Traffic Safety Forecasting in Turkey (2002–2030)

Final Project Report

---

1. Introduction

This project investigates the trends in traffic safety in Turkey over the years 2002 to 2023 and forecasts future trends through 2030. The goal is to determine whether the roads have become safer since 2002, and to estimate whether they will be safer in the near future using statistical modeling and hypothesis testing.

Four key safety indicators were analyzed:

1. **Accidents per Vehicle**: The total number of accidents divided by the total number of vehicles in a given year.
2. **Deaths per Accident**: The total number of fatalities divided by the total number of accidents.
3. **Injuries per Accident**: The total number of injuries divided by the total number of accidents.
4. **Driver-Fault Accident Ratio**: The number of driver-fault accidents divided by the number of accidents that involved injury or death. (This indicator which we didn't use in previus EDA steps was added to further strengthen our predictions)

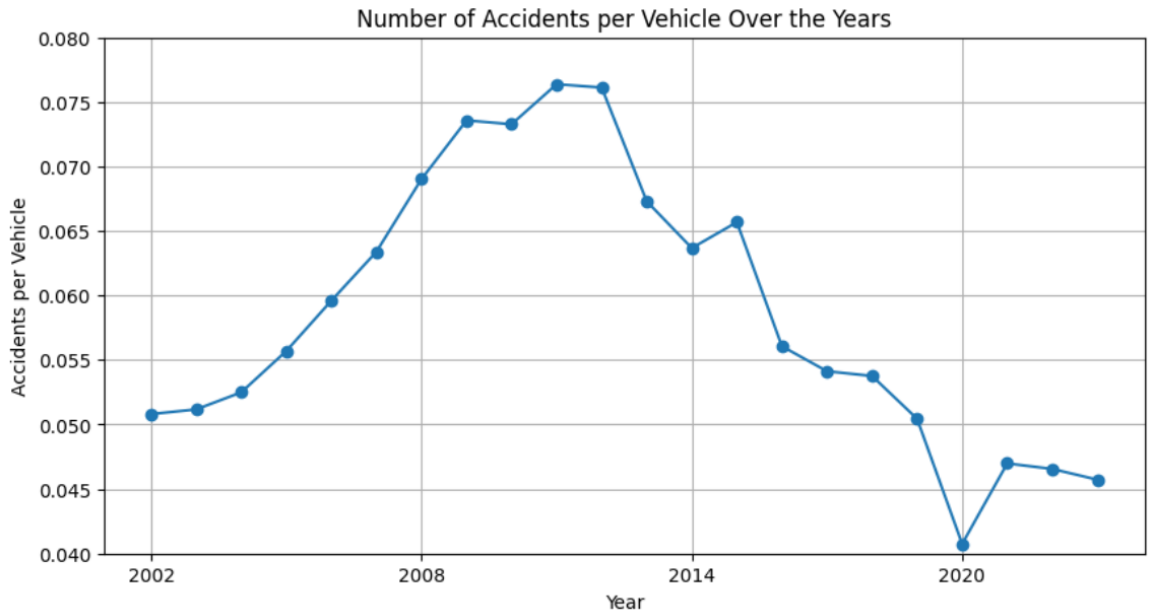
First, EDA was conducted on these four indicators so better visualize their trends in the recent history. Then, hypothesis tests were done to figure out whether there are any significant safety differences between the pre-2015 and post-2015 eras.

After that, we used both linear and polynomial regression models to predict each indicator and then compared their suitability in giving a reliable prediction. Finally, we used t-tests to compare historical and forecasted values to determine whether road safety is likely to improve.

---

2. Exploratory Data Analysis

Here is the results of the exploratory data analysis done on all 4 of our indicators:



Minimum rate of accidents per vehicle: 0.04074606861411521

Maximum rate of accidents per vehicle: 0.07638061228396507

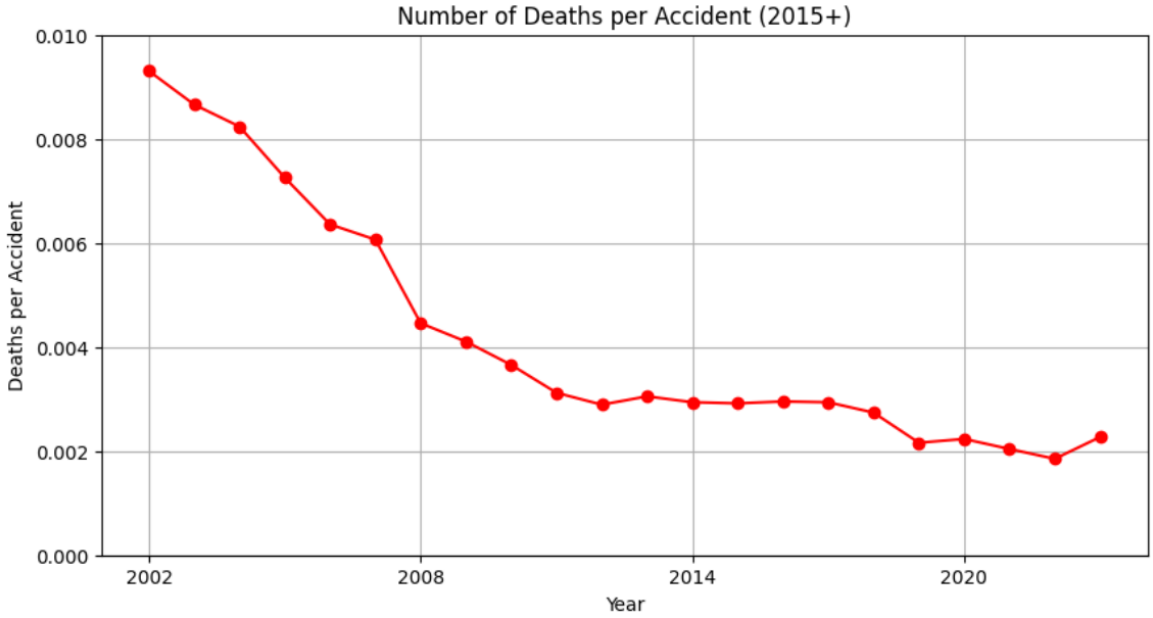
Mean before 2015: 0.06404078110031187

Mean after 2015: 0.051122843773686644

Difference between means: -0.012917937326625231

Std dev before 2015: 0.009459407313590941

Std dev after 2015: 0.007312708410178602



Minimum rate of deaths per accident: 0.0018508350250657567

Maximum rate of deaths per accident: 0.009306989678859968

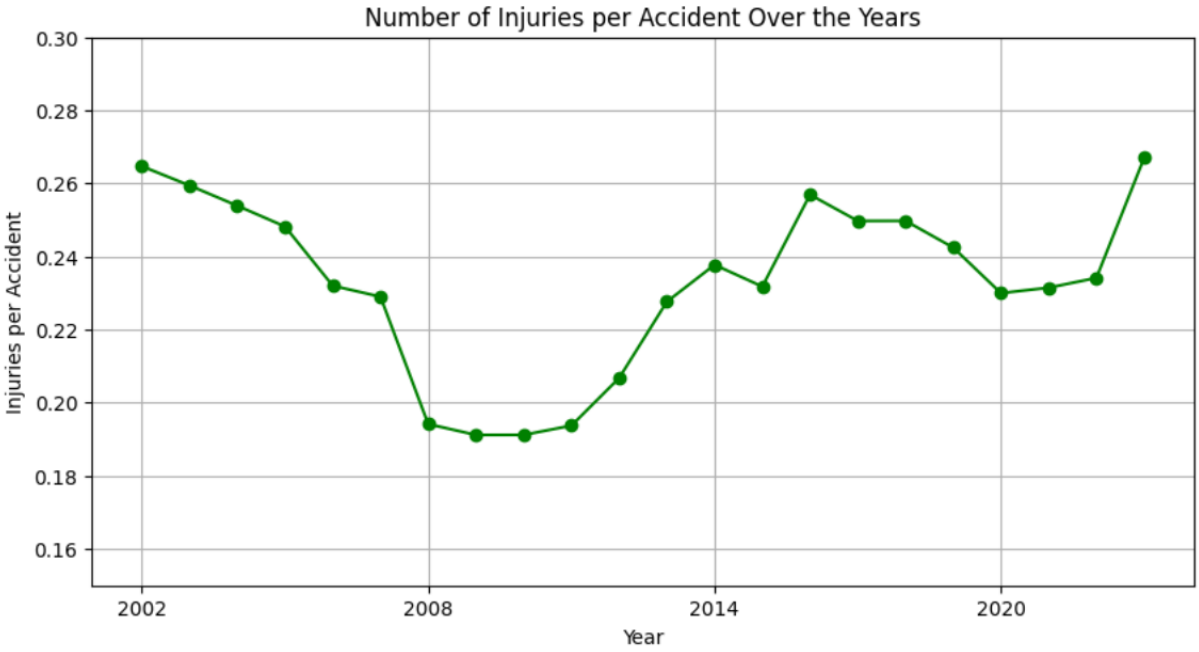
Mean before 2015: 0.005393015838183288

Mean after 2015: 0.002456105279608589

Difference between means: -0.0029369105585746987

Std dev before 2015: 0.0023702147622526423

Std dev after 2015: 0.0004306117759313482 7



Minimum rate of injuries per accident: 0.19118124528882247

Maximum rate of injuries per accident: 0.2669853044129375

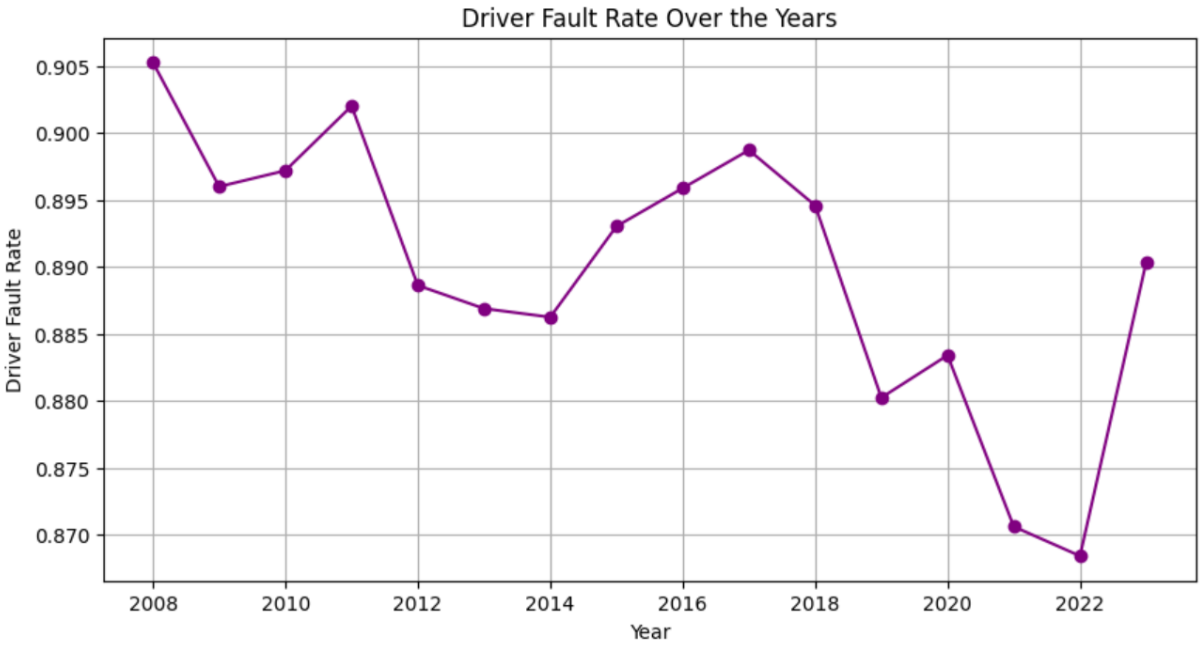
Mean before 2015: 0.22535856957180853

Mean after 2015: 0.2437017416772691

Difference between means: 0.018343172105460576

Std dev before 2015: 0.02726406833327248

Std dev after 2015: 0.013051369215305177



Minimum fault rate: 0.8684

Maximum fault rate: 0.9053

Mean before 2015: 0.8946

Mean after 2015: 0.8861

Difference between means: -0.0085

Std dev before 2015: 0.0075

Std dev after 2015: 0.0111

---

3. Hypothesis Testing

We determine our general hypotheses as such:

- Null Hypothesis: There is no difference between then and now in terms of safety
- Alternate Hypothesis: The roads are more safer now than before

Again, the time peroids will be 2002-2014 for before and 2015-2023 for now. To make a conclusion we will make 4 t-tests, on our 4 safety parameters. If at least 3 of our test accept the alternate hypothesis, then we will accept our general alternate hypothesis(i.e. the roads are safer now). Our t-tests will be one-tailed, meaning we are only checking if there is a decrease in the parameters we are researching.

Accidents per vehicle: t = 3.6071; p-value = 0.0008

Deaths per accident: t = 4.3647 ; p-value = 0.0003

Injuries per accident: t = -2.1026 ; p-value = 0.9751

Fault rate: t = 1.8105 ; p-value = 0.0460

[Accidents per vehicle] --> Significant difference (p = 0.0009) → Reject H_0

[Deaths per accident] --> Significant difference (p = 0.0004) → Reject H_0

[Injuries per accident] --> No significant difference (p = 0.9752) → Fail to reject H_0

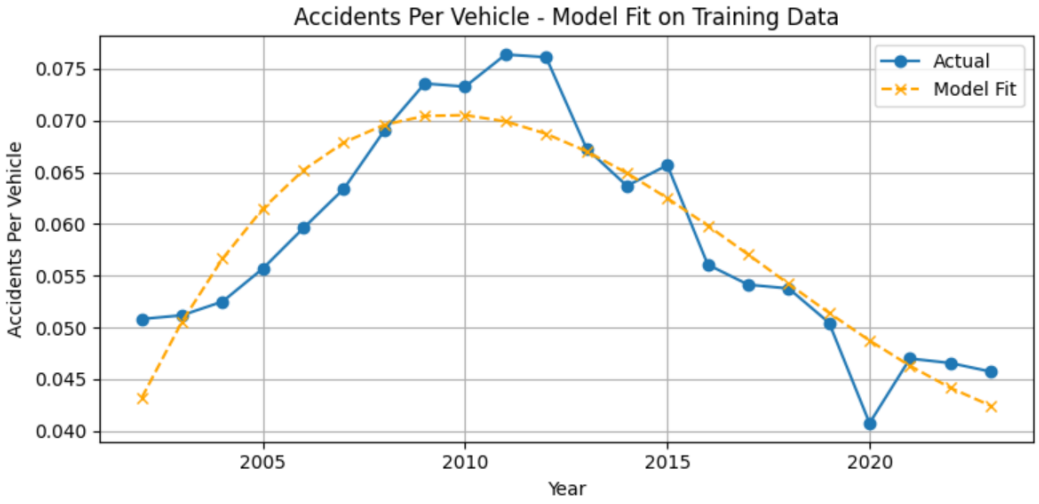
[Fault rate] --> Significant difference (p = 0.0460) → Reject H_0

4. Linear Regression vs. Polynomial Regression

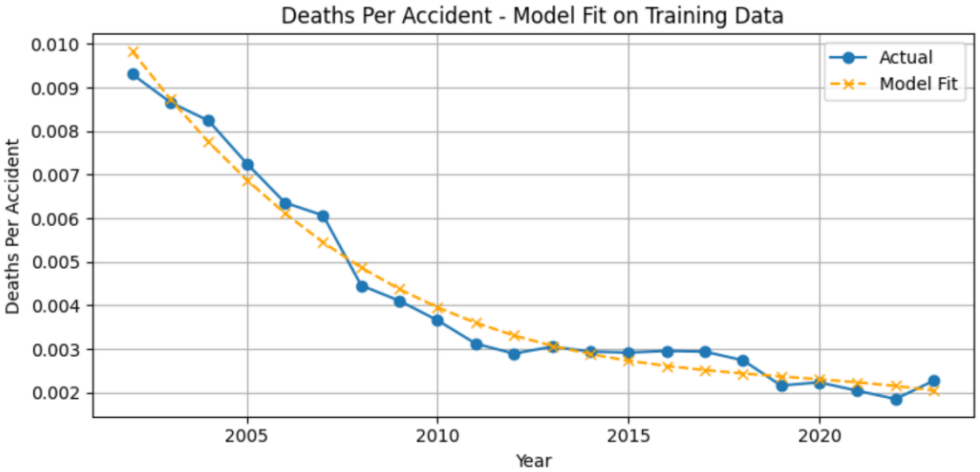
For each of the four safety indicators, both linear regression and polynomial regression (degree 3) models were trained using historical data from 2002 to 2023. To ensure model reliability and reduce overfitting, we applied k-fold cross-validation using 3 folds during training. The models were then evaluated using two standard metrics:

- **R^2 Score (Coefficient of Determination)**: Indicates how well the model explains the variance in the data.
- **Mean Squared Error (MSE)**: Measures the average squared difference between observed and predicted values.

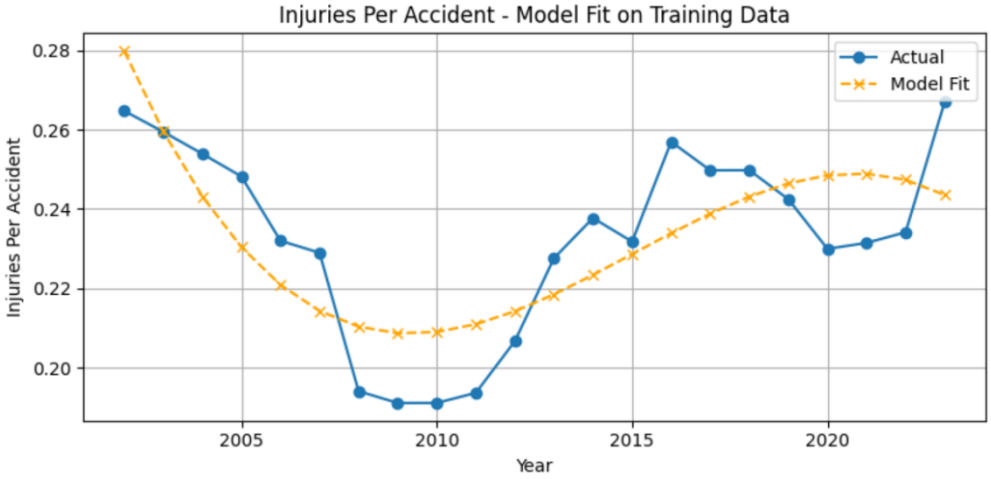
➔ *POLYNOMIAL REGRESSION:*



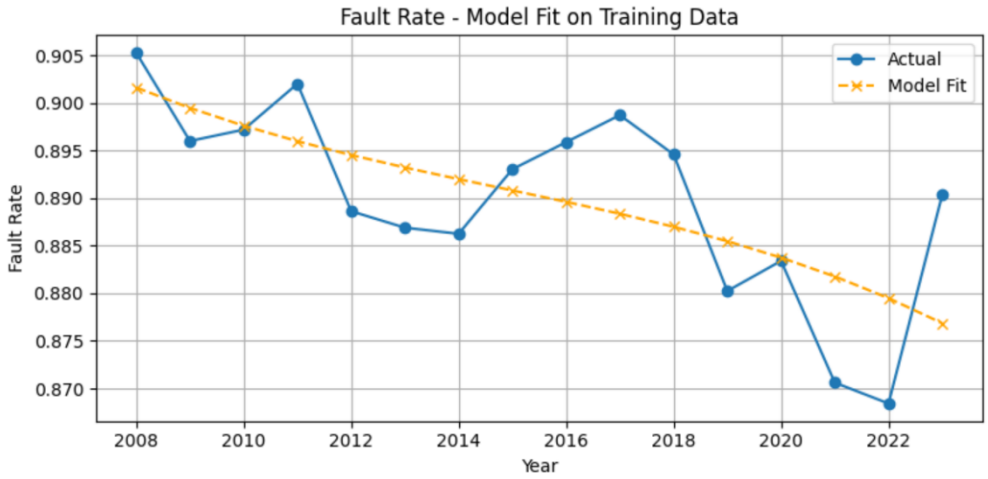
MSE: 0.000018 <---> R^2: 0.836053



MSE: 0.000000 <---> R^2: 0.978317

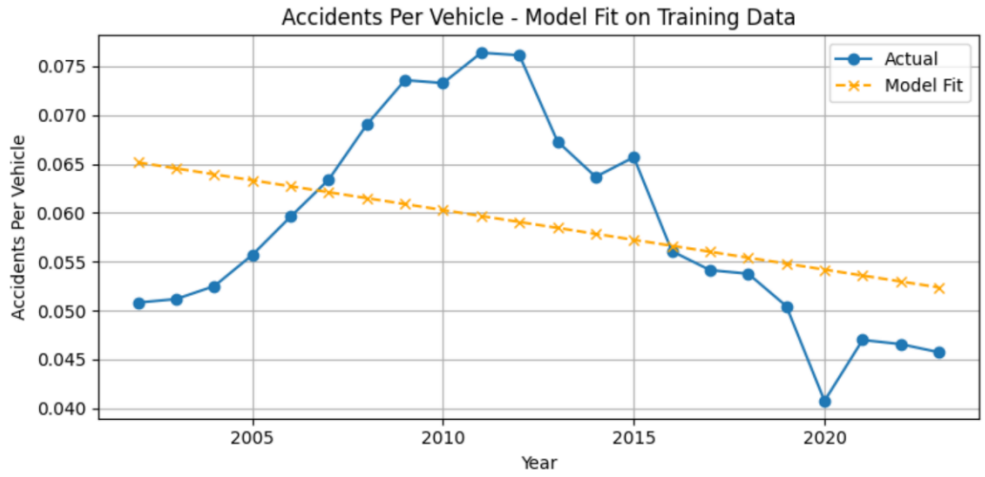


MSE: 0.000210 <---> R^2: 0.617154

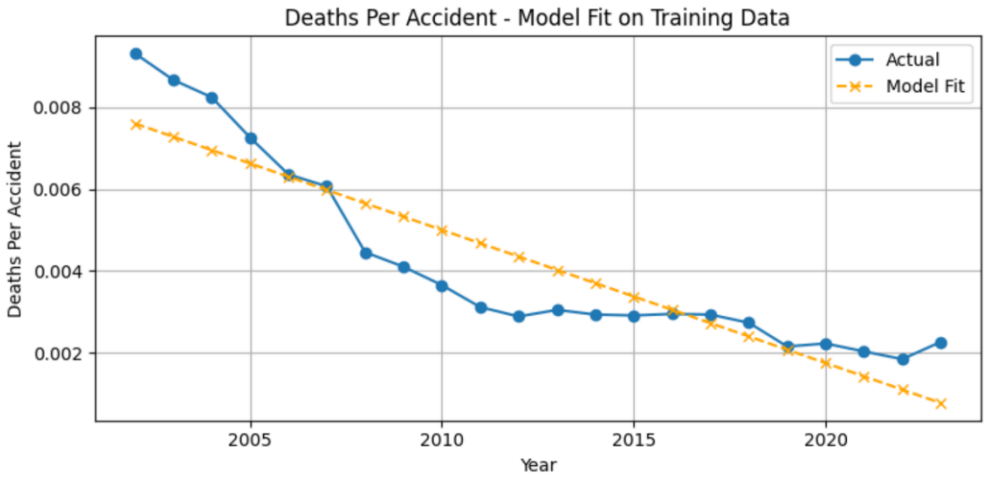


MSE: 0.000052 <---> R^2: 0.481239

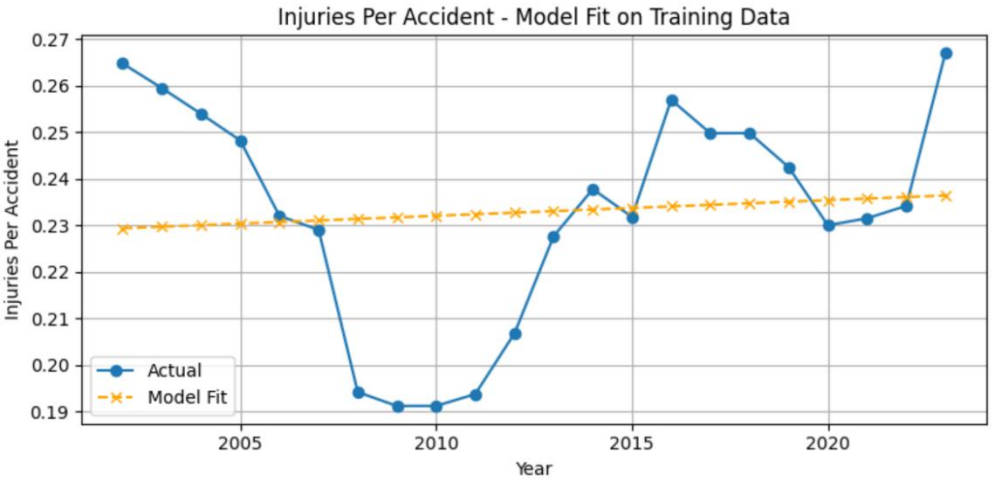
➔ *LINEAR REGRESSION:*



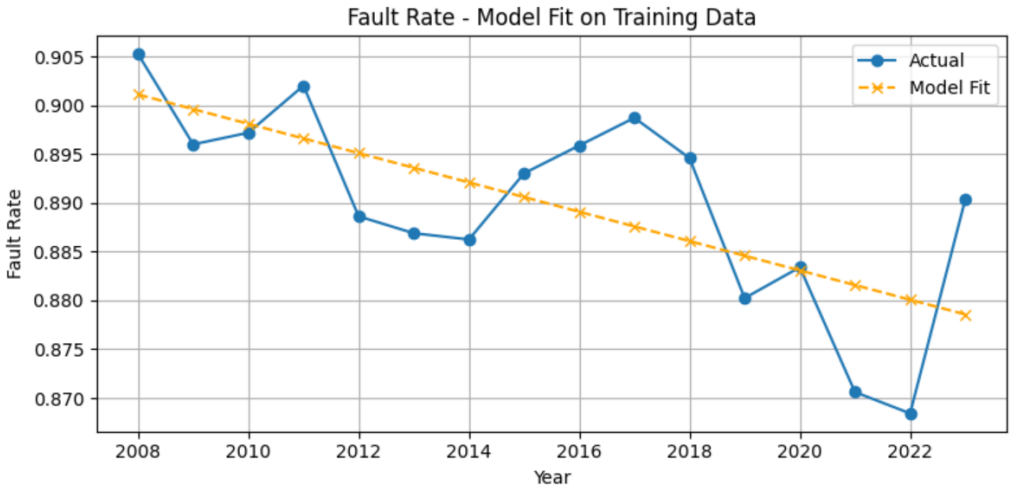
MSE: 0.000094 <---> R^2: 0.137547



MSE: 0.000001 <---> R^2: 0.812806



MSE: 0.000544 <---> R^2: 0.008240



MSE: 0.000053 <---> R^2: 0.476469

From these results, we can see that for most cases the polynomial regression model fits better (higher R^2 and lower MSE). However, due to some inconsistent predictions, we have decided to use linear regression model's predictions for the "injuries per accident" indicator. Rest of the indicators are predicted using the polynomial regression model.

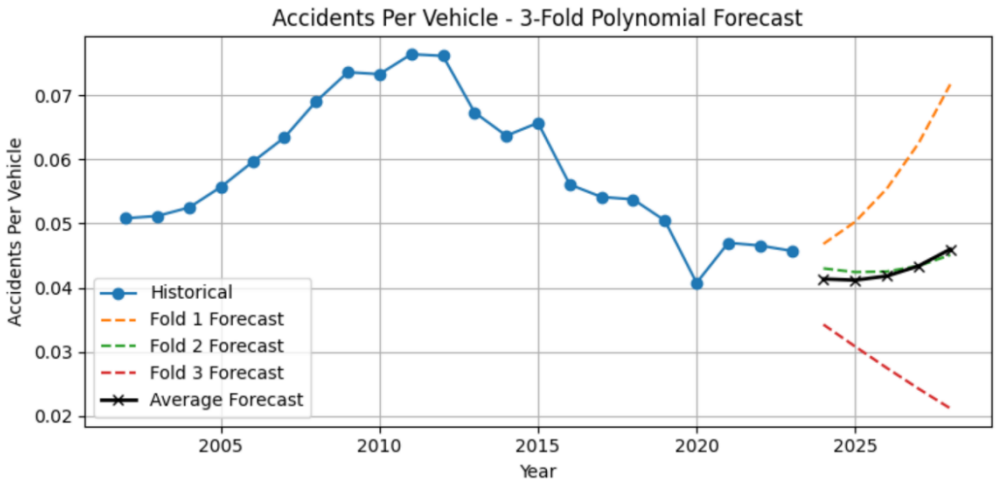
---

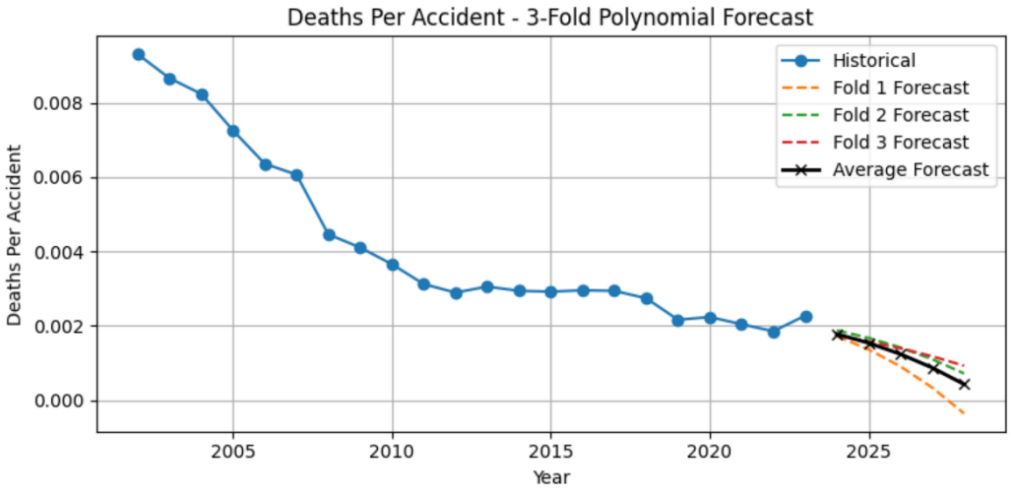
5. Prediction Results (2024–2030)

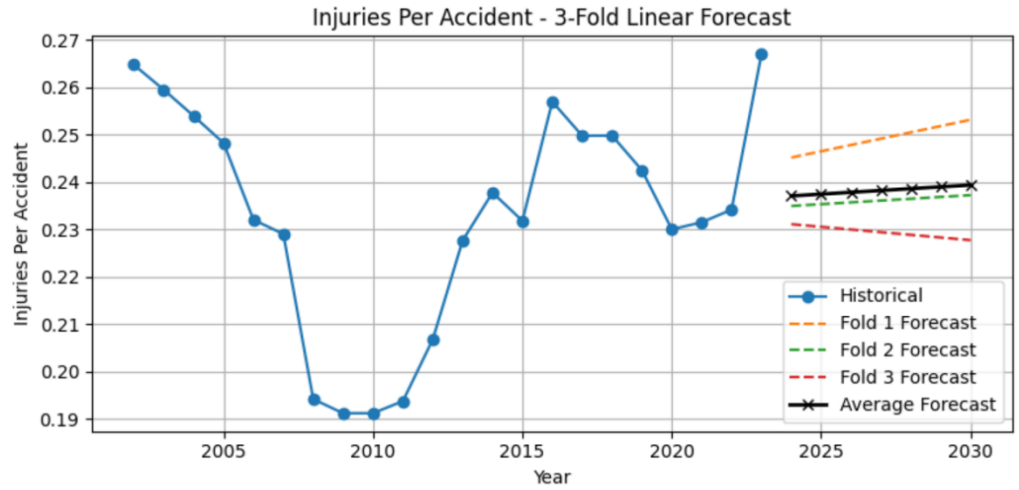
Using the polynomial & linear regression models trained on the full historical data (2002–2023), predictions were made for each indicator for the years 2024 through 2030.

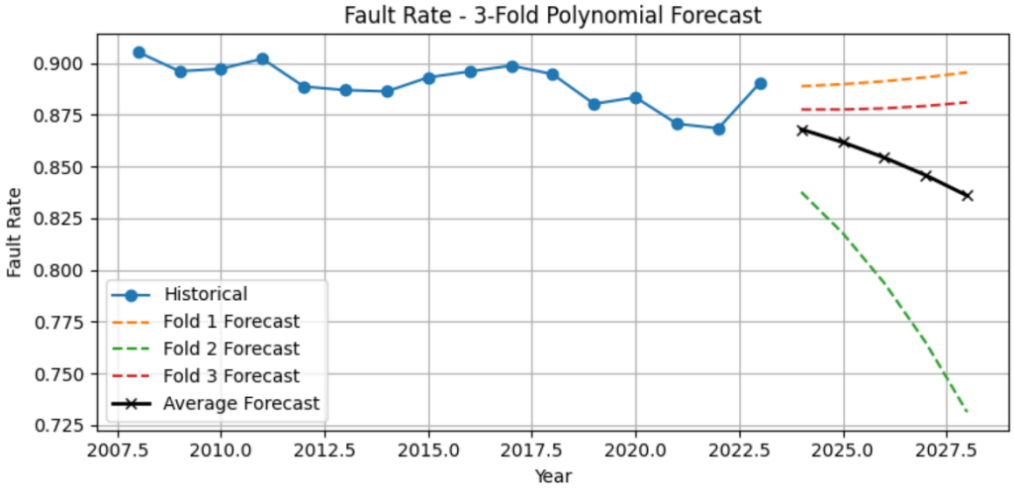
Any predicted values that fell below zero were clipped to zero, as negative values are not meaningful for these indicators.

The following table summarizes the average values for each indicator over three distinct periods: early historical years (2002–2014), recent historical years (2015–2023), and forecasted years (2024–2030). These averages are used in the subsequent statistical tests.




Deaths Per Accident - 3-Fold Polynomial Forecast
Deaths Per Accident
Year
Historical
Fold 1 Forecast
Fold 2 Forecast
Fold 3 Forecast
Average Forecast
0.000
0.002
0.004
0.006
0.008
2005
2010
2015
2020
2025


Injuries Per Accident - 3-Fold Linear Forecast
Injuries Per Accident
Year
Historical
Fold 1 Forecast
Fold 2 Forecast
Fold 3 Forecast
Average Forecast
0.19
0.20
0.21
0.22
0.23
0.24
0.25
0.26
0.27
2005
2010
2015
2020
2025
2030


Fault Rate - 3-Fold Polynomial Forecast
Fault Rate
Year
Historical
Fold 1 Forecast
Fold 2 Forecast
Fold 3 Forecast
Average Forecast
0.725
0.750
0.775
0.800
0.825
0.850
0.875
0.900
2007.5
2010.0
2012.5
2015.0
2017.5
2020.0
2022.5
2025.0
2027.5

Indicator	Historic Mean (2002–2023) /	Predicted Mean (2023–2028)
Accidents per Vehicle	0.059377	0.042737
Deaths per Accident	0.004283	0.001169
Injuries per Accident	0.232863	0.238197
Driver-Fault Accident Ratio	0.889803	0.853129

---

6. T-Test: Statistical Comparison of Historical and Forecasted Values

To determine whether the forecasted safety indicators reflect statistically significant improvements, one-tailed t-tests were conducted. These tests compare the historical means from the 2002–2014 period against the predicted means from 2024–2030.

The hypotheses for each indicator were defined as follows:

- For **Accidents per Vehicle**, **Deaths per Accident**, **Driver-Fault Accident Ratio and Injuries Per Accident**:
    - **Null Hypothesis (H_0)**: The predicted mean value for 2024–2030 is greater than or equal to the historical mean value (2002–2023) (no improvement).
    - **Alternative Hypothesis (H_1)**: The mean value for 2024–2030 is lower than the historical mean (indicating improvement in safety).

The significance level (alpha) was set to 0.05. If the p-value from the t-test was below 0.05, the null hypothesis was rejected in favor of the alternative, indicating a statistically significant change.

If two or more of the indicators show statistically significant improvement, we conclude that road safety in Turkey is projected to improve.

- *Results:*
    - *Accidents per Vehicle:*
        - *t-statistic: 6.7507*
        - *p-value: 0.00000029*
        - ***-> Significant decrease detected (p < 0.05)***
    - *Deaths per Accident:*
        - *t-statistic: 5.4920*
        - *p-value: 0.00000603*
        - ***-> Significant decrease detected (p < 0.05)***

- *Injuries per Accident:*
  - *t-statistic: 17.048168*
  - *p-value: 0.9999*
  - ***=> Fail to reject null hypothesis: No significant decrease predicted.***

- *Fault Rate:*
  - *t-statistic: 5.8049*
  - *p-value: 0.000563*
  - ***-> Significant decrease detected (p < 0.05)***

Since we have rejected the null hypothesis in 3 out of 4 t-tests we have done (which means we have detected a significant decrease in 3 out of 4 indicators' values), we can say that according to this model, the safety of the roads in the near future is expected to increase.

---

7. Conclusion

This study analyzed traffic safety trends in Turkey from 2002 to 2023 and forecasted future trends through 2030 using statistical modeling and hypothesis testing. Four key safety indicators were examined to evaluate whether the country's roads have become safer over time. The results from exploratory data analysis and t-tests confirmed that significant improvements have occurred in most indicators since 2015.

Moreover, regression-based forecasting models predicted continued improvement for three out of four indicators through 2030. With statistically significant decreases in accident rates, fatalities, and driver-fault ratios, we conclude that traffic safety in Turkey is on a positive trajectory and is expected to improve further in the coming years.