

Ciencia de Datos

Práctico N°1: Nivelación con Pandas y Seaborn

Introducción: Kaggle, es una plataforma online de acceso gratuito, lanzada en 2010, para dar soporte a la comunidad de data scientists, permitiendo explorar y publicar tanto base de datos como modelos y desarrollar competencias (con costo para las grandes empresas) para resolver desafíos de la Ciencia de Datos. Desde 2017 es una subsidiaria de Google LLC. Para poder acceder a los recursos de la plataforma es necesario registrarse previamente como usuario.

A su vez, Kaggle proporciona cursos cortos, gratuitos y autocorregidos para adquirir destrezas básicas en el uso de las herramientas usuales en la industria.

Por otro lado, Google proporciona Colaboratory, o Colab para abreviar, un producto que permite escribir y ejecutar código en Python desde un navegador de internet, usando recursos de sus propios servidores.

A los fines de nivelar nuestros conocimientos, las primeras dos tareas consisten en hacer dos cursos y conseguir los correspondientes certificados y luego resolver un desafío creando una notebook en Colab.

Tarea 1: Realizar el curso sobre **Pandas** de Kaggle.

Tarea 2: Realizar el curso sobre **Data Visualization** de Kaggle.

Tarea 3: Para aplicar lo aprendido se proponen las siguientes actividades:

- a Descargar el dataset Superstore Sales Dataset disponible en Kaggle. Indagar el diccionario de las columnas del dataset. Luego almacenar el archivo `csv` en el drive personal de la cuenta institucional de la UNC.
- b Crear una instancia en Colab y averiguar cómo acceder al sistema de archivos de drive para cargar el `csv` usando Pandas, con la variable `Order Date` como índice de fecha.
- c Usar Seaborn para mostrar un `lineplot` con el número de *total* de ventas en cada uno de los días, correspondientes a la categoría **Furniture**, sólo en la ciudad de **Los Angeles**.
- d Crear un `DataFrame` con una variable que sume el número total de ventas efectuadas en cada uno de los días de la semana –de lunes a domingo– (usando el índice de la base) y otra columna con los correspondientes porcentajes de ventas. ¿Cuáles son los porcentajes de los días de más y de menos ventas?
- e Usar Seaborn para mostrar un `barplot` de la frecuencia relativa de ventas en cada uno de los días de la semana.