

Científico de Datos

Nivel Básico

Aliados:



Microsoft

Vigilada Mineducación



Advanced analytics for business

Generalidades

✓ Total clases: 48 horas

✓ 12 semanas

✓ 1 temática semanal (4 horas)

❖ Profe: Juan Esteban Serna Vega

❖ Email: jesernav@eafit.edu.co

❖ Monitor: Stiven Restrepo

❖ Email: yerestrepr@eafit.edu.co

Aliados:



Vigilada Mineducación



Tema: Intro al curso

Programa (resumen)

1	Introducción a Ciencia de datos
2	Python como lenguaje de Ciencia de datos
3	Estadística Básica (Nivelación)
4	Distribuciones de Probabilidad
5	Evolución del análisis predictivo y el machine learning
6	Pre-procesamiento de datos
7	Operaciones básicas con datasets
8	Modelos de Regresión*
9	Modelos de Clasificación*
10	Métodos de segmentación natural y estratégica*
11	Bigdata y Databricks en Ciencia de Datos*
12	Automatización de modelos

**Introducción*

Aliados:



Vigilada Mineducación

Tema: Intro al curso

Sprint 1

1	Introducción a Ciencia de datos
2	Python como lenguaje de Ciencia de datos
3	Estadística Básica (Nivelación)

Sprint 2

4	Distribuciones de Probabilidad
5	Evolución del análisis predictivo y el machine learning
6	Pre-procesamiento de datos

Sprint 3

7	Operaciones básicas con datasets
8	Modelos de Regresión*
9	Modelos de Clasificación*

Sprint 4

10	Métodos de segmentación natural y estratégica*
11	Bigdata y Databricks en Ciencia de Datos*
12	Automatización de modelos

**Introducción*

Aliados:



Vigilada Mineducación



Evaluación

Análisis de caso: Incorporación de analítica avanzada a la realidad del negocio (pdf + datasets)

Empresa: CottonPlus Inc.

Industria: Retail de moda

Entregable

Notebook (ver objetivos según capítulo del caso)

Fecha de 1ª entrega: a partir de Julio 16

Recursos

- [Material de clase \(slides + notebook\)](#)
- Material complementario (DS)
- Contenido asincrónico

Otros Recursos EAFIT

- Office 365
- [Plataforma Interactiva](#)
- Biblioteca

Aliados:



Vigilada Mineducación



Para las clases...



Participar es importante.



Empezamos y terminamos a tiempo (hay break)



Silenciar el micrófono cuando no estemos hablando



Cámara es opcional, pero estaría chévere prenderla.

Tip! Armen grupo en **Wapp, Telegram, Slack**, o donde quieran...

Aliados:



Microsoft

Vigilada Mineducación



Advanced analytics for business

Intro a DS

1. Data Science
2. Conceptos
3. Herramientas

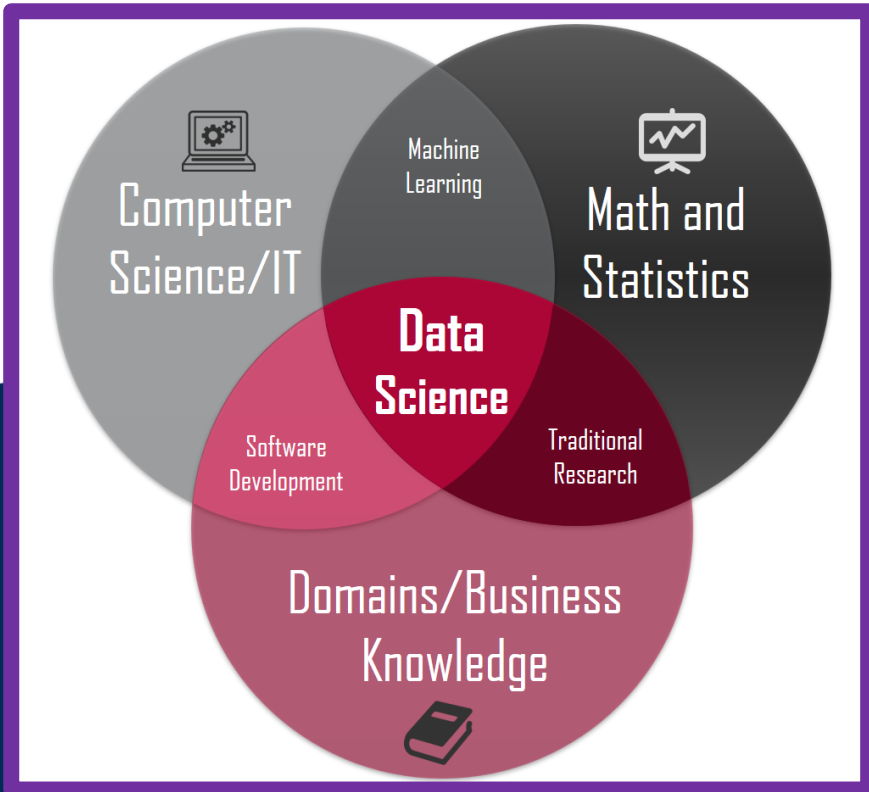
Aliados:



Vigilada Mineducación



¿Qué es DS?



¿Qué sabe un Data Scientist?

Ojo acá!

- 1 Conocimientos de **matemática y estadística**
- 2 Conocimientos de **programación**
- 3 Algunos conocimientos del **tema específico**
- 4 Habilidad de **comunicar** resultados

Aliados:



Vigilada Mineducación



¿Qué es DS?



Aliados:



Vigilada Mineducación



Conceptos

- Probabilidad y Estadística
- Bases de Datos
- Data Mining, Big Data
- Machine Learning - Aprendizaje Automático
- Deep Learning - Redes Neuronales
- Inteligencia Artificial

➤ Metodologías Ágiles

Glosario

Python: lenguaje de programación que utilizaremos durante la carrera. Es un lenguaje de propósito general, pero tiene una gran comunidad en Ciencia de Datos, Machine Learning y Deep Learning

Miniconda: una distribución de Python, hermana menor - y minimalista - de Anaconda.

Conda: un manejador de paquetes. Es la herramienta que utilizaremos por defecto para instalar librerías en python y crear ambientes. Cuando falle, recurriremos pip.

Jupyter: entorno de desarrollo. Viene en dos sabores, Jupyter Notebook y, más recientemente, Jupyter Lab. Ambos trabajan sobre los mismos notebook y son muy parecidos. La principal diferencia es que Lab tiene una interfaz más amigable y más parecida a un navegador.

Notebook: archivo en el cual escribimos código, visualizamos resultados e imágenes, escribimos texto con preguntas, descripciones, conclusiones, etc. Su extensión es .ipynb.

Google Colab: entorno de desarrollo online, orientado a Ciencia de Datos, basado en Jupyter, en el cual podemos desarrollar o correr notebooks en la nube. Gran recurso si no queremos instalar

Aliados:



Vigilada Mineducación



Herramientas

1. Computer Science / IT

- Lenguajes de programación (Python, R)
- Desarrollo Web (HTML)
- Bases de datos (SQL)

2. Math and Statistics

- Matrices
- Probabilidades

3. Business knowledge

- Qué industria te gusta?
- Cómo se aplica DS en esa industria?



kaggle



Google
colab



MOOC Platforms

Aliados:



Vigilada Mineducación



Data Scientist: The Sexiest Job of the 21st Century

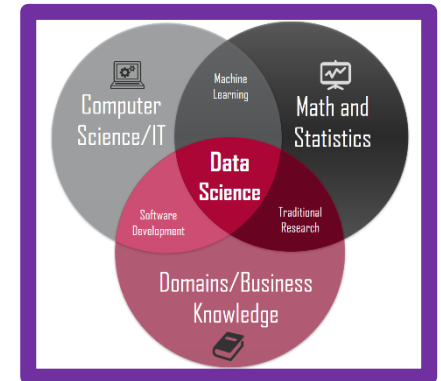
[Ver online](#)

BIG DATA



¿Qué hace un Data Scientist?

1. Estructura a los datos
2. Encuentra patrones en ellos
3. Opina y recomienda a los directivos en temas de productos, procesos y decisiones.



Encuentra la historia enterrada en los datos y la comunica!

Aliados:



Vigilada Mineducación



¿Qué problemas resuelve un Data Scientist?

1. Sistemas de recomendación
2. Predicciones (análisis de comportamiento, compra, precios, tráfico...)
3. Marketing predictivo (Market Basket Analysis, Leads)
4. Detección de fraude
5. Web Scrapping
6. Procesamiento de lenguaje natural (NLP)
7. Industria musical
8. Visión por computador (detección de enfermedades, agro)
9. Genética
10. Robótica
11. Optimización de recursos (pipelines)
12. Smart cities
13. Política
14. IoT

Hay muchos más...

Aliados:



Microsoft

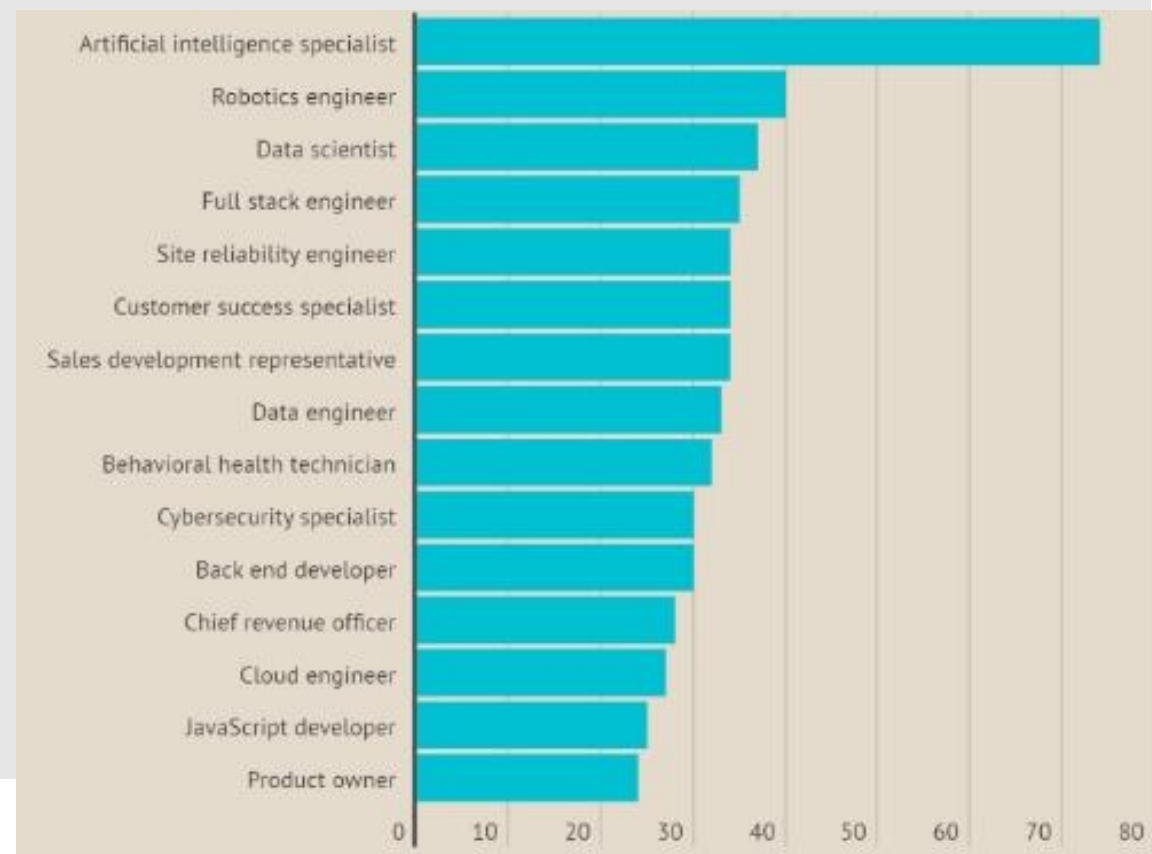
Vigilada Mineducación



Advanced analytics for business

Campos de acción

- --Transformación digital--
- Educación
- Cambio climático
- Ecommerce / Retail
- FinTech / Finance
- Marketing
- Operations
- Salud
- Privacidad y calidad de la info
- Procesos (...)



Aliados:



Vigilada Mineducación



Proyección laboral

[*Ver online*](#)

These are the 10 most in-demand A.I. jobs according to Indeed
—and they all pay at least \$95,000

1. Data scientist

Median yearly salary in the U.S.: \$110,000

2. Senior software engineer

Median yearly salary in the U.S.: \$120,000

3. Machine learning engineer

Median yearly salary in the U.S.: \$125,000

4. Data engineer

Median yearly salary in the U.S.: \$122,060

5. Software engineer

Median yearly salary in the U.S.: \$100,000

6. Software developer

Median yearly salary in the U.S.: \$95,000

7. Software architect

Median yearly salary in the U.S.: \$135,107

8. Senior data scientist

Median yearly salary in the U.S.: \$127,500

9. Full stack developer

Median yearly salary in the U.S.: \$108,730

10. Principal software engineer

Median yearly salary in the U.S.: \$155,000

Buscar Python en:

- G Trends

Aliados:



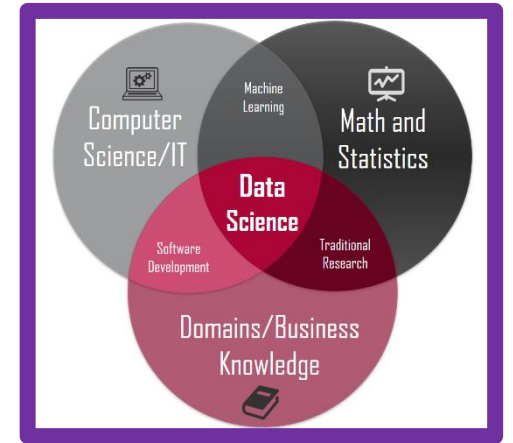
Vigilada Mineducación



Advanced analytics for business

Roadmap de formación

- Programación Python y R
- Machine Learning
- Bases de datos: SQL y NoSQL (MongoDB)
- Despliegue Web: Django – Flask
- Big Data: Hive – MapReduce – Pig
- Otras herramientas (Máquinas virtuales)



Aliados:



Vigilada Mineducación



DS dentro de Artificial Intelligence (AI)

- ML es una aplicación de AI
- ML es una herramienta para DS

Aliados:



Vigilada Mineducación



Material complementario

- <https://learnxinyminutes.com/docs/python3/>
- <https://www.tutorialsteacher.com/python>
- [Todo el mundo debería aprender a programar \(video\)](#)
- Libro: Big Data. Walter Sosa Escudero.



Aliados:



Vigilada Mineducación



Para la próxima clase

1. Tener instalado Python

Ayuda: [Instalación Anaconda + Iniciando con Python](#)

2. Se sugiere abrir cuenta en GitHub

3. Se sugiere explorar herramientas Google Trends, Google Books Ngram Viewer.

Aliados:



Microsoft

Vigilada Mineducación



Advanced analytics for business

Contenido asincrónico

- [Azure Cloud Concepts](#)
- [Azure Cloud Services](#)
- Resto de actividades en plataforma Interactiva Virtual

Aliados:



Microsoft

Vigilada Mineducación



¡Gracias!

Aliados:



Microsoft

Vigilada Mineducación



Advanced analytics for business