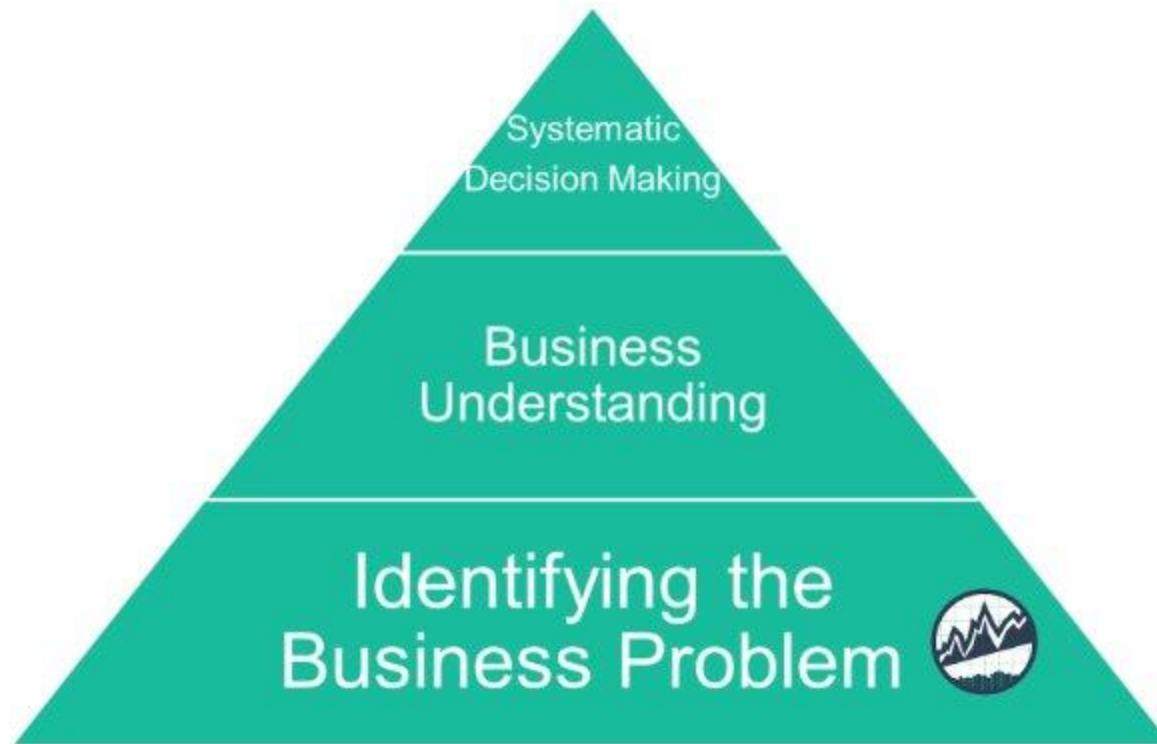


## **Módulo 4** Metodologías de ciencia de datos

*Dr. Carlos Alberto González Martínez*



# Business Science Problem Framework



---

gmc

# Objetivo

El participante identificará el entorno BSPF, informacional y operacional, así como los algoritmos usuales en la ciencia de datos.

# Business Science Problem Framework

## Contenido

### Business Science Problem Framework

1. **Ver el negocio como una máquina**
2. **Comprender los impulsores**
3. **Medir los conductores**
4. **Descubrir problemas y oportunidades**
5. **Codificar algoritmos**
6. **Medir los resultados**
7. **Informar impacto financiero**

# Business Science Problem Framework

## Business Science Problem Framework

Nos basamos en el trabajo de Matt Dancho, titulado *Cómo gestionar con éxito un proyecto de ciencia de datos: el marco del problema de la ciencia empresarial* (2018).

El marco de proyectos de ciencias empresariales (BSPF) surge como apoyo a la necesidad de muchas empresas en la ejecución de proyectos exitosos, sin embargo, la mayoría de los proyectos de ciencia de datos fracasa. BSPF ayuda como guía para proyectos exitosos de ciencia de datos.

---

gmc

# Business Science Problem Framework

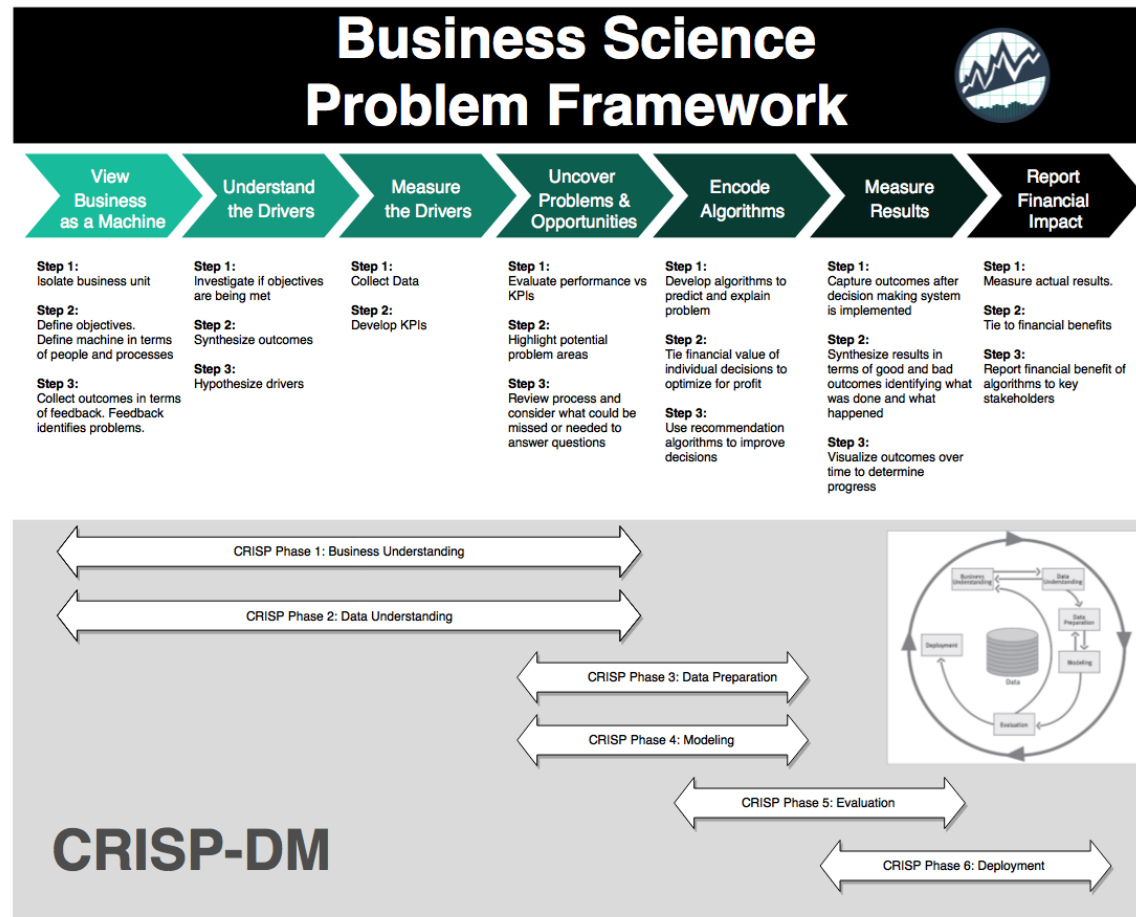
## Business Science Problem Framework

Características de un proyecto de ciencia de datos para tener éxito:

- **Comunicación**, para presentar eficazmente los beneficios a los ejecutivos, mostrando los resultados que se relacionan con los objetivos de la organización.
- **Comprensión empresarial**, que solo ocurre a través de la interacción con las partes interesadas del negocio más cercanas al proceso o problema.
- **Planificación** para alinear a todos los involucrados con el alcance y el plan del proyecto.
- Una lista de verificación de **acciones probadas** que deben considerarse.

# Business Science Problem Framework

## Business Science Problem Framework

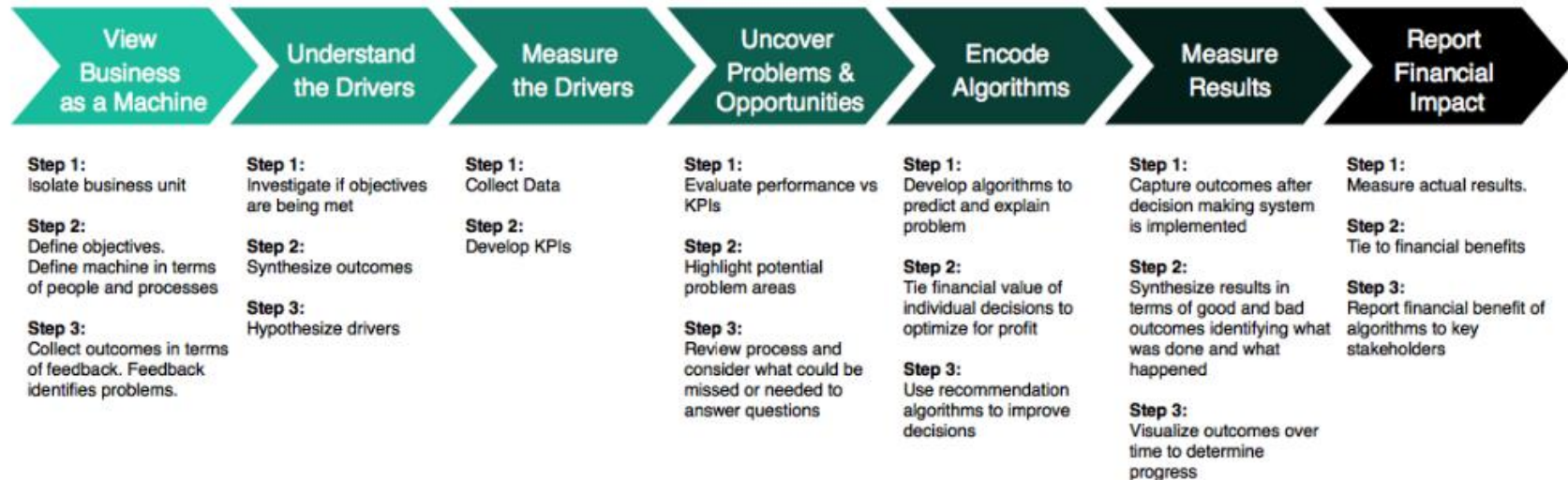




# Business Science Problem Framework

## Business Science Problem Framework

### Para la mitad superior de BSPF



gmc



# Business Science Problem Framework

## 1. Ver el negocio como una máquina

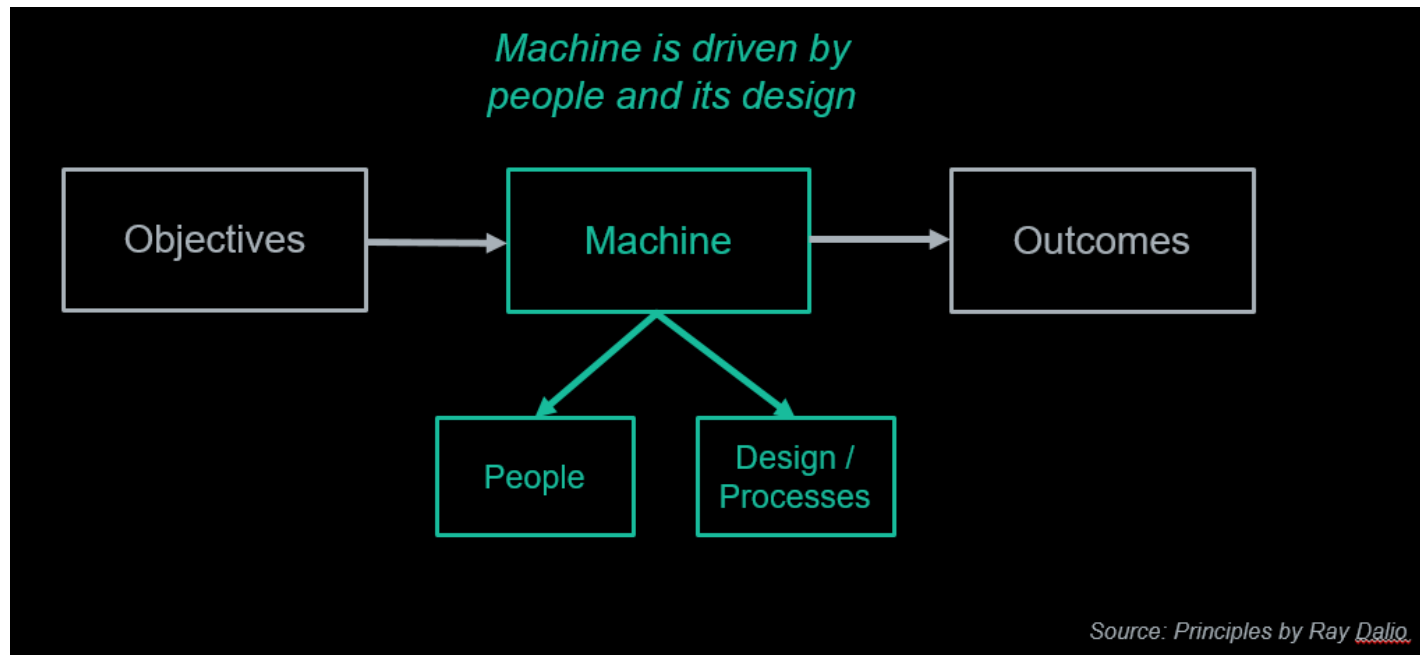
La primera fase es ver el negocio como una máquina.  
Esto involucra:

- 1) Aislar unidades de negocio
- 2) Definir objetivos
- 3) Recolectar resultados

Esto implica dividir el negocio en partes internas (ventas, fabricación, contabilidad, etcétera) y partes externas (clientes, proveedores), visualizando las conexiones.

# Business Science Problem Framework

## 1. Vea el negocio como una máquina



gmc

# Business Science Problem Framework

## 1. Ver el negocio como una máquina

Para ejemplificar la metodología, se va a utilizar el problema del abandono de clientes. Para este ejemplo realizamos la siguiente evaluación:

- 1) **Aislar las unidades de negocio:** *la interacción se produce entre Ventas y el Cliente.*
- 2) **Definir objetivos:** *hacer felices a los clientes*
- 3) **Recolectar resultados:** *estamos perdiendo clientes lentamente. Está reduciendo los ingresos de la organización en 500.000 dólares al año.*

# Business Science Problem Framework

## 2. Comprender los impulsores

A continuación, comenzamos el proceso de comprensión de los controladores. Los pasos clave, son:

- 1) Investigar si se están cumpliendo los objetivos
- 2) Sintetizar resultados
- 3) Hipotetizar los conductores

La clave en esta fase es comenzar con los **objetivos comerciales** : **Satisfacción del cliente**. Cuando los clientes están contentos, siguen regresando. La pérdida de clientes generalmente indica una baja satisfacción.

# Business Science Problem Framework

## 2. Comprender los impulsores

Necesitamos ***sintetizar los resultados***. En nuestro ejemplo hipotético, los clientes se van a la competencia.

El último paso es *formular hipótesis sobre los impulsores*. En esta etapa, es fundamental reunirse con **expertos en la materia (PYME)**.

Para el ejemplo, realizamos la siguiente evaluación:

Investigar si se están cumpliendo los **objetivos**: *No, los clientes no están contentos.*

**Sintetizar resultados**: *el competidor tiene un tiempo de entrega más rápido.*

**Hipotetizar los impulsores**: *el tiempo de entrega está relacionado con la entrega del proveedor, la disponibilidad de inventario, el personal y el proceso de programación.*

# Business Science Problem Framework

## 3. Medir los conductores

Ahora comenzamos el proceso de medición de los controladores. Los pasos clave, son:

- 1) Recolectar datos
- 2) Desarrollar KPI

Primero, necesitamos **recopilar datos** relacionados con los controladores de alto nivel. Estos datos pueden almacenarse en bases de datos o puede ser necesario recopilarlos. Se pueden encontrar en herramientas tipo ERP.

# Business Science Problem Framework

## 3. Medir los conductores

Una vez recopilados los datos, necesitamos desarrollar **indicadores clave de desempeño (KPI)**, que son medidas cuantificables que la organización utiliza para medir el desempeño.

Para nuestro ejemplo de abandono de clientes:

**Plazo de entrega medio:** el nivel es de 2 semanas, que se basa en los comentarios de los clientes sobre la competencia.

**Plazo de entrega promedio del proveedor:** el nivel es de 3 semanas, que se basa en los comentarios relacionados con los proveedores de la competencia.



# Business Science Problem Framework

## 3. Medir los conductores

**Porcentaje de disponibilidad de inventario:** el nivel del 90% está relacionado con el lugar donde los clientes experimentan una demanda insatisfecha. Estos datos provienen de los datos de ERP, que comparan las solicitudes de venta con la disponibilidad del producto.

**Rotación de personal:** el nivel del 15% se basa en los promedios de la industria.

Dos puntos clave en este paso:

1. La recopilación de datos lleva tiempo.
2. Tenga en cuenta que los KPI requieren el conocimiento de los clientes y la industria, para las métricas de proveedores, inventario y facturación.

# Business Science Problem Framework

## 4. Descubrir problemas y oportunidades

Es hora de descubrir problemas y oportunidades. Necesitamos que:

- 1) Evalúe el rendimiento frente a los KPI
- 2) Resalte las áreas potencialmente problemáticas
- 3) Revise nuestro proyecto para ver lo que podría haberse perdido

Para nuestro ejemplo de satisfacción del cliente, revisamos los resultados de los hallazgos de la organización con los KPI, para determinar dónde pueden existir las áreas problemáticas. Ampliamos la tabla de KPI para incluir un valor real actual.

---

gmc

# Business Science Problem Framework

## 4. Descubrir problemas y oportunidades

Nuestro tiempo de entrega promedio es de 6 semanas, en comparación con el tiempo de entrega promedio de la competencia, de 2 semanas, que es la causa del primer pedido para la pérdida de clientes.

El tiempo de entrega promedio de nuestro proveedor está a la par con el de la competencia, lo que no motiva ninguna preocupación.

Nuestro porcentaje de disponibilidad de inventario es del 80%, que es demasiado bajo para mantener un alto nivel de satisfacción del cliente. Esta podría ser una razón por la que la deserción está aumentando.

Nuestra rotación de personal en áreas clave es cero durante los últimos 12 meses, por lo que no es motivo de preocupación.

# Business Science Problem Framework

## 5. Codificar algoritmos

Los pasos clave en esta fase, son:

- 1) Desarrolle algoritmos para predecir y explicar el problema
- 2) Optimice las decisiones para maximizar las ganancias
- 3) Utilice algoritmos de recomendación para mejorar la toma de decisiones

Para el **desarrollo de algoritmos**, se puede seleccionar entre una basta gama de herramientas y lenguajes, como: Python, R, RapidMiner, Modeler, H2O Automated Machine Learning y LIME, para explicaciones de modelos de caja negra, por mencionar algunas.

# Business Science Problem Framework

## 5. Codificar algoritmos

A continuación, **optimice las selecciones de decisiones para maximizar las ganancias**. Investigar la **optimización del umbral** para problemas de clasificación binaria. Además, pruebe el **análisis de sensibilidad** para medir qué características tienen el mayor efecto en la rentabilidad de las decisiones.

Por último, cree **algoritmos de recomendación** que incorporen comentarios de las PYME, junto con las explicaciones de las funciones de LIME (o procedimientos de explicación de funciones similares).

---

gmc

# Business Science Problem Framework

## 6. Medir los resultados

Una vez que un modelo ha sido desarrollado, evaluado y enviado a producción (es decir, implementado), es el momento de medir los resultados. Esto requiere de:

- 1) Capturar resultados
- 2) Sintetizar resultados
- 3) Visualizar los resultados a lo largo del tiempo

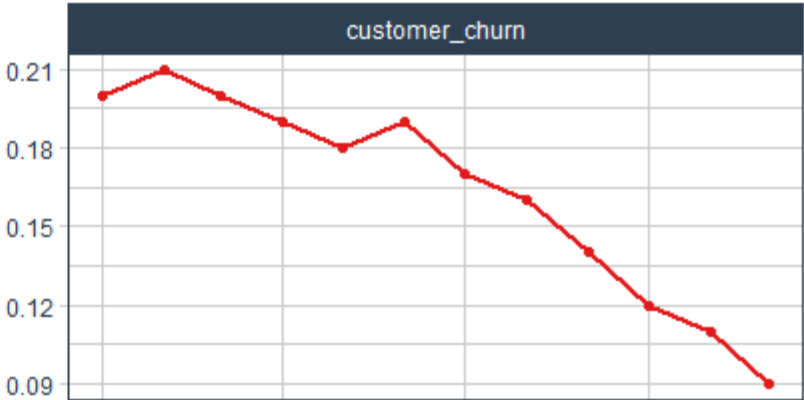
Una vez que el algoritmo se implementa a través de una aplicación web u otra herramienta de toma de decisiones, los resultados deben medirse para mostrar el progreso.

# Business Science Problem Framework

## 6. Medir los resultados

Para nuestro ejemplo de satisfacción del cliente, podemos hacer gráficos como éstos, que exponen la disponibilidad de inventario y la tasa de rotación de clientes. Estamos viendo un aumento de inventario y una disminución de la rotación de clientes. ¡Estos son buenos resultados!

Monitoring Impact of Decision Making Improvements



gmc



# Business Science Problem Framework

## 7. Informar impacto financiero

Ahora estamos en la última fase: informar el impacto financiero. Si hicimos una buena ciencia de datos, implementamos la toma de decisiones sistemática y repasamos los problemas, corrigiéndolos a lo largo del camino. Ahora deberíamos ver resultados positivos. Estos son los pasos:

- 1) Medir los resultados reales
- 2) Vincular los beneficios económicos
- 3) Informar el beneficio financiero a las partes interesadas clave

---

gmc

# Business Science Problem Framework

## 7. Informar impacto financiero

Una vez que se comprenden los resultados, debemos mostrarlos como beneficios financieros. Esto no solo justifica nuestra existencia, sino que muestra a la organización que está mejorando.

**La clave aquí es que los resultados deben transmitirse en términos de impacto financiero.**

Es insuficiente decir que salvamos a 75 empleados o 75 clientes. Más bien, debemos decir que el costo promedio de un empleado perdido o un cliente perdido es de \$ 100,000 por año, por lo que le ahorramos a la organización \$ 7.5M/año. Informe siempre como valor financiero.

Para lo anterior también nos podemos apoyar en gráficas.

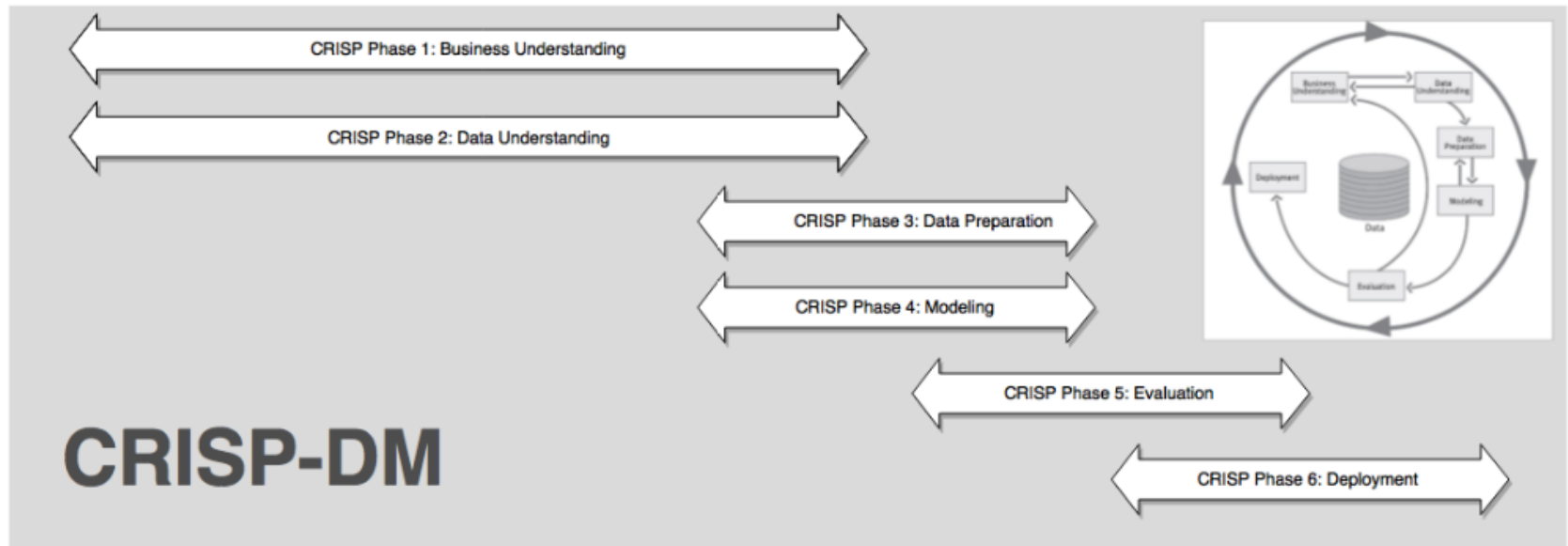
---

gmc

# Business Science Problem Framework

Para la mitad inferior

Se sigue la metodología CRISP-DM



gmc

# Business Science Problem Framework

## Para la mitad inferior

Las siete fases de BSPF fluyen a lo largo de las seis fases de CRISP-DM, que son pasos de **alto nivel** para cualquier problema de ciencia de datos (más allá de los negocios):

1. Comprensión empresarial
2. Comprensión de datos
3. Preparación de datos
4. Modelado
5. Evaluación
6. Despliegue

# Referencias bibliográficas

Dancho, M. (2018). *Cómo gestionar con éxito un proyecto de ciencia de datos: el marco del problema de la ciencia empresarial*. Recuperado de <https://www.business-science.io/business/2018/06/19/business-science-problem-framework.html>

# Contacto

Carlos Alberto González Martínez

*Jefe de departamento de correlaciones, cruces y alertas (C5i)*

gmcmxiv@hotmail.com