

ACÀMICA

¡Bienvenidas/os a Data Science!



Agenda

¿Cómo anduvieron?

¿Cómo seguir?

ENCUESTA FINAL

Break

Actividad: trabajando con mis datos



¿Dónde estamos?



¿Cómo anduvieron?



Hoja de ruta

fase	ADQUISICIÓN Y EXPLORACIÓN		MODELADO				DEPLOY
	Exploración de datos	Feature Engineering	Machine Learning: Clasificación y Regresión	Optimización de parámetros	Procesam. del lenguaje natural	Sistema de recomendación	Publicación de modelos
tiempo	SEM 1	SEM 5	SEM 8	SEM 12	SEM 14	SEM 18	SEM 22
	SEM 2	SEM 6	SEM 9	SEM 13	SEM 15	SEM 19	SEM 23
	SEM 3	SEM 7	SEM 10		SEM 16	SEM 20	SEM 24
	SEM 4		SEM 11		SEM 17	SEM 21	



Cronograma

Usted
Está Aquí

SEM 20

- PCA
- Sistemas de recomendación

SEM 21

- Sistemas de recomendación
- Trabajar en la entrega 6 + Empleabilidad

SEM 22

- Trabajar en la entrega 6 + Empleabilidad
- Puesta en producción

SEM 23

- Puesta en producción
- Trabajar en la entrega 7

SEM 24

- Cierre de cursada

Entrega 6

Entrega 7



Programar como proceso creativo



¿A través de qué creen que avanza la humanidad?





A través de **procesos creativos**
que generan nuevas
herramientas y contenidos.

Procesos creativos

- **Nadie crea algo de cero**

Procesos creativos

- **Nadie crea algo de cero**
- Todo proceso creativo (científico, artístico o técnico) sea apoya en procesos previos.
- Las licencias y derechos de autor surgen para darle margen a los autores y fomentar nuevas creaciones.
- Hoy (lamentablemente) muchas veces funcionan solo para garantizar ganancias de algunas empresas y obstaculizan el avance



Procesos creativos

Recomendamos MUCHO el siguiente documental





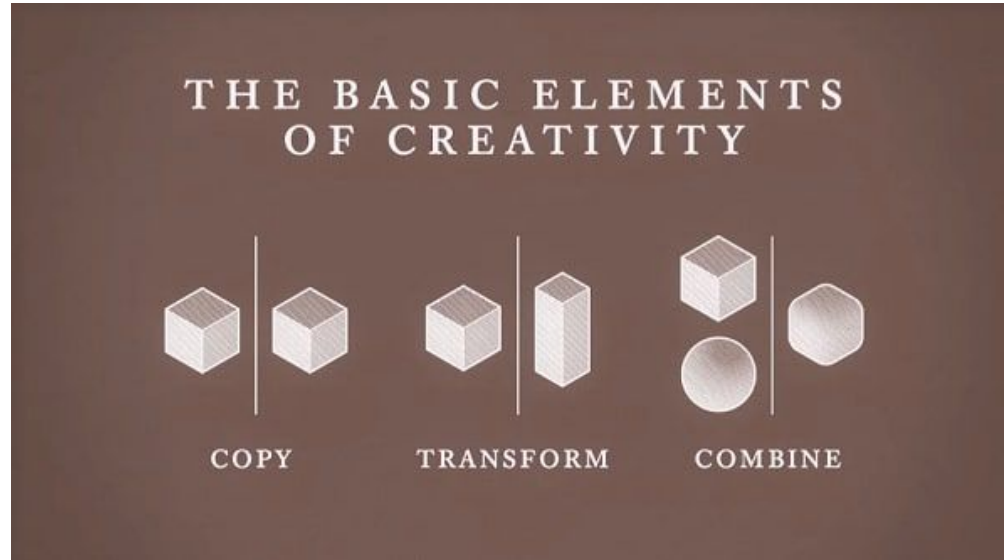


Zeppelin copied without making fundamental changes.
O Led Zeppelin copiava sem fazer mudanças fundamentais.

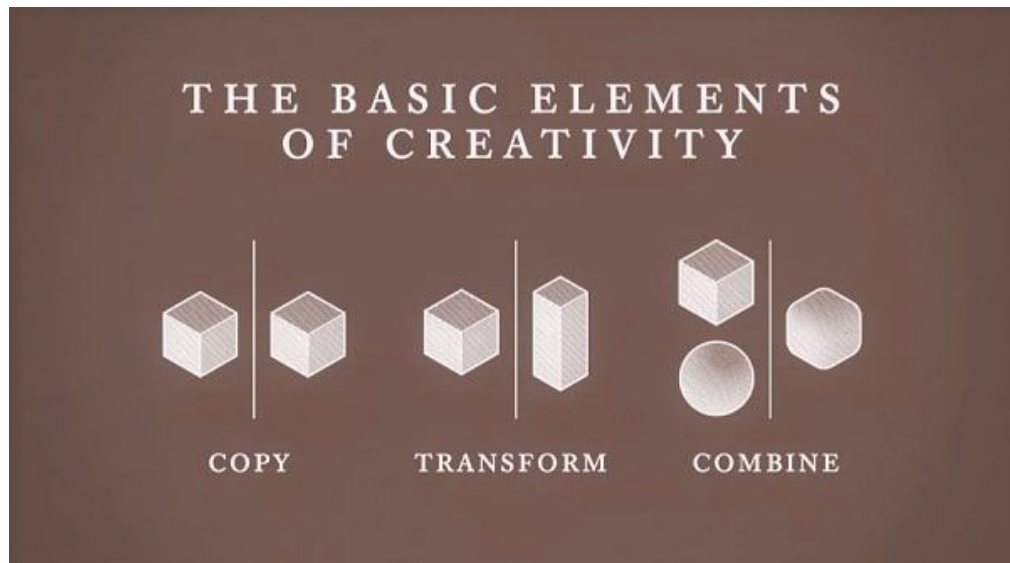
4:38 / 37:30

Everything is a Remix Remastered (2015 HD)

¡Programar ES hacer un remix!



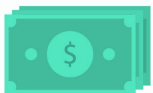
¡Programar ES hacer un remix!



- No empezamos de cero, es imposible (y estaría mal intentarlo).
- Adaptamos algo que otro/a hizo a nuestro contexto.

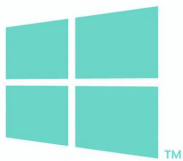
Licencias en programación

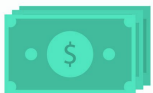




Programas pagos

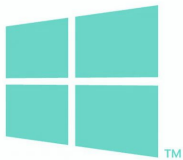
Hay programas que se “compran”, uno como consumidor paga una licencia para poder usarlos:





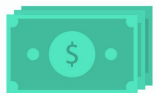
Programas pagos

Hay programas que se “compran”, uno como consumidor paga una licencia para poder usarlos:



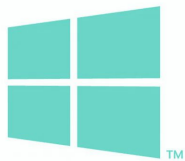
Estos programas **son cajas negras.**

En general, se tiene acceso a versiones ya compiladas de estos programas (binary), pero no tenemos a la vista las líneas de código que los hacen funcionar (source code).



Programas pagos

Hay programas que se “compran”, uno como consumidor paga una licencia para poder usarlos:



Estos programas **son cajas negras**.

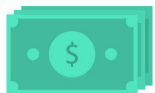
En general, se tiene acceso a versiones ya compiladas de estos programas (binary), pero no tenemos a la vista las líneas de código que los hacen funcionar (source code).



Programas gratuitos

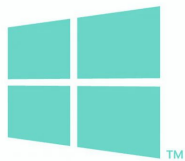
Freeware: son gratuitos (no tenemos que pagar por ellos), pero no tenemos acceso al código de los mismos.





Programas pagos

Hay programas que se “compran”, uno como consumidor paga una licencia para poder usarlos:



Estos programas **son cajas negras**.

En general, se tiene acceso a versiones ya compiladas de estos programas (binary), pero no tenemos a la vista las líneas de código que los hacen funcionar (source code).



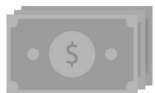
Programas gratuitos

Freeware: son gratuitos (no tenemos que pagar por ellos), pero no tenemos acceso al código de los mismos.



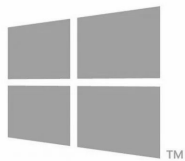
Open source: son gratuitos y el código es de libre acceso.





Programas pagos

Hay programas que se “compran”, uno como consumidor paga una licencia para poder usarlos:



Estos programas **son cajas negras**.

En general, se tiene acceso a versiones ya compiladas de estos programas (binary), pero no tenemos a la vista las líneas de código que los hacen funcionar (source code).



Programas gratuitos

Freeware: son gratuitos (no tenemos que pagar por ellos), pero no tenemos acceso al código de los mismos.



Open source: son gratuitos y el código es de libre acceso.



Hay esperanza...

Open source: son gratuitos y el código es de libre acceso.



No sólo en formas de programas, la gente suele compartir código para las más diversas funciones con la comunidad:



El espíritu colaborativo de la comunidad en el ámbito de la programación, permite socializar el conocimiento y que todos salgamos beneficiados.

**¡HAY QUE
APROVECHARLO
Y COLABORAR!**

¿Podemos copiar el código?



Sobre copiar el código

Cuando hilamos fino, aparecen muchas dudas.

- ¿Podemos copiarlo? ¿Cual es la unidad mínima que podemos copiar?
- ¿Podemos venderlo?
- ¿Cuenta si copiamos la idea de un algoritmo?
- ¿Qué pasa si usamos muchos pedazos de diferentes códigos para armar el nuestro?

¡LICENCIAS!

Sobre copiar el código

Hay varios tipos de licencias para amplio espectro de contenido (código, hardware , imágenes, etc..), solo por mencionar algunas:

- [GNU General Public License](#) Permite copiar, distribuir (en forma física o digital) y cobrar si el código es parte de un producto. La única condición es que el código que escribas a partir de este, lleve también licencia GNU (va en el código mismo).
- [GNU LESSER GENERAL PUBLIC LICENSE](#) Parecido al anterior, pero permite que otro código que usa partes del mismo no sea GNU.
- [BSD license](#) Otra licencia poco restrictiva bastante popular
- [Creative Commons \(CC\)](#) Hay varios tipos de estas licencias, muy usada en diseño. En general hay que darle crédito al autor del material y suelen no permitir la distribución con fines comerciales.

Sobre copiar el código

Hay varios tipos de licencias para amplio espectro de contenido (código, hardware , imágenes, etc..), solo por mencionar algunas:

MIT License Esta es la licencia más permisiva, solo consta de este texto:

Copyright <YEAR> <COPYRIGHT HOLDER>

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

<https://choosealicense.com/>

<https://fossbytes.com/open-sources-license-type/>



[ENCUESTA FINAL](#)



A close-up photograph of a white ceramic cup filled with a latte. The surface of the milk is decorated with intricate latte art, featuring a central heart shape surrounded by concentric, wavy lines. The cup is placed on a matching white saucer. In the background, a white napkin and a silver fork are visible, though they are out of focus. The overall lighting is soft and even, highlighting the textures of the coffee and the smooth surface of the cup.

¡BREAK!



Recursos



Inferencia Bayesiana

Si alguien quiere continuar formándose en Probabilidad y Estadística, un buen comienzo es:

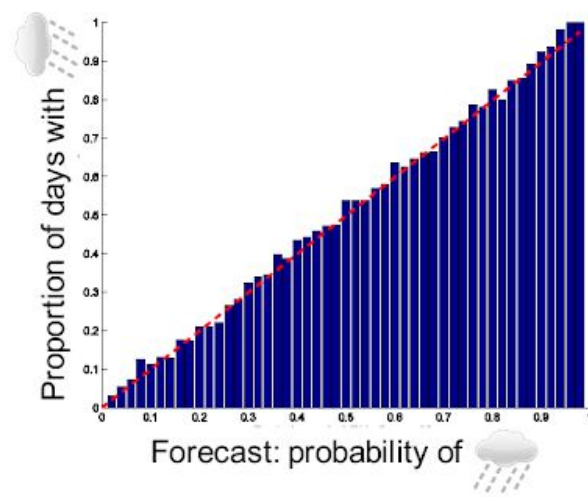
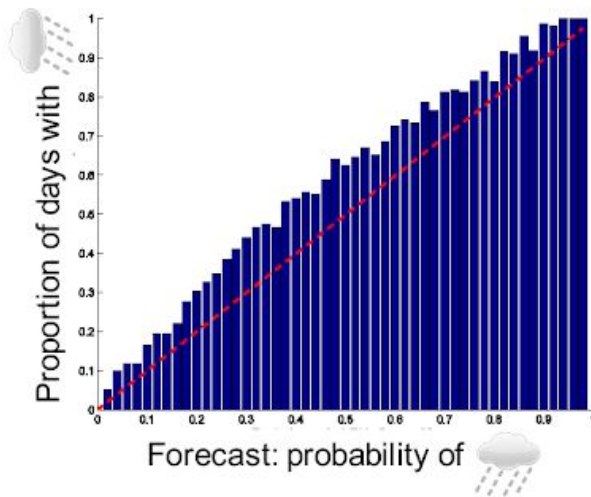
- [Bayesian Methods for Hackers](#)

o el repositorio

- [GitHub - Bayesian Method for Hackers](#) (ambas versiones son oficiales)



Calibración



- En general, `.predict_proba()` devuelve probabilidades mal calibradas.
- ¡Los rankings están bien!



Calibración

[Predicting Good Probabilities With Supervised Learning](#), Alexandru Niculescu-Mizil y Rich Caruana

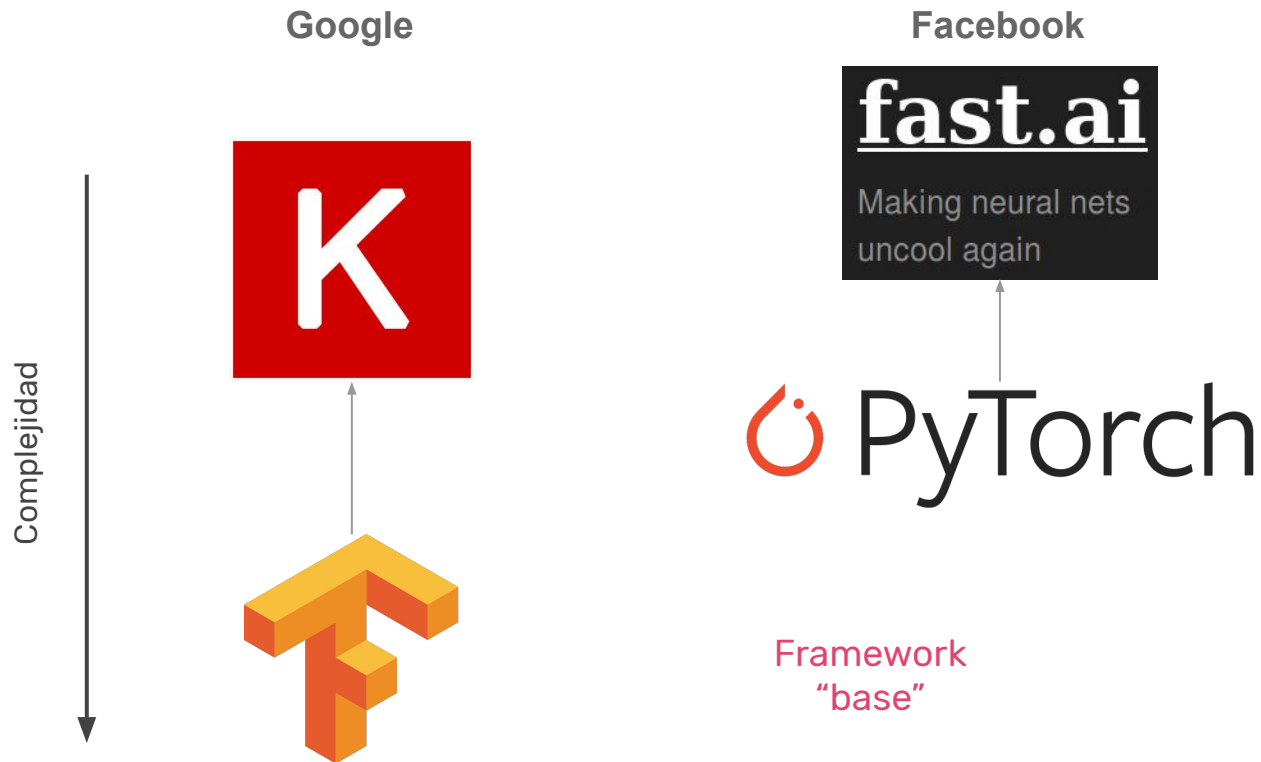
https://scikit-learn.org/stable/auto_examples/calibration/plot_calibration_curve.html#sphx-glr-auto-examples-calibration-plot-calibration-curve-py

https://scikit-learn.org/stable/auto_examples/calibration/plot_compare_calibration.html#sphx-glr-auto-examples-calibration-plot-compare-calibration-py





Redes Neuronales



Redes Neuronales • Cursos online

- [Deeplearning.ai](#): Famosos cursos de Andrew NG
 - Cinco cursos en total
 - Pagos (a través de Coursera)
 - Disponibles en Youtube (sin los ejercicios pero se consiguen)
 - Un poco lentos por momentos
 - El Curso 3 es muy general y muy recomendable
- [Fast.ai](#): además de ser una librería, ofrece cursos
 - Más actuales
 - Gratuitos
 - Más accesibles (con pinzas)
 - Mención interesante a servicios en la nube



Bases de datos

- [Pandas como interfaz SQL](#)

SQL:

```
SELECT fecha, precipitacion FROM datos WHERE fecha BETWEEN '2014-06-01'  
AND '2014-07-01'
```





Libros

- [*The Data Science Design Manual*](#), Steven S. Skiena, 2017: libro actual, accesible, disponible online, cercano a los contenidos de la carrera que les permitirá afianzar y profundizar conocimientos.
- [*Machine Learning*](#), Tom M. Mitchell, 1997: Un clásico. Accesible pero formal. En algunos temas poco actual. Ni mención a Python.
- [*An Introduction to Statistical Learning \(with Applications in R\)*](#), James, Witten, Hastie y Tibshirani.
 - Accesible
 - Hermano menor de *The Elements of Statistical Learning*
 - Ejemplos en R. Pero si quieren los ejemplos en Python, ya saben dónde buscar.
- Y todos los recursos que fuimos mencionando durante las clases



Otros recursos

- [Towards Data Science](#). Muy buenas notas, en general cortas. Tanto de nivel técnico como divulgativo.
- [Kaggle](#). Muchísimos recursos. A veces un poco caótico, pero lo recontra vale. **Recomendación:** elegir una competencia que les interese y participar.
- [Two Minutes Papers](#). Un canal de Youtube donde cuentan trabajos de Inteligencia Artificial (en general redes neuronales). Es super accesible la manera en que cuenta los trabajos, y puede estar bueno para tener una idea del “state of the art”.
- [Medium Machine Learning](#). Pariente de Towards Data Science. Un poco más general, ya que pueden encontrar no solo tutoriales sino también noticias.
- **Meetups, reuniones, eventos, etc.**

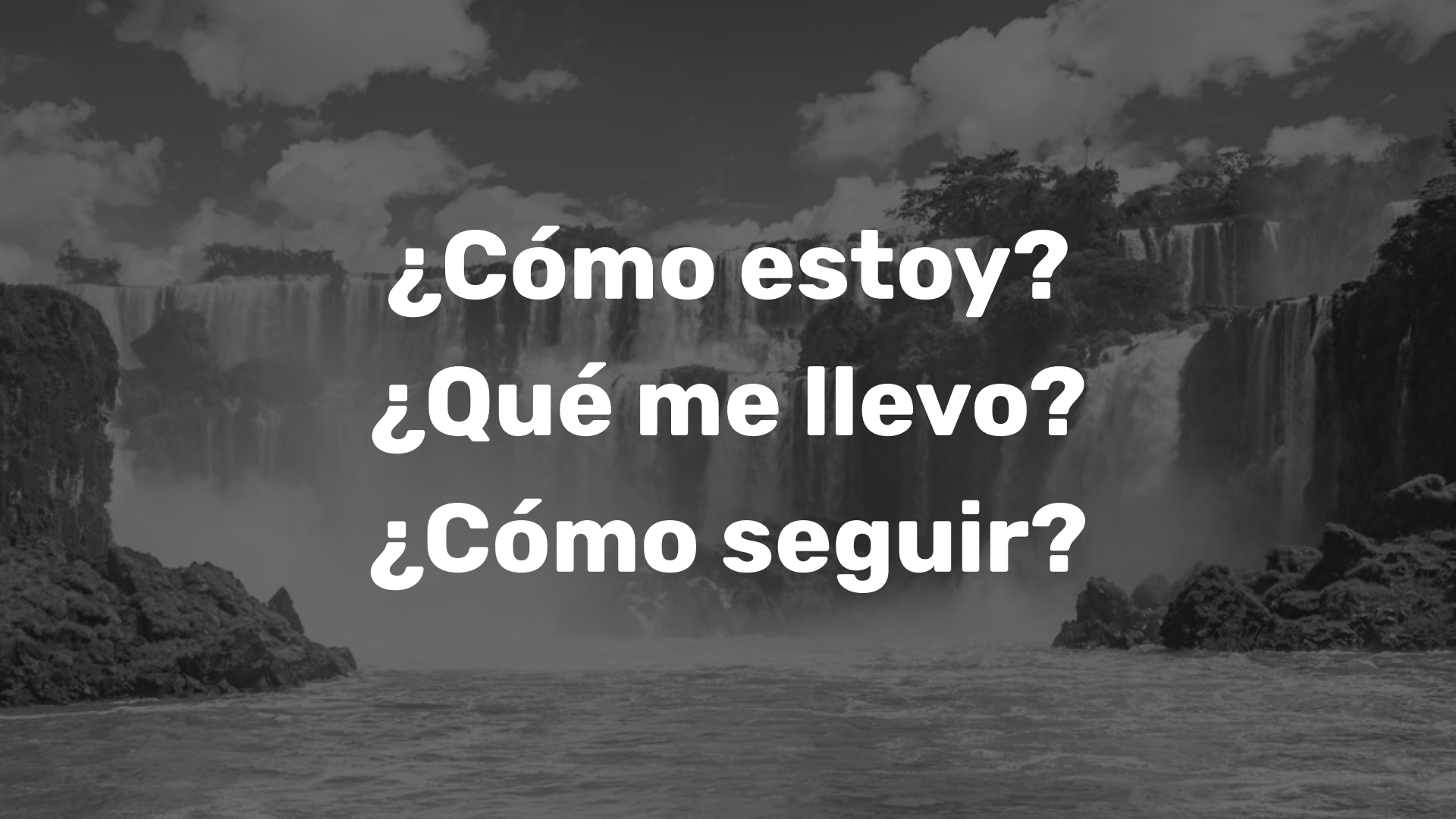


Recomendaciones Personales

- [Charla 1](#) y [Charla 2](#) sobre educación del futuro. Especialmente para Carla que viene pensando en estos temas. Ese proyecto tiene canal de [Youtube](#) y Spotify con MUY buenos contenidos.
- Sobre reflexiones de Sociología y Tecnología (Lu) recomiendo muchísimo a **Yuval Noah Harari**. En el [libro Sapiens](#) desarrolla una [breve historia de la humanidad](#) (Excelente). También tiene otros 2 libros sobre [desafíos de la sociedad del futuro](#). Recomiendo mucho esta [charla \(paseo\) que le da a Zuckerberg](#).
- [Todo es Fake](#) es un podcast de Anfibia Podcasts de Revista Anfibia que analiza el impacto de la tecnología digital en la vida moderna.
- [Vox es un canal de youtube](#) con MUY buen contenido. Les recomiendo mucho [esta serie sobre el futuro del trabajo](#) y los documentales de **Explained** que tienen en Netflix (Del tema tecnológico tienen el de eSports y el de Coding).

Una última reflexión como grupo





¿Cómo estoy?
¿Qué me llevo?
¿Cómo seguir?

~~—Para la próxima—~~





¡FELICIDADES!

ACÀMICA