

Reto: Análisis de deserción de empleados con sus sistemas operativos utilizados

Duración: 2h

Objetivo:

El objetivo de este reto es poner en práctica los contenidos revisados durante el módulo realizando los siguientes pasos:

Recuperar información de diferentes archivos utilizando la plataforma pandas.

Analizar, interpretar y conjuntar de manera correcta la información contenida en diferentes archivos.

Con la información ya conjuntada, realizar algunas consultas para recuperar información relevante.

Guardar los resultados de las consultas en un archivo con diferentes hojas de trabajo.

Introducción:

Actualmente en las empresas, la gente ya no dedica toda su vida a trabajar con una sola organización. Cada vez es más frecuente que veas casos en donde la gente cambia de trabajo. Esta ocasión seguirás trabajando con los datos provistos en el Hackathon HackerEarth 2020, junto con más información proporcionada por el departamento de TI de esos empleados.

Trabajarás con cuatro archivos diferentes: dos de ellos tienen exactamente la misma estructura que en el módulo pasado, con la información de los empleados como su id, edad, nivel educativo, estado civil, entre otros datos. Pero los otros dos son los que contienen información adicional de estos empleados con el sistema operativo que usan en sus computadoras, la versión, el número de tickets levantados, el sistema operativo que usan en sus dispositivos móviles, la versión y también número de tickets de ayuda que levantaron. Estos tickets lo que representan es el número de solicitudes de atención o soporte por cada uno de los empleados generados durante un mes específico, donde te interesa saber cosas, como por ejemplo, el número total de tickets generados por la gente con diferentes sistemas operativos de escritorio.

Instrucciones:

Para poder concluir exitosamente este reto, necesitarás llevar a cabo los siguientes pasos:

1. Descarga los cuatro archivos adjuntados en la plataforma: **Test.csv**, **Test2.csv**, **Train.csv** y **Train2.csv**.
2. Abre los archivos descargados para conocer su estructura y familiarizarse con la información contenida en ellos.
3. Genera una libreta en Google Colab para resolver el reto con scripts de Python.
4. Crea cuatro DataFrames en pandas, cada uno para cargar el contenido de los diferentes archivos.
5. Conjunta de manera correcta la información, de tal forma que queden todos los registros juntos en un solo DataFrame que contenga todas las columnas necesarias.
6. Para todos los registros que no tienen un valor para un sistema operativo, ya sea de su computadora o de su dispositivo móvil, asignar el valor "no registrado".
7. Ejecuta la instrucción adecuada en pandas para encontrar:
 - a. El porcentaje de personas que usan Windows / Mac / Linux / no registrado como su sistema operativo en su computadora.
 - b. El porcentaje de personas que usan iOS / Android / No registrado como su sistema operativo en su dispositivo móvil.
 - c. La edad promedio de la gente que utiliza cada sistema operativo de escritorio.
 - d. El nivel educativo promedio de la gente que utiliza cada sistema operativo de escritorio.
 - e. El número total de tickets generados por la gente con cada sistema operativo de escritorio.
8. Envía el resultado de cada solicitud de información a un archivo XLS[X], donde el resultado de cada consulta debe ir en su propia hoja de trabajo.

Entregable:

1. Guarda todos tus cambios en la libreta de Colab donde trabajaste con la solución al reto. Asegúrate de dejar todas las ejecuciones de cada celda visibles.
2. Descarga la libreta en formato Google Colab (extensión .ipynb) y lo renombrarás de la siguiente manera:
3. SC23_Reto_APELLIDOS_NOMBREdelparticipante.ipynb
4. Entrega también un archivo XLSX que debes nombrar:
5. SC23_Reto_APELLIDOS_NOMBREdelparticipante.xlsx
6. Entrega los dos archivos (.ipynb y .xlsx) y la URL de tu archivo de Google Colab con permisos de visualización, presionando la opción "Añadir publicación" (+) en el apartado Reto de aplicación en el trabajo.