

Evaluación Almacenamiento y Captura de datos

Claudio Aracena

Entrega: 15 de enero del 2023

En grupos de máximo tres personas deberán resolver los dos problemas descritos abajo. Deben entregar el notebook con el código para resolver cada problema. Agregue comentarios al código para entender el razonamiento utilizado.

1. Realice un proceso de web scraping para el sitio <https://jetsmart.com/cl/es/>. Para ello realice las siguientes tareas:
 - a. Utilizar Selenium para emular un navegador y cargar la página requerida (1 punto).
 - b. Buscar vuelos desde Santiago a Arica del 1 de febrero al 28 de febrero utilizando los campos de inputs del buscador de la página (1 punto).
 - c. Una vez en la página de resultados, extraiga la información de precios y fechas de vuelos de ida y de vuelta para todos los días de febrero y genere un dataframe con la información de precios y fechas (2 puntos).
 - d. Crear una base de datos sqlite con la información del dataframe (1 punto).
 - e. Realizar una consulta SQL para seleccionar los vuelos de ida y vuelta con el precio total mínimo (1 punto).
2. Utilizando la API de Twitter, realice las siguientes tareas:
 - a. Extraer los últimos 1.000 tweets de tu usuario preferido que escriba en español (1 punto)
 - b. Hacer un análisis de frecuencia de palabras (100 palabras más frecuentes), habiendo removido primero stopwords y usando tokenización en español (2 puntos).
 - c. Extraer 100 usuarios con perfil público que son seguidos por su usuario de preferencia y extraer los últimos 10 tweets de cada uno de esos usuarios (1 punto).
 - d. Hacer un análisis de frecuencia de palabras igual al de la parte b) (100 palabras más frecuentes), de los tweets extraídos en la parte c) (2 puntos).

Cada pregunta tiene una escala de 1 a 7. El promedio de ambas preguntas resulta en la nota final de la tarea.