



Diplomado para Acceder a Grado

Módulo Ciencia de Datos

Analítica de datos

Dr. José Ramón Iglesias

DSP-ASIC BUILDER GROUP

Director Semillero TRIAC

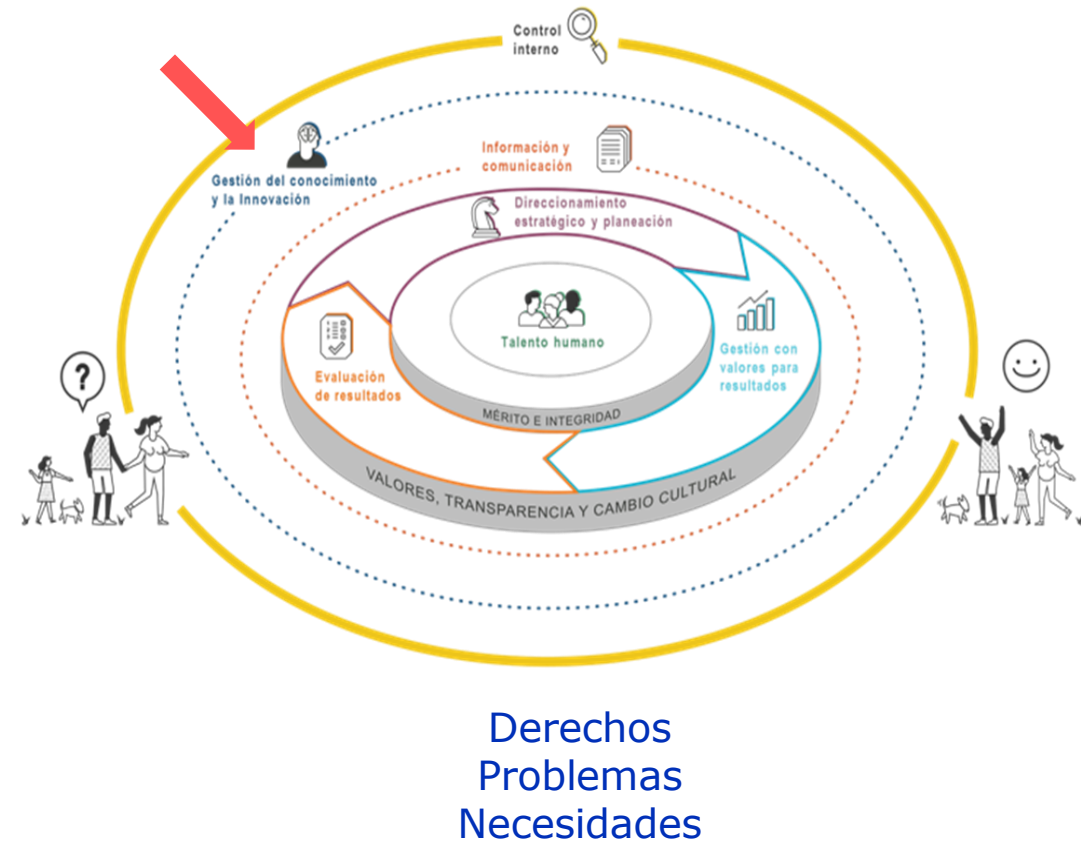
Ingeniería Electrónica

Universidad Popular del Cesar

01. Introducción

Datos que transforman desde la gestión del conocimiento

Generar resultados que atiendan los planes de desarrollo y garanticen los derechos, resuelvan las necesidades y problemas de los ciudadanos con integridad y calidad en el servicio



Analítica institucional

Profundizar en los análisis de la información y los datos que genera la entidad para fortalecer el conocimiento de su desempeño y de su propósito fundamental



Diagnóstico de datos:

Identificación, organización y actualización de los datos producidos por la entidad



Tableros de indicadores:

Generación de indicadores y monitoreo de la gestión de la entidad



Análisis de datos:

Profundizar en los análisis de datos internos y externos, haciendo cruces que permitan extraer nuevas perspectivas de los datos y de la información



Visualizaciones:

Visualización de datos para formular preguntas diferentes y contar historias a través de los datos



02. Analítica de datos

¿Por qué conocer no es suficiente?



¿Qué conocemos de analítica?

Juguemos !!

1 De las siguientes opciones ¿Cómo se define la analítica?

- a. Técnicas computacionales usadas de forma eficiente en datos masivos y distintos (estructurados y no estructurados) para la toma de decisiones.
- b. Conjunto de servicios de integración de sistemas de información.
- c. Analizar patrones a partir de los datos para generar conocimiento de valor para la toma de decisiones.
- d. Algoritmos y técnicas de Aprendizaje de Máquina enfocados en el análisis de patrones y predicción.

1 De las siguientes opciones ¿Cómo se define la analítica?

- a. Técnicas computacionales usadas de forma eficiente en datos masivos y distintos (estructurados y no estructurados) para la toma de decisiones.
- b. Conjunto de servicios de integración de sistemas de información.
- c. Analizar patrones a partir de los datos para generar conocimiento de valor para la toma de decisiones.**
- d. Algoritmos y técnicas de Aprendizaje de Máquina enfocados en el análisis de patrones y predicción.

2. ¿Cómo se podría dividir la analítica?

- a. Descriptiva, Diagnóstica, Predictiva, Descriptiva
- b. Revisión, Gestión, Validación
- c. Descriptiva y Propositiva
- d. Big Data, Minería de Datos, Analítica Avanzada

2. ¿Cómo se podría dividir la analítica?

a. Descriptiva, Diagnóstica, Predictiva, Descriptiva

b. Revisión, Gestión, Validación

c. Descriptiva y Propositiva

d. Big Data, Minería de Datos, Analítica Avanzada

3. ¿Cuál de las siguientes preguntas corresponden a analítica predictiva?

- a. ¿Cuál es la probabilidad de que un niño mejore su estado de nutrición al incluirlo en un programa específico?
- b. ¿Cuál es la probabilidad de que un estudiante deserte en los próximos 12 meses?
- c. a y b
- d. Ninguna de las anteriores

3. ¿Cuál de las siguientes preguntas corresponden a analítica predictiva?

- a. ¿Cuál es la probabilidad de que un niño mejore su estado de nutrición al incluirlo en un programa específico?
- b. ¿Cuál es la probabilidad de que un estudiante deserte en los próximos 12 meses?
- c. **a y b**
- d. Ninguna de las anteriores

4. ¿Cuál de las siguientes preguntas corresponden a analítica prescriptiva?

- a. ¿Cuáles serían las estrategias que generarían mejor impacto en la disminución de la deserción escolar?
- b. ¿Cuántos ciudadanos se han graduado los últimos 5 años?
- c. ¿Cuántos niños han cambiado de estado de nutrición en los últimos 12 meses?
- d. Ninguna de las anteriores

4. ¿Cuál de las siguientes preguntas corresponden a analítica prescriptiva?

- a. ¿Cuáles serían las estrategias que generarían mejor impacto en la disminución de la deserción escolar?
- b. ¿Cuántos ciudadanos se han graduado los últimos 5 años?
- c. ¿Cuántos niños han cambiado de estado de nutrición en los últimos 12 meses?
- d. Ninguna de las anteriores

5. ¿Cómo se podría definir Big Data?

- a. Algoritmos y técnicas de Aprendizaje de Máquina enfocados en el análisis de patrones y predicción.
- b. Base de datos con más de 1TB de datos.
- c. Analizar patrones a partir de los datos para generar conocimiento de valor para la toma de decisiones.
- d. Técnicas computacionales usadas de forma eficiente en datos masivos y distintos (estructurados y no estructurados) para la toma de decisiones

5. ¿Cómo se podría definir Big Data?

- a. Algoritmos y técnicas de Aprendizaje de Máquina enfocados en el análisis de patrones y predicción.
- b. Base de datos con más de 1TB de datos.
- c. Analizar patrones a partir de los datos para generar conocimiento de valor para la toma de decisiones.
- d. **Técnicas computacionales usadas de forma eficiente en datos masivos y distintos (estructurados y no estructurados) para la toma de decisiones**

“Bienestar y la Competitividad sostenibles”



Conocimiento profundo del Estado, el Ciudadano y el Territorio

Analítica y Big Data



Datos

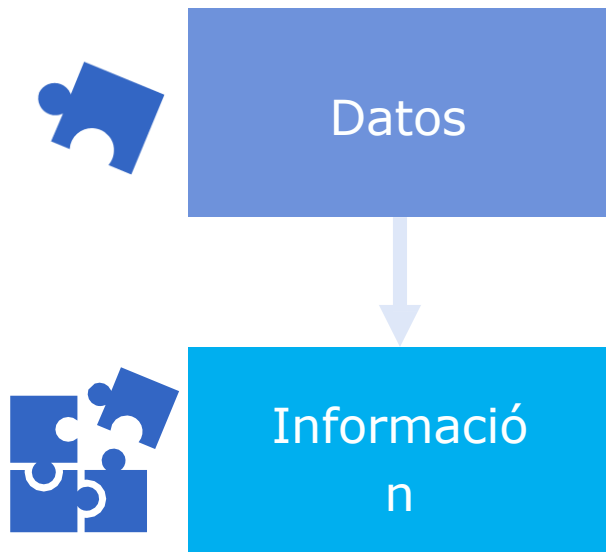
Datos – Información – Decisiones

Un dato es un conjunto discreto, de factores objetivos sobre un hecho real (Davenport et al., 1999)

- 54
- Masculino
- 2022
- 2021-01-05
- MIPG

Es el cruce entre Filas y Columnas en una hoja de cálculo, es la base de análisis de datos.

Datos – Información – Decisiones

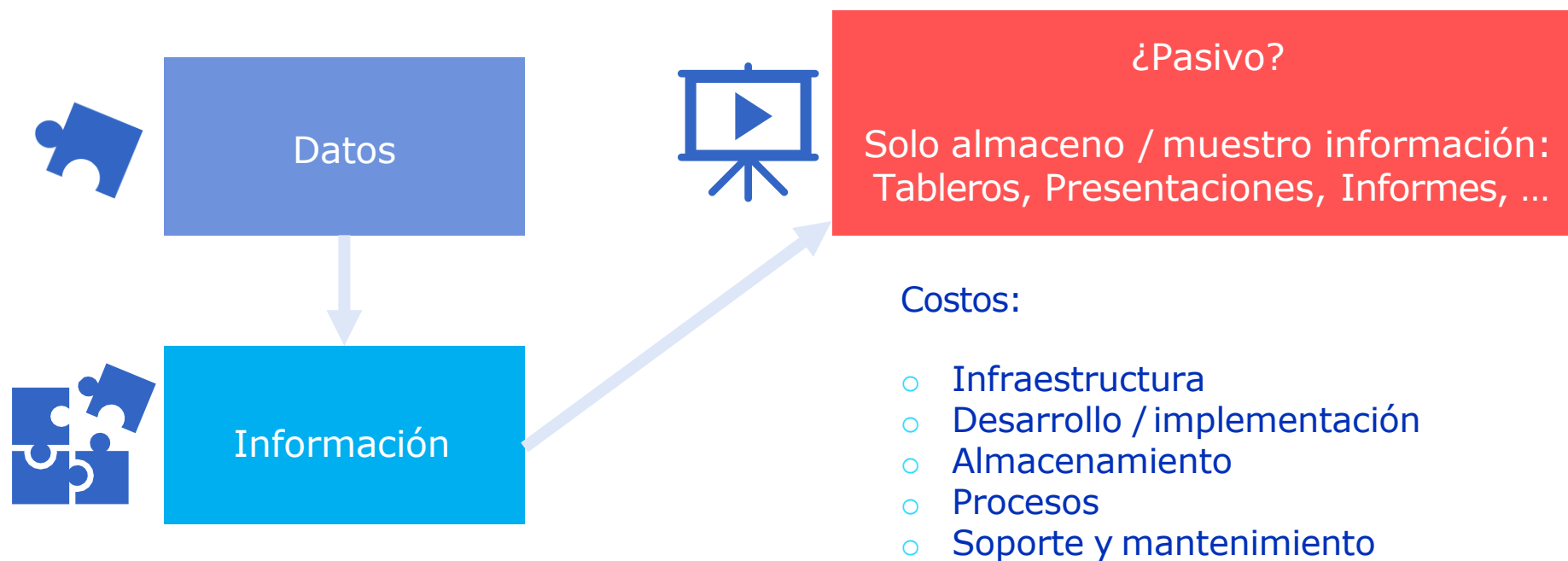


La Información Corresponde a datos con contexto – sentido
(Davenport et al., 1999)

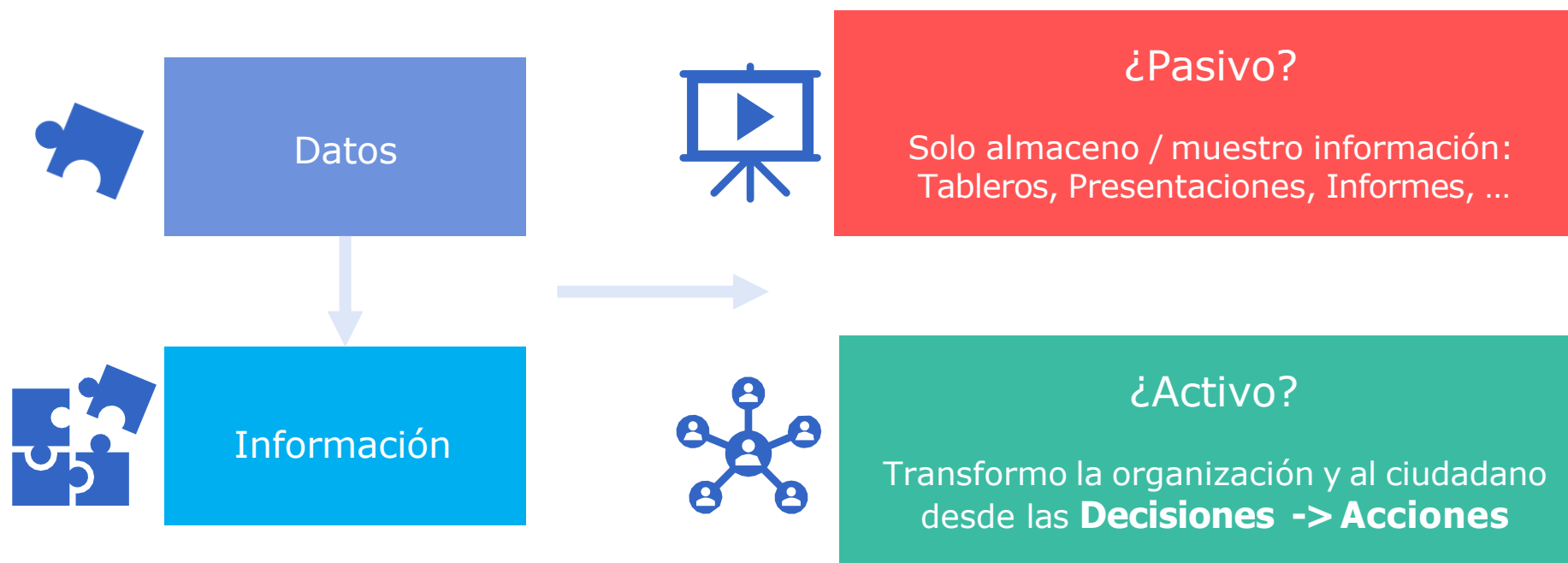
- Edad: 54
- Sexo: Masculino
- Año actual: 2022
- Fecha de actualización: 2021-01-05
- Fuente de información: MIPG

Permite darle sentido a los datos y analizarlos en un contexto específico

Datos – Información – Decisiones



Datos – Información – Decisiones



Decisiones que generen acciones
logrando bienestar y competitividad



Datos

Tipos de datos: privacidad

- **Datos sensibles:** se entiende por datos sensibles aquellos que afectan la intimidad del Titular o cuyo uso indebido puede generar su discriminación, tales como aquellos que revelen el origen racial o étnico, la orientación política, las convicciones religiosas o filosóficas, la pertenencia a sindicatos, organizaciones sociales, de derechos humanos o que promueva intereses de cualquier partido político o que garanticen los derechos y garantías de partidos políticos de oposición así como los datos relativos a la salud, a la vida sexual y los datos biométricos.
- **Datos personales:** Es cualquier pieza de información vinculada a una o varias personas determinadas o determinables o que puedan asociarse con una persona natural o jurídica. Los datos personales pueden ser públicos, semiprivados o privados.



Datos

No es solo tener datos, es usarlos
adecuadamente

Tipos de datos: privacidad

- **Dato privado:** Es el dato que por su naturaleza íntima o reservada sólo es relevante para el titular
- **Dato semiprivado:** Es semiprivado el dato que no tiene naturaleza íntima, reservada, ni pública y cuyo conocimiento o divulgación puede interesar no sólo a su titular sino a cierto sector o grupo de personas o a la sociedad en general, como el dato financiero y crediticio de actividad comercial o de servicios.
- **Dato público:** Es el dato calificado como tal según los mandatos de la ley o de la Constitución Política y todos aquellos que no sean semiprivados o privados, de conformidad con la ley. Son públicos, entre otros, los datos contenidos en documentos públicos, sentencias judiciales debidamente ejecutoriadas que no estén sometidos a reserva y los relativos al estado civil de las personas.



Datos

Tipos de datos: variables

Cualitativos: Hacen referencia a características o cualidades cuya medición no corresponde a números (suelen ser dimensiones o atributos de análisis) Podemos distinguir dos tipos:

- **Nominales:** Son variables que no tienen orden, ejemplo: Sexo, País, ...
- **Ordinales:** Son variables que pueden asociarle un orden, ejemplo: Estrato, Estado Proceso, ...



Datos

Tipos de datos: variables

Cuantitativos: Es aquella que puede ser expresada numéricamente, y pueden ser:

- **Discretos:** Son aquellos que toman valores aislados (normalmente enteros) sin valores entre los mismos, ejemplo: # de Hijos, # de vehículos, ...
- **Continuos:** Este corresponde a aquellos datos con valores intermedios (normalmente reales), ejemplo: Promedio de altura, nivel de glicemia, probabilidad de ocurrencia, ...

Tableros de indicadores:
Generación de indicadores y monitoreo de la gestión de la entidad

Tipos de datos: almacenamiento

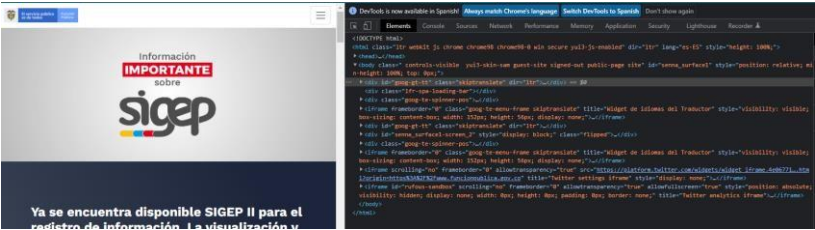
Estructurados: Son datos que están organizados normalmente en tablas u hojas de cálculo, en una estructura definida, por ejemplo: Hojas de Cálculo, Tablas en Bases de datos, ...

	LEY	NOMBRE LEY	REPORTE	TIPO	CTA	SER CTA	OBJ	ORD	SER ORD	SUB ITEM	FUENTE	REC	NOT
4	09-01-01	DEPARTAMENTO DE LA POLICIA PUBLICA - SECTOR GENERAL	A-01-01-01	A	01	01	01				Manejo	10	CRP
5	09-01-01	DEPARTAMENTO DE LA POLICIA PUBLICA - SECTOR GENERAL	A-01-01-02	A	01	01	02				Manejo	10	CRP
6	09-01-01	DEPARTAMENTO DE LA POLICIA PUBLICA - SECTOR GENERAL	A-01-01-03	A	01	01	03				Manejo	10	CRP
7	09-01-01	DEPARTAMENTO DE LA POLICIA PUBLICA - SECTOR GENERAL	A-01-01-04	A	01	01	04				Manejo	10	CRP
8	09-01-01	DEPARTAMENTO DE LA POLICIA PUBLICA - SECTOR GENERAL	A-01-01-05	A	01	01	05				Manejo	10	CRP
9	09-01-01	DEPARTAMENTO DE LA POLICIA PUBLICA - SECTOR GENERAL	A-01-01-06	A	01	01	06				Manejo	10	CRP
10	09-01-01	DEPARTAMENTO DE LA POLICIA PUBLICA - SECTOR GENERAL	A-01-01-07	A	01	01	07				Manejo	10	CRP
11	09-01-01	DEPARTAMENTO DE LA POLICIA PUBLICA - SECTOR GENERAL	A-01-01-08	A	01	01	08				Manejo	10	CRP
12	09-01-01	DEPARTAMENTO DE LA POLICIA PUBLICA - SECTOR GENERAL	A-01-01-09	A	01	01	09				Manejo	10	CRP
13	09-01-01	DEPARTAMENTO DE LA POLICIA PUBLICA - SECTOR GENERAL	A-01-01-10	A	01	01	10				Manejo	10	CRP
14	09-01-01	DEPARTAMENTO DE LA POLICIA PUBLICA - SECTOR GENERAL	A-01-01-11	A	01	01	11				Manejo	10	CRP
15	09-01-01	DEPARTAMENTO DE LA POLICIA PUBLICA - SECTOR GENERAL	A-01-01-12	A	01	01	12				Manejo	10	CRP
16	09-01-01	DEPARTAMENTO DE LA POLICIA PUBLICA - SECTOR GENERAL	A-01-01-13	A	01	01	13				Manejo	10	CRP
17	09-01-01	DEPARTAMENTO DE LA POLICIA PUBLICA - SECTOR GENERAL	A-01-01-14	A	01	01	14				Manejo	10	CRP
18	09-01-01	DEPARTAMENTO DE LA POLICIA PUBLICA - SECTOR GENERAL	A-01-01-15	A	01	01	15				Manejo	10	CRP



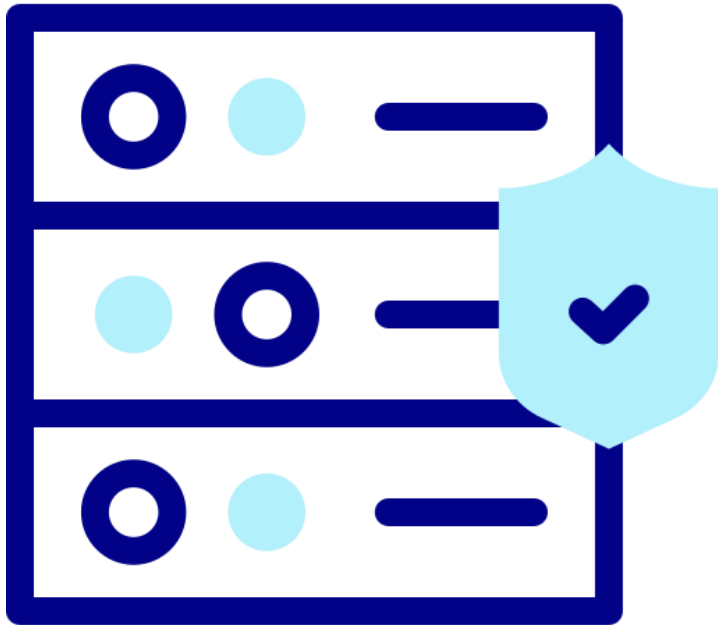
Semiestructurados: Son una combinación de los dos anteriores, en el cual existe un orden general pero los datos pueden ser no estructurados, por ejemplo: Mensajes JSON, XML, Hojas de Excel con imágenes y texto libre en ciertas columnas, ...

No estructurados: Son datos libres sin un formato único y que puede diferir entre registro y registro, por ejemplo: Blogs, Video, Audio, textos en PQRS.



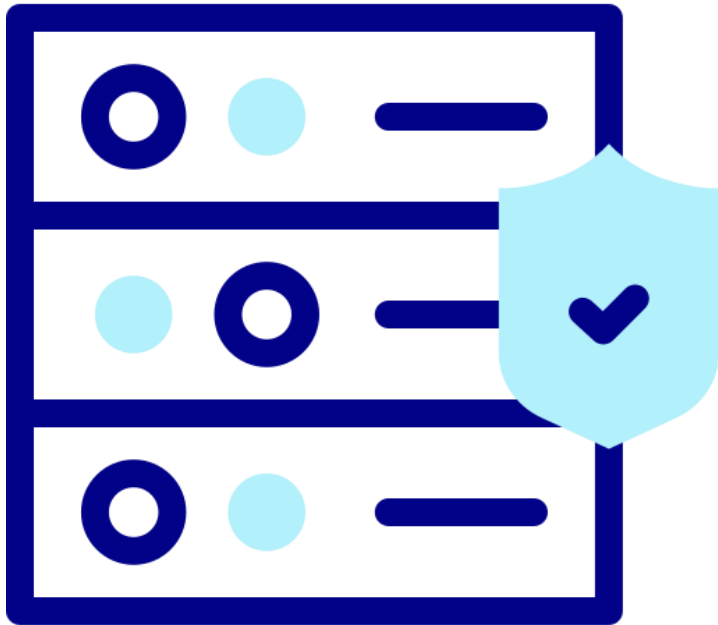
Fuente imágenes: <https://www.funcionpublica.gov.co/>
[OPEN DATA HANDBOOK]. (s.f). Recuperado de <http://opendatahandbook.org/>

Datos: calidad



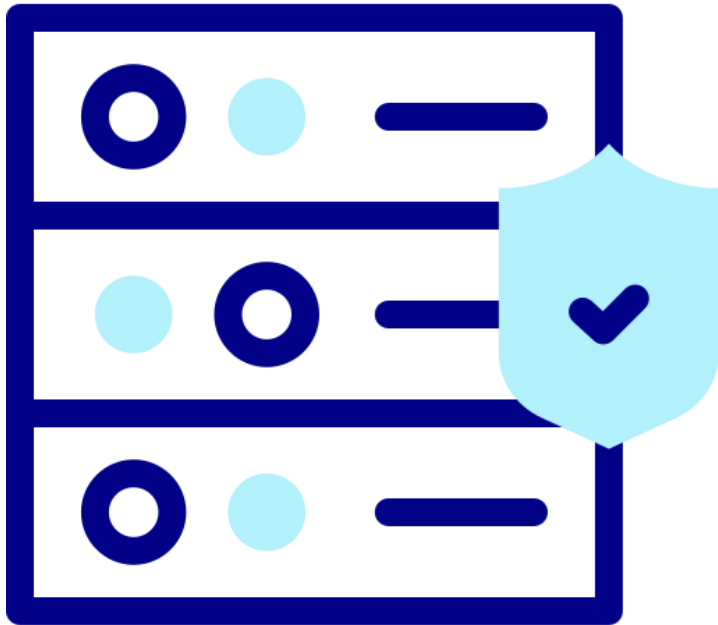
- **Exactitud (Accuracy):** Se mide el grado en el que los datos representan correctamente el objeto del mundo real o un evento que se describe.
Ejemplo: La dirección de envío de pedidos a un cliente en la base de datos de clientes es la dirección real.
- **Completitud (Completeness):** El grado en el que el dato tiene el valor esperado y cumple con los requerimientos marcados. Si un dato es opcional no debe considerarse para lograr el 100% de completitud.
Ejemplo: Podemos establecer que los clientes tendrán sus datos completos si hemos registrado su nombre, primer apellido, segundo apellido, número de identificación, e-mail, dirección, código postal, ciudad y país. El segundo nombre será opcional.

Datos: calidad



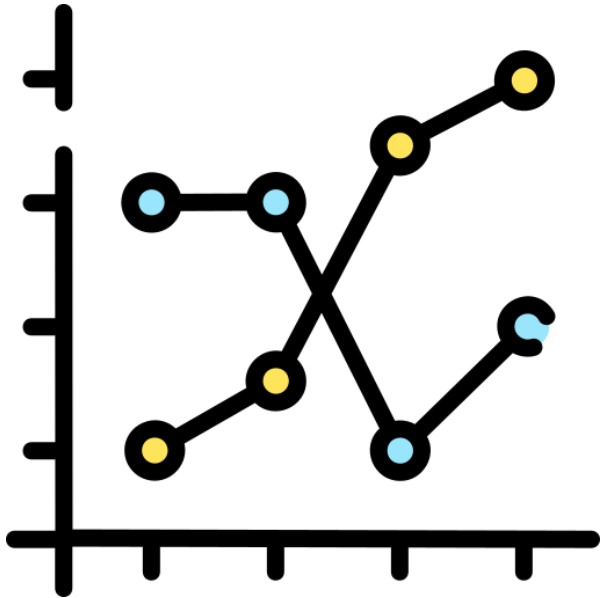
- **Consistencia (Consistency):** Mide si los datos están libres de contradicción y tienen coherencia lógica, de formato o temporal.
Ejemplo: Para un cliente determinado tenemos ventas registradas pero no nos consta ninguna orden de pedido.
- **Pertinencia temporal (Timeliness):** Mide el grado en que los datos están disponibles cuando se requieren.
Ejemplo: Para la asignación de habitaciones en un hotel, la recepción debe contar con el número actualizado de habitaciones disponibles en el momento de registrar la llegada del cliente.

Datos: calidad

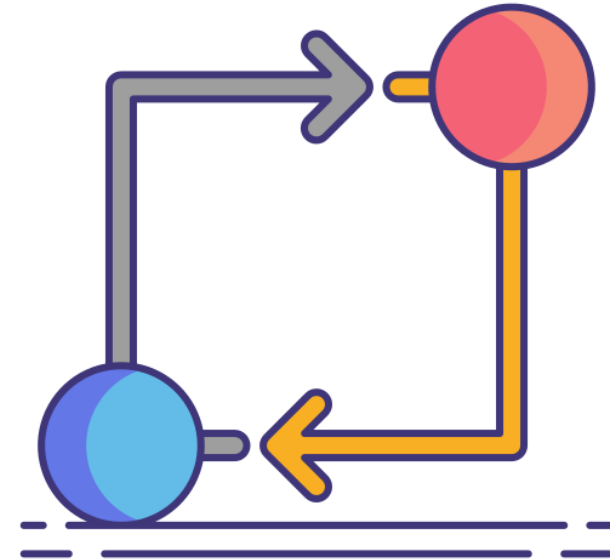


- **Unicidad (Uniqueness):** Cada dato es único. Con esta dimensión se busca corregir la duplicidad inesperada en nuestros dataset.
Ejemplo: En nuestra base de datos podemos tener dos clientes que se registraron como «Fran García» y «Francisco Juan García», siendo la misma persona pero sólo el último contiene todos los datos completos.
- **Validez (Validity):** Medir si un valor se ajusta a una regla de negocio o a un estándar preestablecido en cuanto a formato, tipo de dato, valores posibles o rangos especificados.
Ejemplo: En el seguimiento de entrega de un pedido, la última actualización es posterior a la hora actual. Dan Myers expone este caso en su blog explicando que si existiera una regla de negocio que indique que las actualizaciones no pueden producirse en una fecha y hora superior a la actual del sistema, este problema no se hubiera producido.

Correlación vs Causalidad



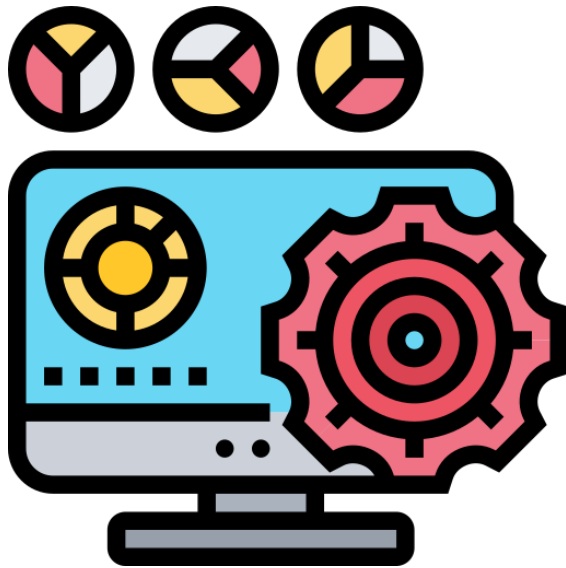
Vs





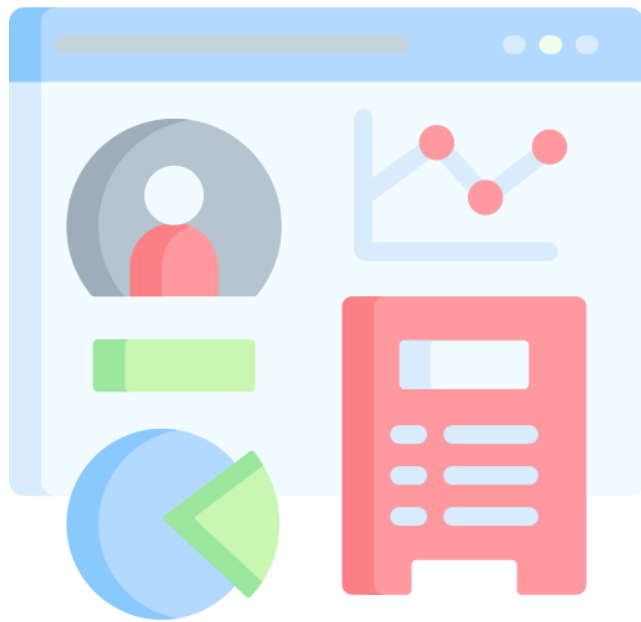
02. Analítica de Datos

Generación de valor desde los datos



Importancia

- Unificar fuentes y conceptos de información
- Tener una única verdad
- Análisis descriptivos
- Entender que paso y por qué paso (desde los datos)
- Medir y mejorar en los distintos procesos
- Identificación de oportunidades
- Fuente para la automatización de decisiones
- Uso de los datos ... Pasivo -> Activo



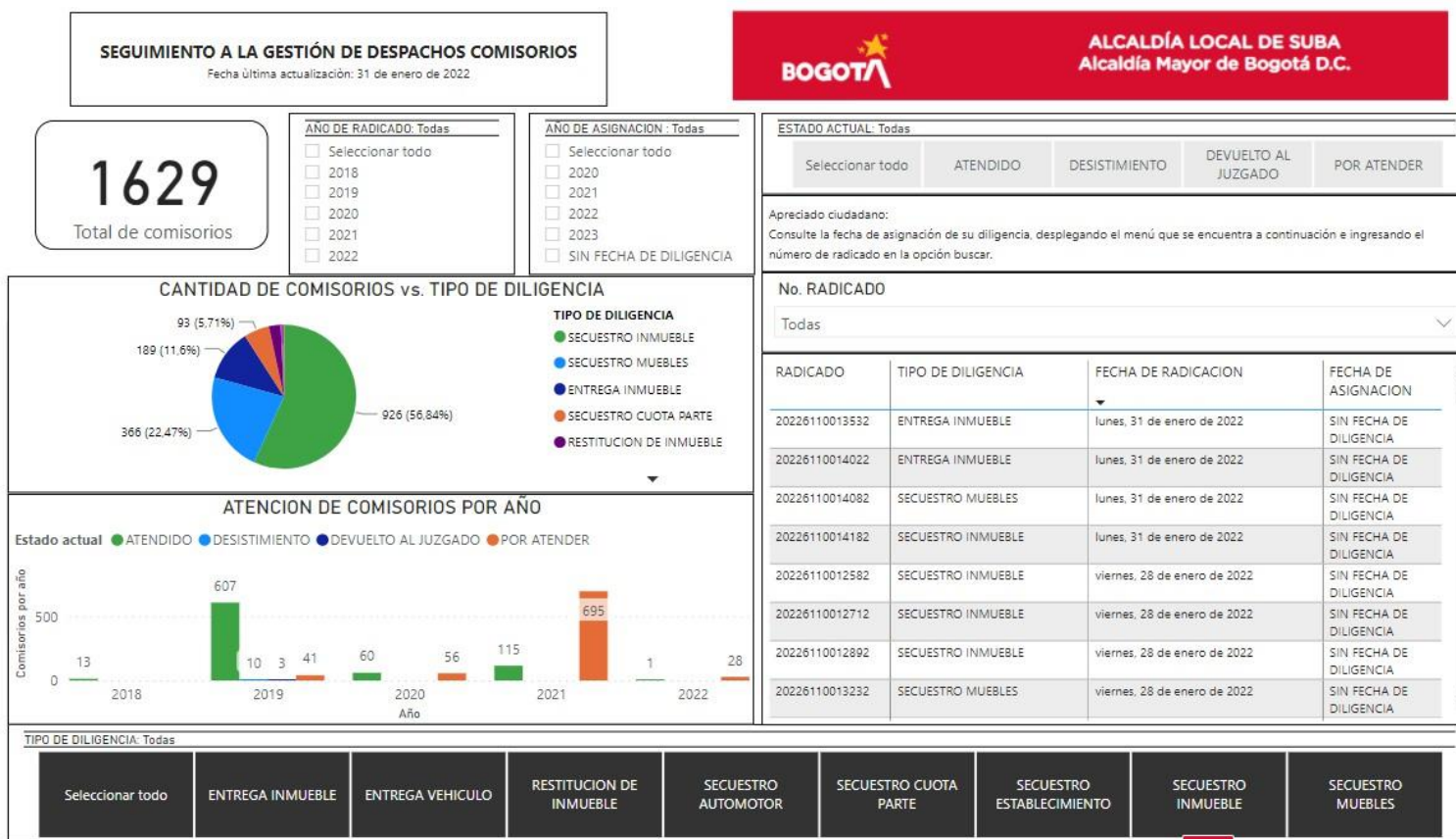
Principales productos

- Modelos de datos (modelos de industria)
- Cálculos e indicadores
- Reportes e informes por pregunta de negocio, proceso u otro.
- Cuadros de mando (dashboard) Operativos
- Cuadros de mando Corporativos y Estratégicos
- KPI (Key Performance Indicators)
- CPI (Corporate Performance Indicators)
- Democratizar la información y el análisis en las áreas



KPI: indicadores clave de desempeño

“Un **KPI** (*key performance indicator*), conocido también como **indicador clave o medidor de desempeño** o **indicador clave de rendimiento**, es una medida del nivel del rendimiento de un proceso. El valor del indicador está directamente relacionado con un objetivo fijado previamente y normalmente se expresa en valores porcentuales.”



CANTIDAD DE COMISORIOS vs. TIPO DE DILIGENCIA

TIPO DE DILIGENCIA

☐ SECUESTRO INMUEBLE

☐ SECUESTRO MUEBLES

☐ ENTREGA INMUEBLE

☐ SECUESTRO CUOTA PARTE

☐ RESTITUCION DE INMUEBLE

ATENCION DE COMISORIOS POR AÑO

Estado actual

☐ ATENDIDO

☐ DESISTIMIENTO

☐ DEVUELTO AL JUZGADO

☐ POR ATENDER

Año

2018

2019

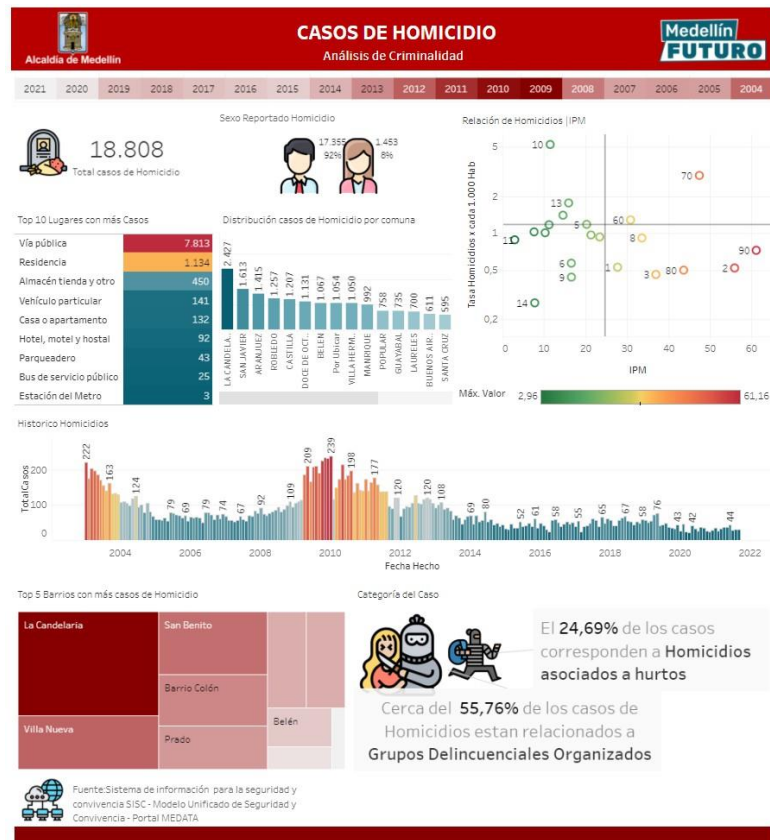
2020

2021

2022

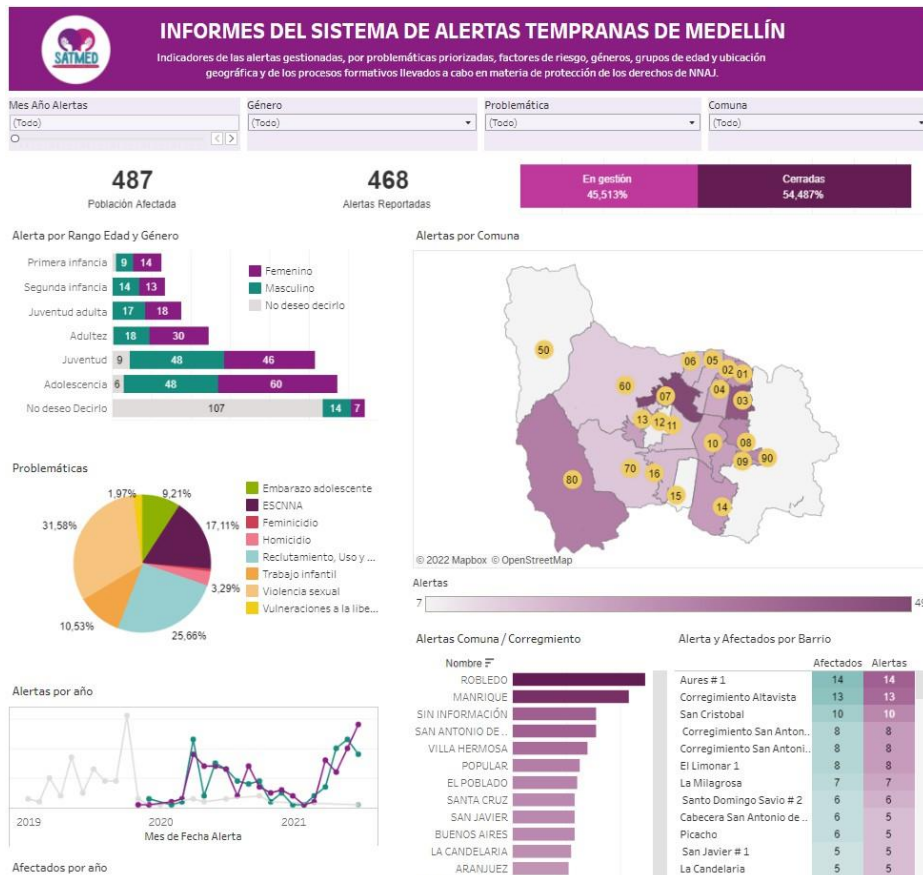
Ejemplos de Tableros

Gestión histórica de radicados y asignaciones en la Alcaldía Local de Suba – Bogotá D.C.



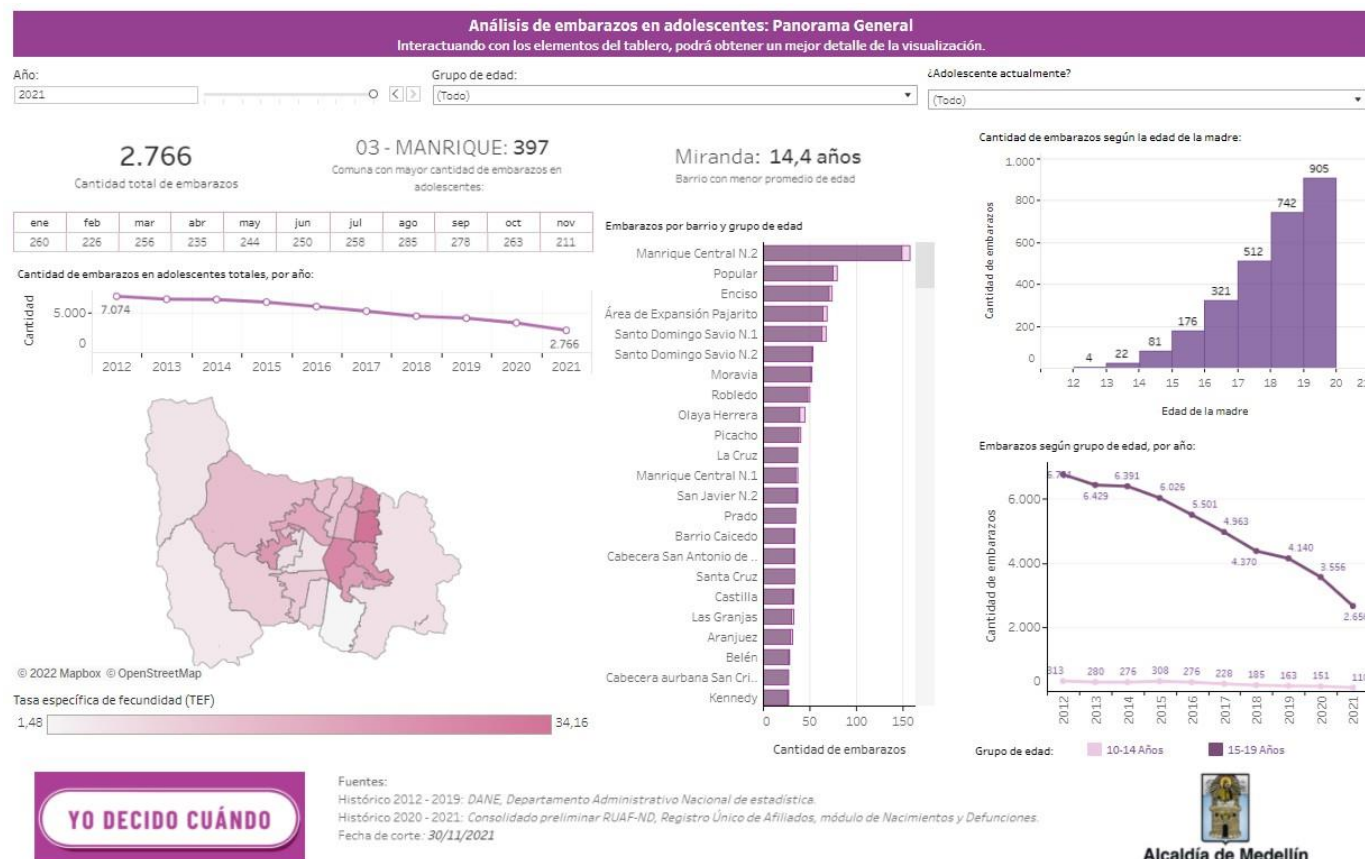
Ejemplos de Tableros

Análisis general de la criminalidad – casos de Homicidio en el municipio de Medellín.



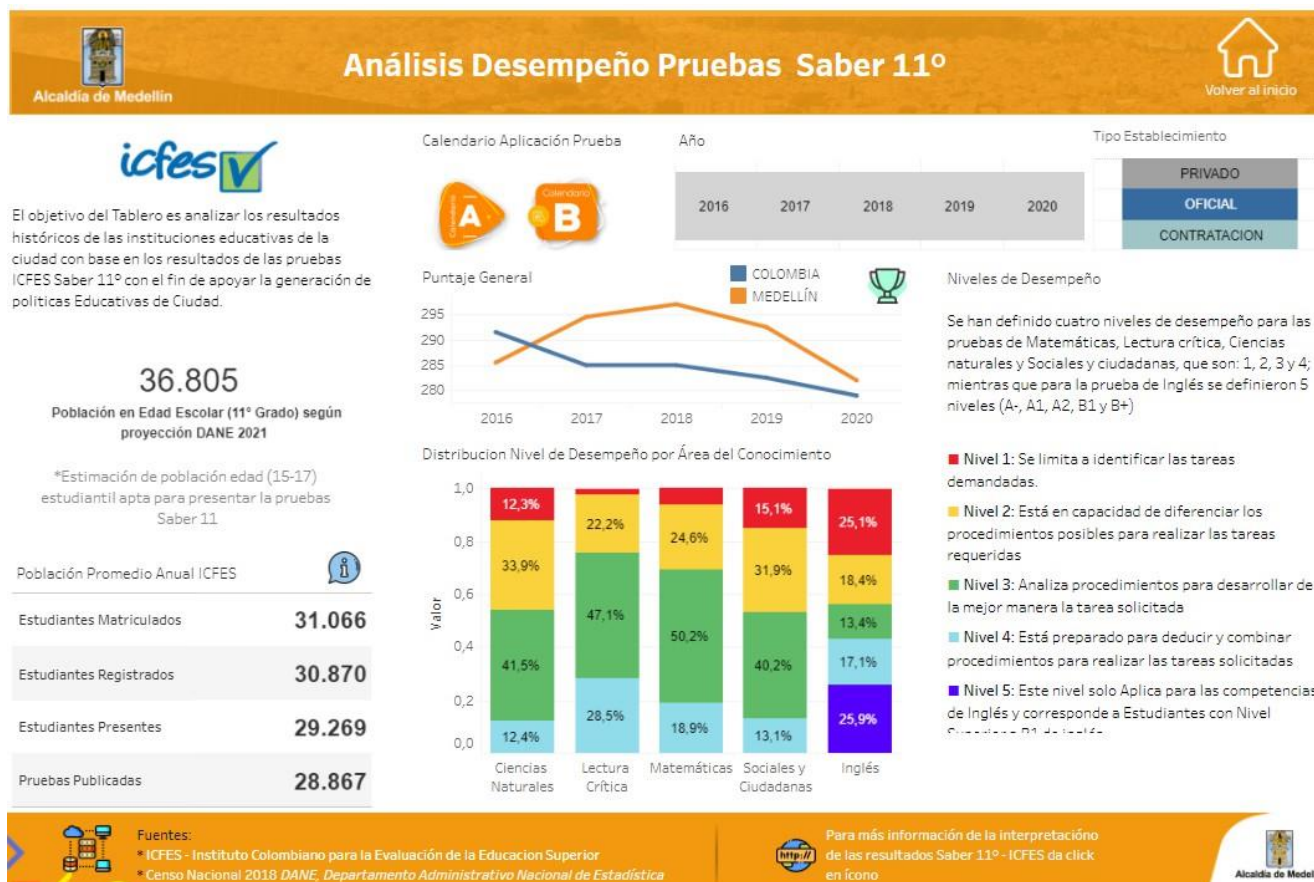
Ejemplos de Tableros

Análisis de las Alertas Tempranas registradas en la plataforma del municipio de Medellín



Ejemplos de Tableros

Identificación de los factores de riesgo de embarazo adolescente y subsecuente en el municipio de Medellín.



Ejemplos de Tableros

Análisis histórico del comportamiento de los estudiantes del municipio de Medellín en las pruebas Saber 11



Inteligencia institucional

¡A tener en cuenta!

- Toma de decisiones basada realmente en los datos (generación de valor)
- Masificación de la habilidad
- Acceso y uso de la información (regulación, fuentes y calidad)
- Uso del análisis de datos en los procesos
- Tecnología que responda de forma rápida, flexible e integrada a los nuevos tipos y análisis de información
- Seguridad de la información

¿Por qué conocer no es suficiente?



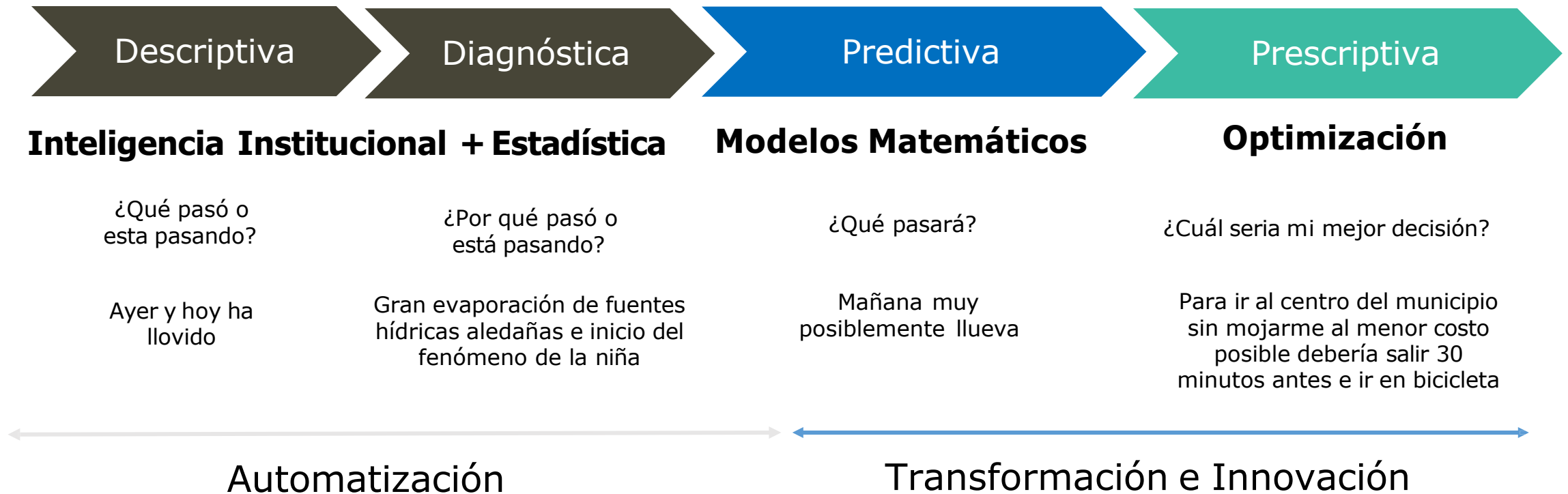
Peter mira esto ...
Señor los números cada vez son
más grandes ...

Yoshi volvimos!!
Vamos, vamos, necesito más
transporte y arboles!

Oye, mira el mercado de pulpa
de árbol ... necesitamos comprar
ya!

Pero al final ... todo era por un
bebé

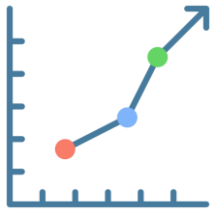
Tipos de analítica



¡Todos lo hacemos a diario!

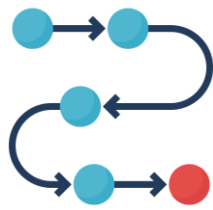
Analítica descriptiva y diagnóstica

Respuesta a preguntas de negocio desde los datos



Datos

Registro de lluvias por día
Registro histórico de
temporadas de lluvias y
fenómeno del niño y la niña



+ Procesamiento =

Cuadro comparativo de
lluvia en la ciudad
Análisis de tendencias,
cruce de información y
correlaciones

**Qué pasó y
esta pasando**

En los últimos 7 días ha
llovido entre las 8 am y
11am

**+ Por qué pasó y
esta pasando**

Debido a una gran
evaporación de fuentes
hídricas cercanas y el inicio
del fenómeno de la niña

Procesos de análisis que dan respuesta del pasado y el hoy, para lo cual se emplean técnicas estadísticas y de inteligencia institucional.



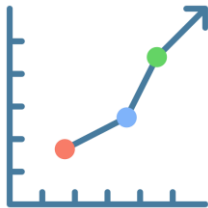
Analítica descriptiva y diagnóstica

¡A tener en cuenta!

- Tipo de Datos
- Estacionalidad
- Causalidad y Correlación
- Propósito, proceso y acción
- Cómo apoya mis procesos
- KPIs
- Gestión y gobierno de los datos
- Retroalimentación a los procesos
- Entendimiento más allá del dato

Analítica predictiva

Los datos históricos al servicio del futuro probable



Datos

Ayer y hoy llovió

f_x

+

Procesamiento

$P(\text{mañana llueva} \mid \text{hoy y ayer llovió})$

=



Futuro Probable

Muy posiblemente llueva
(Probabilidad del 90% con un nivel
de confianza del 95%)

Procesos de análisis que dan respuesta de futuro, para lo cual se emplean técnicas estadísticas, matemáticas y de Inteligencia Artificial.

Analítica predictiva

Tipos de técnicas empleadas

Regresión

Interpolación de los datos a partir del análisis entre la relación de las variables:

- Llenado de valores perdidos.
- ¿Qué ocurrió en cierto momento en el tiempo?

Clasificación

Asignarle un valor a un registro a partir de sus propiedades o características:

- Detección de fraude.
- ¿Qué ciudadanos reaccionarán a una estrategia específica?

Predicción

Identificar un futuro probable a partir de técnicas de análisis de datos:

- Demanda de servicios.
- Probabilidad de ser víctima de hurto en cierta calle.



Analítica predictiva

¡A tener en cuenta!

- Propósito, proceso y acción
- Como se va a vivir en el día a día
- Criterios de aceptación (medidas de confianza y error)
- Entrenamiento y validación
- Modelos y premisas a considerar
- Criterios de validación de reentrenamiento
- Uso en los sistemas transaccionales
- Posibles sesgos en los modelos y los datos

Analítica prescriptiva

Optimización de escenarios probables para seleccionar la “mejor” decisión.



Datos

Ayer y hoy llovió, y mañana posiblemente llueva

f_x

+

Procesamiento

Minimizar el tiempo y costo de desplazamiento sin mojarse

=



Opción óptima

Debería salir 30 minutos antes en bicicleta para llegar al centro de mi ciudad

Procesos de análisis, simulación y optimización que dan respuesta acerca de la “mejor” opción de acuerdo con los criterios (drivers) definidos.



Analítica prescriptiva

¡A tener en cuenta!

- Criterios / drivers de optimización
- Modelo a optimizar
- Meta parámetros
- ¿Cómo impactará el proceso?
- Criterios de aceptación (medidas de confianza y error)
- Algoritmo y premisas a considerar
- Sugerir más no obligar a seleccionar la respuesta óptima



Todo se basa en las preguntas

¿Cuáles son las miradas y preguntas que realizo día a día?

Ejemplo preguntas

Ciudadano

Descriptiva

- ¿Quiénes son mis ciudadanos / empresas?
- ¿Qué los hace únicos?
- ¿Quiénes están debajo de la línea de pobreza?
- ¿A quiénes he apoyado con ayudas económicas los últimos 12 meses?

Predictiva

- ¿Cuál es el nivel de vulnerabilidad de mis ciudadanos?
- ¿Cuál es la probabilidad de embarazo subsecuente en niñas y adolescentes menores de 20 años?
- ¿Cuáles son los factores clave que inciden en el embarazo subsecuente?

Prescriptiva

- ¿Cuál debería ser la oferta institucional para disminuir al menos 5 puntos la vulnerabilidad de nuestros NNAJ?
- ¿Cuál debería ser la ruta de atención que minimiza el riesgo de embarazo adolescente y subsecuente?

Ejemplo preguntas

Ciudadano

Descriptiva

- ¿Cuántos ciudadanos tengo? ¿En que estratos? ¿Dónde están ubicados?
- ¿Cuántas empresas por tamaño operan en mi municipio?
- ¿Cuál es el tamaño de mi parque automotor?

Predictiva

- ¿Cuál es la probabilidad de que ocurra un homicidio / feminicidio en el polígono XXX durante los próximos 7 días?
- ¿Cuál es el estado de salud de un ciudadano?
- ¿Cuál es la probabilidad de que una persona cambie de estado de salud?

Prescriptiva

- ¿Cuáles cursos debería ofrecer para apoyar a mi población?
- ¿Cuál debería ser el protocolo de atención en salud que optimice la recuperación de mi paciente / ciudadano?

Ejemplo preguntas

Ciudadano

Descriptiva

- ¿Cuántos niños y niñas están en riesgo de desnutrición por déficit o exceso?
- ¿Quiénes están en programas de apoyo alimentario?
- ¿Cuántos de ellos han cambiado de estado de nutrición?

Predictiva

- ¿Quiénes están en riesgo de desnutrición y no los conocemos?
- ¿Cuáles son los factores más relevantes en la desnutrición de los NN de mi municipio?
- ¿Cuál es la probabilidad de que un niño mejore su estado de nutrición al incluirlo en un programa específico?

Prescriptiva

- ¿Cuál debería ser la ruta de atención a los niños y niñas en estado de desnutrición aguda o en riesgo de desnutrición por déficit o exceso?
- ¿Cuál sería el impacto de introducir una nueva estrategia de atención a la niñez?

Analítica: implementación

Pasos recomendados en un ciclo de implementación

Necesidad: Pregunta de negocio / Proceso a impactar

Identificación de las fuentes de datos, calidad de información, conceptos y transformaciones (ETL)

Modelación y validación + Mapeos

Reportes/Tableros Base y Habilitación de la capacidad en el negocio

Automatización, puesta en producción + Validación uso y cambios de definición

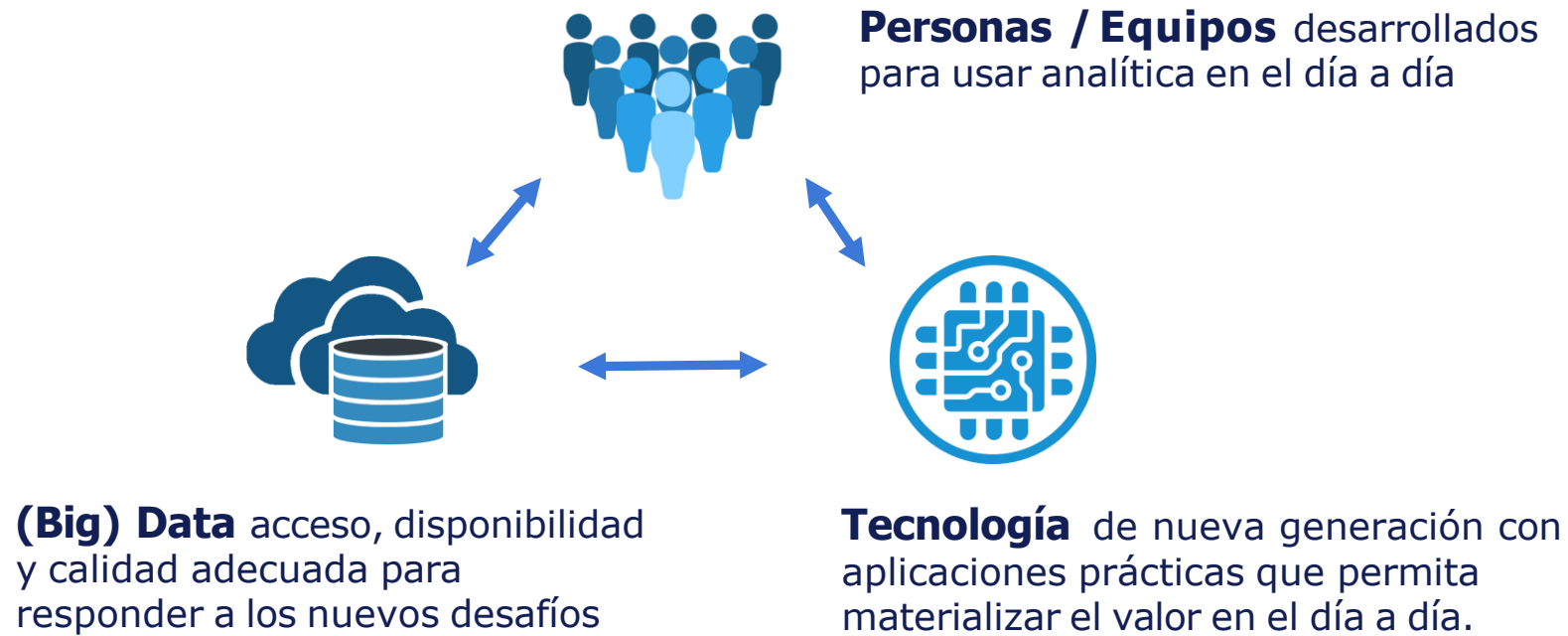


Preguntémonos

https://miro.com/welcomeonboard/TFd1YTJzSEdmTDFGN20za3ozMjJIdDNoaWVUUFMwRFZoYjFlbnNnM3o0ZTJUa1pzbERiaFBxRkN2UWxiMHFyY3wzNDU4NzY0NTIwMjAyNjQ0NTcy?share_link_id=822644236781

¿y como lo logramos..?

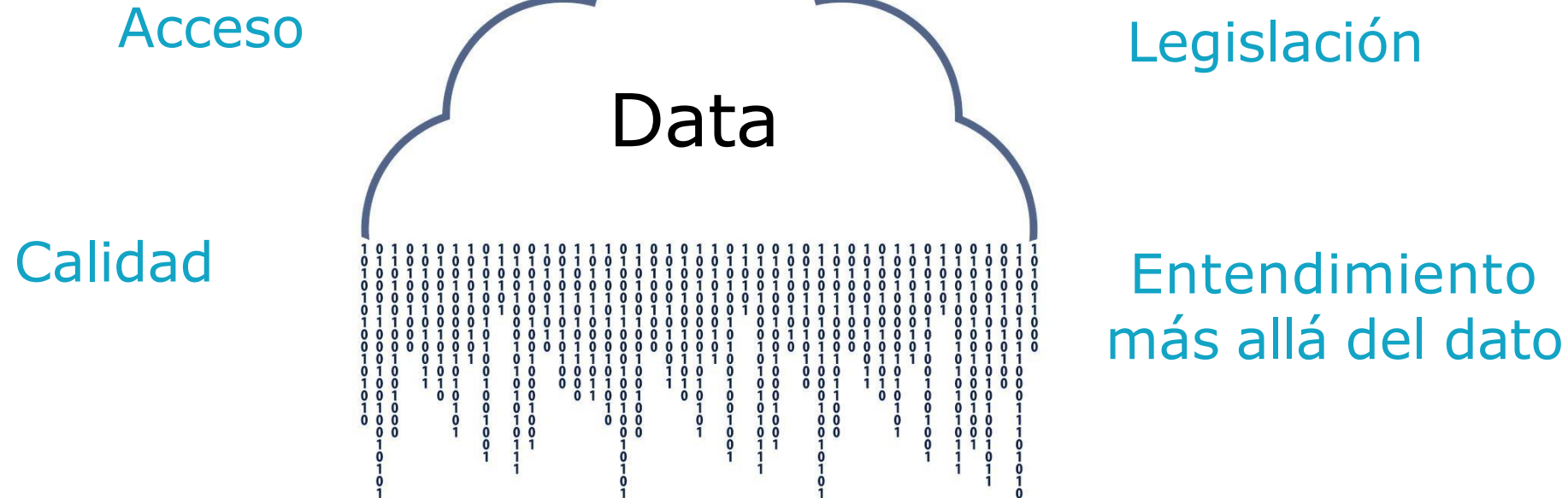
Es necesario...



Interacción y transformación de los procesos

Analítica: datos

OpenData – MDM/MDG – bases de conocimiento –
Uso y Admón.



Analítica: tecnología

Visualización – Integración – Cloud – Realtime



Analítica: equipos

Gestión – Conocimiento – Acciones – Retroalimentación





Analítica

¡A tener en cuenta!

- Confundir estrategia con aplicación.
- Falta de voluntad y entendimiento de la estrategia de implementación por las áreas
- La responsabilidad es de una sola área (TI).
- Implementación solo operativa – solo gerencial



Analítica

¡A tener en cuenta!

- Selección de arquitectura y proveedores solo por el costo.
- Implementaciones Todo o Nada (con alta inversión inicial).
- Seleccionar tecnología y no soluciones



Analítica

¡A tener en cuenta!

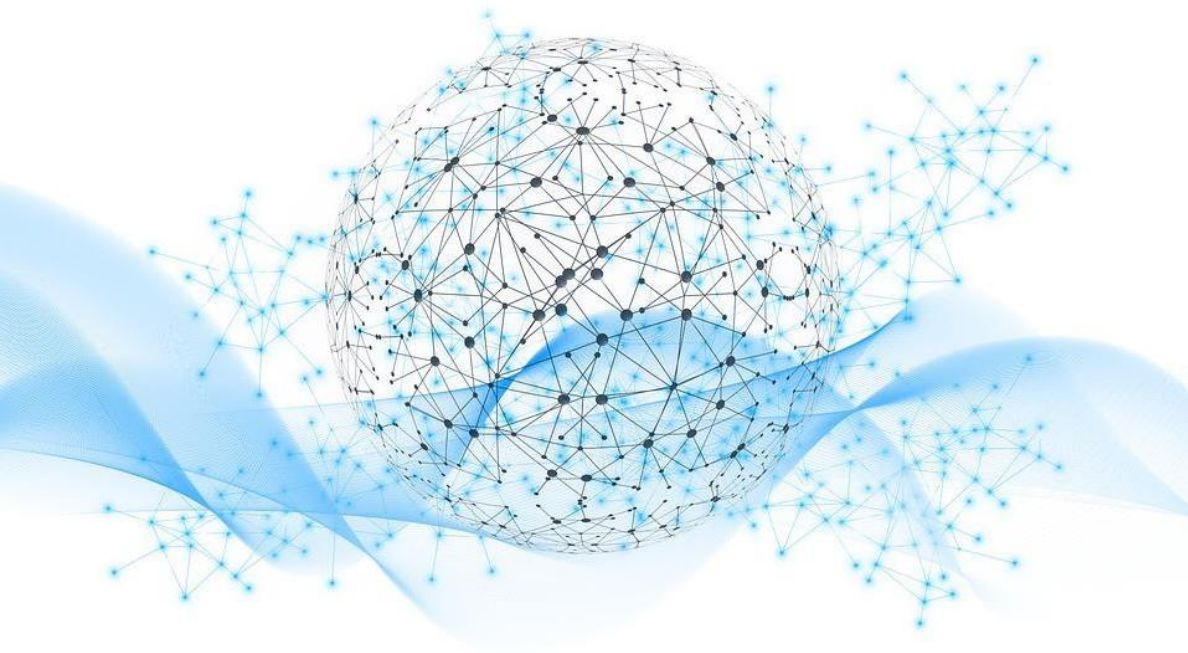
- Calidad y disponibilidad de la información
- Comenzar por el tablero o por la información no por la acción o proceso.
- Falta de una hoja de ruta al interior del área
- No asociar la Analítica a la cultura organizacional



Analítica

¡A tener en cuenta!

- Querer hacerlo todo (falta de democratización y empoderamiento)
- Gestión del cambio.



03. Minería de datos y Big Data

Diferencias, usos y aplicaciones

Analítica – Minería de datos – Big Data

Analítica: Analizar patrones a partir de los datos para generar conocimiento de valor para la toma de decisiones

Minería de datos:
Algoritmos y técnicas de Aprendizaje de Maquina, estadística y otros enfocados en el análisis de patrones y predicción de los mismos

BigData: Técnicas computacionales usadas de forma eficiente en datos masivos y distintos (estructurados y no estructurados) para la toma de decisiones

Minería de datos



Algoritmos y técnicas de Aprendizaje de Maquina, Estadística y otros, enfocados en el análisis de patrones y predicción de los mismos.

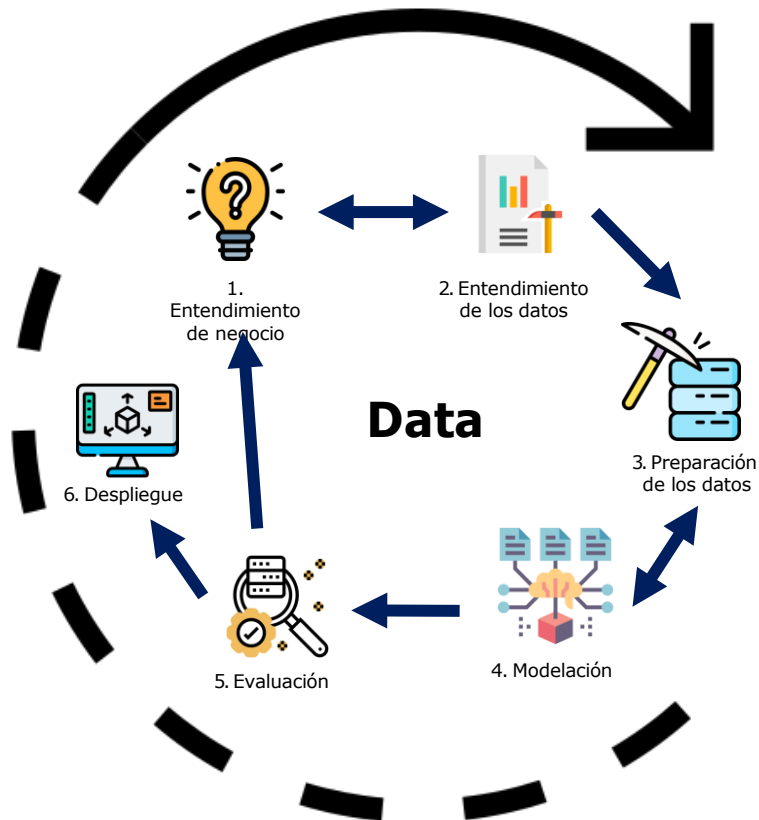
Minería de datos

Metodología CRISP-DM

Cross-industry **s**tandard **p**rocess for **d**ata **m**ining

Estandar abierto para el Desarrollo de proyectos de minería de datos.

1. Entendimiento de negocio
2. Entendimiento de los datos
3. Preparación de los datos
4. Modelación (matemática)
5. Evaluación
6. Despliegue

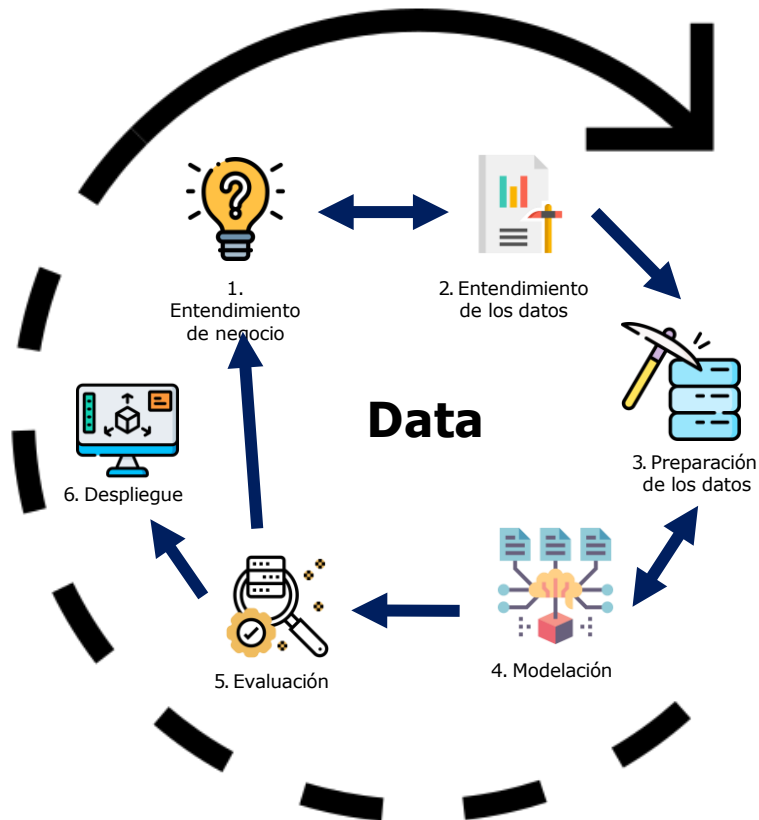


Minería de datos

Metodología CRISP-DM: etapa 1

Etapa 1: Entendimiento del negocio

Esta fase inicial se enfoca en la comprensión de los objetivos y exigencias del proyecto desde una perspectiva de negocio. Posteriormente convierte ese conocimiento de los datos en la definición de un problema de minería de datos y en un plan preliminar diseñado para alcanzar los objetivos.

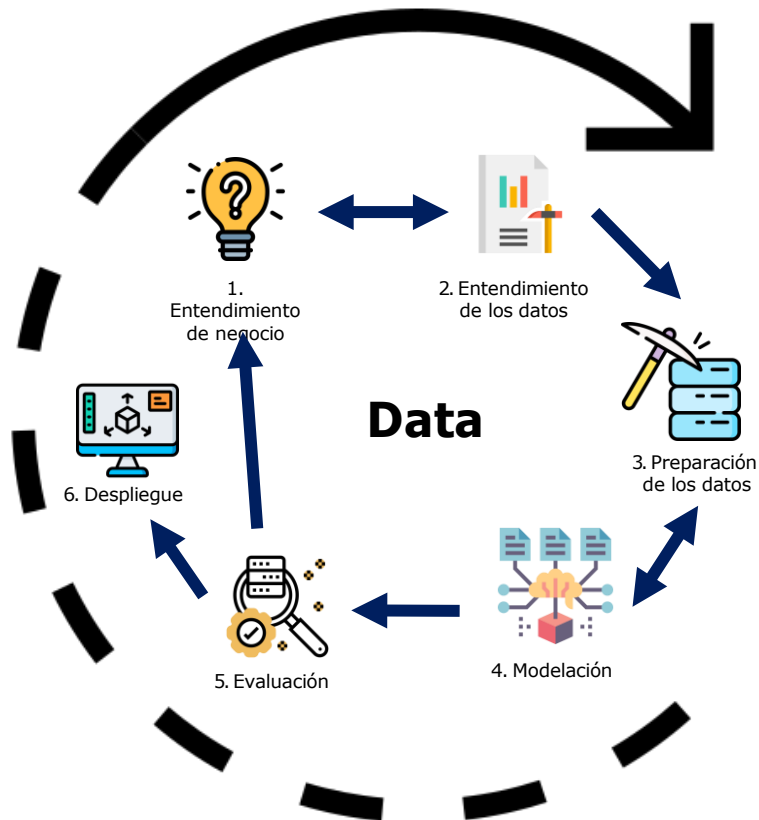


Minería de datos

Metodología CRISP-DM: etapa 2

Etapa 2: Entendimiento de los datos

La comprensión de los datos se encarga de la recolección de datos inicial y continúa con las actividades que permiten familiarizarse primero con los datos, identificar sus problemas de calidad, descubrir conocimiento preliminar en los mismos, y/o descubrir subconjuntos interesantes para formular hipótesis. En esta fase se tienen en cuenta también las fuentes de datos que hasta el momento no se estaban utilizando (fuentes externas, ...).

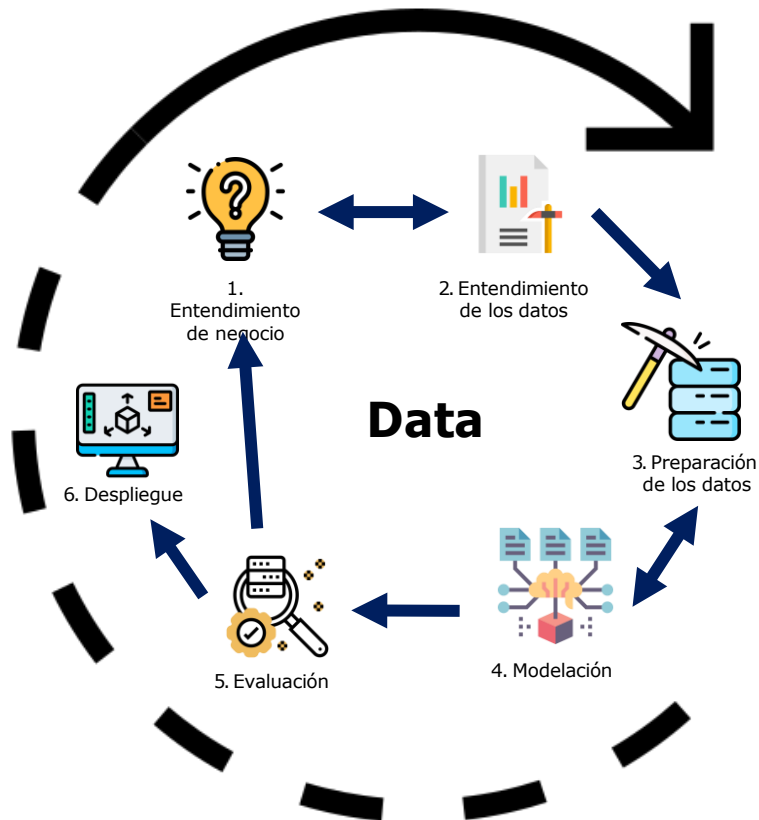


Minería de datos

Metodología CRISP-DM: etapa 3

Etapa 3: Preparación de los datos

La fase de preparación de los datos cubre todas las actividades necesarias para construir el conjunto de datos final (los datos que serán provistos por las herramientas de modelado). Las tareas de preparación incluyen la selección de los datos, la limpieza de éstos, la construcción de nuevas variables, la integración de los datos y el formateo de los mismos.

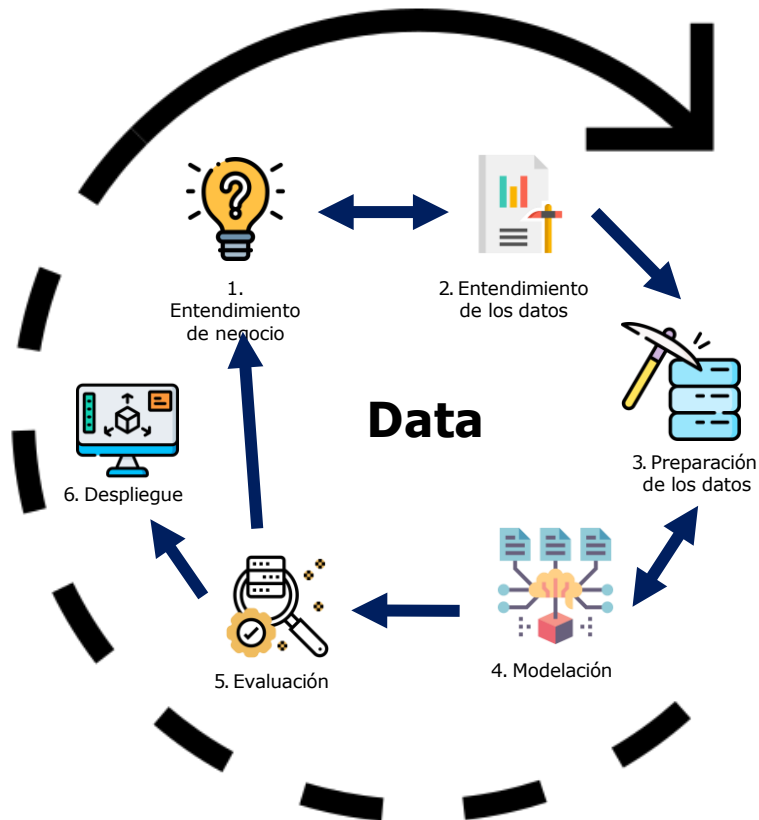


Minería de datos

Metodología CRISP-DM: etapa 4

Etapa 4: Modelado

Durante esta fase, se aplican las técnicas de minería de datos a nuestros datos. Se aplican varias técnicas de modelado y los parámetros de uso de las mismas se afinan hasta alcanzar los valores óptimos. Algunas técnicas de modelado necesitan requerimientos específicos sobre el formato de los datos, que podrán llevarnos de nuevo a la fase de preparación de los datos.

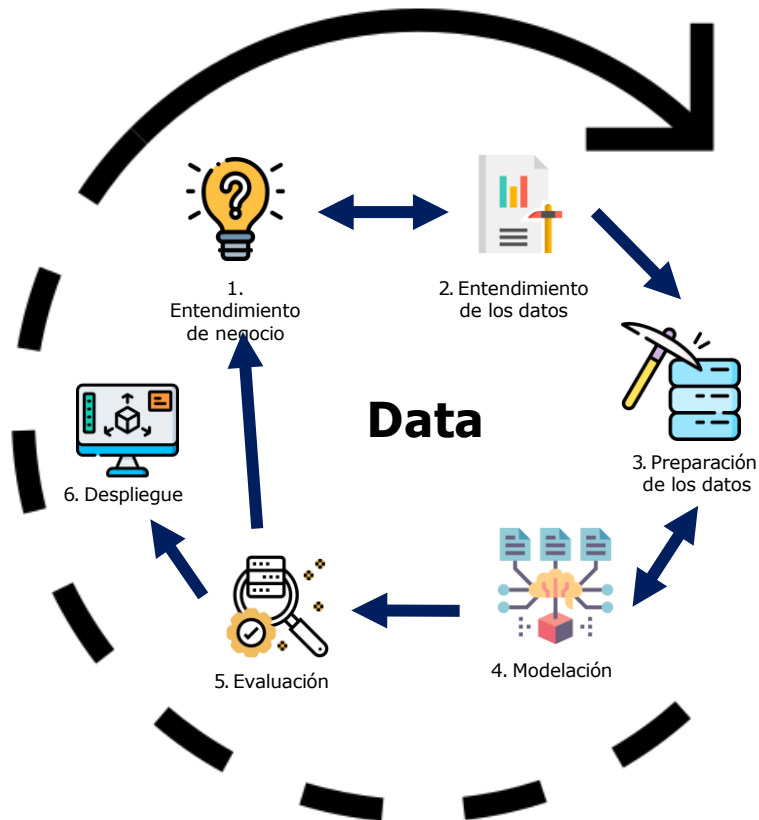


Minería de datos

Metodología CRISP-DM: etapa 5

Etapa 5: Evaluación

En este caso se evalúan los modelos anteriores para determinar si son útiles a las necesidades de negocio. En esta etapa los modelos ya están contruidos y deben tener una alta calidad desde una perspectiva de análisis de datos.

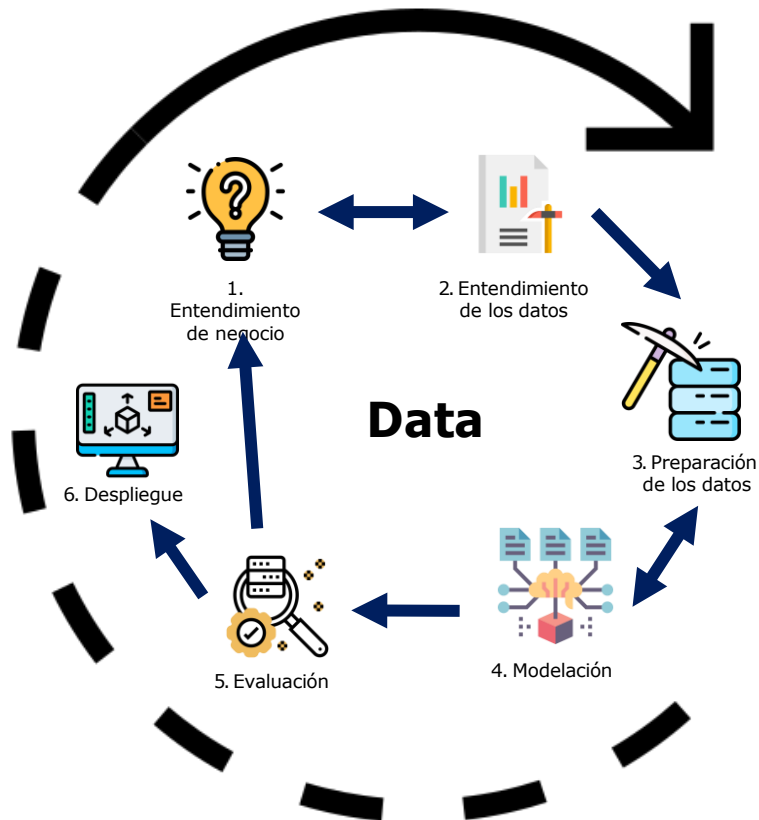


Minería de datos

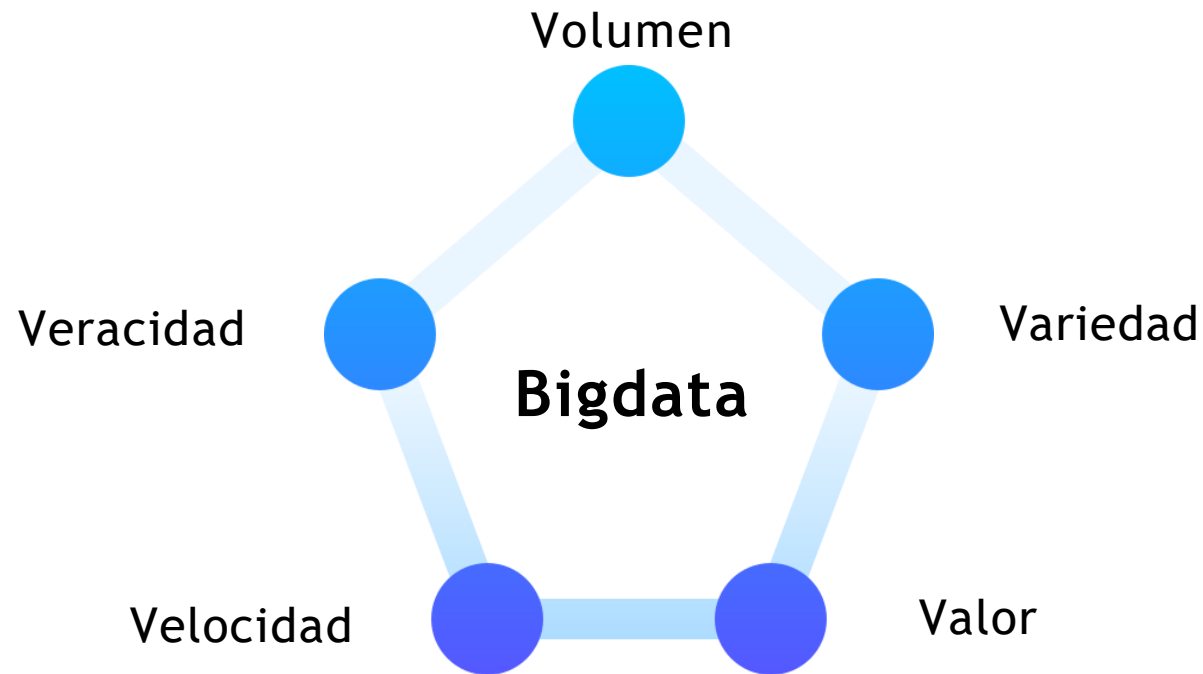
Metodología CRISP-DM: etapa 6

Etapa 6: Despliegue

La fase de despliegue implica la explotación de los modelos dentro de un entorno de producción. La creación de un modelo no es generalmente el final del proyecto, ya que su creación es un proceso vivo dentro del proceso de decisiones de una organización (podría ser necesario rehacer el modelo para tener en cuenta nuevo conocimiento en el futuro).

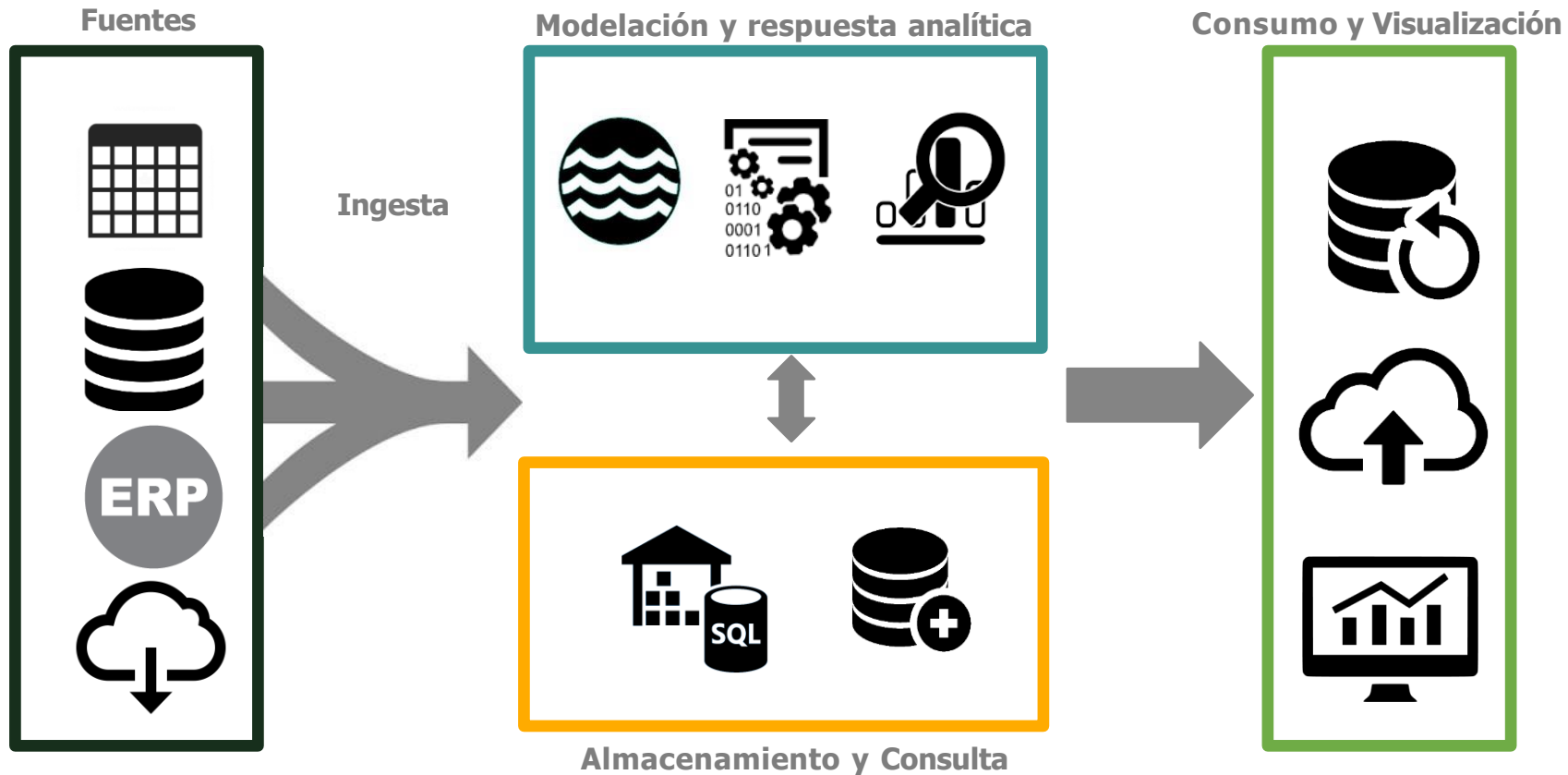


Big Data



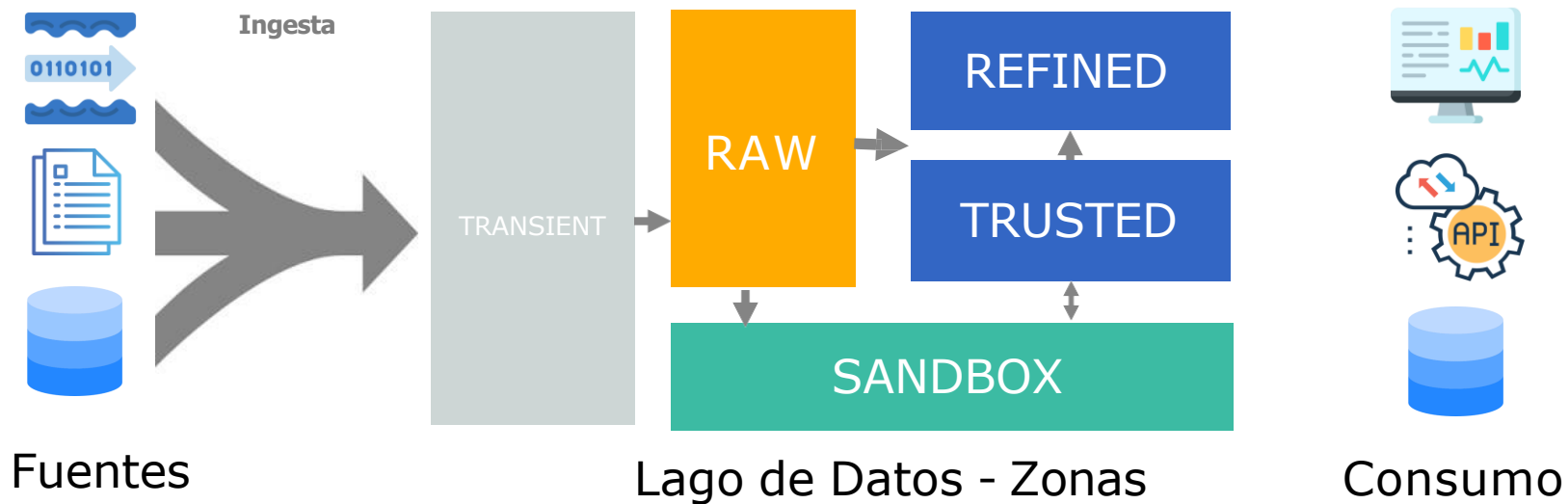
Técnicas computacionales usadas de forma eficiente en datos masivos y distintos (estructurados y no estructurados) para la toma de decisiones

Big Data: arquitectura



Big Data: lago de datos

Flujo de datos



Gobierno: Metadata, Seguridad, Segregación de roles y funciones, Accesos.



Modelar

https://miro.com/welcomeonboard/TFd1YTJzSEdmTDFGN20za3ozMjJIdDNoaWVUUFMwRFZoYjFlbnNnM3o0ZTJUa1pzbERiaFBxRkN2UWxiMHFyY3wzNDU4NzY0NTIwMjAyNjQ0NTcy?share_link_id=822644236781



Validemos conocimientos

Juguemos !!

El municipio SIO está en proceso de apoyo a los campesinos de la región, para lo cual realizó una encuesta para caracterizar las necesidades de apoyo económico, conocimiento e infraestructura en sus veredas. Luego de realizar el estudio surgió la pregunta de ¿Cuál debería ser la necesidad priorizada por la alcaldía? De acuerdo con ello, ¿Cuál es la analítica que respondería a esa pregunta?

Luego de tabular los datos, se requiere entender la distribución de ayudas por cada una de las veredas en los últimos 12 meses, ¿Cuál sería la analítica que da respuesta a esta necesidad? _____

Adicionalmente se identificó que, en una vereda el 80% del total de habitantes se llamaban Juan y adicionalmente requerían apoyo económico, se puede afirmar que existe una alta causalidad entre llamarse Juan y requerir apoyo económico? Si _____No _____

Dados los buenos resultados, el alcalde de SIO decidió implementar Analítica institucional al interior del municipio, para lo cual le solicito al área de tecnología que comprase la infraestructura y programas necesarios.

Dado lo anterior, en el cual solo se está invirtiendo en tecnología y equipos, ¿crees que la estrategia funcionará?
Si _____ No _____

Qué otros componentes faltan por desarrollar en el municipio para lograr tener analítica:

- a) Cloud
- b) Equipos / Personas, Datos
- c) Open data
- d) Infraestructura

Al realizar la implementación, el equipo técnico de la alcaldía hizo un llamado para que se capacitarán a las personas involucradas en realizar la analítica del municipio, adicional a identificar los flujos de datos y definir un adecuado gobierno de datos; sin embargo, una vez realizado lo anterior, sin previo aviso se envió un correo a toda la alcaldía indicando que desde ese momento no se iba a utilizar el proceso actual en físico si no que todo se virtualizaba en la nueva herramienta SIO-Data con un pdf con vagas instrucciones. Dado esto, que crees que faltó en el proceso de lanzamiento de SIO-Data:

- a) Gestión del cambio
- b) Mas celulares
- c) Un nombre más novedoso
- d) Una salida institucional

Luego del lanzamiento, se construyeron varios tableros de indicadores los cuales fueron enviados por medio de un link por correo institucional a los servidores de la alcaldía. ¿Qué crees que debería hacer la alcaldía con este tablero?

- a) Monitorear los resultados constantemente y tomar acciones con base en ello
- b) Archivar el correo
- c) Revisarlo cada año al momento de reportar resultados en el FURAG.
- d) Ninguna de las anteriores

Luego de un tiempo, en la reunión semanal de gobierno, las personas indicaron que los datos del tablero estaban erróneos, generando discusiones y falta de claridad en las cifras. ¿Cómo crees que se pudo haber evitado esta situación?

- a) Enviando correos diarios a los equipos.
- b) Confiando ciegamente en los datos del tablero.
- c) Validando periódicamente la validez de los resultados y conceptos de negocio, realizando los cambios pertinentes para alinearse a la realidad de la alcaldía.
- d) Solicitando ayuda a un municipio vecino.

Luego de estos resultados el municipio quisiera determinar quiénes de los ciudadanos podrían necesitar acompañamiento en salud, para lo cual ellos deberían:

- a) Solicitar apoyo de otros municipios
- b) Realizar un modelo de analítica predictiva, identificando los posibles ciudadanos susceptibles de acompañamiento en salud.
- c) Generación de un tablero de control con todos los ciudadanos del municipio
- d) Ninguna de las anteriores

Luego de todo esto, las personas involucradas en este proyecto se fueron del municipio y el proyecto quedo estancado debido a que:

- a) No realizaron una adecuada gestión del conocimiento
- b) Ya había pasado un año
- c) No se deseaba seguir cosechando excelentes resultados
- d) Ninguna de las anteriores



04. Analítica institucional

Conclusiones

Analítica institucional

Profundizar en los análisis de la información y los datos que genera la entidad para fortalecer el conocimiento de su desempeño y de su propósito fundamental



Diagnóstico de datos:

Identificación, organización y actualización de los datos producidos por la entidad



Tableros de indicadores:

Generación de indicadores y monitoreo de la gestión de la entidad



Análisis de datos:

Profundizar en los análisis de datos internos y externos, haciendo cruces que permitan extraer nuevas perspectivas de los datos y de la información

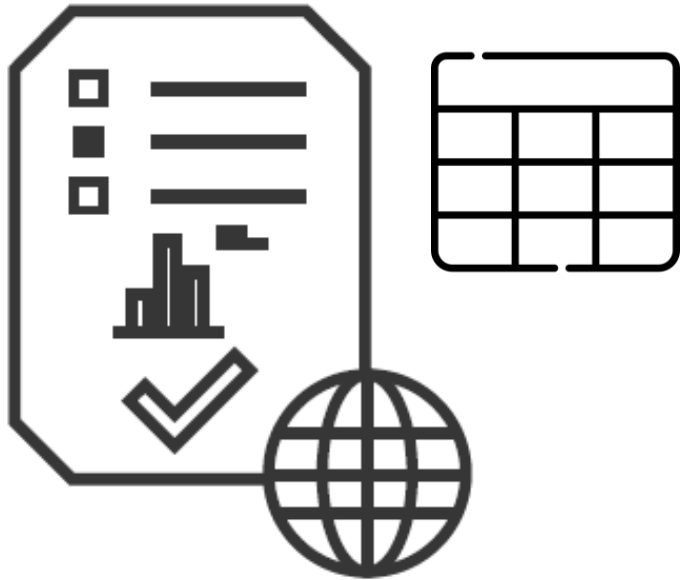


Visualizaciones:

Visualización de datos para formular preguntas diferentes y contar historias a través de los datos

Analítica institucional

Diagnóstico de datos



Identificación, organización y actualización de los datos producidos por la entidad

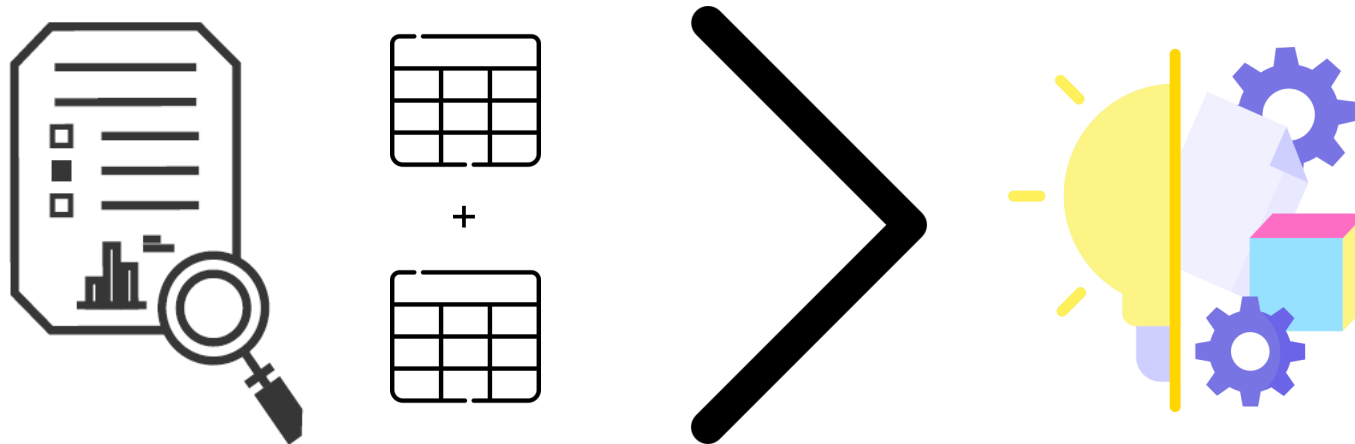
- Datos
- Tipo de Dato (privacidad y almacenamiento)
- Calidad de los datos
- Uso y propósito
- Flujo de la información

Gobierno de los Datos: Roles, responsabilidades, identificación de los datos de la entidad y estrategia de registro y actualización.

Analítica institucional

Análisis de datos

Profundizar en los análisis de datos internos y externos, haciendo cruces que permitan extraer nuevas perspectivas de los datos y de la información



Analítica Descriptiva: Conocer que esta sucediendo en nuestro territorio, entidad, con nuestro ciudadano / empresa

Analítica institucional

Tableros de indicadores



Generación de indicadores y monitoreo de la gestión de la entidad

- Indicadores SMART
- Tableros de gestión operativos
- Tableros Gerenciales que apoye la toma de decisiones

Analítica Descriptiva: Generación de KPIs y Tableros para la gestión diaria

Analítica institucional

Visualizaciones

Visualización de datos para formular preguntas diferentes y contar historias a través de los datos

- ¿Qué sucede y puede suceder?
- ¿Cuáles decisiones son las más adecuadas?
- Preguntas más allá del hoy

Analítica Descriptiva + Predictiva: Identificación de comportamientos y toma de decisiones basado en ellos





CONCLUSIONES

- Es necesario considerar la analítica como parte de la estrategia y filosofía de trabajo en la entidad.
- La analítica no es ciencia de cohetes, todos lo hacemos, lo importante es hacerlo en el día a día.
- El conocimiento debe replicarse, generar acciones y ser un motor movilizador.
- Es relevante contar con una hoja de ruta de implementación de Analítica (descriptiva, diagnóstica, predictiva y prescriptiva) que permita sumar esfuerzos.