

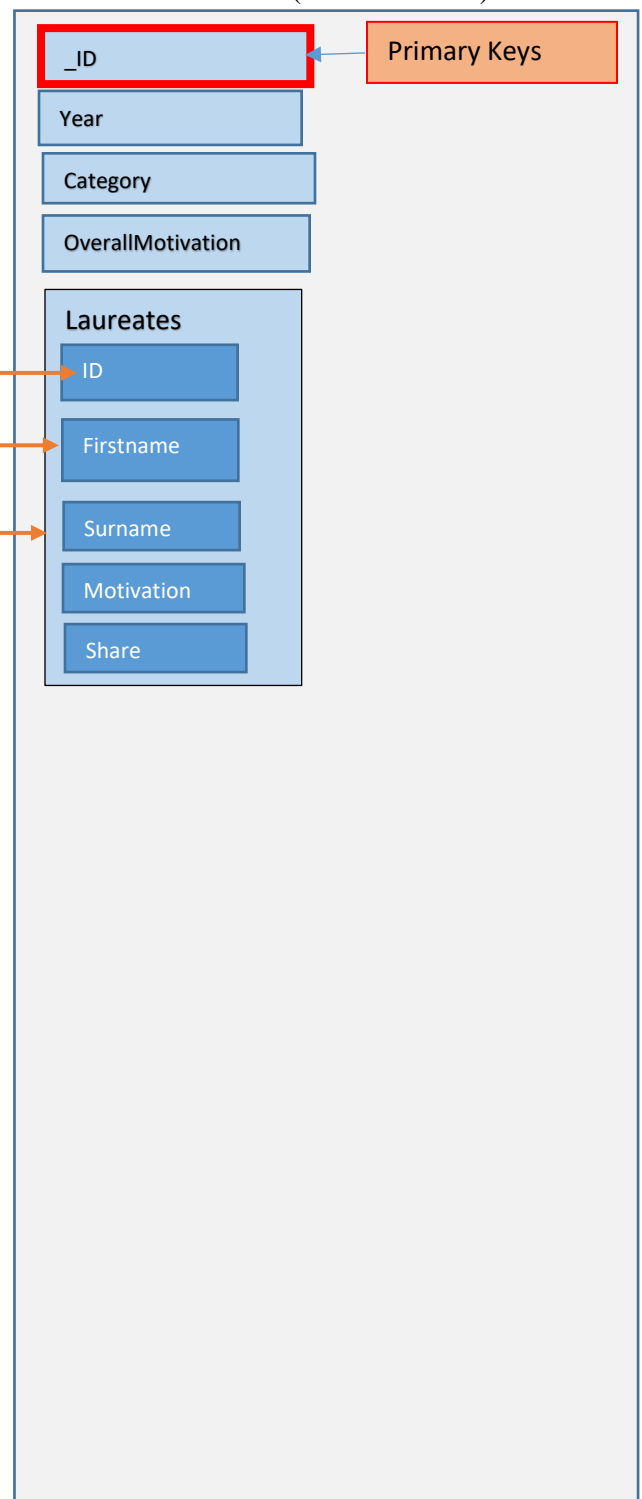
The data contains information about who has been awarded the Nobel Prize, when, in what prize category and the motivation, as well as basic information about the Nobel Laureates such as birth data and the affiliation at the time of the award.

I. Database model: “Nobel” database contains 2 collections: Laureates and Prizes

1. “**Laureates**” collection (934 documents)



2. “**Prizes**” collection (590 documents)



In summary, 2 collections have the relationship 1 to many via “ID” field which link from “Laureates” collection (Primary key) to “Prizes” collection.

1. **“Laureates” collection** includes 3 main groups of information:

1.1. *_ID & ID code for identifying each laureate*

1.2. *Personal information:*

- Name: “Firstname” and “lastname”
- Gender: 3 types of gender in this field, “male”, “female” for individuals & “org” for organisations laureates
- Born: to specify date of birth & place of birth, with 4 fields “Born”, “BornCity”, “BornCountry”, “BornCountryCode”
- Died information on laureates’ date and place of death

1.3. *Prizes information: data in array to provide details on:*

- Year,
- Category,
- Share,
- Motivation &
- Affiliation of prizes. In which, Affiliation contains an array with more detail on:
 - + Name
 - + City
 - + Country

Note: 1 laureate may have more than 1 prizes

2. **“Prizes” collection** includes 3 main groups of information:

2.1. *_ID for object identification*

2.2. *Overall information, including “year”, “category”, & “overallmotivation:*

2.3. *Laureates information:*

- ID, which can be matched with ID in “Laureates” collection
- “firstname” & “surname” of laureates as listed in “Laureates” collection
- Detail on “motivation
- “share” disclose numbers of laureates shares for each prize

II. Data Analysis Intuition:

Categorical Data:

```
_id: ObjectId('63d92672b9986af2fac4d33c')
year: "2018"
category: "physics"
overallMotivation: ""for groundbreaking inventions in the field of laser physics""
▼ laureates: Array
  ▼ 0: Object
    id: "960"
    firstname: "Arthur"
    surname: "Ashkin"
    motivation: ""for the optical tweezers and their application to biological systems""
    share: "2"
```

```
_id: ObjectId('63d9265cb9986af2fac4cf95')
```

```
id: "2"  
firstname: "Hendrik Antoon"  
surname: "Lorentz"  
born: "1853-07-18"  
died: "1928-02-04"  
bornCountry: "the Netherlands"  
bornCountryCode: "NL"  
bornCity: "Arnhem"  
diedCountry: "the Netherlands"  
diedCountryCode: "NL"  
gender: "male"  
prizes: Array  
  ▼ 0: Object  
    year: "1902"  
    category: "physics"  
    share: "2"
```

```
motivation: ""in recognition of the extraordinary service they rendered by their re..."  
  ▼ affiliations: Array  
    ▼ 0: Object  
      name: "Leiden University"  
      city: "Leiden"  
      country: "the Netherlands"
```

Quantitative Data: Only the number of shares of each prize is quantitative data in both collections

```
_id: ObjectId('63d92672b9986af2fac4d33c')  
year: "2018"  
category: "physics"  
overallMotivation: ""for groundbreaking inventions in the field of laser physics""  
  ▼ laureates: Array  
    ▼ 0: Object  
      id: "960"  
      firstname: "Arthur"  
      surname: "Ashkin"  
      motivation: ""for the optical tweezers and their application to biological systems""  
      share: "2"
```

```
_id: ObjectId('63d9265cb9986af2fac4cf95')  
id: "2"  
firstname: "Hendrik Antoon"  
surname: "Lorentz"  
born: "1853-07-18"  
died: "1928-02-04"  
bornCountry: "the Netherlands"  
bornCountryCode: "NL"  
bornCity: "Arnhem"  
diedCountry: "the Netherlands"  
diedCountryCode: "NL"  
gender: "male"  
  ▼ prizes: Array  
    ▼ 0: Object  
      year: "1902"  
      category: "physics"  
      share: "2"  
      motivation: ""in recognition of the extraordinary service they rendered by their re..."  
      ▼ affiliations: Array
```

III. Exploration of data:

1. “Prizes”:

Category: There are 6 types of Categories prizes and it is shared quite evenly among them, except economics area.

Row Labels	Count of category
chemistry	110
economics	50
Literature	110
Medicine	109
Peace	99
Physics	112
Grand Total	590

Year: The period of this data is from 1901 to 2018.

2. “Laureates”:

Born: The data have laureates born from 1817 to 1997, in which some document does not have date of birth.

BornCity: There are 614 different born cities in total.

BornCountry: There are 120 different born countries in total, including some countries have changed names during the reported period.

Died: The list includes laureates died during the period of 1903-2018.

DiedCity: There are 296 different died cities in total.

Gender: There are 3 types of genders, including “male” & “female” for individual laureates & “org” for organisations laureates

Prizes: there are 858 prizes in total affiliated for the list of laureates