

Módulo 1 - Desafío 1

Introducción

En este módulo vimos varios conceptos de estadística descriptiva básica, trabajamos sobre varias nociones de programación en Python y sobre aspectos de la línea de comando (navegar directorios, modificar archivos, etc.). Ahora, la idea es poner todo en práctica. Ese es el objetivo de este primer Desafío del curso.

Una reconocida universidad de Estados Unidos nos ha encargado un informe de diagnóstico sobre la performance promedio de los [score SAT](#) por estado con el objetivo de identificar principales características del rendimiento escolar promedio. La exploración inicial se realizará a partir de las herramientas de estadística descriptiva y visualización aprendidas en el módulo y, sobre dicha base, se busca determinar el área en el que la universidad establecerá una nueva unidad educativa.

Recuerda explorar los datos e indagar sobre posibles mensajes que pueda transmitir el análisis para facilitar la toma de decisiones por parte de nuestro cliente, haciendo hincapié en brindar mensajes cortos, claros, con visualizaciones (aquí, una imagen vale más que mil palabras!).

Objetivo (material a entregar): generar un ***Notebook Jupyter*** que concentre el análisis de los datos, los mensajes a comunicar y las visualizaciones. Esto constituirá el 'entregable' de tu primer desafío.

Entrega y presentación: Clase 7 del Módulo 1, fecha tentativa: **viernes 11/08 o sábado 12/08**, según días de cursada.

Dataset

El dataset (provisto por la universidad) contiene las medias del SAT score en matemática y expresión verbal para cada estado y para el distrito de Columbia para el año 2015.

Requisitos

El trabajo debe:

- describir los datos de forma sintética, aprovechando al máximo la información disponible.
- realizar un análisis exploratorio de los datos, incluyendo:
 - usar ***matplotlib*** como librería para generar gráficos y visualizaciones básicas,

- usar ***numpy*** para calcular medidas resumen: media, mediana, moda, desvío estándar, etc., y
- analizar si la distribución de los puntajes SAT se asemeja a una distribución normal.

El trabajo también puede contribuir a responder dos inquietudes que surgieron al interior de la universidad. Alguno directivos discutieron sobre la posibilidad de que las condiciones de ingresos afecten positiva (o negativamente) la performance del SAT score. Otros mencionaron que aquellos estados con mayor proporción de migrantes podrían presentar, en promedio, un menor rendimiento en la comprensión oral y expresión escrita debido a que cuentan con un idioma materno distinto al inglés. En línea con ello, el trabajo puede:

- Explorar si se encuentra algún vínculo entre:
 - el rendimiento del SAT score general y datos de ingresos promedios del estado. Una primera aproximación puede surgir a partir de comparar los resultados del SAT score con variables como el ingreso promedio o el [producto bruto](#) por estado, la [distribución del ingreso](#) (índice de Gini) y la incidencia de la pobreza, o...
 - el rendimiento en writing y/o en matemática en relación con la [cantidad](#) o la [densidad](#) de la población y/o la incidencia de [migración](#) en cada estado.
 - Estas alternativas presentadas pueden ser sustituidas por otras que consideren de interés.
- Replicar los gráficos en ***seaborn*** o introducir mapas con los datos por estado con ***plotly*** (estos pueden ser útiles para identificar fácilmente diferencias entre estados). [Aquí](#) hay un ejemplo para mapas en plotly.
- Crear un post en un blog de al menos 500 palabras y 2 gráficos realizando una descripción de los datos, tu análisis y tu enfoque. Incluí un link en el notebook que realizaste.

¿Cómo empezar?

Podés ver las consignas (guía paso a paso) en el Notebook subido al Moodle. A continuación también hay algunas sugerencias:

- cargá los datos
- corré algunos comandos básicos de numpy para empezar a describir los datos
- escribí un pseudocódigo antes de empezar a codear. Suele ser muy útil para darle un esquema lógico al análisis
- leé la documentación de cualquier tecnología o herramienta de análisis que uses. A veces no hay tutoriales para todo y los documentos y las ayudas son fundamentales para entender el funcionamiento de las herramientas utilizadas
- documentá todos los pasos, transformaciones, comandos y análisis que realices.

Recursos útiles

[Encontrando los datos que necesitas](#)

[Cómo dar una buena y breve charla](#)

Evaluación

Los profesores usarán la siguiente escala para calificar tu trabajo y las habilidades técnicas adquiridas en módulo:

Puntaje	Descripción
0	Incompleto
1	No cumple con las expectativas
2	Cumple con las expectativas
3	Excede con creces las expectativas