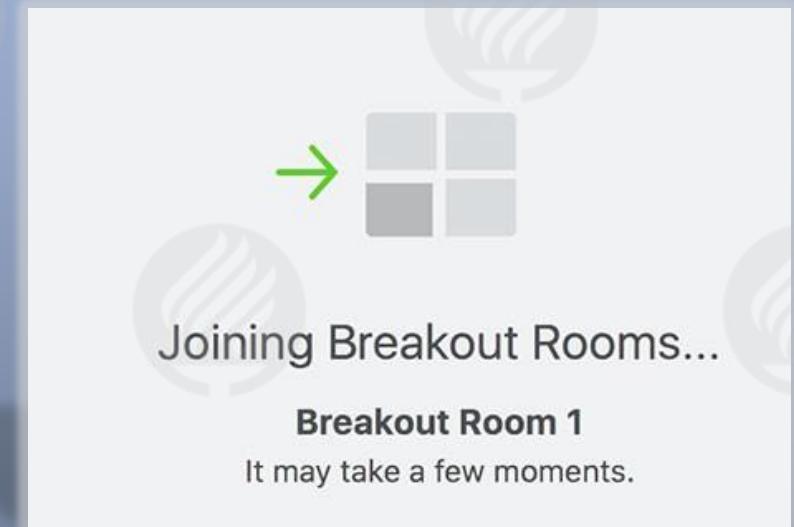
4. Del **dataframe original** obtén:

- **Histograma** del puntaje (score) ¿Cuál es el rango de puntaje más frecuente?
- **Box plot** de la esperanza de vida (Healthy life expectancy) ¿Cómo se interpreta?
- **Subplots** de 2 x 3 con scatter plots del puntaje (score) versus los 6 factores ¿Qué factor influye más en la evaluación? ¿Cuál influye menos?
- Compruebe lo anterior con un **heatmap**

5. Combina con el dataframe **countries** (world_table_country.csv) para graficar la felicidad promedio por continente

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



- ¿Cuál consideras fue el grado de **dificultad** de la práctica?
- ¿Qué **plataforma de visualización** ocupaste con más frecuencia?
- ¿Te fue posible **personalizar** alguno de los gráficos generados?
- ¿Qué **otros reportes** gráficos podrías generar con los archivos proporcionados?

2 | Práctica 2

Análisis del progreso de la 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



- Con la aprobación de las primeras vacunas contra el **COVID-19**, ha crecido el interés por el progreso de la **vacunación mundial**
- Diariamente se recaba **información** sobre el número de **personas vacunadas**
- Para esta práctica utilizarás una muestra reciente de esta información, recuperada del portal **kaggle.com**, para explorar los datos por país y obtener **representaciones visuales significativas**

2.2 Práctica | Instrucciones



- Descarga el archivo ***country_vaccinations.csv***
- Genera una libreta en **Google Colab** para esta práctica, cuya estructura esté basada en los análisis solicitados
- Lee, en un **dataframe**, todos los registros del archivo

2.3 Práctica | Trabajo en equipo



Del **dataframe** anterior:

1. Usa los atributos y funciones estudiadas para familiarizarte con la **estructura y contenido** del dataframe (`shape`, `columns`, `head()`, `nunique()`, `isna()`)
2. A través de la función **describe()** responde:
 - ¿Cuál es el **máximo valor** de vacunaciones diarias por millón (`daily_vaccinations_per_million`): relación entre el **número de vacunaciones y la población total** para una fecha determinada?
 - El **promedio** de personas vacunadas por día (`daily_vaccinations`)

2.3 Práctica | Trabajo en equipo



3. Obtén un nuevo dataframe que contenga el máximo valor de **people_vaccinated_per_hundred** para cada país y ordénalo
4. Elimina aquellos registros que contenga **Nan** en el resultado anterior. Grafica el dataframe resultante

2.3 Práctica | Trabajo en equipo



5. Señala a **México** con una cruz roja en la gráfica. Para esto tendrás que saber dos cosas:
 - la **posición** que ocupa México en la lista - coordenada en x con `np.where(df.index == 'Mexico')`
 - su valor de **people_vaccinated_per_hundred** - coordenada en y con `df.loc['Mexico']`
6. Centrémonos en lo que ha ocurrido en **México**. Para ello obtén un nuevo dataframe con el resultado del filtrar a nuestro país del dataframe original. Haz que la fecha sea el índice

2.3 Práctica | Trabajo en equipo

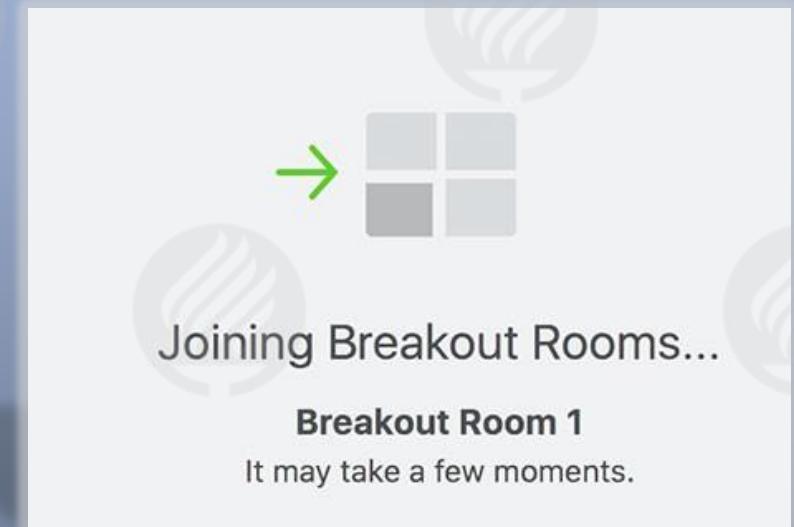


7. Del dataframe anterior grafica las columnas **people_vaccinated** y **people_fully_vaccinated** para **comparar** el acumulado de personas que aún no tienen el conjunto completo de inmunizaciones (normalmente 2) y las que sí lo tienen

8. Grafica la columna **daily_vaccinations** para ver el **volumen** diario de vacunación

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



- ¿Te fue posible **concluir** todos los ejercicios de la práctica?
- ¿Por qué se debía utilizar el **máximo** y no la suma para la agrupación por país?
- ¿Qué **otra información** importante podrías visualizar gráficamente con los datos proporcionados?
- ¿Qué elementos de la práctica podrías incorporar en los **reportes** que generas en tu organización?

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



- A través de los ejercicios prácticos pudiste constatar que ya tienes los conocimientos para **explorar y visualizar grandes volúmenes de información**
- Las representaciones visuales que obtuviste te dan ciertas pistas para continuar con la identificación de **características relevantes** en un proyecto de ciencia de datos
- ¿Cómo **integrar** estos resultados en un **dashboard** o cuadro de mando y **alinearlos con los KPI's o métricas de tu organización**?

3.1 Reflexión General: Tips y Reflexiones de la práctica



- Familiarízate siempre con tus datos, conoce su **estructura y contenido**
- Elige la **plataforma de visualización** que se ajuste a tus necesidades (Pandas es útil para gráficos simples y Seaborn admite perspectivas más complejas, pero en ambos casos debes considerar adentrarte en la complejidad de Matplotlib para personalizar)
- Utiliza diversos **tipos de gráficos** para explorar tus datos
- Intenta identificar **patrones** y experimentar con diversos **escenarios**

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 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 continuarás trabajando con el archivo de empleados provisto en el Hackathon **HackerEarth 2020**
- Aunque ya conoces su estructura, aplicarás las **técnicas de exploración** efectuadas en esta sesión
- Graficarás los datos utilizando la **plataforma de visualización** de tu preferencia, previa **agrupación y filtrado de la información en los casos necesarios**
- Con lo anterior podrás tener una **visión global** de los datos y sacar **conclusiones significativas**

Plan de aplicación: Reto | Rúbrica



| Criterio | Puntaje |
|------------------------------------------------|------------------|
| Análisis de estructura y contenido | 10 |
| Visualización y personalización de la estética | 10 |
| Gráficas para exploración de datos | 10 |
| Interpretación de resultados | 5 |
| Total | 35 puntos |

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 completaste dos prácticas que reforzaron tus conocimientos para:
 - Explorar la información de un único o varios archivos de datos
 - Obtener representaciones visuales relevantes
 - Construir gráficas superpuestas y subgráficas
 - Realizar anotaciones y personalización de la estética

El siguiente paso es replicar estos análisis en la información que manipulas como parte de tus responsabilidades laborales

Cierre: Concepto clave

Prácticamente todas las disciplinas utilizan **visualización de datos** en la actualidad; pues con ello se aporta significado a la información.

La **visualización** te permitirá generar nuevas oportunidades, solucionar problemas substanciales y comunicar resultados de manera efectiva.



Tecnológico de Monterrey | 2021

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

Gracias | Programas LIVE