



UNIVERSIDAD NACIONAL AGRARIA  
**LA MOLINA**  
FACULTAD DE ECONOMÍA Y PLANIFICACIÓN  
UNIDAD DE EXTENSIÓN Y PROYECCIÓN SOCIAL  
Ingeniería en Gestión Empresarial, Economía y Estadística Informática

En convenio con



# BIG DATA & DATA SCIENCE

PROGRAMA DE ESPECIALIZACIÓN



# Business Data Discovery

Empezar con el negocio





# Procesos de Comprensión Analítica



# Temas



Procesos de Comprensión  
Analítica



Entendimiento de Negocio



Recolección de Datos

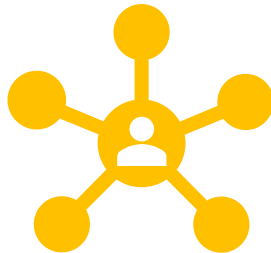


Generación de Análisis

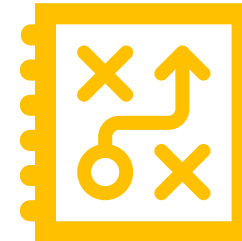


Medición de KPIs

Los **datos y los análisis** ya no se utilizan solo para respaldar la toma de decisiones; cada vez se usan más en lugares donde nunca antes habían existido.



Dan forma y moldean las experiencias de los clientes externos e internos, basados en las preferencias predichas de cómo cada individuo y grupo quiere interactuar con la organización.



Impulsan procesos de negocio, no solo recomendando la siguiente mejor acción, sino también activando esas acciones automáticamente.

# ¿Por qué es útil un proceso para Data & Analytics?



Un proceso es **una secuencia detallada de actividades** necesarias para realizar tareas comerciales específicas



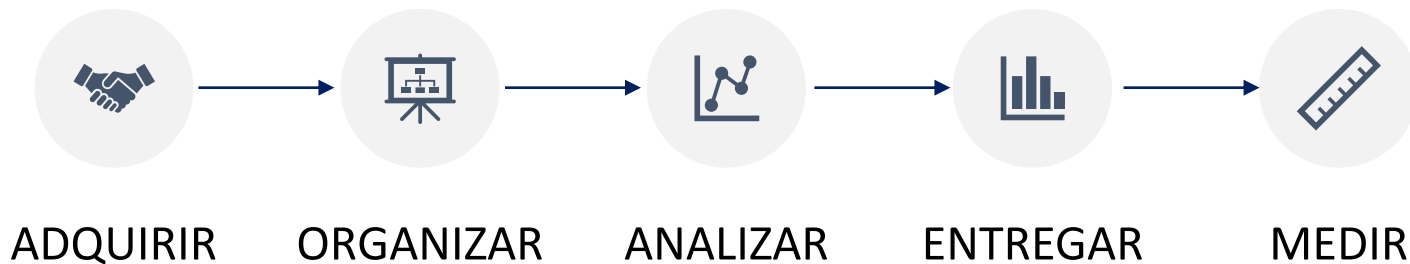
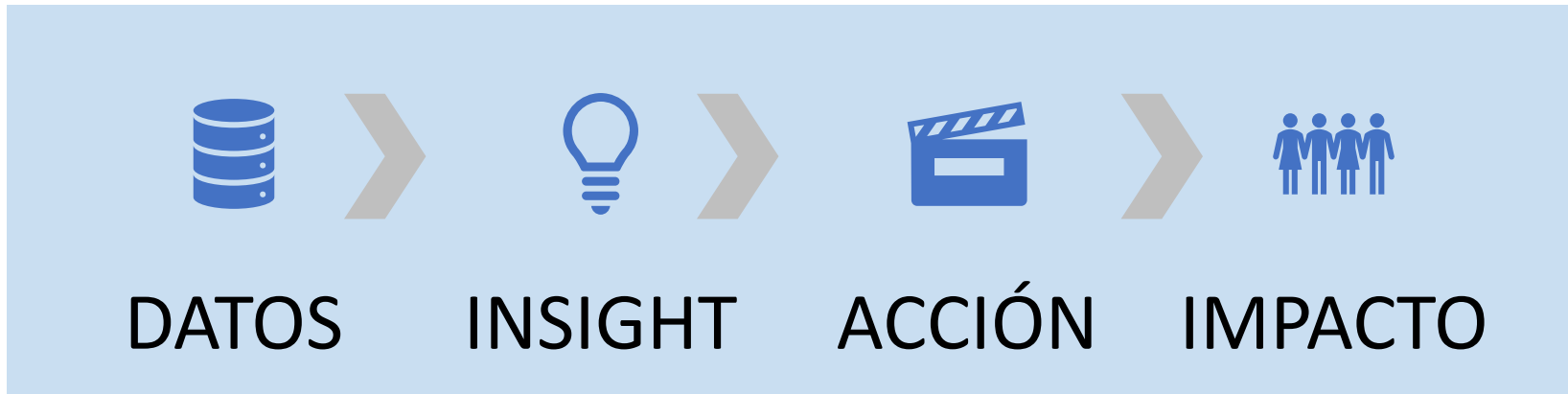
Se utiliza para **estandarizar procedimientos y establecer mejores prácticas.**



La **tecnología y las herramientas** están **cambiando rápidamente.**  
Un proceso estandarizado puede proporcionar continuidad y estabilidad del flujo de trabajo.

- Basado en discusiones con Luis Morinigo, Dir. IoT, NewSignature

# Procesos de Comprensión Analítica

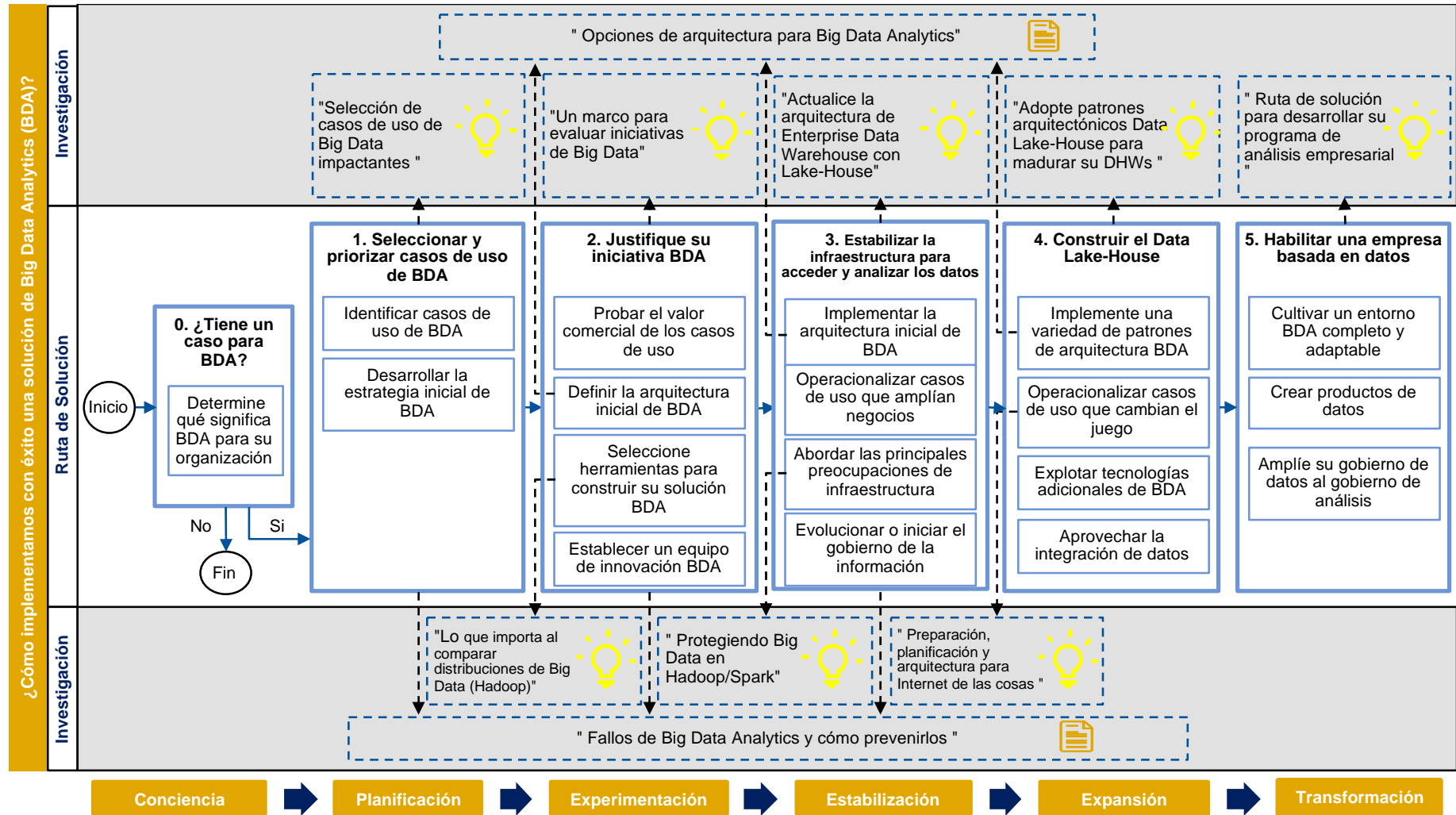


# Un marco para comprender la gestión de datos frente a las necesidades de estrategia de datos

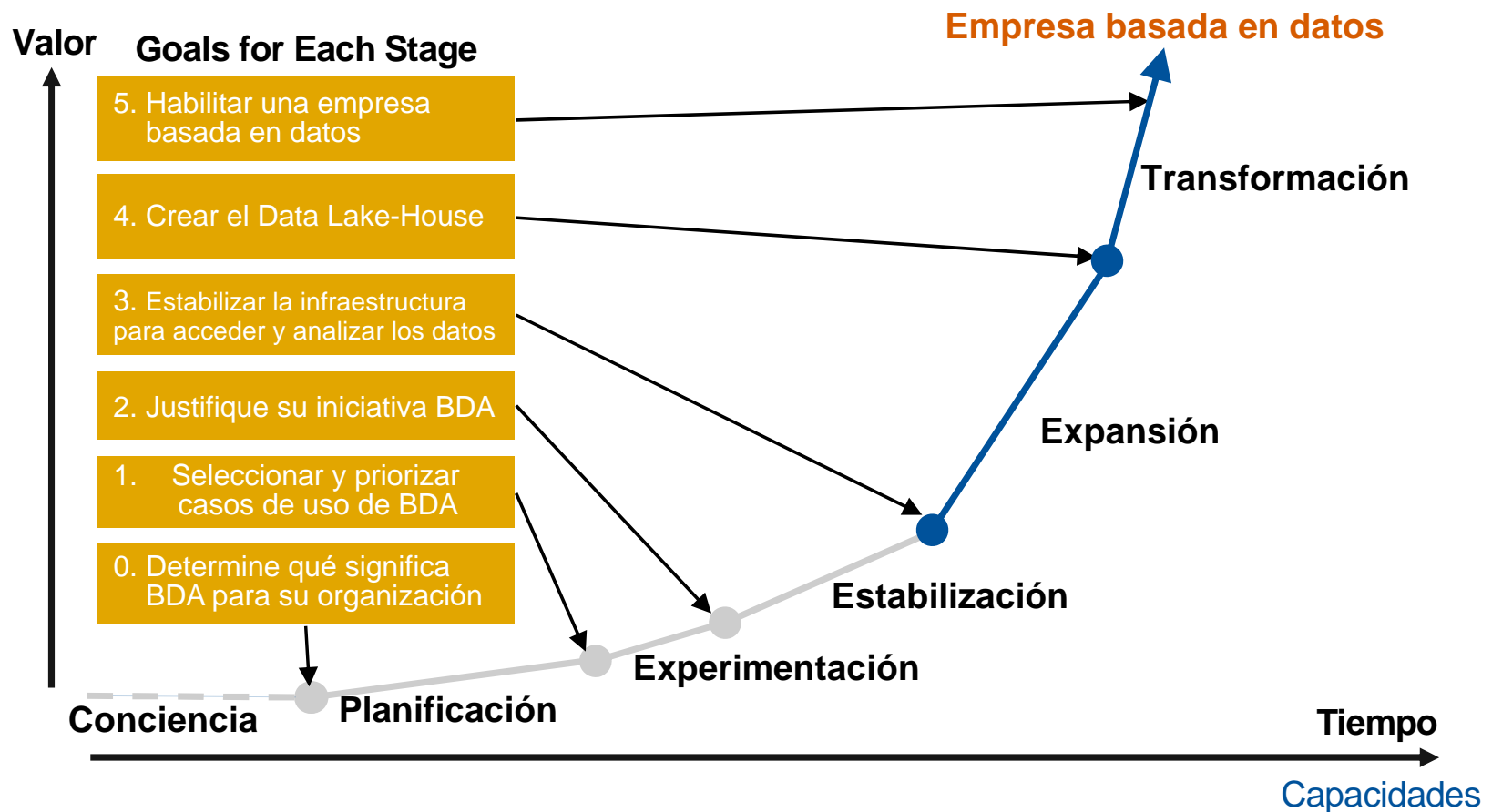




# Procesos de Comprensión Analítica



# Procesos de Comprensión Analítica





# Catégoria de Casos de usos



Conocimiento del  
cliente



Producto y proceso  
eficiente



Productos y  
servicios digitales



Excelencia  
operacional



Digital marketing



Gestión de riesgos  
y cumplimiento



# Casos de uso de conocimiento de cliente

Análisis de  
sentimiento del  
cliente

Prevención de  
abandono de  
clientes (Churn)

Focalización de  
clientes

Segmentación de  
clientes

Customer  
lifetime value

Comportamiento  
del cliente

Interacciones con  
el cliente

Atención al  
cliente

Motores de  
recomendación

Optimización  
personalizada del  
sitio web

# Productos y servicios digitales



DATOS COMO  
PRODUCTO Y DATOS  
COMO SERVICIO



MONITOREO  
REMOTO DE  
TUBERÍAS EN ÁREAS  
DESPOBLADAS



HOGAR  
CONECTADO



OPTIMIZACIÓN Y  
PERSONALIZACIÓN  
DE CONTENIDO



USO DE SENSORES PARA  
DETECTAR ESTACIONAMIENTO  
DISPONIBLE EN LAS ÁREAS  
METROPOLITANAS

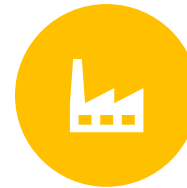
# Excelencia operacional



APLICACIÓN Y  
SOLUCIÓN DE  
PROBLEMAS DEL  
DISPOSITIVO



OPTIMIZACIÓN  
DE  
OPERACIONES  
COMERCIALES



OPTIMIZACIÓN  
DE FABRICACIÓN



OPTIMIZACIÓN  
DE  
OPERACIONES  
EN TIEMPO  
REAL.



CONTROL DE  
CALIDAD Y  
ASEGURAMIENT  
O



GESTIÓN DE  
DATA CENTER



MONITOREO DE  
CAMBIOS



MONITOREO Y  
GESTIÓN DE  
REDES

# Gestión de Riesgos y Cumplimiento



LA VIGILANCIA DEL  
CUMPLIMIENTO  
(DETECCIÓN DE  
PROBLEMAS Y PRUEBA  
DE CUMPLIMIENTO)



AUDITORÍA



VIGILANCIA WEB PARA  
CIBERAMENAZAS



ANÁLISIS DE VIDEO PARA  
VIGILANCIA



DETECCIÓN DE LAVADO  
DE DINERO RIESGO DEL  
CLIENTE (ENTIENDA  
QUIÉN ES EL CLIENTE)

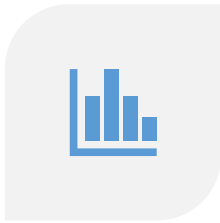


DETECCIÓN DE  
MALWARE INVISIBLE

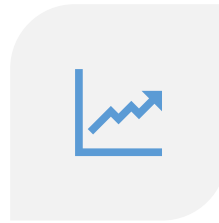


DETECCIÓN DE CUENTA DE  
USUARIO COMPROMETIDA  
PROTECCIÓN DE MARCAS

# Producto y proceso eficiente



ANÁLISIS DE  
INSTRUMENTACIÓN DE  
PRODUCTO



ANÁLISIS DE DEMANDA Y  
FORECASTING



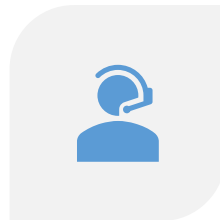
INTELIGENCIA DE MERCADO  
PARA TOMAR DECISIONES  
COMERCIALES ESTRATÉGICAS  
SOBRE PRODUCTOS Y  
PROCESOS.



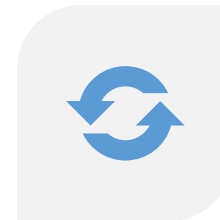
MANTENIMIENTO  
PREVENTIVO DE EQUIPOS  
Y ESTRUCTURAS.



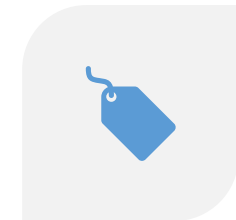
COLOCACIÓN DEL EQUIPO  
PARA UNA SALIDA ÓPTIMA



SOPORTE Y RESOLUCIÓN  
DE PROBLEMAS



CICLO DE VIDA DEL  
PRODUCTO Y CICLOS DE  
VIDA DEL PRODUCTO  
INDIVIDUAL



PRECIO DINAMICO





## Caso de uso - Dynamic Forecasting

- El pronóstico dinámico tiene en cuenta los grandes datos disponibles, que van desde las condiciones del tráfico hasta los pronósticos del tiempo, los precios de los productos básicos y los sensores.
- El pronóstico dinámico se puede aprovechar para un pronóstico más preciso del cliente y para pronosticar la disponibilidad y el precio de productos volátiles como el combustible o el acero.
- El análisis de pronósticos de transporte ofrece metodologías de asignación robustas y específicas para el transporte que pueden traducir la demanda del cliente, la ubicación y el producto en demandas de envío específicas del carril.



## Caso de uso - Stress Testing

- Big data incluye eventos en tiempo real, cambios en la posición del inventario o fallas en la calidad del producto o patrones de demanda.
- Las pruebas de resistencia considerarán una multitud de escenarios y evaluarán el desempeño de la cadena de suministro.
- Las pruebas de resistencia pueden usarse para la gestión de riesgos y la planificación de contingencias.



## Caso de uso - Track and Trace

- Big data incluirá información de sensores y dispositivos de navegación.
- La información se puede utilizar para controlar las posiciones de inventario, los niveles de servicio al cliente y la utilización de la capacidad.
- También se puede utilizar para la trazabilidad de la genealogía y la autenticidad requeridas para productos de marca y productos farmacéuticos de alto valor.

## Caso de uso - Product Quality Management

- Big data incluirá datos de garantía y devolución que identifiquen qué configuraciones y especificaciones son de mayor riesgo para una marca o línea de productos.
- Esto también puede extenderse a las tendencias de las redes sociales para identificar comentarios no informados sobre el uso del producto.
- Los conocimientos se utilizarán para monitorear y optimizar los parámetros de calidad del producto y el proceso de los fabricantes contratados / fuentes de suministro aguas arriba, a través de la producción y hacia el mercado secundario.



## Caso de uso - Manufacturing Process Optimization

- Big data incluirá múltiples variables, como las condiciones del material y del equipo, y otras variables externas que afectan el entorno de fabricación, como la humedad o los costos de energía.
- Los conocimientos correlacionarán los datos "según lo construido" y "según lo fabricado" con los datos "según lo planeado" para determinar dónde se debe gestionar y eliminar la variabilidad del proceso, lo que lleva a la optimización en tiempo real de los procesos de producción para producir el producto de la más alta calidad al más bajo costo o nivel de consumo de recursos.



## Caso de uso - Product Portfolio Management

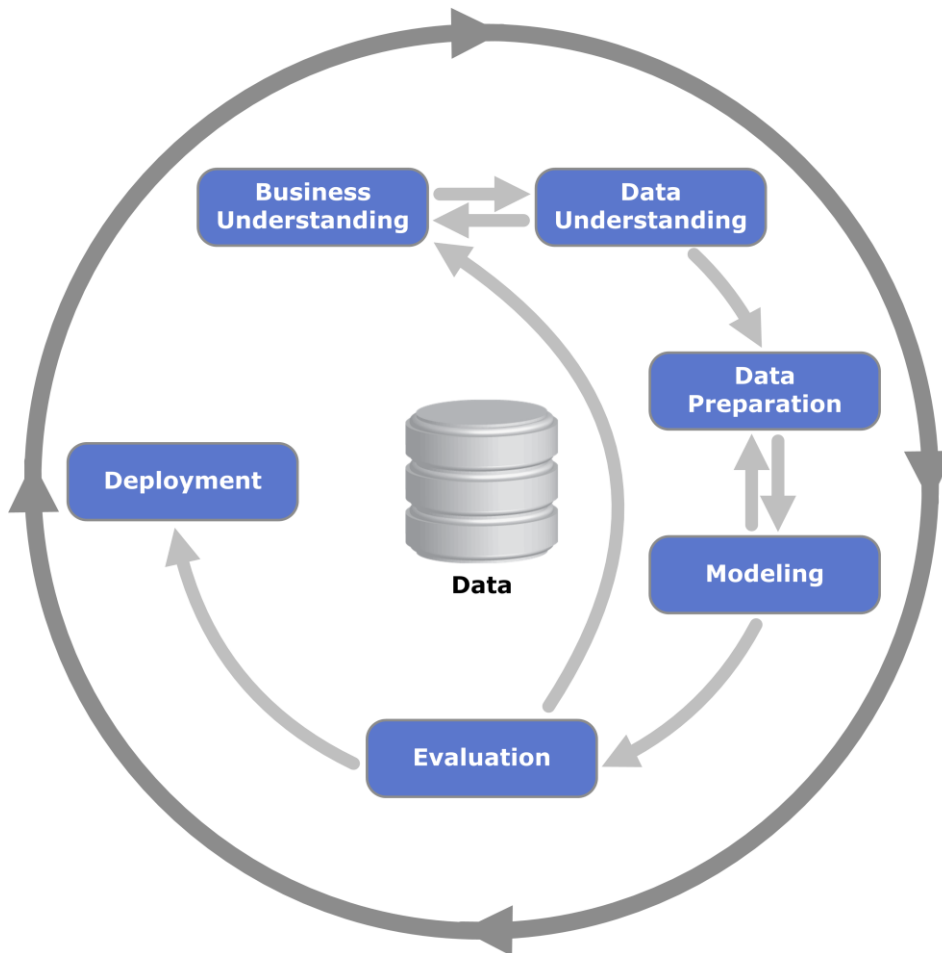
- Big data incluye datos estructurados que abarcan SKU, ventas históricas y contribución al margen. También abarca datos externos estructurados y no estructurados, como los cambios en la demografía y las estrategias de los competidores.
- Utilizando Big Data, las empresas analizarán el rendimiento de cada SKU, con el objetivo de gestionar la jubilación de SKU y la racionalización de SKU para mejorar la rentabilidad general.



# Evaluación de caso de uso

Producto y Proceso / Casos de uso de eficiencia	Complejidad de Big Data Analytics	Crecimiento	Mejora de servicio al cliente	Administración de personal	Lanzamiento de productos	Mejora de operaciones	Manejo de costos	Mejora de ganancias	Innovación en I + D	Marketing
Caso de uso 1	1				✓		✓			
Caso de uso 2	1		✓				✓			✓
Casa de uso 3	2	✓		✓	✓	✓	✓		✓	
Caso de uso 4	3			✓			✓			✓
Caso de uso 5	4	✓	✓			✓	✓	✓	✓	
Caso de uso 6	2	✓		✓	✓	✓		✓	✓	
Caso de uso 7	5		✓					✓		

# Método CRISP-DM



La metodología CRISP-DM (Cross Industry Standard Process for Data Mining) es un marco de proceso para diseñar, crear, construir, probar e implementar soluciones de aprendizaje automático. El proceso se organiza en seis fases. Las fases se pueden ver en el siguiente diagrama.

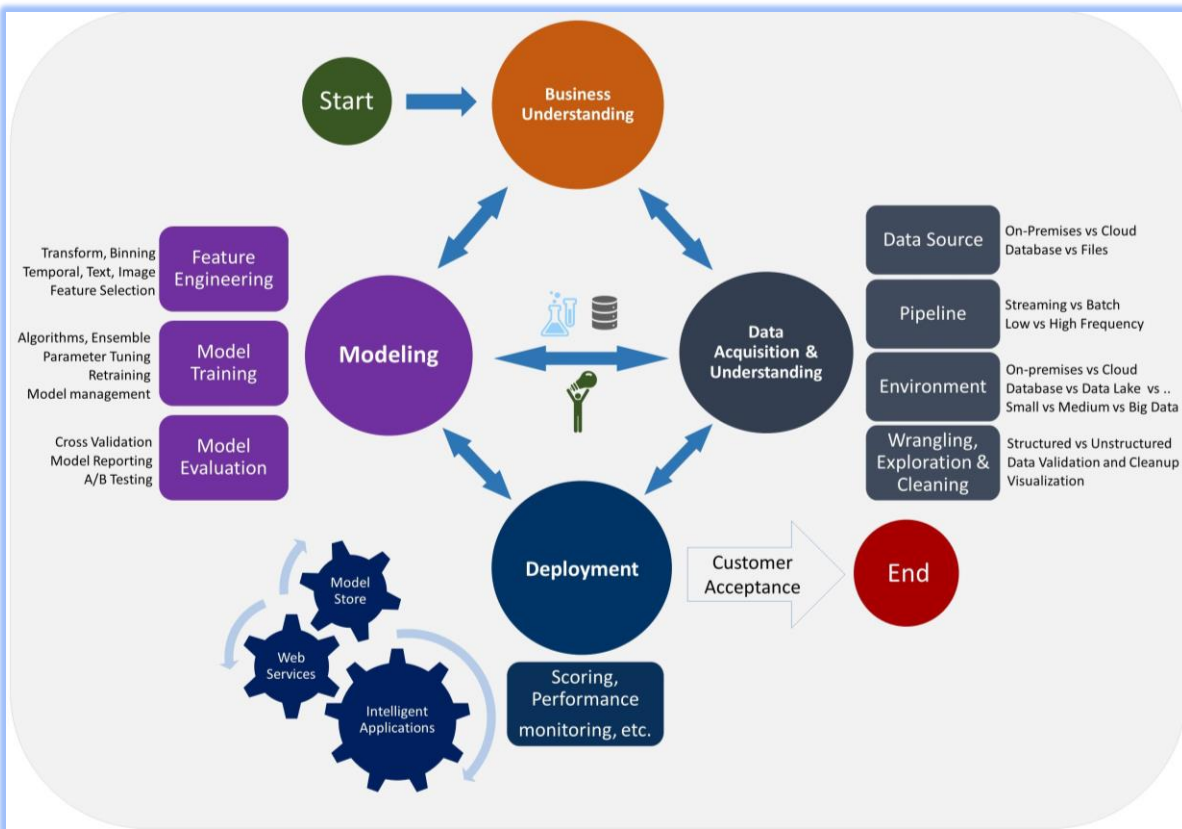


# Modelo de referencia CRISP-DM

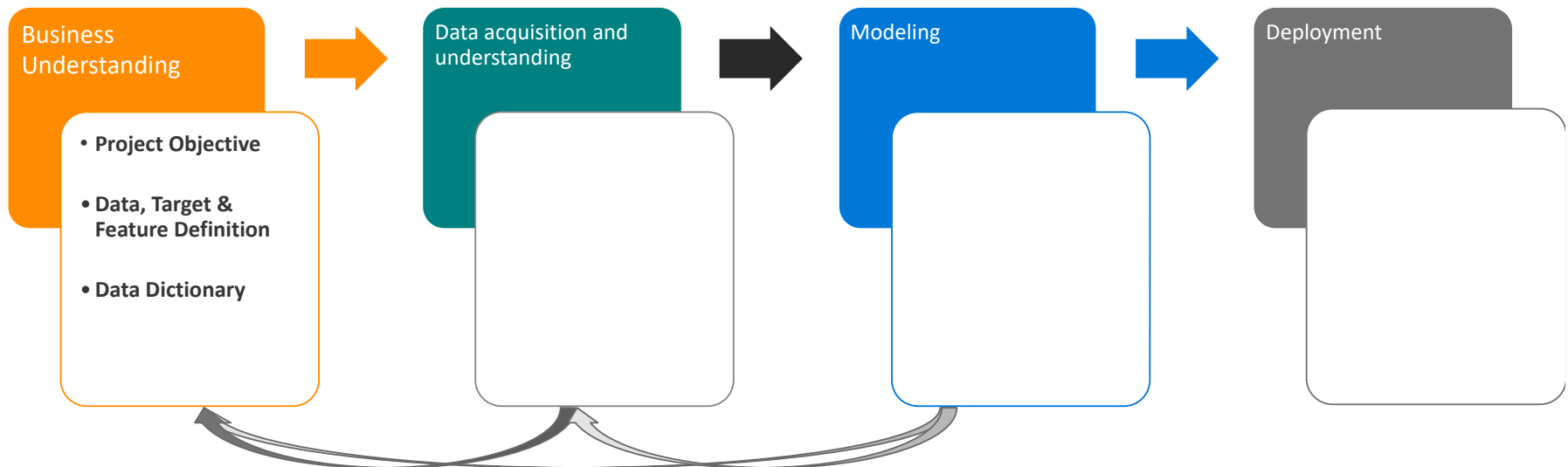
Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
<b>Determine Business Objectives</b> Background Business Objectives Business Success Criteria  <b>Assess Situation</b> Inventory of Resources Requirements, Assumptions, and Constraints Risks and Contingencies Terminology Costs and Benefits  <b>Determine Data Mining Goals</b> Data Mining Goals Data Mining Success Criteria  <b>Produce Project Plan</b> Project Plan Initial Assessment of Tools and Techniques	<b>Collect Initial Data</b> Initial Data Collection Report  <b>Describe Data</b> Data Description Report  <b>Explore Data</b> Data Exploration Report  <b>Verify Data Quality</b> Data Quality Report	<b>Select Data</b> Rationale for Inclusion/Exclusion  <b>Clean Data</b> Data Cleaning Report  <b>Construct Data</b> Derived Attributes Generated Records  <b>Integrate Data</b> Merged Data  <b>Format Data</b> Reformatted Data  Dataset Dataset Description	<b>Select Modeling Techniques</b> Modeling Technique Modeling Assumptions  <b>Generate Test Design</b> Test Design  <b>Build Model</b> Parameter Settings Models Model Descriptions  <b>Assess Model</b> Model Assessment Revised Parameter Settings	<b>Evaluate Results</b> Assessment of Data Mining Results w.r.t. Business Success Criteria Approved Models  <b>Review Process</b> Review of Process  <b>Determine Next Steps</b> List of Possible Actions Decision	<b>Plan Deployment</b> Deployment Plan  <b>Plan Monitoring and Maintenance</b> Monitoring and Maintenance Plan  <b>Produce Final Report</b> Final Report Final Presentation  <b>Review Project</b> Experience Documentation

# Data Science Lifecycle Microsoft

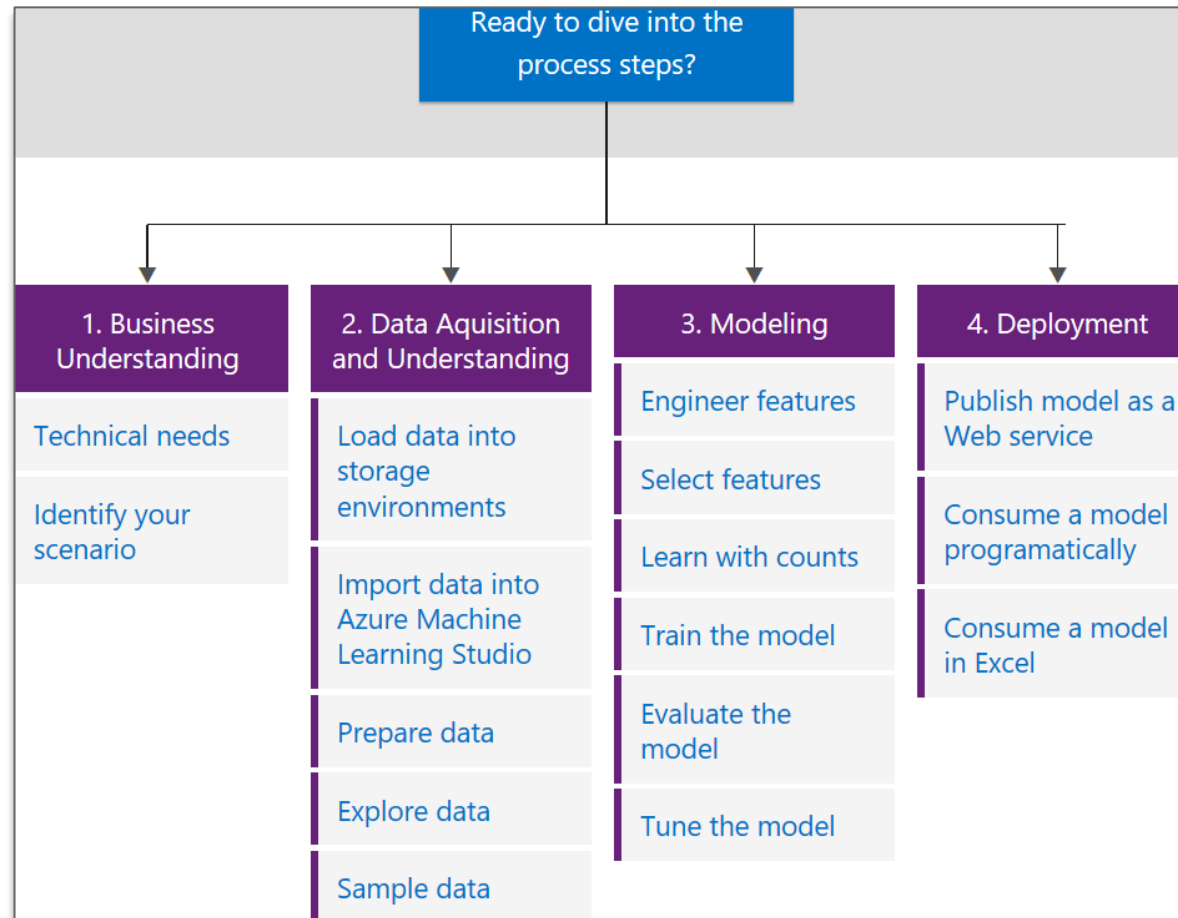
- El Team Data Science Process (TDSP) proporciona un ciclo de vida para estructurar el desarrollo de sus proyectos de ciencia de datos. El ciclo de vida describe los pasos completos que siguen los proyectos exitosos.
- Este ciclo de vida ha sido diseñado como un proceso ágil e iterativo para desarrollar, implementar y administrar aplicaciones de IA en la nube.



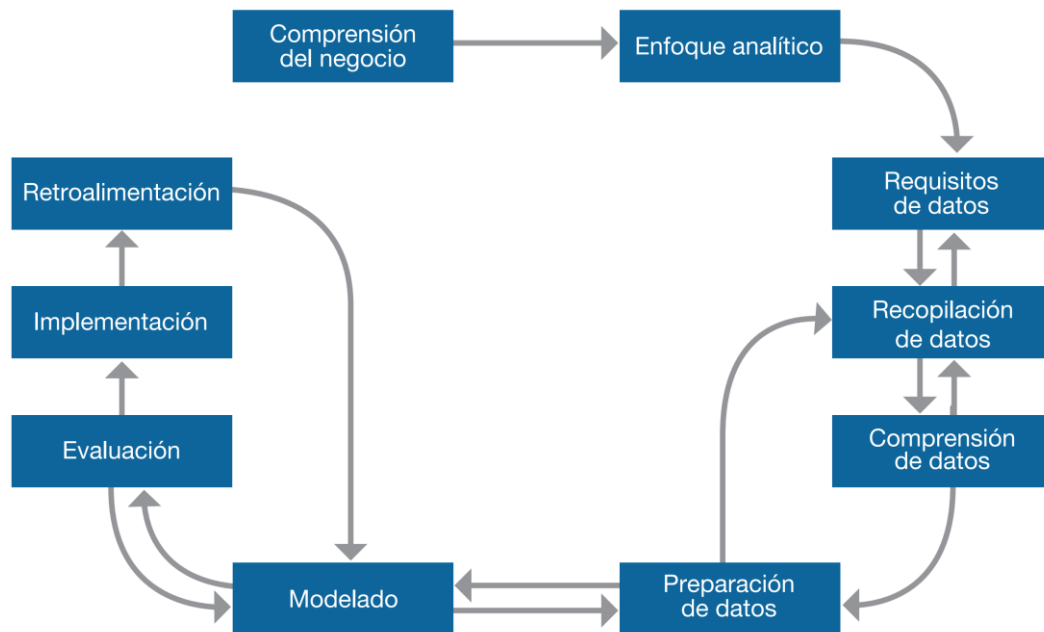
Las etapas del ciclo de vida de TDSP pueden integrarse con entregables y puntos de control específicos



# Modelo TDSP



# Metodología fundamental para la ciencia de datos



Esta metodología tiene algunas similitudes con las metodologías reconocidas para la minería de datos, pero pone el énfasis en varias de las nuevas prácticas en la ciencia de datos, como el uso de grandes volúmenes de datos, la incorporación de la analítica de texto en el modelado predictivo y la automatización de algunos procesos.