

Cartilla
**Guía para la identificación de soluciones
analíticas**

Proceso Información, Innovación y tecnología
Subproceso Gobernanza de datos, Analítica y Estudios
Económicos
Versión 2
Código CT-IIT-0120
Año 2024

El contenido de este documento corresponde a Información Pública

TABLA DE CONTENIDO

Introducción	3
1. Objetivo	4
2. Siglas y definiciones	4
3. Contenido	6
3.1 Conceptos básicos de analítica.	¡Error! Marcador no definido.
3.2 Asociación de las soluciones de analítica con los objetivos de negocio.....	6
3.3 Antecedentes y datos disponibles.....	10
3.4 Preguntas de análisis para identificar necesidades de analítica.....	11
3.5 Referencias.	11

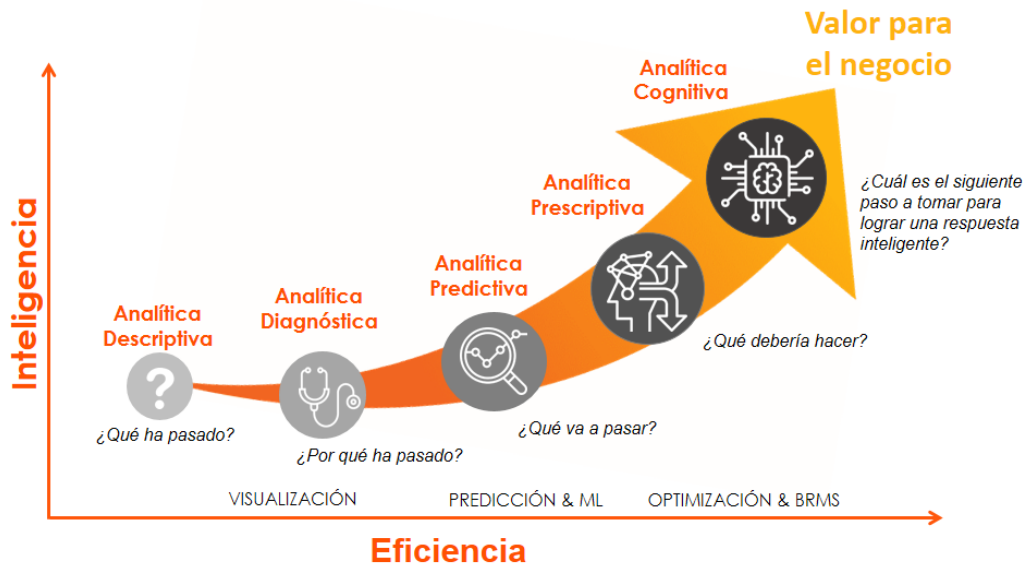
Introducción

Business Analytics (BA) es la práctica y el arte de utilizar datos cuantitativos para la toma de decisiones, que incluye una variedad de métodos de análisis de datos, que implican poco más que contar, verificar reglas y aritmética básica. El siguiente nivel de análisis de negocios, ahora denominado Inteligencia de Negocios (BI, por sus siglas en inglés), se refiere a la visualización de datos e informes para comprender "qué sucedió y qué está sucediendo". Esto se realiza mediante el uso de gráficos, tablas y paneles para mostrar, examinar y explorar datos.

La inteligencia de negocio ha evolucionado hacia herramientas y prácticas más amigables y efectivas, como la creación de paneles interactivos que permiten al usuario no solo acceder a datos en tiempo real, sino también interactuar directamente con ellos. Los tableros efectivos son aquellos que se conectan directamente a los datos y brindan a los líderes una herramienta para ver rápidamente lo que podría no ser evidente de inmediato en una base de datos grande y compleja. Adaptado(Shmueli, 2018)

El Análisis de Negocios ahora típicamente incluye BI, así como métodos sofisticados de análisis de datos, como modelos estadísticos y algoritmos de minería de datos utilizados para explorar datos, cuantificar y explicar relaciones entre mediciones, y predecir nuevos datos.

En función de las técnicas utilizadas y el estado de madurez del proceso en materia de análisis de información, el análisis de datos puede clasificarse en las siguientes categorías:



Tomado de: *Decide soluciones, Infografía: tipos de Analítica Avanzada de datos y su valor para el negocio (2019)* [Online]. <https://medium.com/@Decide/infograf%C3%ADa-tipos-de-anal%C3%ADtica-avanzada-de-datos-y-su-valor-para-el-negocio-cbbcdab6c347>.

Para información complementaria relacionada con los tipos de algoritmos de analítica, se pueden consultar recursos en la web. Algunas lecturas recomendadas son las siguientes:

- Algoritmos de aprendizaje automático: <https://azure.microsoft.com/es-es/overview/machine-learning-algorithms/#techniques>.
- Principales algoritmos usados en machine learning:

<https://www.aprendemachinelearning.com/principales-algoritmos-usados-en-machine-learning/>.

1. Objetivo

Facilitar el entendimiento de conceptos básicos de analítica de datos por parte los funcionarios de la UAE DIAN, para orientar y dar criterios de análisis al momento de generar una solicitud a la Subdirección de Información y Analítica. Dicha solicitud debe presentarse según lo establecido en el procedimiento PR-IIT-0472, teniendo en cuenta el enfoque del negocio, orientado al cumplimiento de los objetivos estratégicos y la mejora de los procesos.

2. Siglas y definiciones

- **Algoritmo:** es un conjunto de órdenes que establecen un proceso lógico para ejecutar actividades y encontrar una solución a uno o varios problemas. (next_u, 2022). Todo algoritmo tiene una entrada, conocida como *input* y una salida, conocida como *output*, y entre medias, están las instrucciones o secuencia de pasos a seguir. Estos pasos deben estar ordenados y, sobre todo, deben ser una serie finita de operaciones que permitan conseguir una determinada solución.
Fuente: UAE DIAN – Subdirección de Información y Analítica adaptado (Ferrovial, 2023) .
- **Analítica descriptiva:** este tipo de análisis se utiliza para analizar y describir las características de los datos con un objetivo establecido.
Fuente: UAE DIAN – Subdirección de Información y Analítica.
- **Analítica diagnóstica:** en todas las entidades u organizaciones pueden ocurrir eventos, tanto internos como externos, que impactan de forma negativa o positiva el cumplimiento de sus objetivos. El propósito del análisis diagnóstico es identificar las posibles causas de dichos eventos y entender lo que está sucediendo, con el fin de fortalecer la toma de decisiones, bien sea para solucionar las dificultades o aprovechar las oportunidades que se presentan.
Fuente: UAE DIAN – Subdirección de Información y Analítica.
- **Analítica predictiva:** como su nombre indica, este tipo de análisis se utiliza para predecir resultados futuros. Basándose en datos pasados o históricos, se determina un conjunto de resultados probables para el futuro. En este tipo de análisis, se necesitan grandes volúmenes de datos que se extraen con habilidad utilizando modelos predictivos. Esto garantiza que se puedan hacer predicciones precisas.
Fuente: UAE DIAN – Subdirección de Información y Analítica.
- **Analítica prescriptiva:** es una combinación de todos los demás tipos de análisis. En lugar de centrarse en la supervisión de los datos, esta técnica hace hincapié en las perspectivas de acción. Los datos se recogen utilizando modelos descriptivos y predictivos, así como una combinación de matemáticas e informática.
Fuente: UAE DIAN – Subdirección de Información y Analítica.
- **Analítica cognitiva:** se trata de la más alta evolución de la analítica y recoge un amplio número de técnicas comúnmente englobadas dentro de la inteligencia artificial. Apoyándose en estas técnicas, se desarrollan soluciones capaces de entender, aprender y reconocer patrones escondidos en

datos en muy diversos formatos (ejemplo generación de lenguaje natural, movilidad autónoma, robótica, etc.).

Fuente: UAE DIAN – Subdirección de Información y Analítica.

- **Base de datos:** es un conjunto de datos no redundantes, almacenados en un soporte informático, organizados en forma independiente de su utilización y accesibles simultáneamente por distintos usuarios y aplicaciones. Además de organizar la información de una forma lógica, se encargan de modelar un escenario real subyacente a los datos, incorporando reglas y detalles. De esta forma, constituyen un soporte a los datos que atraviesan uno o varios procesos dentro de una empresa, considerando, además, múltiples usuarios y aplicaciones.
Fuente: UAE DIAN – Subdirección de Información y Analítica.
- **Información estructurada:** se refiere a aquella que está definida y sujeta a un formato concreto que facilita su procesamiento. Por ejemplo, la información organizada y estructurada en bases de datos relacionales u hojas de cálculo se considera estructurada.
Fuente: Extraídos de MINTIC- MGGTI.G.GEN.01 – Documento Maestro del Modelo de Gestión y Gobierno de TI, 2019.
- **Información no estructurada:** es aquella que no posee una estructura predefinida, no están organizada de acuerdo con algún patrón. Entre esta se encuentra la información de tipo multimedia (video, voz, imagen), información generada en las redes sociales, foros, e-mails, presentaciones Power Point o documentos Word.
Fuente: Extraídos de MINTIC- MGGTI.G.GEN.01 – Documento Maestro del Modelo de Gestión y Gobierno de TI, 2019.
- **Inteligencia artificial:** disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico.
Fuente: Real academia española, Diccionario de la lengua española | inteligencia | Definición. Recuperado de: <https://dle.rae.es/inteligencia>.
- **Machine learning o aprendizaje automático:** el *Machine Learning* es una disciplina del campo de la Inteligencia Artificial que, a través de algoritmos, dota a los ordenadores de la capacidad de identificar patrones en datos masivos y elaborar predicciones (análisis predictivo). Este aprendizaje permite a los computadores realizar tareas específicas de forma autónoma, es decir, sin necesidad de ser programados.
Fuente: Iberdrola. Machine Learning, Definiciones, tipos y aplicaciones prácticas. Recuperado de: <https://www.iberdrola.com/innovacion/machine-learning-aprendizaje-automatizo>.
- **Modelo analítico:** un algoritmo aplicado a un conjunto de datos o una representación matemática de un proceso o procedimiento del mundo real para resolver un problema y obtener una solución óptima. Los modelos analíticos tienen como resultado un valor, indicador o una respuesta al ser aplicados sobre nuevos datos. Se espera que este resultado se acerque lo mayor posible al resultado real.
Fuente: (Consulting, 2020)
- **Dashboard/Tablero de control:** la inteligencia de negocios combina principios de gestión y herramientas analíticas para mejorar la toma de decisiones en organizaciones mediante la

optimización y el análisis de datos empresariales. Este enfoque busca incrementar la eficiencia operativa y apoyar la toma de decisiones.

Fuente: Rouhani, S., Asgari, S., & Mirhosseini, S. V. (2012). Review study: Business intelligence concepts and approaches. American Journal of Scientific Research, 50(1), 62-75.

3. Contenido.

3.1 Metodologías aplicables de analítica

La elección de los métodos o aplicaciones de analítica depende del tipo de datos, los objetivos del análisis y el contexto. Puede hacer uso de las herramientas de analítica en el momento en que requiera tomar decisiones que contribuyan al cumplimiento de los objetivos estratégicos.

Posibilitando una mejor comprensión de los sesgos, preferencias y mecanismos que influyen en la toma de decisiones de los contribuyentes, la economía del comportamiento en la DIAN es una poderosa herramienta para la analítica de datos, que se basa mayoritariamente en la utilización de métodos experimentales y cuasi-experimentales para la obtención de información sobre el contribuyente a través de evaluaciones de impacto. Para mayor referencia sobre guía de procedimientos, consultar frameworks [BASIC + SMART](#).

A continuación, se presentan algunas de las herramientas analíticas para resolver problemas identificados en los procesos, estas herramientas que faciliten el entendimiento, (Shmueli, 2018) preparación, procesamiento y visualización de datos y permiten identificar patrones, tendencias y relaciones en los datos, para la generación de información de valor.

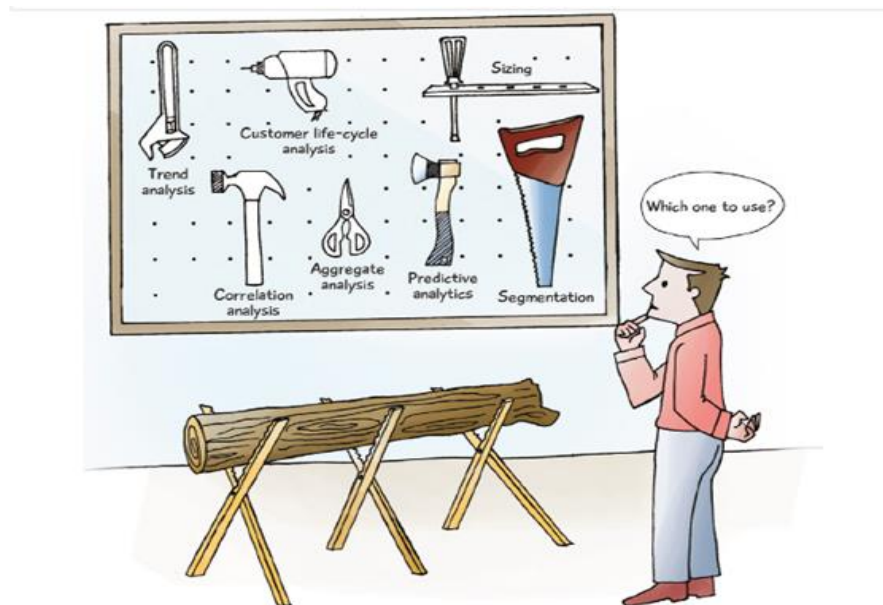


Ilustración 1. Tomada de <https://learning-oreilly-com.ezproxy.uniandes.edu.co/library/view/behind-every-good/9780814449219/xhtmll/Chapter03.html#ch3>.

METODOLOGÍA	DESCRIPCIÓN	APLICACIÓN
Análisis descriptivo agregado	Consiste en describir las tendencias claves en los datos existentes y observar las situaciones que conduzcan a la creación de hipótesis o conclusiones generales.	Análisis descriptivo, perfilado y análisis de campañas.
Análisis de Componentes Principales (PCA)	Es una técnica de reducción de dimensionalidad que se utiliza para simplificar conjuntos de datos complejos identificando las características más importantes y representativas.	Simplifica la base de datos, ya sea para elegir un menor número de variables para pronosticar una variable objetivo, o para comprender una BDD de una forma más simple. La técnica PCA forma parte de los algoritmos de aprendizaje no supervisado de Machine Learning, y se utilizaría solo con aquellas variables numéricas disponibles. <i>Fuente:</i> https://rpubs.com/Cristina_Gil/PCA
Análisis de correlación	Muestra todas las correlaciones por pares entre variables. Las correlaciones fuertes (positivas o negativas) contienen mucha superposición en la información, son buenos candidatos para la reducción de datos eliminando una de las variables.	Una forma sencilla de encontrar redundancias es observar una matriz de correlación.
Análisis de Clúster	A partir de un conjunto de variables, se busca crear grupos de casos que compartan características comunes. Es un proceso iterativo en el que no existen soluciones únicas (no supervisado).	Segmentación de clientes en grupos, determinar las distintas variables, agrupar temas, descubrir zonas con ciertas características.
Análisis de Redes	Se utiliza para analizar relaciones y conexiones en una red, como redes sociales, redes de colaboración científica, etc.	Puede revelar nodos influyentes, comunidades y patrones de interacción.
Análisis de Regresión	En el análisis de regresión, se tiene un número de variables predictoras (explicativas) y una variable de respuesta continua (resultado o destino), y se tiene que encontrar una relación entre estas variables que permita predecir un resultado. <i>Fuente:</i> (Raschka & Mirjalili, 2019)	Tipo de aprendizaje supervisado para predicción de resultados continuos. <i>Fuente:</i> (Raschka & Mirjalili, 2019)
Análisis de Sentimiento	Utilizado en el procesamiento de lenguaje natural, busca determinar el sentimiento o la opinión general expresada en un conjunto de texto, generalmente en comentarios o publicaciones en redes sociales.	Conocer cómo responden las personas: detectar el enfado o satisfacción de los usuarios que se comunican con la Entidad, cómo valoran los

METODOLOGÍA	DESCRIPCIÓN	APLICACIÓN
		servicios, si están contentos o descontentos, entre otros.
Análisis de tendencias	El análisis de tendencias se define como una técnica estadística y analítica utilizada para evaluar e identificar patrones, tendencias o cambios en los datos a lo largo del tiempo. Consiste en examinar datos históricos para descubrir la dirección o las tendencias de un proceso o procedimiento en específico.	Datos de series temporales, visualización de datos, identificación de patrones, tendencias o ciclos recurrentes, análisis de regresión, extrapolación y predicción.
Aprendizaje Automático (Machine Learning)	Machine learning es una rama de la inteligencia artificial (IA) y la informática que se centra en el uso de datos y algoritmos para imitar la forma en la que aprenden los seres humanos, con una mejora gradual de su precisión. https://www.ibm.com/es-es/topics/machine-learning	El aprendizaje automático ofrece una alternativa más eficiente para capturar el conocimiento en datos, mejorar el rendimiento de los modelos predictivos y tomar las decisiones en datos.
Análisis predictivo/Series temporales	Analiza datos secuenciales en el tiempo para identificar patrones, tendencias y estacionalidades. Es útil en la predicción de valores futuros en función de datos pasados.	Es útil en la predicción de valores futuros en función de datos pasados.
Aprendizaje Profundo (Deep Learning):	Es una rama del aprendizaje automático que se basa en redes neuronales artificiales para analizar y comprender datos complejos, como imágenes y texto.	Traducción automática, redes neuronales para la identificación de imágenes que tienen letras visibles, detección de fraude, reconocimiento visual, generación automática de escritura a mano.
Economía del comportamiento	Posibilitando una mejor comprensión de los sesgos, preferencias y mecanismos que influyen en la toma de decisiones de los contribuyentes, la economía del comportamiento es una poderosa herramienta para la analítica de datos.	Diseñar productos y servicios que permitan tomar mejores decisiones usando información sobre los factores que afectan el comportamiento y las decisiones de los contribuyentes. Facilitar el perfilamiento de riesgos y comportamientos del contribuyente. Implementar estrategias de comunicación más efectivas, y que permitan una evaluación de impacto adecuada.
Tablero de control	Es una herramienta visual que presenta información clave para la toma de decisiones de manera accesible y oportuna. Presenta métricas relevantes para el negocio, permitiendo analizar y responder a cambios en las operaciones o cualquier otro aspecto crítico de la organización. Su diseño	Al enfrentarse a requerimientos analíticos, optar por un tablero de control es ideal cuando se busca sintetizar datos de múltiples fuentes en hallazgos accionables, facilitando la toma de decisiones y el

METODOLOGÍA	DESCRIPCIÓN	APLICACIÓN
	personalizable facilita el seguimiento de objetivos y el alcance de metas estratégicas establecidas.	seguimiento a los procesos. Estas herramientas son esenciales para visualizar tendencias, patrones y métricas clave, permitiendo una gestión y estrategia basadas en datos eficaces y ágiles.

Fuente: Subdirección de información y Analítica.

3.2 Asociación de las soluciones de analítica con los objetivos de negocio.

Generalmente la necesidad de aplicación de analítica puede originarse por alguno de los siguientes factores:

- Recursos disponibles en la organización para nuevos proyectos (datos, recursos de personal, tecnológicos o financieros, capacidad de conocimiento).
- Problemas, dificultades o alta complejidad en la operación de los procesos y/o cumplimiento de objetivos.

Para realizar una solicitud se debe aplicar lo establecido en el procedimiento PR-IIT-0472 Gestión de Solicitudes y Proyectos de Analítica de Datos.

3.3 Antecedentes y datos disponibles [3].

✓ Describir la solución actual.

Para facilitar el análisis y el entendimiento por parte de otras dependencias, se deben describir las **soluciones actuales** que se ejecutan en la entidad para resolver el problema o dar cumplimiento al objetivo. En el mismo sentido, se sugiere analizar las **ventajas y desventajas de la solución actual**, tanto desde el punto de vista operativo como en el cumplimiento (o incumplimiento) de objetivos (ejemplo: nivel de recaudo actual, cantidad de contribuyentes o cobertura de control, etc.); este análisis permitirá encontrar buenos argumentos para justificar la necesidad de analítica, así como también puede orientar mejor la solicitud y favorecer el diseño de la solución.

✓ Antecedentes de análisis de datos.

Se sugiere hacer un inventario de análisis cuantitativos de datos (tablas dinámicas, gráficas de tendencia, histogramas, etc.) que se hayan efectuado en el pasado sobre el mismo conjunto de datos o sobre datos relacionados con la necesidad de negocio a responder; igualmente se sugiere identificar los hallazgos, tendencias o conclusiones resultantes de dichos análisis.

✓ Diagnóstico de datos disponibles.

Para que un requerimiento de analítica de datos sea viable, debe contarse con un conjunto de datos recolectados, los cuales pueden ser estructurados o no estructurados. Las siguientes preguntas de análisis pueden ser útiles al momento de evaluar los datos disponibles.

- ¿Qué datos están disponibles para el análisis? ¿Son datos que corresponden a información estructurada o no estructurada?
- ¿Qué tipo de orígenes de datos están disponibles para el análisis? ¿Cómo se almacenan los datos?
- ¿La Subdirección de información y analítica dispone de acceso directo a los almacenes de datos o a las bases de datos fuente?
- ¿Los datos están completos (sin columnas o registros faltantes o nulos)? ¿Los datos son confiables/correctos/exactos?
- ¿Es necesario adquirir o utilizar datos externos?

Este análisis permitirá identificar si la solicitud de una solución analítica es viable, o si es necesario hacer algún alistamiento de datos o gestionar el acceso a los mismos, entre otras posibles acciones, previas al desarrollo de la solicitud.

3.4 Preguntas de análisis para identificar necesidades de analítica.

Una vez identificados los conceptos básicos, la necesidad de negocio y los antecedentes en términos de análisis de datos, se recomienda plantearse las siguientes preguntas con el fin de identificar el o los tipos de soluciones analíticas que serían más apropiadas para dar respuesta a sus requerimientos.

1. ¿El producto o salida principal corresponde a un listado o tabla de datos?
2. ¿Se espera que el producto incluya conclusiones, patrones o tendencias identificados por el equipo de analítica?
3. ¿Se espera que el producto permita a los dueños de proceso visualizar y analizar datos para generar conclusiones?
4. ¿Se espera que el producto permita tomar decisiones en tiempo real durante la ejecución de un proceso?
5. ¿Se requiere que el producto de información esté disponible para los funcionarios que ejecutan el proceso?
6. ¿Se requiere que el producto de información se genere de forma periódica (semanal, mensual, anual, etc.)?

Las conclusiones del anterior análisis serán muy útiles para dar un mejor entendimiento a la Subdirección de información y analítica sobre la necesidad de negocio a la que se debe dar respuesta.

3.5 Referencias.

- [1] Decide soluciones, *Infografía: tipos de Analítica Avanzada de datos y su valor para el negocio* (2019) [Online] Disponible: <https://medium.com/@Decide/infograf%C3%ADa-tipos-de-anal%C3%ADtica-avanzada-de-datos-y-su-valor-para-el-negocio-cbbcdab6c347>.
- [2] Plan estratégico institucional DIAN 2020 – 2022.
- [3] IBM Corporation (1994-2012). Manual CRISP-DM de IBM SPSS Modeler (p. 4-12).
- [4] (OpenAI, 2023)
- [5] <https://learning-oreilly-com.ezproxy.uniandes.edu.co/library/view/behind-every-good/9780814449219/xhtml/Chapter03.html#ch3>.

4. CONTROL DE CAMBIOS

Versión	Vigencia		Descripción de los cambios	Tipo de información
	Desde	Hasta		
1	09/02/2022	23/04/2024	Versión inicial	Esta versión corresponde a Información Pública
2	24/04/2024		Se incorporaron algunas herramientas analíticas.	Esta versión corresponde a Información Pública

Elaboró:	Yuri Angélica Reina Elaboración Técnica	Gestor II	Subdirección de Información y Analítica
	Tito Alejandro Menjura Elaboración Metodológica	Gestor II	Coordinación de Procesos y Riesgos Operacionales
Revisó:	María Fernanda Osorio Moreno	Subdirectora	Subdirección de Información y Analítica
Aprobó:	María Fernanda Osorio Moreno	Subdirectora	Dirección de Gestión Estratégica y de Analítica