

>>> Data Science: Las oportunidades en mundo de datos

Name: M. Sc. Lesly Zerna<sup>†</sup>

Date: 8 / 10 / 2018

---

<sup>†</sup>lesly.zerna.bo07@gmail.com

```
>>> Outline
```

1. Introducción
2. Ciencia de Datos
3. Oportunidad en los Datos
4. Visualización de Datos
5. Para terminar...

```
>>> Valor
```



<https://compraraccionesdebolsa.com/mercados-bolsa/minas-de-oro/>

```
>>> Valor
```

En el mundo actual con...  
INTERNET, *Web 2.0*, Redes Sociales,

>>> Valor

En el mundo actual con...  
INTERNET, *Web 2.0*, Redes Sociales,  
dónde está el valor?

## >>> Un mundo de datos



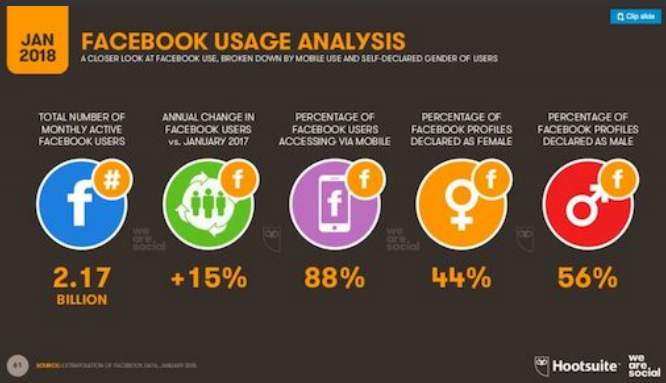
<https://towardsdatascience.com/why-data-science-succeeds-or-fails-c24edd2d2f9>

>>> Datos...



<https://ilifebelt.com/cuantos-usuarios-tiene-facebook-en-el-mundo/2018/02/>

## >>> Datos y mas datos



<https://ilifebelt.com/cuantos-usuarios-tiene-facebook-en-el-mundo/2018/02/>



```
>>> Datos y datos
```



[https://www.pcmag.com/news/363951/  
how-data-science-at-netflix-turned-hollywood-on-its-head](https://www.pcmag.com/news/363951/how-data-science-at-netflix-turned-hollywood-on-its-head)

>>> Trabajar con Datos

Algunas habilidades:

- \* Critical thinking
- \* Resolución de problemas

## >>> Trabajar con Datos

Algunas habilidades:

- \* Critical thinking
- \* Resolución de problemas
- \* Matemáticas
- \* Programación

## >>> Trabajar con Datos

Algunas habilidades:

- \* Critical thinking
- \* Resolución de problemas
- \* Matemáticas
- \* Programación
- \* Comunicación

## >>> Trabajar con Datos

Algunas habilidades:

- \* Critical thinking
- \* Resolución de problemas
- \* Matemáticas
- \* Programación
- \* Comunicación
- \* (extra) Inteligencia Artificial / Machine Learning
- \* ...

## >>> Trabajar con Datos

Algunas habilidades:

- \* Critical thinking
- \* Resolución de problemas
- \* Matemáticas
- \* Programación
- \* Comunicación
- \* (extra) Inteligencia Artificial / Machine Learning
- \* ...
- \* obviamente datos -> Ciencia de Datos

```
>>> Ciencia de Datos
```

```
* Inteligencia Artificial???
```

```
>>> Ciencia de Datos
```

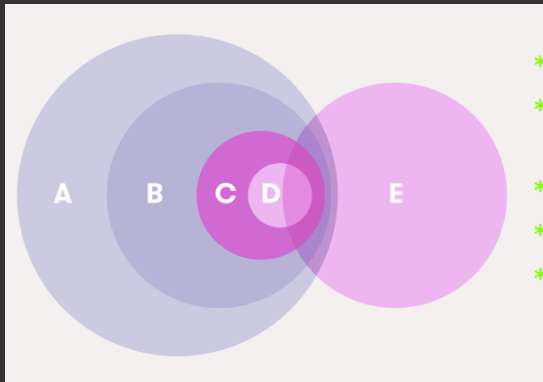
- \* Inteligencia Artificial???
- \* Matemáticas?



```
>>> Ciencia de Datos
```

- \* Inteligencia Artificial???
- \* Matemáticas?
- \* Machine Learning???
- \* ...

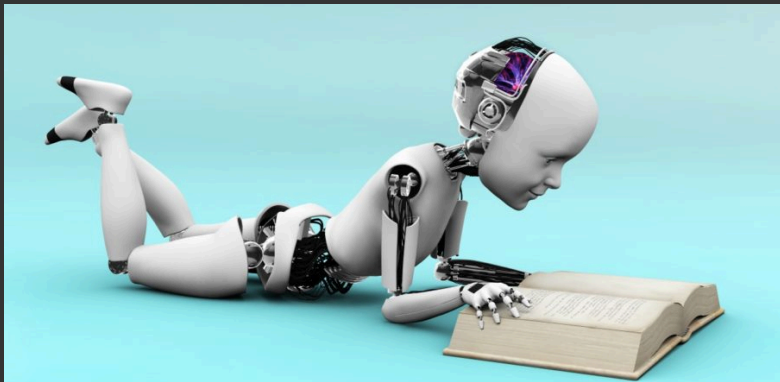
>>> ¿Qué es qué?



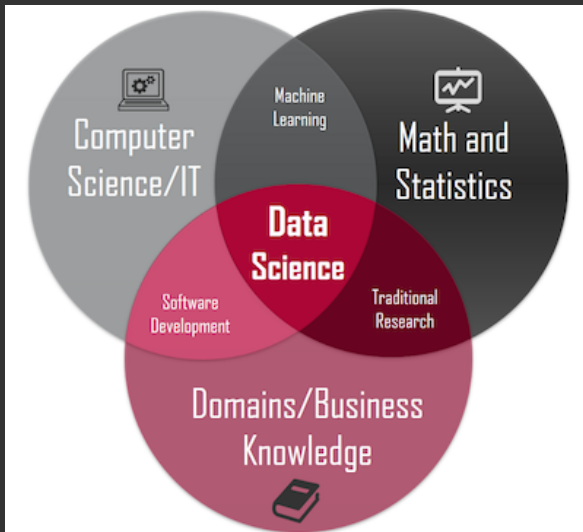
- \* A: Computer Science
- \* B: Artificial Intelligence
- \* C: Machine Learning
- \* D: Deep Learning
- \* E: Data Science

<https://course.elementsofai.com/1/2>

# >>> Machine Learning



## >>> Ciencia de Datos



<https://towardsdatascience.com/why-data-science-succeeds-or-fails-c24edd2d2f9>

```
>>> Ciencia de Datos
```

La persona que trabaja con datos...

## MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

### MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

### DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative



### PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

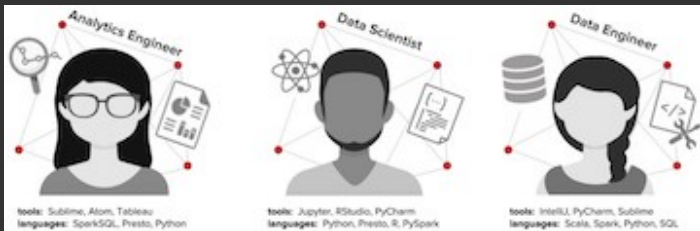
### COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

MarketingDistillery.com is a group of practitioners in the area of e-commerce marketing. Our fields of expertise include marketing strategy and optimization: customer tracking and on-site analytics; predictive analytics and econometrics; data warehousing and big data systems; marketing channel insights in Paid Search, SEO, Social, CRM and brand.

Marketing  
DISTILLERY

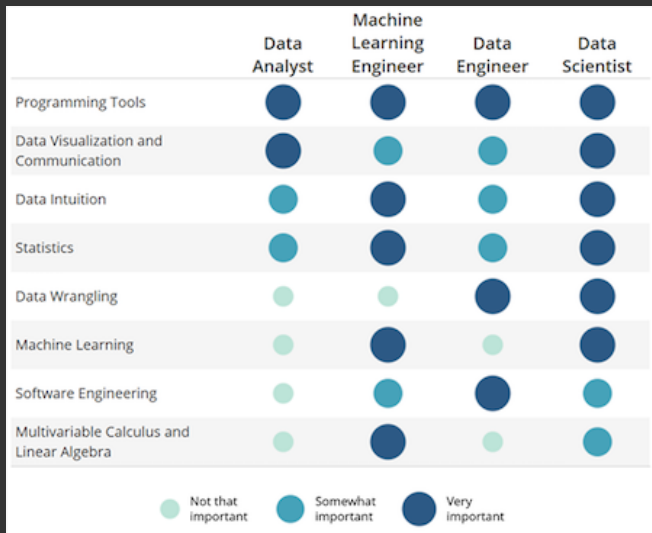
## >>> Científico de Datos



https:

//medium.com/netflix-techblog/notebook-innovation-591ee3221233

## >>> Ciencia de Datos



<https://blog.udacity.com/2014/11/data-science-job-skills.html>



>>> Otros conceptos

- \* Data Analysis

- \* Automatización en la extracción de información

## >>> Otros conceptos

- \* Data Analysis

- \* Automatización en la extracción de información

- \* Big Data

- \* Análisis de la información para toma de decisiones

## >>> Diferentes áreas

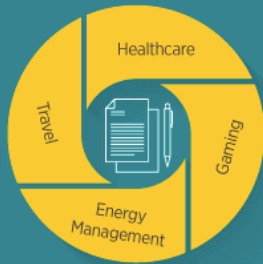
Data Science algorithms are used in industries like:



Big Data is used in industries like:



Data Analytics is used in industries like:



[https://www.simplilearn.com/  
data-science-vs-big-data-vs-data-analytics-article](https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article)

```
>>> Trabajar con Datos
```

¿dónde encontrarlos?

```
>>> Trabajar con Datos
```



```
¿dónde encontrarlos?  
en todo lado
```

```
>>> Trabajar con Datos
```

```
¿dónde encontrarlos?  
en todo lado  
pero...
```

```
>>> Trabajar con Datos
```


```
¿dónde encontrarlos?  
en todo lado  
pero...  
datos estructurados...
```


  

Competitions   Datasets   Kernels   Discussion   Learn   ...   [Sign In](#)

## Kaggle is the place to do data science projects

[See how it works](#) 



**Sign up with just one click:**

We won't share anything without your permission

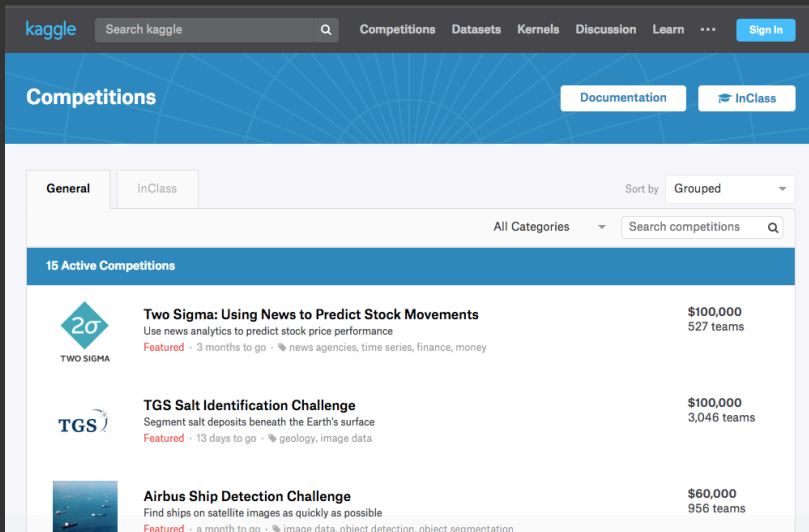
[Google](#)   [Facebook](#)   [Yahoo](#)

**Manually create an account:**




[Sign Up](#)

<https://www.kaggle.com/>



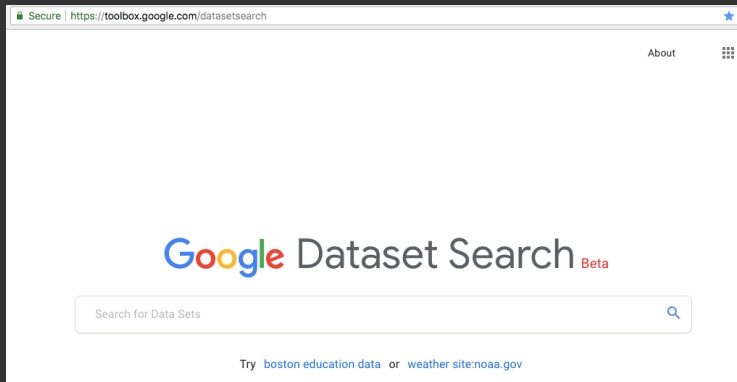


The screenshot shows the Kaggle website's 'Competitions' section. At the top, there's a navigation bar with the Kaggle logo, a search bar, and links for Competitions, Datasets, Kernels, Discussion, Learn, and a Sign In button. Below this is a blue banner with the word 'Competitions' and buttons for 'Documentation' and 'InClass'. The main content area has tabs for 'General' and 'InClass', with 'General' selected. A 'Sort by' dropdown is set to 'Grouped'. Below the tabs is a search bar for competitions. A blue header indicates '15 Active Competitions'. Three competitions are listed: 'Two Sigma: Using News to Predict Stock Movements' with a prize of \$100,000 and 527 teams; 'TGS Salt Identification Challenge' with a prize of \$100,000 and 3,046 teams; and 'Airbus Ship Detection Challenge' with a prize of \$60,000 and 956 teams. Each entry includes a logo, title, description, featured status, time to go, and tags.

Logo	Competition Title	Prize	Teams
	<b>Two Sigma: Using News to Predict Stock Movements</b> Use news analytics to predict stock price performance <i>Featured</i> · 3 months to go · news agencies, time series, finance, money	\$100,000	527 teams
	<b>TGS Salt Identification Challenge</b> Segment salt deposits beneath the Earth's surface <i>Featured</i> · 13 days to go · geology, image data	\$100,000	3,046 teams
	<b>Airbus Ship Detection Challenge</b> Find ships on satellite images as quickly as possible <i>Featured</i> · a month to go · image data, object detection, object segmentation	\$60,000	956 teams

<https://www.kaggle.com/>

>>> Data Set



`https://toolbox.google.com/datasetsearch`

>>> Oportunidad en los Datos

- \* Encontrar información

## >>> Oportunidad en los Datos

- \* Encontrar información
- \* Tomar decisiones

## >>> Oportunidad en los Datos

- \* Encontrar información
- \* Tomar decisiones
- \* Crear nuevos productos

>>> Cómo empezar

Entender los datos

>>> Cómo empezar

Entender los datos  
Estadística Descriptiva

```
>>> Estadística Descriptiva
```

Datos



# >>> Estadística Descriptiva

## Datos

- \* Cuantitativos
- \* Categóricos

>>> Estadística Descriptiva

Datos

- \* Cuantitativos

# >>> Estadística Descriptiva

## Datos

- \* Cuantitativos
  - \* Discretos
  - \* Continuos

# >>> Estadística Descriptiva

## Datos

- \* Cuantitativos
  - \* Discretos
  - \* Continuos
- \* Categóricos

# >>> Estadística Descriptiva

## Datos

- \* Cuantitativos

- \* Discretos

- \* Continuos

- \* Categóricos

- \* Ordinales

- \* Nominales

>>> Estadística Descriptiva

Cuantitativos

# >>> Estadística Descriptiva

Cuantitativos

- \* Discretos

# >>> Estadística Descriptiva

Cuantitativos

- \* Discretos

- \* Continuos



>>> Estadística Descriptiva

Categoricos

# >>> Estadística Descriptiva

## Categoricos

- \* Ordinales:

- \* Ordenados

- \* Ejemplo: Rating, Ranking ...

# >>> Estadística Descriptiva

## Categoricos

- \* Ordinales:

- \* Ordenados

- \* Ejemplo: Rating, Ranking ...

- \* Nominales

- \* No ordenados

- \* Ejemplo: razas de perro, ...

# >>> Estadística Descriptiva

## Categoricos

- \* Ordinales:

- \* Ordenados

- \* Ejemplo: Rating, Ranking ...

- \* Nominales

- \* No ordenados

- \* Ejemplo: razas de perro, ...

```
>>> Estadística- Info de los Datos
```

Centro

- \* Promedio

```
>>> Estadística- Info de los Datos
```

Centro

- \* Promedio
- \* Mediana

```
>>> Estadística- Info de los Datos
```

Centro

- \* Promedio
- \* Mediana
- \* Moda

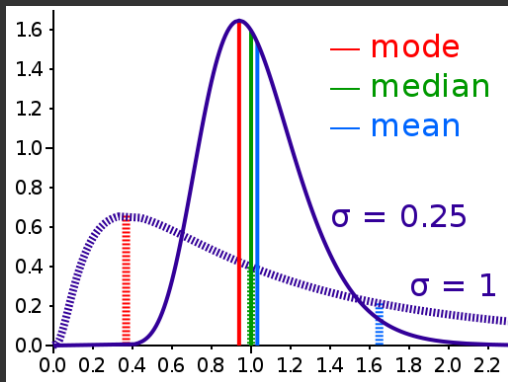
## >>> Datos: Tendencias Centrales



<https://www.bbc.com/bitesize/articles/z99jpbk>



## >>> Tendencias Centrales



[https://commons.wikimedia.org/wiki/File:Comparison\\_mean\\_median\\_mode.svg](https://commons.wikimedia.org/wiki/File:Comparison_mean_median_mode.svg)

>>> Otros conceptos importantes:

**Sample Variance**

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

**Sample Standard Deviation**

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

<http://statisticslectures.com/topics/variancesample/>

>>> Otros conceptos importantes:

- \* Desviación Standard

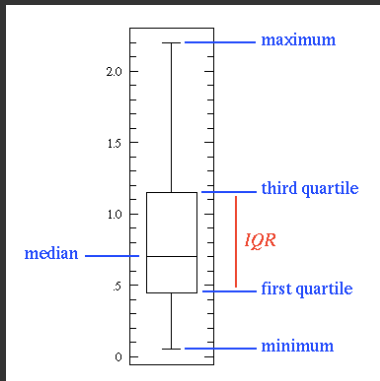
```
>>> Otros conceptos importantes:
```

- \* Desviación Standard

- \* Varianza

## >>> Conceptos de Dispersión

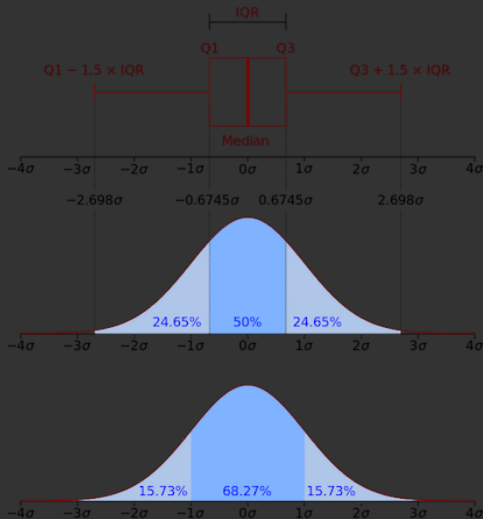
### Box-Plot



<http://www.physics.csbsju.edu/stats/box2.html>

## >>> Conceptos de Dispersión

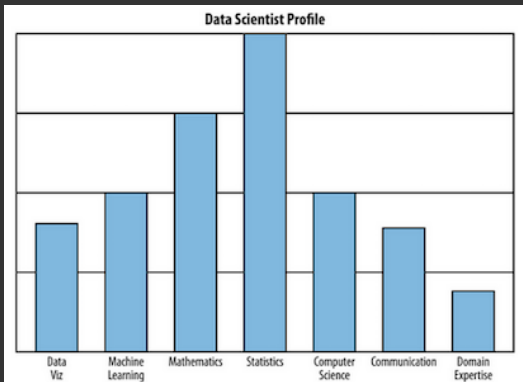
### Box-Plot



[https://en.wikipedia.org/wiki/Box\\_plot](https://en.wikipedia.org/wiki/Box_plot)

# >>> Conceptos de Dispersión

## Histograma



[https://machinelearningmastery.com/  
data-science-skills-simple-method-can-use-evaluate-use-results-build](https://machinelearningmastery.com/data-science-skills-simple-method-can-use-evaluate-use-results-build/)

```
>>> Estadísticas
```

Entender los DATOS



```
>>> Estadísticas
```

- \* Estadística Descriptiva

>>> Estadísticas

- \* Estadística Descriptiva
- \* Estadística Inferencial

```
>>> Estadísticas
```

- \* Estadística Descriptiva
- \* Estadística Inferencial
  - \* Conclusiones

>>> Herramientas

Manos a la obra!

- \* Básico:

>>> Herramientas

Manos a la obra!

- \* Básico:

- \* Excel

- \* Google SpreadSheet

>>> Herramientas

Manos a la obra!

- \* Básico:

- \* Excel

- \* Google SpreadSheet

- \* Complejo:

```
>>> Herramientas
```

Manos a la obra!

- \* Básico:

- \* Excel

- \* Google SpreadSheet

- \* Complejo:

- \* Hadoop

- \* Apache Spark

```
>>> Herramientas
```

Manos a la obra!

- \* Básico:

- \* Excel

- \* Google SpreadSheet

- \* Complejo:

- \* Hadoop

- \* Apache Spark

- \* Otras:



```
>>> Herramientas
```

Manos a la obra!

- \* Básico:

- \* Excel

- \* Google SpreadSheet

- \* Complejo:

- \* Hadoop

- \* Apache Spark

- \* Otras:

- \* Tableau

>>> Herramientas

Manos a la obra!

- \* Programación

```
>>> Herramientas
```

Manos a la obra!

- \* Programación

- \* Python

- \* R

```
>>> Herramientas
```

Manos a la obra!

- \* Programación

- \* Python

- \* R

- \* Consultas Base de Datos

```
>>> Herramientas
```

Manos a la obra!

- \* Programación

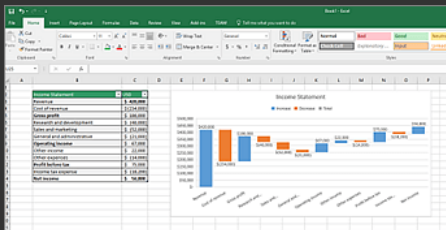
- \* Python

- \* R

- \* Consultas Base de Datos

- \* SQL

# >>> Herramientas



```
>>> Visualización de Datos
```

¿Porqué es importante?

>>> Visualización de Datos

¿Porqué es importante?

- \* Reportes

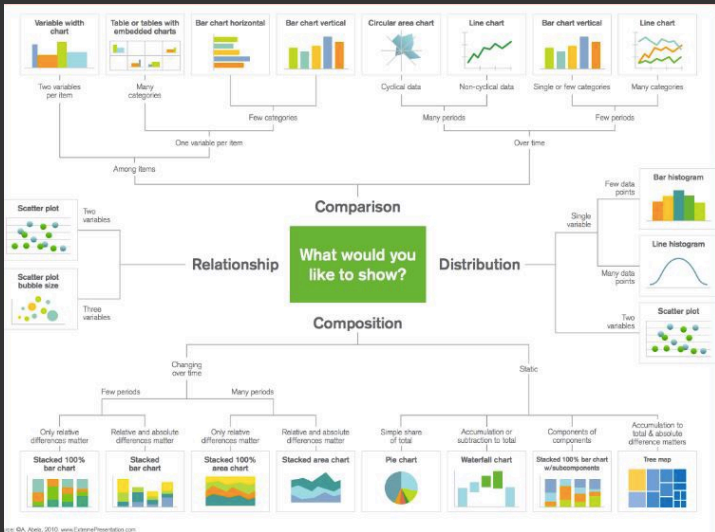


## >>> Visualización de Datos

¿Porqué es importante?

- \* Reportes
- \* Presentaciones frente a personas no técnicas

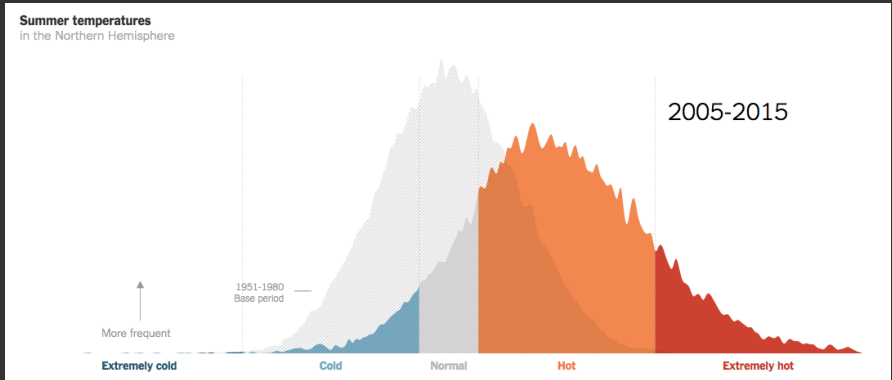
```
>>> Explicar con los datos
```



<https://towardsdatascience.com/>

5-quick-and-easy-data-visualizations-in-python-with-code-a2284bae952

## >>> Visualización de Datos



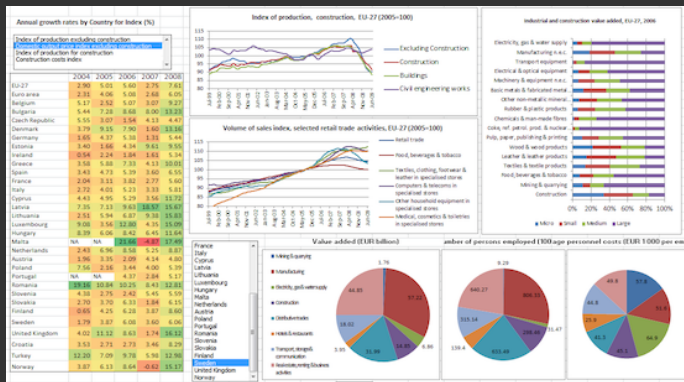
<https://www.nytimes.com/interactive/2017/07/28/climate/more-frequent-extreme-summer-heat.html>

# >>> Ejemplos



<https://apandre.wordpress.com/tools/excel/dashboards-with-excel/>

# >>> Ejemplos

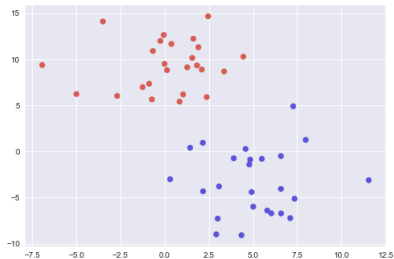


<https://apandre.wordpress.com/tools/excel/dashboards-with-excel/>

## >>> Ejemplos

```
: # Create a random set of examples
from sklearn.datasets.samples_generator import make_blobs
X, Y = make_blobs(n_samples=50, centers=2, random_state=23, cluster_std=2.90)

plt.scatter(X[:, 0], X[:, 1], c=Y, cmap=cmap)
plt.show()
```



[https://github.com/l3s777/CCBOL18\\_DataScience](https://github.com/l3s777/CCBOL18_DataScience)

```
>>> Productos Data-driven
```

- \* Decisiones:
  - \* Spotify
  - \* Netflix ...

>>> Productos Data-driven

- \* Decisiones:

- \* Spotify

- \* Netflix ...

- \* La necesidad del cliente



>>> Productos Data-driven

- \* Decisiones:

- \* Spotify

- \* Netflix ...

- \* La necesidad del cliente

- \* DATOS: poder de conocer

>>> Resumen

\* Datos y Datos

>>> Resumen

- \* Datos y Datos

- \* Matemáticas

>>> Resumen

- \* Datos y Datos
- \* Matemáticas
- \* Diferentes perfiles

>>> Resumen

- \* Datos y Datos
- \* Matemáticas
- \* Diferentes perfiles
- \* Oportunidades en los datos

>>> Resumen

- \* Datos y Datos
- \* Matemáticas
- \* Diferentes perfiles
- \* Oportunidades en los datos
- \* Entender los DATOS

>>> Resumen

- \* Datos y Datos
- \* Matemáticas
- \* Diferentes perfiles
- \* Oportunidades en los datos
- \* Entender los DATOS
- \* Estadística

```
>>> Referencias
```

- \* Cursos online: Udacity, Coursera



>>> Referencias

- \* Cursos online: Udacity, Coursera
- \* Google, Microsoft, IBM, Amazon Tools

## >>> Referencias

- \* Cursos online: Udacity, Coursera
- \* Google, Microsoft, IBM, Amazon Tools
- \* Comunidad Tecnológica

>>> Contacto

@leslysandra  
lesly.zerna.bo07@gmail.com

Muchas gracias!