



## Reto | Análisis de deserción de empleados

**Duración:** 2h

### Objetivo:

Con el desarrollo del reto integrarán todos los conocimientos adquiridos en éste y los módulos previos del diplomado, cumpliendo los siguientes objetivos:

- Realizar un análisis de los datos para conocer su estructura y contenido, y así relacionarla con los requisitos del reto
- Desarrollar funciones que nos permitan obtener datos de una manera eficiente
- Crear una aplicación web con diversas secciones que nos permitan visualizar datos desde diversos puntos de vista
- Visualizar información relevante en gráficas para realizar procesos de análisis

### Introducción:

En los retos de los módulos previos, trabajaste con los datos provistos en el **Hackathon HackerEarth 2020**, tomando como hipótesis que esta información resultará explicativa del fenómeno de deserción laboral que tanto afecta en la actualidad a las empresas y organizaciones.

Ya has completado algunas observaciones elementales y sintetizado dicha información mediante técnicas de filtrado y agrupación. Tu trabajo en esta fase consiste en complementar el análisis de datos con la representación en un dashboard usando **Streamlit** de manera eficiente y atractiva para un usuario final.

### Instrucciones:

Los siguientes pasos te guiarán en el proceso de desarrollo del reto para que logres completarlo con éxito:

1. Genera una libreta en **Google Colab** para el reto.
2. Instala **streamlit**, **ngrok** y crea el **túnel** de comunicación para visualizar la aplicación.



3. Descarga el archivo: **Employees.csv** y guarda, en un dataframe (**employees**), para evitar tráfico innecesario construir una función principal donde recuperen n datos (500 por default) para la etapa de pruebas y desarrollo, NOTA: usar el atributo **cache**.
4. Observa la estructura y contenido del dataframe con los atributos
5. Crear el programa **employees.py** que contendrá la aplicación (usar **%%writefile** )
6. Crear título de la aplicación, encabezados y texto de descripción del proyecto.
7. Crear un **sidebar** en la aplicación
8. En sidebar crear un control **checkbox** que permita mostrar u ocultar el dataframe completo
9. Crear un buscador de empleados con cajas de texto y botones de comando, que permitan buscar por Employee\_ID, Hometown o Unit, mostrar dataframe con resultados encontrados y total de empleados. Nota: Usar funciones con cache.
10. En el sidebar incluir un control **selectbox** que permita filtrar los empleados por su nivel educativo, mostrar el dataframe filtrado y total de empleados. Nota: Usar funciones con cache.
11. En el sidebar crear un control **selectbox** con las ciudades que participaron en el estudio, mostrar los empleados por ciudad en un dataframe filtrado y total de empleados. Nota: Usar funciones con cache.
12. Crear un **selectbox** para filtrar por la unidad funcional (Unit) a la que pertenece. Nota: Usar funciones con cache.
13. Crear un histograma de los empleados agrupados por edad.
14. Crear una gráfica de frecuencias para las unidades funcionales (Unit) para conocer cuántos empleados hay en cada Unidad
15. **Analizar** los datos con una gráfica que nos permita visualizar las ciudades(**Hometown**) que tienen el mayor índice de deserción
16. **Analizar** la información con una gráfica que permita visualizar la edad y la tasa de deserción
17. **Analizar** con una gráfica que determine la relación entre el tiempo de servicio y la tasa de deserción
18. Al finalizar comparte tu libreta (**Google Colab**) con el instructor, descarga tu script y súbelo a Canvas (sigue las recomendaciones de la sección **Formato de entrega de actividad**)