

La Mujer en Ciencia de Datos

Sindia Sherly Tarazona Tocto

sindia.tarazona.t@uni.pe



DSW

EDUCAR, INSPIRAR Y EMPODERAR



27 años
Huanuqueña
Cristiana
Me encanta jugar vóleibol



Formación Profesional

Bachiller en Ingeniería Estadística
Universidad Nacional de Ingeniería

Diplomatura en Business Intelligence & Business Analytics
Universidad Nacional de Ingeniería – FIEECS



Experiencia Profesional

Especialista de Modelos Analíticos
Banco Pichincha Perú

Otras empresas en las que tuve la oportunidad de trabajar:
Yanbal Internacional, Experian Perú, BCP, Colegios Peruanos Innova Schools



Voluntariado

Co-organizadora en la Comunidad R-ladies Lima

Educar, Inspirar y Empoderar

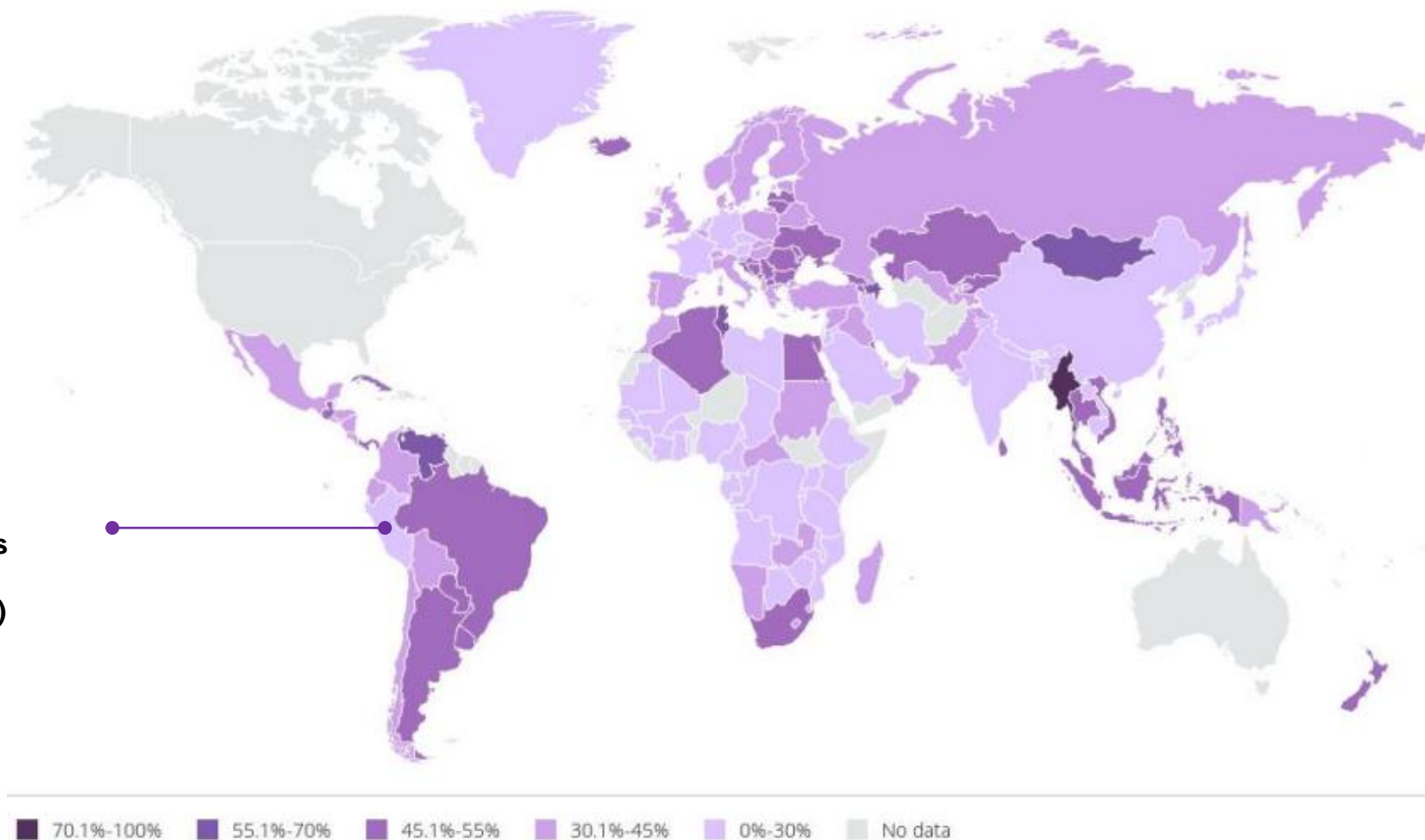
Según un informe del Instituto de Estadística de la UNESCO, menos del 30% de los investigadores del mundo son mujeres.

Numerosos estudios han encontrado que las mujeres en los campos STEM publican menos, se les paga menos por su investigación y no progresan tanto como los hombres en sus carreras.



Mapa de género en ciencias - 2017

30% de investigadores peruanos son mujeres (2017)



Fuente: UNESCO Institute for Statistics, June 2019

<http://uis.unesco.org/sites/default/files/documents/fs55-women-in-science-2019-en.pdf>

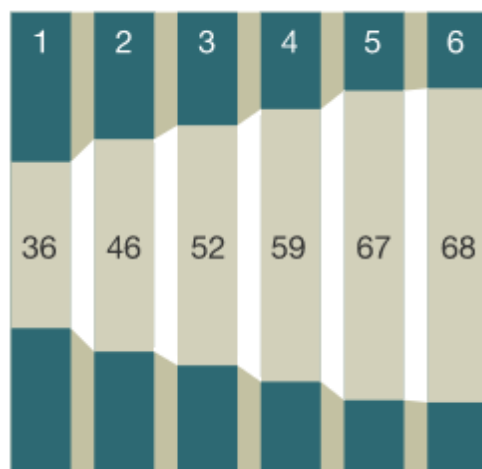
Educar, Inspirar y Empoderar

Estudio: Mujeres en el lugar de trabajo 2018

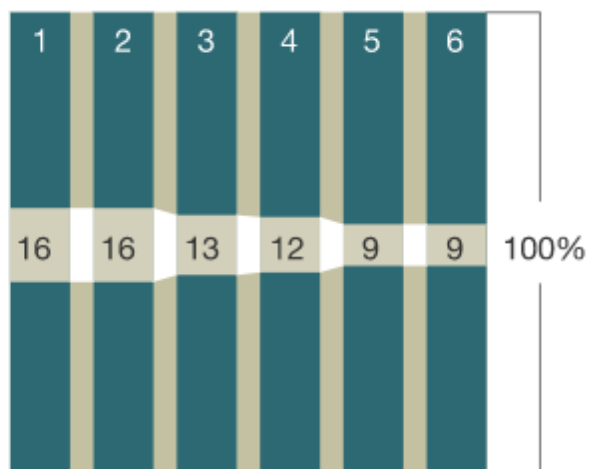
Las mujeres están subrepresentadas en cargos de alta gerencia (líderes)

1 = Entry level 3 = Senior manager/director 5 = Senior vice president
2 = Manager 4 = Vice president 6 = C-suite

White men

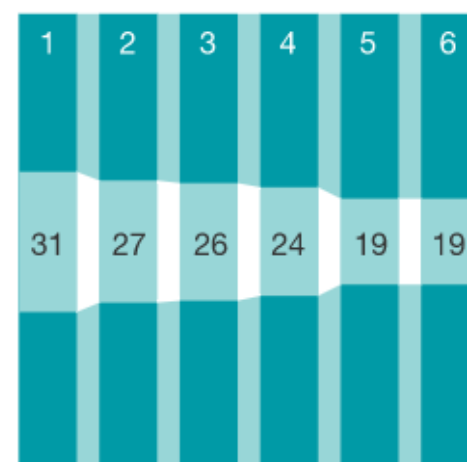


Men of color

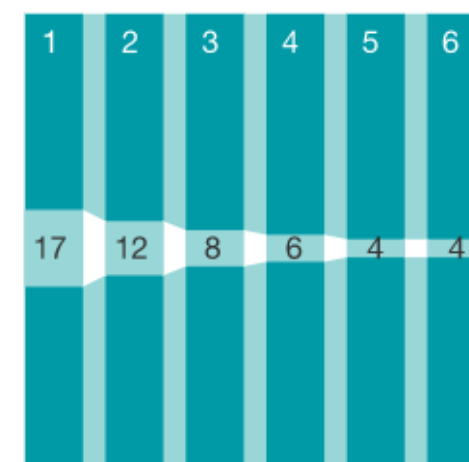


1 = Entry level 3 = Senior manager/director 5 = Senior vice president
2 = Manager 4 = Vice president 6 = C-suite

White women



Women of color



Entry level —————> C-suite

Fuente: Mckinsey <https://www.mckinsey.com/featured-insights/gender-equality/women-in-the-workplace-2018>

Educar, Inspirar y Empoderar

¿Y en el campo de Ciencia de Datos, cómo es la brecha de género?

**El campo de Data Scientist está conformado predominantemente por varones.
Solo 30% son mujeres.**

Fuente: Estudio realizado evaluando perfiles de DS en LinkedIn -2019

Elaboración:

<https://365datascience.com/data-scientist-profile-2019/>

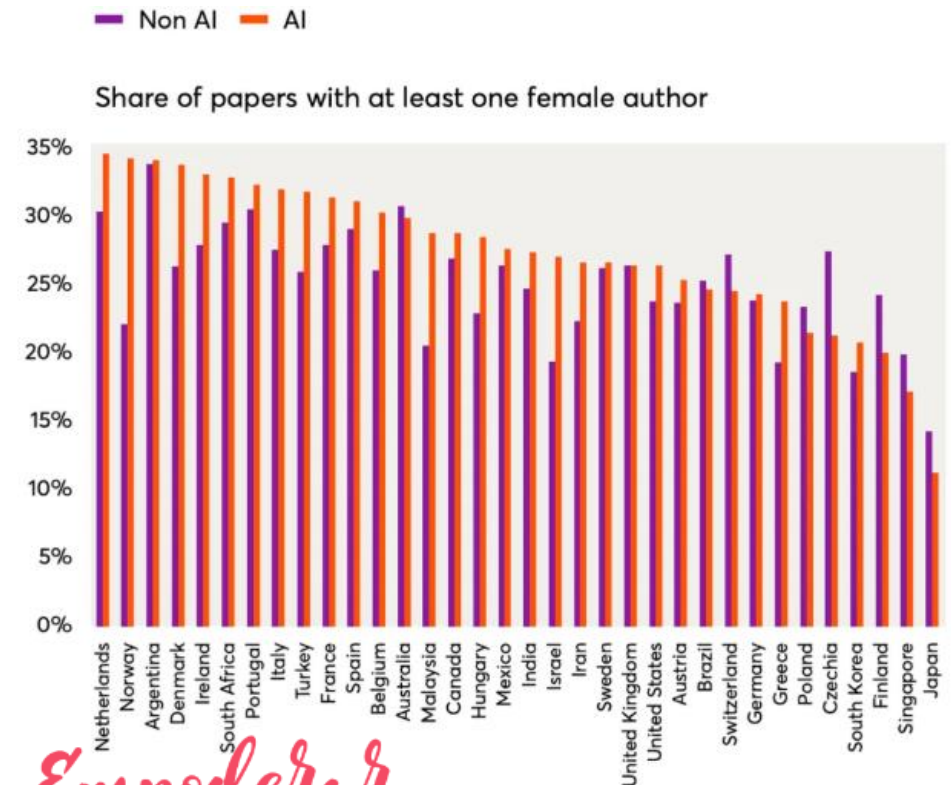
365 DataScience

Solo el 13.83% de los autores de artículos de IA son mujeres

Fuente: Nesta - a UK-based innovation foundation

<https://syncedreview.com/2020/03/13/exploring-gender-imbalance-in-ai-numbers-trends-and-discussions/>

Países como Argentina, México y Brasil encabezan la lista en la región.



Educar, Inspirar y Empoderar

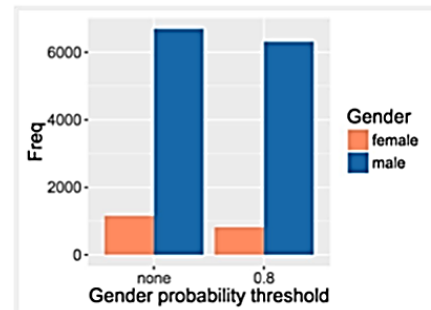
2016 – Creación de paquetes



Gender Statistics

Only ~ 15% of R package authors are women

(Estimates produced with CRAN package genderizeR, as on March 2016)

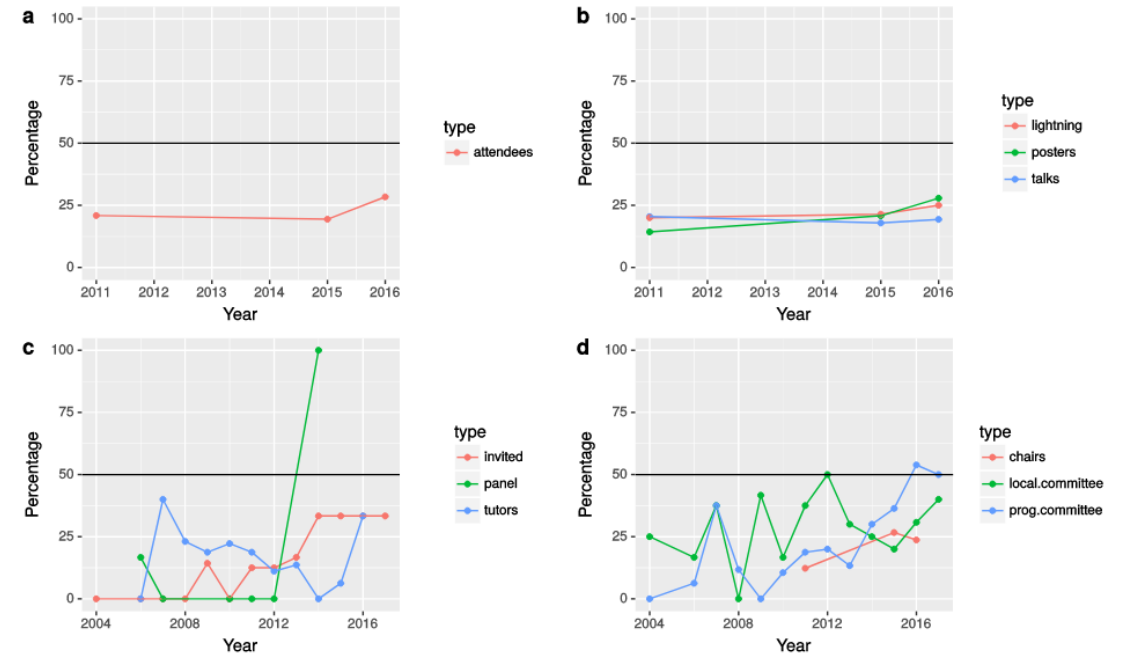


En el 2010, solo 9% de autores de paquetes eran mujeres.

En el 2016, solo 15% de autores de paquetes eran mujeres.

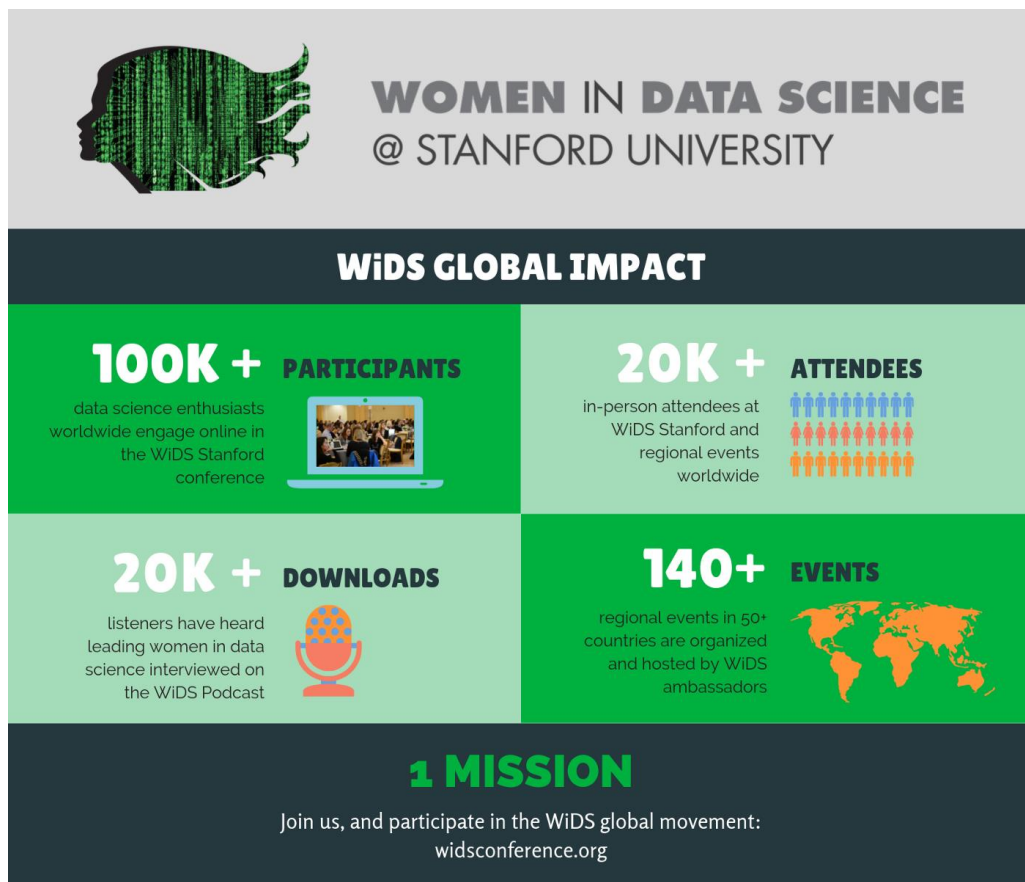
Fuente: R forwards <http://forwards.github.io/data/>

2016 – Asistencia al UseR!



- Antes del 2016, las asistentes mujeres eran menos del 25%.
- Menos del 25% eran oradoras mujeres o presentadoras de posters.
- Cada vez se aprecia mayor participación femenina en el comité organizador.

Iniciativas y/o comunidades globales que buscan reducir la brecha de género en Ciencia de Datos



Iniciativas locales que buscan reducir la brecha de género en Ciencia de Datos



Educación, Inspiración y Empoderamiento

1

Solicitud de nuevo modelo

Clave: Hacernos las preguntas correctas

Diagnóstico

- ¿Qué necesidad busca cubrir?
- ¿Quién es tu sponsor?
- ¿Tiene antecedentes?
- ¿Cuál es la población?
- ¿De qué datos dispongo? ¿De qué periodos son? ¿Cuál es el volumen?
- ¿Dónde se almacena los datos que necesito?
- ¿Cuál será el diseño (historia, ventana, target, periodos de backtesting)?
- ¿Qué plataformas dispongo para realizar el proyecto?
- ¿Trabajaré con toda la población o sacaré muestras?
- ¿Es necesario una segmentación?



Finaliza: Elaboración del Gantt del proyecto



2

Armar el set de variables iniciales

- ¿Existe alguna tabla maestra de variables?
- ¿Qué variables sugiere el negocio?
- ¿Qué transformaciones debo realizar (recencias, frecuencias, max, min, ratios, variaciones, etc.)?



Mi experiencia en proyectos de Data Science - ¿Cómo abordarlo?

3 Análisis Univariado

¿Qué tipo de variables dispongo?
¿Cuáles son sus descriptivos(cantidad, cantidad de missings, min, max, mediana, var, p1, p2, ...etc.)

4 Limpieza de datos

¿Hasta qué porcentaje de valores missings se tolera?
¿Tengo datos outliers? ¿Cómo los trataré?
¿Qué variables debo excluirlas?

5 Modelado

¿Cuál será la proporción de partición 60/40, 70/30, 80/20?
¿De qué forma haré la selección de variables?
¿Qué modelos testaré?
¿Qué indicadores usaré para medir el performance del modelo? ¿Cuánto es el bench?



Mi experiencia en proyectos de Data Science - ¿Cómo abordarlo?

6 Validación con negocio

¿Les hace sentido las variables?
¿Tienen alguna sugerencia?



7 Piloto

¿Cuál será el Diseño?
¿Cuánto tiempo durará?



8 Implementación

¿Plataforma de implementación?
¿Frecuencia de actualización?



9 Seguimiento

¿Cuál será la frecuencia del seguimiento?



Educar, Inspirar y Empoderar

¡Gracias!

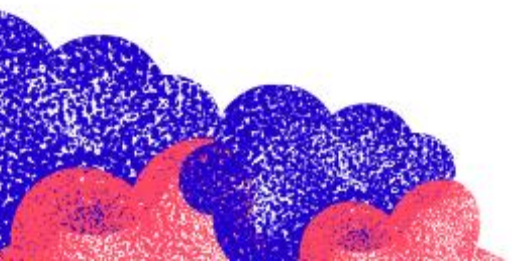
**El mundo necesita de ti...
Más mujeres creando valor a partir de los datos**

Educar, Inspirar y Empoderar

ESCRIBE EN EL CHAT , TUS



¿Preguntas?





@datasciencewomen



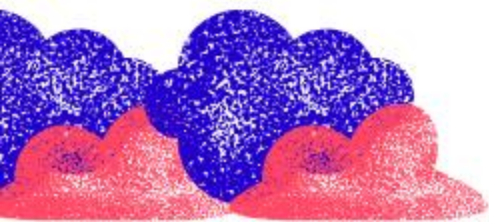
dswomen@datascience.pe



@datasciencewomen



dswomen@datascience.pe



EDUCAR, INSPIRAR Y EMPODERAR

