

# Ciclo de vida en proyectos de Data Science: Expectativa vs Realidad



Melina Solovey

Data Scientist, Pi Data Strategy & Consulting



WOMEN IN DATA SCIENCE



DATA SCIENCE FEM

# ¿QUÉ ES DATA SCIENCE?

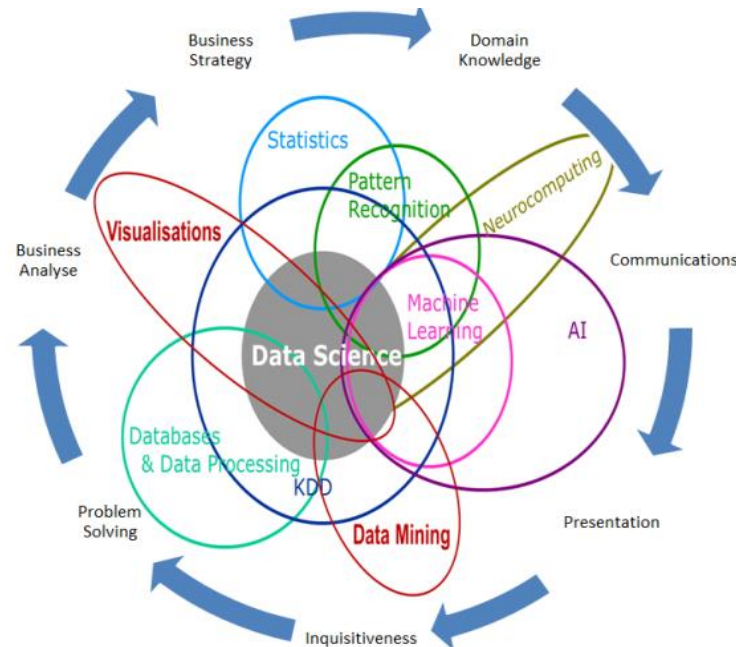


WOMEN IN DATA SCIENCE



DATA SCIENCE FEM

La ciencia de datos es un campo interdisciplinario que involucra análisis e inferencia de datos, desarrollo de algoritmos y aplicación de sistemas o tecnologías para resolver problemas analíticamente complejos.





# ¿QUÉ NO ES DATA SCIENCE?



WOMEN IN DATA SCIENCE



DATA SCIENCE FEM

Data Science  $\neq$  Business Intelligence

Data Science  $\neq$  Big Data

Data Science  $\neq$  Machine Learning

Machine Learning no es una rama de data science. Data Science usa a ML como herramienta.



No es la salvación de las empresas que nunca midieron nada.  
“entra basura, sale basura”



Presente los datos utilizando algunos gráficos de Excel sin ninguna información sobre los datos.



WOMEN IN DATA SCIENCE



DATA SCIENCE FEM



# EXPECTATIVA



WOMEN IN DATA SCIENCE



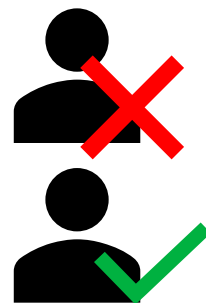
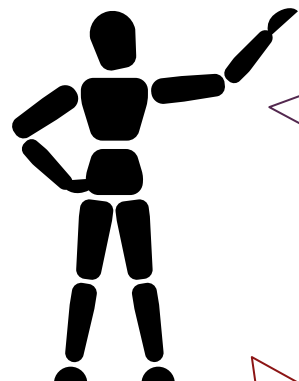
DATA SCIENCE FEM

Queremos  
hacer un churn  
de nuestros  
clientes...

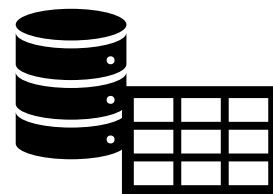
...hemos generado  
este dataset,  
donde ya hemos  
clasificado a  
nuestros clientes...

...creemos que las  
variables A, B y C  
inciden en este  
proceso...

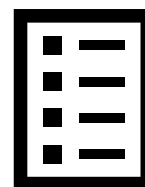
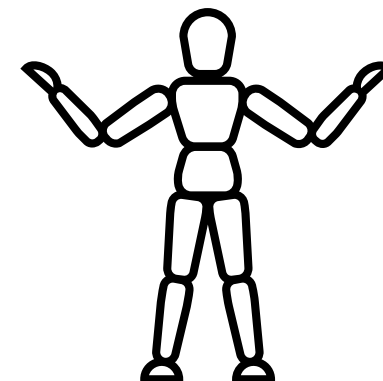
...generamos este  
diccionario de  
datos.



Comenzaremos  
a trabajar!!!



A	B	C	D	E	F
/	/	/	-	-	\



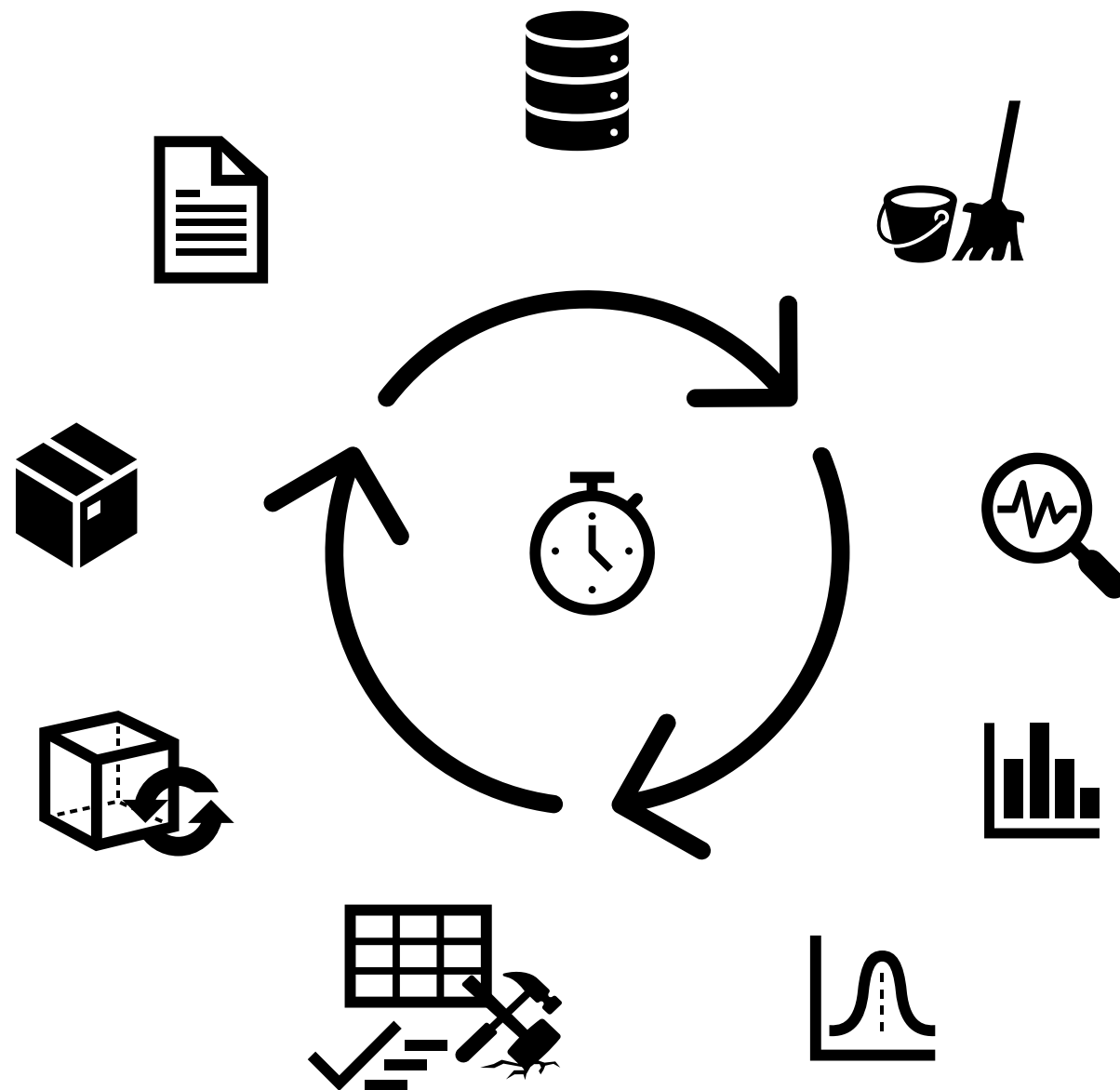


# Modelar

100%







# REALIDAD

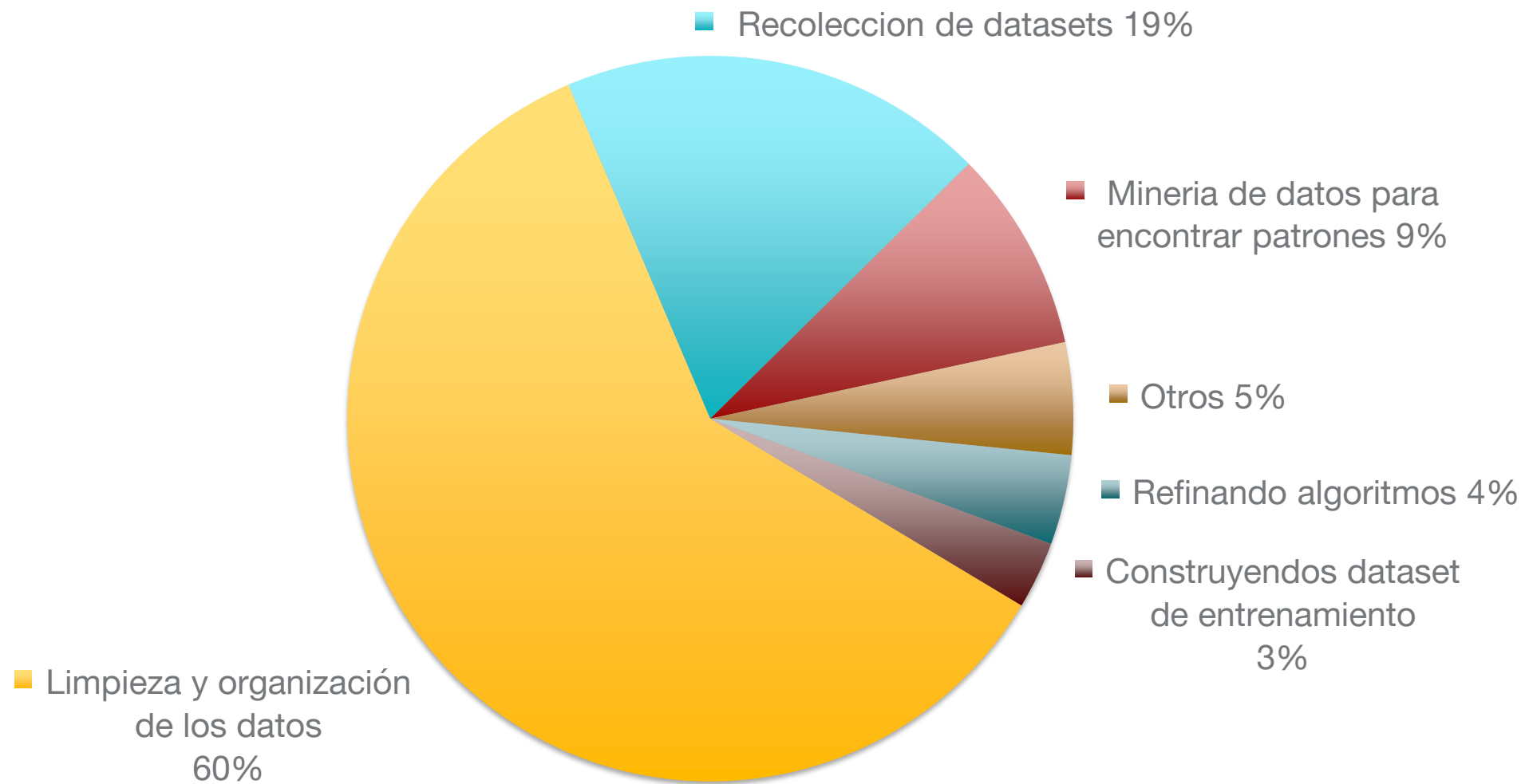


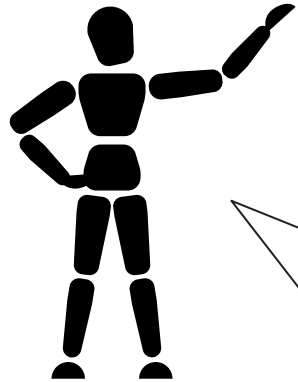
**WOMEN IN DATA SCIENCE**



**DATA SCIENCE FEM**







Queremos optimizar este proceso...

Donde están los datos? En que formato? Como los vamos a disponibilizar?

Que valor quieren optimizar? Sobre que métrica trabajaremos?

Ya tenemos el objetivos? Hay que crearlo?  
Hay que empezar a recolectarlo?

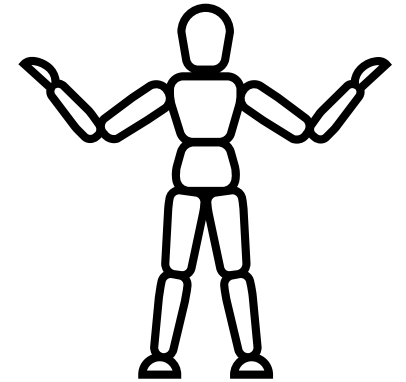
Que tamaño tienen los datos?

Sobre que plataforma vamos a trabajar?

Como vamos a disponibilizar el modelo? Como será el reentrenamiento?

Quien va a ser el stakeholder?

Como quieren consumir el modelo?  
Predicción diaria? A demanda?





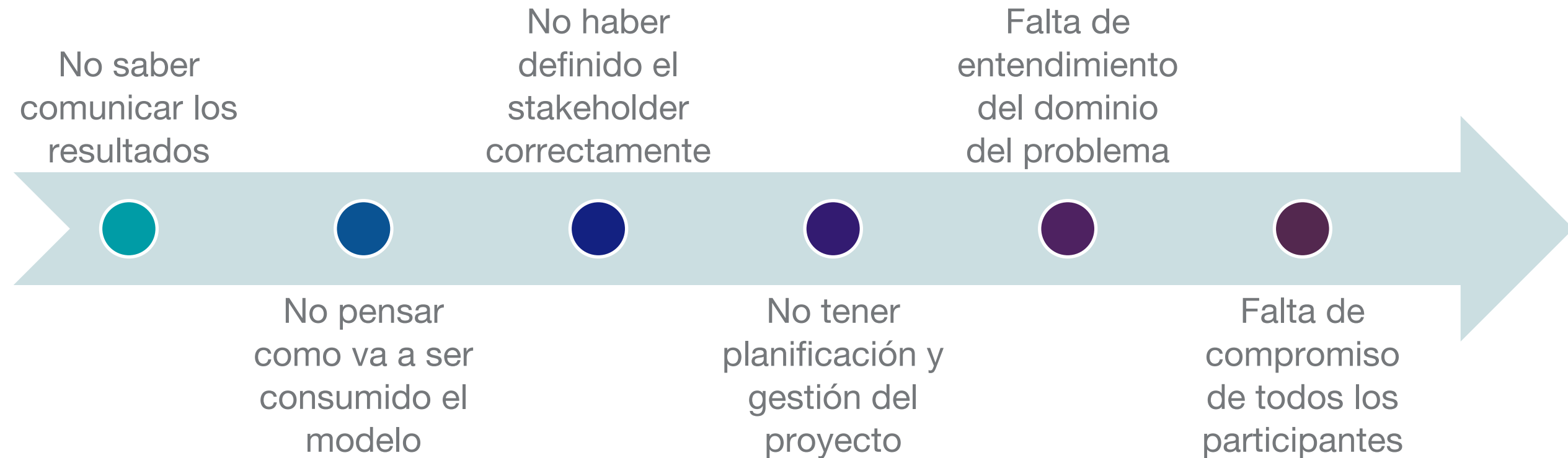


WOMEN IN DATA SCIENCE

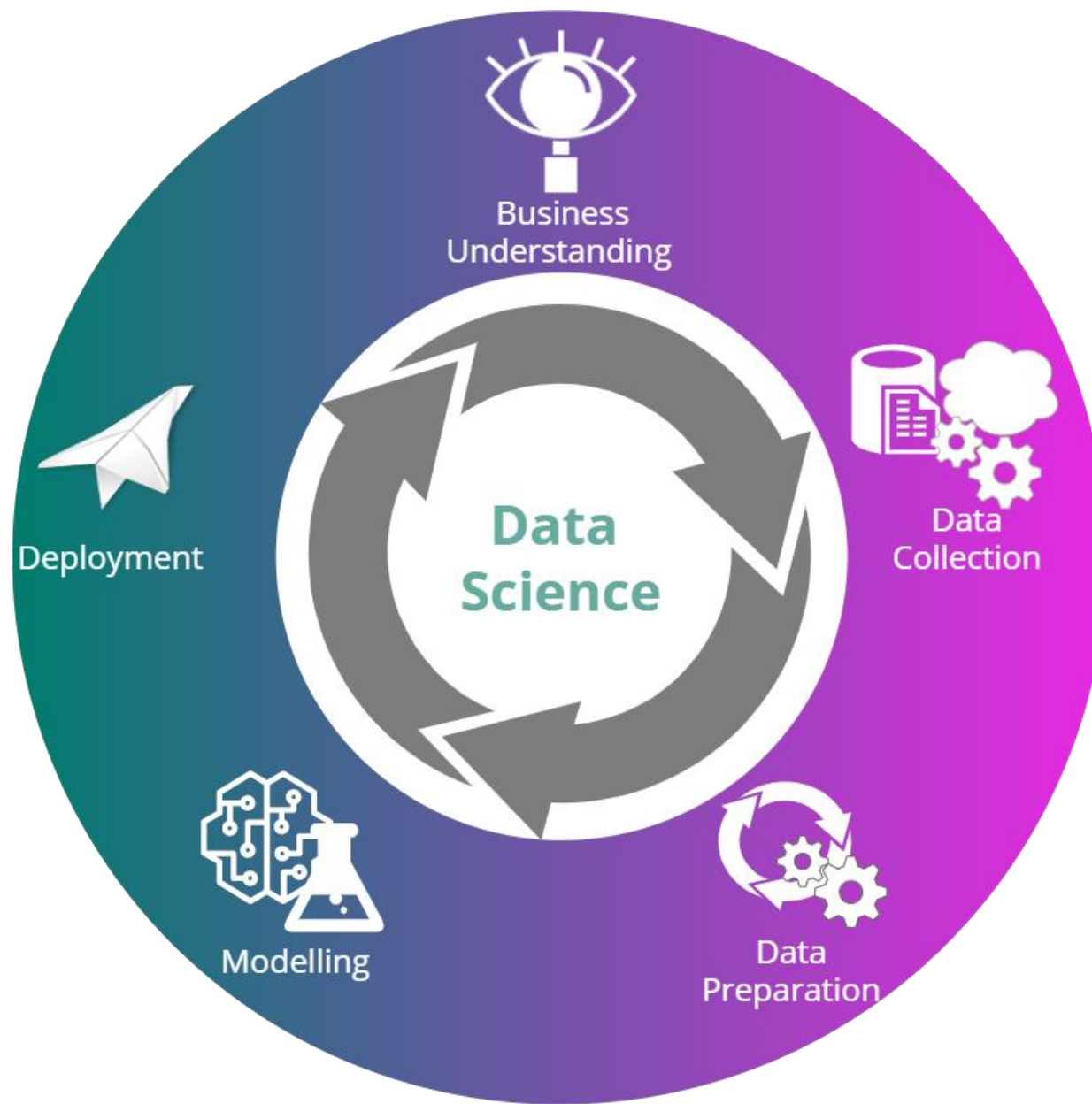


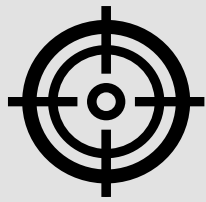
DATA SCIENCE FEM

# Camino al fracaso









Objetivos



Especificar  
variables claves

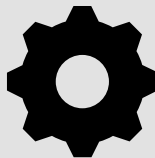


Identificar los  
orígenes de datos



Cómo hacerlo

- Identificar las **variables de negocio claves** que el análisis necesita predecir
- Definir los **objetivos del proyecto** formulando y refinando preguntas "precisas" que sean relevantes, específicas y sin ambigüedades.
- Definir:
  - Grupo de proyecto
  - Roles & Responsabilidades
- Definir las métricas de éxito **-SMART-**
- Encontrar los datos relevantes



Artefactos



Documento  
de Proyecto



Diccionario  
de datos



Listado de  
orígenes



WOMEN IN DATA SCIENCE



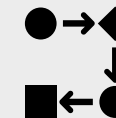
DATA SCIENCE FEM



Objetivos



Datos limpios y de  
alta calidad

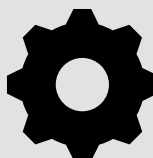


Arquitectura del  
pipeline de datos



Cómo hacerlo

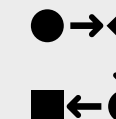
- **Disponibilizar los datos** en el entorno analítico de destino
- **Explorar los datos** para determinar si la calidad de los datos es adecuada para responder a la pregunta
- **Configurar el pipeline de datos** para sumar o actualizar los datos periódicamente.



Artefactos



Reporte de  
calidad de datos



Definición de  
Arquitectura



Data  
Collection



WOMEN IN DATA SCIENCE



DATA SCIENCE FEM

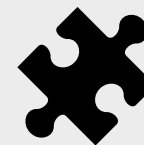




Objetivos



Análisis de  
datos

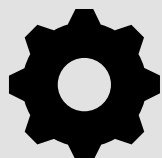


Construcción  
de features



Cómo hacerlo

- **Manejo de datos faltantes**
- **Corregir valores no válidos**
- **Eliminar duplicados**
- **Estructurar los datos para introducirlos en un algoritmo**
- **Feature engineering:** crear nuevas features a partir de los datos sin procesar para facilitar el entrenamiento del modelo



Artefactos



Reporte de Análisis



Set de Features



Checkpoint





Objetivos

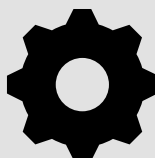


Modelo



Cómo hacerlo

- **Entrenamiento de modelo:** encontrar el modelo que responde a la pregunta con mayor precisión comparando sus métricas de éxito. Actividades:
  - *Dividir los datos en test y train*
  - *Construir los modelos usando el dataset de train*
  - *Evaluar el entrenamiento y la performance con el dataset de test*
  - *Seleccionar la “mejor” solución*
- Determinar si su modelo es **adecuado para poner producción**



Artefactos



Modelo



Reporte del modelo



Checkpoint



Modelling



WOMEN IN DATA SCIENCE



DATA SCIENCE FEM



Objetivos

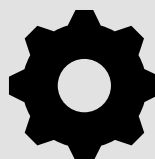


Deploy del modelo



Cómo hacerlo

- ***Poner en funcionamiento el modelo:*** implementar el modelo y el pipeline en un entorno de producción o similar al de producción, para el consumo de aplicaciones



Artefactos



Detalles del  
Despliegue



Documento de la  
arquitectura de la solución



WOMEN IN DATA SCIENCE

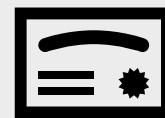


DATA SCIENCE FEM





Objetivos

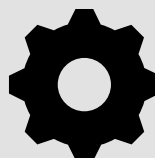


Finalizar los entregables  
del proyecto



Cómo hacerlo

- **Validación del sistema:** confirme que el modelo implementado y el pipeline satisfacen las necesidades del cliente
- **Transferencia del proyecto:** transfiera el conocimiento del proyecto a la entidad que ejecutará el sistema en producción



Artefactos



Informe de cierre del  
proyecto para el cliente



Business



WOMEN IN DATA SCIENCE



DATA SCIENCE FEM

# ROLES



**WOMEN IN DATA SCIENCE**



**DATA SCIENCE FEM**



## Data Scientists

### Tareas

- Implementa y optimiza algoritmos
- Extrae datos de su fuente
- Procesa los datos usando diferentes métodos estadísticos
- Diseña nuevos test o experimentos
- Visualiza y presenta gráficamente los datos



## Data Engineers

### Tareas

- Analiza y evalúa plataformas y arquitecturas de mercado para recopilar, almacenar, procesar y analizar conjuntos de datos
- Realiza la consultoría de los procesos de negocios, considerando UX, seguridad, costos, etc
- Gestiona el presupuesto asignado para plataformas de datos



## Architects

### Tareas

- Trabaja en equipos con otros programadores
- Analiza la arquitectura existente y diseña y gestiona la nueva arquitectura
- Implementa y desarrolla la arquitectura
- Elabora test unitarios sobre la plataforma de datos
- Elabora propuestas para nuevas herramientas según la necesidad del cliente



## Project Manager

### Tareas

- Crea y coordina los equipos
- Define los tiempos de desarrollo e implementación de los proyectos, de acuerdo con las necesidades del clientes
- Gestiona las relaciones con proveedores, contrataciones, gestión de partners, etc
- Crear informes y presentaciones







## Data Scientists

### Skills

- Formación en matemáticas, estadística o machine learning
- Gran capacidad comunicativa
- Capacidad lógica y de intuición
- Gran capacidad organizativa y metódica



## Data Engineers

### Skills

- Experiencia en lenguajes de programación
- Experiencia para manejar tecnología de datos



## Architects

### Skills

- Capacidad analítica
- Capacidad de aprendizaje autónomo
- Gran interés y experiencia en lenguajes de programación y tecnologías más recientes
- Experiencia en herramientas cloud
- Experiencia en storage y data mining

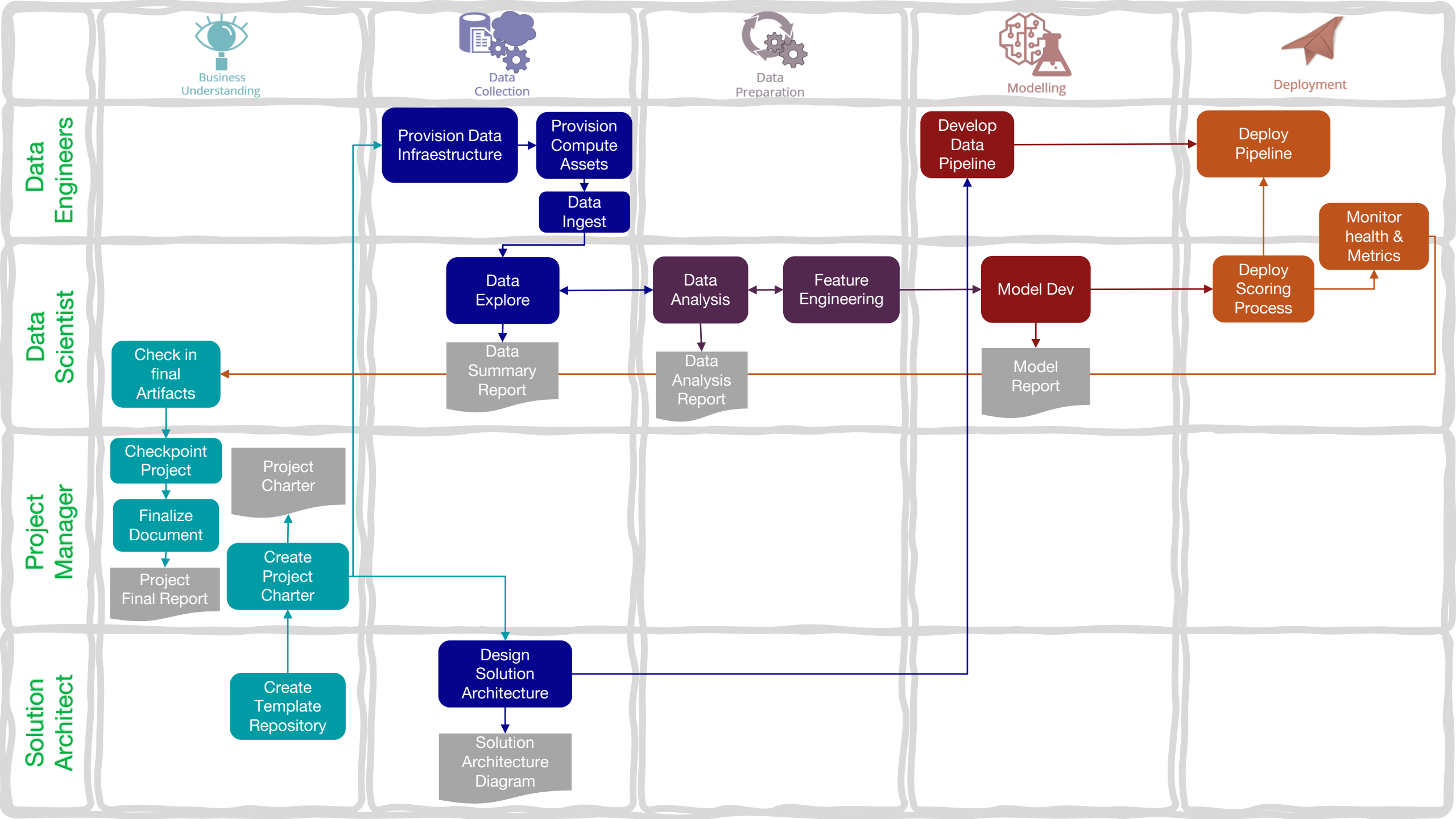


## Project Manager

### Skills

- Comunicación y habilidades interpersonales
- Capacidad para cumplir los plazos de entrega y gestionar la entrega de proyectos
- Excelentes habilidades de redacción de informes y presentación
- Pensamiento crítico y resolutivo





# Buenas prácticas





# MUCHAS GRACIAS



**WOMEN IN DATA SCIENCE**



**DATA SCIENCE FEM**



¿Qué otra buena  
práctica sumarias a  
la última lista?



WOMEN IN DATA SCIENCE



DATA SCIENCE FEM