Introducción a Big Data

M2 – Data Science.

Fundación Telefónica Movistar
CURSO DE INTRODUCCIÓN A BIG DATA.

M2. Data Science.

Seguramente cuando buscamos la definición o las competencias del científico de datos o de la ciencia de datos, nos encontramos que cada autor y cada experto tiene sus propias definiciones.

Trataremos de explorar las principales tareas, definirlas y catalogarlas para tener, nosotros también nuestra propia idea acerca de que se trata todo esto.

Data Science es un campo interdisciplinario que involucra métodos científicos, procesos y sistemas para extraer conocimiento o un mejor entendimiento de datos en sus diferentes formas, ya sea estructurados o no estructurados.

Surgimiento del término.

2001

William Cleveland escribe el paper "Data Science: An action plan to expand the field of statistics"

2008

DJ Patil y a Jeff Hammerbacher acuñan el término *Data Scientist* para describir sus propios puestos de trabajo

La práctica de la ciencia de datos requiere diferentes habilidades. Cada científico de datos tiene un perfil diferente acorde a sus habilidades y orientación educativa, por este motivo, se hace imprescindible el formar equipos de trabajo acorde al objetivo del proyecto.

Perfiles.

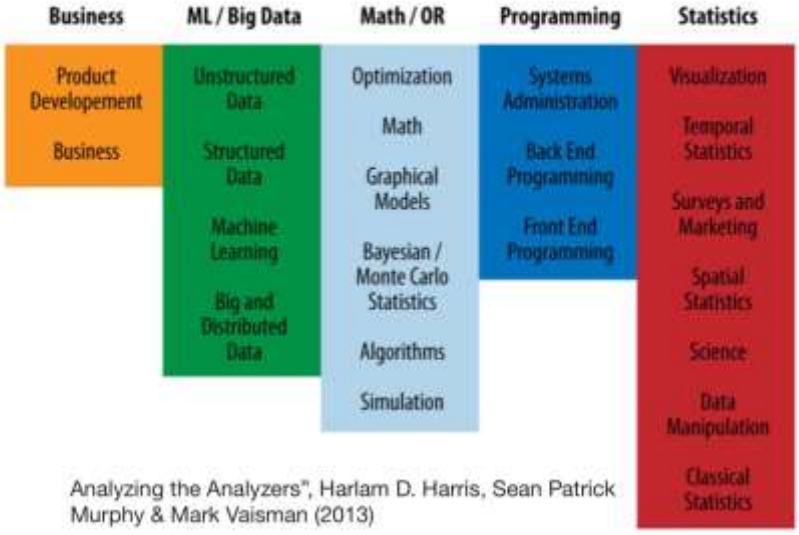
En el estudio publicado bajo el nombre "Analyzing the Analyzers", Harlam D. Harris, Sean Patrick Murphy y Mark Vaisman realizaron una encuesta a *data scientists* para analizar los diferentes perfiles.

En la pregunta de **autoidentificación**, los encuestados tenían que responder cuánto se identificaban con cada una de las 11 categorías, que después agruparon en 4 Grupos:

Data Developer	Developer	Engineer	
Data Researcher	Researcher	Scientist	Statistician
Data Creative	Jack of All Trades	Artist	Hacker
Data Businessperson	Leader	Businessperson	Entrepeneur

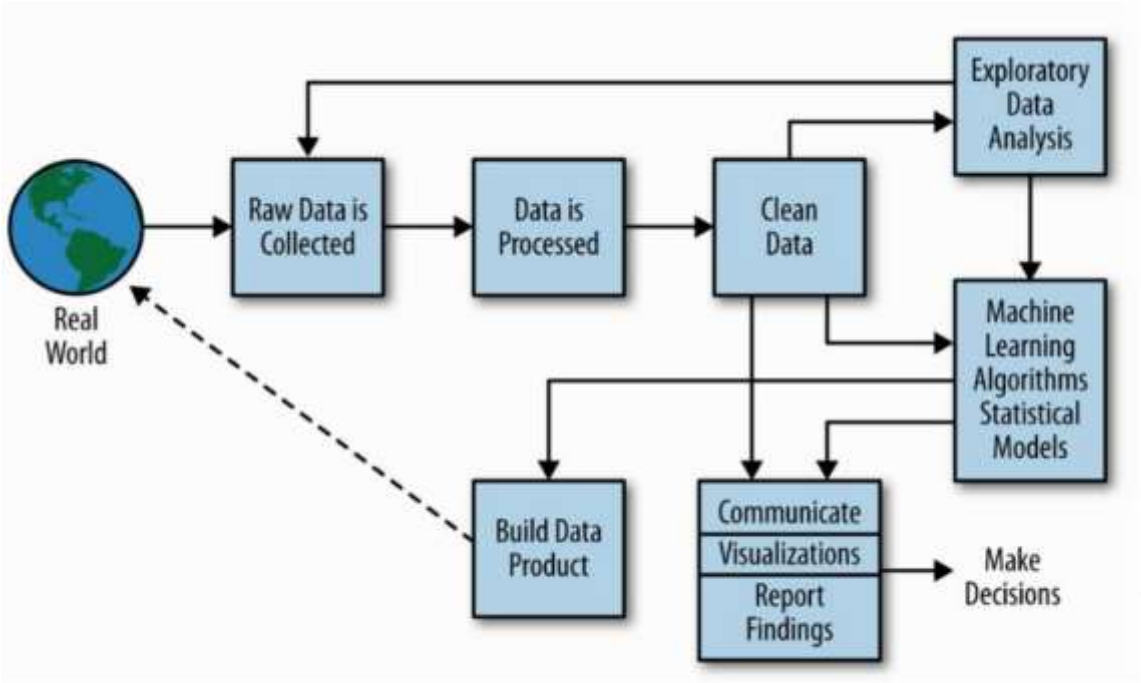
Analyzing the Analyzers", Harlam D. Harris, Sean Patrick Murphy & Mark Vaisman (2013)

Adicionalmente, le pidieron a los encuestados que ordenen sus **habilidades** de acuerdo a cuáles dominan más.



Analyzing the Analyzers", Harlam D. Harris, Sean Patrick Murphy & Mark Vaisman (2013)

El proceso de un proyecto de ciencia de datos.



Rachel Schutt and Cathy O'Neil. "Doing Data Science."

** Cada punto del proceso se explica y debate en clase **