



Curso

Data Analytics con Python



Carlos Rondan

Data Scientist Senior



Mentor



Data Scientist Senior

24 años

Carlos Rondan

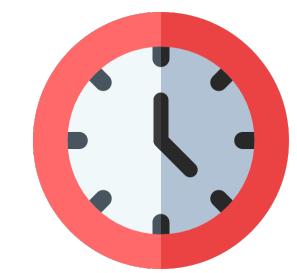
Experiencia

- Master en Data Science & Big Data
- Experto en Analítica multivariante y Machine Learning
- 4 años de experiencia en los sectores Banca, inmobiliario, tech y Salud (seguros y prestación)
- VP Analytics & CO-Founder de Futura
- Certificado en CAPM® (Gestión de Proyectos - PMI)
- Ingeniero Industrial



“El reto del Científico de Datos es equilibrar lo práctico y lo complejo. Dominar a fondo una técnica es más valioso que aprender muchas herramientas.”

Reglas



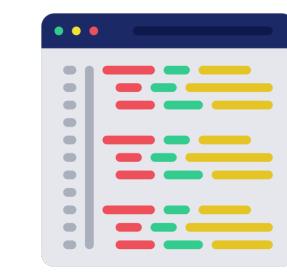
**Ser puntual con
cada una de las
sesiones**



**Mantener el
micro en
silencio**



**Usar el chat
para las
preguntas**



**Completa
todas las tareas
a tiempo**



20%

Asistencia

Se enviará en la clase un formulario de asistencia en los primeros 15 min.



30%

Quiz de Clase

Al inicio de Cada sesión se evaluará con un quiz correspondiente a la sesión pasada



50%

Trabajo Final módulo

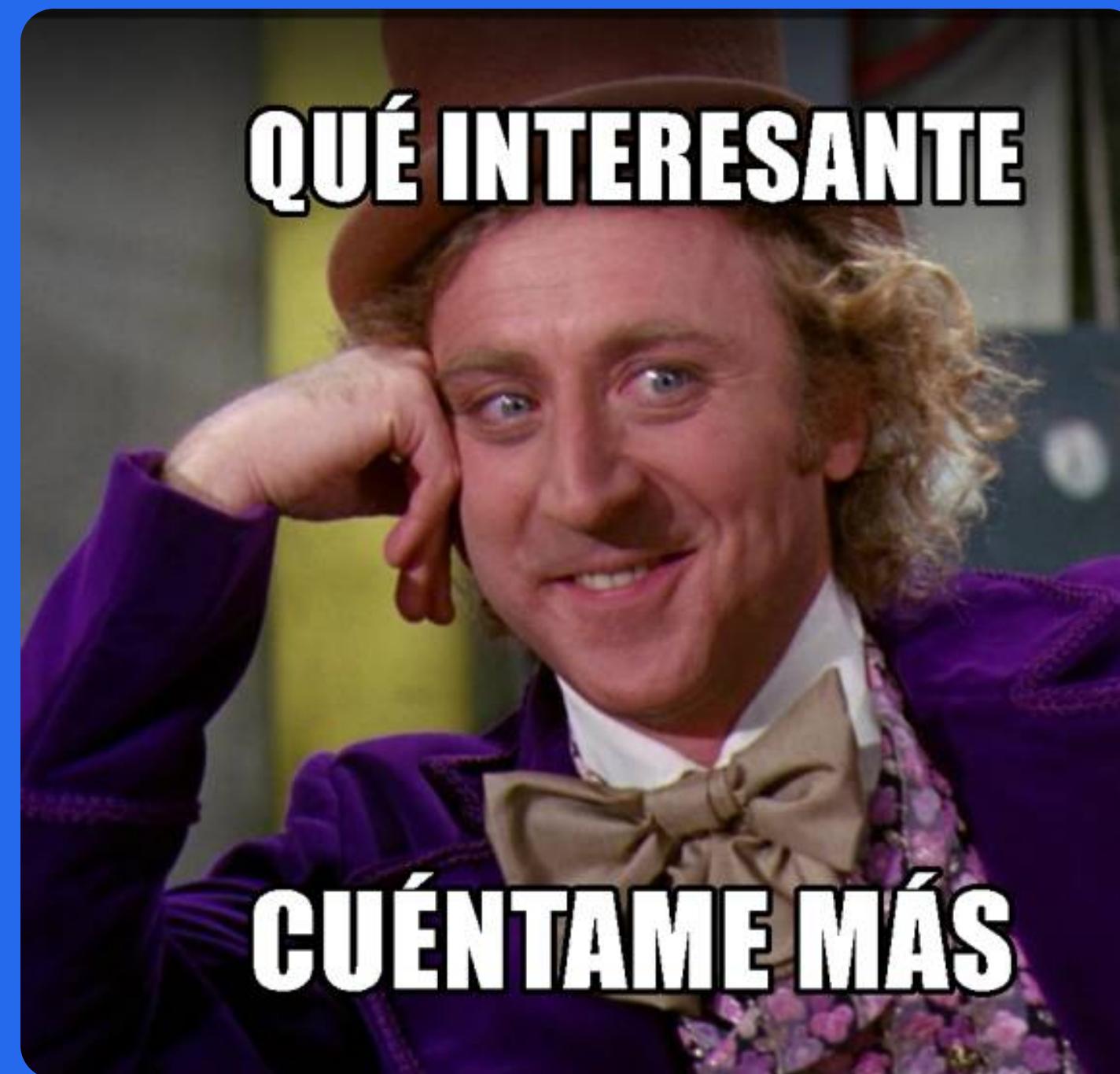
Cada módulo contará con un proyecto a realizar.

Calificación Final del Curso



Requisitos mínimos para obtener el Certificado de Estudios

¡Queremos conocerte!
Cuéntanos un poco más sobre ti...



Línea de carrera:

**Pero antes hablemos
de los roles en Data...**



Carlos Rondan

Data Scientist Senior

Agenda

1 Flujo de trabajo y los roles clave

2 Qué es la Ciencia de Datos

3 Habilidades esenciales para un DS

4 Herramientas y tecnologías clave

5 Construyendo tu línea de carrera



Roles Clave

Roles en un equipo de Data

- Data Analyst

Analiza datos estructurados, genera reportes y visualizaciones para responder preguntas del negocio. **Análisis descriptivo y exploración de datos**

- Business Analyst

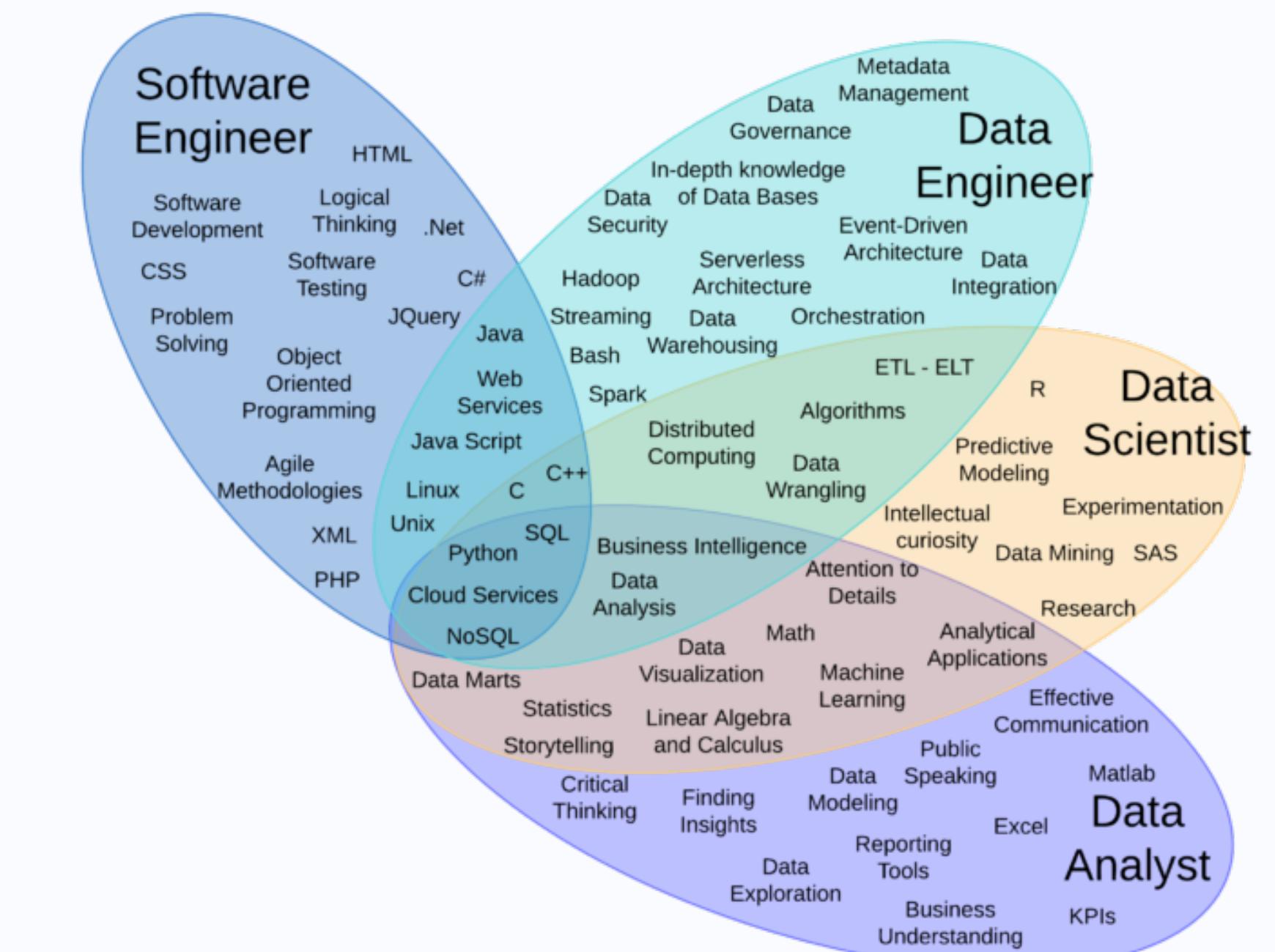
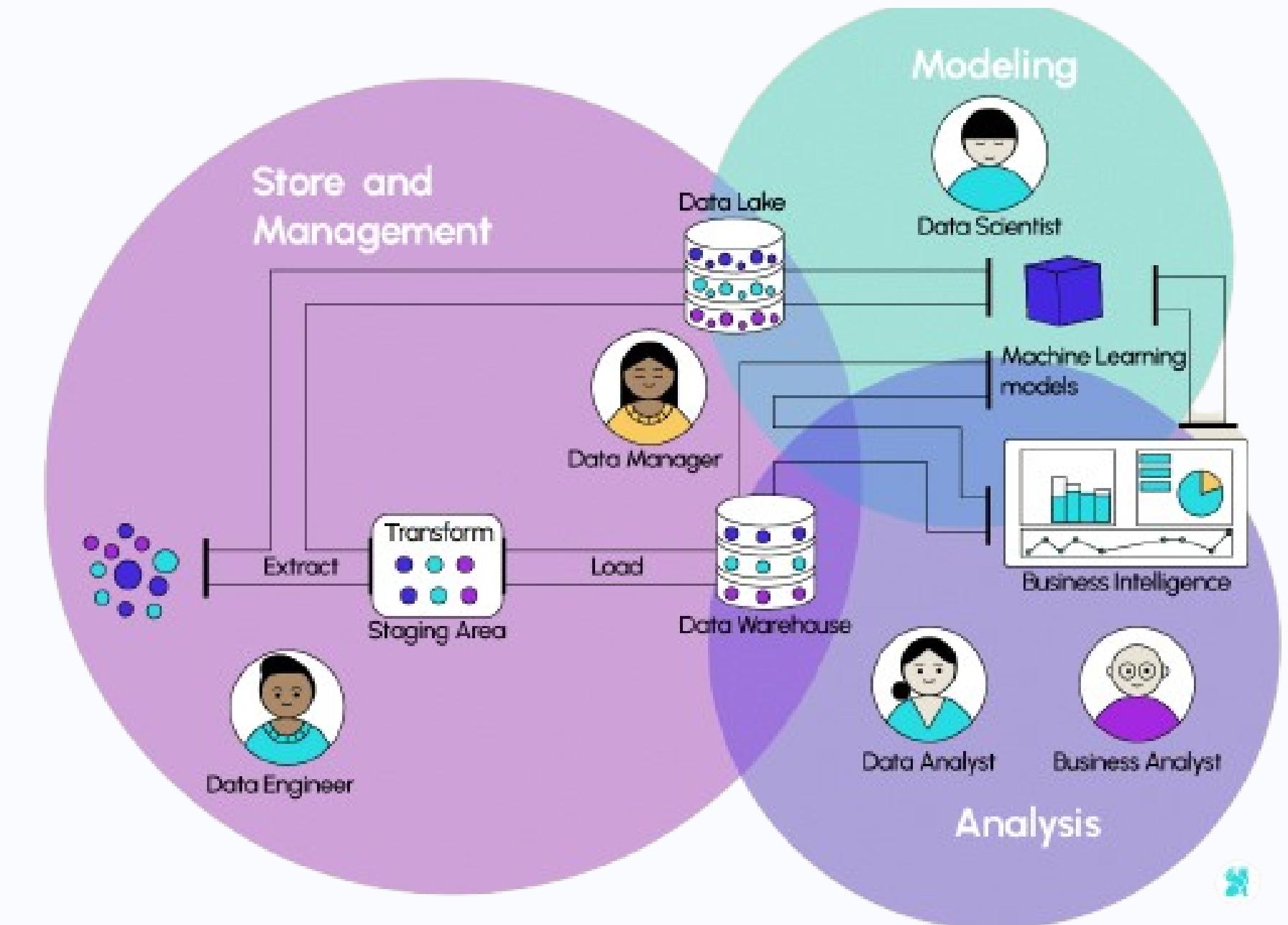
Traduce necesidades del negocio en requerimientos técnicos; prioriza problemas clave. [Alinear los objetivos de datos con las metas empresariales](#)

- Data Engineer

Diseña pipelines, gestiona infraestructura de datos y asegura la calidad de los mismos. [Preparar y almacenar datos para su análisis.](#)

- Data Scientist

Desarrolla modelos predictivos y prescriptivos utilizando modelos estadísticos y técnicas avanzadas de Machine Learning. **Generar insights y predicciones complejas para resolver problemas específicos.**

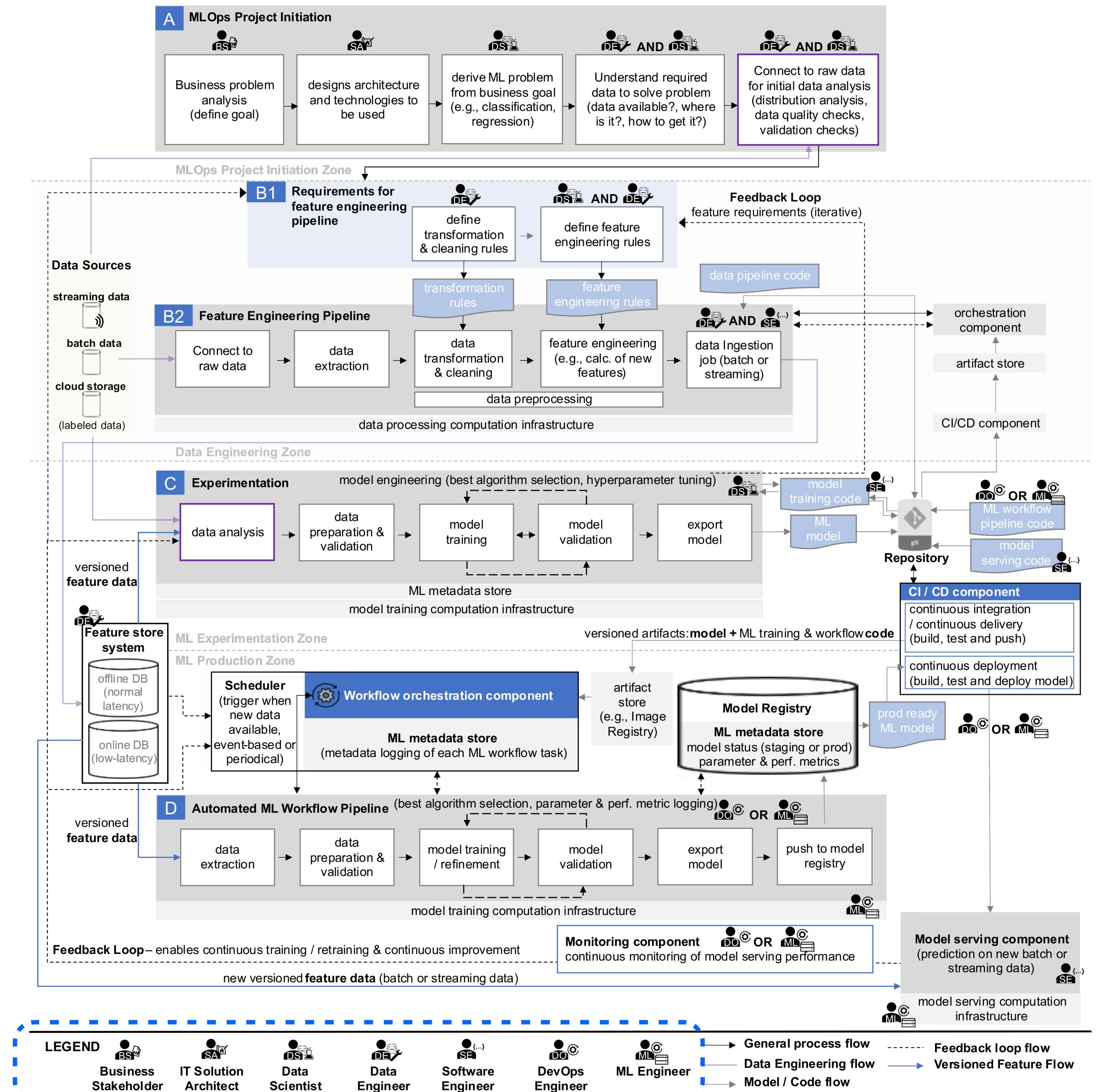


1. Flujo de trabajo de MLOps y roles clave

Este diagrama representa el flujo completo de MLOps (Machine Learning Operations), desde la definición del problema hasta la implementación del modelo en producción, destacando los roles involucrados en cada etapa y las actividades clave.

MLOps integra personas, procesos y herramientas para garantizar un desarrollo continuo, confiable y escalable de sistemas de machine learning.

Fuente: "End-to-end MLOps architecture and workflow with functional components and roles," ArXiv, 2022.



¿Qué será Ciencia de Datos?

Ser un experto en Python o R

Dominar la estadística avanzada y matemática

Hacer modelos de Machine Learning todo el tiempo

Trabajar solo con grandes volúmenes de datos (Big Data)

Visualizar datos bonitos con gráficos llamativos

Un científico de datos solo necesita saber programar bien

Los científicos de datos son magos que predicen el futuro

Cualquier trabajo relacionado con datos es Ciencia de Datos

Implica trabajar únicamente con inteligencia artificial

Ciencia de Datos

Solo es útil para científicos y grandes empresas tecnológicas

Es simplemente programar algoritmos

2. ¿Qué es ciencia de Datos?

"La Ciencia de Datos es un campo que combina herramientas, técnicas y principios matemáticos, estadísticos e informáticos para extraer conocimiento y valor de datos estructurados y no estructurados. Se centra en la preparación de datos, el análisis predictivo, el aprendizaje automático y la toma de decisiones informada."

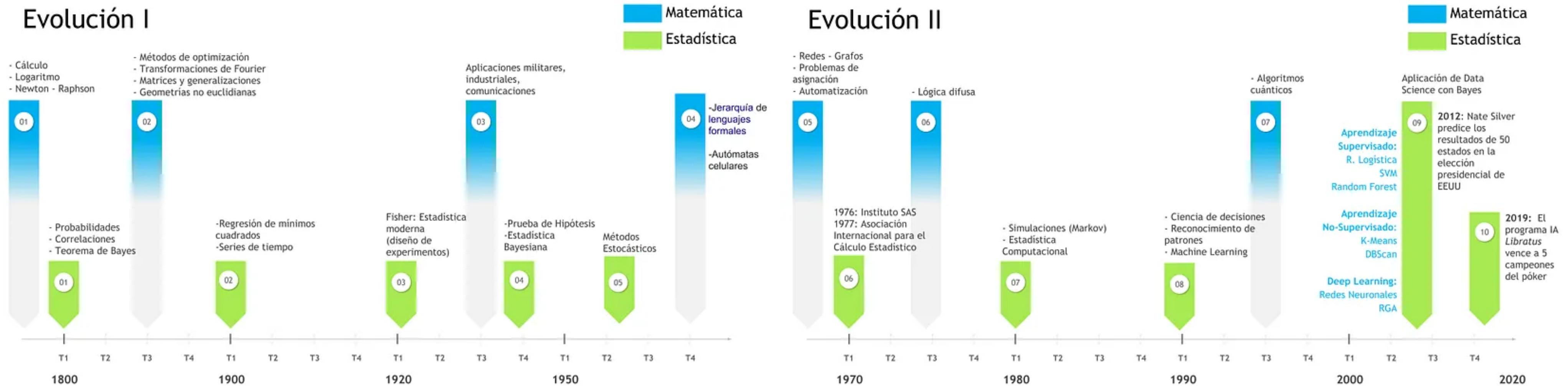
Gartner

"La Ciencia de Datos es el proceso de descubrir patrones y conocimientos en datos mediante el uso de métodos como la estadística, el aprendizaje automático y la analítica avanzada. Esto permite a las organizaciones tomar decisiones informadas basadas en datos."

Microsoft

"La Ciencia de Datos es la disciplina que utiliza datos y herramientas avanzadas de analítica, aprendizaje automático y visualización para descubrir patrones y resolver problemas complejos en diversas industrias."

AWS



Para qué aprender Data Science?

1. Relevancia en casi todos los sectores
2. Alta demanda laboral y buenas perspectivas profesionales
3. Toma de decisiones basadas en datos
4. Desarrollo de habilidades valiosas
5. Poder resolver problemas reales
6. Adaptabilidad al Futuro
7. Innovación y creatividad



Fintech: Detección de fraudes



Gaming: Análisis o efectos visuales



Salud: Prevención o diagnóstico de enfermedades



Retail o Consumo minorista: Crea mejor experiencia para el cliente recomendando productos o servicios adaptados a los gustos o preferencias , optimiza precios o analiza comportamiento de los consumidores.



Educación: Mejorar la educación adaptando las herramientas



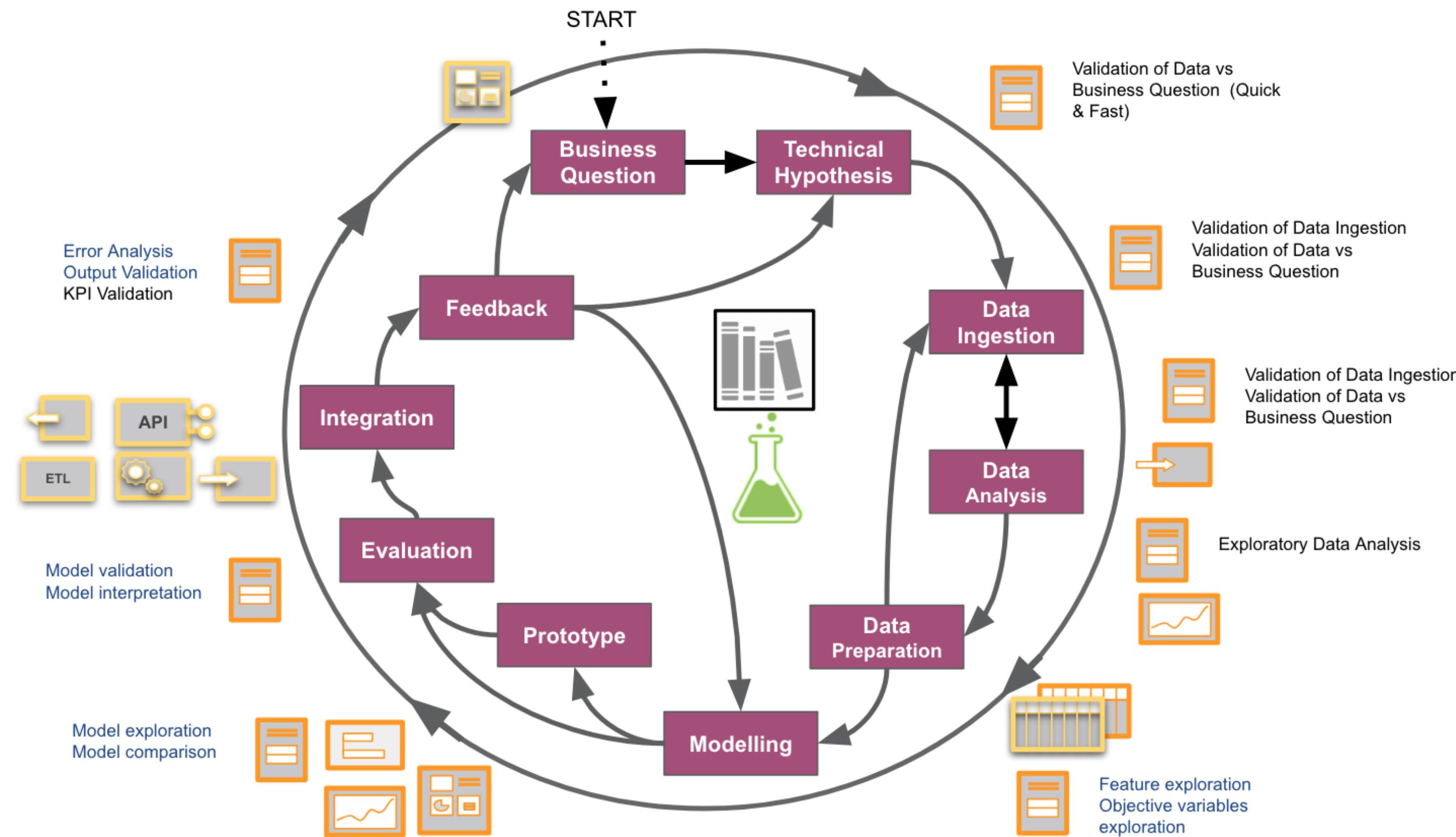
Industria y energía: Mantenimiento preventivo de maquinaria y optimización energética



Logística o transporte: Optimización de rutas y conducción autónoma

Etapas de un Proyecto de Ciencia de Datos

En general estos son los pasos para un proyecto de Ciencia de Datos. Está basado en la metodología CRISP-DM



3. Habilidades esenciales

Habilidades Técnicas para un Data Science



Programación

Python y R son los lenguajes más usados. Conocimientos de bibliotecas como pandas, NumPy, Scikit-learn y TensorFlow son clave.



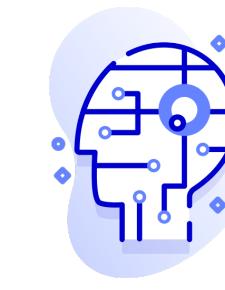
Estadística y Matemáticas

- Probabilidad y Álgebra Lineal
- Modelos Regresionales y de clasificación
- Series Temporales
- Modelos de Supervivencia
- Analítica Multivariante: PCA, MCA, Análisis Factorial, etc..
- Identificar relaciones causales
- Requiere supuestos explícitos sobre la distribución de los datos



Manejo de Bases de Datos

- Bases de Datos Relacionales (SQL)
- Bases NoSQL



Machine Learning

- Supervisado: Árboles de decisión, SVM.
- No Supervisado: Clustering K-means, DBSCAN, Gaussian Mixture Models.
- Deep Learning: Redes neuronales convolucionales (CNNs) y recurrentes (RNNs).
- Maximizar la capacidad predictiva
- no requiere tantas suposiciones sobre la distribución de los datos
- patrones complejos y no lineales

3. Habilidades esenciales

Habilidades Blandas para un Data Science



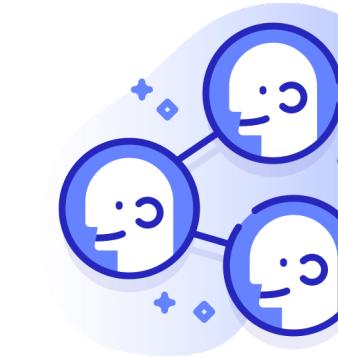
Comunicación Efectiva

Explicar modelos avanzados como PCA o clustering de forma sencilla para stakeholders no técnicos.



Pensamiento crítico

Identificar suposiciones en modelos y cómo afectan las decisiones.



Colaboración

Trabajar con equipos multidisciplinarios

ROADMAP DE APRENDIZAJE

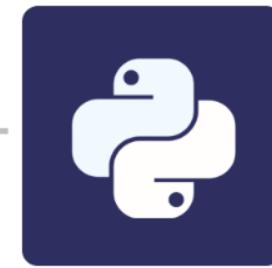


72 HORAS

(3h por sesión, dos veces a la semana)



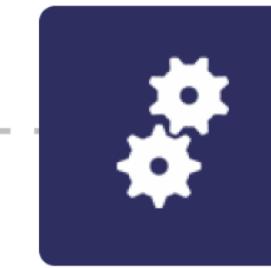
INTRODUCCIÓN A
LA CIENCIA DE
DATOS



PYTHON PARA CIENCIA
DE DATOS



VISUALIZACIÓN Y
ANÁLISIS DE
DATOS



TRANSFORMACIÓN DE
DATOS Y WOEs



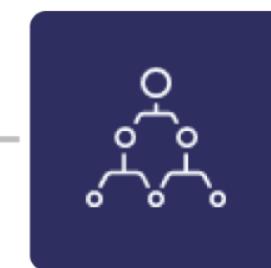
ANÁLISIS
EXPLORATORIO DE
DATOS



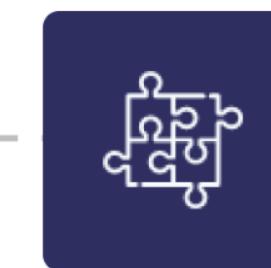
MODELOS DE
REGRESIÓN



MODELO LOGIT &
KNN



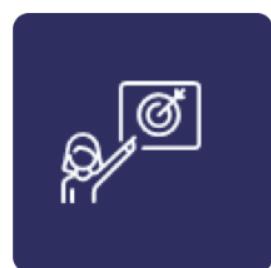
TECNICAS DE
CLASIFICACIÓN
(RANDOM FOREST)



MODELOS DE
SEGMENTACIÓN
(K-MEANS Y CLUSTER JERÁRQUICO)



COMPONENTES
PRINCIPALES Y REDUCCIÓN
DE DIMENSIONES



SUPPORT VECTOR
MACHINE & MODELOS
ENSAMBLADOS



MARKET BASKET
ANALYSIS



TEXT MINING &
DEEP LEARNING



ANÁLISIS DE SERIES
DE TIEMPO



COMPARACIÓN DE
MODELOS



CASOS DE ESTUDIO

FINANCIERO, SEGUROS Y COBRANZAS

- Marketing Bancario
- Predictor de Ingresos
- Credit Scoring
- Fuga de Clientes Telco
- Seguro: Pago de Primas de renovación
- Contactabilidad Cobranzas

RETAIL

- Big Mart Analytics
- Black Friday

SALUD Y RECURSOS HUMANOS

- Health Analytics
- Human Resources Analytics

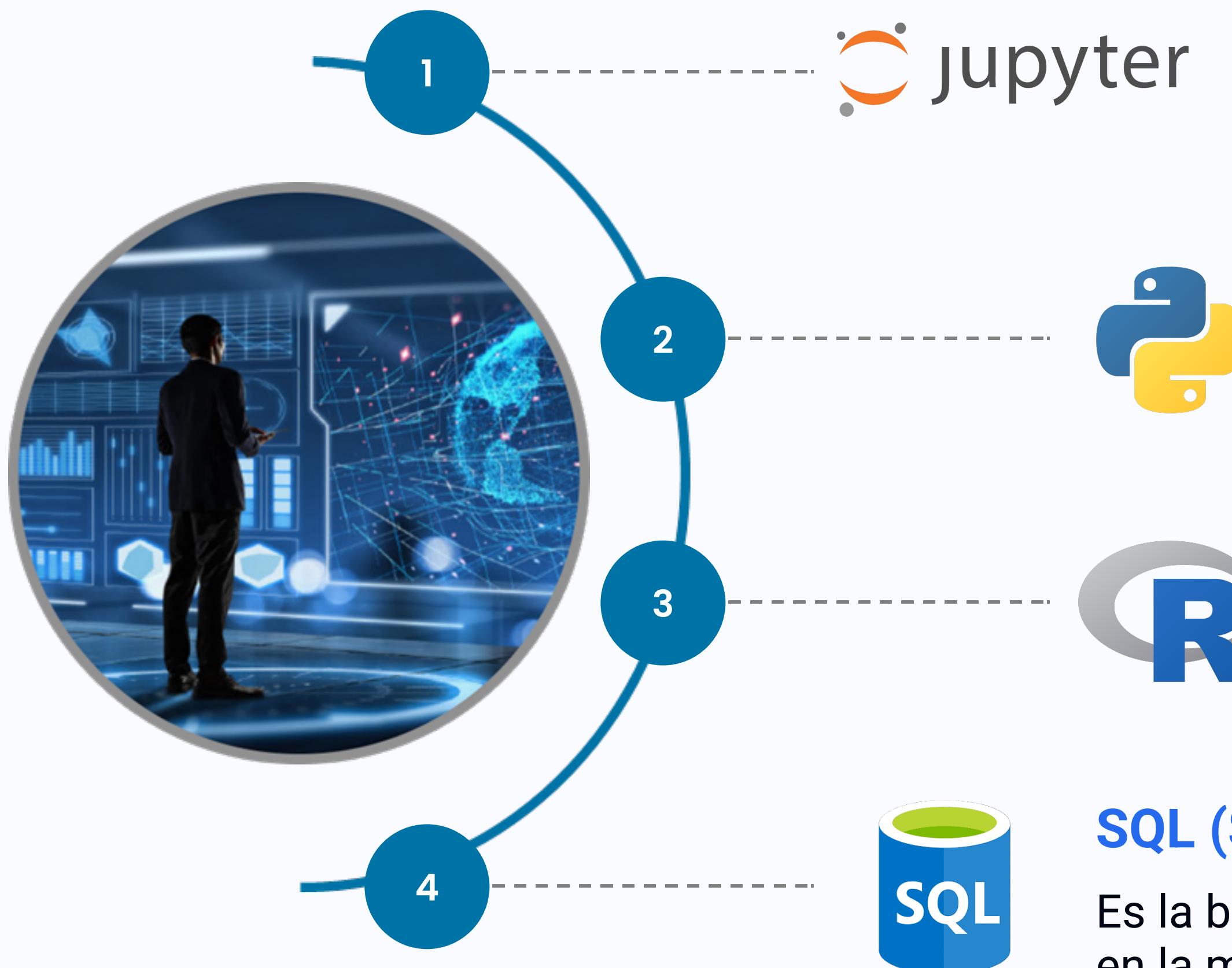
INMOBILIARIO

- Análisis de Precios de Viviendas
- Adquisición de Créditos Hipotecarios



4. Herramientas y tecnologías clave

Análisis y modelado



Jupyter Notebooks

Integración de código, gráficos y notas.

Python y sus librerías avanzadas

- **Statsmodel**: Para modelos estadísticos avanzados y pruebas de hipótesis.
- **Scikit-learn**: Machine Learning estándar.

R

Especialmente para técnicas multivariantes como ANOVA, análisis discriminante y regresión.

SQL (SQL server / Mysql / Postgresql)

Es la base para manejar y consultar datos estructurados en la mayoría de los proyectos.

4. Herramientas y tecnologías clave

Procesamiento de Big Data



Apache Spark y Hadoop

Procesamiento distribuido para manejar grandes volúmenes.

Big query / Athena

- consulta eficiente en la nube.

4. Herramientas y tecnologías clave

Visualización de Datos y Business Intelligence

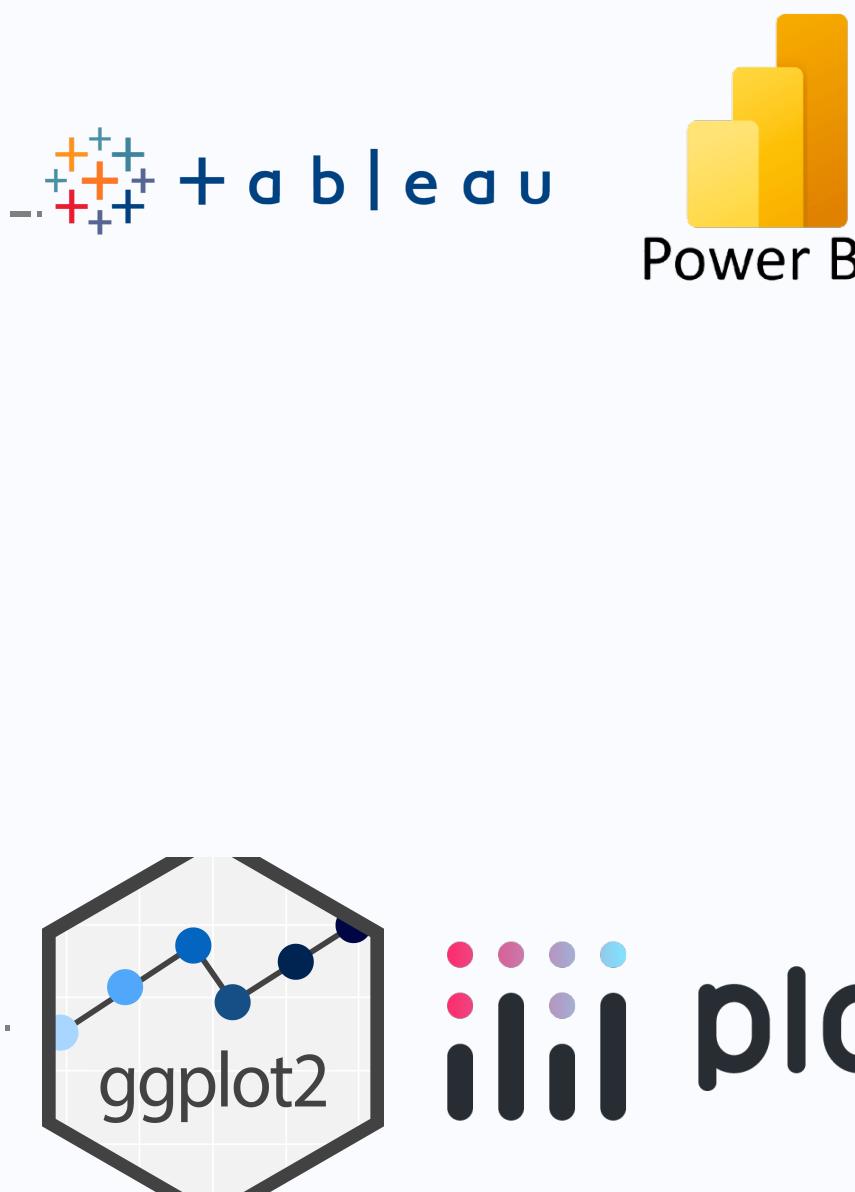


Tableau y Power BI

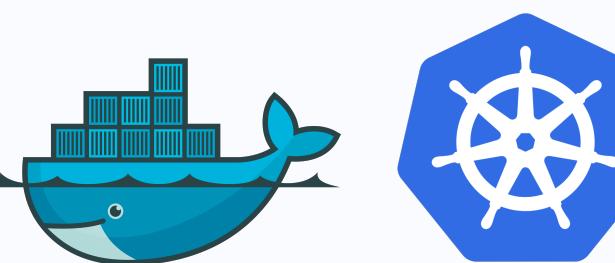
Visualización interactiva para stakeholders.

Librerías de visualización

Para gráficos estadísticos personalizados

4. Herramientas y tecnologías clave

Despliegue y MLOPS



Docker y Kubernetes

Contenerización y orquestación de modelos.



MLflow

Seguimiento de experimentos y despliegue de modelos.



Github

Control de versiones para scripts y notebooks.

5. Construye tu línea de Carrera

Qué es lo que haría yo?

Etapas clave



Aprende Python, SQL y visualización (Power BI o Tableau)



Especialízate en estadística, analítica multivariante y machine learning



Construye un portafolio de proyectos.

- Proyectos de Kaggle, Datos abiertos Perú, etc...
- Publica tus proyectos en GitHub con una estructura modular y limpia

Certificaciones y Conexión con la comunidad



Cursos en Coursera o edX. Por ejemplo:

- MicroMasters® Program in Statistics and Data Science



- Participa en conferencias o eventos relacionados con Data
- Comparte tus análisis o visualizaciones en Linkedin
- Participa en Datathones, aún cuando no tengas experiencia. Apóyate de gente que si lo tenga

Gracias



Carlos Rondan

