

# Tipología y ciclo de vida de los datos

## Práctica 1

Universitat Oberta de Catalunya

Máster Ciencia de Datos

Mar Bonora Ortega

### 1. Contexto

La información recopilada consiste en un listado de las mujeres a las cuales se les ha otorgado el Premio Nobel de cualquier disciplina a lo largo de la historia. Por todos es sabido que existe una brecha de género en la sociedad en la que vivimos, por eso es interesante utilizar datos como estos para estudiar el avance de las mujeres en cuestiones de igualdad con los hombres. Además, es interesante utilizar los datos de este premio y no otros ya que es un premio muy antiguo con muchos años de historia.

Se ha elegido Wikipedia como sitio web fuente de la información ya que es una enciclopedia libre como ellos mismos se declaran, donde cualquier usuario puede aportar información sobre cualquier tema. Aunque hay que poner especial atención en la veracidad de los datos aportados por esta misma razón, en el caso de la información que nos concierne está debidamente citada de las fuentes oficiales de las que proviene. Por último, al ser una enciclopedia libre, no hay problemas para utilizar información de la página para proyectos propios.

### 2. Título

Mujeres ganadoras del Premio Nobel a lo largo de la historia

### 3. Descripción del dataset

Como se ha comentado en apartados anteriores, el dataset generado contiene la información básica de las ganadoras del Premio Nobel a lo largo de la historia.

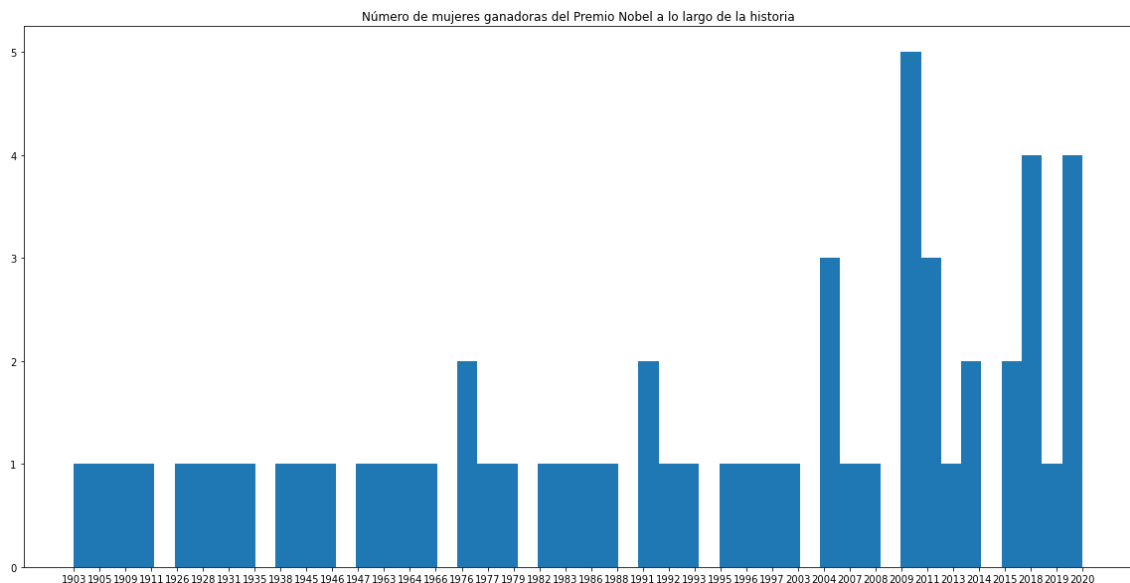
En la página que hemos utilizado como fuente de datos, es decir, Wikipedia, se aporta una tabla con las galardonadas, el año en que fueron premiadas, el país de origen, la categoría, la motivación y una imagen de la premiada (la cual obviaremos porque no nos aporta información para el estudio que nos interesa llevar a cabo).

### 4. Representación gráfica

A continuación se muestra la cabeza de la tabla, es decir, las primeras filas de la tabla, para que se pueda tener una idea general del dataset generado.

	Año	Premiada	País	Categoría	Motivación
0	1903	Curie, MarieMarie Curie(compartido con Pierre ...	Francia Polonia	Física	«En reconocimiento a los extraordinarios servi...
1	1905	Suttner, Bertha vonBertha von Suttner	Austria-Hungría	Paz	Presidenta honoraria de la Oficina Internacion...
2	1909	Lagerlof, SelmaSelma Lagerlöf	Suecia Suecia	Literatura	«En reconocimiento al elevado idealismo, la ví...
3	1911	Curie, MarieMarie Curie	Francia Polonia	Química	Por su descubrimiento del radio y el polonio.[...
4	1926	Deledda, GraziaGrazia Deledda	Italia	Literatura	«Por sus escritos inspiradamente idealistas, q...

También se ha generado un histograma para que se pueda ver el número de mujeres que han sido premiadas cada año a lo largo de la historia, y de esta forma ir haciéndonos una primera idea de la estructura del conjunto de datos que hemos obtenido.



## 5. Contenido

Los campos que lo componen son los siguientes:

- Año: año en el que el Premio Nobel en cuestión fue entregado
- Ganadora: nombre de la ganadora del Premio Nobel
- País: país de origen de la ganadora del Premio Nobel
- Categoría: categoría de premio que ha sido otorgado, por ejemplo Física o Matemáticas
- Motivación: frase que describe los motivos por los cuales se le otorga el Premio Nobel de la categoría correspondiente a la ganadora en cuestión

Los datos recogidos datan desde el año 1903 hasta el año 2020, por lo que nos satisface el nivel de actualidad y veracidad de los mismos. Por último, los datos son datos públicos que han sido recogidos a partir los datos oficiales de Fundación Nobel.

## 6. Agradecimientos

El conjunto de datos ha sido obtenido de Wikipedia, que es una enciclopedia libre en línea que cualquiera puede editar, por ello es un sitio para la divulgación de la información y los datos que encontramos no necesitan un trato especial en el sentido de principios éticos y legales.

El artículo que incluye el listado de mujeres ha sido redactado a base de contribuciones de varios usuarios, y el propietario de Wikipedia es Jimmy Wales. Además, como se ha comentado con anterioridad estos datos provienen de la fuente oficial de datos de la Fundación Nobel, la cual tampoco necesita un tratamiento especial de sus datos debido a su carácter público.

En el propio artículo se comentan los datos a grandes rasgos y ya se responden a preguntas que nos interesaría responder como quién fue la primera mujer premiada o si

alguna mujer ganó el premio en más de una ocasión – Marie Curie en ambos casos. Pero al recoger este conjunto de datos, se pretende hacer un análisis más profundo, más allá de las curiosidades que podamos obtener del mismo, respecto al avance de las mujeres hacia la igualdad.

Por supuesto, en el artículo se cita como enlaces externos, además del sitio web oficial de la Fundación Nobel de donde se han extraído los datos, una fotogalería de las premiadas “Mujeres Nobel” y un artículo que trata sobre las pocas mujeres que fueron galardonadas en comparación con la cantidad que fueron nominadas, el cual se llama “Cómo la Academia rechazó a la Mujer”. Estos artículos hacen un análisis más profundo del tema que nos atañe y los tendremos en cuenta para el estudio de los datos recogidos.

## **7. Inspiración**

Vivimos en una sociedad patriarcal, en la cual vamos avanzando poco a poco hacia una igualdad de género todavía no real en todos los aspectos de la vida diaria, como pueden ser los aspectos sociales y económicos. El acceso de las mujeres a la educación ha estado restringido durante mucho tiempo en todo el mundo y sigue estándolo a día de hoy en ciertas partes del planeta. Es por eso que utilizar el Premio Nobel, premio que podría ser considerado el más importante a lo largo de la historia de la humanidad, puede darnos una imagen concreta de cómo estamos avanzando hacia la igualdad de género.

El hecho de que haya paridad en los premiados no es solo una cuestión de cuotas, es un resultado de igualdad de oportunidades en la actualidad y de igualdad en referentes y puestos de poder para las generaciones futuras, lo cual es un proceso iterativo ya que una cuestión conduce a la otra. De ahí radica la importancia de este tipo de estudios.

A partir de este conjunto de datos se pretende intentar comprender el avance de las mujeres en materias académicas. Para ello se busca responder a preguntas como las siguientes:

- ¿Ha aumentado el porcentaje de premiadas con el paso de los años? ¿De forma lineal, o ha habido años excepcionales tanto positiva como negativamente?
- ¿Aumentan las premiadas de la misma manera las mujeres en todos los países? ¿Hay una diferencia en los llamados “países ricos”, es decir, Europa y Norte América?
- ¿Aumentan las premiadas de la misma manera en todas las categorías? ¿Hay una diferencia en las categorías tradicionalmente masculinizadas, como por ejemplo la Física?
- ¿Quién fue la primera mujer en ganar un Premio Nobel? ¿En qué categoría? ¿Fue galardonada en solitario o fue compartido el premio?
- ¿Qué galardonadas (y cuántas) han obtenido el premio compartido?
- ¿Alguna mujer ha ganado un Premio Nobel en más de una ocasión?

Como ya se ha comentado en el apartado anterior, algunas de estas preguntas ya han sido respondidas en otros estudios o artículos, ya que el tema que se pretende estudiar es una cuestión muy importante en los tiempos que corren.

## **8. Licencia**

La licencia seleccionada para el dataset resultante sería “Released Under CC0: Public Domain License”. La razón es que, como se ha comentado, los datos que conforman el

dataset son datos públicos que han sido obtenidos de una enciclopedia libre, que a su vez los ha recogido del sitio oficial de la Fundación Nobel. Por lo tanto, el dataset formado tendrá una licencia completamente pública para el uso de cualquier persona que la pueda necesitar.

#### 9. Código

En el repositorio de Git se puede encontrar el código generado para llevar a cabo la realización de la práctica. El lenguaje utilizado ha sido Python.

#### 10. Dataset

El conjunto de datos obtenido con el desarrollo de la práctica tiene formato CSV y ha sido subido a Zenodo. El enlace del DOI es: <https://doi.org/10.5281/zenodo.5653299>

#### 11. Contribuciones

Debido a problemas personales, he considerado que sería mejor realizar la Práctica en solitario.

Contribuciones	Firma
Investigación previa	Mar Bonora Ortega
Redacción de las respuestas	Mar Bonora Ortega
Desarrollo de código	Mar Bonora Ortega