







Equipo de trabajo - Estudiantes



Logística Módulo 1



Introducción al análisis de datos



#### Equipo UNAL





### Profesor:

• Felipe Restrepo Calle, PhD. <a href="https://dis.unal.edu.co/~ferestrepoca/">https://dis.unal.edu.co/~ferestrepoca/</a>



### Asistentes docentes:

- Alejandro Gallego, PhD(c)
- Santiago Toledo, PhD(c)
- Juan Sebastián Lara, MSc.



# Asistentes logística y gestión:

- José Ricardo Castro
- Elkin Aponte Gomez



#### Estudiantes





# Información demográfica



Información académica



Conocimientos previos

Estudiantes



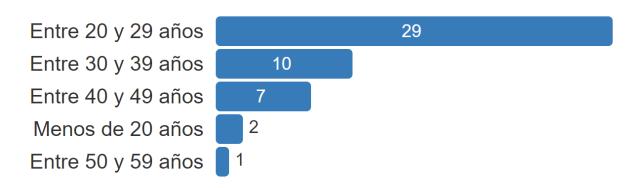
## Información demográfica

Número de estudiantes: 141

Número de grupos: 7

Participación en encuesta: 49 35% hasta las 17:30

## Rango de edad:



### Sexo:





Estudiantes \_\_\_\_\_ Información demográfica

## País:



# Cuidad/Departamento:

Value	Count	Frequency (%)	
Bogotá	32	65.3%	
Cundinamarca	4	8.2%	
Nariño	4	8.2%	
Medellín	3	6.1%	
Santander	2	4.1%	
valle del guamuez	1	2.0%	
Durham, NC	1	2.0%	
Cartagena	1	2.0%	
Ibagué	1	2.0%	





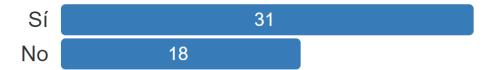
#### Estudiantes



## Información académica

Count Frequency (%)

### Profesional:



### Títulos:

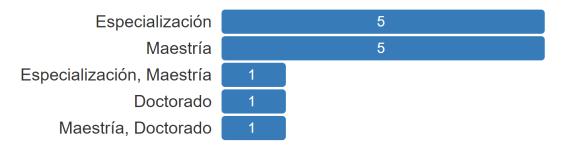
**Value** 

Ingeniería de Sistemas y similares	15	30.6%
Economía	7	14.3%
Ingeniería Industrial	5	10.2%
Ingeniería Electrónica/Eléctrica	3	6.1%
Ingeniería Agrícola/Agronómica	2	4.1%
Física	2	4.1%
Ingeniería Mecánica	2	4.1%
Otras Ingenierías	2	4.1%
Ingeniería Química	2	4.1%
Matemáticas	2	4.1%
Other values (6)	7	14.3%

# Posgrado:



# Nivel posgrado:

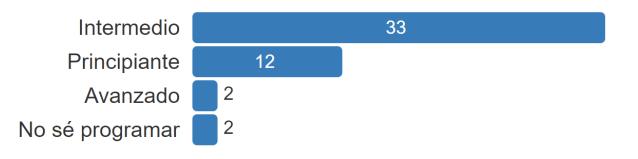


Ingeniería de telecomunicaciones, Negocios internacionales, Estadística, Psicología, Ciencia Política, Licenciatura en ciencias sociales

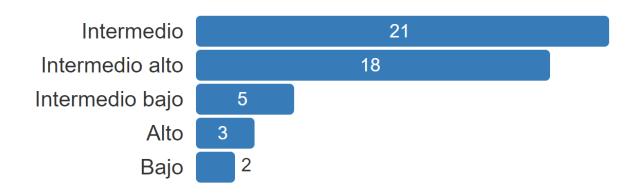




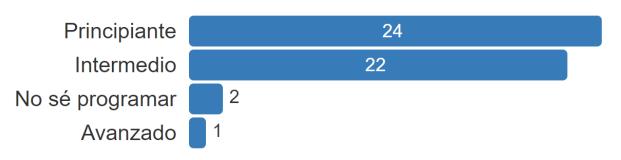
## Nivel programación:



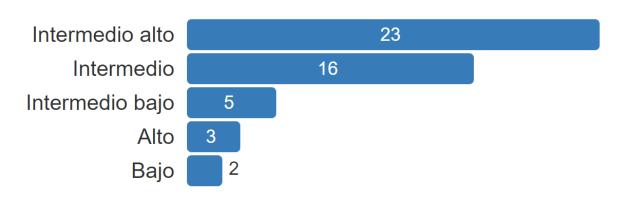
## Nivel Probabilidad y estadística:



# Nivel Python:



# Nivel Álgebra Lineal:











Equipo de trabajo - Estudiantes

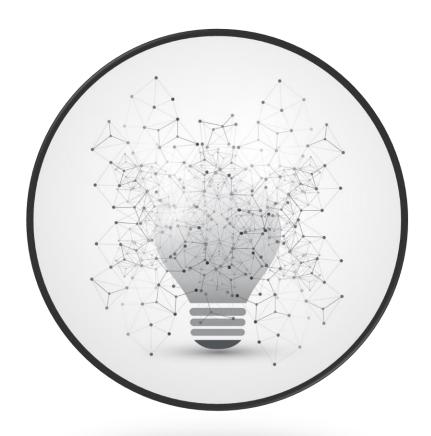


Logística Módulo 1



Introducción al análisis de datos

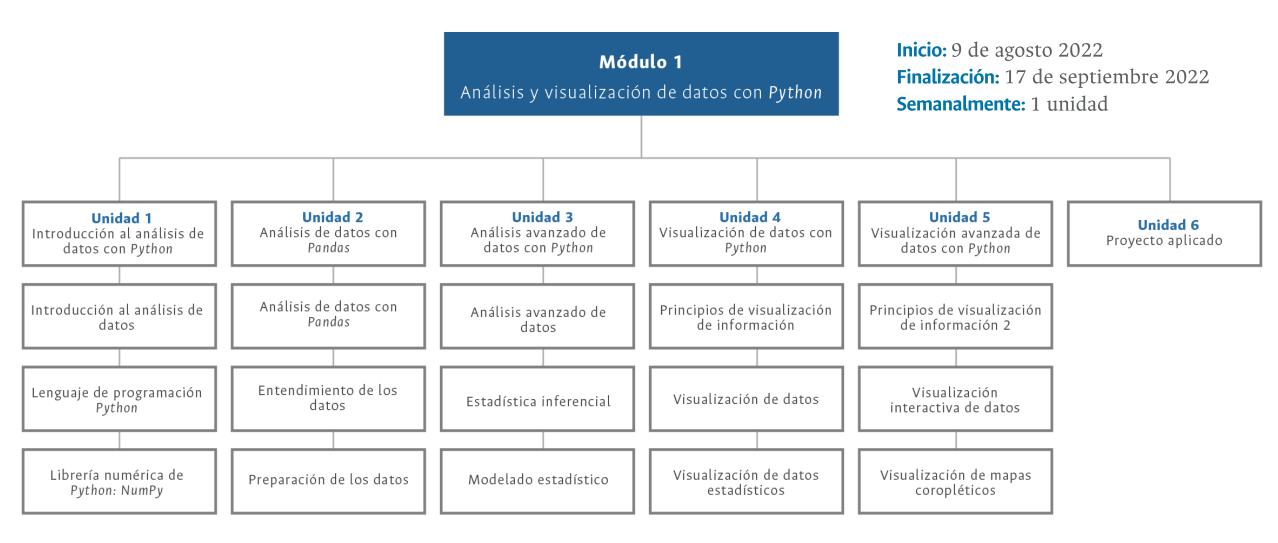




- Programación de computadores. Deseable: Python
- Estadística básica
- Álgebra lineal básica



# Mapa de contenidos del módulo







### Actividades académicas



### Clase sincrónica

- **Martes 18:00 19:45** (1h 45 min)
- **Zoom:** https://zoom.us/j/98240140136
- Fundamentos conceptuales plenaria
- Salas de chat en vivo



# Trabajo autónomo:

- **7 horas/semana** aproximadamente
- **Estudio** de OVAs y recursos educativos
- **Talleres guiados** (Notebooks)
- Quices y tareas
- Participación/Colaboración en foros



### **Tutorías sincrónicas:**

- **Sábados en la mañana** (1h 30 min)
- Asistente docente
- Grupos pequeños
- Énfasis práctico







Aula virtual https://uecp.edunext.io/



Comunicación con compañeros y docente (tiempo de respuesta: hasta 1 día hábil)

Chat en vivo durante clases sincrónicas https://campuswire.com/



**Jupyter Notebooks - Google Colaboratory** 

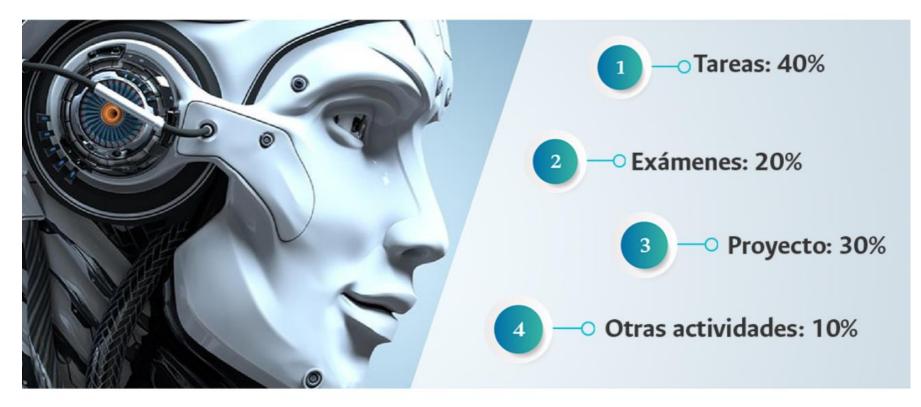
Programación interactiva en Python desde navegador web Requiere cuenta de correo **gmail** 



Entorno de evaluación automática

Quices y tareas





- 1 Tarea por Unidad\*
- 2 Quices por Unidad\*
- Proyecto aplicado
- 5% Colaboración Campuswire 5% Autoevaluación

\*Importante: fechas de entrega firmes 2 semanas máximo 1 semana recomendado







Equipo de trabajo - Estudiantes

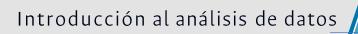


Logística Módulo 1



Introducción al análisis de datos











### Ciencia de datos

- Campo de estudio que busca extraer conocimiento a partir de conjuntos de datos, para lo cual se aplican métodos científicos y computacionales.
- Involucra: análisis de datos, aprendizaje de máquina (Machine learning) y Big Data



#### Análisis de datos

Proceso en el que se obtienen, inspeccionan, limpian y transforman conjuntos de datos, para extraer y comunicar información valiosa que apoye el proceso de toma de decisiones.

### Aprendizaje de máquina

Campo de la inteligencia artificial que estudia algoritmos y técnicas computacionales que permiten que los computadores aprendan a partir de los datos.

#### Big Data

Campo de estudio que busca maneras de almacenar y procesar grandes volúmenes de datos, que generalmente no pueden ser manejados con métodos tradicionales.

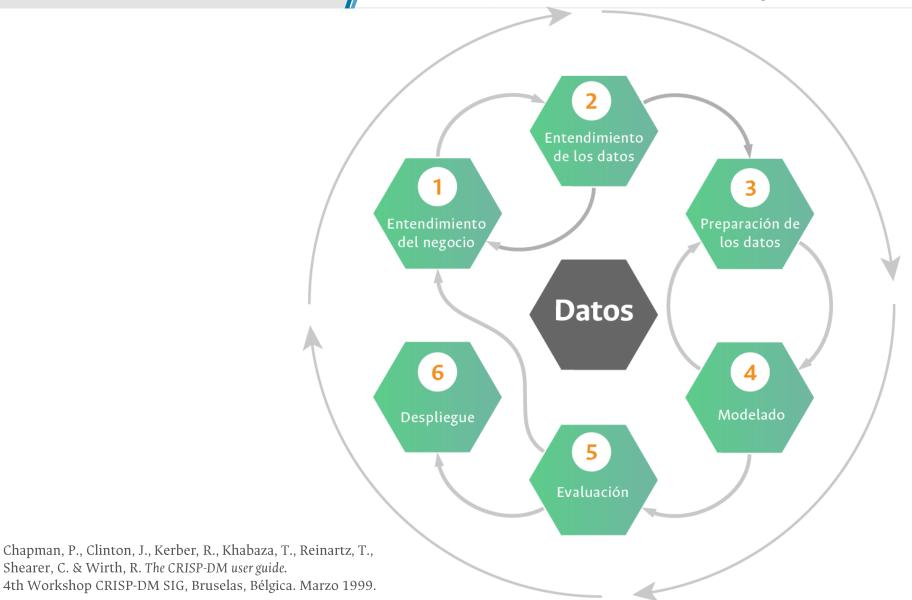
**PLATAFORMAS DE DATOS** 

**♥**mongoDB

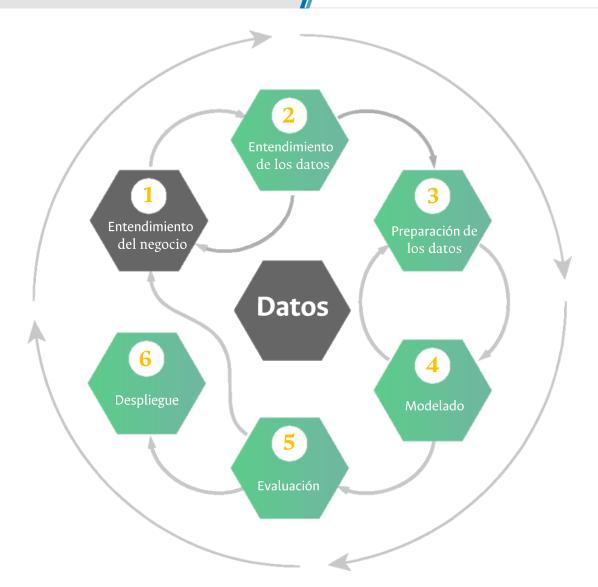




## Metodología CRISP-DM (Cross-Industry Standard Process for Data Mining)



## Metodología CRISP-DM / Entendimiento del negocio

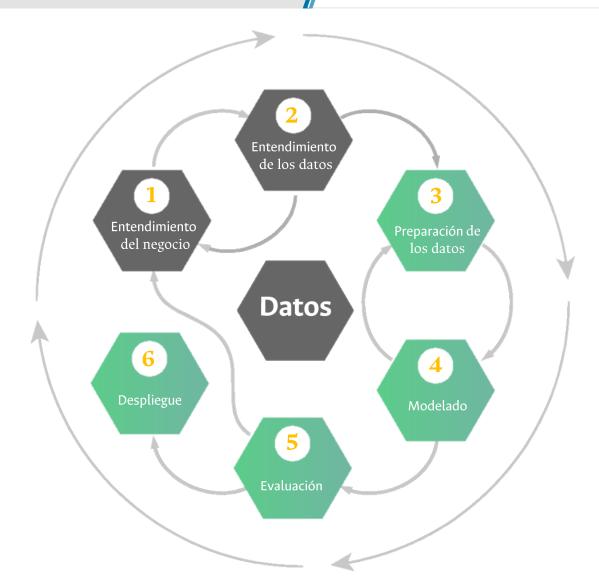


### Entendimiento del negocio

- Conocer el dominio de aplicación.
- Definir y comprender el problema a resolver.
- Evaluar la situación y formular los objetivos del proyecto que se alineen con los objetivos de la organización.



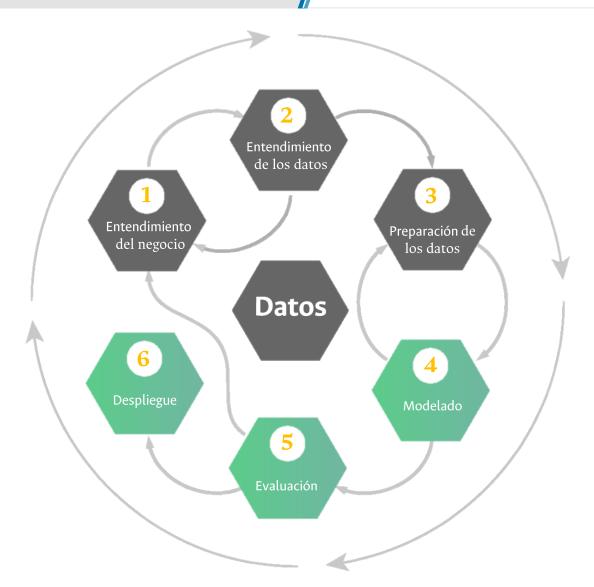
## Metodología CRISP-DM / Entendimiento de los datos



#### Entendimiento de los datos

- Adquirir/consolidar los conjuntos de datos.
- Realizar el análisis de datos exploratorio.
- Usar **estadística descriptiva** para caracterizar, inspeccionar, describir y resumir los datos.
- Usar técnicas de **visualización de la información** para representar de manera gráfica los detalles ocultos en los datos.

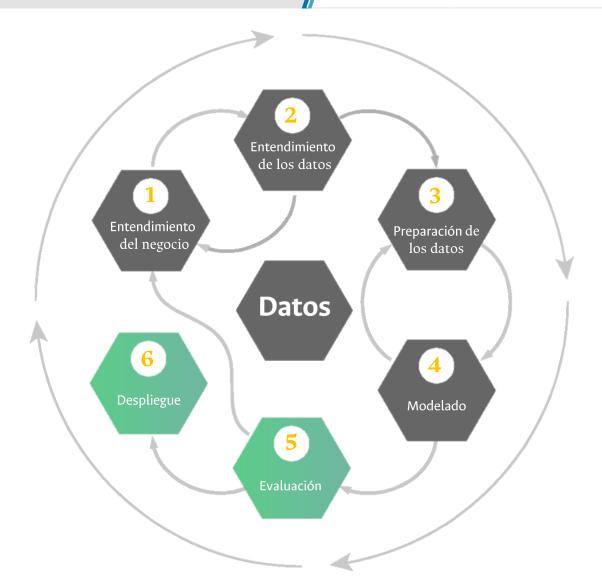
## Metodología CRISP-DM / Preparación de los datos



### Preparación de los datos

- Analizar la calidad de los datos.
- Limpiar los conjuntos de datos.
- Seleccionar características.
- Pre-procesar y transformar los datos para su análisis y modelado posterior.

### Metodología CRISP-DM / Modelado

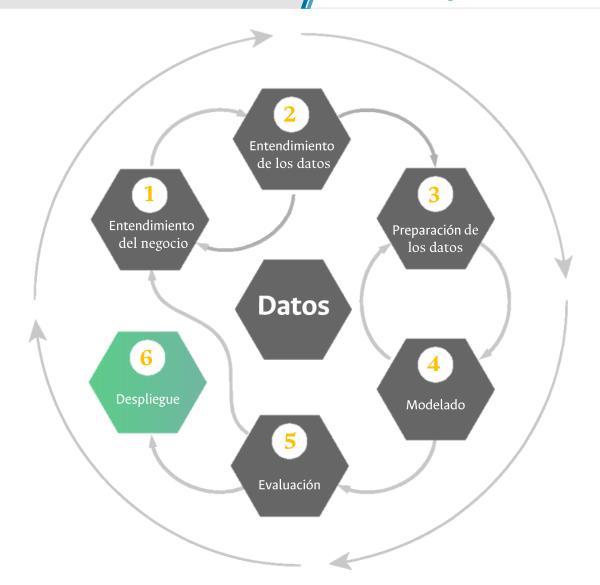


### Modelado

- Elegir las técnicas de modelado apropiadas, según el tipo de problema y desarrollar modelos computacionales.
- Métodos estadísticos.
- Técnicas de aprendizaje de máquina (machine learning).



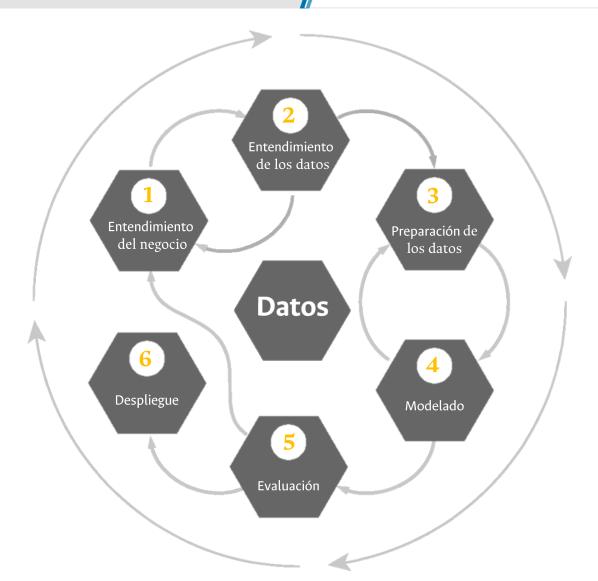
### Metodología CRISP-DM / Evaluación



### **Evaluación**

- Evaluar el desempeño del modelo o los análisis realizados.
- Validar los resultados con respecto a métricas de desempeño y los criterios de éxito definidos al inicio del proyecto.

### Metodología CRISP-DM / Despliegue



### **Despliegue**

- En este punto la organización puede aprovechar los resultados del proceso.
- Soportar la toma de decisiones basado en datos e información.
- Reportar resultados.
- Poner en marcha el modelo computacional en un ambiente de producción.
- Monitorear los modelos analíticos de forma periódica.

# Metodología CRISP-DM





# Referencias

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R. (Marzo de 1999). The CRISP-DM user guide. Cuarto taller CRISP-DM SIG, Bruselas.

Nguyen, M. & Altintas, I. (2020). Machine Learning with Big Data [Aprendizaje automático con Big Data]. UC San Diego. https://www.coursera.org/learn/big-data-machine-learning

Diez, D., Çetinkaya-Rundel, M., Barr, C.D. (2019). OpenIntro Statistics. (4.ª ed.). OpenIntro. https://www.openintro.org/book/os/

Kapil, A. R. (2018) Data exploration and preparation. DataVedas. https://www.datavedas.com/data-exploration-and-preparation/

Recurso computacional en línea de la UCLA. (2014). Probabilidad y estadística Ebook. UCLA. http://wiki.stat.ucla.edu/socr/index.php/EBook

Python. (s.f.). Python. https://www.python.org/









# INGENIERÍA

#### **Autor**

Felipe Restrepo Calle, PhD

#### Asistente docente

Alberto Nicolai Romero Martínez

#### Diseño instruccional

Claudia Patricia Rodríguez Sánchez

#### Diseño gráfico

Clara Valeria Suárez Caballero Milton R. Pachón Pinzón

#### **Diagramadora PPT**

Daniela Duque

2022

