Study about caesareans in Spain

María Dolores Pérez

Dataset

I have used the data from INE (Instituto Nacional de Estadistica).

Official data about births from 2012 until 2016.

Motivation

I would like to study if the number of caesareans depends on:

- The place of birth
- The age of the mother

I would like to build a decision tree with a subset of the data, in order to can predict if a birth will require a caesarean.

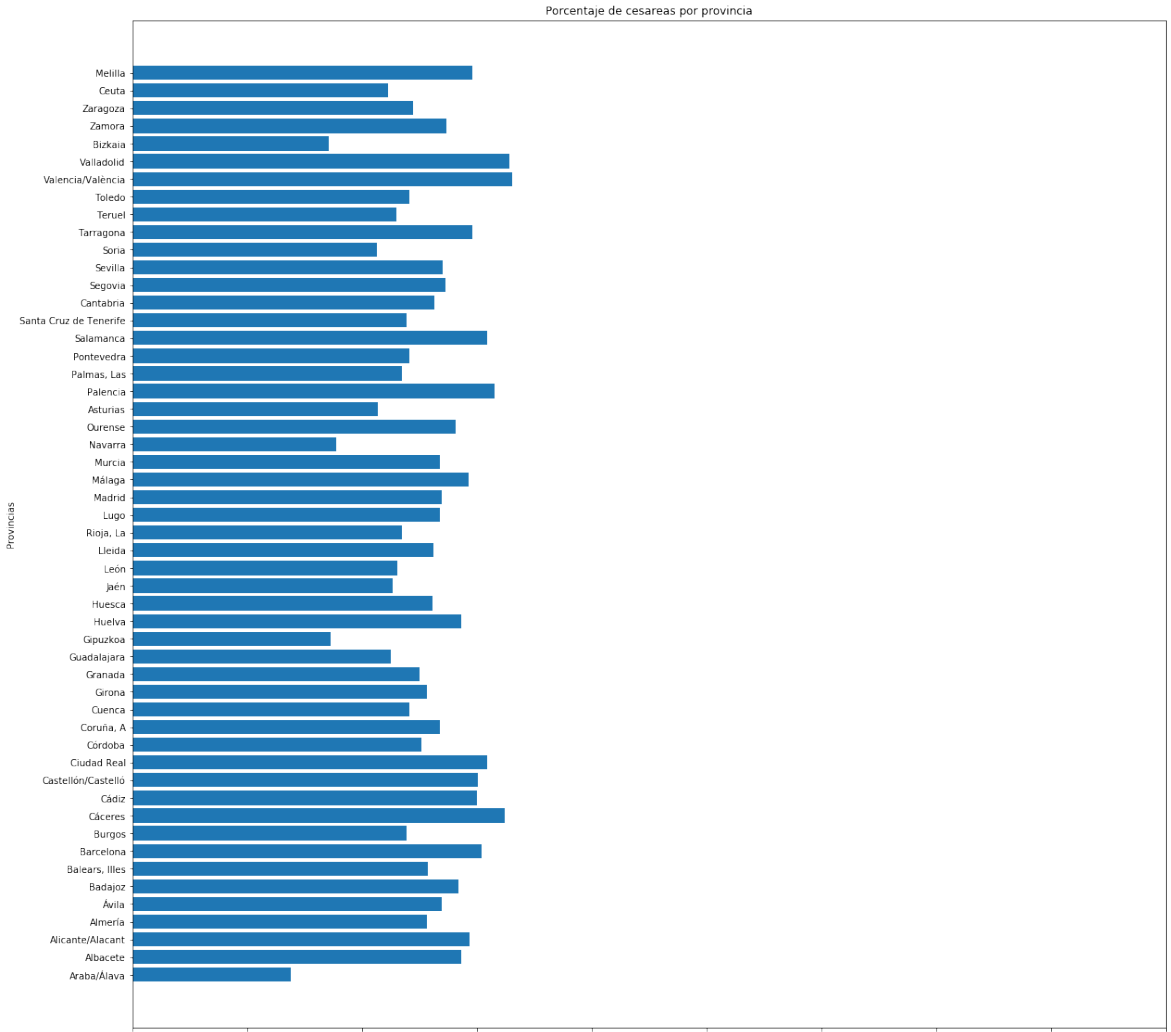
Research Questions

Research:

1. Has the age of the mother any relationship with the need of cesarean?

2. Is there any province where the number of cesarean is bigger than others?

3. Is there any patron in the circumstances of the cesarean? I want to build a decision tree with a subset of the data used



Findings. Distribution of caesareans by province

The chart shows the percentage of caesareans/births for every Spanish province.
It is a big chart, but you can see there are great differences between some of the provinces (in the next slide)

Findings. Distribution of caesareans by province

Let see with the figures, the difference between the three basque provinces and , for example, Valencia and Caceres

Percentage of the three basque provinces:

Alava: 13,83%

Guipuzkoa: 17,28%

Bizkaia: 17,14%

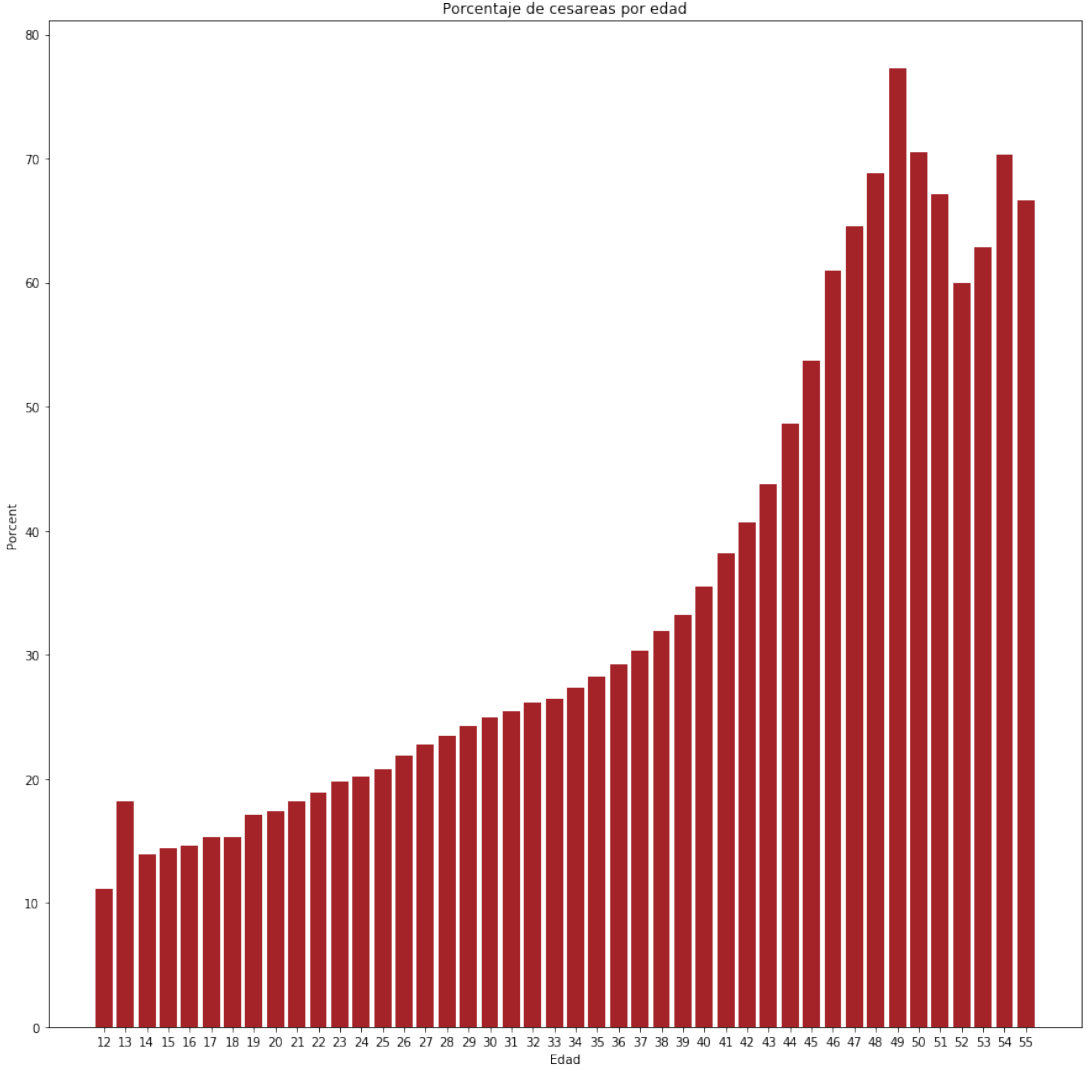
Percentage (avg) for Spain: 25,96

Highest percentages

Valencia: 33,05

Caceres: 32,45

We can see that actually the likelihood of caesarean is significantly bigger in Valencia and Caceres, than in the Basque provinces.



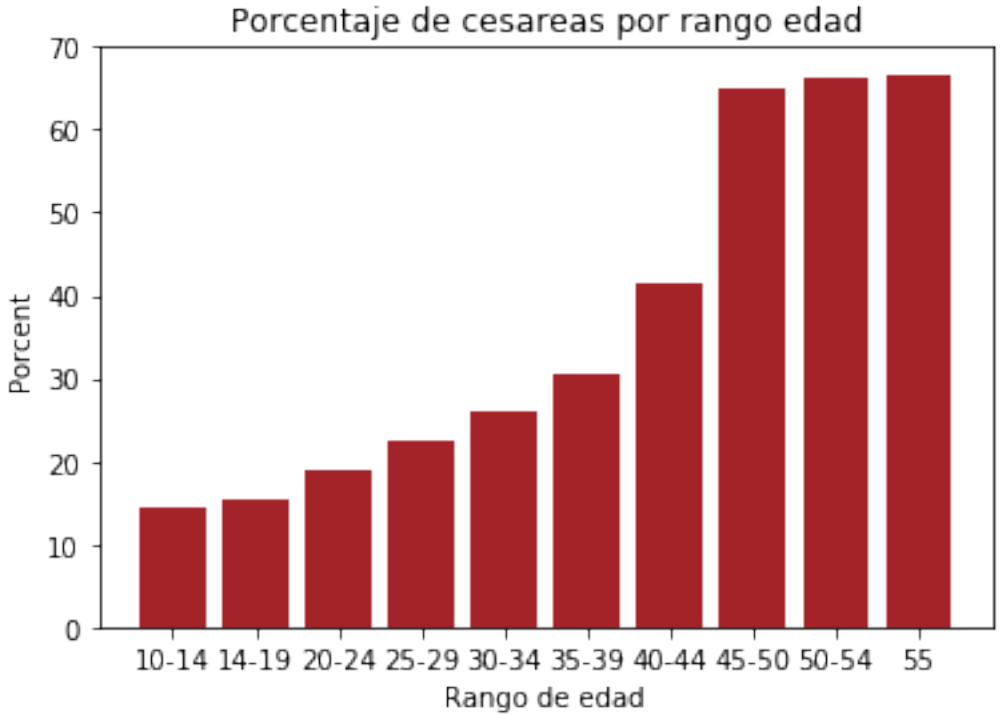
Findings. The age of the mother

Here is a bar chart with the relation between the percentage caesarean/ total births and the age of the mother

It´s easy to see the relation (with a few exceptions) between the age of the mother and the likelihood of caesarean

Findings. The age of the mother

To make this chart smaller, I have group the age of the mother in ranges of 5 years:



The relations keep being clear, but some information with the age is more than 50 is lost.

Decision Tree

I have build a decision tree to determine the need of caesarean. The features used in the training are:

- Province
- Multiplicity of the birth
- Normal or with complications
- Number of previous births (and number of previous birth alive)
- Mother age
- Sex of the new born
- If the new born lived more than 24 hours
- Weight of the new born
- Weeks of the gestation

Decision Tree

Based on this features the accuracy of the tree was: 80%
With this results we can see that is a kind of “patron” in the case of caesarean.

Limitations of my conclusions

I think that the need of caesarean is something very complex and here I can study only a few features related to caesarean. Although the data covers five years in a whole country, there are only circumstances of the birth. There are not information about the gestation, the health of the mother….

Future work

Some ideas for future work:

- To take in account more years
- To take in account more countries of Europe
- To add data from the gestation
- To study data about the health and habits of the mother
- To study the history of women in the same family

Acknowledgements

I am very grateful to the teachers of this course. They explain very clearly and transmit the love for data analytics.

I am also very grateful to my former colleges in the Statistics Centre for all I learnt there.

References

I haven´t used any other references but the material of the course.