

# Nobel Prize Database

Data source: <https://www.kaggle.com/datasets/nobelfoundation/nobel-laureates>

## Table original look

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	Year	Category	Prize	Motivation	Prize Share	Laureate ID	Laureate Type	Full Name	Birth Date	Birth City	Birth Country	Sex	Organization Name	Organization City	Organization Country	Death Date	Death City	Death Country
1	1901	Chemistry	The Nobel Prize in recognition		1/1	160	Individual	Jacobus Henricus van't Hoff	1852-08-03	Rotterdam	Netherlands	Male	Berlin University	Berlin	Germany	1911-03-01	Berlin	Germany
2	1901	Literature	The Nobel Prize in special recognition		1/1	569	Individual	Sully Prudhomme	1839-03-01	Paris	France	Male				1907-09-07	Châtenay	France
3	1901	Medicine	The Nobel Prize for his work		1/1	293	Individual	Emil Adolf von Behring	1854-03-01	Hansdorf (Prussia)	Germany	Male	Marburg University	Marburg	Germany	1917-03-31	Marburg	Germany
4	1901	Peace	The Nobel Peace Prize		1/2	462	Individual	Jean Henry Dunant	1828-05-08	Geneva	Switzerland	Male				1910-10-30	Heiden	Switzerland
5	1901	Peace	The Nobel Peace Prize		1/2	463	Individual	Frédéric Passy	1822-05-02	Paris	France	Male				1912-06-12	Paris	France
6	1901	Physics	The Nobel Prize in recognition		1/1	1	Individual	Wilhelm Conrad Röntgen	1845-03-27	Lennep (Prussia)	Germany	Male	Munich University	Munich	Germany	1923-02-10	Munich	Germany
7	1902	Chemistry	The Nobel Prize in recognition		1/1	161	Individual	Hermann Emil Fischer	1852-10-10	Euskirchen	Prussia (Germany)	Male	Berlin University	Berlin	Germany	1919-07-15	Berlin	Germany
8	1902	Literature	The Nobel Prize for the greatest		1/1	571	Individual	Christian Matthias Thoma	1817-11-03	Garding	Schleswig (Denmark)	Male				1903-11-01	Charlottenburg	Germany
9	1902	Medicine	The Nobel Prize for his work		1/1	294	Individual	Ronald Ross	1857-05-18	Almora	India	Male	Liverpool University	Liverpool	United Kingdom	1932-09-16	Putney	United Kingdom
10	1902	Peace	The Nobel Peace Prize		1/2	464	Individual	Élie Ducommun	1833-02-01	Geneva	Switzerland	Male				1906-12-07	Bern	Switzerland
11	1902	Peace	The Nobel Peace Prize		1/2	465	Individual	Charles Albert Goblet	1843-05-02	Tramelan	Switzerland	Male				1914-03-16	Bern	Switzerland
12	1902	Physics	The Nobel Prize in recognition		1/2	2	Individual	Hendrik Antoon Lorentz	1853-07-18	Arnhem	Netherlands	Male	Leiden University	Leiden	Netherlands	1928-02-04		Netherlands
13	1903	Physics	The Nobel Prize in recognition		1/2	3	Individual	Pieter Zeeman	1865-05-20	Zonnemaai	Netherlands	Male	Amsterdam	Amsterdam	Netherlands	1943-10-09	Amsterdam	Netherlands
14	1903	Chemistry	The Nobel Prize in recognition		1/1	162	Individual	Svante August Arrhenius	1859-02-01	Vik	Sweden	Male	Stockholm	Stockholm	Sweden	1927-10-02	Stockholm	Sweden
15	1903	Literature	The Nobel Prize as a tribute to		1/1	572	Individual	Bjørnstjerne Bjørnson	1832-12-08	Kvikne	Norway	Male				1910-04-26	Paris	France
16	1903	Medicine	The Nobel Prize in recognition		1/1	295	Individual	Niels Ryberg	1860-12-10	Thorshavn	Faroe Islands	Male	Finsen	Copenhagen	Denmark	1904-09-24	Copenhagen	Denmark
17	1903	Peace	The Nobel Peace Prize		1/1	466	Individual	William Randolph Hearst	1828-03-17	Fareham	United Kingdom	Male				1908-07-22	London	United Kingdom
18	1903	Physics	The Nobel Prize in recognition		1/2	4	Individual	Antoine Henri Becquerel	1852-12-12	Paris	France	Male	École Polytechnique	Paris	France	1908-08-25		France
19	1903	Physics	The Nobel Prize in recognition		1/4	5	Individual	Pierre Curie	1859-05-15	Paris	France	Female	École municipale	Paris	France	1906-04-19	Paris	France

## Facts:

I have settled a prize\_ID for each unique row to identify the prize for each laureate.

Each laureate has a unique laureate\_ID.

One laureate may have several prizes (several entries).

Laureate\_type is binary, either individual, or organization.

Only when laureate\_type='Individual', the laureate then can have a university. Otherwise when laureate\_type='Organization', the laureate does not have: birth info, death info, university, sex.

One university may cultivate multiple laureates.

Some laureates may do not have death info.

Prize\_share\_num is not one unique number for each category in each year. Perhaps in one category during some year, personA has 1/4 of the prize, personB has 1/4, and personC has 1/2. So this number should only depends on prize\_ID or other equal combinations.

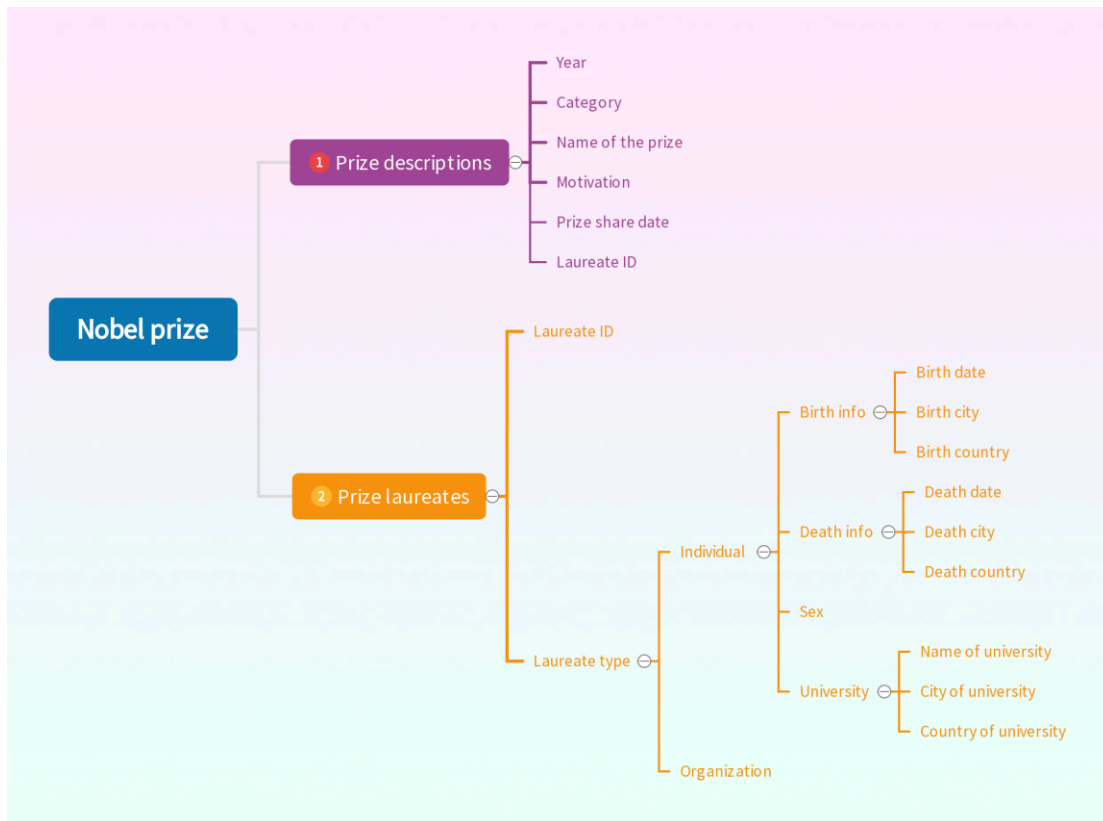
Motivation is a descriptive sentence which for each prize is not necessarily unique (several laureates contribute to one aim, thus they have the same motivation).

I cannot guarantee that if one laureate has several prizes, then his/her prize motivations would not be the same. Perhaps he/she continually make contributions to that field. That depends on how the Nobel Prize officials think.

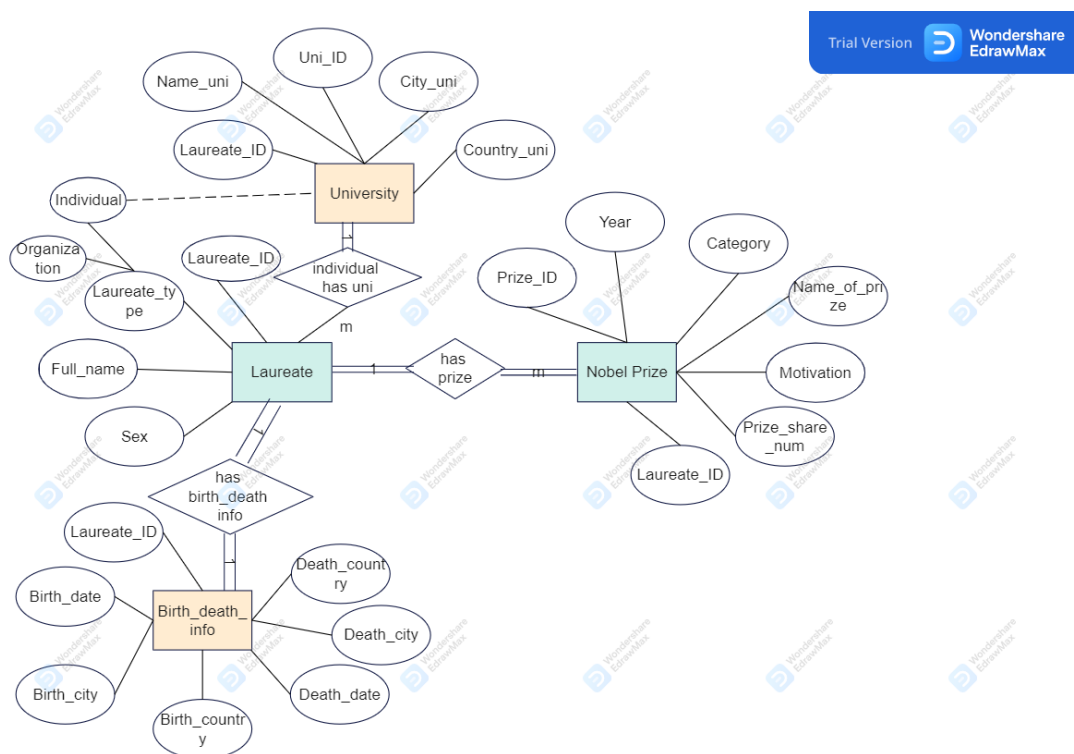
I cannot guarantee that one laureate may not win several prizes in one year.

The column 'name\_of\_prize' is not just a combination of the category and the year. Some fields look like this: 'The Sveriges Riksbank Prize in Economic Sciences 2010'. So this column cannot be deleted or be dismantled.

## Tree structure



## ER Diagram – Apr.18 ver

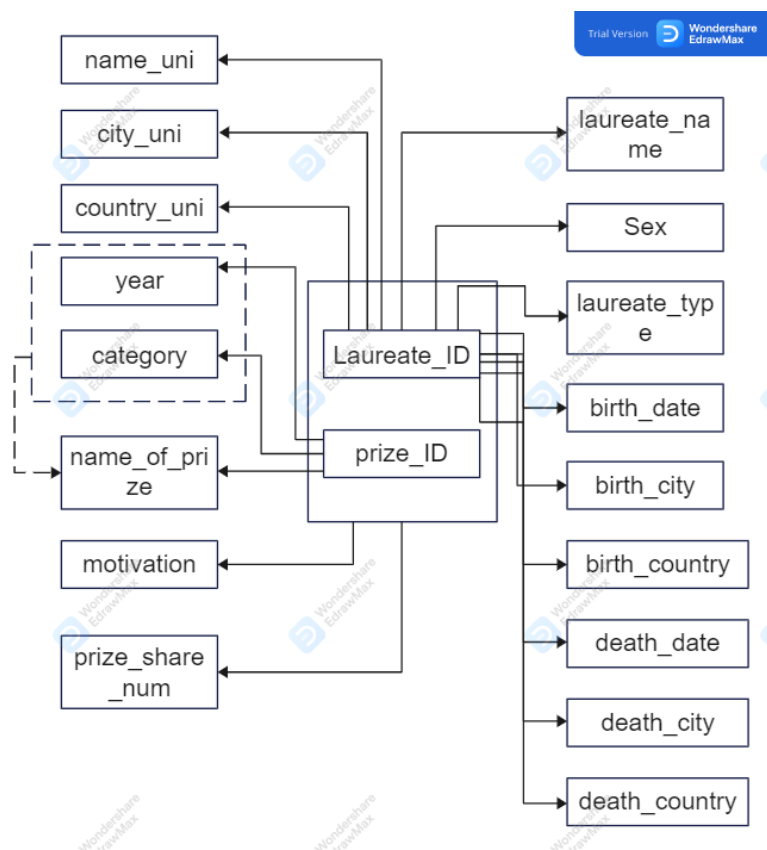


Candidate keys: prize\_ID, {laureate\_ID, name\_of\_prize}, {full\_name, name\_of\_prize}



# Normalisation

## 1NF



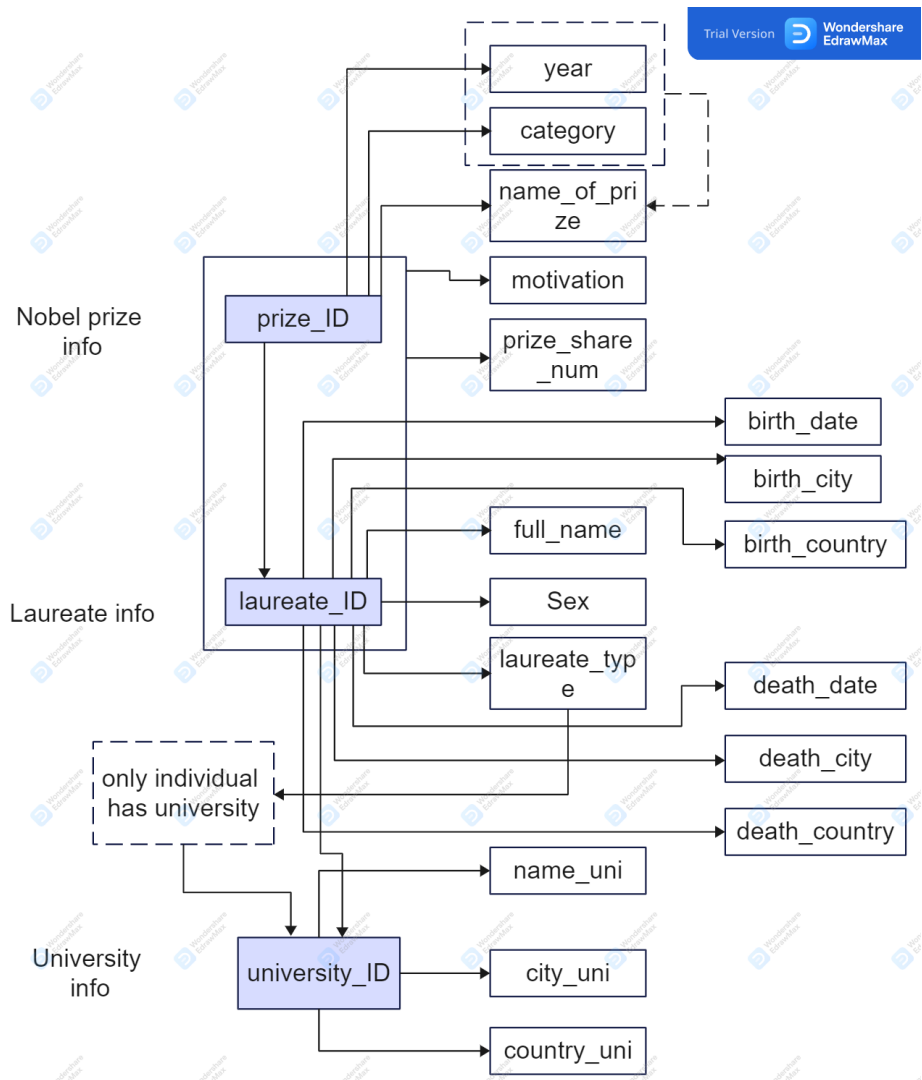
Prize_Id	Laureate_ID	Year	Category	Name_of_prize	Motivation	Prize_share_num	...(and all others)
1	253	1911	physics	Physics 1911	"..."	1	...
2	34	1911	chemistry	Chemistry 1911	"..."	2	...

Discussion of 1NF:

Candidate keys: laureate\_ID, prize\_ID. Should consider whether to add a university\_ID since one university may have several laureates, we add this ID to let the table more searchable.

In this table, all the attributes satisfy the atomic condition, which are not separable any more.

# 2NF



Prize_ID	Year	Category	Name_of_prize	Motivation	Prize_share_num	Laureate_ID
1						160
2						569
3						293
4						462

Laureate_ID	Full_name	Sex	Laureate_type	University_ID	Birth_date	...	Death_date	...	
1		Male	Individual	3					
2		N/A	Organisation	N/A	N/A		N/A		

University_ID	Name_uni	City_uni	Country_uni
1			
2			

---

#### Discussion of 2NF:

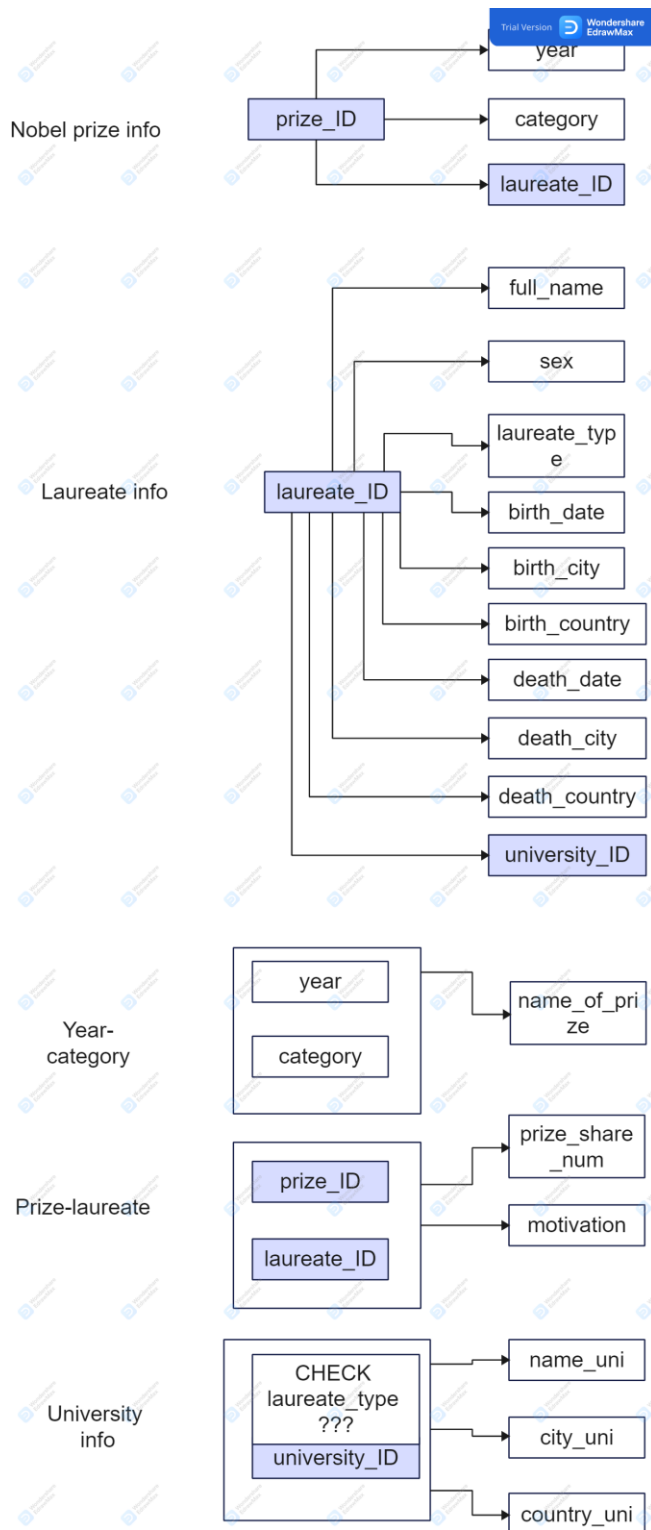
Previously, the primary keys determine different attributes. So we should separate those up.

Primary keys: {prize\_ID, laureate\_ID}, prize\_ID, laureate\_ID, university\_ID

Other non-key attributes are irreducibly dependent on these primary keys in each table now. In each table, we cannot see any non-key attribute which is only dependent on part of the primary keys in that table.

This step is the first step to dismantle our tables into more reasonable fully structured small ones.

### 3NF



Prize_ID	Year	Category	Laureate_ID
1			160
2			569
3			293

Laureate_ID	Full_name	Sex	Laureate_type	University_ID	Birth_date	...	Death_date	...	
1		Male	Individual	1					
2		N/A	Organisation	N/A	N/A		N/A		

Year	Category	Name_of_prize
1901	Physics	The Nobel Prize in Chemistry 1901
1901	Literature	The Nobel Prize in Literature 1901

Prize_ID	Laureate_ID	Prize_share_num	Motivation
1	160	1	"in recognition of the extraordinary services he has rendered by the discovery of the laws of chemical dynamics and osmotic pressure in solutions"
2	569	2	"in special recognition of his poetic composition, which gives evidence of lofty idealism, artistic perfection and a rare combination of the qualities of both heart and intellect"

I have meet a problem about how to choose primary key(s) for the University table. The condition here is: each university should only have laureate whose laureate\_type is 'Individual'. If I only list a column here called 'laureate\_type', then it is unmeaningful since all the items should be 'Individual'.

Laureate_type	University_ID	Name_uni	City_uni	Country_uni
Individual	1	Berlin University	Berlin	Germany
Individual	2	Marburg University	Marburg	Germany

#### Discussion of 3NF:

Previously, (1) laureate\_type determines whether there exists a university entry or not – only 'individual's, instead of 'organization's, have an entry of university information. We should seperate this into different tables. (2) name\_of\_prize is largely determined by year and category. Although through prize\_ID we can know the name\_of\_prize, the name is not generated that way and may cause check difficulties if the name only combines with prize\_ID. So I consider to split a new table for the name\_of\_prize.

After eliminating transitive relationships, the tables look more straightforward now.



---

## BCNF

Discussion of BCNF:

BCNF says that 'there should not exist dependent relationships among primary keys'.

Primary keys: prize\_ID, laureate\_ID, {year, category}, {prize\_ID, laureate\_ID}, university\_ID

There are dependent relationships.

Once we know the prize\_ID, we then know who the laureate is. {year, category} also depends on prize\_ID.

Laureate\_type depends on laureate\_ID.

University\_ID is rather an independent attribute, since laureate may change universities, but one university may have several laureates.

I don't know how to split my tables into BCNF standards, since I need to completely restructure the tables based on new primary keys. Please give me some suggestions!

## 4NF

Don't know yet. Will make it more complete later. Please give me some suggestions!