Femicide In Turkey 2008-2020

Final project

Business Strategy for Data Scientist

Sahar Farhat

Data Story

- How many women Have been killed Between 2008 and 2020?

- What is the age range of the victims?
- What are he reasons behind the killings?
- How were they killed?
- What is the relationship between the victim and the killer?
- Where the killer punished, or has it been a restriction order?

- Killings predictions from 2021 until 2030

Data information

- The data are taken from news sites on the internet and cannot be 100% accurate. Some data are unfortunately incomplete and not available since they were not made public later. The dataset was created by Yağmur Sezen Demir and Büşra Durak by pulling the data on the internet into Excel via Selenium.

- There may be changes in the data due to the fact that some expressions used in Turkish are the same or not at all in English. For example, in Turkish, the phrase 'Religious Wedding Husband' is taken as her husband in English, but not her husband in Turkish.

Feature Explanation

- Id: identifiger
- city: which city she was killed
- age: adult / underage
- date: when it happened
- protectionorder: protection order (yes,no,undetectable)
- why1, why2: reason of being killed
- killer1, killer2: by whom
- killingway1, killingway2, killingway3: how she died
- statusofkiller: status of killer
- year: which year it happened

General information from the dataset
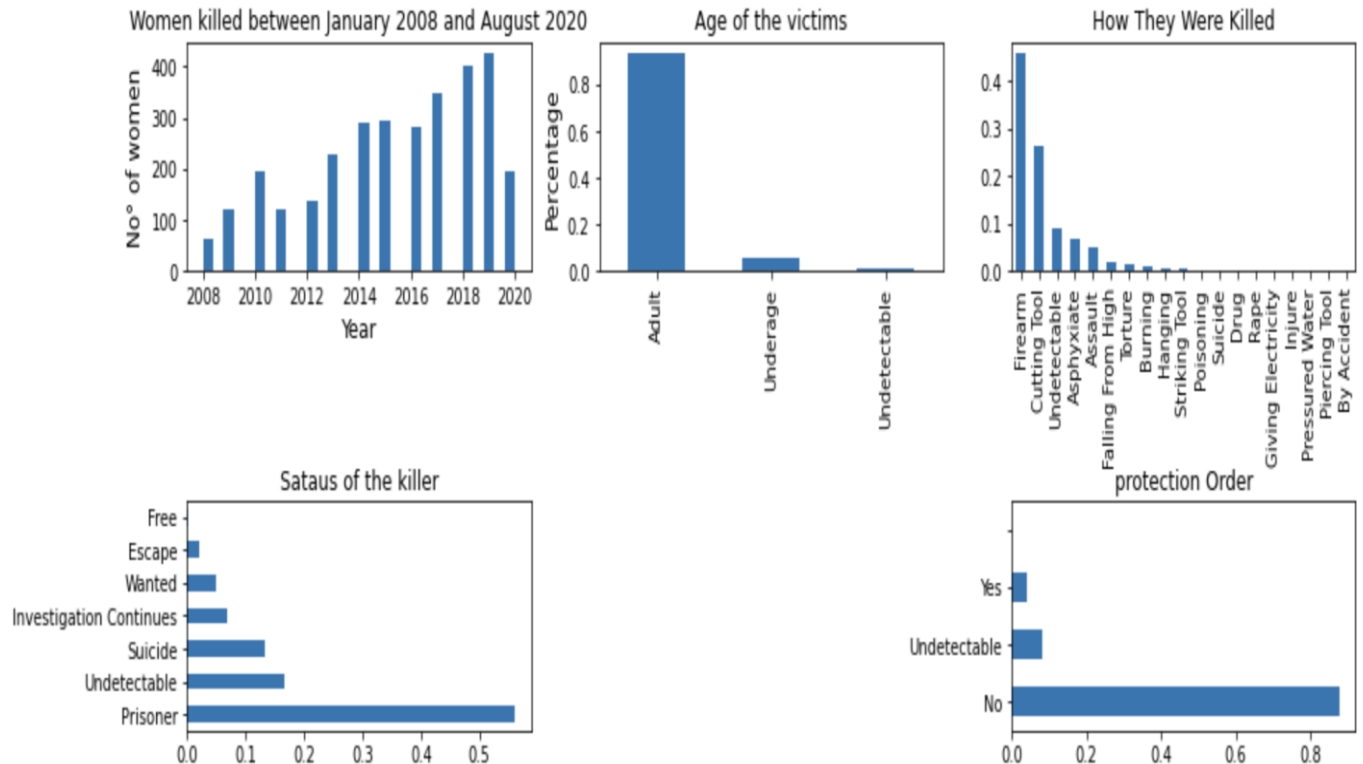
- The killing peaked between 2018 and 2019

- A total of 194 women were killed in the first 8 months of 2020.

- 80% of the victims were adults

- At the first 7 killings in the data, it is seen that almost half of them were firearms

- More than 50% of the killers are in prison

- More than 80% of the victims are not under protection order

# What are he reasons behind the killings?
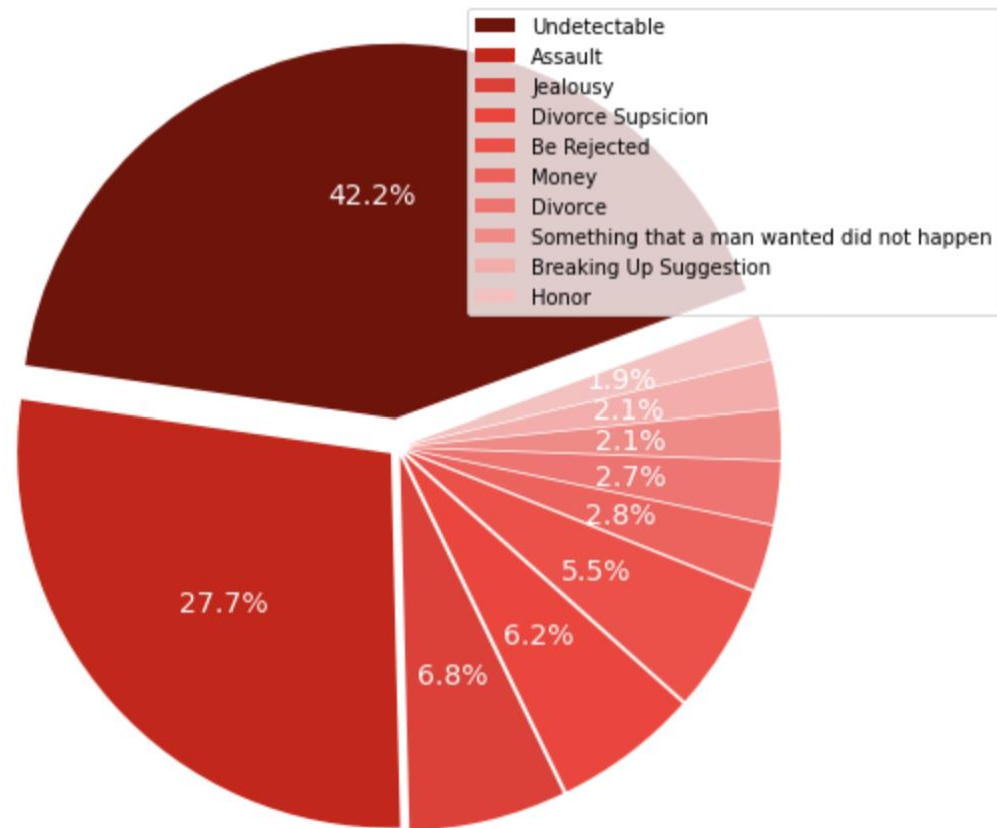
- 42.2% of the femicide reasons are undetected. Either the dataset is not filled with enough information (due to fear of telling what happened for example) or that a lot of women were victims only for unvaluable reasons.

- The rest major reasons are Assault or jealousy
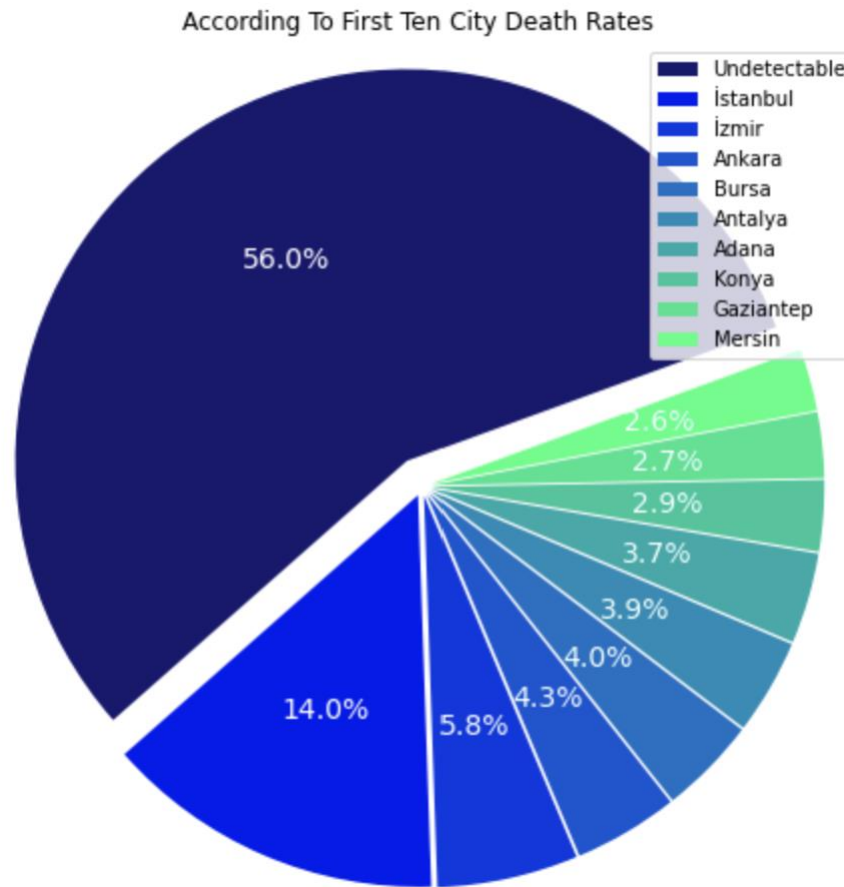
The identity of the killer

The identity of the killer

Women Who Were Killed By People With Emotional Relationship
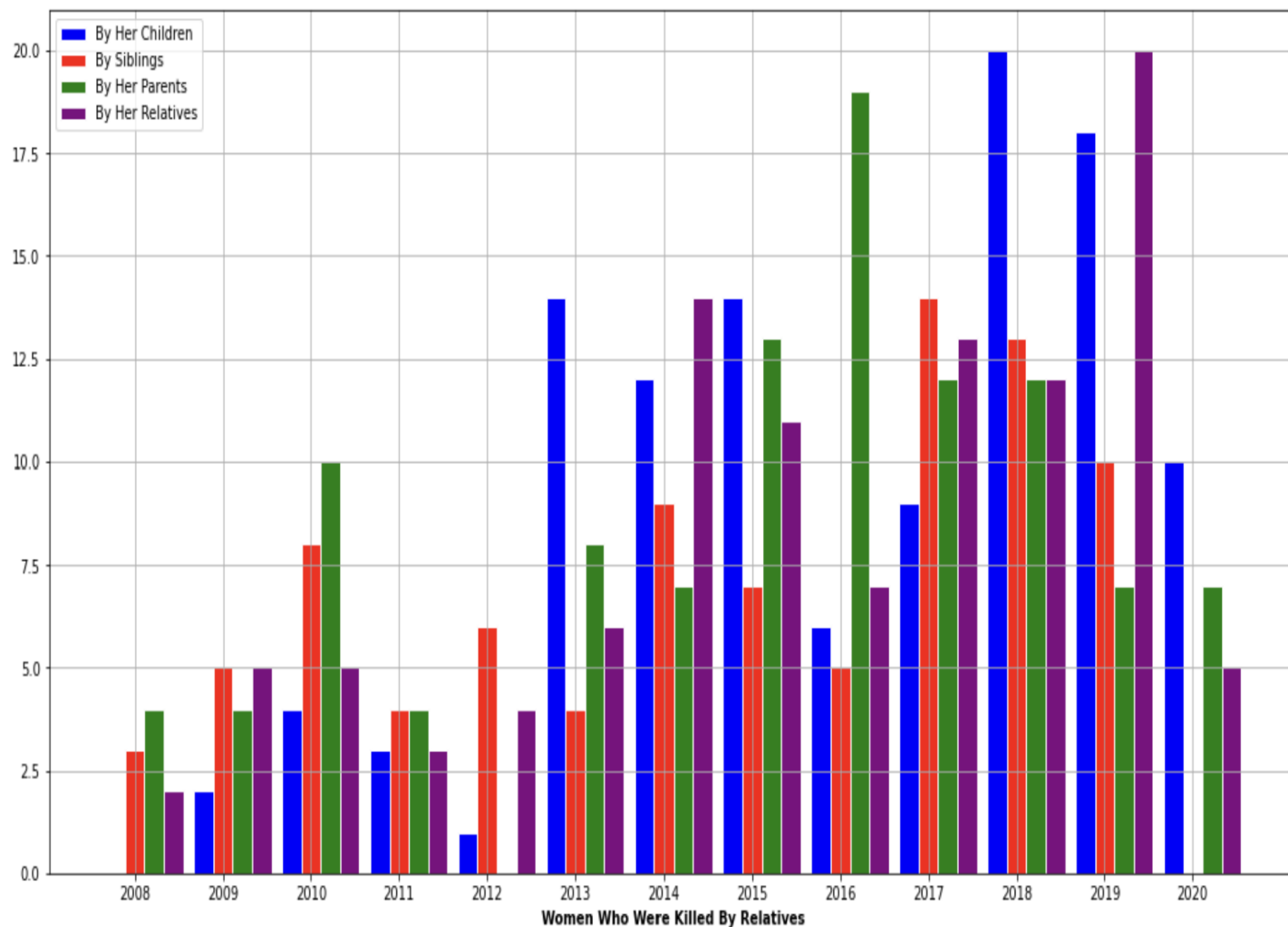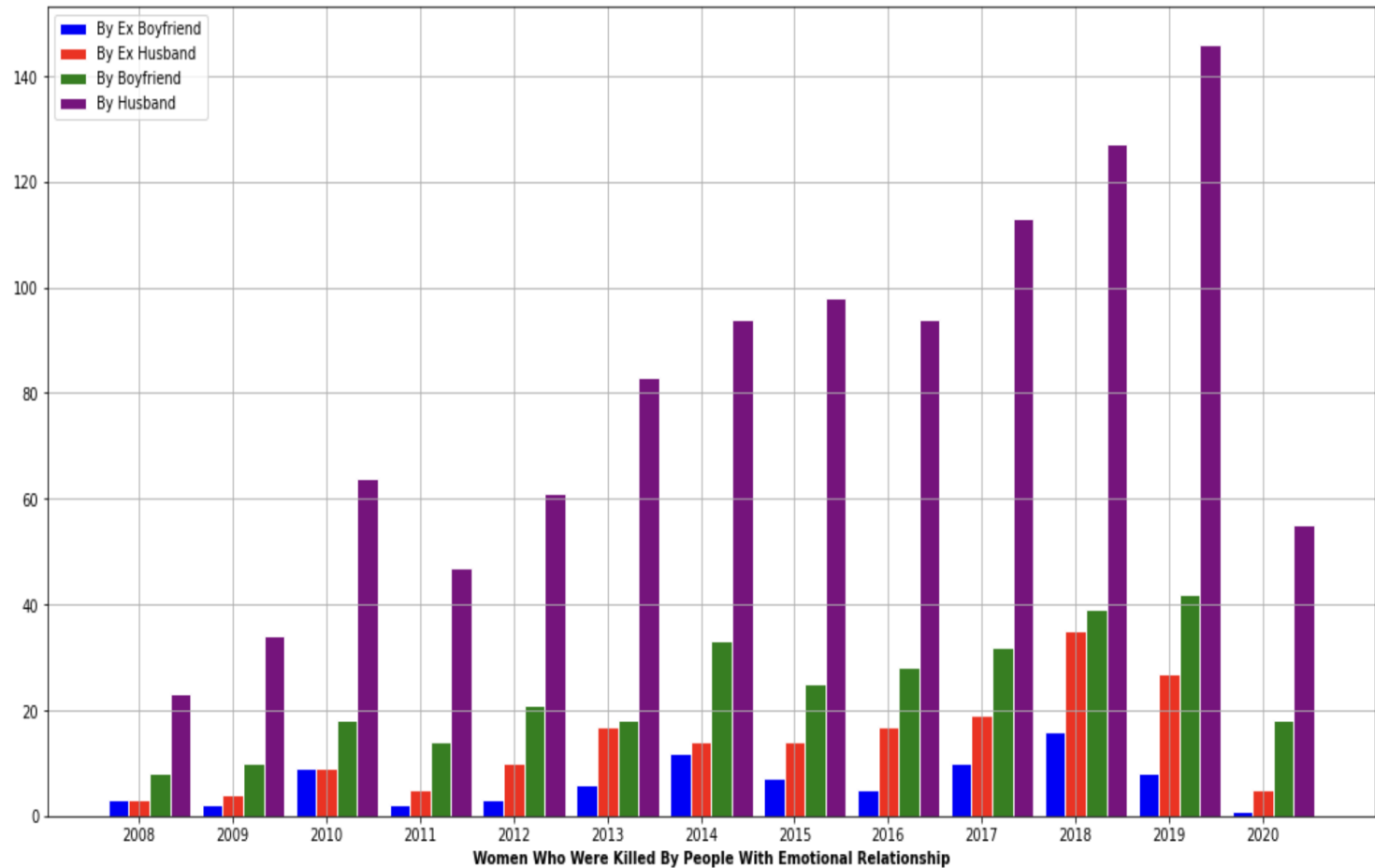
- The percentage of women killed by their own children over the years is weirdly high, an average between 10 and 20 case.

- The women killed by their parents are also with a high importance

- Except for unspecified data, a total of 1040 women were killed by their husbands.

- The total count of femicide from 2008 to 2020 is 3110

Death rate according to cities

According To First Ten City Death Rates

Legend:
- Undetectable
- İstanbul
- İzmir
- Ankara
- Bursa
- Antalya
- Adana
- Konya
- Gaziantep
- Mersin

56.0%
14.0%
5.8%
4.3%
4.0%
3.9%
3.7%
2.9%
2.7%
2.6%

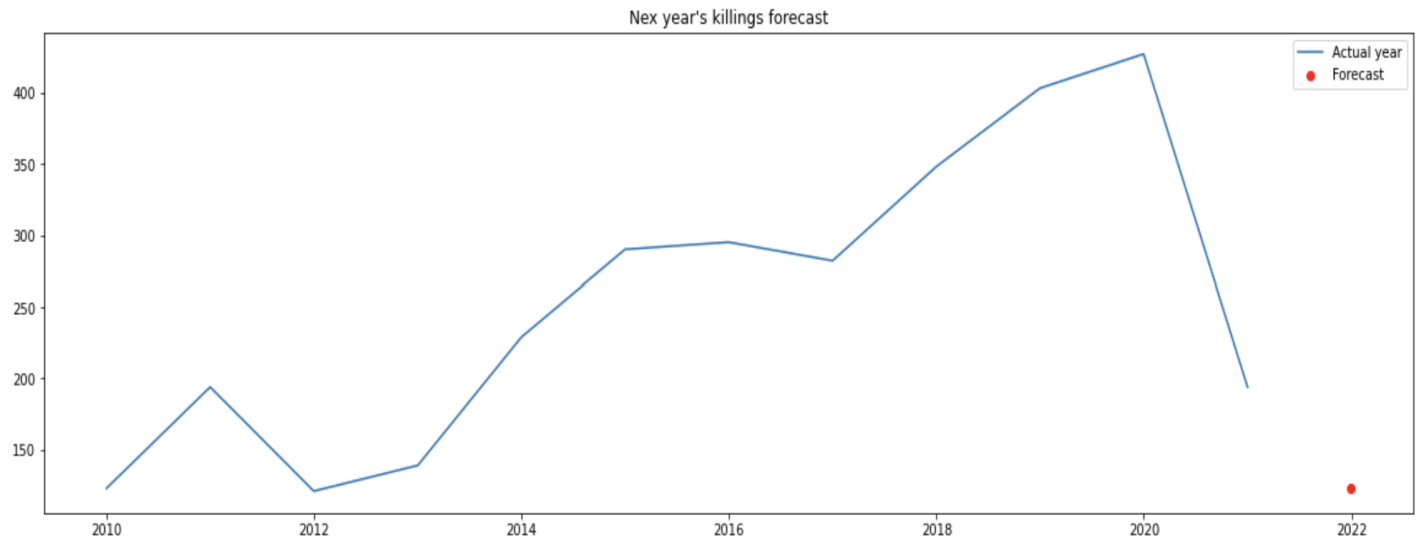Prediction Model
form 2021 until 2030

Regression Models

- A multiple regression models where tested (Linear Regression, XG Boost Regressor and Random Forest Regressor)

The regression Model	RMSE
Linear Regression	73.0749
XG Boost Regressor	3.250
Random Forest Regressor	33.295

So, we will choose XGB Regressor

XGB Regressor Prediction

The prediction of the regressor is not precise enough

Auto-Arima Prediction

For the auto-ARIMA we started by splitting the data into test and training and we dropped the new feature that we have created to be able to apply the regression models. From the SARIMAX report we concluded that the term is statically significant with a p value= 0.051. All the results implied that the auto-ARIMA is a well fitted model that can forecast the sales accurately.

Auto-Arima Prediction

Next years pridiction from auto-ARIMA
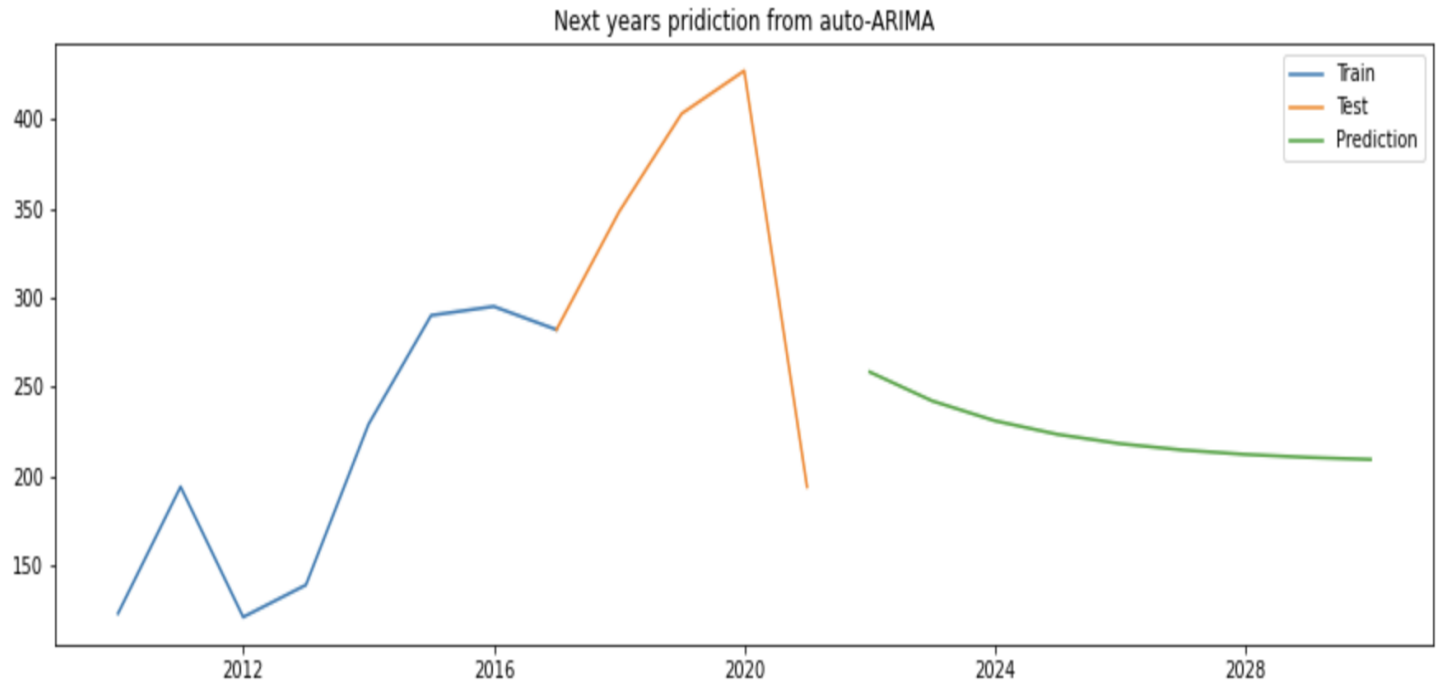
Auto-Arima Prediction

```
Next years pridiction from auto-ARIMA
2021-12-31     258.305393
2022-12-31     242.075960
2023-12-31     230.959738
2024-12-31     223.345771
2025-12-31     218.130645
2026-12-31     214.558585
2027-12-31     212.111931
2028-12-31     210.436115
2029-12-31     209.288279
2030-12-31     208.502077
Freq: A-DEC, Name: Killings, dtype: float64
```

Conclusion

- The Femicide case seems to be growing dramatically

- Though the dataset is not recent, and it is missing a lot of values it still gives a general idea about the gravity of thee situation

- If there is more data the report will help to have more precise results to help decrease the number of yearly deaths (for example more precision about the murder location can help taking more restrictive measures).