

Evaluación Almacenamiento y Captura de datos

Claudio Aracena

Entrega: 1 de marzo del 2020

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://www.portalinmobiliario.com/>. Para ello realice las siguientes tareas:
 - a. Utilizar Selenium para emular un navegador y cargar la página requerida (1 punto).
 - b. Buscar arriendo de departamentos en la comuna de Santiago utilizando los campos de inputs del buscador de la página (1 punto).
 - c. Extraer los primeros 100 departamentos con su precio, superficie, cantidad de dormitorios, cantidad de baños y dirección (1 punto).
 - d. Crear un dataframe con la información extraída (1 punto).
 - e. Crear una base de datos con la información del dataframe (1 punto).
 - f. Realizar una consulta SQL para saber el precio promedio de arriendo por metro cuadrado (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 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.