MODS 206 - Applied Econometrics

- Vinicius Mori
- Guilherme Kenzo Yamaguti
- Gustavo Hess
- Luiz Fonteles
- Fredo Alejos

Describing the dataset

city	station	date	weather	dow	hour	speed	vehicleType
1	1	2021-04-01	2.264056	4.0	13.957664	39.421352	2.255187
1	1	2021-04-02	2.380476	5.0	12.697818	39.434616	2.258038
1	1	2021-04-03	1.749168	6.0	11.838524	39.523284	2.250378
1	1	2021-04-04	2.144361	7.0	11.329315	39.538990	2.251946
1	1	2021-04-05	2.144267	1.0	13.409771	39.525909	2.249219

energyConsumption	trafficCongestion	carCrash	pedestrianDeath	treated
8.586321	2.884873	0.003347	0.000335	1
8.586430	2.772947	0.002332	0.000833	1
8.562825	3.000000	0.003175	0.000454	1
8.565154	3.000000	0.002136	0.000610	1
8.564945	2.880901	0.003455	0.000329	1

treated_time	size	population	popDensity	cityName
-61	192	379909	1978.692708	Azzurropoli
-60	192	379909	1978.692708	Azzurropoli
-59	192	379909	1978.692708	Azzurropoli
-58	192	379909	1978.692708	Azzurropoli
-57	192	379909	1978.692708	Azzurropoli

→ The dataset shows us 6 different cities with 10 stations each. Initially this table had one line per vehicle as it passed by the stations. Then it was grouped by day and station and we believe the other columns were averaged.

Describing the dataset

- **Weather:** weather condition at that date & station;
- **dow:** day of week;
- **hour:** hour of day (0–23) when the observation was aggregated;
- **speed:** average vehicle speed at that station and hour;
- **vehicleType:** dominant vehicle class contributing to the speed/energy metrics;
- **energyConsumption:** estimated energy usage by all vehicles in that hour (per station);
- **trafficCongestion:** calculation of congestion level;
- **carCrash:** number of reported car crashes during that hour at that station;
- **pedestrianDeath:** number of reported pedestrian deaths during that hour at that station.

Describing the dataset

- **Treated:** variable that indicated in which stations the policy was applied;

- **treated_time:** this show the time (-60 to 60) where 0 is the time the treatment was applied;

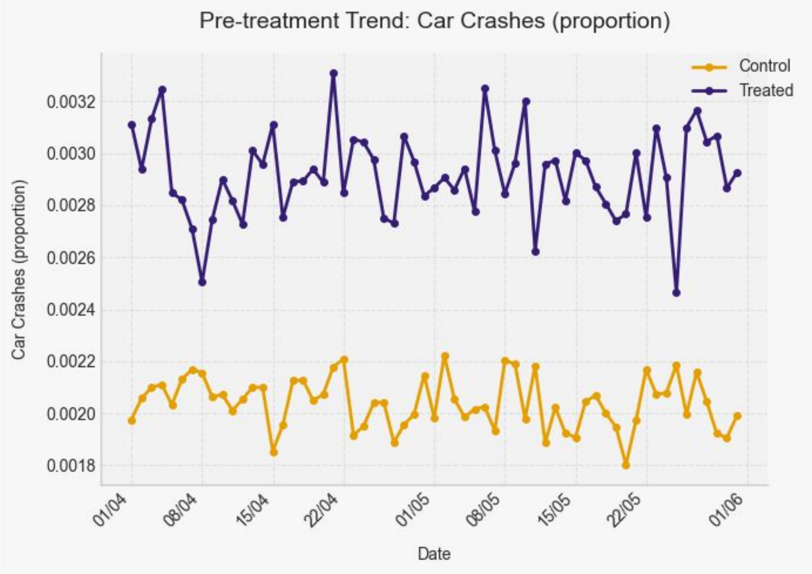
- **Size:** area of the city;

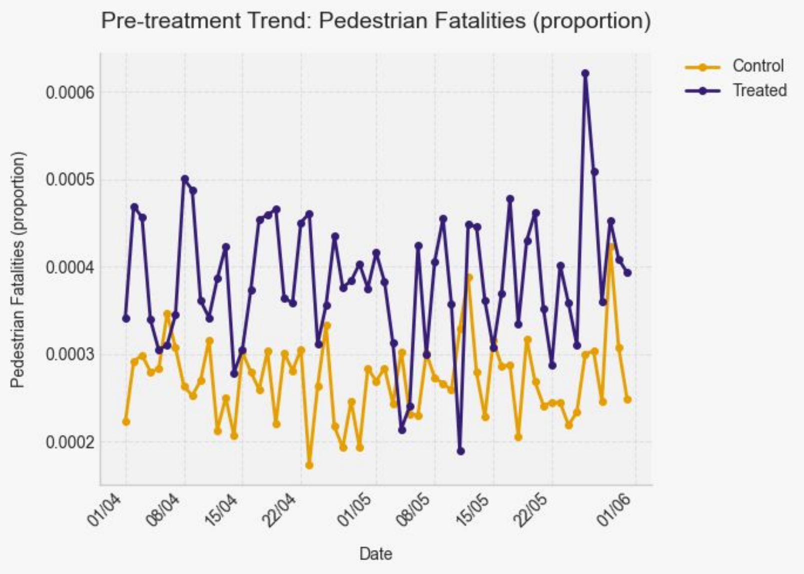
- **Population:** total number of inhabitants within the city;

- **popDensity:** population per square kilometer;

- **cityName:** name of the city.

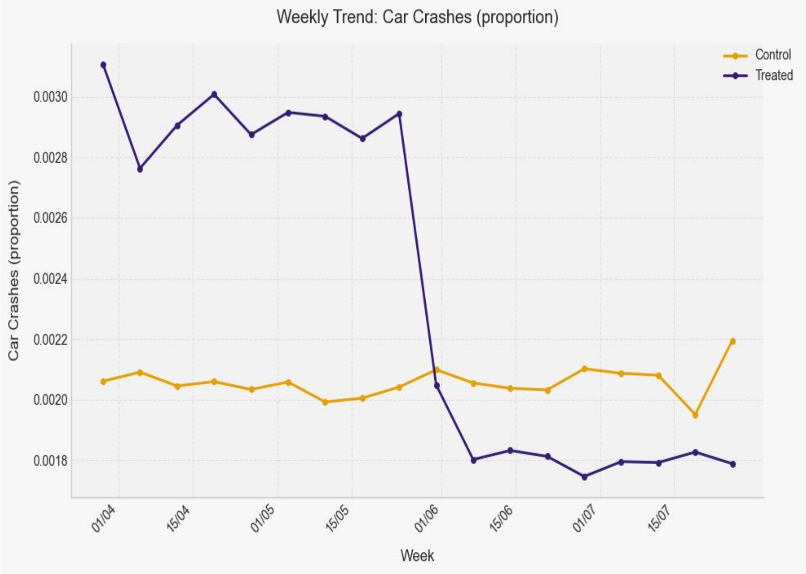
What was found

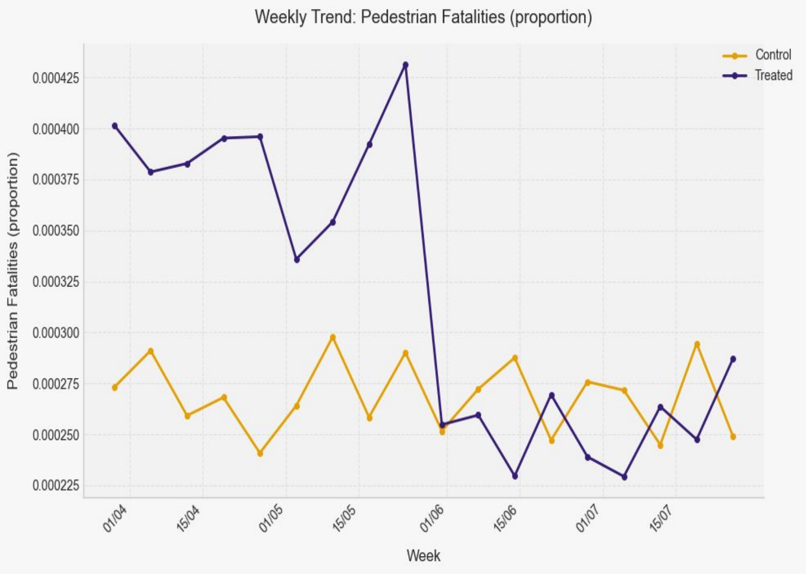




On average, pedestrian deaths and car crashes in the treated group are higher than in the control group before measure implementation.

What was found



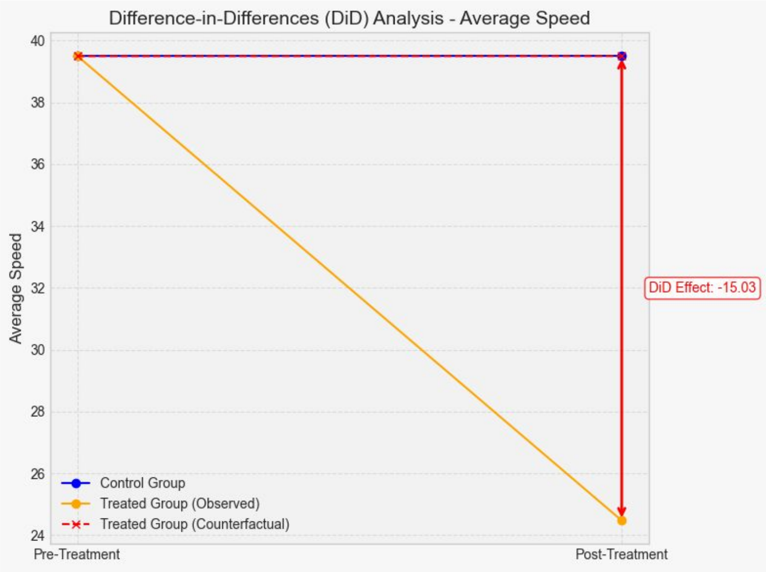


After the implementation, we saw a substantial reduction in car crashes and pedestrian fatalities.

What was found

These results were most likely obtained via speed limit implementation used in the implementation group.

DiD




```
Outcome variable: speed
Intercept: +39.5007
Coefficients:
  treated: -0.0045
  post   : +0.0021
  DiD    : -15.0281  <--- Average Treatment Effect
```

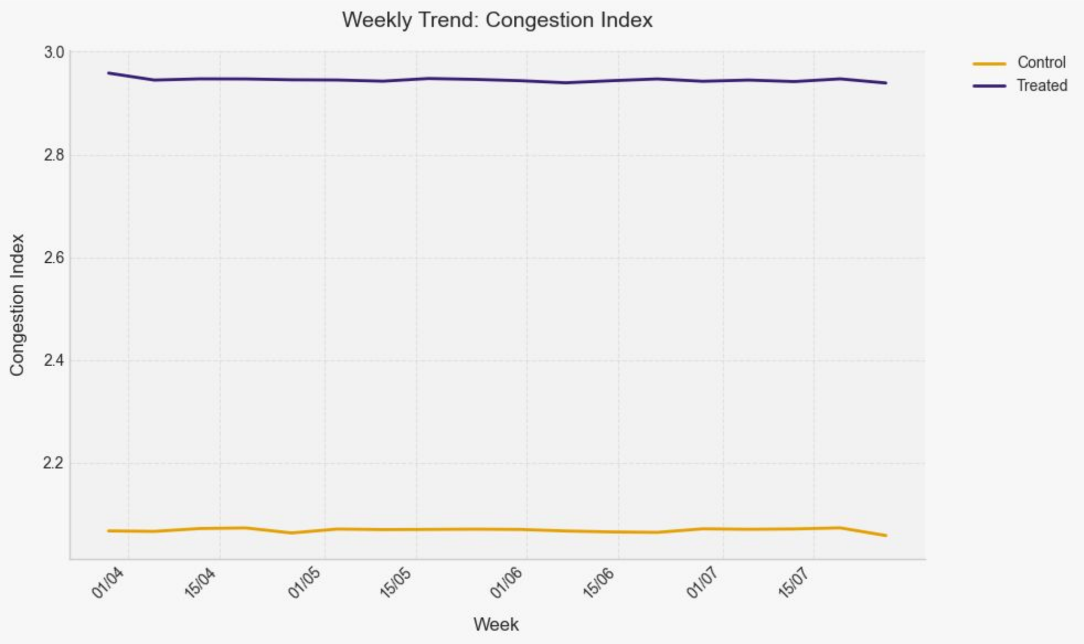
We performed a DiD analysis to estimate the treatment's causal effect on average speed. Results show a significant decrease of around 15,03 units after intervention

Traffic Congestion Analysis

```
weather: 0.04623102782768779
dow: -0.00019306717568966266
speed: 0.0002886777125975107
vehicleType: -0.032088361635219505
carCrash: 101.92108887871935
pedestrianDeath: 108.03215878028597
size: -0.004416150520058948
population: 4.596742785915615e-06
popDensity: -0.0002999905042601328
```

We performed a **regression** to understand how was traffic congestion correlated with other parameters and found out it **varies mainly with car crashes and pedestrian deaths**.

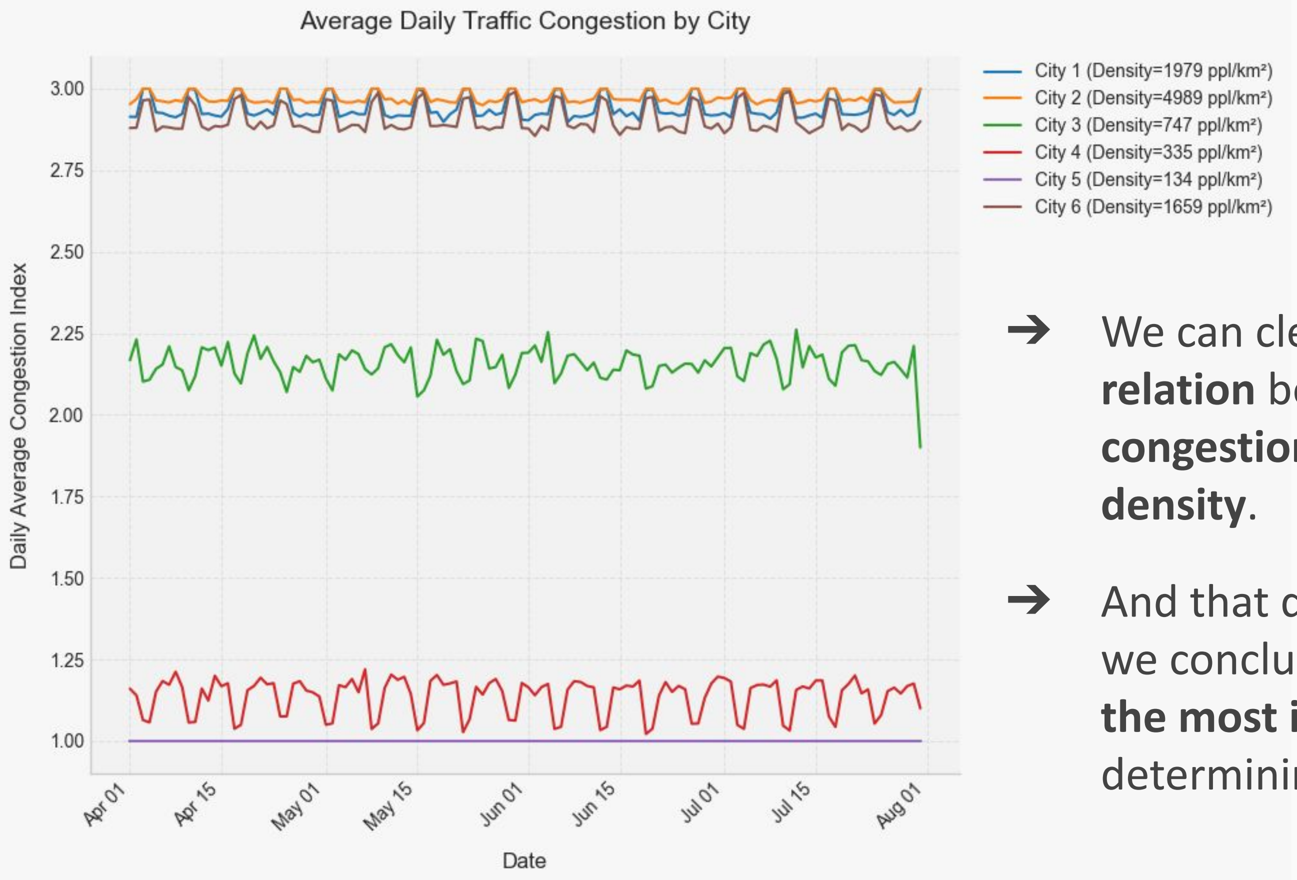
Traffic Congestion Analysis



➔ And even though car crashes and pedestrian deaths were reduced in the implemented group, **traffic remained the same**.

➔ That means that even though they are correlated, it’s not really cause and effect, it is most likely that they are **two effects that share a cause**.

Traffic Congestion Analysis



- ➔ We can clearly see the **relation** between **traffic congestion** and **population density**.
- ➔ And that didn’t change -- so we concluded the **density is the most important factor** in determining congestions.

Energy Consumption – Was it affected by the policy?

Hypothesis:

The policy reduced average speeds in treated areas, causing more stop-and-go driving. This may increase energy consumption, indirectly.

$$
\begin{aligned}
\text{EnergyConsumption} = \beta_0 & \\
& + \beta_1 \cdot \text{treated time} \\
& + \beta_2 \cdot \text{speed} \\
& + \beta_3 \cdot \text{traffic Congestion} \\
& + \beta_4 \cdot \text{vehicleType} \\
& + \beta_5 \cdot \text{weather} \\
& + \beta_6 \cdot \text{carCrash} \\
& + \beta_7 \cdot \text{pedestrianDeath} \\
& + \beta_8 \cdot \text{dow} \\
& + \epsilon
\end{aligned}
$$

Key Results:

- Treated_time: **not significant** ($p = 0.483$)
- Speed: **−0.1612**, **$p < 0.001$** → lower speed → higher energy

Conclusion:

The policy did **not directly** increase energy consumption, but **indirectly** did so by reducing speed — confirming our hypothesis.

Policy implementation effects in traffic congestion

- ➔ When graphing traffic congestion data, no significant change can be seen after the policy implementation.
- ➔ However, this may initially be due to the difference of scale in both cases.

Policy implementation in traffic congestion

➔ After standardizing the data for better visualization at scale, we still cannot see a noticeable change in congestion for the treatment group.

DiD estimation for traffic congestion

- The parameters for estimating traffic congestion are based on the DiD estimate shown below.

$$TrafficCongestion = \alpha + \beta * T_{it} + FE_i + FE_t + e_{it}$$

- Where beta is the diff-in-diff estimator
- The fixed effects in time are weather and day of week, which can have an important impact in the traffic congestion of each day.
- The fixed effects in category is the traffic level in the station, the size and the population density in the city, which was represented with a logarithm since the change in traffic congestion gets slower in big amounts.

DiD estimation for traffic congestion

→ The R-squared got the high value of 0.701 indicating a good estimation of traffic congestion with the variables mentioned.

OLS Regression Results

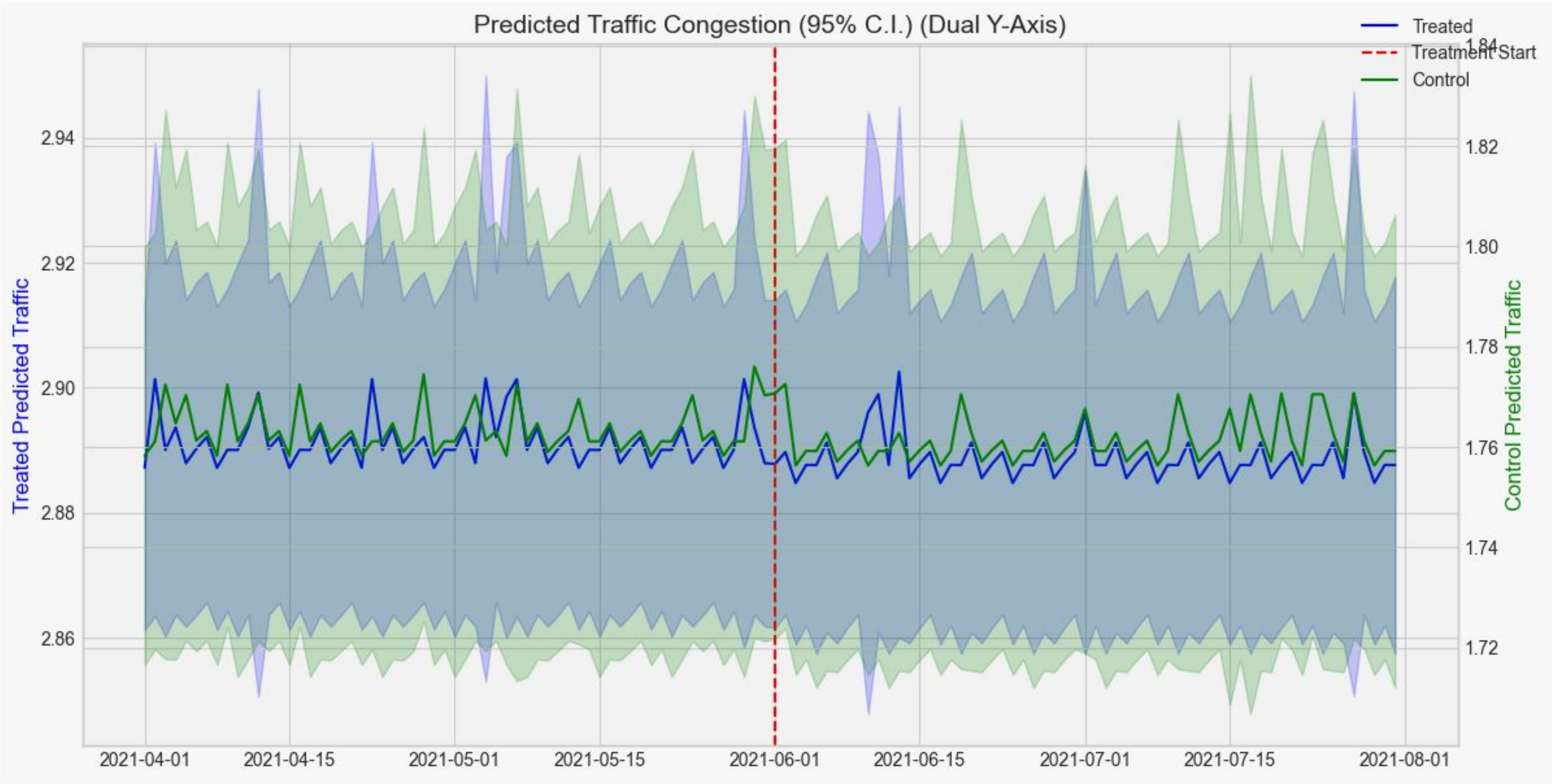
Dep. Variable:	trafficCongestion	R-squared:	0.701
Model:	OLS	Adj. R-squared:	0.701
Method:	Least Squares	F-statistic:	946.3
Date:	Wed, 07 May 2025	Prob (F-statistic):	0.00
Time:	00:27:07	Log-Likelihood:	-2139.8
No. Observations:	5490	AIC:	4316.
Df Residuals:	5472	BIC:	4435.
Df Model:	17		

DiD estimation for traffic congestion

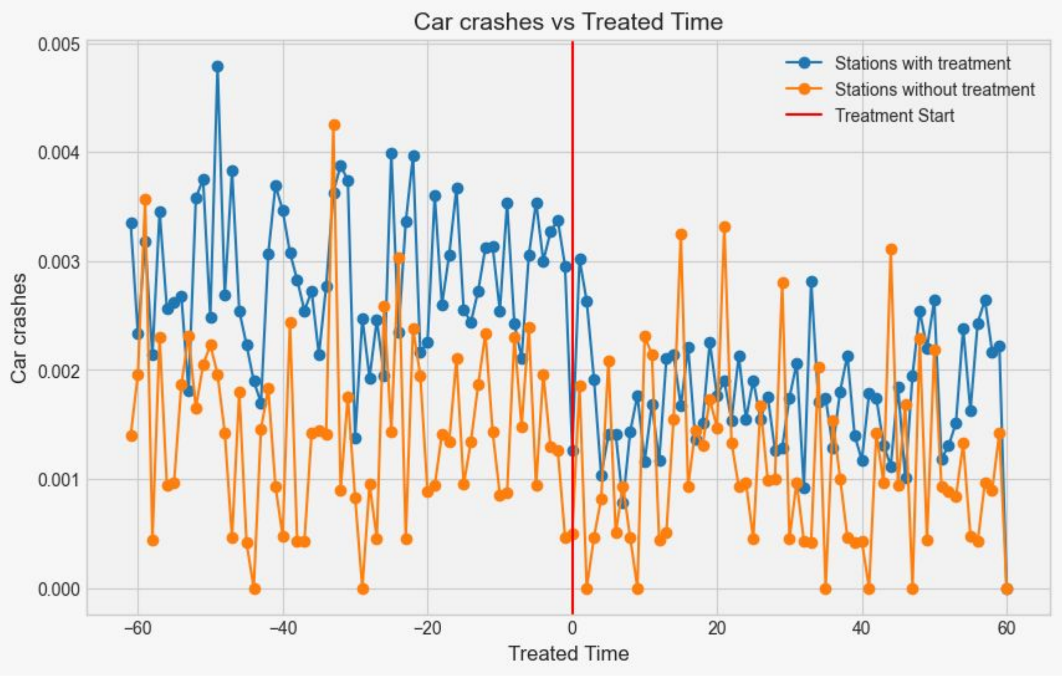
- ➔ The beta (group:post) coefficient has a value of -0.0005.
- ➔ The p-value of beta (group:post) is way higher than 0.005 (0.972), indicating a clear non-effect in traffic congestion due to the new policy implementation.

	coef	std err	z	P>|z|	[0.025	0.975]
group	0.2005	0.009	22.968	0.000	0.183	0.218
group:post[T.True]	-0.0005	0.014	-0.035	0.972	-0.028	0.027
log_popDensity	-0.4488	0.018	-24.302	0.000	-0.485	-0.413
size	-0.0075	0.000	-48.146	0.000	-0.008	-0.007

DiD estimation for traffic congestion

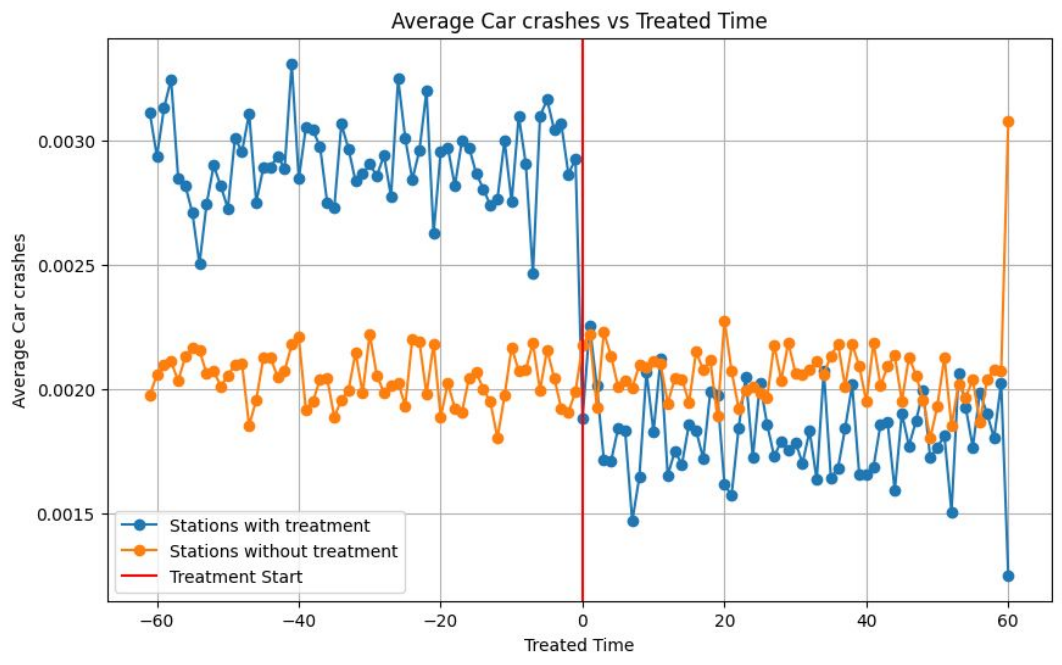


Policy implementation effects in car crashes



➔ After graphing and comparing two records of two stations with different policies, we can already notice a small change after implementation in car crashes frequency.

Policy implementation effects in car crashes



➔ However, after using the general average of both groups, the difference is much more visible.

DiD estimation for car crashes

$$CarCrashes = \alpha + \beta * T_{it} + FE_i + FE_t + e_{it}$$

- ➔ Where beta is the diff-in-diff estimator
- ➔ The fixed effects in time and category are the same as in the traffic congestion analysis.

DiD estimation for car crashes

→ The R-squared got a value of 0.416, indicating a decent estimation of car crashes per day in each station.

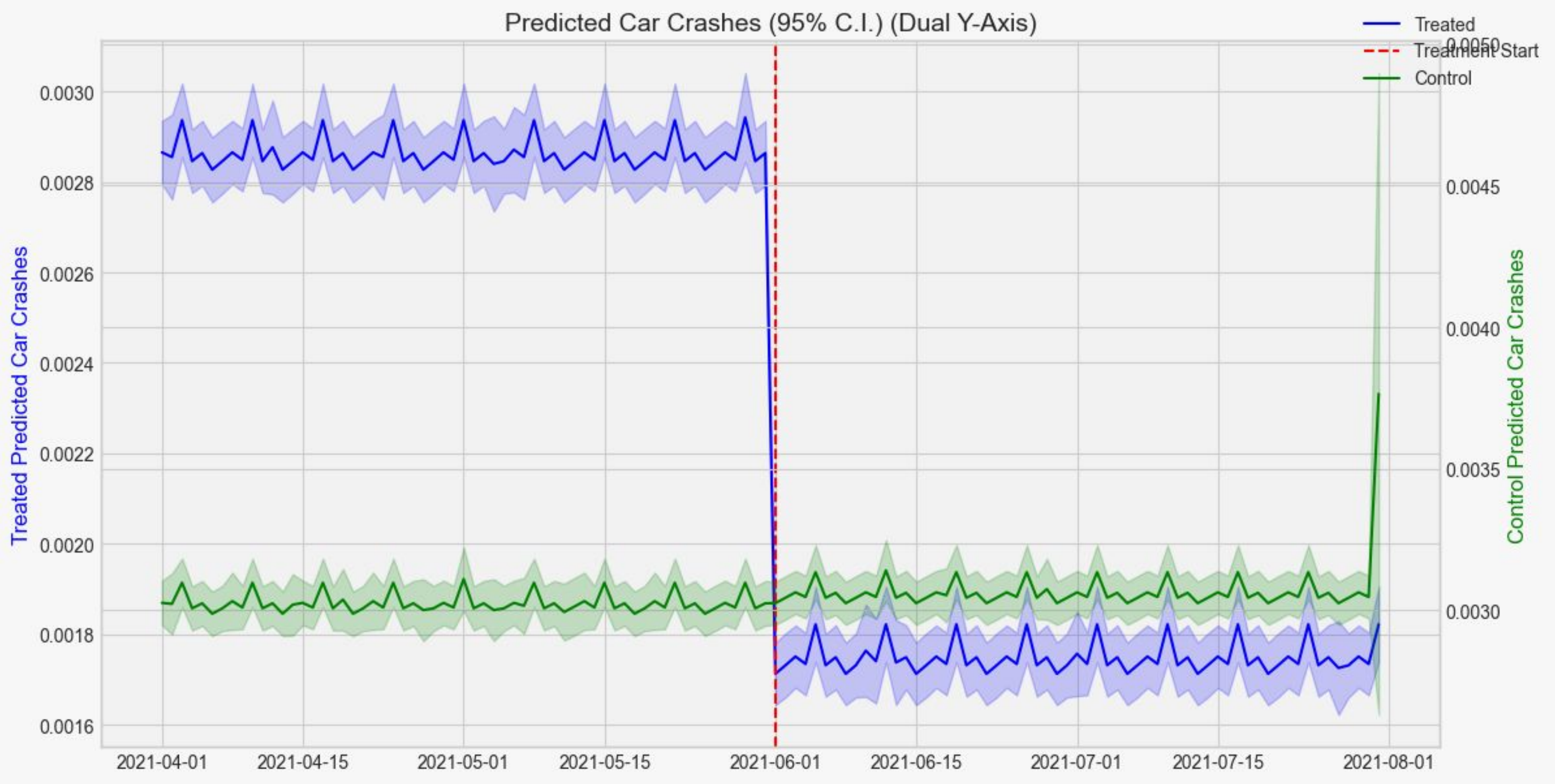
OLS Regression Results			
Dep. Variable:	carCrash	R-squared:	0.416
Model:	OLS	Adj. R-squared:	0.415
Method:	Least Squares	F-statistic:	362.1
Date:	Wed, 07 May 2025	Prob (F-statistic):	0.00
Time:	14:20:23	Log-Likelihood:	40745.
No. Observations:	7320	AIC:	-8.145e+04
Df Residuals:	7302	BIC:	-8.133e+04
Df Model:	17		
Covariance Type:	HC3		

DiD estimation for car crashes

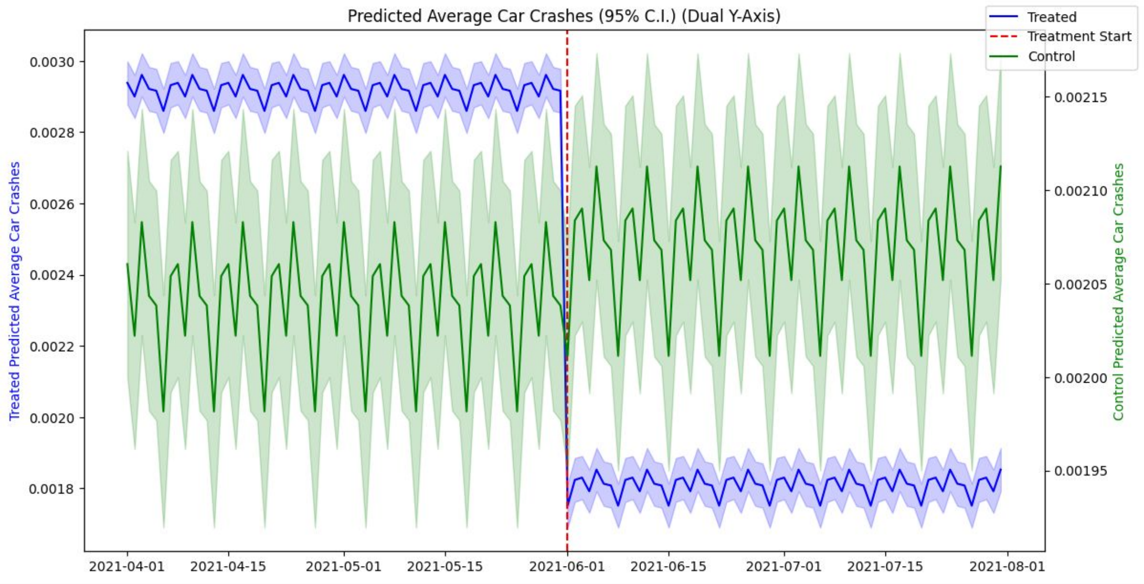
- ➔ The beta (group:post) coefficient has a value close to zero
- ➔ However, the p-value of beta (group:post) is much lower than 0.001 (0.000), indicating a clear effect in car crashes due to the new policy implementation.

	coef	std err	z	P>|z|	[0.025	0.975]
group	0.0003	2.94e-05	9.658	0.000	0.000	0.000
group:post[T.True]	-0.0011	4.11e-05	-27.679	0.000	-0.001	-0.001
size	-2.066e-06	2.55e-07	-8.093	0.000	-2.57e-06	-1.57e-06
log_popDensity	0.0004	2.93e-05	15.025	0.000	0.000	0.000

DiD estimation for car crashes



DiD estimation for car crashes

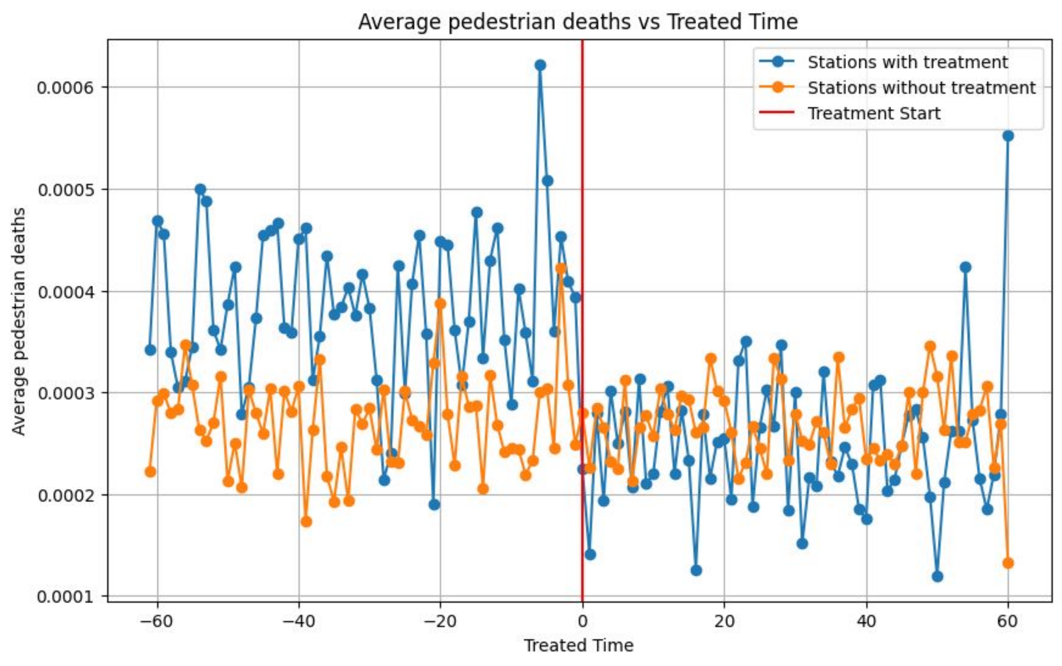


➔ Doing the same procedure with the average of car crashes, we obtain a R-squared of 0.888, showing a good estimation.

Policy implementation effects in pedestrian deaths

- ➔ Graphing the pedestrian deaths in each station, we can see that there's a lot of days where there weren't any fatal accident.
- ➔ This limit could generate troubles with the expected value of error.

Policy implementation effects in pedestrian deaths



- ➔ That's why the average is more helpful in the analysis, since there are less constraints.
- ➔ We can notice a decrease in pedestrian deaths after the implementation of policy.

DiD estimation for pedestrian deaths

$$AveragePedestrianDeaths = \alpha + \beta * T_{it} + FE_t + e_{it}$$

- ➔ Where beta is the diff-in-diff estimator
- ➔ The fixed effects in time are the same as in the traffic congestion and car crashes analysis.
- ➔ There are not fixed effects for category since they are all mixed in the average.

DiD estimation for pedestrian deaths

➔ The R-squared got a value of 0.462, indicating a pretty decent estimation of pedestrian deaths per day in average.

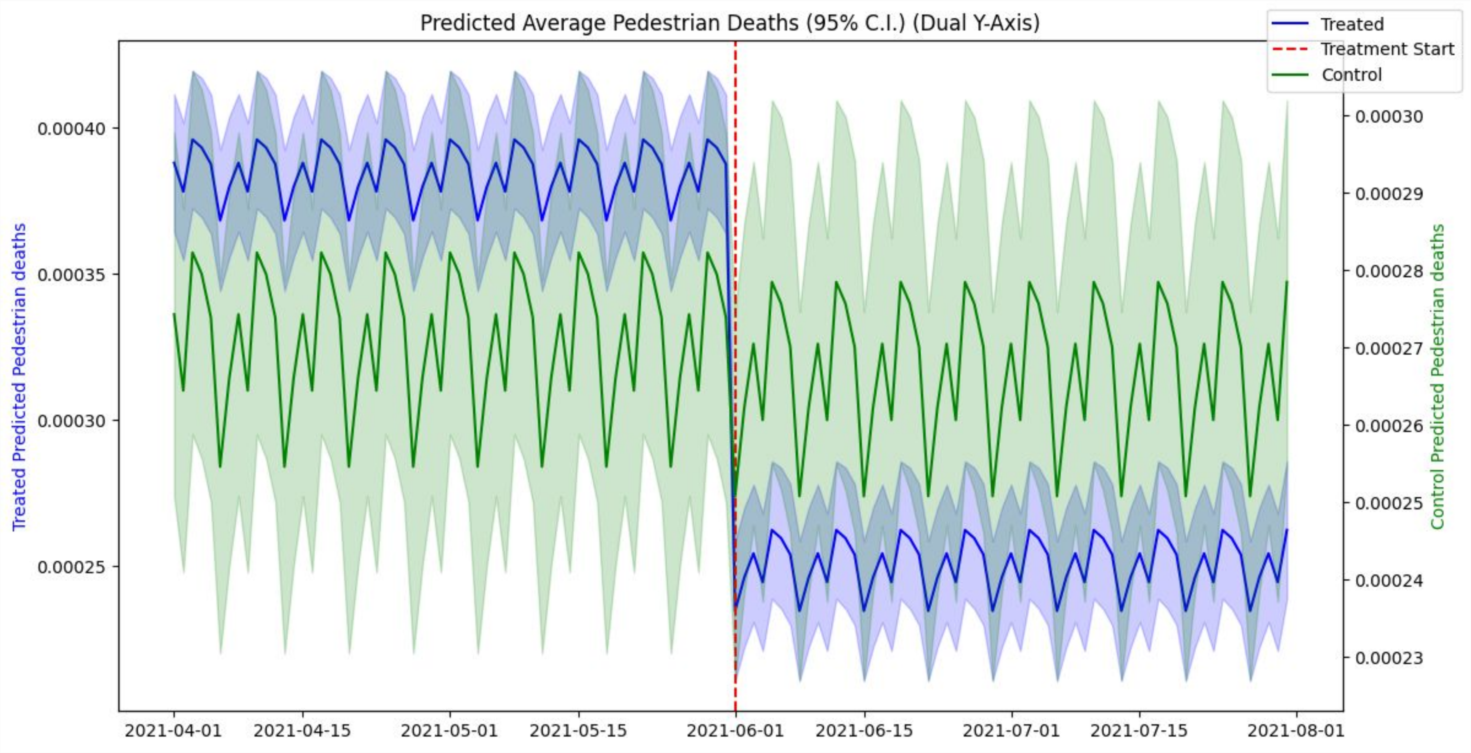
```
                            OLS Regression Results
==============================================================================
Dep. Variable:        pedestrianDeath   R-squared:                       0.462
Model:                            OLS   Adj. R-squared:                  0.442
Method:                 Least Squares   F-statistic:                     22.35
Date:                Wed, 07 May 2025   Prob (F-statistic):           3.32e-27
Time:                        14:41:07   Log-Likelihood:                 2032.5
No. Observations:                 244   AIC:                            -4045.
Df Residuals:                     234   BIC:                            -4010.
Df Model:                           9
Covariance Type:            nonrobust
```

DiD estimation for pedestrian deaths

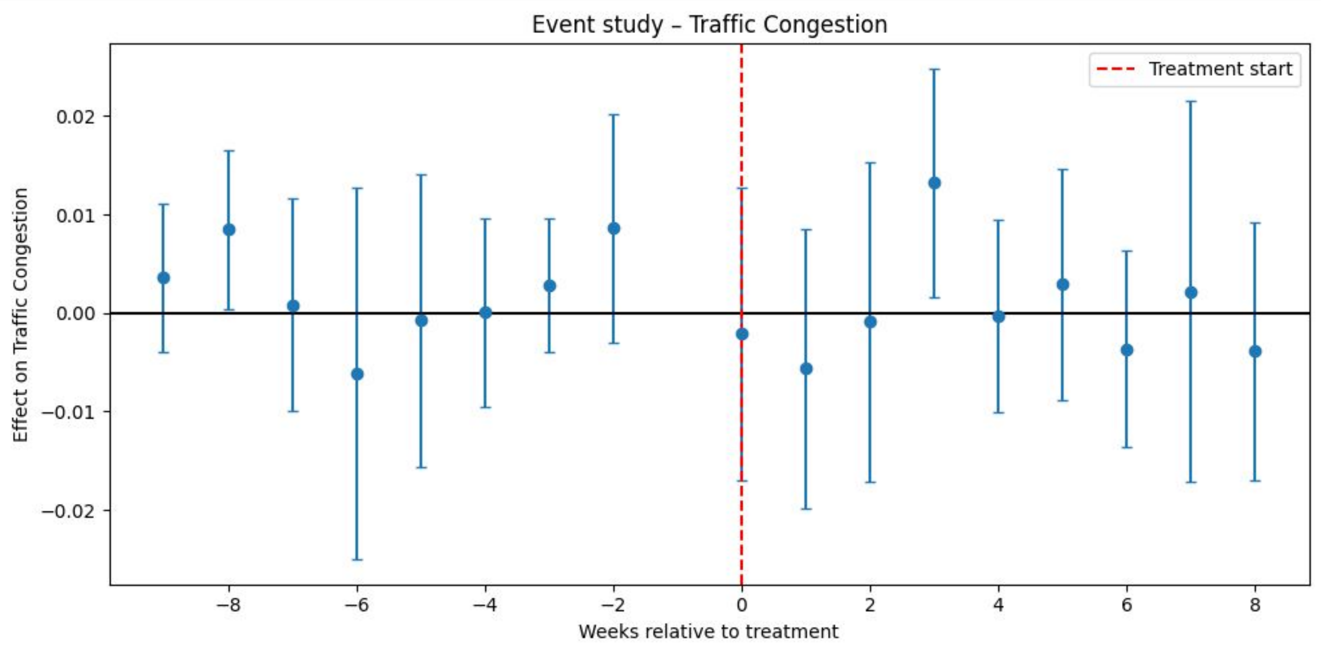
- ➔ The beta (group:post) coefficient has a value close to zero
- ➔ However, the p-value of beta (group:post) is much lower than 0.001 (0.000), indicating a clear effect in pedestrian deaths due to the new policy implementation.

	coef	std err	z	P>|z|	[0.025	0.975]
group	0.0001	1.08e-05	10.539	0.000	9.24e-05	0.000
group:post[T.True]	-0.0001	1.53e-05	-8.510	0.000	-0.000	-9.98e-05

DiD estimation for pedestrian deaths

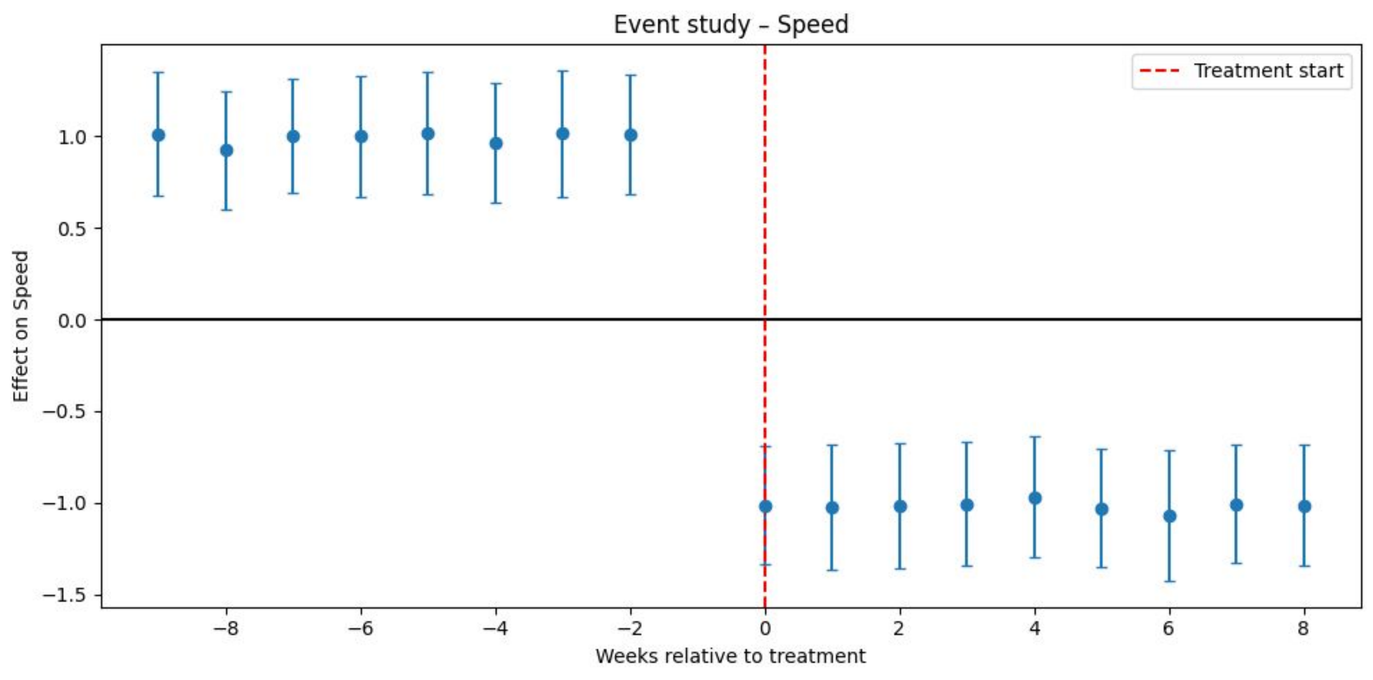


Traffic Congestion



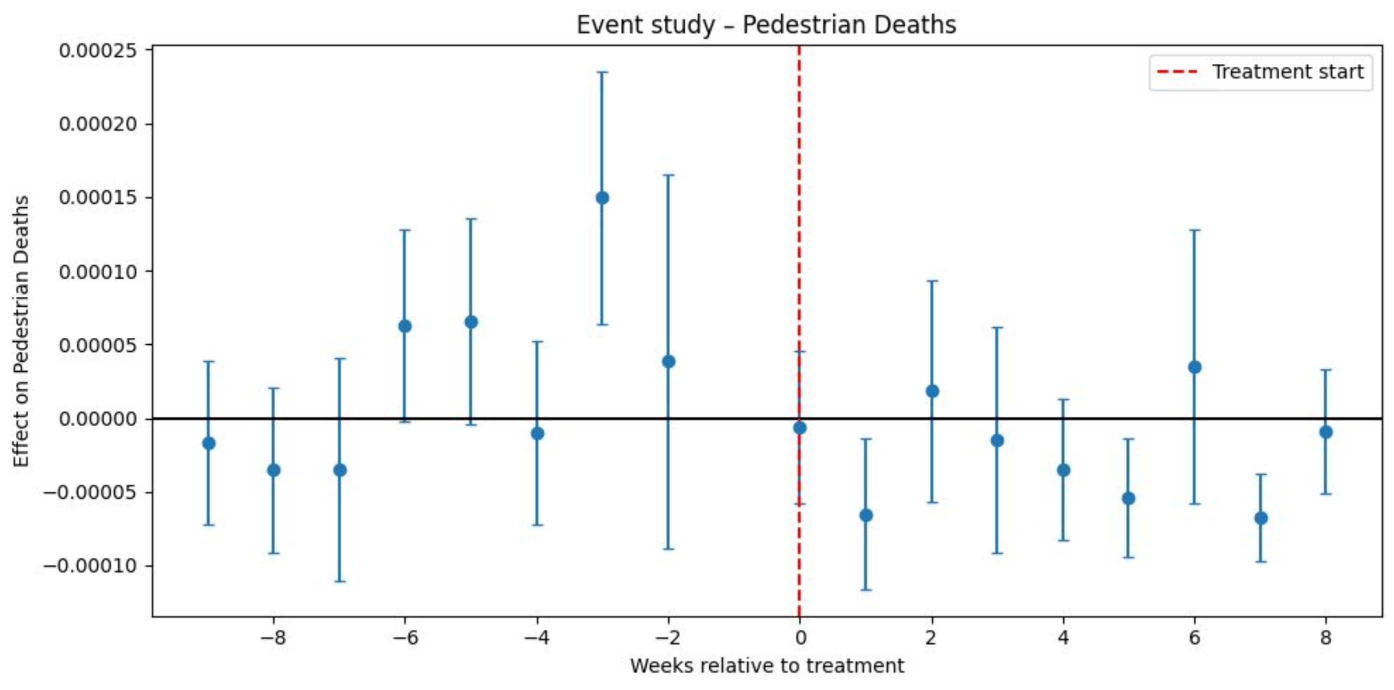
- Pre-treatment coefficients ≈ 0, thus parallel trends hold.
- Post-treatment effects remain indistinguishable from zero.
- Policy shows **no detectable impact** on congestion.

Speed



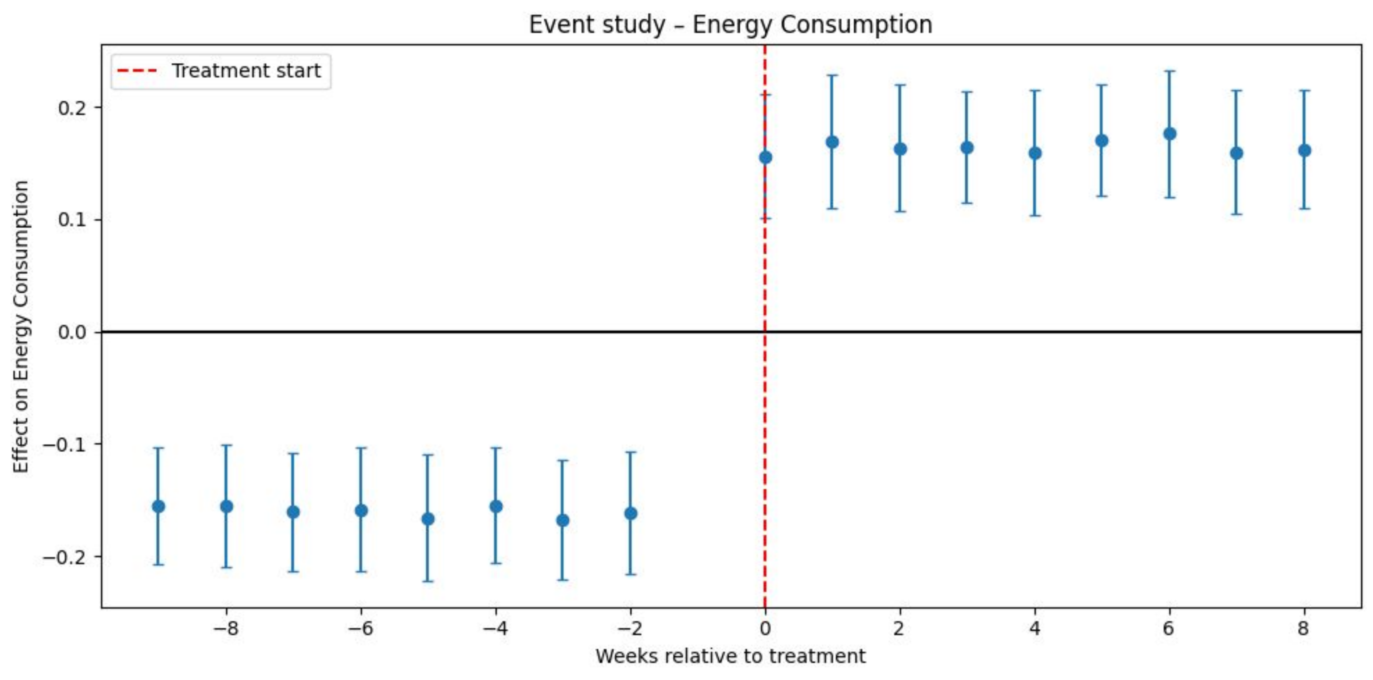
- Treated sites faster than control before intervention (violation).
- Speeds drop below control after week 0.
- Apparent effect, **but causal claim weak** due to non-parallel pre-trends.

Pedestrian Deaths



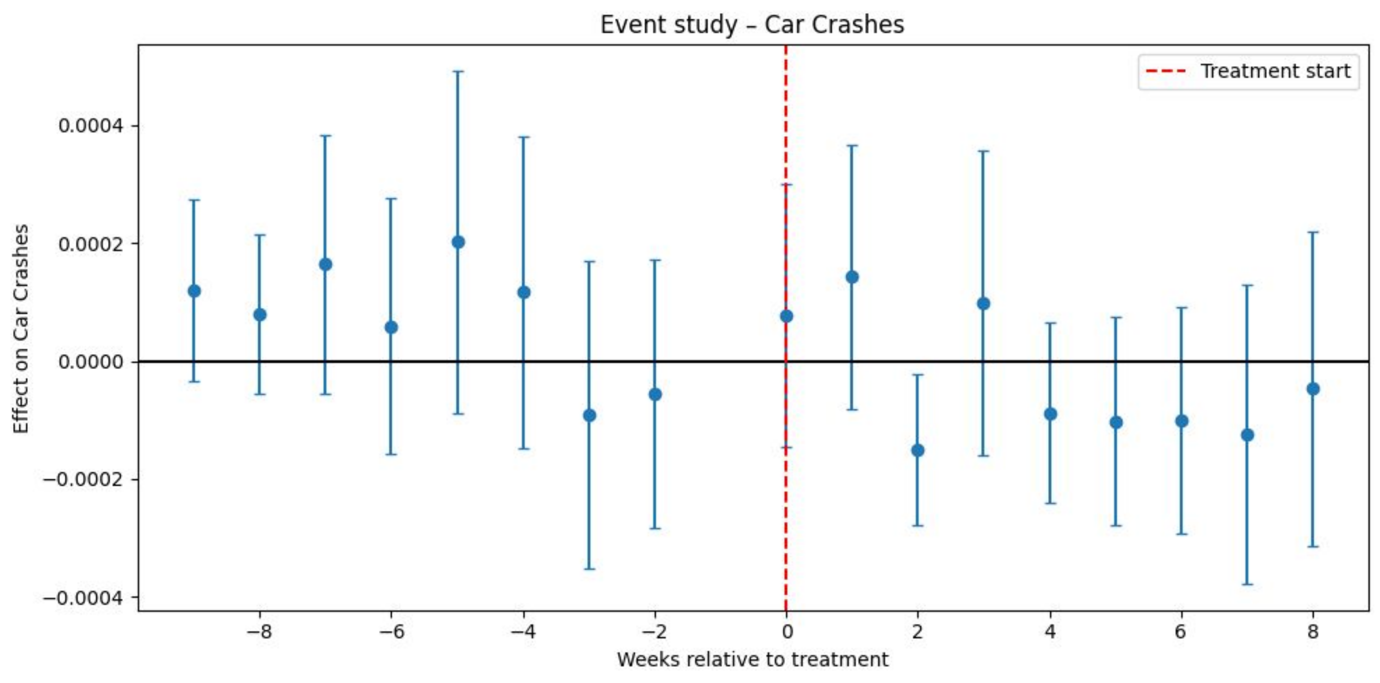
- Pre-trend flat; coefficients flip from + to – post-treatment.
- Many post-treatment points significantly < 0.
- Policy cuts pedestrian fatalities.

Energy Consumption



- Stable negative coefficients pre-treatment, stable positive after.
- All post-points significantly > 0, clear pattern.
- Policy increases energy use (≈ +0.15 kWh).

Car Crashes



- Small positive pre-period uptick, but generally flat.
- Sharp negative shift after week 0; several CIs below zero.
- Significant reduction in crashes attributable to policy.

Key Insights

- The speed limit policy implemented on June 21st, impacted key traffic metrics in treated stations.
- While there was no significant effect on traffic congestion, clear changes were observed in safety-related outcomes.

According to the Differences-in-Differences analysis:

- **Car crashes** and **pedestrian deaths** significantly decreased after the policy, with strong statistical evidence.
- **Energy consumption** slightly increased, indirectly, due to reduced speeds causing less efficient driving patterns.
- **Traffic congestion** remained unchanged, likely driven more by population density than by the policy itself.

THANK YOU!