DIVORCE PREDICTION

Turkish sample - 2019

Elisa Suárez CI: 5.233.124-9

CODERHOUSE

Comisión: 32695 - 2022

Motivation

What are the key indicators for divorce?

Which questions/factors are most significant when predicting divorce?

Audience

The Ohio State University: Decision Psychology Group

If divorce predictors of married couples can be estimated early, many divorces can be prevented.

We just need to pay attention to the data analysis and take a moment to reflect, and maybe, change our course of action towards our partner.

The mission of the OSU Decision Psychology Group is to promote understanding of the psychological underpinnings of judgments that people form and decisions that they make.

As such, it provides a valuable perspective on phenomena and mechanisms related to a wide variety of complex problems and across areas of psychology.

https://psychology.osu.edu/about/programs/decision

Metadata

- Dataset has 55 columns, not null data, and the data type is int64.
- Of the 100 participants, 49% were divorced and 51% were married couples.
- There were 49% males and 51% were females in the study group.

- The ages of the participants ranged from 20 to 63 (X̄= 36.04, SD = 9.34)
- 74 (43.5%) were married for love, and 96 (56.5%) were married in an arranged marriage
- While 127 (74.7%) of the participants had children, 43 (25.3%) had no children.
- Memory usage is 73.2 KB
- No null values.

Data Collection

Research data were collected using the face-to-face interview technique and via Google Drive.

Divorced participants answered the scale items by considering their marriages. And, of the married participants, only those with happy marriages, without any thought of divorce, were included in the study.

Attribute information

Questions (55) are ranked on a scale of 0-4 with 0 being the lowest and 4 being the highest.

The last category states if the couple has divorced.

Visualizations

There are different forms and types of marriages in Turkey, depending on the region, the culture of the parents and the level of education determine the type and form of marriage.

For example, in areas with a lower educational level, traditional marriages abound and this type depends on the well-known matchmaker, a woman who takes the bride to whomever she wants to marry.

If the educational level is higher, the role of the matchmaker disappears and the bride is the one who makes the decision. The parents do not interfere in it, they do not impose a specific girl on the boyfriend, as in traditional marriage.


How well I know my spouse?
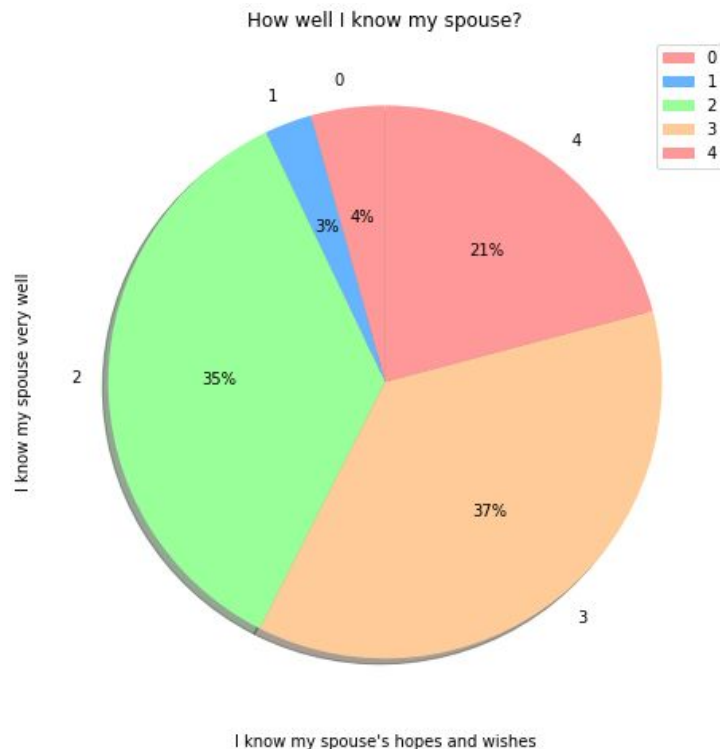
Visualizations

According to the 'Love and trust values' graph down below, the majority thinks that having the same values in trust and about what love should be is part of a happy marriage.
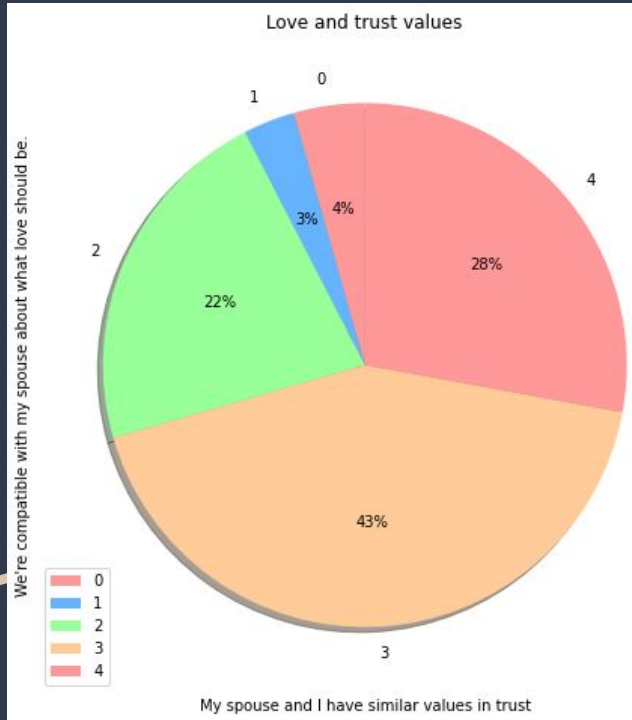
In this graph we can see similar values respect the 'Love and trust' graph about 'Love and personal freedom values' in the marriage.

Love and trust values

My spouse and I have similar values in trust

Love and personal freedom values

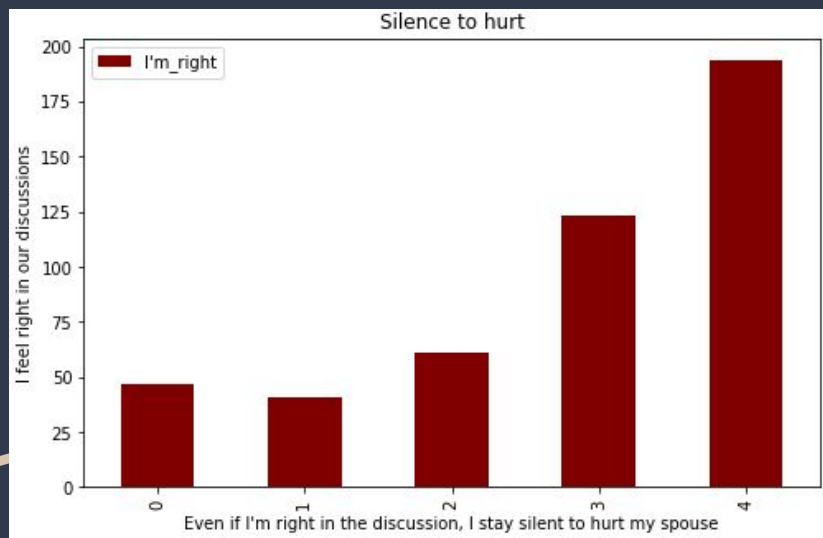My spouse and I have similar values in terms of personal freedom
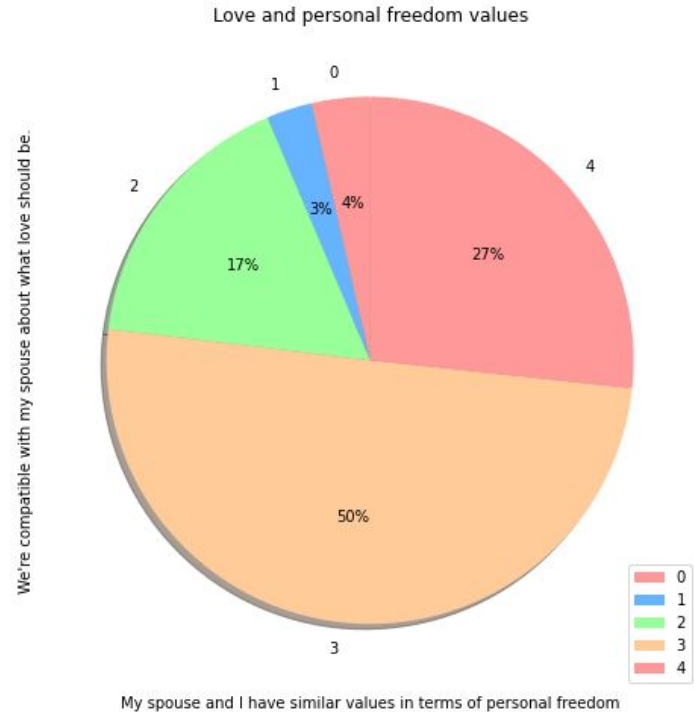
Visualizations

In the 'Sudden discussions' graph, we can see that when couples perceive they are starting a discussion, it is already happening, all of the sudden, and they feel they do not even know why.
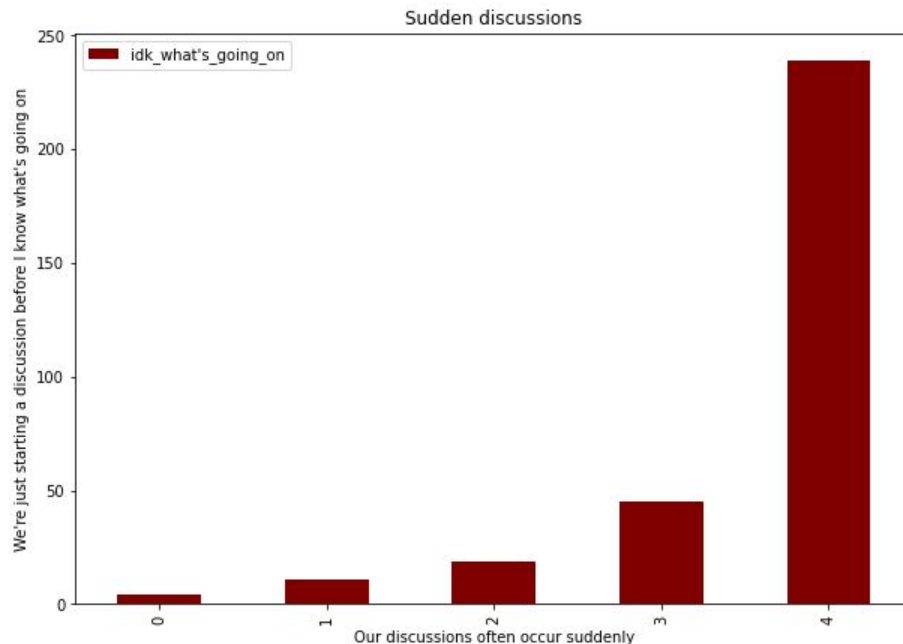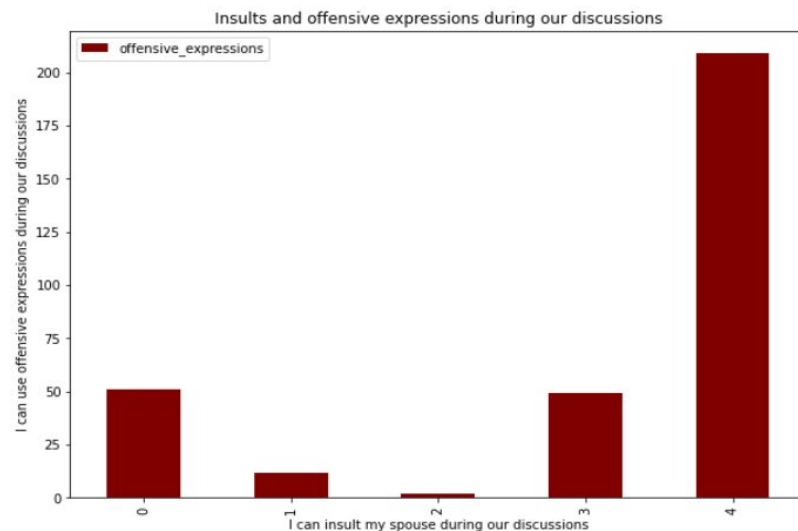
Visualizations

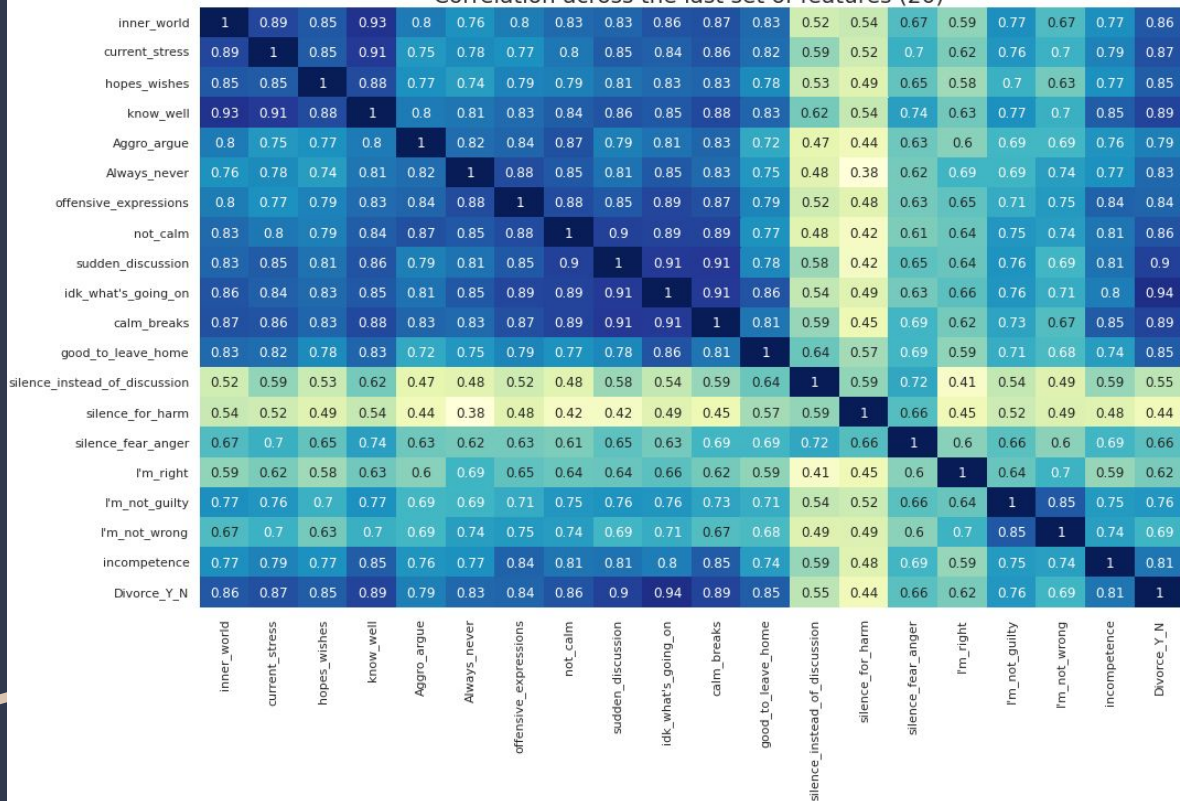Sometimes, being silent in discussions hurts more than talking...

And sometimes, arguing, insults and offensive expressions escape from our mouths...

Insights

Correlation across the last set of features (20)

❏ People who does not have quality time at home, do not take care of each other when the partner is sick, and do not know their spouse's favorite food. They do not share the same goals or freedom values. They do not trust each other, and do not share ideas about marriage, love or happiness.

❏ I truly do not understand why "We don't have time at home as partners" (No_home_time) and "We are like two strangers who share the same environment at home rather than family" (2_strangers) are so poorly correlated.

Insights

Correlation across the first set of features (17)
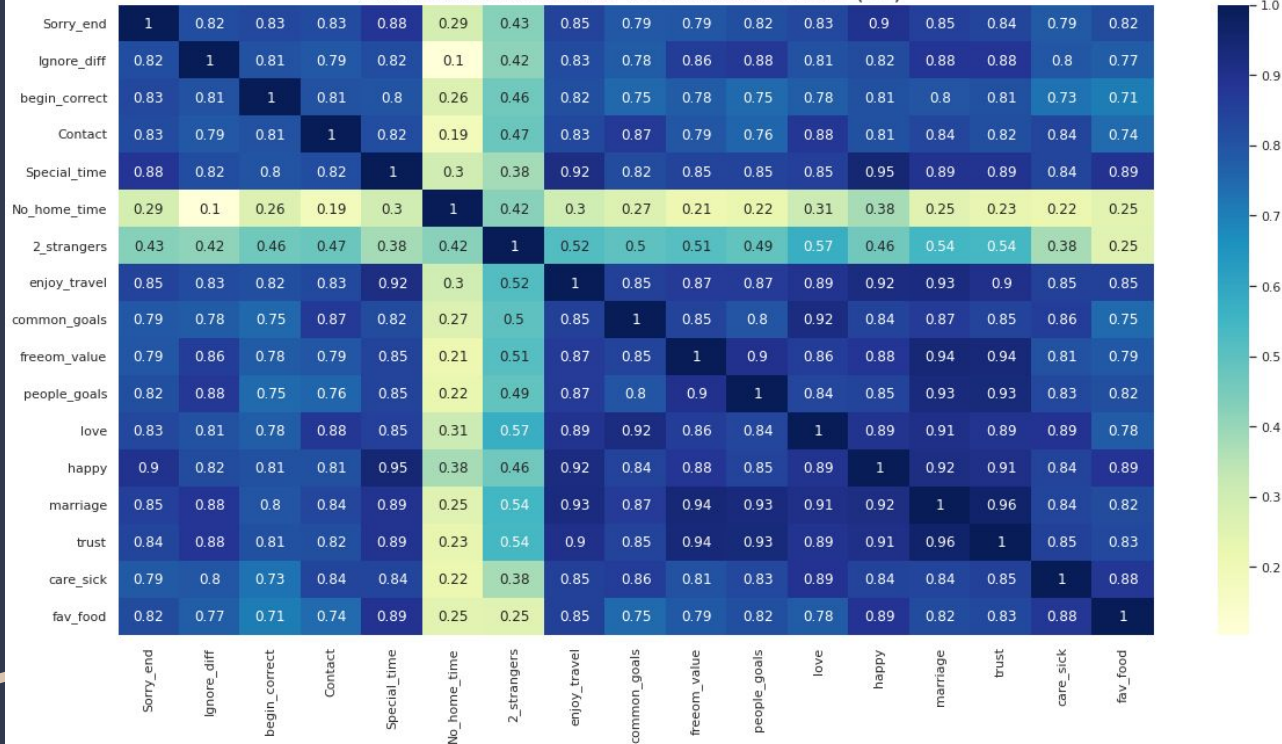
- ❑ The graphics showed that the "special time" is correlated to sharing the same views about marriage and love should be. Also, people associate special time with enjoy traveling together.

- ❑ Their freedom values are aligned with their ideas of trust and marriage and with their goals.

- ❑ At the other end, not having time at home as partners is almost not related with ignoring their differences and "not having contact.

Limitations

- Given the economic and migration crisis that was taking place in that country in 2019, the study had to be carried out with a small sample of 170 participants. And it was hard to obtain more data.

- Due the sample size, the predictive models metrics are really similar between them.

- Also, in order to make an analysis of the situation, sometimes we have to not think as an occidental mind, but like an median orient one.

Turkey: Key Facts and Figures
July 2019

UNHCR The UN Refugee Agency

Key Figures

4.0 million
Number of Persons of Concern (as of 31 July 2019)

- Syria — 3.6 M
- Afghanistan — 170,000
- Iraq — 142,000
- Iran — 39,000
- Somalia — 5,700
- Other Nationalities — 11,700

Apprehension and Interception Figures in 2019 (as of 31 July 2019)
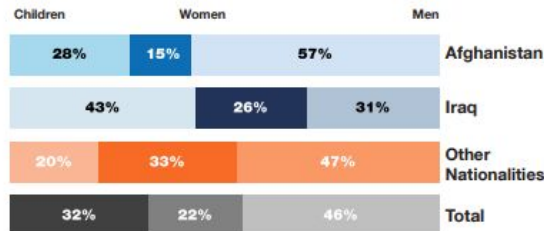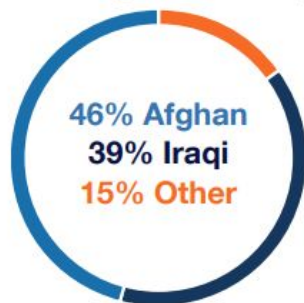
Dead / Missing in 2019 in Turkish Territorial Water: **28**

Total Interceptions at sea in 2019: **18,365**

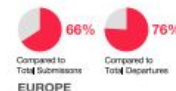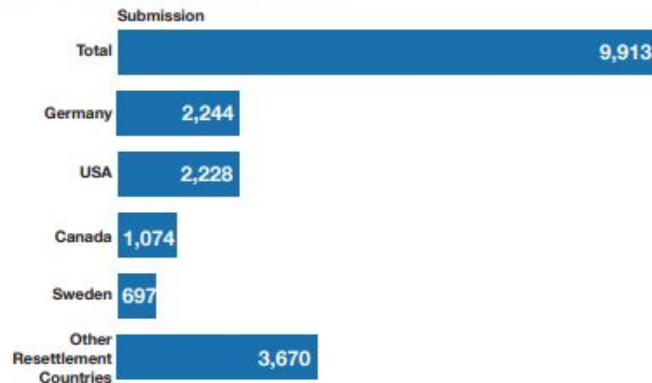Total Apprehension Incidents at western sea borders in 2019: **541**

2019 Registration Statistics
368,230 Individuals Registered with UNHCR (as of 10 September 2018)*

- 46% Afghan
- 39% Iraqi
- 15% Other

	Children	Women	Men	
Afghanistan	28%	15%	57%	
Iraq	43%	26%	31%	
Other Nationalities	20%	33%	47%	
Total	32%	22%	46%	

* UNHCR ended registration process in Turkey on 10 September 2018.
the registration process will continue with the procedure carried out by the Turkish authorities.

2019 Resettlement Statistics
9,913 Individuals Submitted (as of 31 July 2019)

Submission

- Total — 9,913
- Germany — 2,244
- USA — 2,228
- Canada — 1,074
- Sweden — 697
- Other Resettlement Countries — 3,670

USA & CANADA
- 33% Compared to Total Submissions
- 23% Compared to Total Departures

EUROPE
- 66% Compared to Total Submissions
- 76% Compared to Total Departures

AUSTRALIA & NEW ZEALAND
- <1% Compared to Total Submissions
- <1% Compared to Total Departures

98% Accuracy

97% F1 score

Conclusions

- ❏ Logistic Regression Model

- ❏ Random Forest Classifier

- ❏ Support Vector Machine Classifier

Conclusions

❏ **Random forests** are for supervised machine learning, where there is a labeled target variable. It can be used for solving regression (numeric target variable) and classification (categorical target variable) problems.

❏ Also, they are an ensemble method, meaning they combine predictions from other models.

❏ In a random forest classification, multiple decision trees are created using different random subsets of the data and features. Each decision tree is like an expert, providing its opinion on how to classify the data.

❏ Predictions are made by calculating the prediction for each decision tree, then taking the most popular result.

❏ It has high computational and memory cost, especially for large and complex data sets, and high dependency and correlation, especially for small and similar data sets. That is why, in this particular case, this is not the chosen model to recommend.

Conclusions

❏ The objective of the **Support Vector Machine** algorithm is to find a hyperplane in an N-dimensional space(N — the number of features) that distinctly classifies the data points.

❏ To separate the two classes of data points, there are many possible hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin, i.e the maximum distance between data points of both classes.

❏ Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence.

❏ Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we maximize the margin of the classifier.

❏ SVMs offer good precision and make fast predictions, but they are not suitable for large data sets due to their high training time. So,

❏ if it were to be used with a larger sample, it could be more computationally expensive.

Conclusions

❏ **Logistic regression** is fast and relatively uncomplicated, and it's convenient for you to interpret the results. It's essentially a method for binary classification.

❏ It estimates the probability of an event occurring, such as divorced or not divorced, based on a given dataset of independent variables.

❏ The RL is used when we want to investigate whether one or several variables explain a dependent variable that takes on a qualitative character.

❏ This method belongs to the family of supervised machine learning models. It is also considered a discriminative model, which means that it attempts to distinguish between classes (or categories).

❏ Since the outcome is a probability, the dependent variable is bounded between 0 and 1.

❏ Due to its simplicity of execution, reaching the same results as the other models with this small sample, and having been the model that most challenged my knowledge of mathematics and statistics, and since this is a project for educational purposes, I recommend this Regression Model Logistics.