

DigitalHouse >
Coding School

DATA SCIENCE

UNIDAD 1
MÓDULO 1

Presentación del
Desafío 1

Agosto 2017

PRESENTACIÓN DE PROYECTOS



**Un buen proyecto...
es manejable y se relaciona con
un dominio de trabajo conocido**

¿Cuál es el alcance de la idea del proyecto?

¿Qué problema estás intentando resolver?

¿Quién puede beneficiarse de las conclusiones?

¿Quién puede beneficiarse o perjudicarse?

¿Qué datos existen para resolver este problema?

¿De dónde es posible obtenerlos? ¿Qué formato tienen los datasets disponibles? ¿Qué observaciones pueden hacerse?

- 1 **Identificar un problema a resolver con datos**
- 2 **Identificar los objetivos esperados**
- 3 **Identificar el dataset que (potencialmente) constituye el insumo de trabajo**
- 4 **Definir estrategias de abordaje y 'criterios de éxito'**
- 5 **Identificar audiencia (potencial) para presentar los resultados**



[Tips para la presentación](#)

MÓDULO 1 - DESAFÍO





Fundamentos:
Numpy, Stats y
Visualización

01

- Introducción al programa y a la disciplina
- Resumen Python
- Numpy
- Estadística Descriptiva con Numpy
- Introducción a la Visualización de Datos
- Presentaciones de resultados del Desafío 1

Desafío del Módulo

Provistos de un dataset de puntajes SAT de todo EEUU, los participantes realizarán un análisis exploratorio utilizando Numpy y Matplotlib aplicando técnicas básicas de estadística descriptiva.

- **¿Qué?**

Diagnóstico sobre la performance promedio de los score SAT por Estado a partir de las principales características del rendimiento escolar.

- **¿Para quién?**

Universidad XX, Estados Unidos

- **¿Para qué?**

Identificar el Estado en el que la Universidad XX establecerá una nueva unidad educativa

- **¿Cómo?**

Contenidos del Módulo 1 (estadística descriptiva y distribuciones, NumPy y Matplotlib)

- **Complementos**

Información adicional (datos por Estado)

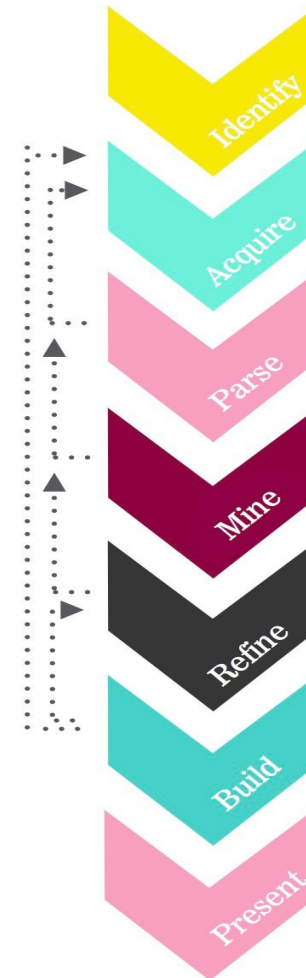
Otras librerías (Seaborn, Plotly)

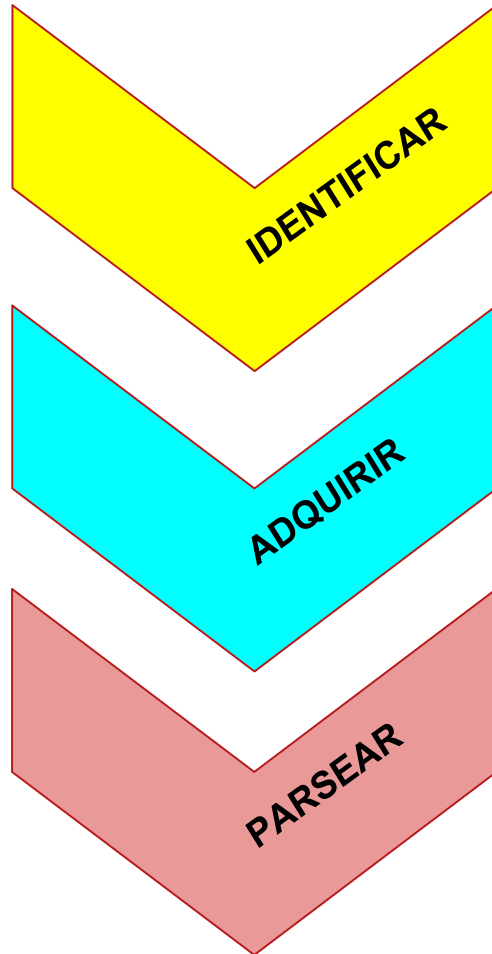


A futuro, te proponemos

- Que todos los participantes intervengan activamente en la discusión y el trabajo, utilizando algunos pasos del Flujo de Trabajo de Data Science.
- Que discutan los resultados y arriben a conclusiones consensuadas.
- Que preparen una notebook con los principales hallazgos y la motivación por incluir algún dato complementario en caso que lo consideren pertinente.
- Que preparen una presentación corta con los principales resultados (10').

- El “Flujo de trabajo de Data Science” constituye, en última instancia, un set de standards sumamente útil y una referencia para tener en cuenta en los **desafíos del curso**.
- Repasemos las diferentes etapas, que están explicadas en detalle en el documento “**Flujo de Trabajo en Data Science.pdf**”





IDENTIFICAR EL PROBLEMA

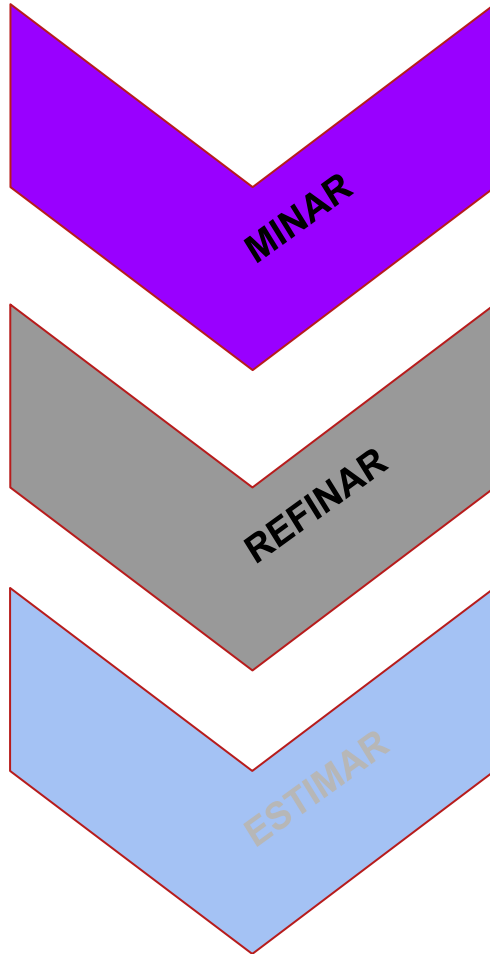
- Identificar los objetivos del producto/negocio/problema
- Identificar y generar hipótesis sobre metas y criterios para el éxito del análisis
- Generar un set de preguntas para identificar el dataset “correcto”.

ADQUIRIR LOS DATOS

- Identificar el dataset “correcto”
- Importar los datos y generar las estructuras de datos adecuadas
- Determinar las herramientas más apropiadas para trabajar con los datos

PARSEAR LOS DATOS

- Explorar toda la documentación relacionada con los datos
- Realizar Análisis Exploratorio de los Datos (AED)
- Verificar la calidad de los datos



MINAR LOS DATOS

- Dar formato, limpiar, homogeneizar y filtrar los datos
- Crear nuevas columnas derivadas de los datos originales (recodificaciones, cálculos, etc.)

REFINAR LOS DATOS

- Identificar tendencias y outliers
- Aplicar y calcular estadísticos descriptivos e inferenciales
- Documentar y transformar los datos

ESTIMAR UN MODELO

- Seleccionar un modelo apropiado (forma funcional, estimación, etc.)
- Estimar el modelo
- Evaluar y refinar el modelo



PRESENTAR LOS RESULTADOS

- Resumir los resultados del análisis con alguna narrativa o historia
- Presentar las limitaciones, los supuestos y las fortalezas del/los modelo/s estimados
- Identificar preguntas derivadas y nuevos problemas para seguir profundizando el análisis