



Introducción al análisis y estructura de datos

Clase sincrónica

Mapa general de la carrera

Revisión modular - Data Analytics

Introducción al
análisis de
datos

Programación
con Python
para el análisis
de datos

Análisis
estadístico con
Python

Visualización y
comunicación
de insights



Mapa general de la carrera

Revisión modular - Data Science

Introducción al
análisis de
datos

Programación
con Python
para el análisis
de datos

Análisis
estadístico con
Python

Visualización y
comunicación
de insights

Machine
Learning

Modelos
avanzados y
Redes
neuronales



SQL para el análisis de datos

El autoaprendizaje de la unidad Kickoff corresponde a un ingreso independiente del módulo en el LMS, si no tienes acceso solicitarlo a tu Coordinador/a Académico/a.

| Unidad | Clases (sincrónico) | Autoaprendizaje (asincrónico) | Tutoría (sincrónico) |
|--|---------------------|-------------------------------|----------------------|
| <i>Kickoff</i> | <i>2 horas</i> | <i>Desde 6 horas</i> | <i>2 horas</i> |
| Introducción al análisis y estructura de datos | 2 horas | Desde 6 horas | 2 horas |
| Consultas en SQL (Parte I) | 2 horas | Desde 6 horas | 2 horas |
| Consultas en SQL (Parte II) | 2 horas | Desde 6 horas | 2 horas |
| Consultas en SQL (Parte III) | 2 horas | Desde 6 horas | 2 horas |
| Consultas en SQL (Parte IV) | 2 horas | Desde 6 horas | 2 horas |
| <i>Prueba</i> | <i>0 horas</i> | <i>Desde 6 horas</i> | <i>0 horas</i> |
| <i>Receso</i> | <i>0 horas</i> | <i>0 horas</i> | <i>0 horas</i> |

¿Qué aprenderemos en este módulo?

Al finalizar este módulo, serás capaz de construir una base de datos en PostgreSQL y hacer consultas para la obtención de información y generación de insights que permitan resolver problemas de negocios.



***Comprender el rol del
analista de datos,
principales problemas que
enfrenta y herramientas que
utiliza.***

- Unidad 1: Introducción al análisis y estructura de datos.
- Unidad 2: Consultas en SQL
 - Parte I
 - Parte II
 - Parte III
 - Parte IV



Te encuentras aquí



¿Qué aprenderás en esta sesión?

- *Comprender el rol del analista de datos, su ecosistema y los problemas que suele enfrentar*
- *Identificar tipos de datos disponibles para resolver un problema.*

¿Qué te sugiere el
nombre “Analista de
datos”?



Analista de datos

De esto se trata

¿Qué queremos
obtener?

¿Qué dificultades
puede tener?

¿Qué herramientas
utilizamos?

¿Qué respuestas
queremos entregar y
cómo?



/* Data Analyst */

El rol Data Analyst

Descripción

Con los datos, el Data Analyst debe:

- Extraerlos, limpiarlos y organizarlos.
- Analizarlos a partir de requerimientos.
- Presentar la información para la toma de decisiones.



El rol Data Analyst

¿Qué hace?

Extraer, interpretar, limpiar
y procesar diferentes tipos
de datos.

Analizar los datos con
herramientas apropiadas
para ello.

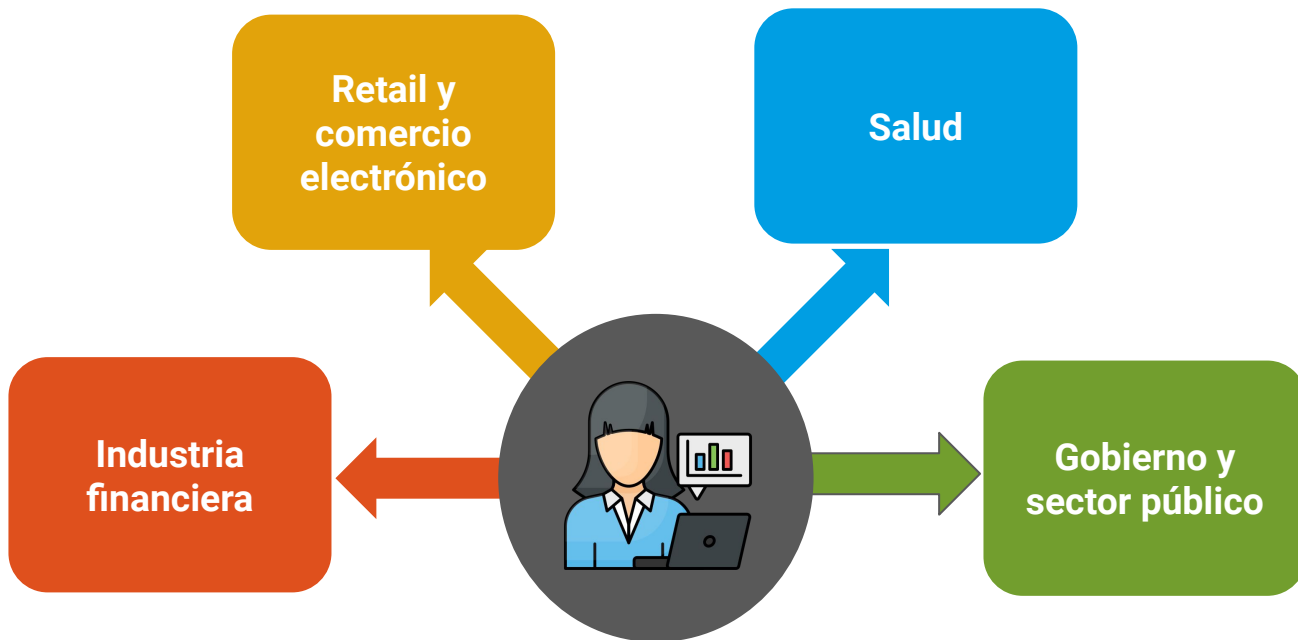


Informar adecuada y
efectivamente los datos
más relevantes, y los
resultados del análisis.

Interpretar los resultados
que arroja el análisis.

El rol Data Analyst

¿Dónde trabaja?

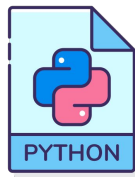


¿Qué herramientas utiliza?

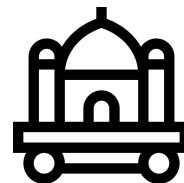
El ecosistema del Data Analyst



Lenguaje de programación especializado en el manejo de bases de datos.



Poderoso y moderno lenguaje de programación, con interesantes extensiones para el análisis de datos.



Tableau

Software de visualización de datos, mediante gráficas muy atractivas y eficientes.



¿Conoces alguna de
estas herramientas?

¿Conoces otras?



Analyst / Engineer / Scientist

Diferencias y semejanzas

Analistas de datos:
estudian datos
numéricos para
entregar información,
para la toma de
decisiones.



Data Engineer:
diseñan, construyen y
mantienen sistemas
de almacenamiento y
procesamiento de
datos.



Data Scientist:
observan y analizan
gran cantidad de
datos complejos para
descubrir e interpretar
patrones.



/* Analítica y problemas de negocios*/

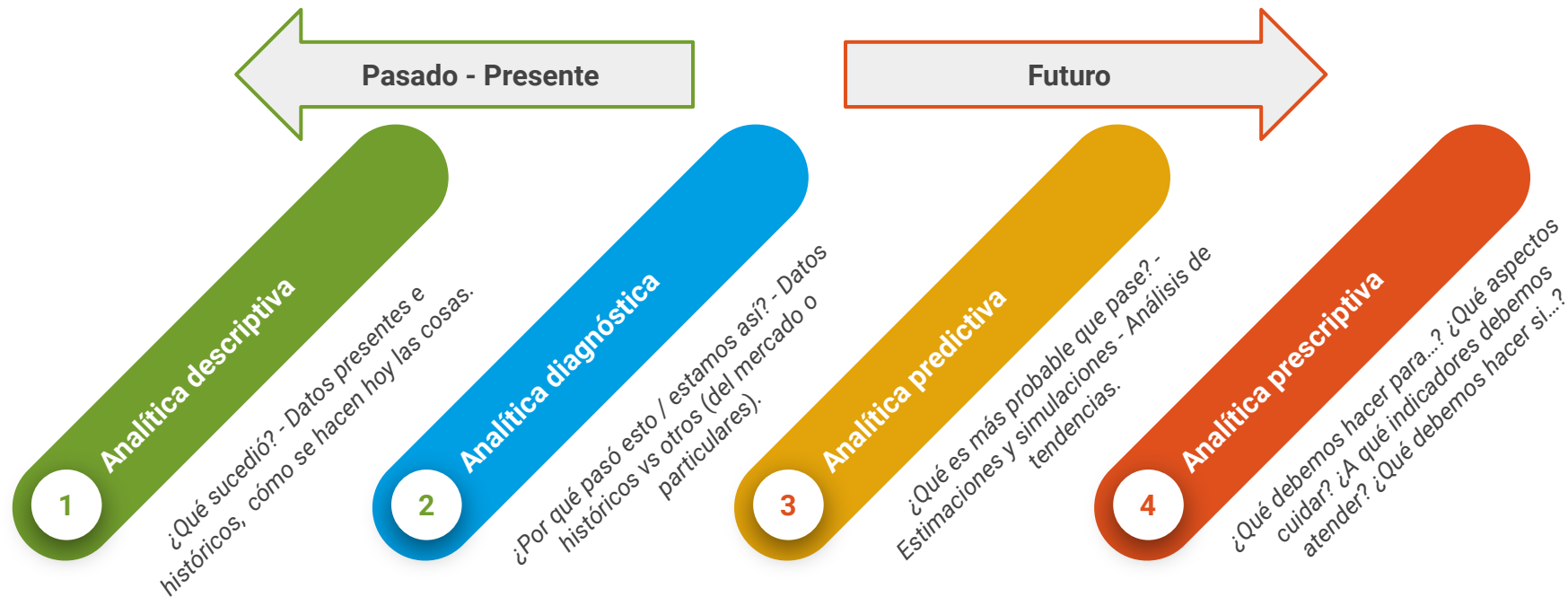
Tipos de analítica

¿Qué son y para qué sirven?

- Son diferentes los enfoques y niveles de profundidad que se utilizan en el análisis de datos.
- Se emplean para abordar diferentes problemas de negocios, desde entender la situación actual hasta tomar decisiones estratégicas para mejorar el desempeño y los resultados.



Tipos de analítica

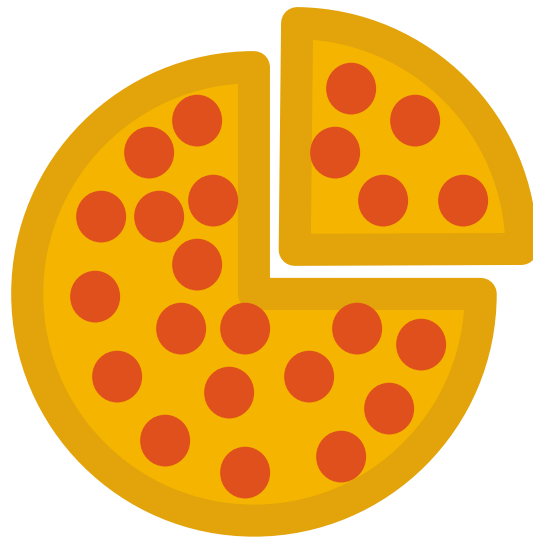


Una situación, distintos enfoques

Ejemplo: Caso

Una pizzería desea diseñar un plan de desarrollo para el año, considerando diferentes aspectos:

- ¿Cuál es la pizza más vendida?
- ¿Qué edades tienen los consumidores?
- ¿Compran en local, o con despacho?
- ¿Por teléfono, web, app?
- Ubicación geográfica de los despachos.
- Tiempos de espera.
- Monto promedio consumido.



Una situación, distintos enfoques

Ejemplo: Reportería

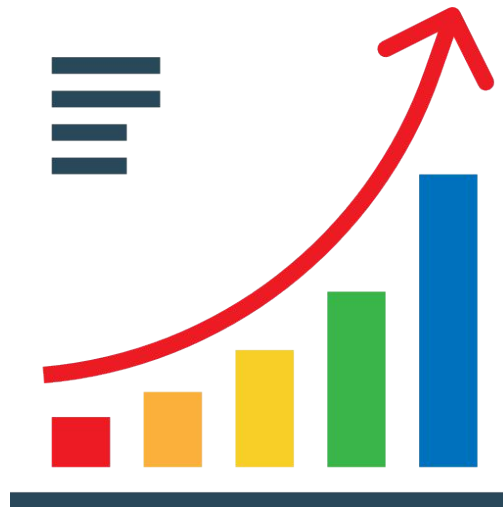
- Se enfoca en presentar información:
 - De manera visual y fácil de entender
 - Puede incluir gráficos, tablas y dashboards
- Se busca facilitar la toma de decisiones, accediendo en forma efectiva a la información.
- Todo análisis de datos requiere que los datos se presenten de manera adecuada.
- Inicialmente, la reportería es un caso de **analítica descriptiva**.



Una situación, distintos enfoques

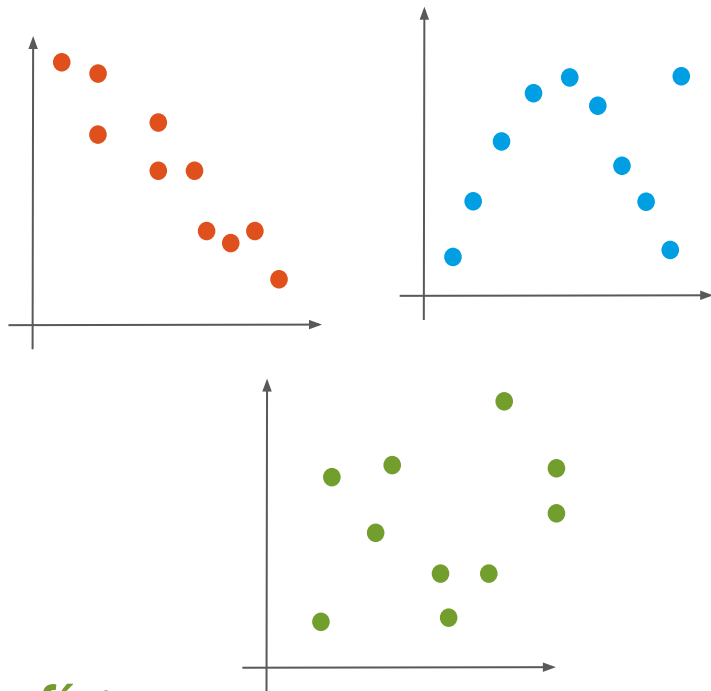
Ejemplo: Correlaciones

- ¿Se puede relacionar la edad del cliente con su consumo?
- ¿Tiempo de espera v/s monto dispuesto a gastar?
- Tiempo “vitrineando” v/s consumo?



Una situación, distintos enfoques

"A más X, más Y"



- Medida estadística que indica la relación entre dos variables.
- ¿Es posible calcular/aproximar una variable a partir de la otra?
- Una **correlación positiva** indica que a medida que aumenta una variable, la otra también aumenta.
- En cambio, una **correlación negativa** indica que a medida que aumenta una variable, la otra disminuye.

Una situación, distintos enfoques

Correlaciones y patrones

- La determinación de correlaciones es una forma de **analítica descriptiva**.
- Se utilizan indicadores estadísticos y herramientas como la **regresión**, que veremos más adelante.
- Una forma básica de observar correlaciones es mediante diagramas de puntos, que permiten sugerir formas o patrones.
- Si se logra establecer y modelar la correlación, estaremos realizando una **analítica predictiva**



Una situación, distintos enfoques

Ejemplo: Segmentaciones



- Los compradores entre 20 y 40 años compran más por app que los mayores.
- Pero en el mismo rango etario, los hombres suelen comprar más por app que las mujeres.
- Y entre estos hombres, se observan diferencias notorias en los medios de pago (débito, crédito, cupones).
- **¿Podemos identificar y diseñar estrategias específicas?**



Una situación, distintos enfoques

Ejemplo: Segmentaciones y tipos de analítica

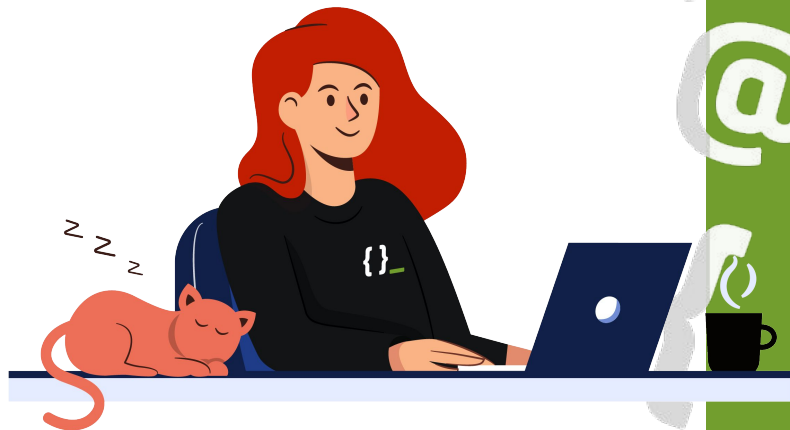
- La segmentación se relaciona con la analítica descriptiva y diagnóstica.
- La **analítica descriptiva** se utiliza para comprender los datos a través de la visualización y el resumen de las características y tendencias de los datos.
- La **analítica diagnóstica** se usa para identificar las causas subyacentes de las tendencias y patrones encontrados en los datos.
- Es fundamental que la segmentación no sea antojadiza, sino que refleje diferencias esenciales en el negocio.



Una situación, distintos enfoques

Ejemplo: A/B test

- ¿Vendemos la pizza a \$X cada unidad? ¿O la vendemos a \$2X pero promocionando “2x1”?
- Si una oferta es por tiempo limitado, ¿conviene decir solo eso? ¿O es mejor decir “solo hasta hoy a las 23:59”?



Una situación, distintos enfoques

Ejemplo: A/B test



Fuente:

<https://sebastianpendino.com>

{desafío}
latam_

- Herramienta para mejorar el desempeño de un negocio que permite experimentar con diferentes opciones, obteniendo una comprensión clara y confiable de qué variante es más efectiva.
- La información obtenida puede ser utilizada para tomar decisiones basadas en datos sobre cómo mejorar la experiencia de los usuarios, aumentar las conversiones, reducir el abandono de carritos de compra, etc.

Una situación, distintos enfoques

Ejemplo: A/B Test y tipos de analítica

En un A/B Test, se divide un grupo de usuarios en A y B.

Se presenta a cada grupo diferentes versiones de un producto o una página web, para medir las diferencias en el comportamiento de los usuarios entre los dos grupos.

**Analítica
predictiva**



Para predecir cuál de las dos versiones tendrá un mejor rendimiento en términos de conversión, retención de usuarios, etc.

**Analítica
prescriptiva**



Para identificar la mejor acción a tomar, basándose en los resultados del experimento.



Analíticas y problemas

- La separación entre tipos de analítica no es tan rígida, depende del tipo de problema.
- De todos modos, es fundamental en cada situación de negocios:
 - Recopilar los datos actuales (¿Qué pasó?)
 - Interpretar los datos y relacionarlos (¿Por qué pasó?)
 - Prever escenarios y ensayar posibles respuestas (¿Qué pasaría si...?)
 - Tomar decisiones y acciones, basados en lo anterior (¿Qué debemos hacer?)



/* Metodologías de trabajo */

¿Qué es para ti una
metodología de trabajo?

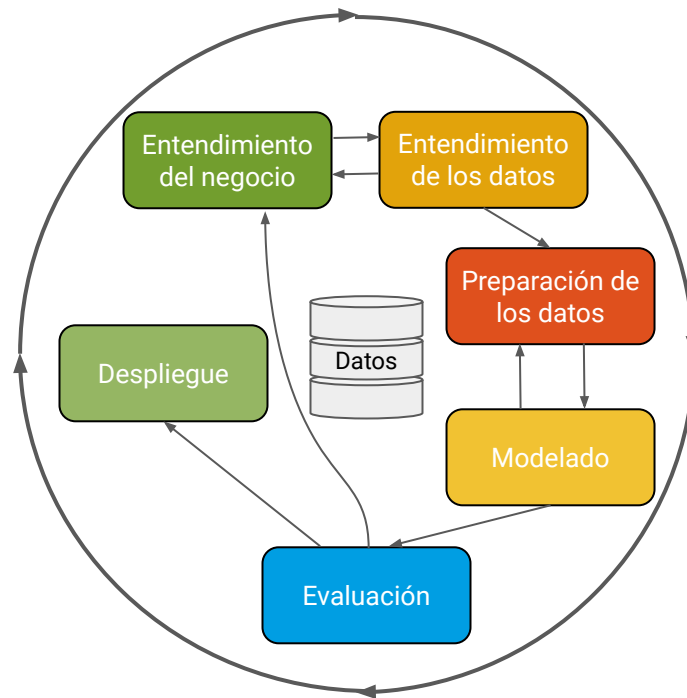
¿Conoces alguna que
sea propia del análisis de
datos?



Metodología CRISP - DM

Esquema general

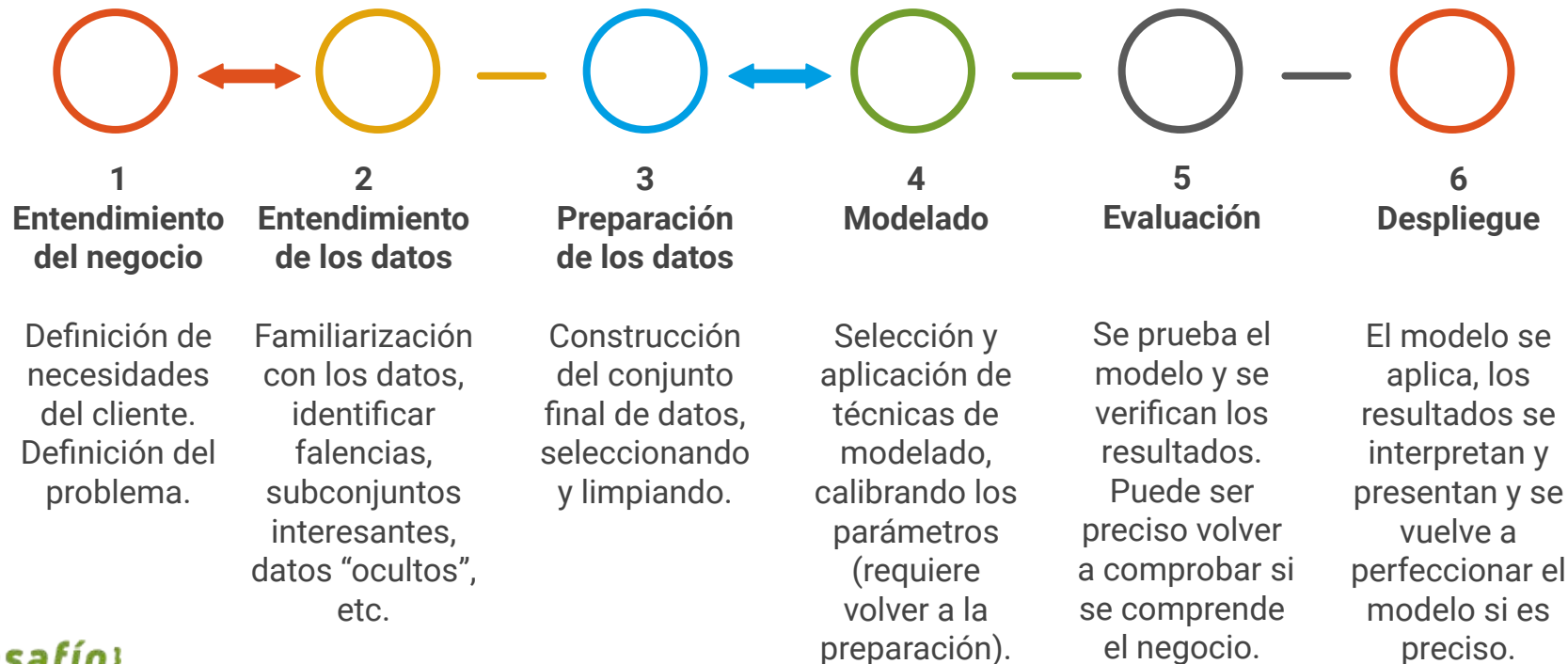
- Es un marco de trabajo para afrontar cualquier problema relacionado con datos.
- A pesar de que en la práctica no es necesario ser explícito en la aplicación de todos los pasos, siempre deben ser considerados todos.



Fuente: [Crisp](#)

Metodología CRISP - DM

Pasos



/* Datos: tipos, tablas y fuentes */

Entidades y atributos

Contexto

- Cuando hablamos de **entidades** nos referimos a un conjunto de “objetos” que tienen similitudes, por ejemplo un vehículo.
 - Un vehículo puede ser una auto, una moto, una bicicleta, una patineta, etc.
 - El objeto es vehículo.
- Estos tipos vehículos tienen particularidades, que se conocen como **atributos**.
 - **Moto de 2 a 4** ruedas, impulsada por un motor de combustión o eléctrico.
 - **Auto de 4 a 6** ruedas o más, impulsado por un motor de combustión o eléctrico.
 - **Bicicleta de 2** ruedas impulsada por la fuerza generada por una persona.

Entidades y atributos

Aplicado en figuras geométricas



Tablas

¿Qué son?

- Es el lugar donde almacenamos nuestros datos, muy parecido a lo que sería hoja de cálculo de excel o numbers donde almacenamos datos de diferentes tipos.
- Una tabla se compone de **filas** y **columnas**. Cada fila corresponde a una **instancia**, y en cada columna registramos un **atributo**.

Tablas

Tipos de datos

- Cada atributo o campo corresponde a un tipo de valor (value) que le puede ser asignado.
- Existen diversos tipos de datos; es conveniente siempre utilizar el más adecuado para cada atributo.

| string | number | string | boolean | dinero |
|--------|--------|----------|---------|--------|
| id | lados | color | regular | precio |
| chan | 7 | rojo | true | \$5 |
| blo | 8 | verde | false | \$7 |
| atsin | 6 | amarillo | false | \$10 |
| welbu | 12 | celeste | true | \$4 |

Ejercicio: Identificar elementos de una tabla



Identificar elementos de una tabla

- ¿A qué corresponden las identidades de la tabla?
- ¿Cuál es el valor del tercer atributo de la segunda entidad de la tabla?
- ¿Cuál es el valor del segundo atributo de la primera entidad de la tabla?
- ¿A qué tipo de dato corresponde cada columna?

| id | fe.nac | inscrito | Prom |
|-----------|------------|----------|------|
| camila135 | 21/08/1987 | true | 5.6 |
| daniel536 | 31/05/1990 | true | 4.7 |
| isabel43 | 18/07/1988 | false | 6.3 |

Organización de datos

Estructurados



- Están organizados de manera uniforme y siguen un esquema preestablecido, permitiendo su almacenamiento y análisis eficiente.
- Por ejemplo, información almacenada en bases de datos relacionales y tablas, los cuales se organizan generalmente en columnas y filas.
- Un ejemplo de una tabla de datos estructurados podría ser una tabla de clientes, en la que cada fila representa a un cliente y cada columna contiene información sobre ese cliente, como su nombre, dirección, número de teléfono y fecha de registro, etc.

Organización de datos

Semi Estructurados

- Están organizados y contienen información, pero no se ajustan a un esquema o estructura predeterminada.
- Si bien son más complejos de analizar, de ellos se pueden extraer conclusiones útiles mediante técnicas de procesamiento de lenguaje natural, análisis de texto y herramientas de extracción de metadatos.
- Por ejemplo: Hojas de cálculo no consistentes, bases de datos no relacionales (noSQL), datos en CSV, archivos Json y XML, archivos de página web.



| | V | |
|------|----|--|
| 5 | | |
| | F | |
| ocho | Sí | |

Organización de datos

No Estructurados



- No tienen un formato predefinido o no están organizados de una forma lógica.
- No se pueden consultar con la misma facilidad que los datos estructurados y requiere un proceso de conversión para su uso.
- Por lo general, se requiere de técnicas avanzadas, como procesamiento de lenguaje natural, para poder trabajar con ellos.
- Por ejemplo: texto, el cual podría ser directamente un artículo, una tabla con las opiniones de clientes, una consulta de datos traídos de twitter, o incluso imágenes.

Tablas

Fuentes de datos



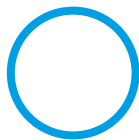
SQL

Archivo estándar de bases de datos. Es un archivo de texto plano, interpretable por un motor de bases de datos.



XLSX

Archivo en formato Microsoft Excel.



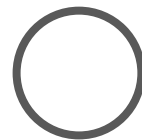
CSV

Significa Comma separated values, y permite que las columnas queden separadas por comas, y las líneas por saltos.



XML

Formato de etiquetas para diseño web, con marcado.



APIs

Interfaz de programación de aplicaciones.

¿Con qué tipo de datos
se trabajará en este
módulo?



Organización de datos

El Data Analyst



- Trabajaremos con datos estructurados, para aprender.
- Sin embargo, en su labor diaria, gran parte el trabajo de un Analista de datos será “limpiar” y preparar los datos para poder trabajar con ellos en forma efectiva

Desafío - Nuestro primer análisis



Desafío

"Nuestro primer análisis"

- Descarga el archivo "Desafío".
- Tiempo de desarrollo asincrónico: desde 4 horas.
- Tipo de desafío: grupal.
- Revisemos los requerimientos desde plataforma.

¡AHORA TE TOCA A TI! 💪



Ideas fuerza

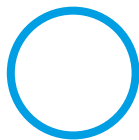
El Data Analyst



Extrae, analiza y presenta datos para la **toma de decisiones**.



Utiliza herramientas como **SQL**, **Python** y **Tableau**.

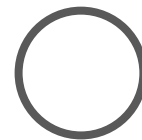


Aplica analíticas: **descriptiva**, **diagnóstica**, **predictiva** y **prescriptiva**.



Aborda problemas de **A/B test**, **Correlaciones**, **Segmentaciones** y **Reportería**.

Aplica metodologías de trabajo, como **CRISP-DM**.



Trabaja principalmente con datos **estructurados en tablas**.

¿Qué concepto te gustaría
reforzar?

¿Con qué te quedas de la
sesión?



Recursos asincrónicos

¡No olvides revisarlos!

Esta semana contarás con los siguientes recursos:

- Infografía “Otras herramientas”.
- Guía de estudio “Tipos de analítica y otras metodologías”.
- Desafío “Nuestro primer análisis”.





Próxima sesión...

- *Tutoría.*

{desafío}
latam_

*Academia de
talentos digitales*

