Data Science

Emiliano Actis Dato
Analytics Client Architect
actisdato@ar.ibm.com
@emiactisdato

Quién soy?

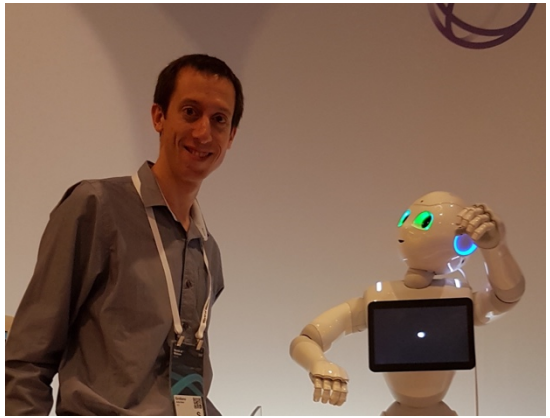
Emiliano Actis Dato

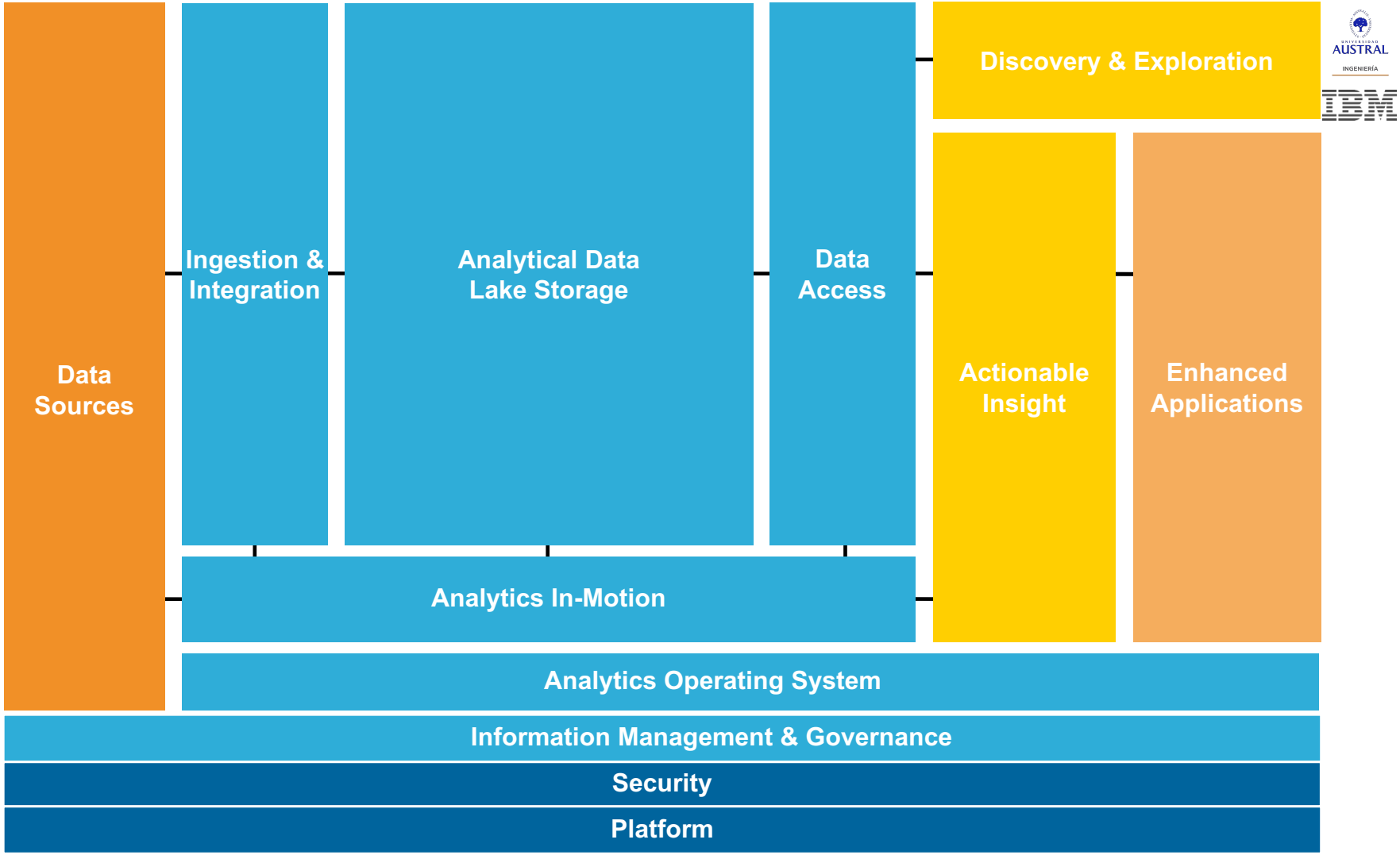
Ingeniero en Electrónica

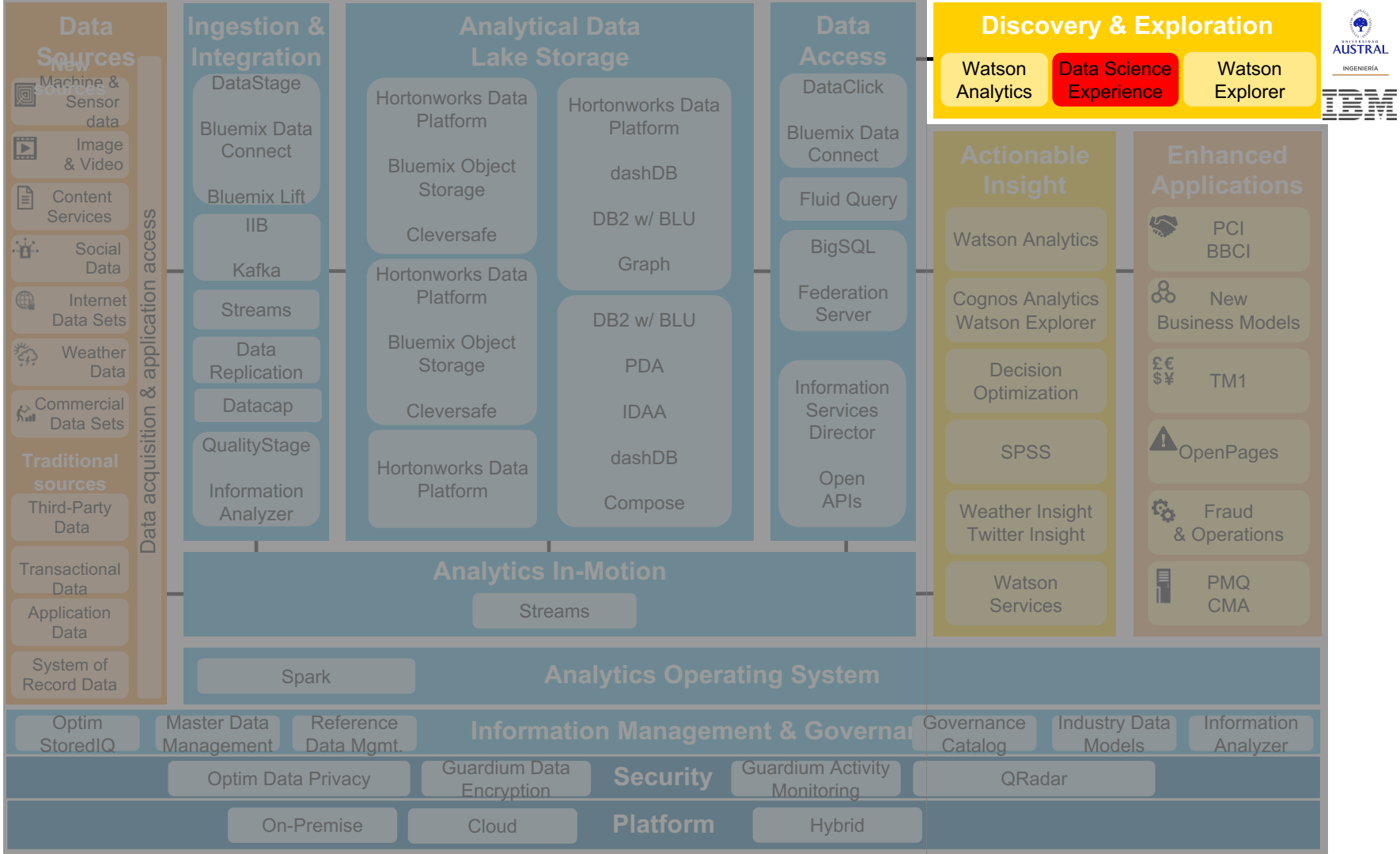
UNLP 1995-2001

11 años en IBM

Analytics Client Architect

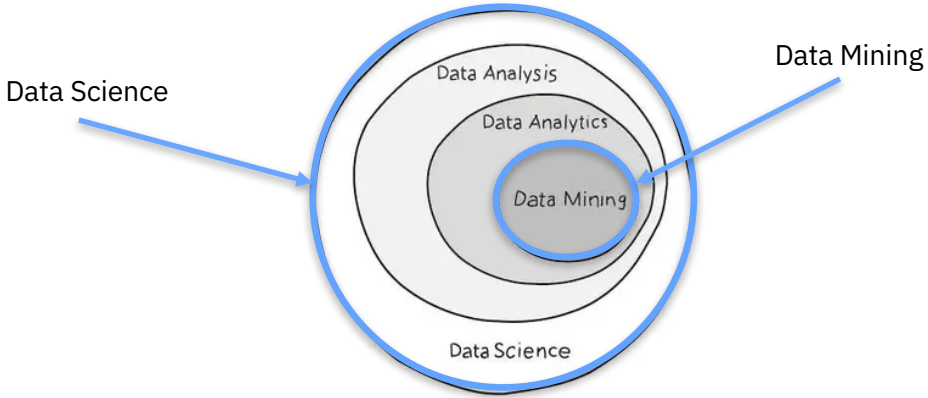



Data Sources
Ingestion & Integration
Analytical Data Lake Storage
Data Access
Discovery & Exploration
Actionable Insight
Enhanced Applications
Analytics In-Motion
Analytics Operating System
Information Management & Governance
Security
Platform


Data Sources
New sources
Machine & Sensor data
Image & Video
Content Services
Social Data
Internet Data Sets
Weather Data
Commercial Data Sets
Traditional sources
Third-Party Data
Transactional Data
Application Data
System of Record Data
Data acquisition & application access
Ingestion & Integration
DataStage
Bluemix Data Connect
Bluemix Lift
IIB
Kafka
Streams
Data Replication
Datacap
QualityStage
Information Analyzer
Analytical Data Lake Storage
Hortonworks Data Platform
Bluemix Object Storage
Cleversafe
Hortonworks Data Platform
Bluemix Object Storage
Cleversafe
Hortonworks Data Platform
Hortonworks Data Platform
dashDB
DB2 w/ BLU
Graph
DB2 w/ BLU
PDA
IDAA
dashDB
Compose
Data Access
DataClick
Bluemix Data Connect
Fluid Query
BigSQL
Federation Server
Information Services Director
Open APIs
Discovery & Exploration
Watson Analytics
Data Science Experience
Watson Explorer
Actionable Insight
Watson Analytics
Cognos Analytics Watson Explorer
Decision Optimization
SPSS
Weather Insight Twitter Insight
Watson Services
Enhanced Applications
PCI BBCI
New Business Models
TM1
OpenPages
Fraud & Operations
PMQ CMA
Analytics In-Motion
Streams
Spark
Analytics Operating System
Optim StoredIQ
Master Data Management
Reference Data Mgmt.
Information Management & Governa
Governance Catalog
Industry Data Models
Information Analyzer
Optim Data Privacy
Guardium Data Encryption
Security
Guardium Activity Monitoring
QRadar
On-Premise
Cloud
Platform
Hybrid

Qué es Data Science?

Métodos para extraer conocimiento o un mejor entendimiento de los datos, cualquiera sea su formato.



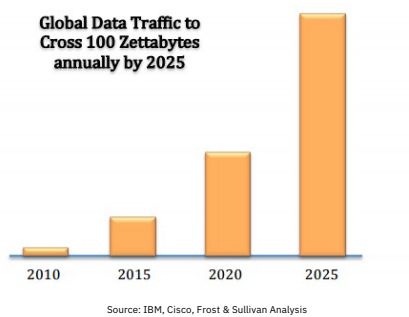
Small Data

lluvia	humedad	viento	golf
si	alta	alto	no
no	media	alto	no
si	baja	medio	no
no	alta	bajo	si

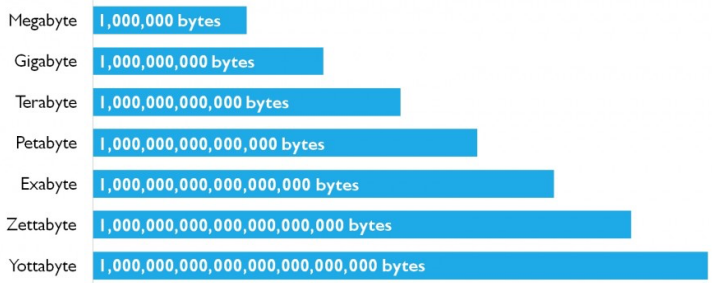
Big Data

Por qué es Data Science Importante?

Cantidad de datos

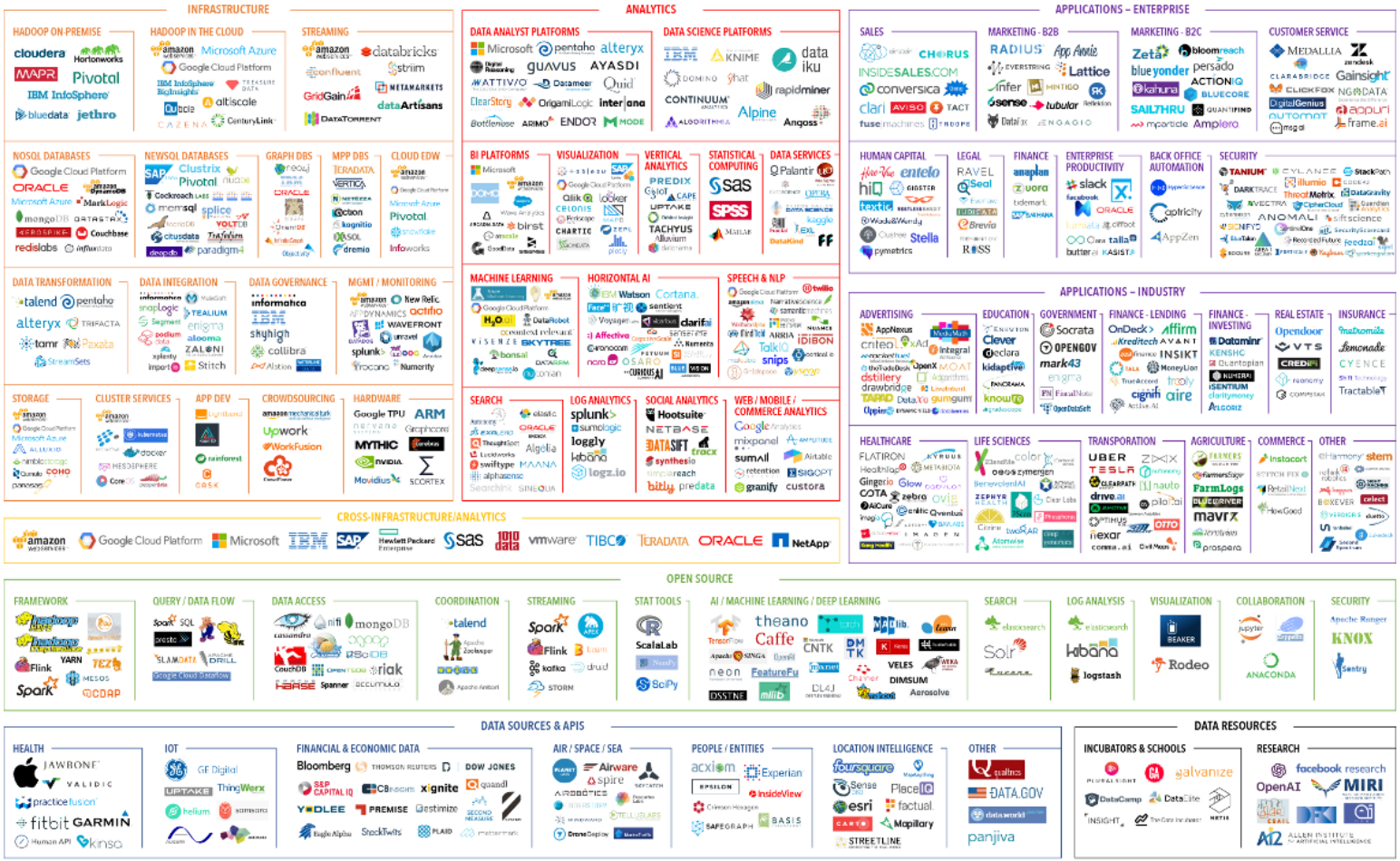


Source: IBM, Cisco, Frost & Sullivan Analysis



Qué hacen las empresas con los datos?

BIG DATA LANDSCAPE 2017



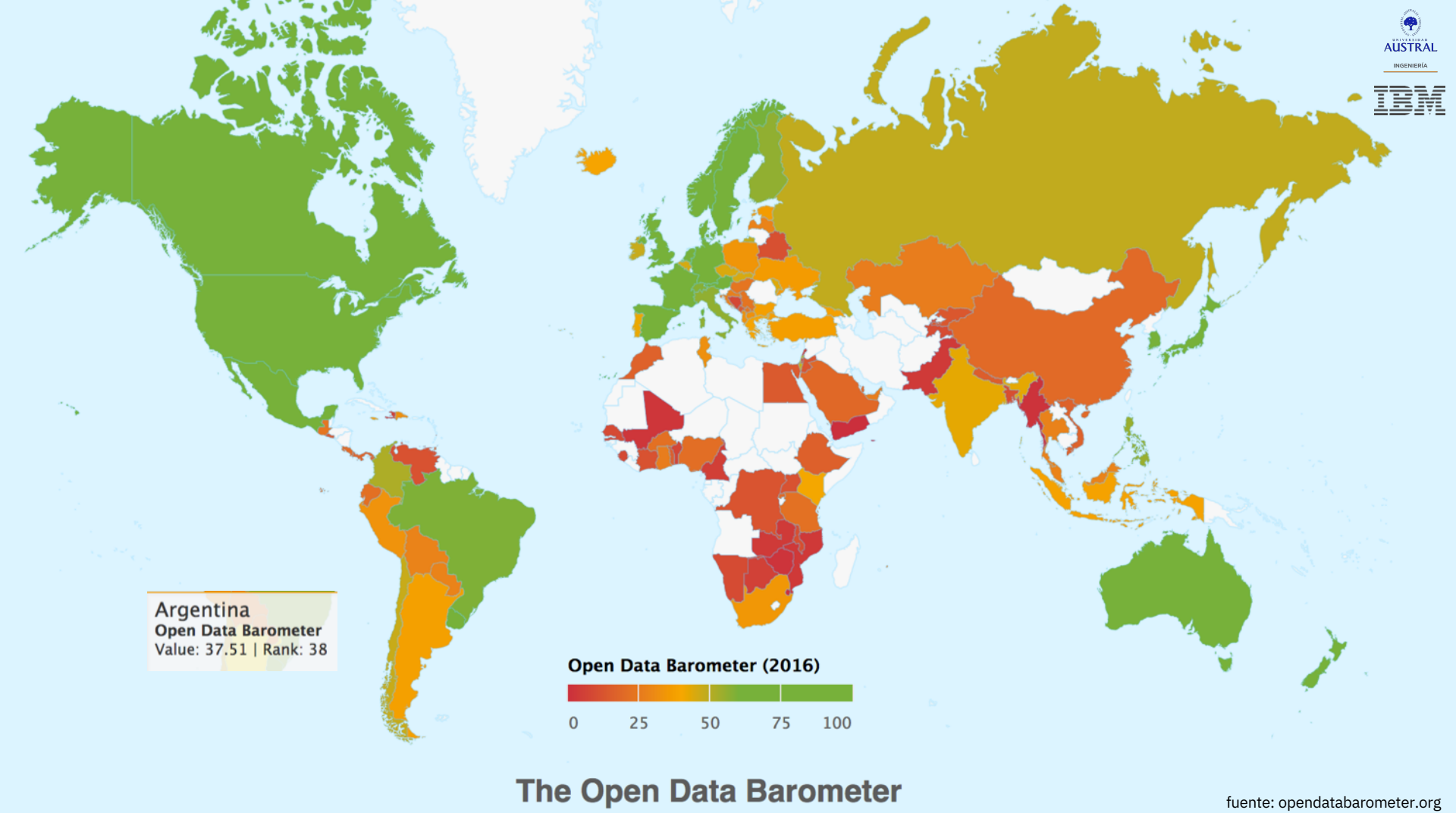
Last updated 4/5/2017

FIRSTMARK
EARLY STAGE VENTURE CAPITAL

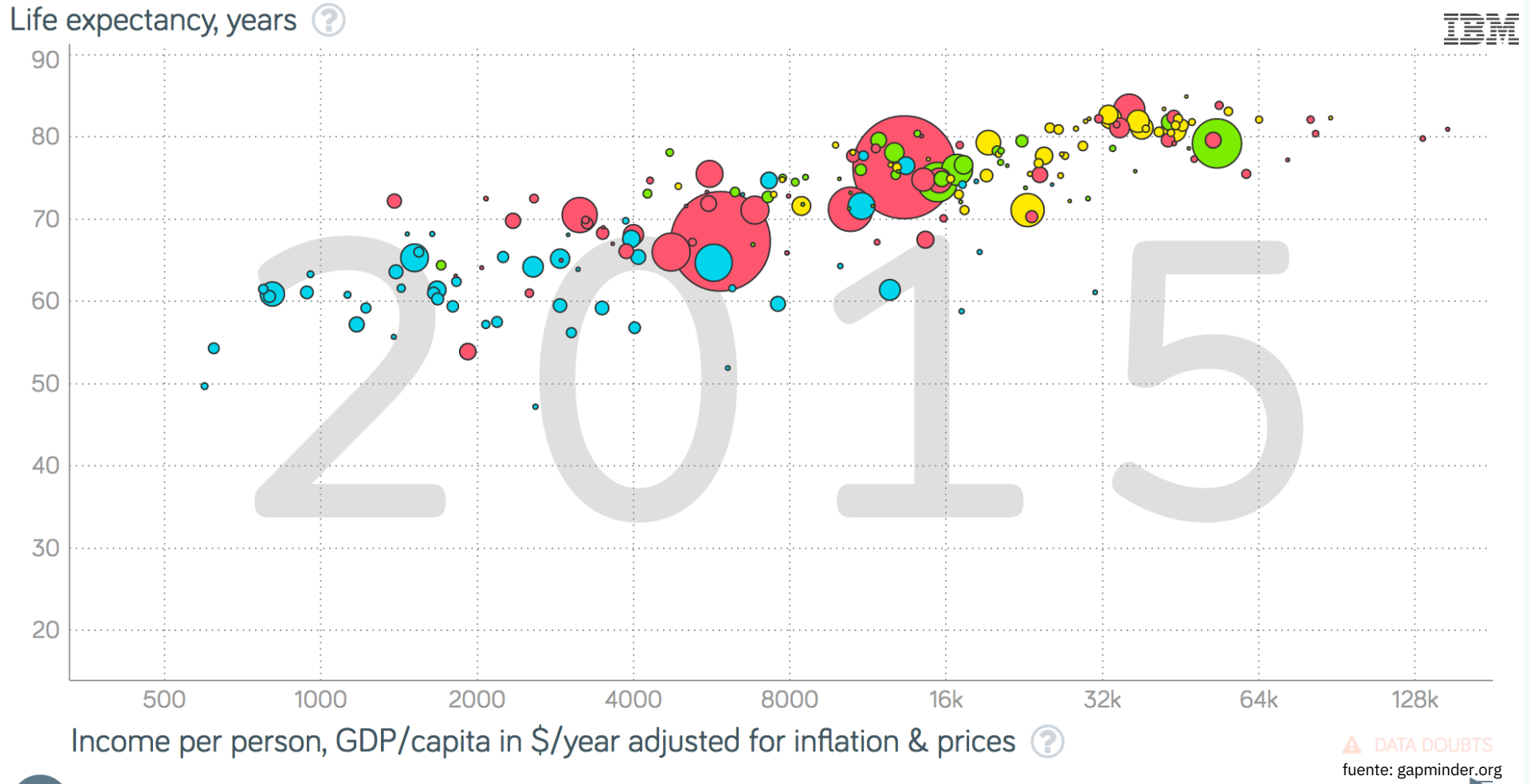
Data Science: Tendencia

fuente: indeed.com

fuente: google trends



fuente: opendatabarometer.org



fuente: gapminder.org

El Data Scientist

- Talento
- Set de herramientas rígido
- Fragmentado y perdiendo tiempo
- Silo analítico

Data Science Experience (DSX)

Aprender	Crear	Colaborar
		

Artículos, Ejemplos, DataSets

ARTICLE
How can data scientists collaborate to build...
SOURCE
IBM
DATE
Jun 24, 2016

ARTICLE
What is machine learning?
SOURCE
IBM
DATE
Jun 24, 2016

NOTEBOOK
Insights from Twitter data about car makers
SOURCE
IBM
DATE
Jun 22, 2016

NOTEBOOK
Insights from New York car accident reports
SOURCE
IBM
DATE
Jun 16, 2016

DATA SET
Country Surface Area (sq. km)
SOURCE
IBM
DATE
Jun 16, 2016

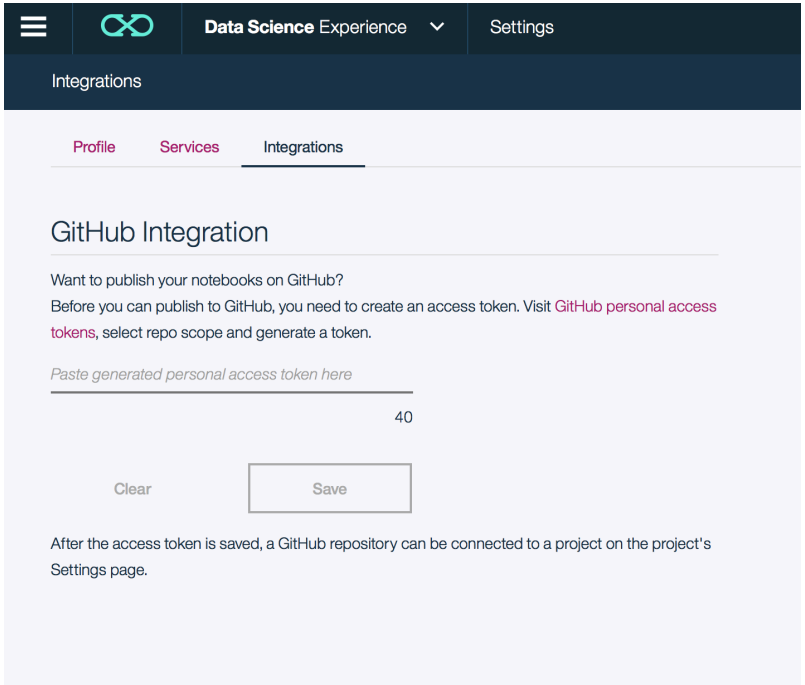
NOTEBOOK
Improved Flight delay prediction
SOURCE
IBM
DATE
Jun 06, 2016

NOTEBOOK
Load data from different sources
SOURCE
IBM
DATE
Jun 02, 2016

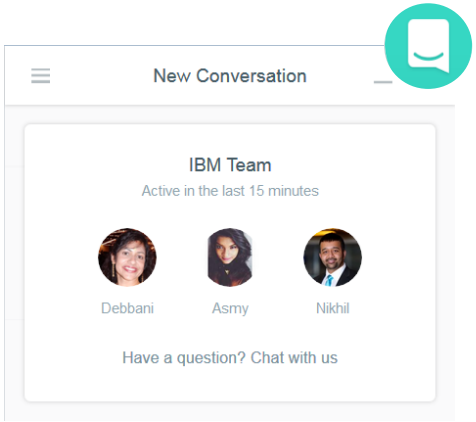
NOTEBOOK
Learn basics about notebooks and Apache Spark
SOURCE
IBM
DATE
Jun 02, 2016

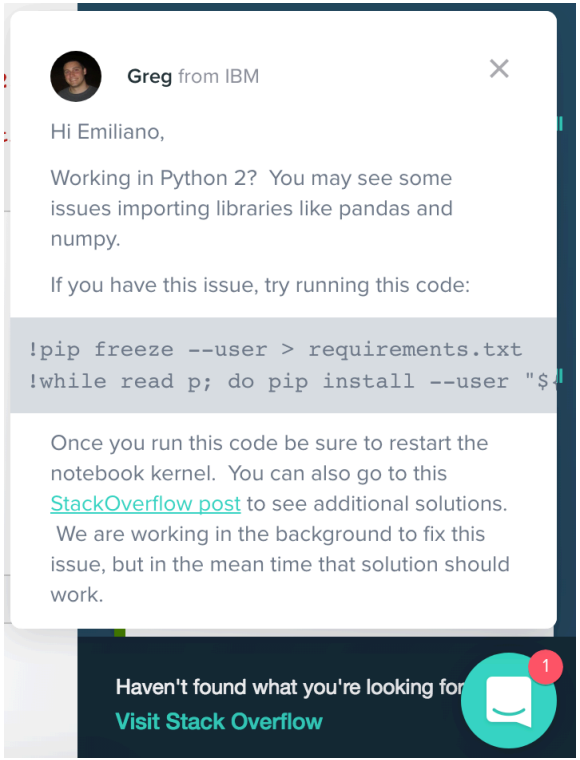
NOTEBOOK
Analyze precipitation data
SOURCE
IBM
DATE
Jun 02, 2016

GitHub

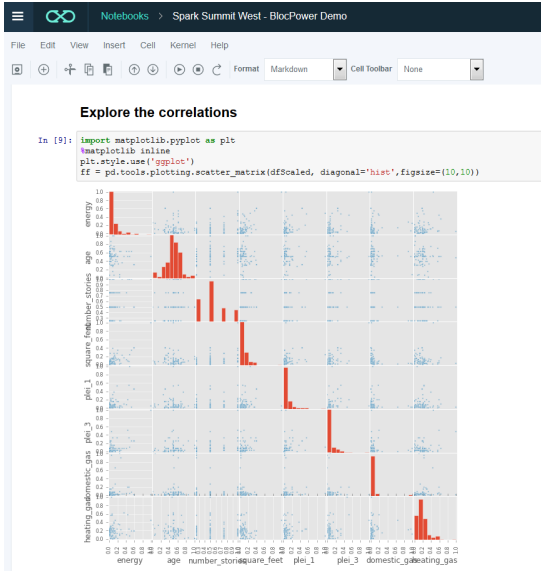

Data Science Experience
Settings
Integrations
Profile
Services
Integrations
GitHub Integration
Want to publish your notebooks on GitHub?
Before you can publish to GitHub, you need to create an access token. Visit GitHub personal access tokens, select repo scope and generate a token.
Paste generated personal access token here
40
Clear
Save
After the access token is saved, a GitHub repository can be connected to a project on the project's Settings page.

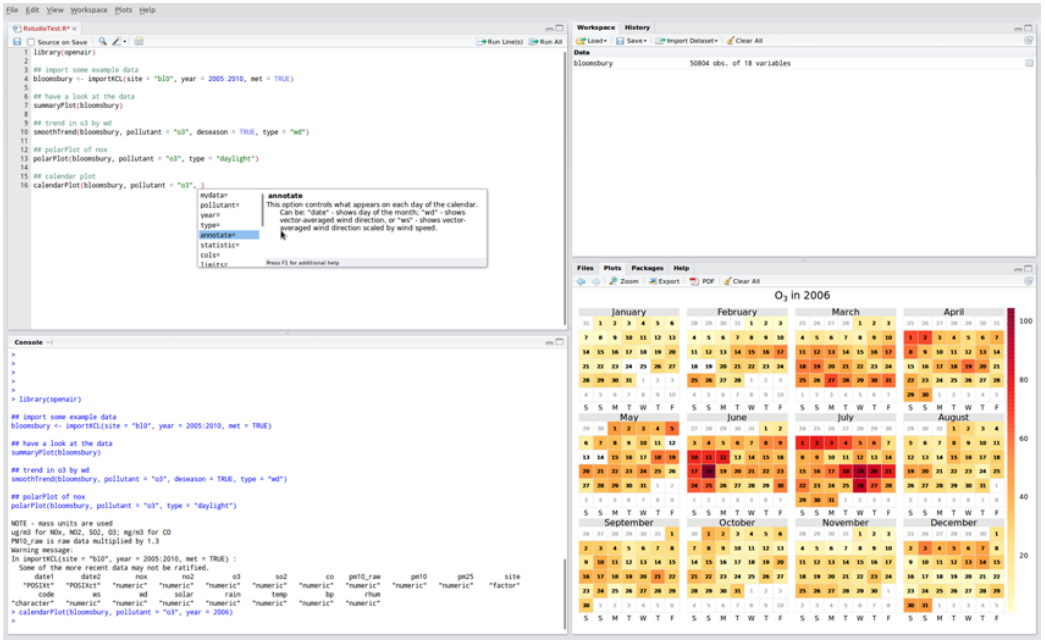
Chat Online


New Conversation
IBM Team
Active in the last 15 minutes
Debbani
Asmy
Nikhil
Have a question? Chat with us


Greg from IBM
Hi Emiliano,
Working in Python 2? You may see some issues importing libraries like pandas and numpy.
If you have this issue, try running this code:
!pip freeze --user > requirements.txt
!while read p; do pip install --user "$
Once you run this code be sure to restart the notebook kernel. You can also go to this StackOverflow post to see additional solutions. We are working in the background to fix this issue, but in the mean time that solution should work.
Haven't found what you're looking for
Visit Stack Overflow
1

Jupyter Notebooks





python R Scala

Complementos

Model Builder

Select Data

Prepare

Train

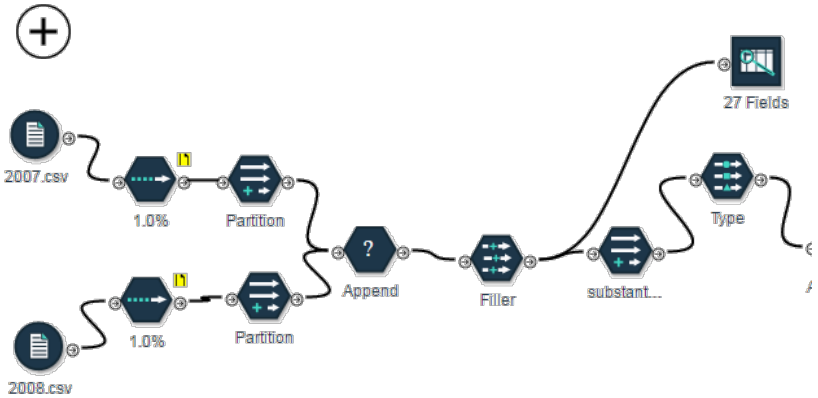
Select Model

Evaluate

Deploy

Predict

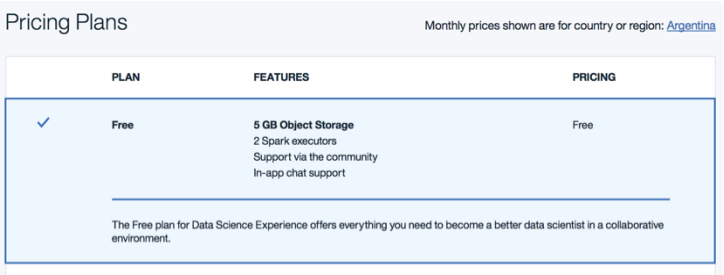
Flow



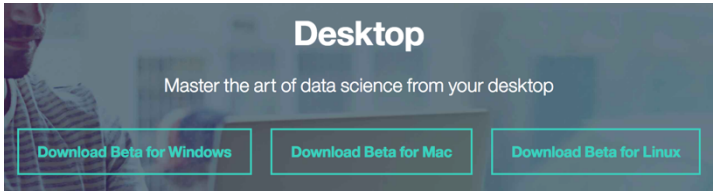
Cómo consigo DSX?

Tres opciones:

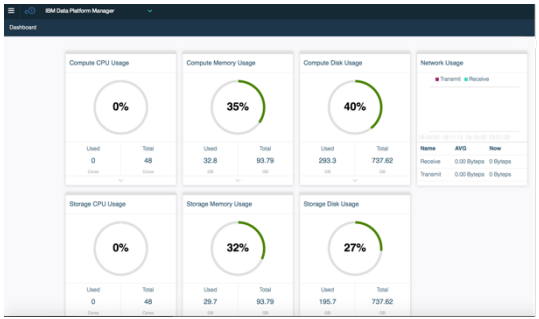
- DSX en Bluemix: http://datascience.ibm.com



- DSX Desktop: http://datascience.ibm.com/desktop



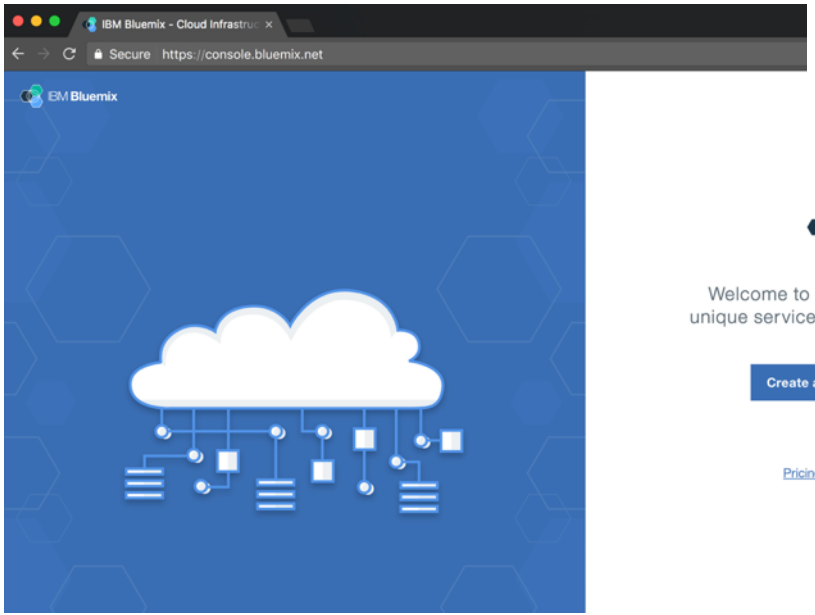
- DSX Local: https://datascience.ibm.com/local

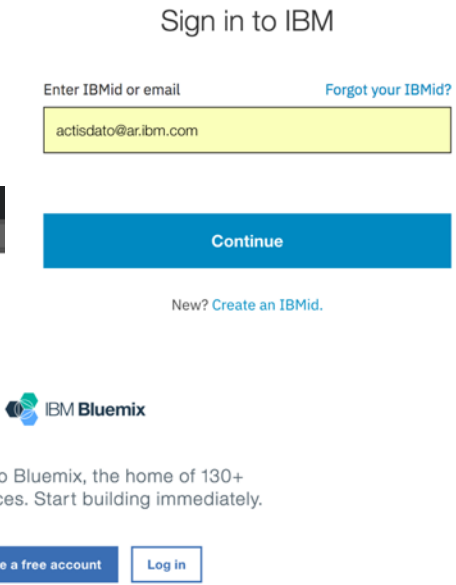


DSX en Bluemix en 3 pasos

Paso 1: Abrir cuenta en Bluemix

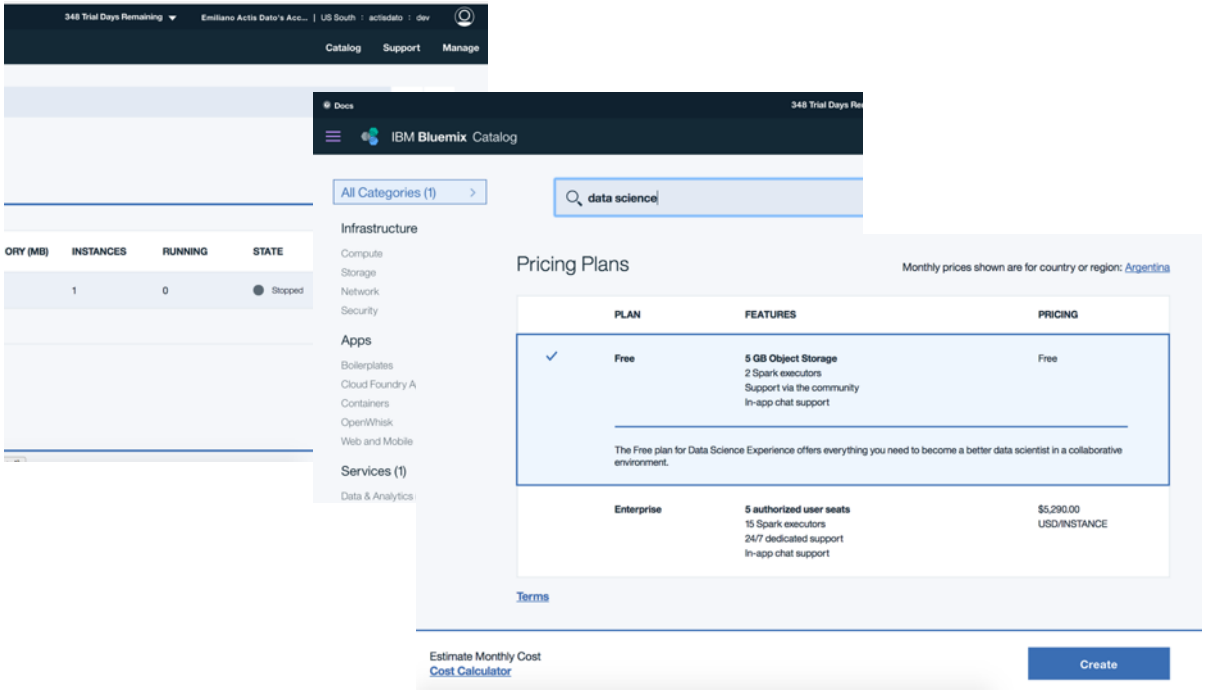
- http://bluemix.net





DSX en Bluemix en 3 pasos

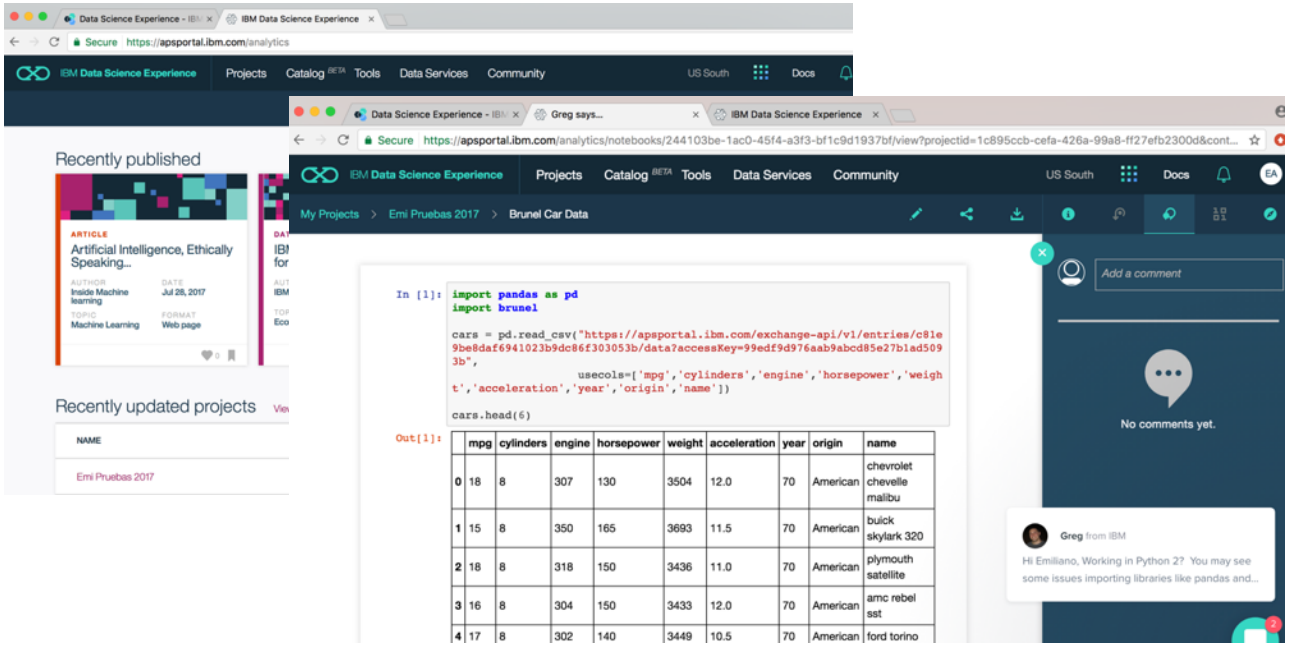
Paso 2: Crear Servicio Data Science Experience



DSX en Bluemix en 3 pasos

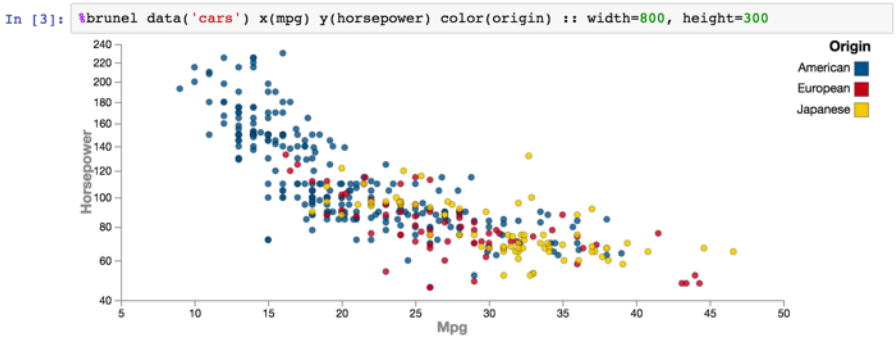
Paso 3: Usar Data Science Experience

- http://datascience.ibm.com

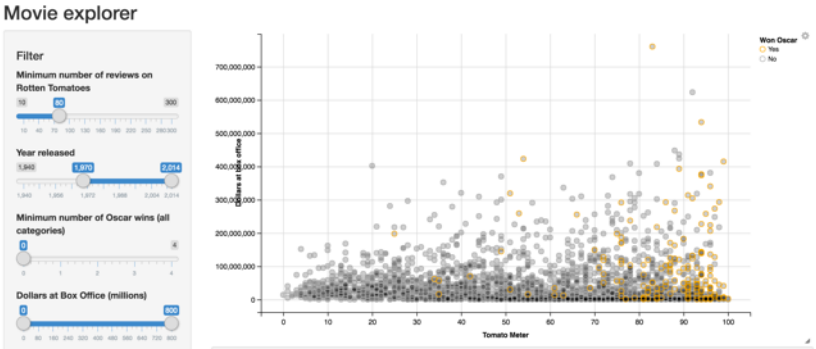


Demo

- Demo 1: Brunel Visualization



- Demo 2: Shiny Movie Explorer



1. Hay muchos datasets disponibles

2. DSX Cloud gratis por 30 días (http://datascience.ibm.com)
   - DSX Desktop gratis sin límite de tiempo

3. El momento es ahora!

Emiliano Actis Dato
Analytics Client Architect
actisdato@ar.ibm.com
@emiactisdato