

Data Scientist Roadmap 2024

[Guide here](#)

Welcome to your comprehensive **Data Science Roadmap**! If you've ever wondered, about "**Steps or Path to Become a Data Scientist**", you're in the right place. This guide is perfect for **Data Science for Beginners** and seasoned professionals alike, covering everything from mastering **Python for Data Science** and **R for Data Science**, to understanding the importance of **Data Cleaning** and **Data Visualization**.



Table of Content

- [What is Data Science?](#)
- [Why Data Science?](#)
- [What Does a Data Scientist Do?](#)
- [Why Become a Data Scientist](#)
- [Skills Required to Become a Data Scientist](#)
- [Data Scientist vs Data Analyst](#)
- [Average Salary of a Data Scientist](#)
- [Data Scientist Roadmap\[2024\]](#)
- [1\) Mathematics](#)
- [2\) Probability](#)
- [3\) Statistics](#)
- [4\) Programming](#)

- [5\) Machine Learning](#)
- [6\) Deep Learning](#)
- [7\) Feature Engineering](#)
- [8\) Natural Language Processing](#)
- [9\) Data Visualization Tools](#)
- [10\) Deployment](#)
- [11\) Other Points to Learn](#)
- [12\) Keep Practicing](#)
- [Data Scientist Roadmap: Education Routes](#)

What is Data Science?

Data science is the field of study that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data. It combines various disciplines such as statistics, machine learning, data analysis, and visualization to uncover hidden patterns, trends, and correlations in data. Data science plays a crucial role in decision-making, forecasting, and problem-solving across industries, driving innovation and enabling organizations to make data-driven decisions.**.**

So briefly it can be said that Data Science involves:

- Statistics, computer science, mathematics
- Data cleaning and formatting
- Data visualization

Nowadays it is known to everyone how popular is **Data Science**. Now the questions that arise are, **Why Data Science?, how to start? Where to start? What topics one should cover? etc._** Do you need to learn all the concepts from a book or you should go with some online tutorials or you should learn Data Science by doing some projects on it? So in this article, we are going to discuss all these things in detail.

¿Qué es la ciencia de datos?

Ciencia de datos es el campo de estudio que utiliza métodos científicos, procesos, algoritmos y sistemas para extraer conocimiento e información de datos estructurados y no estructurados. Combina varias disciplinas como la estadística, el aprendizaje automático, el análisis de datos y la visualización para descubrir patrones, tendencias y correlaciones ocultas en los datos. La ciencia de datos desempeña un papel crucial en la toma de decisiones, la previsión y la resolución de problemas en todas las industrias, impulsando la innovación y permitiendo a las organizaciones tomar decisiones basadas en datos.

En pocas palabras, se puede decir que la ciencia de datos implica:

- Estadística, informática, matemáticas
- Limpieza y formato de datos
- Visualización de datos

Hoy en día todo el mundo sabe lo popular que es la **ciencia de datos**. Ahora las preguntas que surgen son: **¿Por qué la ciencia de datos?, ¿cómo empezar? ¿Por dónde empezar? ¿Qué temas se deben cubrir? etc.** **¿Necesitas aprender todos los conceptos de un libro o deberías recurrir a algunos tutoriales en línea o deberías aprender ciencia de datos haciendo algunos proyectos al respecto?** En este artículo, analizaremos todas estas cuestiones en detalle.

Why Data Science?

So before jumping into the complete Roadmap of Data Science, one should have a clear goal in their mind about why they want to learn Data Science. Is it for the phrase “**The Sexiest Job of the 21st Century**”? Is it for your college academic projects? or is it for your long-term career? or do you want to switch your career to the data scientist world? So first make a clear goal.

Why do you want to learn Data Science? For example, if you want to learn Data Science for your college **Academic projects** then it's enough to just learn the **beginner things in Data Science**. Similarly, if you want to build your **long-term career** then you should learn **professional** or **advanced things** also. You have to cover all the prerequisite things in detail. So it's in your hand and it's your decision why you want to learn Data Science.

¿Por qué Ciencias de Datos?

Entonces, antes de saltar a la hoja de ruta completa de la Ciencia de Datos, uno debe tener un objetivo claro en mente sobre por qué quiere aprender Ciencias de Datos. ¿Es por la frase “**El Trabajo Más Sexy del Siglo XXI**”? ¿Es para tus proyectos académicos universitarios? ¿O es para tu carrera a largo plazo? ¿O quieres cambiar tu carrera al mundo de la ciencia de datos? Entonces, primero establece un objetivo claro.

¿Por qué quieres aprender Ciencias de Datos? Por ejemplo, si quieres aprender Ciencias de Datos para tus **proyectos académicos** universitarios, entonces es suficiente con aprender las **cosas iniciales en Ciencias de Datos**. Del mismo modo, si quieres construir tu **carrera a largo plazo**, entonces también debes aprender **cosas profesionales** o **cosas avanzadas**. Tienes que cubrir todos los requisitos previos en detalle. Entonces, está en tus manos y es tu decisión por qué quieres aprender Ciencias de Datos.

What Does a Data Scientist Do?

A Data Scientist gathers and analyzes complex data to guide business decisions. They collect, clean, and explore data, develop machine learning models, and deploy them for real-world use.

Data Scientists also monitor and maintain models, communicate findings to non-technical stakeholders, and collaborate across teams to align with organizational goals.

¿Qué hace un científico de datos?

Un científico de datos recopila y analiza datos complejos para orientar las decisiones comerciales. Recopila, limpia y explora datos, desarrolla modelos de aprendizaje automático y los implementa para su uso en el mundo real.

Los científicos de datos también supervisan y mantienen modelos, comunican los hallazgos a las partes interesadas no técnicas y colaboran entre equipos para alinearse con los objetivos de la organización.

Why Become a Data Scientist

In the global landscape, **data is the new oil, driving innovation and reshaping industries**. Organizations crave skilled professionals to extract insights from this vast information ocean, and here's where data scientists play a crucial role.

Por qué convertirse en un científico de datos

En el panorama global, los datos son el nuevo petróleo, que impulsa la innovación y transforma las industrias. Las organizaciones necesitan profesionales capacitados para extraer información de este vasto océano de información, y aquí es donde los científicos de datos desempeñan un papel crucial.

High Demand

- US Bureau of Labor Statistics forecasts a **23% job growth** for data scientists (**2020-2030**), surpassing the average.
- Similar global trends indicate a surge in demand.

Gran demanda

- La Oficina de Estadísticas Laborales de EE. UU. pronostica un crecimiento laboral del 23 % para científicos de datos (**2020-2030**), lo que supera el promedio.
- Tendencias globales similares indican un aumento en la demanda.

Lucrative Salaries

- Handsome rewards for expertise; US data scientists earn over **\$120,000 annually**.
- In India, experienced professionals can fetch upwards of **₹15 lakhs (USD 18,750)**.

Salarios lucrativos

- Recompensas atractivas por la experiencia; los científicos de datos estadounidenses ganan más de **\$120 000 al año**.
- En India, los profesionales experimentados pueden ganar más de **₹15 lakhs (USD 18 750)**.

Impactful Work

- **Tangible societal impact:** Develop algorithms for **disease detection**, **optimize energy grids**, or **predict natural disasters**.

Trabajo de impacto

- **Impacto social tangible:** Desarrollar algoritmos para la **detección de enfermedades**, **optimizar redes de energía** o **predecir desastres naturales**.

Skills Required to Become a Data Scientist

Usually, data scientists come from various educational and work experience backgrounds, and most should be proficient in, or in an ideal case be masters in four key areas.

1. **Domain Knowledge**
2. **Math Skills**
3. **Computer Science**
4. **Communication Skill**

Habilidades necesarias para convertirse en un científico de datos

Por lo general, los científicos de datos provienen de diversos contextos educativos y de experiencia laboral, y la mayoría debería ser competente o, en un caso ideal, tener una maestría en cuatro áreas clave.

1. **Conocimiento del dominio**
2. **Habilidades matemáticas**
3. **Ciencias de la computación**
4. **Habilidades de comunicación**

Domain Knowledge

Most people think that domain knowledge is not important in data science, but it is very important. Let's take an example: If you want to be a data scientist in the banking sector, and you have much more information about the banking sector like stock trading, finance, etc. This is going to be very beneficial for you and the bank itself will give more preference to these types of applicants than a normal applicant.

Conocimiento del dominio

La mayoría de las personas piensa que el conocimiento del dominio no es importante en la ciencia de datos, pero sí lo es. Tomemos un ejemplo: si quieres ser un científico de datos en el sector bancario y tienes mucha más información sobre el sector bancario, como la bolsa de valores, las finanzas, etc. Esto te resultará muy beneficioso y el propio banco dará más preferencia a este tipo de solicitantes que a un solicitante normal.

Math Skills

Linear Algebra, Multivariable Calculus & Optimization Techniques, are three things that are very important as they help us in understanding various machine learning algorithms that play an important role in Data Science. Similarly, understanding **Statistics** is very significant as this is a part of Data analysis. **Probability** is also significant to statistics and it is considered a prerequisite for mastering machine learning.

Habilidades matemáticas

Álgebra lineal, cálculo multivariable y técnicas de optimización son tres cosas muy importantes, ya que nos ayudan a comprender varios algoritmos de aprendizaje automático que desempeñan un papel importante en la ciencia de datos. De manera similar, comprender las **estadísticas** es muy importante, ya que son parte del análisis de datos. La **probabilidad** también es importante para las estadísticas y se considera un requisito previo para dominar el aprendizaje automático.

Computer Science

There is much more to learn in computer science. But when it comes to the programming language one of the major questions that arise is:

Python or R for Data Science?

There are various reasons to choose which language for Data Science as both have a rich set of libraries to implement complex machine learning algorithms, visualization, and data cleaning.

Please refer to [R vs Python in Data Science](#) to know more about this. Knowing both of these languages will provide an extra boost in your career as a data scientist.

Apart from the programming language, the other computer science skills you have to learn are:

- Basics of Data Structure and Algorithm
- SQL
- MongoDB
- Linux
- Git
- Distributed Computing
- Machine Learning and Deep Learning, etc.

Ciencias de la computación

Hay mucho más que aprender en ciencias de la computación. Pero cuando se trata del lenguaje de programación, una de las principales preguntas que surgen es:

¿Python o R para la ciencia de datos?

Hay varias razones para elegir qué lenguaje para la ciencia de datos, ya que ambos tienen un amplio conjunto de bibliotecas para implementar algoritmos complejos de aprendizaje automático, visualización y limpieza de datos. Consulta [R vs Python en la ciencia de datos](#) para obtener más información sobre esto. Conocer ambos lenguajes te dará un impulso adicional en tu carrera como científico de datos.

Además del lenguaje de programación, las otras habilidades de ciencias de la computación que debes aprender son:

- Conceptos básicos de estructura y algoritmos de datos
- SQL
- MongoDB
- Linux
- Git
- Computación distribuida
- Aprendizaje automático y aprendizaje profundo, etc.

Communication Skills

It includes both **written** and **verbal communication**. What happens in a data science project is after concluding the analysis, the project has to be **communicated to others**. Sometimes this

may be a **report** you send to your boss or team at work. Other times it may be a **blog post**. Often it may be a **presentation** to a group of colleagues.

Regardless, a data science project always involves some form of communication of the project's findings. So it's necessary to have communication skills for becoming a data scientist.

Habilidades de comunicación

Incluye tanto la comunicación **escrita** como la **verbal**. Lo que sucede en un proyecto de ciencia de datos es que, después de concluir el análisis, el proyecto debe ser **comunicado a otros**. A veces, esto puede ser un **informe** que envías a tu jefe o equipo en el trabajo. Otras veces puede ser una **publicación de blog**. A menudo puede ser una **presentación** a un grupo de colegas.

De todos modos, un proyecto de ciencia de datos siempre implica alguna forma de comunicación de los hallazgos del proyecto. Por lo tanto, es necesario tener habilidades de comunicación para convertirse en un científico de datos.

Learning Resources

There are plenty of resources and videos available online and it's confusing for someone where to start learning all the concepts. Initially, as a beginner, if you get overwhelmed with so many concepts then don't be afraid and stop learning. Have patience, explore, and stay committed to it.

Some useful learning resource links are available at **GeeksforGeeks**:

- [Mathematics](#)
- [Python](#)
- [R](#)
- [SQL](#)
- [Data Structure](#)
- [Machine Learning](#)

Recursos de aprendizaje

Hay muchos recursos y videos disponibles en línea y puede resultar confuso para alguien saber por dónde empezar a aprender todos los conceptos. Al principio, si eres principiante y te sientes abrumado por tantos conceptos, no temas y debes de aprender. Ten paciencia, explora y mantén el compromiso.

Hay algunos enlaces a recursos de aprendizaje útiles disponibles en **GeeksforGeeks**:

- [Matemáticas](#)
- [Python](#)
- [R](#)
- [SQL](#)
- [Estructura de datos](#)
- [Aprendizaje automático](#)

Data Scientist vs Data Analyst

Here is a quick comparison of Data Scientist and Data Analyst

Aspect	Data Scientist	Data Analyst
Scope	Broader focus: machine learning, predictive modeling.	Focus: analyzing data, and providing insights.
Focus	Uncovering patterns, and predicting trends.	Summarizing historical data, providing insights.
Responsibilities	End-to-end processes, complex models.	Proficient in tools, statistical methods, and reporting.
Tools	Advanced: machine learning, Python/R.	Tools: Excel, Tableau, Power BI.
Data Types	Structured, unstructured, large datasets.	Primarily structured data, occasional smaller sets.
Outcome	Extract actionable insights, and solve complex problems.	Summarize data, and provide insights for decision-making.
Overlap	Some overlap and Analysts contribute to the early stages.	Distinct roles, potential for collaboration.

Científico de datos vs. analista de datos

A continuación, se muestra una comparación rápida entre científico de datos y analista de datos

Aspecto	Científico de datos	Analista de datos
Alcance	Enfoque más amplio: aprendizaje automático, modelado predictivo.	Enfoque: analizar datos y brindar información.
Enfoque	Descubrir patrones y predecir tendencias.	Resumir datos históricos y brindar información.

Aspecto	Científico de datos	Analista de datos
Responsabilidades	Procesos integrales, modelos complejos.	Competente en herramientas, métodos estadísticos e informes.
Herramientas	Avanzado: aprendizaje automático, Python/R.	Herramientas: Excel, Tableau, Power BI.
Tipos de datos	Conjuntos de datos grandes, estructurados y no estructurados.	Principalmente datos estructurados, conjuntos más pequeños ocasionales.
Resultado	Extraer información procesable y resolver problemas complejos.	Resumir datos y brindar información para la toma de decisiones.
Superposición	Algunas áreas se superponen y los analistas contribuyen en las primeras etapas.	Roles diferenciados, potencial de colaboración.

Average Salary of a Data Scientist

The average salary of a data scientist varies depending on several factors, including **experience**, **location**, and **skillset**. However, it's generally a high-paying profession with strong growth prospects. Here's a breakdown:

Global Average

- The worldwide average annual salary for a data scientist is around **\$105,000**. (Source: **Glassdoor**)

United States

- In the US, the average annual salary for a data scientist is **\$124,678**. (Source: **Indeed**)
- The median salary is **\$103,500**, according to the Bureau of Labor Statistics. (Source: **BLS**)
- Entry-level data scientists can expect to earn around **\$86,000**, while experienced data scientists with specialized skills can make upwards of **\$156,000**. (Source: **Glassdoor**)

India

- In India, the average annual salary for a data scientist is **₹7,08,012**. (Source: **PayScale**)
- Freshers can expect to start at around **₹5,77,893**, while experienced professionals can earn as much as **₹19,44,566**. (Source: **KnowledgeHut**)



Salary Graph for a Data Scientist

Factors Affecting Salary

Multiple factors might affect your salary as a data scientist:

- **Experience:** As with most professions, experience plays a significant role in determining a data scientist's salary. The more experience you have, the higher your earning potential.
- **Location:** Salaries for data scientists tend to be higher in major tech hubs like San Francisco, New York, and Bangalore compared to smaller cities or rural areas.
- **Skills and Expertise:** Data scientists with specialized skills in areas like machine learning, natural language processing, or specific programming languages can command higher salaries.
- **Company Size and Type:** Large tech companies and startups may offer different salary structures and benefits packages.

Salario promedio de un científico de datos

El salario promedio de un científico de datos varía según varios factores, incluidos la **experiencia**, la **ubicación** y el **conjunto de habilidades**. Sin embargo, generalmente es una profesión bien remunerada con fuertes perspectivas de crecimiento. A continuación, se muestra un desglose:

Promedio global

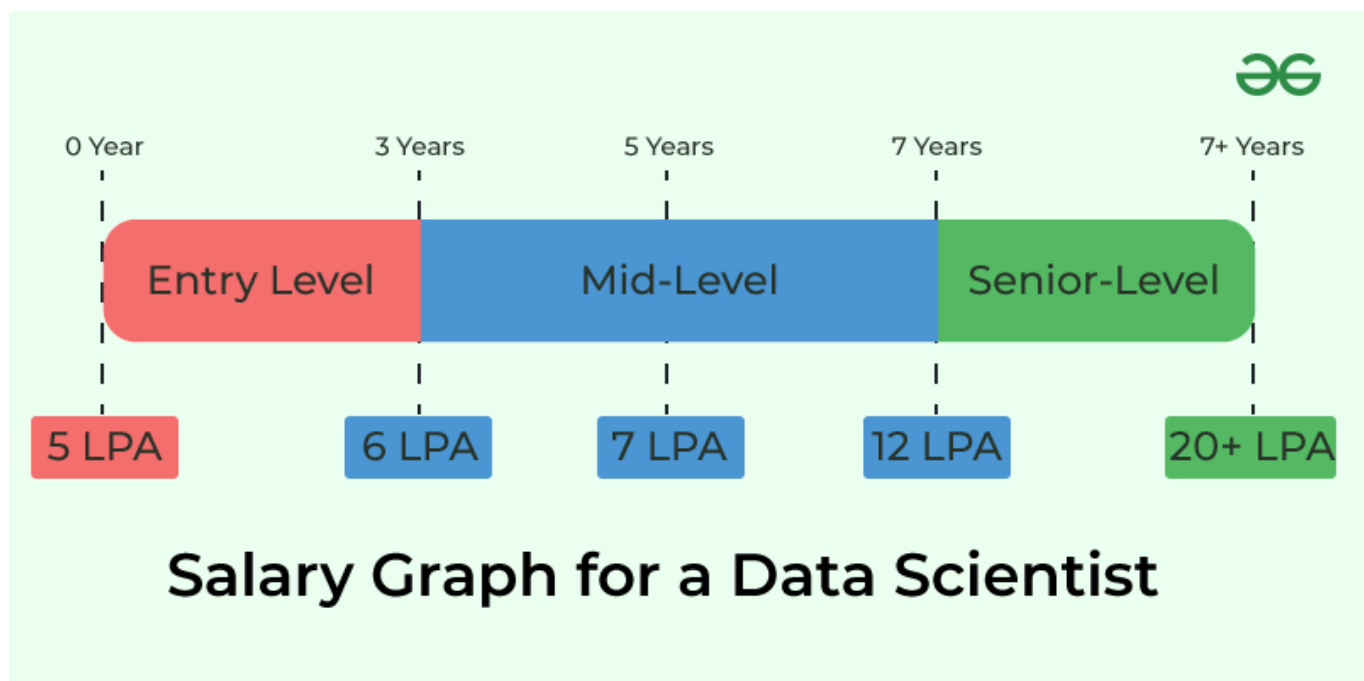
- El salario anual promedio mundial para un científico de datos es de aproximadamente **\$105,000**. (Fuente: **Glassdoor**)

Estados Unidos

- En los EE. UU., el salario anual promedio para un científico de datos es de **\$124,678**. (Fuente: **Indeed**)
- El salario medio es de **\$103,500**, según la Oficina de Estadísticas Laborales. (Fuente: **BLS**)
- Los científicos de datos principiantes pueden esperar ganar alrededor de **\$86,000**, mientras que los científicos de datos experimentados con habilidades especializadas pueden ganar más de **\$156,000**. (Fuente: **Glassdoor**)

India

- En India, el salario anual promedio de un científico de datos es de **₹7,08,012**. (Fuente: **PayScale**)
- Los principiantes pueden esperar comenzar con alrededor de **₹5,77,893**, mientras que los profesionales experimentados pueden ganar hasta **₹19,44,566**. (Fuente: **KnowledgeHut**)



Factores que afectan el salario

Hay varios factores que pueden afectar tu salario como científico de datos:

- **Experiencia:** Como ocurre con la mayoría de las profesiones, la experiencia desempeña un papel importante a la hora de determinar el salario de un científico de datos. Cuanta más experiencia tengas, mayor será tu potencial de ingresos.
- **Ubicación:** Los salarios de los científicos de datos tienden a ser más altos en los principales centros tecnológicos como San Francisco, Nueva York y Bangalore, en

comparación con ciudades más pequeñas o áreas rurales.

- **Habilidades y experiencia:** Los científicos de datos con habilidades especializadas en áreas como el aprendizaje automático, el procesamiento del lenguaje natural o lenguajes de programación específicos pueden exigir salarios más altos.
- **Tamaño y tipo de empresa:** Las grandes empresas tecnológicas y las nuevas empresas pueden ofrecer diferentes estructuras salariales y paquetes de beneficios.

Data Scientist Roadmap 2024

This **data science career roadmap** provides a structured path to master the critical **concepts and skills** needed for success. Remember, data science is **dynamic**, so staying **current** with trends and technologies is **key**. Gaining **real-world experience** through projects and internships can **boost** your skills and **credibility** as a data scientist. Follow this roadmap, **continuously learn**, and **adapt to advancements** for a **rewarding data science journey**

Hoja de ruta para científicos de datos 2024

Esta hoja de ruta profesional en ciencias de datos proporciona una ruta estructurada para dominar los conceptos y habilidades fundamentales necesarios para el éxito. Recuerde que la ciencia de datos es dinámica, por lo que mantenerse al día con las tendencias y tecnologías es clave. Adquirir experiencia en el mundo real a través de proyectos y pasantías puede mejorar sus habilidades y credibilidad como científico de datos. Siga esta hoja de ruta, aprenda continuamente y adaptese a los avances para tener un viaje gratificante en la ciencia de datos.

1) Mathematics

Math skills are very important as they help us understand various machine-learning algorithms that play an important role in Data Science.

- **Part 1:**
 - [Linear Algebra](#)
 - Analytic Geometry
 - Matrix
 - Vector Calculus
 - [Optimization](#)
- **Part 2:**
 - [Regression](#)
 - [Dimensionality Reduction](#)
 - [Density Estimation](#)
 - [Classification](#)

1) Matemáticas

Las habilidades matemáticas son muy importantes porque nos ayudan a comprender varios algoritmos de aprendizaje automático que desempeñan un papel importante en la ciencia de datos.

- **Parte 1:**
 - [Álgebra lineal](#)
 - Geometría analítica
 - Matriz
 - Cálculo vectorial
 - [Optimización](#)
- **Parte 2:**
 - [Regresión](#)
 - [Reducción de dimensionalidad](#)
 - [Estimación de densidad](#)
 - [Clasificación](#)

2) Probability

Probability is also significant to statistics, and it is considered a prerequisite for mastering machine learning.

- [Introduction to Probability](#)
- 1D Random Variable
- The function of One Random Variable
- [Joint Probability Distribution](#)
- [Discrete Distribution](#)
 - [Binomial](#) ([Python](#) | [R](#))
 - [Bernoulli](#)
 - [Geometric](#) etc
- [Continuous Distribution](#)
 - [Uniform](#)
 - [Exponential](#)
 - Gamma
- [Normal Distribution](#) ([Python](#) | [R](#))

2) Probabilidad

La **probabilidad** también es importante para la estadística y se considera un requisito previo para dominar el aprendizaje automático.

- [Introducción a la probabilidad](#)
- Variable aleatoria unidimensional
- La función de una variable aleatoria
- [Distribución de probabilidad conjunta](#)
- [Distribución discreta](#)
 - [Binomial](#) ([Python](#) | [R](#))
 - [Bernoulli](#)
 - [Geométrica](#) etc.
- [Distribución continua](#)
 - [Uniforme](#)
 - [Exponencial](#)
 - Gamma
- [Distribución normal](#) ([Python](#) | [R](#))

3) Statistics

Understanding **Statistics** is very significant as this is a part of Data analysis.

- [Introduction to Statistics](#)
- Data Description
- Random Samples
- Sampling Distribution
- Parameter Estimation
- [Hypotheses Testing](#) ([Python](#) | [R](#))
- ANOVA ([Python](#) | [R](#))
- [Reliability Engineering](#)
- [Stochastic Process](#)
- Computer Simulation
- Design of Experiments
- [Simple Linear Regression](#)
- [Correlation](#)
- Multiple Regression ([Python](#) | [R](#))
- Nonparametric Statistics
 - Sign Test
 - [The Wilcoxon Signed-Rank Test](#) ([R](#))

- The Wilcoxon Rank Sum Test
- [The Kruskal-Wallis Test \(R\)](#)
- Statistical Quality Control
- [Basics of Graphs](#)

3) Estadísticas

Entender las **Estadísticas** es muy importante ya que son parte del análisis de datos.

- [Introducción a la estadística](#)
- Descripción de los datos
- Muestras aleatorias
- Distribución de muestreo
- Estimación de parámetros
- [Prueba de hipótesis](#) (Python | [R](#))
- ANOVA ([Python](#) | [R](#))
- [Confiabilidad Ingeniería](#)
- [Proceso estocástico](#)
- Simulación por computadora
- Diseño de experimentos
- [Regresión lineal simple](#)
- [Correlación](#)
- Regresión múltiple ([Python](#) | [R](#))
- Estadísticas no paramétricas
- Prueba de signos
- [La prueba de rangos con signo de Wilcoxon \(R\)](#)
- La prueba de suma de rangos de Wilcoxon
- [La prueba de Kruskal-Wallis \(R\)](#)
- Control de calidad estadístico
- [Conceptos básicos de Gráficos](#)

4) Programming

One needs to have a good grasp of programming concepts such as **Data structures and Algorithms**. The programming languages used are **Python**, **R**, [Java](#), [Scala](#). [C++](#) is also useful in some places where performance is very important.

- **Python:**
 - Python Basics

- [List](#)
 - [Set](#)
 - [Tuples](#)
 - [Dictionary](#)
 - [Function](#), etc.
- [NumPy](#)
- [Pandas](#)
- [Matplotlib/Seaborn](#), etc.
- **R:**
 - [R Basics](#)
 - [Vector](#)
 - [List](#)
 - [Data Frame](#)
 - [Matrix](#)
 - [Array](#)
 - [Function](#), etc.
 - [dplyr](#)
 - [ggplot2](#)
 - [Tidyr](#)
 - [Shiny](#), etc.
- **DataBase:**
 - SQL
 - [MongoDB](#)
- **Other:**
 - Data Structure
 - [Time Complexity](#)
 - Web Scraping ([Python](#) | [R](#))
 - [Linux](#)
 - [Git](#)

4) Programación

Es necesario tener un buen conocimiento de conceptos de programación como **Estructuras de datos y algoritmos**. Los lenguajes de programación utilizados son **Python**, **R**, [Java](#), [Scala](#). [C++](#) también es útil en algunos lugares donde el rendimiento es muy importante.

- **Python:**
 - Conceptos básicos de Python

- [Lista](#)
- [Conjunto](#)
- [Tuplas](#)
- [Diccionario](#)
- [Función](#), etc.
- [NumPy](#)
- [Pandas](#)
- [Matplotlib/Seaborn](#), etc.
- **R:**
 - [Conceptos básicos de R](#)
 - [Vector](#)
 - [Lista](#)
 - [Marco de datos](#)
 - [Matriz](#)
 - [Array](#)
 - [Función](#), etc.
 - [dplyr](#)
 - [ggplot2](#)
 - [Tidyr](#)
 - [Shiny](#), etc.
- **Base de datos:**
 - SQL
 - [MongoDB](#)
- **Otros:**
 - Estructura de datos
 - [Complejidad temporal](#)
 - Web Scraping ([Python](#) | [R](#))
 - [Linux](#)
 - [Git](#)

5) Machine Learning

ML is one of the most vital parts of data science and the hottest subject of research among researchers so each year new advancements are made in this. One at least needs to understand the basic algorithms of **Supervised and Unsupervised Learning**. There are multiple libraries available in Python and R for implementing these algorithms.

- **Introduction:**

- How Model Works
- Basic Data Exploration
- First ML Model
- Model Validation
- [Underfitting & Overfitting](#)
- Random Forests ([Python](#) | [R](#))
- [scikit-learn](#)
- **Intermediate:**
 - [Handling Missing Values](#)
 - [Handling Categorical Variables](#)
 - [Pipelines](#)
 - [Cross-Validation](#) ([R](#))
 - [XGBoost](#) ([Python](#) | [R](#))
 - Data Leakage

5) Aprendizaje automático

El aprendizaje automático es una de las partes más importantes de la ciencia de datos y el tema de investigación más candente entre los investigadores, por lo que cada año se realizan nuevos avances en este campo. Al menos es necesario comprender los algoritmos básicos del aprendizaje supervisado y no supervisado. Hay varias bibliotecas disponibles en Python y R para implementar estos algoritmos.

- **Introducción:**
 - Cómo funciona el modelo
 - Exploración básica de datos
 - Primer modelo de ML
 - Validación del modelo
 - [Subajuste y sobreajuste](#)
 - Bosques aleatorios ([Python](#) | [R](#))
 - [scikit-learn](#)
- **Intermedio:**
 - [Manejo de datos faltantes ... Valores](#)
 - [Manejo de variables categóricas](#)
 - [Pipelines](#)
 - [Validación cruzada](#) ([R](#))
 - [XGBoost](#) ([Python](#) | [R](#))
 - Fuga de datos

6) Deep Learning

Deep Learning uses [TensorFlow](#) and [Keras](#) to build and train neural networks for structured data.

- [Artificial Neural Network](#)
- [Convolutional Neural Network](#)
- [Recurrent Neural Network](#)
- TensorFlow
- Keras
- [PyTorch](#)
- A Single Neuron
- [Deep Neural Network](#)
- [Stochastic Gradient Descent](#)
- Overfitting and Underfitting
- Dropout Batch Normalization
- Binary Classification

6) Aprendizaje profundo

El aprendizaje profundo utiliza [TensorFlow](#) y [Keras](#) para crear y entrenar redes neuronales para datos estructurados.

- [Red neuronal artificial](#)
- [Red neuronal convolucional](#)
- [Red neuronal recurrente](#)
- TensorFlow
- Keras
- [PyTorch](#)
- Una sola neurona
- [Red neuronal profunda](#)
- [Gradiente estocástico Descenso](#)
- Sobreajuste y subajuste
- Normalización de lotes con abandono
- Clasificación binaria

7) Feature Engineering

In Feature Engineering discover the most effective way to improve your models.

- Baseline Model
- Categorical Encodings
- Feature Generation
- Feature Selection

7) Ingeniería de características

En Ingeniería de características, descubra la forma más eficaz de mejorar sus modelos.

- Modelo de línea base
- Codificaciones categóricas
- Generación de características
- Selección de características

8) Natural Language Processing

In [NLP](#) distinguish yourself by learning to work with text data.

- Text Classification
- Word Vectors

8) Procesamiento del lenguaje natural

En [PLN](#) destaque aprendiendo a trabajar con datos de texto.

- Clasificación de texto
- Vectores de palabras

9) Data Visualization Tools

Make great data visualizations. A great way to see the power of coding!

- Excel VBA
- **BI (Business Intelligence):**
 - [Tableau](#)
 - [Power BI](#)
 - Qlik View
 - Qlik Sense

9) Herramientas de visualización de datos

Realice excelentes visualizaciones de datos. ¡Una excelente manera de ver el poder de la codificación!

- Excel VBA
- **BI (Business Intelligence):**
 - [Tableau](#)
 - [Power BI](#)
 - Qlik View
 - Qlik Sense

10) Deployment

The last part is doing the deployment. Definitely, whether you are fresher or 5+ years of experience, or 10+ years of experience, deployment is necessary. Because deployment will definitely give you a fact is that you worked a lot.

- [Microsoft Azure](#)
- [Heroku](#)
- [Google Cloud Platform](#)
- [Flask](#)
- [Django](#)

10) Implementación

La última parte es la implementación. Definitivamente, ya sea que seas nuevo o tengas más de 5 años de experiencia, o más de 10 años de experiencia, la implementación es necesaria, ya que la implementación definitivamente te dará un hecho: has trabajado mucho.

- [Microsoft Azure](#)
- [Heroku](#)
- [Google Cloud Platform](#)
- [Flask](#)
- [Django](#)

11) Other Points to Learn

- Domain Knowledge
- Communication Skill
- [Reinforcement Learning](#)
- **Different Case Studies:**
 - Data Science at Netflix

- Data Science at Flipkart
- Project on Credit Card Fraud Detection
- Project on Movie Recommendation, etc.

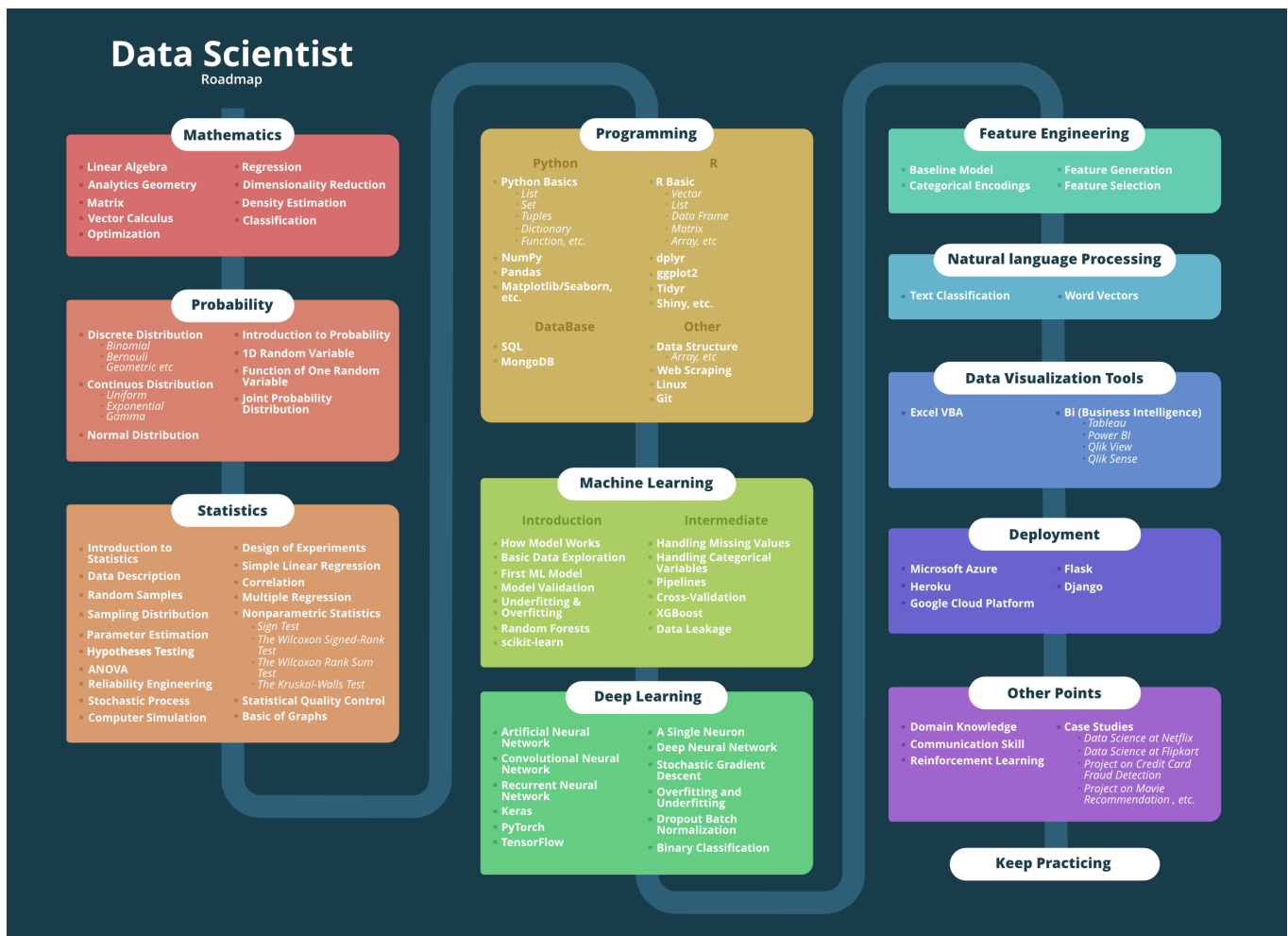
11) Otros puntos a aprender

- Conocimiento del dominio
- Habilidad de comunicación
- [Aprendizaje de refuerzo](#)
- **Diferentes estudios de caso:**
 - Ciencia de datos en Netflix
 - Ciencia de datos en Flipkart
 - Proyecto sobre detección de fraudes con tarjetas de crédito
 - Proyecto sobre recomendación de películas, etc.

12) Keep Practicing

*****“Practice makes a man perfect” which tells the importance of continuous practice in any subject to learn anything.****

So keep practicing and improving your knowledge day by day. Below is a complete diagrammatical representation of the Data Scientist Roadmap.



12) Sigue practicando

****“La práctica hace al maestro”, lo que demuestra la importancia de la práctica continua en cualquier materia para aprender cualquier cosa.****

Así que sigue practicando y mejorando tus conocimientos día a día. A continuación, se muestra una representación esquemática completa de la hoja de ruta del científico de datos.

Data Scientist Roadmap: Education Routes

Regardless of your academic path, unlock success through **lifelong learning and skill mastery**. Dive into coding languages like Python and R, conquer statistics and machine learning fundamentals, whether your background is in **computer science, math, or beyond**. Gain **hands-on experience** through data science projects, internships, and **powerful networking**. Build a **robust skillset** and stay ahead of the curve with the **latest data science trends**.

1. Educational Background:

- **Bachelor's Degree:**

- Most data scientists have at least a bachelor's degree in fields like **computer science, statistics, mathematics, or engineering.**
- Non-traditional backgrounds are okay, but having a solid foundation in quantitative subjects is beneficial.
- **Advanced Degrees:**
 - Many data scientists pursue master's or **Ph.D. degrees**, especially for specialization or research.
 - Degrees in **data science, machine learning, artificial intelligence**, or related fields are increasingly available.

2. Core Skills:

- **Programming Languages:**
 - Learn languages commonly used in data science, like **Python** or **R**.
 - Use libraries and frameworks such as **NumPy, Pandas, scikit-learn, TensorFlow**, and **PyTorch**.
- **Statistics and Mathematics:**
 - Understand statistical concepts and mathematical foundations, including linear algebra and calculus.
- **Data Manipulation and Analysis:**
 - Master data manipulation and analysis with tools like **SQL** and **Pandas**.
- **Machine Learning:**
 - Gain expertise in **machine learning algorithms, covering supervised and unsupervised learning, regression, classification, clustering**, etc.
- **Data Visualization:**
 - Communicate insights through visualization tools like **Matplotlib, Seaborn, or Tableau**.
- **Big Data Technologies:**
 - Familiarize yourself with big data technologies like **Hadoop** and **Spark**.

3. Projects and Practical Experience:

- Work on real-world projects to apply knowledge and build a portfolio.
- Participate in Kaggle competitions or similar challenges.
- Contribute to open-source projects or collaborate on data-related projects.

4. Networking:

- Attend data science meetups, conferences, and networking events.
- Join **online communities, forums**, and **social media groups** related to data science.

5. Continuous Learning:

- Stay updated with the latest trends and technologies in data science.

- Take **online courses**, **attend workshops**, and **pursue certifications** for skill enhancement.

6. Internships and Work Experience:

- **Seek internships** or **entry-level positions** for practical experience.
- Get exposure to real-world data science problems.

7. Soft Skills:

- Develop communication skills to convey findings effectively to non-technical stakeholders.
- Cultivate problem-solving, critical thinking, and attention to detail.

Hoja de ruta del científico de datos: rutas educativas

Independientemente de tu trayectoria académica, alcanza el éxito mediante el **aprendizaje permanente y el dominio de las habilidades**. Sumérgete en lenguajes de codificación como Python y R, domina las estadísticas y los fundamentos del aprendizaje automático, ya sea que tengas experiencia en **informática, matemáticas o más**. Obtén **experiencia práctica** a través de proyectos de ciencia de datos, pasantías y **redes potentes**. Desarrolla un **conjunto de habilidades sólido** y mantente a la vanguardia con las **últimas tendencias en ciencia de datos**.

1. Antecedentes educativos:

- **Licenciatura:**
 - La mayoría de los científicos de datos tienen al menos una licenciatura en campos como **informática, estadística, matemáticas o ingeniería**.^{**}
 - Los antecedentes no tradicionales están bien, pero tener una base sólida en materias cuantitativas es beneficioso.
- **Títulos avanzados:**
 - Muchos científicos de datos cursan estudios de máster o doctorado, especialmente para especializarse o realizar investigaciones.
 - Cada vez hay más títulos disponibles en **ciencia de datos, aprendizaje automático, inteligencia artificial** o campos relacionados.

2. Habilidades básicas:

- **Lenguajes de programación:**
 - Aprenda lenguajes que se usan comúnmente en la ciencia de datos, como **Python o R**.
 - Use bibliotecas y marcos como **NumPy, Pandas, scikit-learn, TensorFlow y PyTorch**.
- **Estadísticas y matemáticas:**

- Comprenda los conceptos estadísticos y los fundamentos matemáticos, incluidos el álgebra lineal y el cálculo.
 - **Manipulación y análisis de datos:**
 - Domina la manipulación y el análisis de datos con herramientas como **SQL y Pandas**.
 - **Aprendizaje automático:**
 - Adquiere experiencia en algoritmos de aprendizaje automático, que abarcan aprendizaje supervisado y no supervisado, regresión, clasificación, agrupamiento, etc.
 - **Visualización de datos:**
 - Comunica información a través de herramientas de visualización como **Matplotlib, Seaborn o Tableau**.
 - **Tecnologías de big data:**
 - Familiarízate con tecnologías de big data como Hadoop y Spark.**
3. **Proyectos y experiencia práctica:**
- Trabaja en proyectos del mundo real para aplicar conocimientos y crear un portafolio.
 - Participa en competencias de Kaggle o desafíos similares.
 - Contribuir a proyectos de código abierto o colaborar en proyectos relacionados con datos.
4. **Networking:**
- Asistir a reuniones, conferencias y eventos de networking sobre ciencia de datos.
 - Unirse a **comunidades en línea, foros y grupos de redes sociales** relacionados con la ciencia de datos.
5. **Aprendizaje continuo:**
- Mantenerse actualizado con las últimas tendencias y tecnologías en ciencia de datos.
 - Tomar **cursos en línea, asistir a talleres y obtener certificaciones** para mejorar las habilidades.
6. **Pasantías y experiencia laboral:**
- **Buscar pasantías o puestos de nivel inicial** para obtener experiencia práctica.
 - Obtener exposición a problemas de ciencia de datos del mundo real.
7. **Habilidades blandas:**
- Desarrollar habilidades de comunicación para transmitir los hallazgos de manera eficaz a las partes interesadas no técnicas.
 - Cultivar la resolución de problemas, el pensamiento crítico y la atención al detalle.

Conclusion

In the 21st century, data science has emerged as a crucial profession, often dubbed *****“The Sexiest Job”***** by Harvard Business Review. With the rise of **Big Data** and frameworks like **Hadoop**, data science focuses on processing vast amounts of data. This field’s significant growth underscores its importance for future readiness.

The comparison between data science and data analyst roles highlights data scientists’ broader scope and responsibilities in predicting trends and solving complex problems. To become a data scientist, a strong educational background, core skills in programming and statistics, practical experience through projects, and continuous learning are essential.

The global demand for data scientists is high, offering lucrative salaries and impactful work opportunities. The roadmap for learning data science covers key domains like mathematics, programming, machine learning, deep learning, natural language processing, data visualization, and deployment. Continuous practice, networking, and soft skills development are emphasized for success in this dynamic field.

Conclusión

En el siglo XXI, la ciencia de datos ha surgido como una profesión crucial, a menudo denominada **“El trabajo más atractivo” por Harvard Business Review. Con el auge de Big Data** y marcos como **Hadoop**, la ciencia de datos se centra en procesar grandes cantidades de datos. El crecimiento significativo de este campo subraya su importancia para la preparación futura.

La comparación entre la ciencia de datos y los roles de analista de datos destaca el alcance y las responsabilidades más amplios de los científicos de datos en la predicción de tendencias y la solución de problemas complejos. Para convertirse en un científico de datos, son esenciales una sólida formación académica, habilidades básicas en programación y estadística, experiencia práctica a través de proyectos y aprendizaje continuo.

La demanda global de científicos de datos es alta y ofrece salarios lucrativos y oportunidades laborales impactantes. La hoja de ruta para aprender ciencia de datos cubre dominios clave como matemáticas, programación, aprendizaje automático, aprendizaje profundo, procesamiento del lenguaje natural, visualización de datos e implementación. La práctica continua, la creación de redes y el desarrollo de habilidades blandas se enfatizan para el éxito en este campo dinámico.

FAQs on Data Scientist Roadmap – A Complete Guide

What qualifications do you need to be a data scientist?

Master Python, R, SQL, and Java for data science, blend math foundations with efficient data handling (Pandas, SQL), and hone soft skills. Pursue relevant degrees or alternative

paths, build a standout portfolio, network, and stay updated for success in this dynamic field.

Is data science an IT job?

Data science is more closely tied to statistics, mathematics, and business intelligence than traditional IT. While it leverages technology heavily, its primary focus lies on the analysis and interpretation of data, making it a distinct field with its own set of skills and goals.

Is it hard to become a Data Scientist?

Becoming a data scientist requires a lot of skills and dedication. It involves mastering technical skills like mathematics, programming and various tools. There is a lot of competition and is evolving at high rate. Learning data science depends on your dedication and approach.

Is data science dead in 10 years?

No, Global data explosion requires skilled interpreters—data scientists. Applications expand across sectors, from healthcare to art. Automation aids but can't replace vital data scientist skills. Continuous learning is crucial in the evolving data science landscape.

Can you become a data scientist without a degree?

Master **Python, R, stats, math, Pandas, SQL, ML**, and **data viz**. Build a strong portfolio, contribute to open source, network at meetups. Stay connected online, keep learning, and persistently showcase skills to break into the field.

How long does it take to become a data scientist?

Enter data science in 6 months to a year with a strong background. Traditional degrees take **4 years**, bootcamps **3 months to a year**. Analyst roles may come sooner, specialized positions 2-5 years or more. Consistent practice speeds progress. Estimates: **Entry (6 months – 1 year)**, **Junior (1-2 years)**, **Mid-level (2-5 years)**, **Senior (5+ years)**.

What are the benefits of becoming a data scientist?

Some of the benefits of becoming data scientist include:

- High demand & salary
- Versatile career
- Personal growth
- Global opportunities

Preguntas frecuentes sobre la hoja de ruta del científico de datos: una guía completa

¿Qué cualificaciones se necesitan para ser un científico de datos?

Domina Python, R, SQL y Java para la ciencia de datos, combina los fundamentos matemáticos con el manejo eficiente de datos (Pandas, SQL) y perfecciona las habilidades interpersonales. Obtén títulos relevantes o caminos alternativos, crea una cartera destacada, crea redes y mantente actualizado para tener éxito en este campo dinámico.

¿La ciencia de datos es un trabajo de TI?

La ciencia de datos está más estrechamente vinculada a las estadísticas, las matemáticas y la inteligencia empresarial que la TI tradicional. Si bien aprovecha mucho la tecnología, su enfoque principal radica en el análisis y la interpretación de datos, lo que la convierte en un campo distinto con su propio conjunto de habilidades y objetivos.

¿Es difícil convertirse en un científico de datos?

Convertirse en un científico de datos requiere muchas habilidades y dedicación. Implica dominar habilidades técnicas como las matemáticas, la programación y varias herramientas. Hay mucha competencia y está evolucionando a un ritmo rápido. Aprender ciencia de datos depende de tu dedicación y enfoque.

¿La ciencia de datos está muerta en 10 años?

No, la explosión global de datos requiere intérpretes capacitados: científicos de datos. Las aplicaciones se expanden en todos los sectores, desde la atención médica hasta el arte. La automatización ayuda, pero no puede reemplazar las habilidades vitales de los científicos de datos. El aprendizaje continuo es crucial en el cambiante panorama de la ciencia de datos.

¿Puede convertirse en un científico de datos sin un título?

Domine **Python, R, estadísticas, matemáticas, Pandas, SQL, ML y visualización de datos**. Cree una cartera sólida, contribuya al código abierto, establezca contactos en reuniones. Manténgase conectado en línea, siga aprendiendo y demuestre constantemente sus habilidades para ingresar al campo.

¿Cuánto tiempo lleva convertirse en un científico de datos?

Ingresa a la ciencia de datos en 6 meses a un año con una sólida formación. Los títulos tradicionales demoran **4 años**, los campamentos de entrenamiento **3 meses a un año**. Los roles de analista pueden llegar antes, los puestos especializados de 2 a 5 años o más. La práctica constante acelera el progreso. Estimaciones: **Principiante (6 meses – 1 año)**, **Junior (1-2 años)**, **Nivel medio (2-5 años)**, **Senior (5+ años)**.

¿Cuáles son los beneficios de convertirse en un científico de datos?

Algunos de los beneficios de convertirse en un científico de datos incluyen:

- Alta demanda y salario
- Carrera versátil
- Crecimiento personal
- Oportunidades globales