

# 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

Vamos a usar un dataset que contiene los [scores SAT](#) alrededor de los Estados Unidos. Vamos a estar explorando estos datos para ver qué podemos analizar usando las herramientas aprendidas en estadística descriptiva. El cliente que encargó este trabajo (una importante universidad norteamericana) espera ver unas buenas visualizaciones y gráficos, además de los análisis estadísticos.

**Objetivo:** generar un Notebook Jupyter que resuma los análisis realizados y los gráficos contruidos.

## Requisitos

El trabajo debe:

- describir de la manera más sintética (pero con el mayor aprovechamiento de la información disponible) posible los datos
- 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.
- analizar si la distribución de los puntajes SAT se asemeja a una distribución normal

## Puntos extra

- replicar los gráficos en seaborn
- 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

## Material a entregar

Un notebook con el código que genera los estadísticos y los gráficos debidamente comentado.

## Fecha de entrega

Clase 8 del Módulo 1 (fecha tentativa: 18/03)

## Código básico

Para este proyecto vamos usar un Notebook Jupyter. Va a usar matplotlib para gráficos y visualizaciones de los datos. En general, los gráficos son muy útiles para resumir la información, para realizar un primer análisis rápido de los datos. Vamos a discutir luego mucho más sobre cómo usar gráficos y visualizaciones para divulgar tu trabajo.

Podés ver las consignas en el notebook subido al Moodle.

## Dataset

El dataset (provisto por la universidad que encarga el trabajo) contiene las medias del SAT score en matemática y expresión verbal para cada estado y para el distrito de Columbia en 2001.

## ¿Cómo empezar? Sugerencias

- cargá los datos
- corré algunos comandos básicos de numpy para empezar a describir los datos
- escribí un pseudocódigo antes de empezar a codera. Suele ser muy útil para darle un esquema y una lógica generales 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 necesitás](#)
- [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