

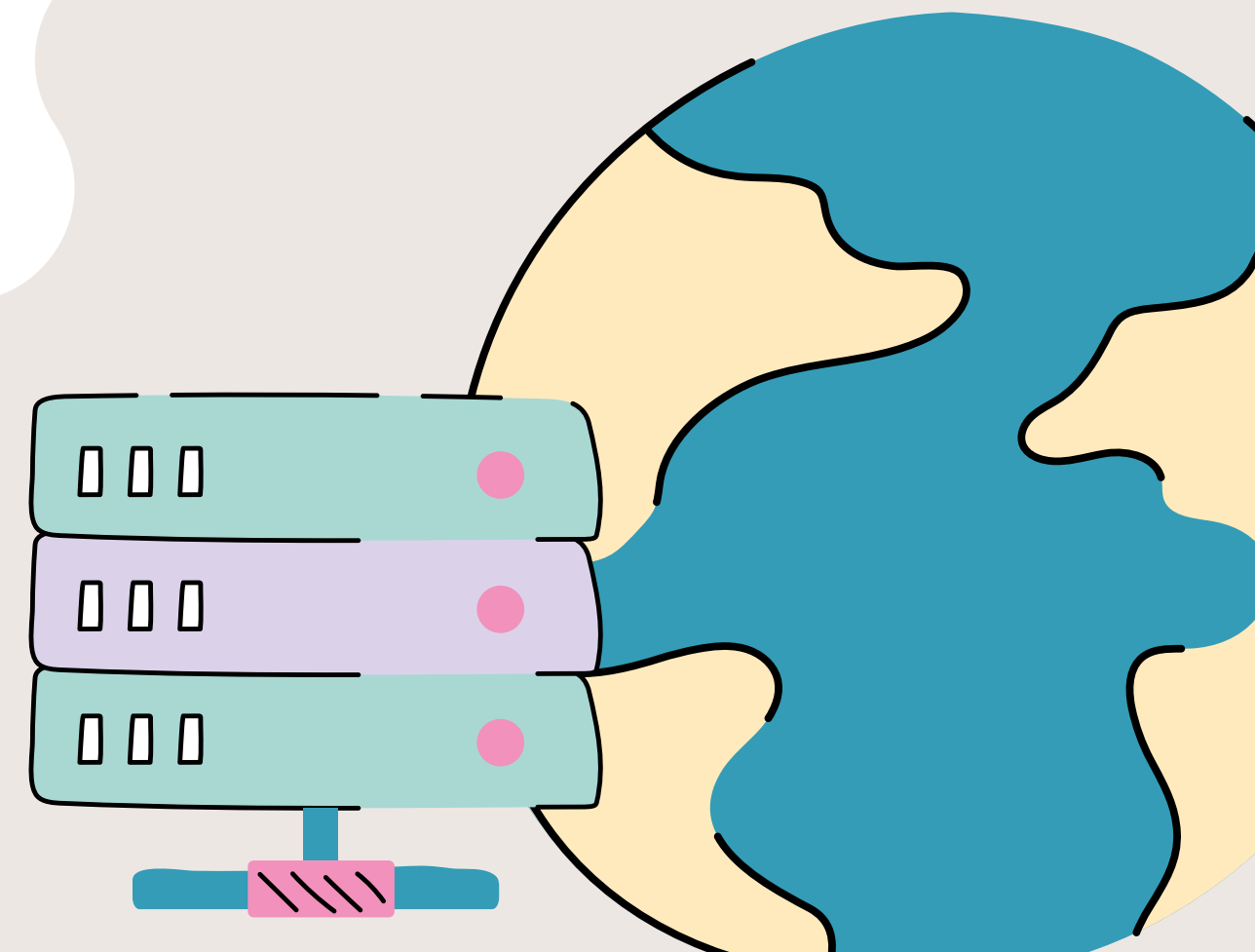


Roadmap Data Scientist

**Guía esencial para tu carrera como Data
Scientist**



Por Juan Duran



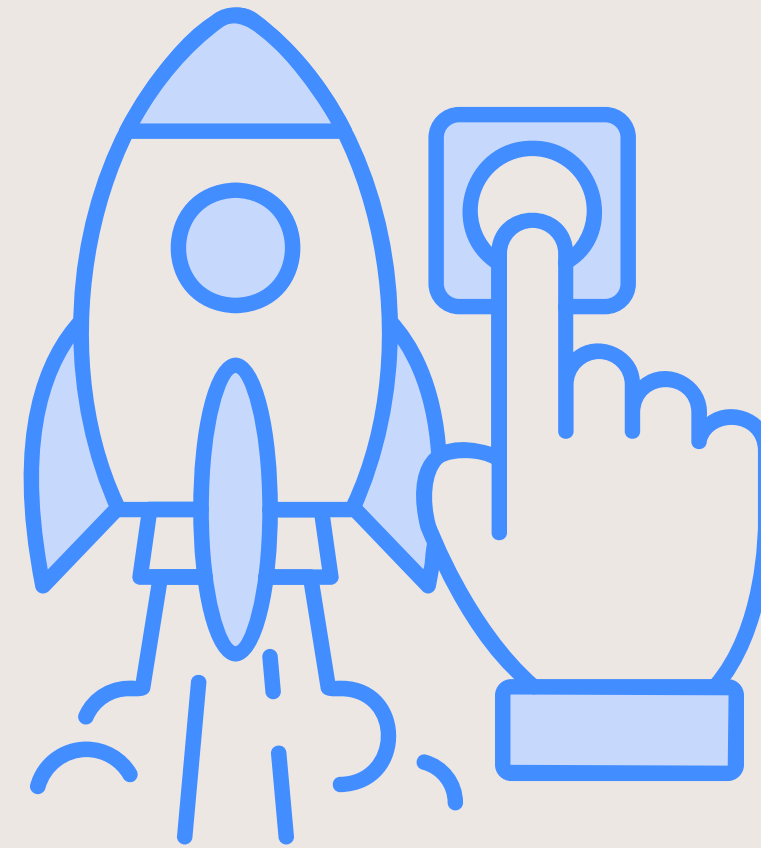
Introducción

Bienvenido a tu viaje para convertirte en **Data Scientist!** 🌟

Durante esta presentación exploraremos los pasos clave, **habilidades técnicas y blandas** necesarias, así como el **desarrollo profesional** en este campo. La ciencia de datos combina estadísticas, programación y conocimiento del dominio para extraer información valiosa.

Los Data Scientists transforman **datos** en información accionable, ayudando a las empresas a tomar **decisiones** informadas y descubrir nuevas oportunidades. Este camino puede parecer desafiante, pero con la guía adecuada y un enfoque estructurado, es posible alcanzar el éxito.

La ciencia de datos implica **creatividad, pensamiento crítico** y la capacidad de **comunicar** hallazgos de manera efectiva. Exploraremos oportunidades de desarrollo profesional, incluyendo certificaciones, cursos y proyectos personales para destacar en este campo competitivo.



Puntos clave



Fundamentos

Comprender **estadística**, **probabilidad** y análisis de datos es esencial. Familiarizarse con **Python** y **R** es crucial. La limpieza y preparación de datos son fundamentales.



Modelado y Algoritmos

Estudiar **algoritmos** de machine learning es vital. Practicar con **Scikit-learn** y **TensorFlow** mejora la precisión. Evaluar y ajustar modelos asegura robustez.



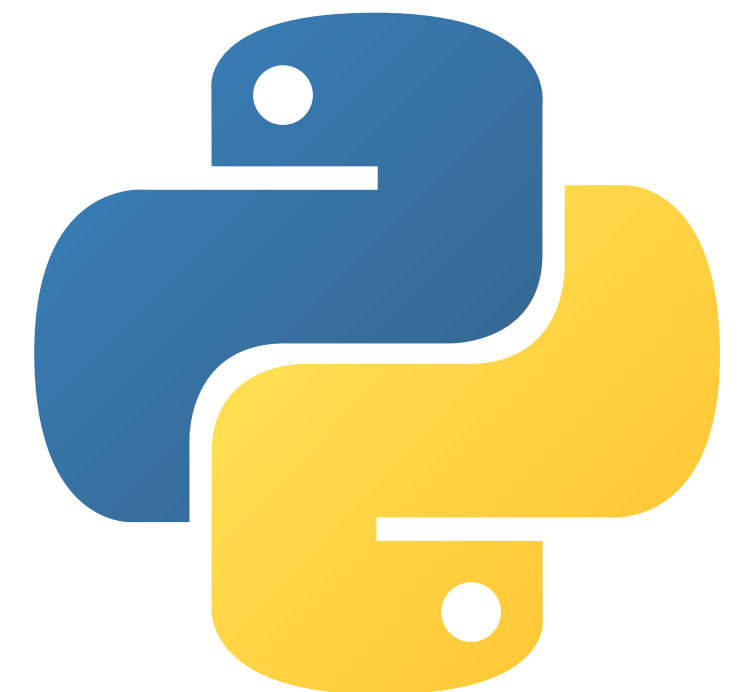
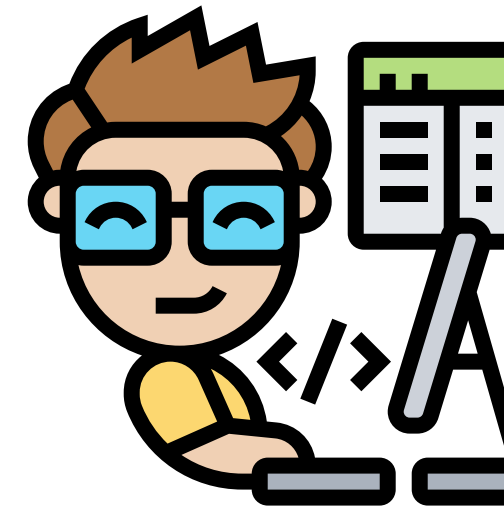
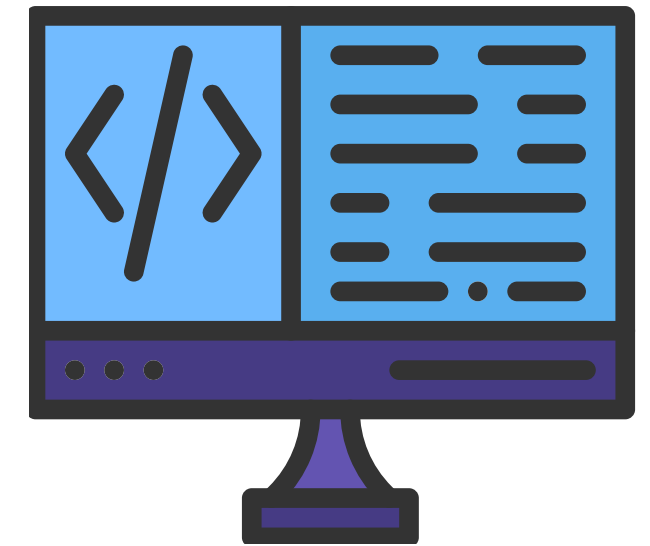
Visualización

Desarrollar habilidades de **presentación** es crucial. Utilizar herramientas como **Tableau** y **Power BI** facilita la interpretación. Contar historias con datos mejora la comunicación.

Programación en Python

Python es una herramienta esencial en la ciencia de datos por su **simplicidad** y **versatilidad**.

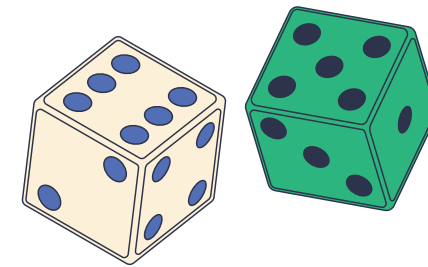
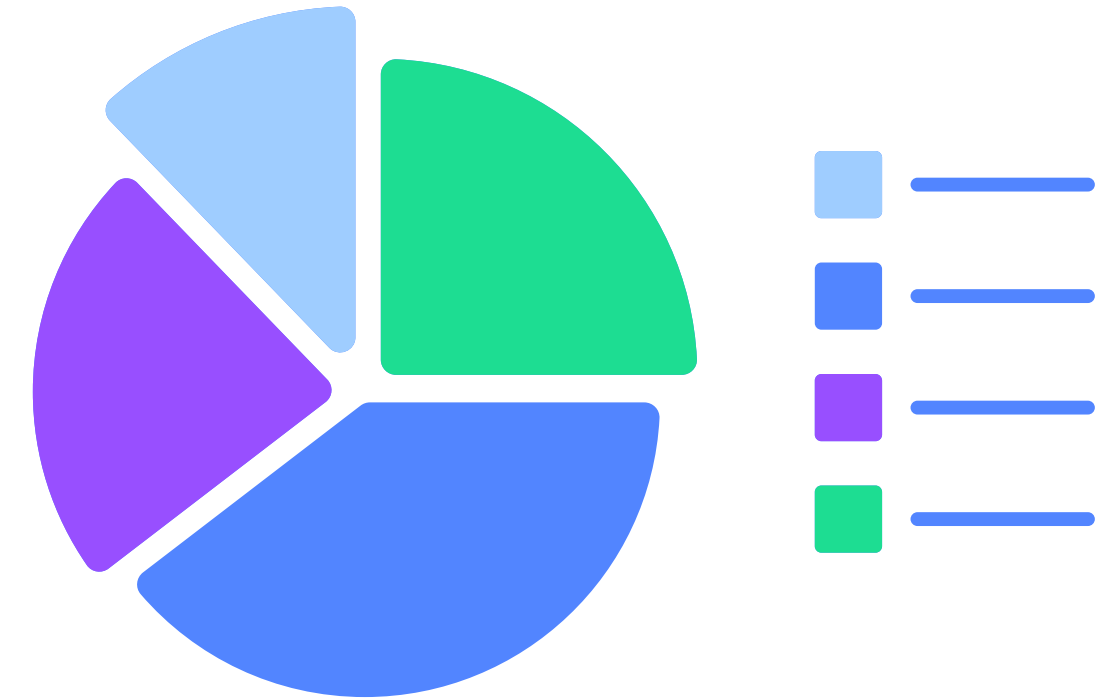
1. Python es el **lenguaje** más utilizado en la ciencia de datos debido a su simplicidad y versatilidad.
2. Aprender a escribir **scripts** eficientes y a utilizar bibliotecas como Pandas, NumPy y Scikit-learn es fundamental.
3. Pandas se utiliza para la **manipulación** y **análisis** de datos, NumPy para operaciones matemáticas y Scikit-learn para implementar algoritmos de machine learning.
4. Python también es **compatible** con bibliotecas de visualización como Matplotlib y Seaborn, lo que facilita la creación de gráficos y visualizaciones.



Estadística y probabilidad

La **estadística** y la **probabilidad** son la base para analizar datos y construir modelos precisos.

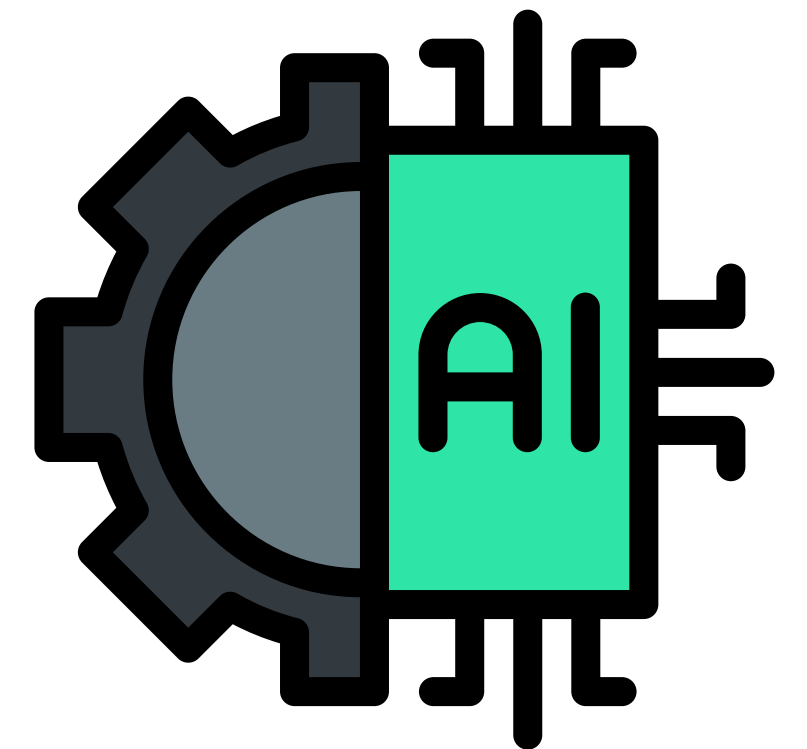
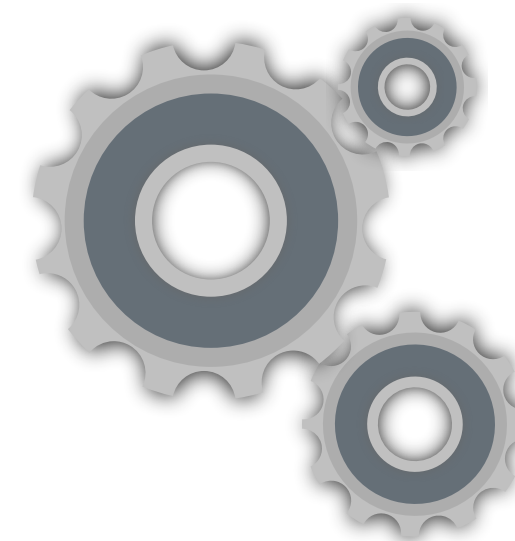
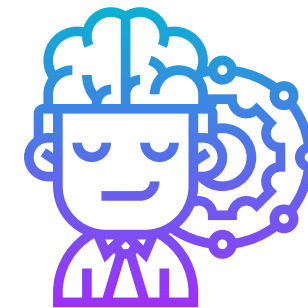
1. Comprender los conceptos fundamentales de la estadística y la probabilidad es esencial para analizar datos y construir **modelos** precisos.
2. Esto incluye conocimientos sobre **distribuciones**, pruebas de hipótesis y análisis de regresión.
3. La estadística **descriptiva** y la **inferencial** son herramientas clave para interpretar y comunicar los resultados del análisis de datos.
4. La **probabilidad** es fundamental para entender y aplicar **algoritmos** de machine learning, especialmente en la evaluación de modelos.



Machine Learning

El machine learning permite a los sistemas **aprender** y mejorar a partir de los **datos**.

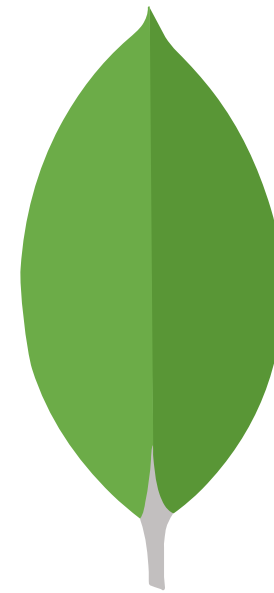
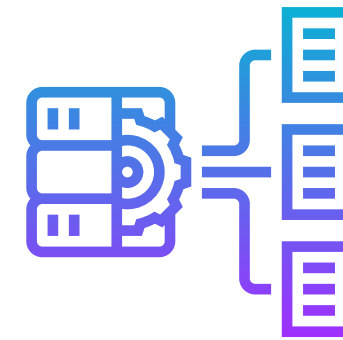
1. Aprender sobre los diferentes algoritmos de machine learning, cómo **implementarlos** y cómo evaluar su rendimiento es crucial.
2. Esto incluye algoritmos **supervisados** y **no supervisados**, así como técnicas de deep learning.
3. La **implementación** práctica de estos algoritmos en proyectos reales ayuda a consolidar el conocimiento y a desarrollar habilidades aplicadas.
4. Evaluar y mejorar el **rendimiento** de los modelos mediante técnicas como la validación cruzada y el ajuste de hiperparámetros asegura que los modelos sean robustos.



Bases de datos y SQL

El manejo de **bases de datos** es crucial para trabajar con grandes **volúmenes** de datos.

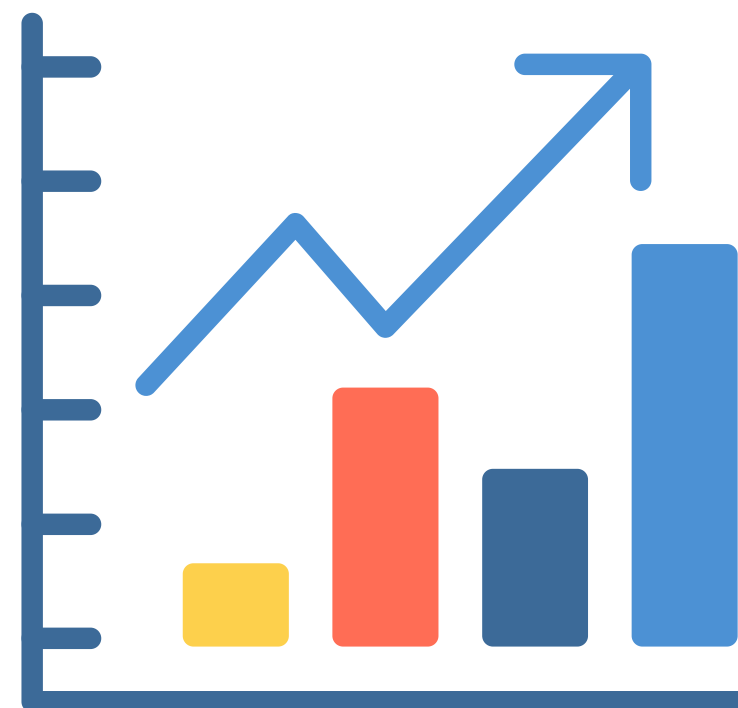
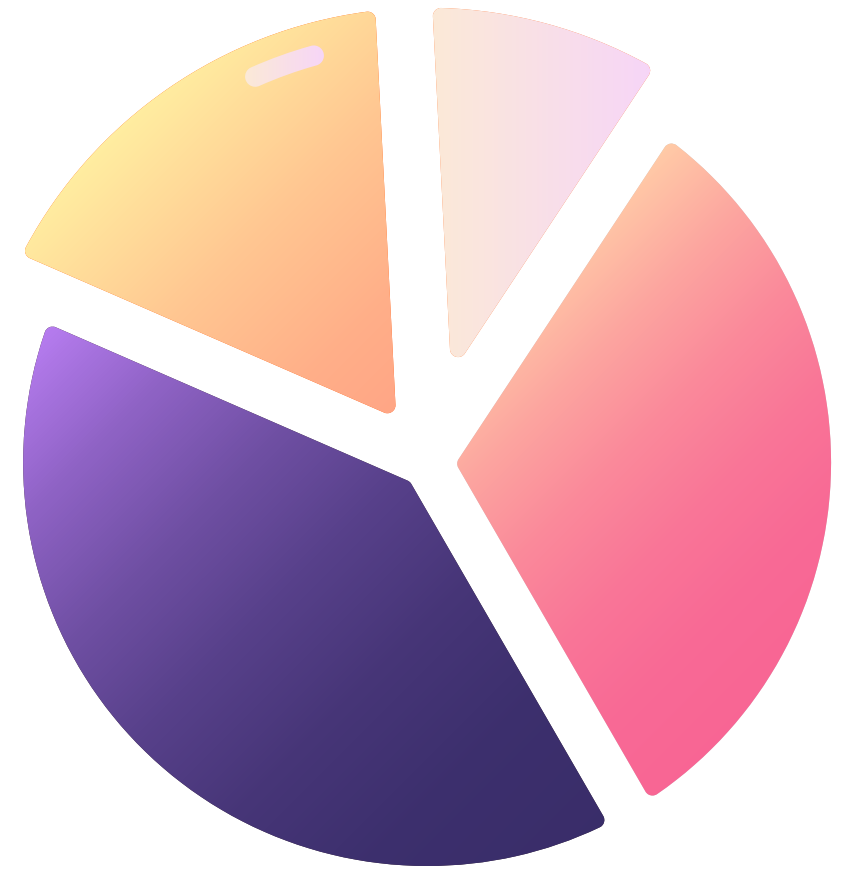
1. Saber cómo manejar y **consultar** bases de datos es crucial para trabajar con grandes volúmenes de datos.
2. **SQL** es el **lenguaje** estándar para interactuar con bases de datos relacionales.
3. La capacidad de escribir **consultas** eficientes y optimizar el rendimiento de las bases de datos es una habilidad muy valorada en la industria.
4. Conocer bases de datos **NoSQL**, como **MongoDB**, también puede ser útil para trabajar con datos no estructurados.



Visualización de datos

La **visualización** de datos es clave para comunicar hallazgos de manera efectiva.

1. Desarrollar habilidades para crear visualizaciones impactantes que **comuniquen** claramente los hallazgos y tendencias en los datos.
2. Esto incluye el uso de gráficos, mapas y **dashboards** interactivos.
3. **Herramientas** como Matplotlib, Seaborn y Plotly en Python son ampliamente utilizadas para la visualización de datos.
4. La capacidad de **contar historias con datos** ayuda a contextualizar los hallazgos y a hacerlos más accesibles para audiencias no técnicas.



Soft Skills

Comunicación Efectiva

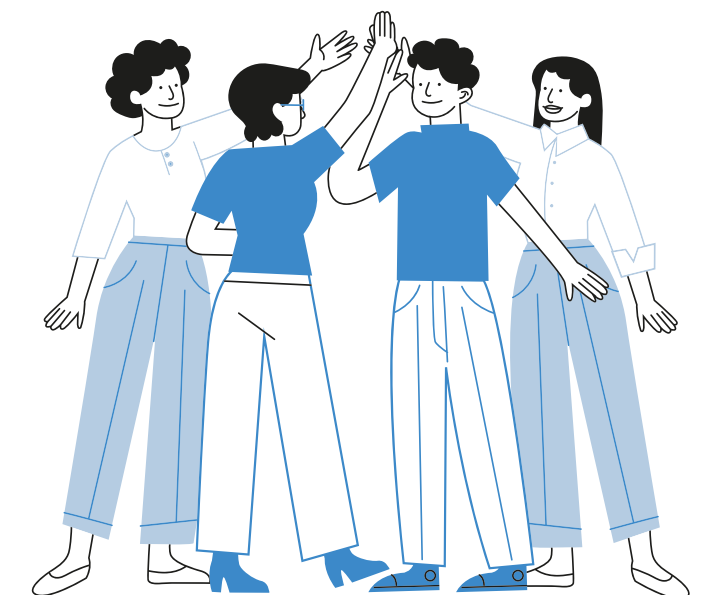
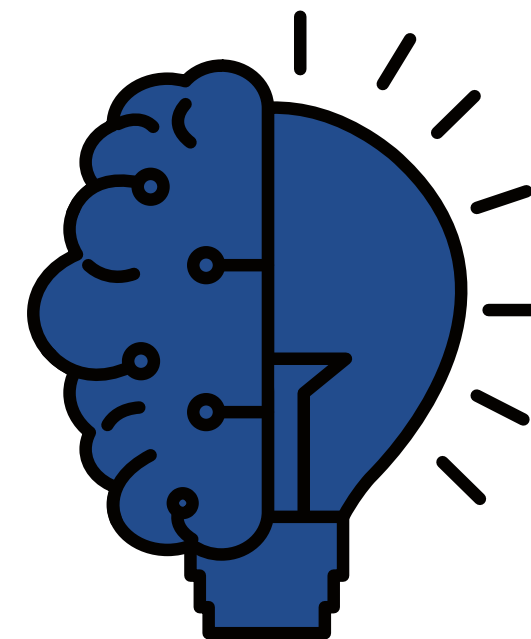
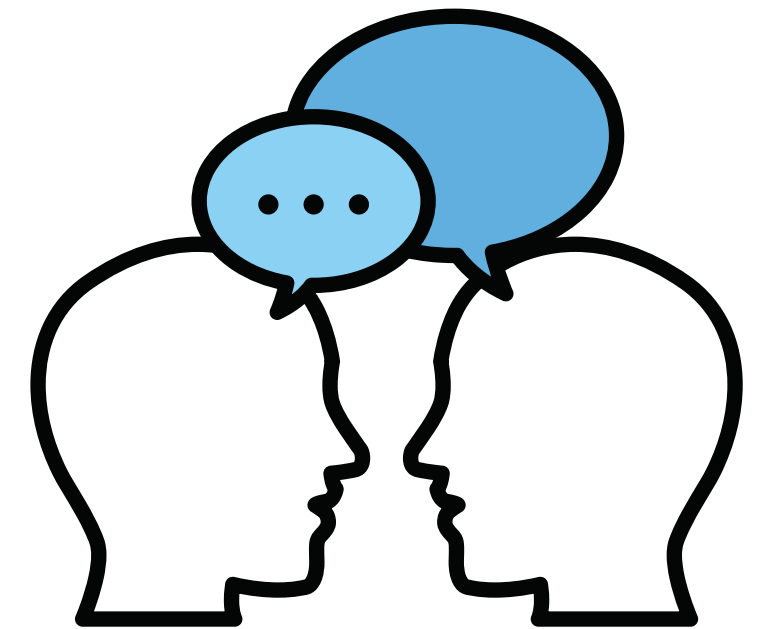
Ser capaz de explicar conceptos complejos de manera sencilla y clara es esencial para trabajar en equipo y presentar hallazgos a audiencias no técnicas.

Pensamiento Crítico

Evaluar problemas y soluciones de manera lógica y objetiva es crucial para tomar decisiones informadas en la ciencia de datos.

Trabajo en Equipo

Colaborar eficazmente con otros profesionales y equipos multidisciplinarios es esencial para el éxito en proyectos de ciencia de datos.

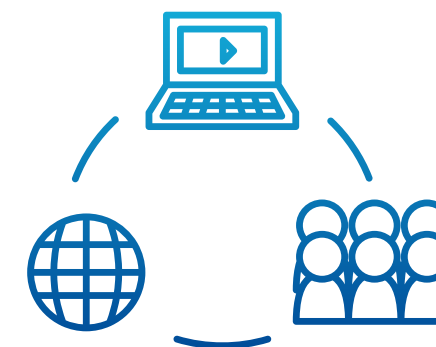


Desarrollo profesional (1)

Certificaciones y cursos 🎓

Obtener **certificaciones** reconocidas en ciencia de datos y machine learning, como las ofrecidas por **Coursera**, **edX** y **Udacity**, puede mejorar tu perfil profesional y demostrar tu compromiso con el aprendizaje continuo.

Los cursos en línea y los **bootcamps** también son excelentes maneras de adquirir nuevas habilidades y mantenerse actualizado con las últimas tendencias y tecnologías. Además, estas certificaciones pueden abrir puertas a nuevas **oportunidades** laborales y aumentar tu credibilidad en el campo. Invertir en tu educación continua es clave para mantenerte competitivo en el mercado laboral.



Desarrollo profesional (2)

Proyectos personales

Trabajar en proyectos personales para aplicar y demostrar tus habilidades es una excelente manera de construir un **portafolio**. Estos proyectos pueden incluir análisis de datos públicos, participación en competencias de **Kaggle** y contribuciones a proyectos de código abierto. Documentar y compartir tus proyectos en plataformas como **GitHub** y **LinkedIn** puede ayudarte a destacar ante posibles empleadores.

Además, estos proyectos te permiten experimentar con nuevas **técnicas** y herramientas, y demostrar tu capacidad para resolver problemas del mundo real. Un portafolio sólido es una excelente manera de mostrar tu experiencia y habilidades a futuros empleadores.



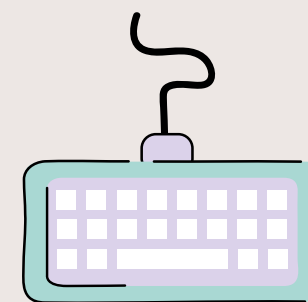
Desarrollo profesional (3)

Networking

Conectar con otros profesionales del campo y participar en comunidades y eventos de ciencia de datos es crucial para el desarrollo profesional. Esto incluye asistir a conferencias, **meetups** y webinars, así como participar en foros y grupos en línea.

El networking te permite **aprender** de otros, **compartir** conocimientos y **descubrir** oportunidades de colaboración y empleo. Además, construir una red de contactos sólida puede proporcionarte mentores y colegas que te apoyen en tu crecimiento profesional. Participar activamente en la comunidad de ciencia de datos te mantiene informado sobre las últimas **tendencias** y desarrollos en el campo.





Gracias



Por Juan Duran

“Coding, Gaming and Leveling Up”