Arizona Crime Data

Benjamin Henning

Dataset – FBI Crime Data (1985-2019)

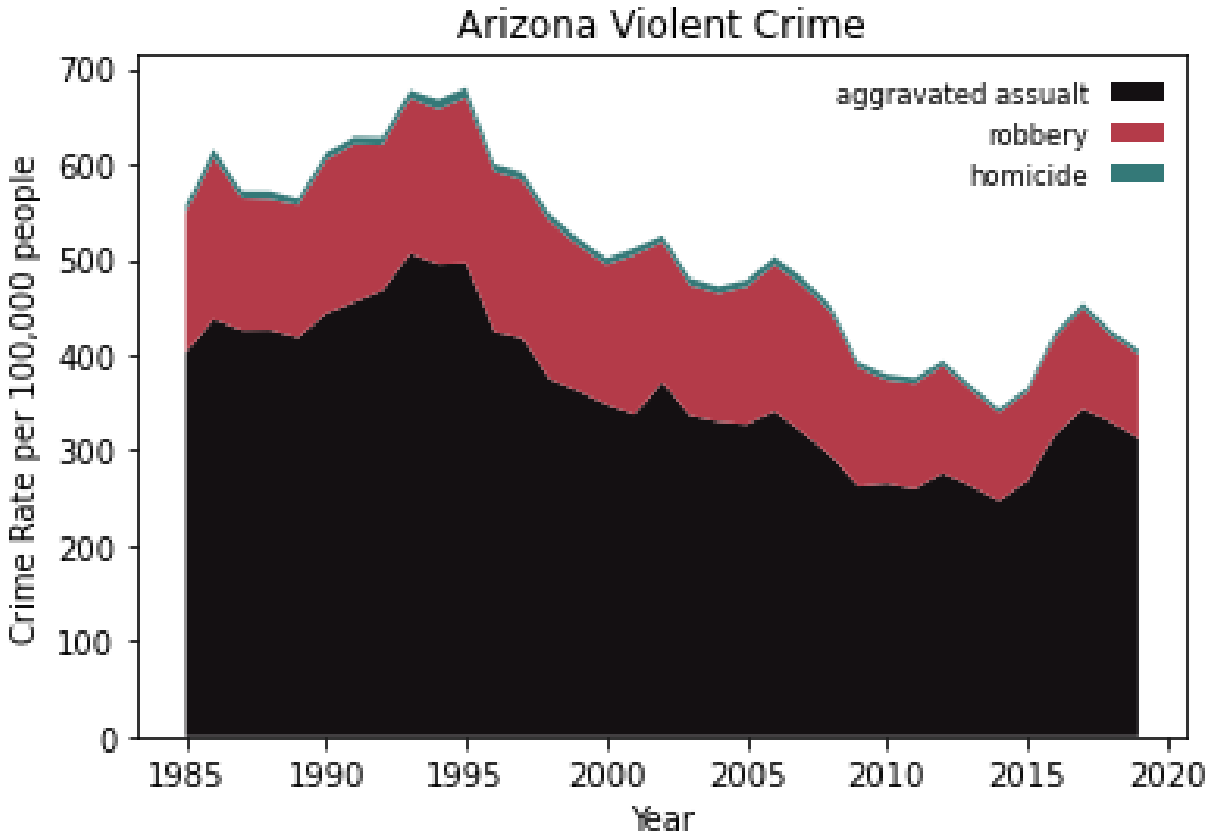
https://crime-data-explorer.fr.cloud.gov/explorer/state/arizona/crime

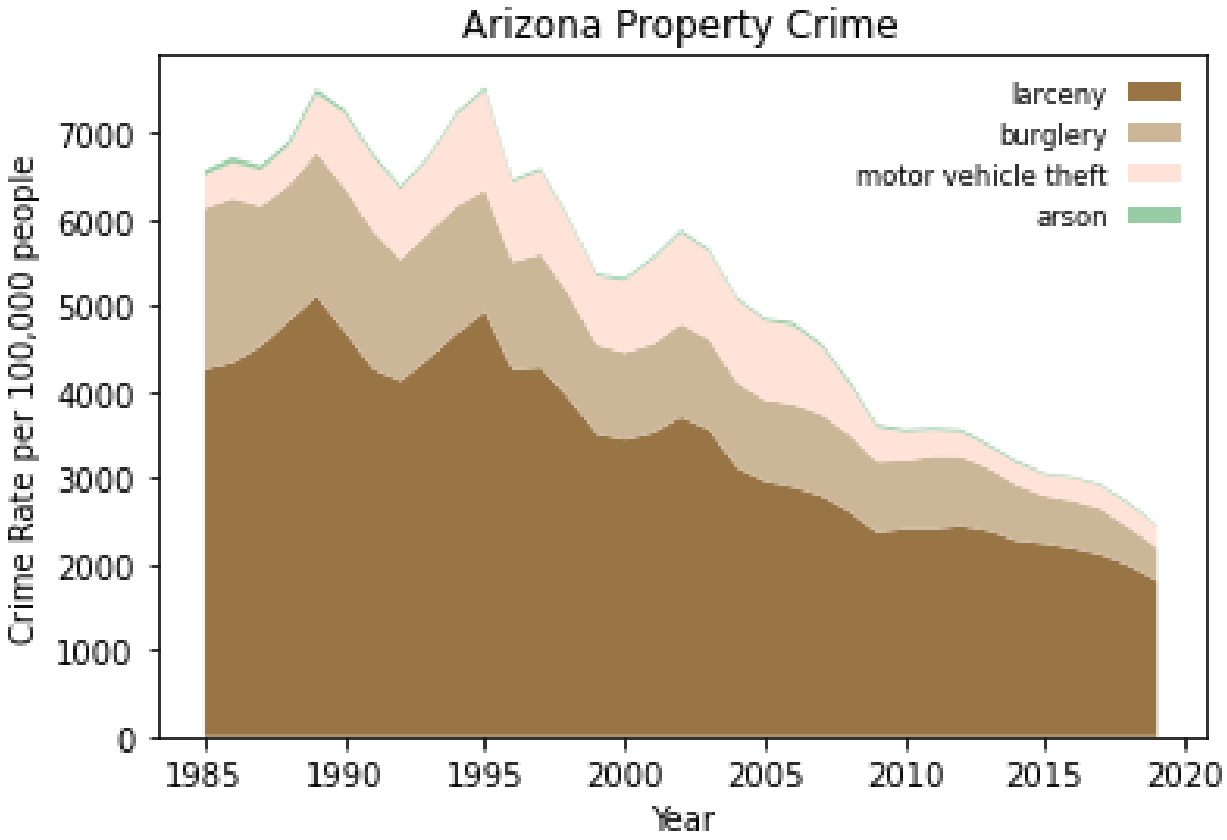
total_property.csv

	A	B	C	D
1	Year	Rate	Location	
2	2015	2500.5	United States	
3	1985	4666.4	United States	
4	2000	3618.3	United States	
5	1997	4316.5	United States	
6	1998	4052.5	United States	
7	1999	3743.6	United States	
8	2001	3658.1	United States	
9	2002	3630.6	United States	
10	2003	3591.2	United States	
11	2004	3514.1	United States	
12	2005	3431.5	United States	
13	2006	3346.6	United States	
14	2007	3276.4	United States	
15	2008	3214.6	United States	
16	2009	3041.3	United States	

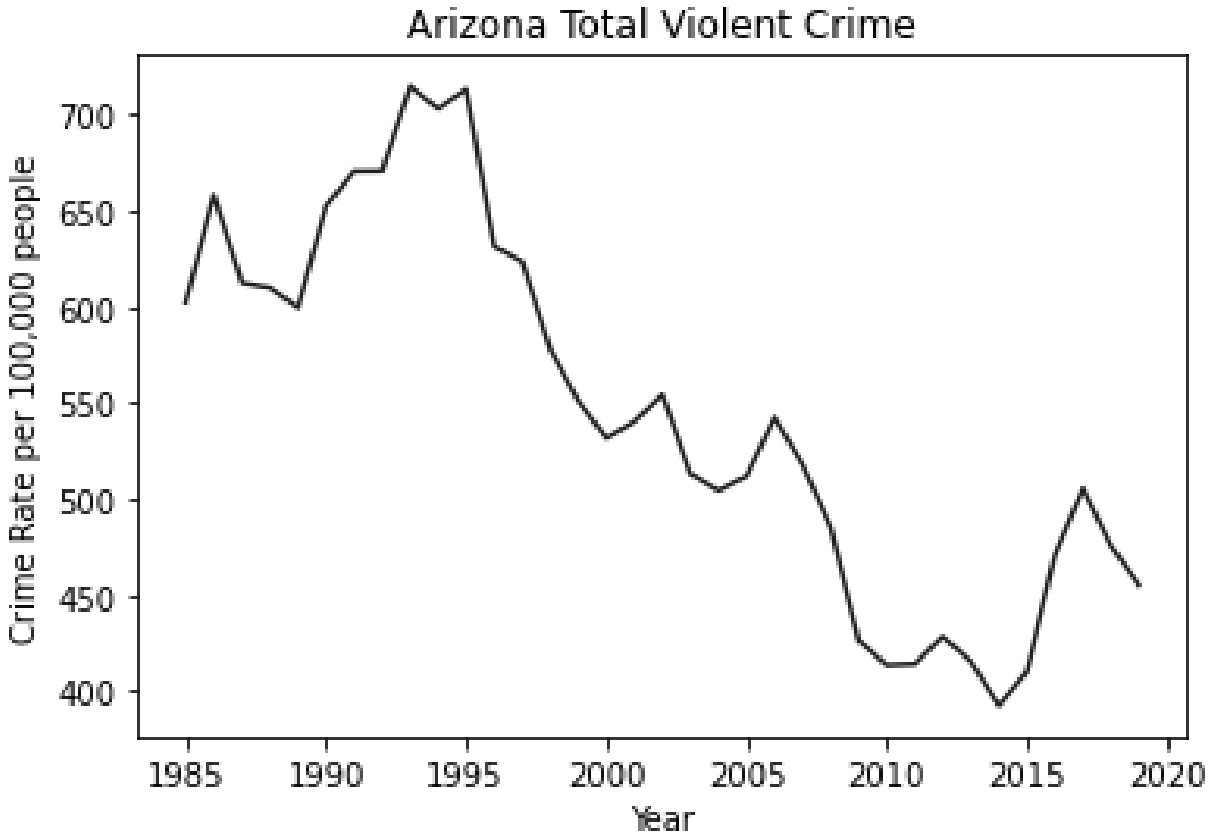
```
========== Total Property Crime Rates ==========
        Rate       Location
Year
1985  4666.4  United States
1985  6513.7        Arizona
1986  6663.1        Arizona
1986  4881.8  United States
1987  6576.0        Arizona
1987  4963.0  United States
1988  6861.2        Arizona
1988  5054.0  United States
1989  7460.2        Arizona
1989  5107.1  United States
1990  7236.4        Arizona
1990  5073.1  United States
1991  6734.9        Arizona
1991  5140.2  United States
1992  4903.7  United States
1992  6357.8        Arizona
1993  4740.0  United States
1993  6716.7        Arizona
1994  7221.4        Arizona
1994  4660.2  United States
1995  4590.5  United States
1995  7500.1        Arizona
1996  6435.5        Arizona
1996  4451.0  United States
1997  6571.3        Arizona
1997  4316.5  United States
1998  5997.0        Arizona
```

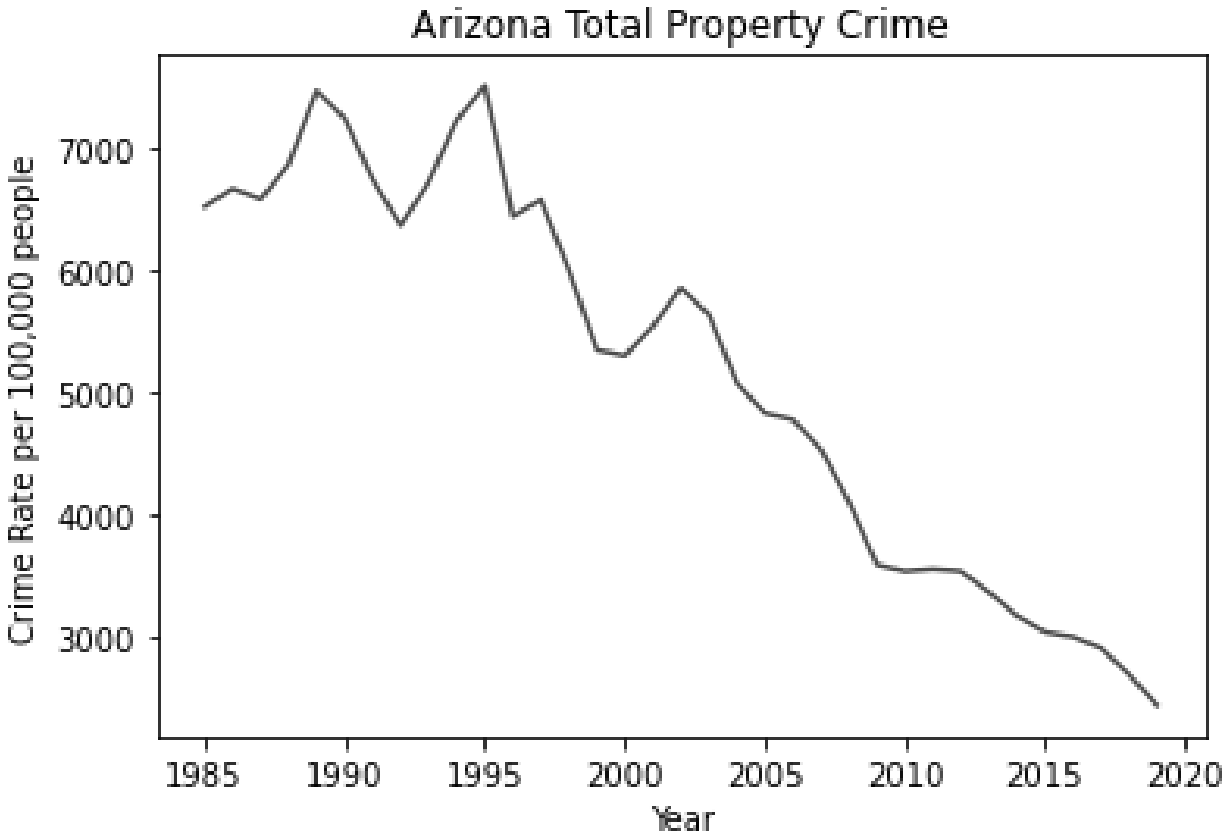
Arizona Crime Data (1985-2019) – Stackplot Visualization



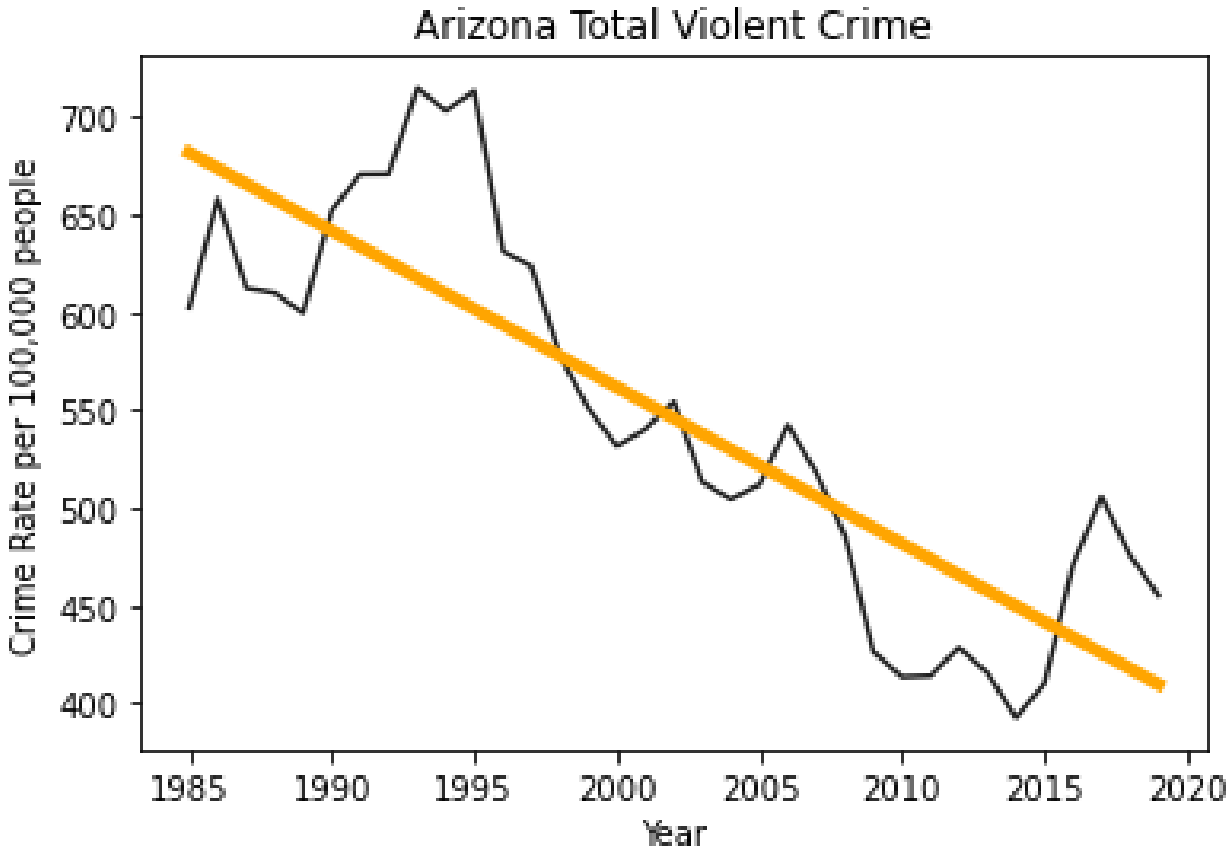


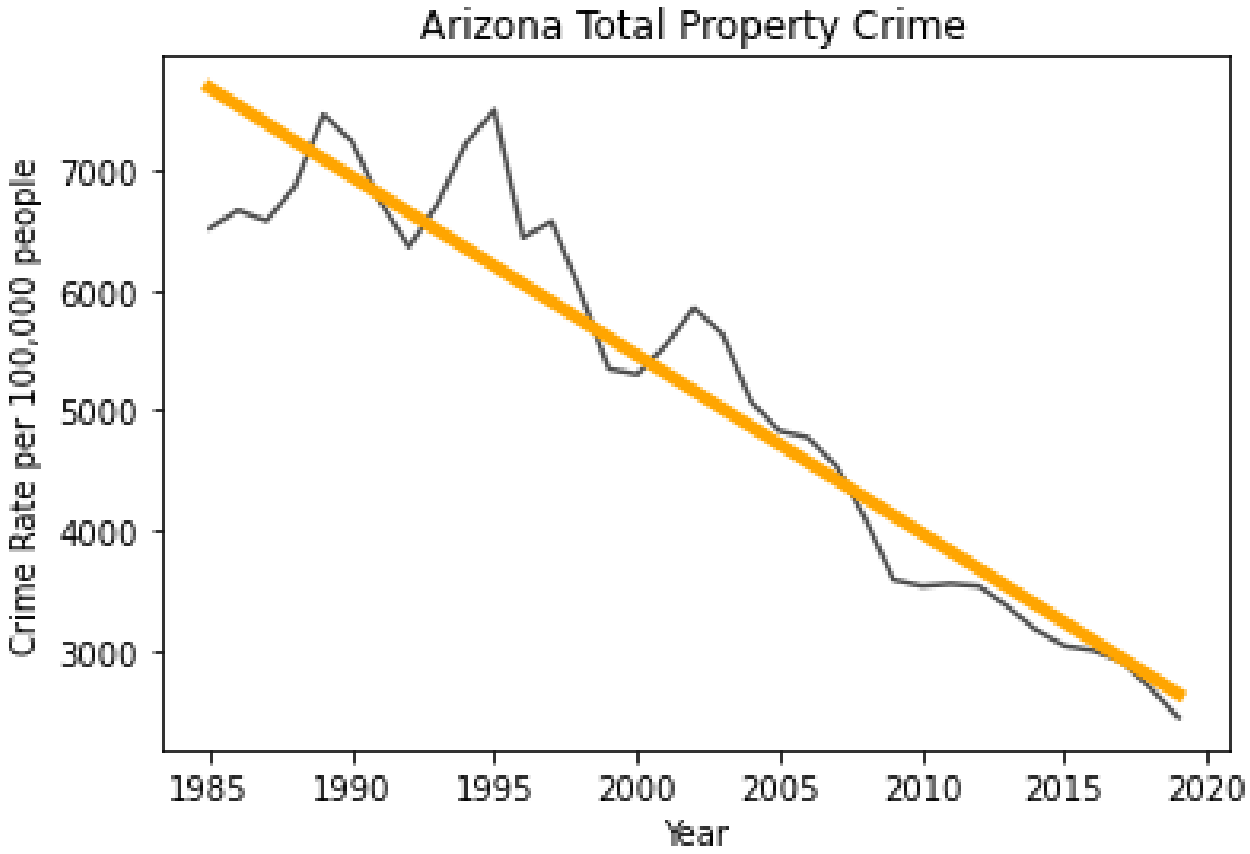
Arizona Crime Data (1985-2019) - Totals



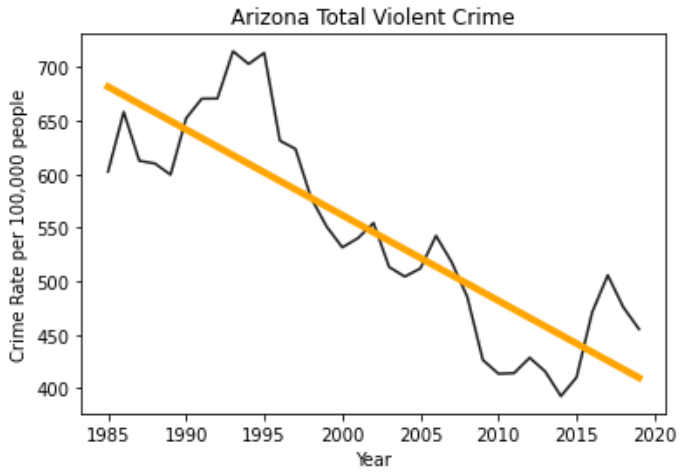


Model 1 – Simple Linear Regression using OLS





Model 1 – Simple Linear Regression using OLS

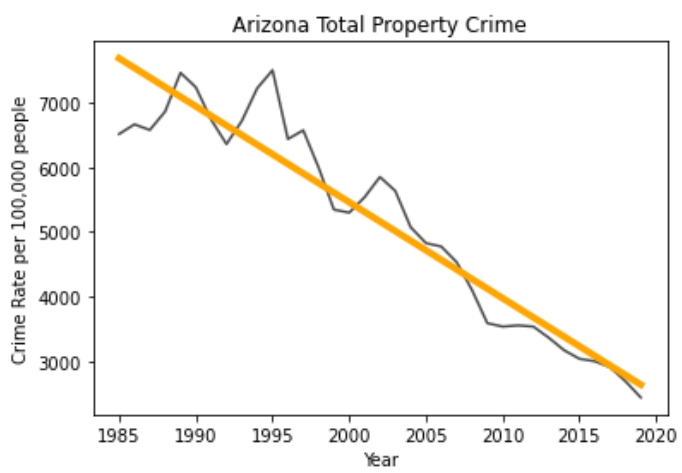


OLS Regression Results

Dep. Variable:	y	R-squared:	0.722
Model:	OLS	Adj. R-squared:	0.714
Method:	Least Squares	F-statistic:	85.72
Date:	Sun, 06 Dec 2020	Prob (F-statistic):	1.07e-10
Time:	06:02:02	Log-Likelihood:	-186.67
No. Observations:	35	AIC:	377.3
Df Residuals:	33	BIC:	380.5
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.656e+04	1729.775	9.574	0.000	1.3e+04	2.01e+04
x1	-7.9994	0.864	-9.258	0.000	-9.757	-6.242

Omnibus:	2.298	Durbin-Watson:	0.412
Prob(Omnibus):	0.317	Jarque-Bera (JB):	2.029
Skew:	0.489	Prob(JB):	0.363
Kurtosis:	2.341	Cond. No.	3.97e+05



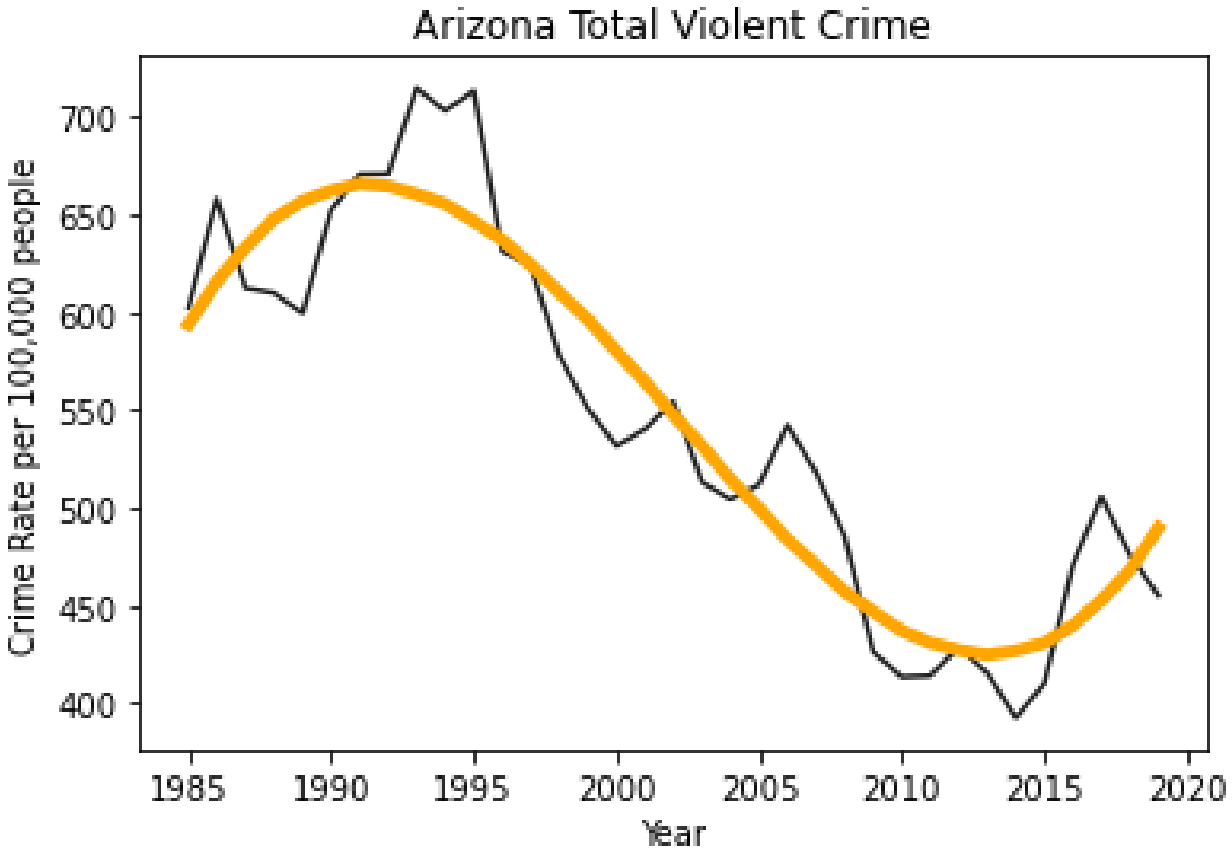
OLS Regression Results

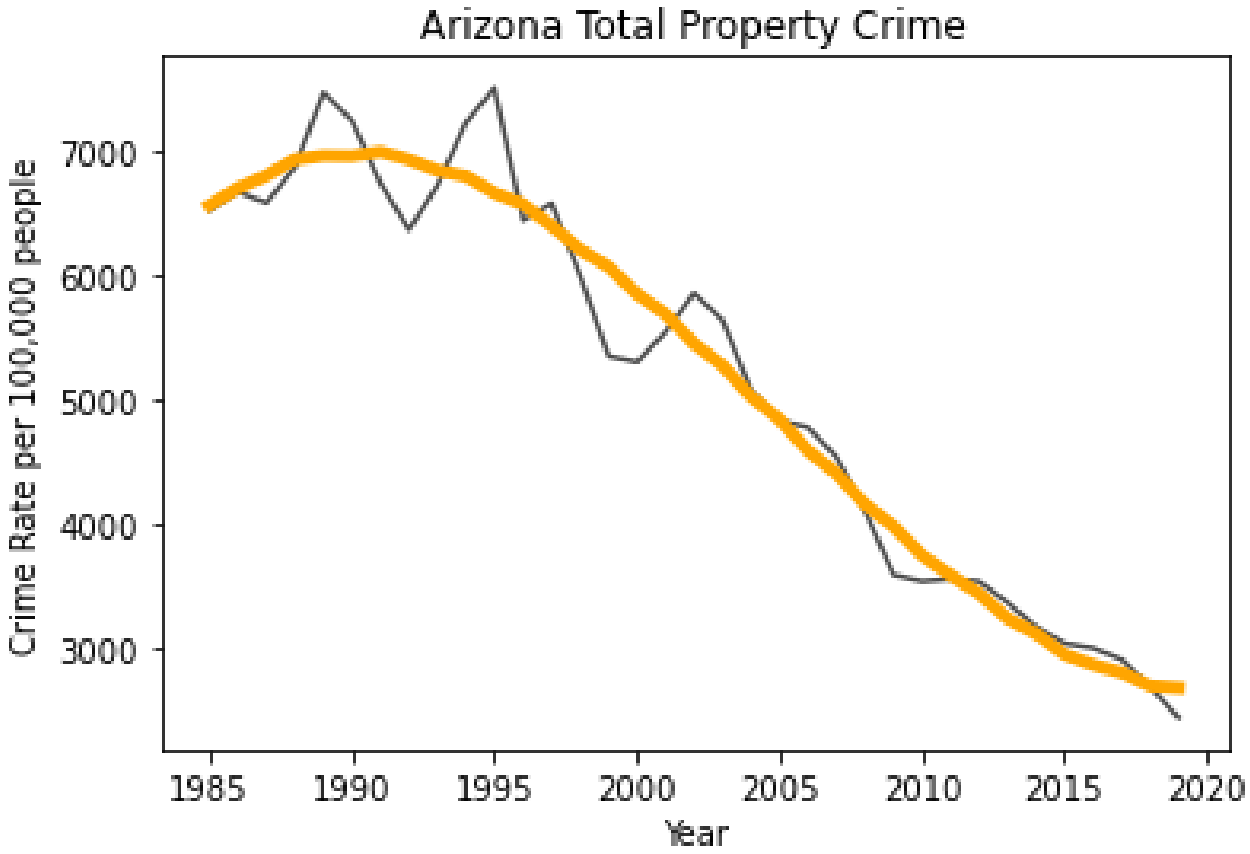
Dep. Variable:	y	R-squared:	0.905
Model:	OLS	Adj. R-squared:	0.902
Method:	Least Squares	F-statistic:	313.5
Date:	Sun, 06 Dec 2020	Prob (F-statistic):	2.04e-18
Time:	06:02:02	Log-Likelihood:	-266.19
No. Observations:	35	AIC:	536.4
Df Residuals:	33	BIC:	539.5
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.022e+05	1.68e+04	18.013	0.000	2.68e+05	3.36e+05
x1	-148.3671	8.380	-17.706	0.000	-165.415	-131.319

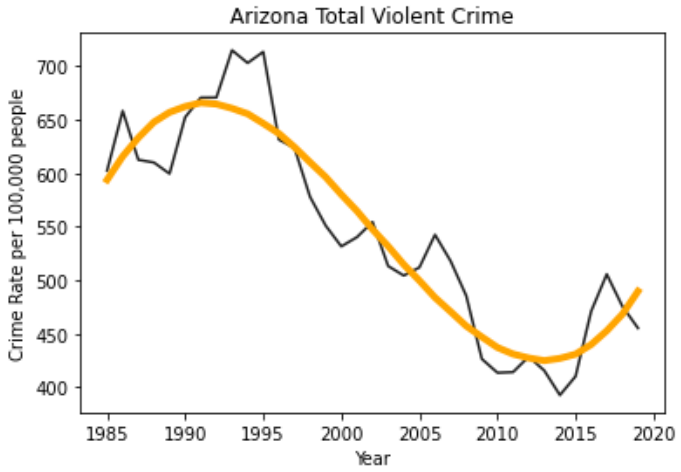
Omnibus:	1.804	Durbin-Watson:	0.498
Prob(Omnibus):	0.406	Jarque-Bera (JB):	0.787
Skew:	0.180	Prob(JB):	0.675
Kurtosis:	3.641	Cond. No.	3.97e+05

Model 2 – 5th Degree Polynomial Regression using OLS





Model 2 – 5th Degree Polynomial Regression using OLS

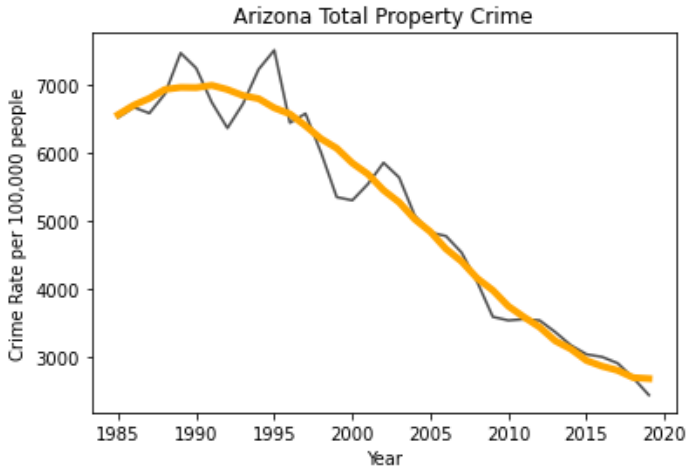


OLS Regression Results

Dep. Variable:	y	R-squared:	0.880
Model:	OLS	Adj. R-squared:	0.864
Method:	Least Squares	F-statistic:	54.85
Date:	Sun, 06 Dec 2020	Prob (F-statistic):	2.27e-13
Time:	06:21:12	Log-Likelihood:	-172.01
No. Observations:	35	AIC:	354.0
Df Residuals:	30	BIC:	361.8
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	2.645e+07	4.25e+06	6.227	0.000	1.78e+07	3.51e+07
x1	5.631e+05	9e+04	6.256	0.000	3.79e+05	7.47e+05
x2	-281.2715	44.965	-6.255	0.000	-373.101	-189.442
x3	0.0468	0.007	6.255	0.000	0.032	0.062
x4	-3.654e-10	5.2e-09	-0.070	0.944	-1.1e-08	1.03e-08

Omnibus:	2.451	Durbin-Watson:	0.883
Prob(Omnibus):	0.294	Jarque-Bera (JB):	1.876
Skew:	0.402	Prob(JB):	0.391
Kurtosis:	2.201	Cond. No.	8.25e+14



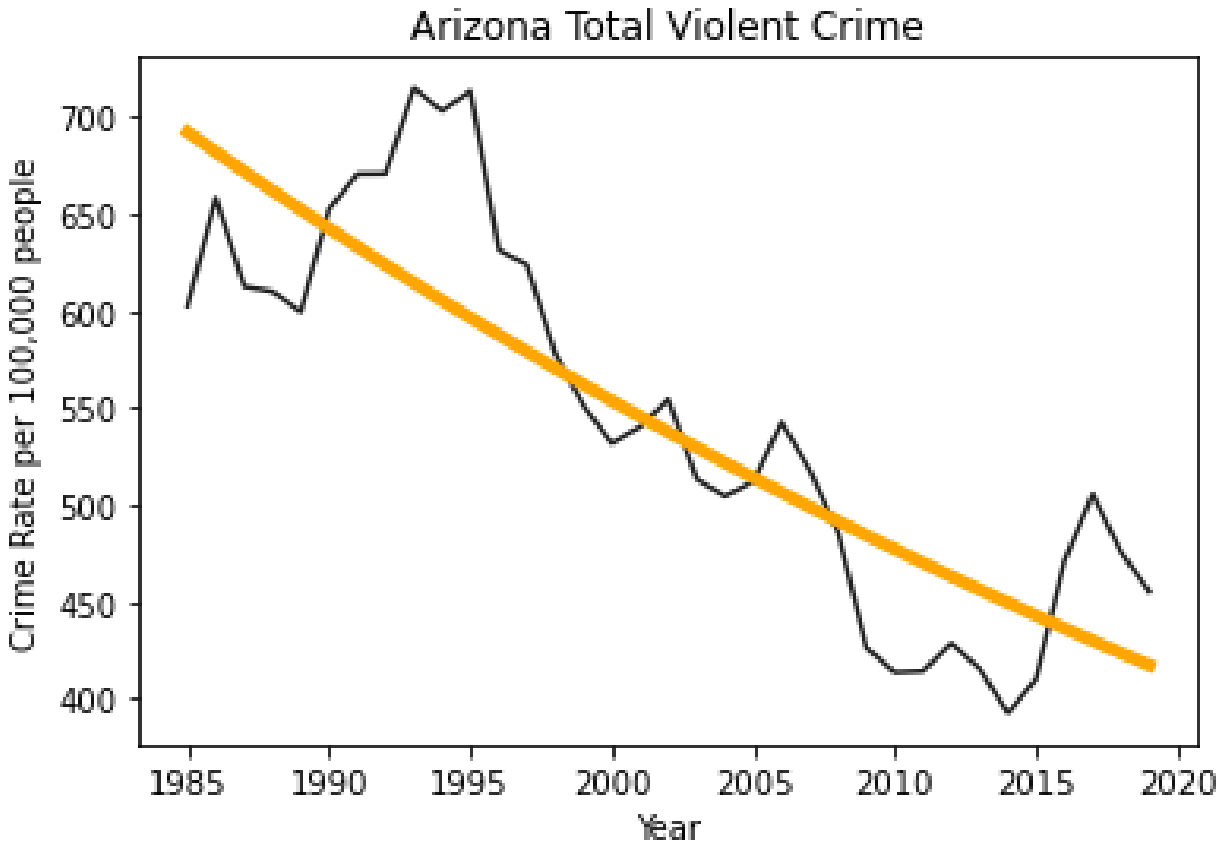
OLS Regression Results

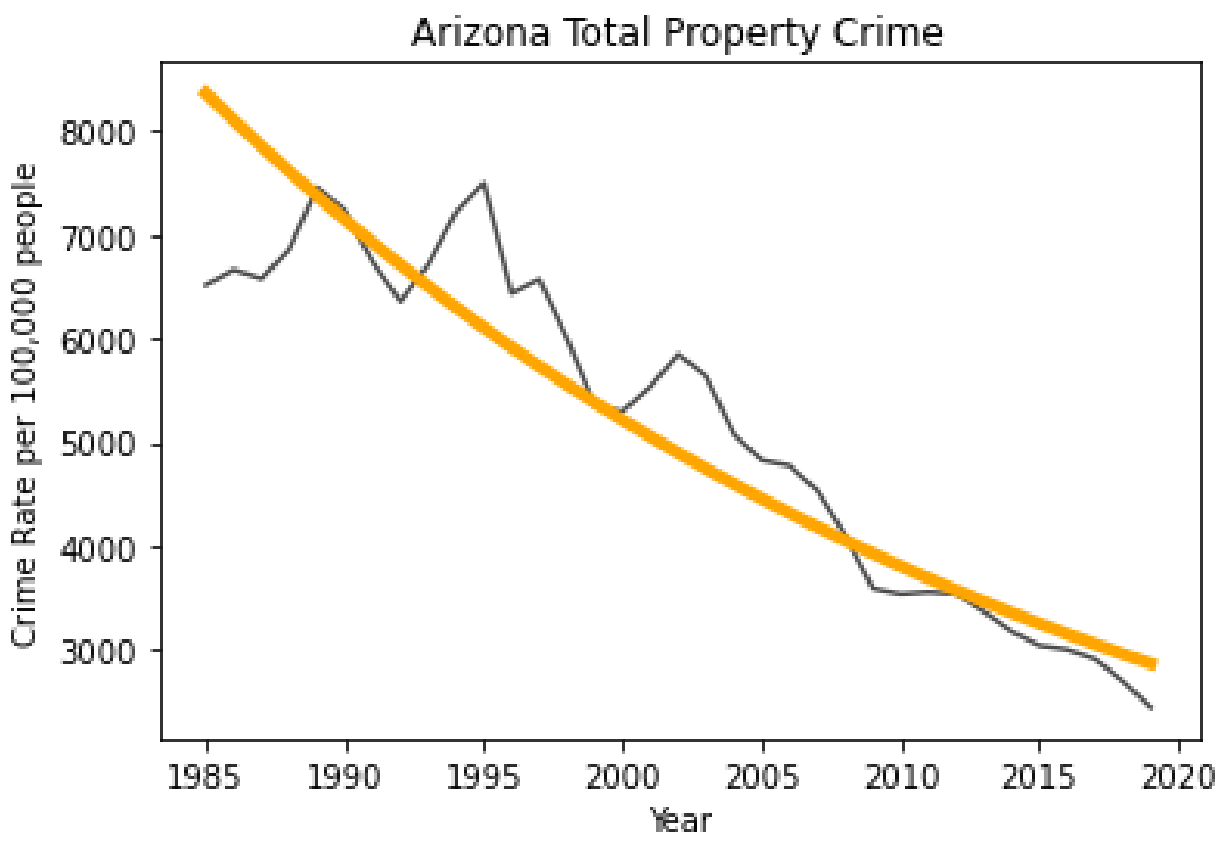
Dep. Variable:	y	R-squared:	0.962
Model:	OLS	Adj. R-squared:	0.957
Method:	Least Squares	F-statistic:	190.4
Date:	Sun, 06 Dec 2020	Prob (F-statistic):	7.39e-21
Time:	06:21:12	Log-Likelihood:	-250.07
No. Observations:	35	AIC:	510.1
Df Residuals:	30	BIC:	517.9
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.754e+08	3.95e+07	4.439	0.000	9.47e+07	2.56e+08
x1	3.98e+06	8.37e+05	4.753	0.000	2.27e+06	5.69e+06
x2	-1984.9769	418.222	-4.746	0.000	-2839.101	-1130.853
x3	0.3300	0.070	4.739	0.000	0.188	0.472
x4	-1.567e-08	4.84e-08	-0.324	0.748	-1.15e-07	8.32e-08

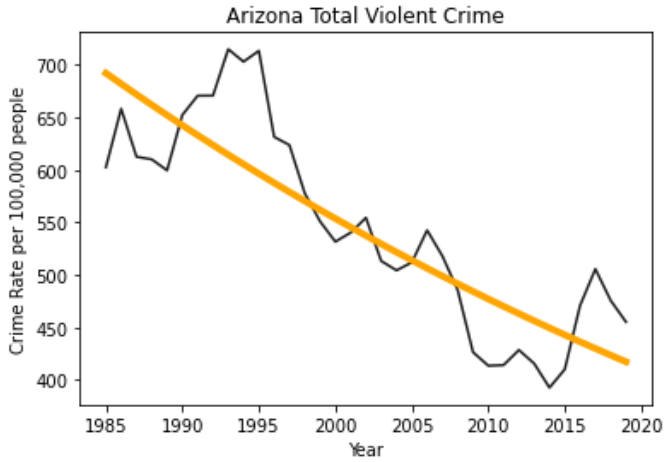
Omnibus:	1.941	Durbin-Watson:	1.184
Prob(Omnibus):	0.379	Jarque-Bera (JB):	0.902
Skew:	0.125	Prob(JB):	0.637
Kurtosis:	3.746	Cond. No.	8.25e+14

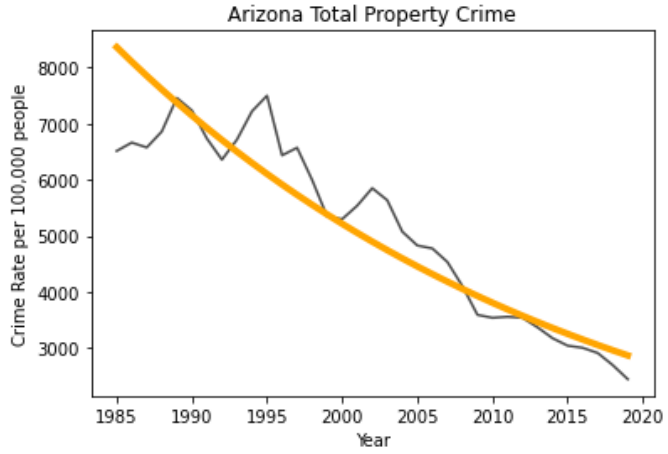
Model 3 – Log transformation of linear regression using OLS





Model 3 – Log transformation of linear regression using OLS





OLS Regression Results

Dep. Variable:	y	R-squared:	0.727
Model:	OLS	Adj. R-squared:	0.719
Method:	Least Squares	F-statistic:	88.08
Date:	Tue, 08 Dec 2020	Prob (F-statistic):	7.74e-11
Time:	04:48:32	Log-Likelihood:	33.849
No. Observations:	35	AIC:	-63.70
Df Residuals:	33	BIC:	-60.59
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	36.0783	3.174	11.365	0.000	29.620	42.537
x1	-0.0149	0.002	-9.385	0.000	-0.018	-0.012

Omnibus:	2.954	Durbin-Watson:	0.410
Prob(Omnibus):	0.228	Jarque-Bera (JB):	1.568
Skew:	0.192	Prob(JB):	0.456
Kurtosis:	2.037	Cond. No.	3.97e+05

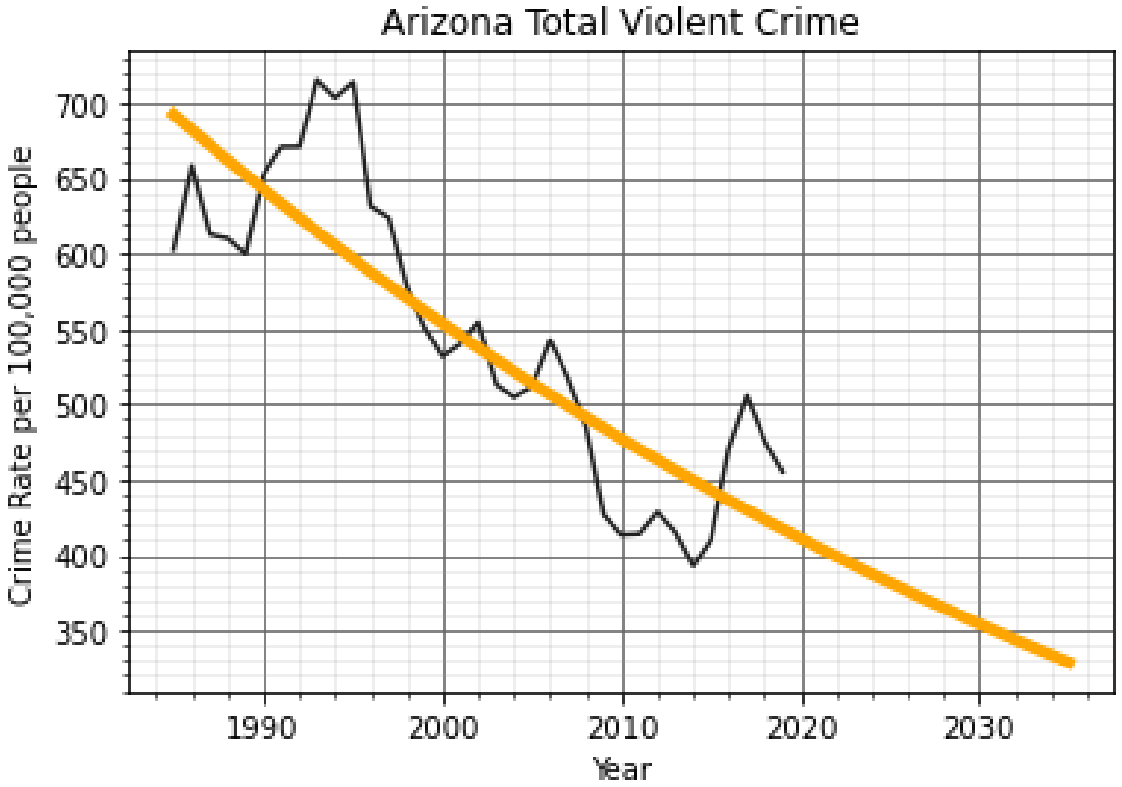
OLS Regression Results

Dep. Variable:	y	R-squared:	0.897
Model:	OLS	Adj. R-squared:	0.894
Method:	Least Squares	F-statistic:	287.8
Date:	Tue, 08 Dec 2020	Prob (F-statistic):	7.30e-18
Time:	04:48:32	Log-Likelihood:	28.306
No. Observations:	35	AIC:	-52.61
Df Residuals:	33	BIC:	-49.50
Df Model:	1		
Covariance Type:	nonrobust		

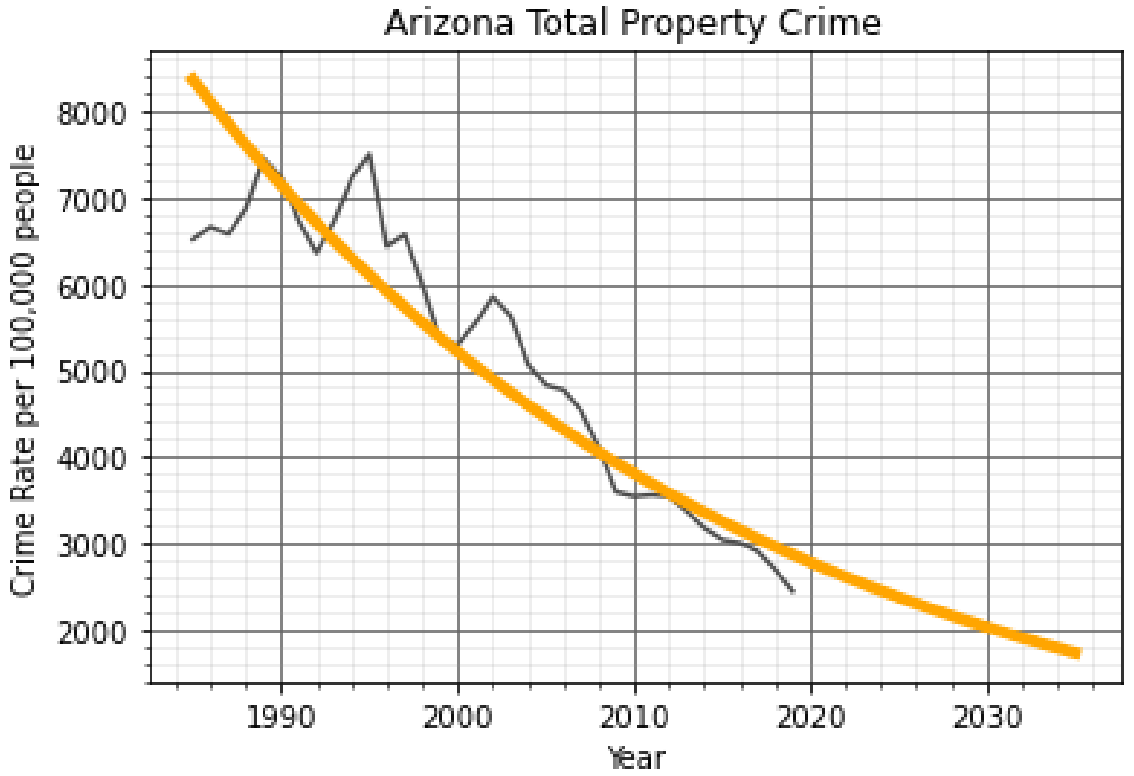
	coef	std err	t	P>|t|	[0.025	0.975]
const	71.5935	3.719	19.250	0.000	64.027	79.160
x1	-0.0315	0.002	-16.966	0.000	-0.035	-0.028

Omnibus:	0.270	Durbin-Watson:	0.298
Prob(Omnibus):	0.874	Jarque-Bera (JB):	0.421
Skew:	-0.171	Prob(JB):	0.810
Kurtosis:	2.586	Cond. No.	3.97e+05

15-year predictions using the Log transformation model

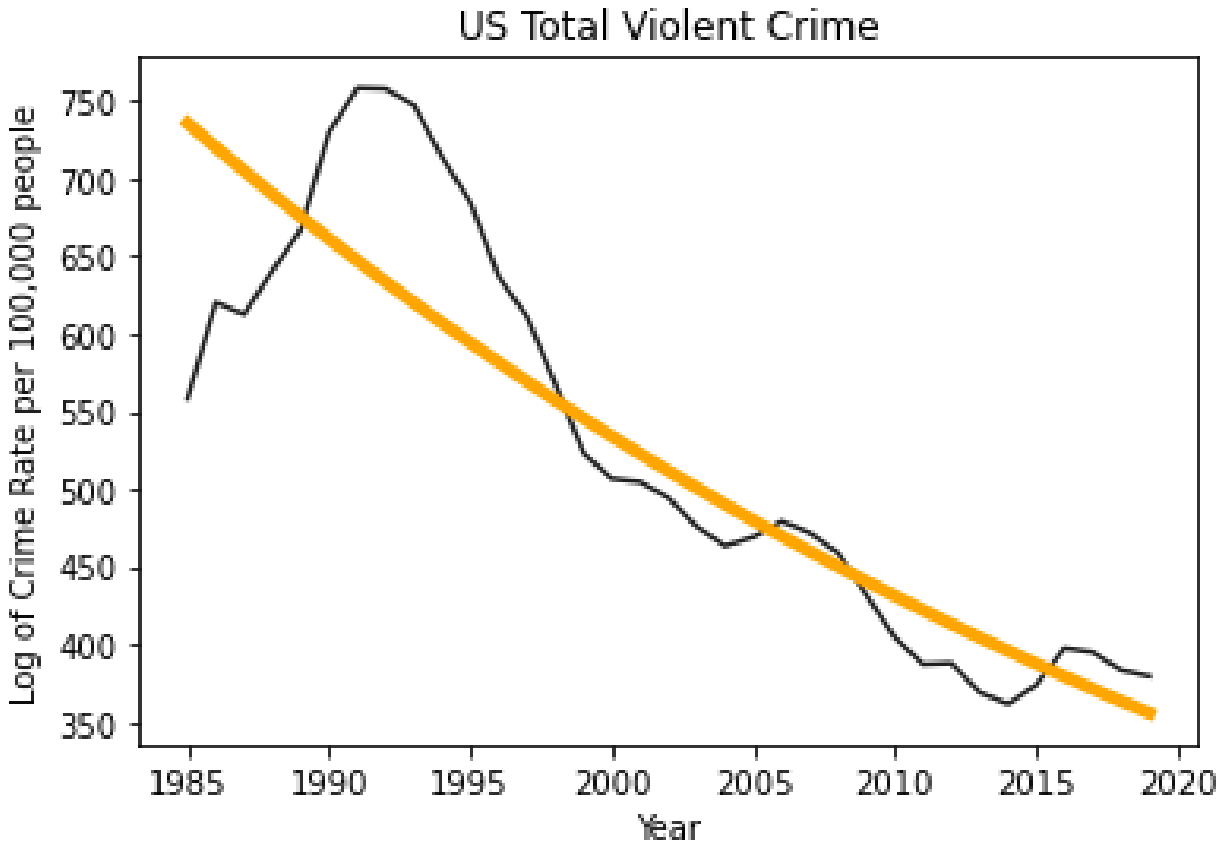


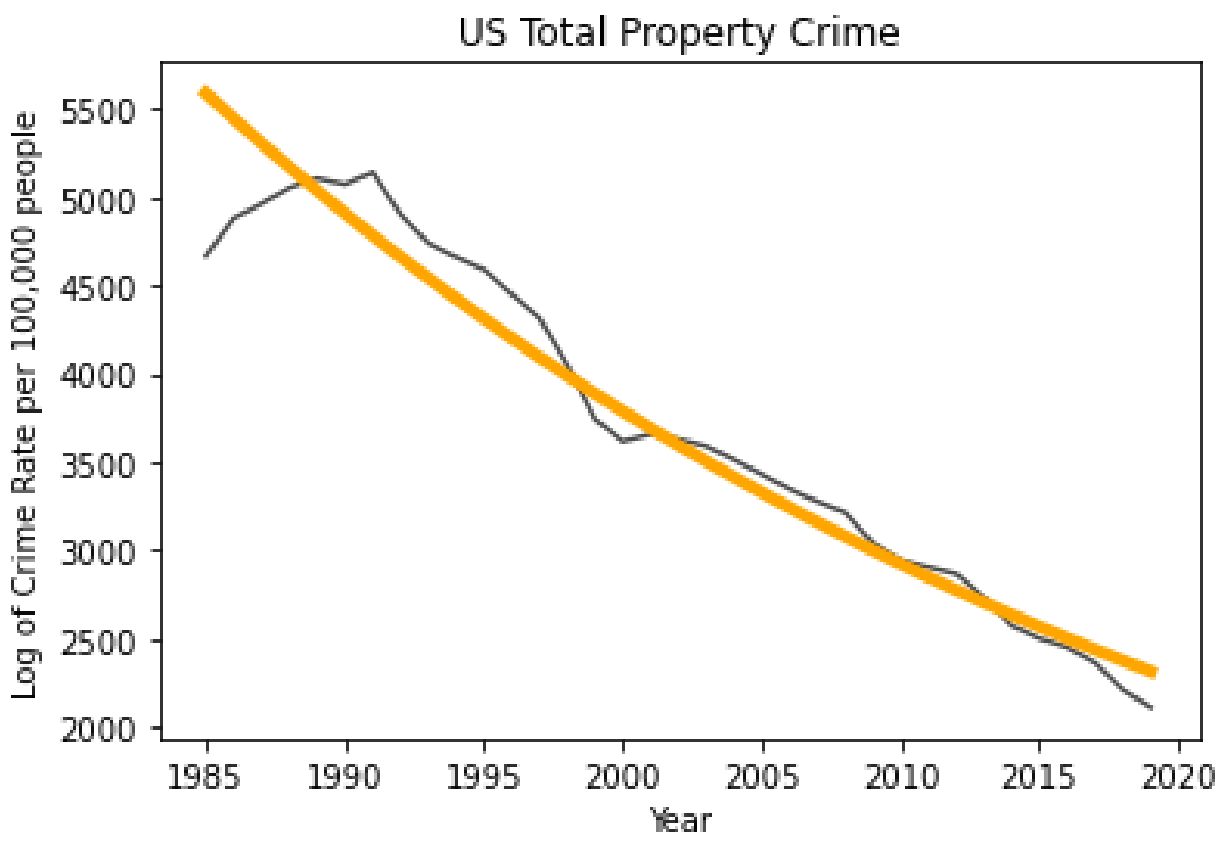

```
2019 - 2865.0 per 100,000
2025 - 2371.4 per 100,000  17% decrease
2030 - 2025.7 per 100,000  29% decrease
2035 - 1730.3 per 100,000  40% decrease
```



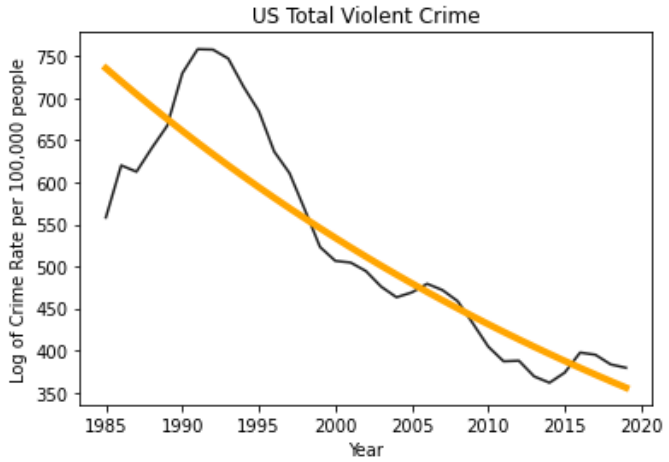

```
2019 - 417.4 per 100,000
2025 - 381.7 per 100,000  9% decrease
2030 - 354.3 per 100,000  15% decrease
2035 - 328.9 per 100,000  21% decrease
```

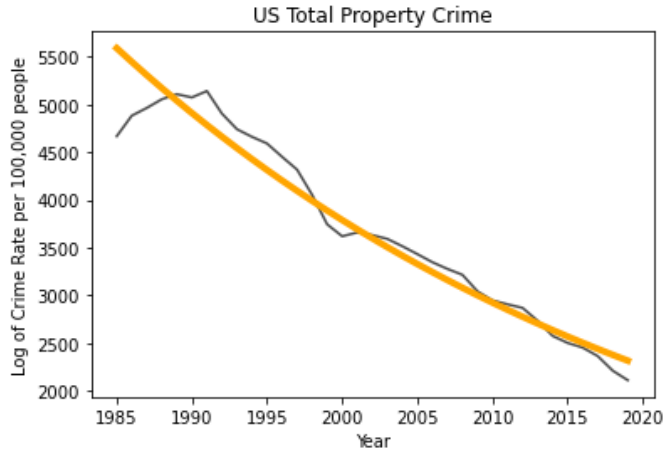
Applying the model to other datasets





Applying the model to other datasets





OLS Regression Results

Dep. Variable:	y	R-squared:	0.824
Model:	OLS	Adj. R-squared:	0.819
Method:	Least Squares	F-statistic:	155.0
Date:	Tue, 08 Dec 2020	Prob (F-statistic):	5.15e-14
Time:	07:05:03	Log-Likelihood:	31.077
No. Observations:	35	AIC:	-58.15
Df Residuals:	33	BIC:	-55.04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	49.0094	3.436	14.263	0.000	42.019	56.000
x1	-0.0214	0.002	-12.448	0.000	-0.025	-0.018

Omnibus:	0.810	Durbin-Watson:	0.203
Prob(Omnibus):	0.667	Jarque-Bera (JB):	0.200
Skew:	-0.137	Prob(JB):	0.905
Kurtosis:	3.250	Cond. No.	3.97e+05

OLS Regression Results

Dep. Variable:	y	R-squared:	0.959
Model:	OLS	Adj. R-squared:	0.958
Method:	Least Squares	F-statistic:	778.3
Date:	Tue, 08 Dec 2020	Prob (F-statistic):	1.59e-24
Time:	07:05:02	Log-Likelihood:	52.546
No. Observations:	35	AIC:	-101.1
Df Residuals:	33	BIC:	-97.98
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	60.0949	1.861	32.298	0.000	56.309	63.880
x1	-0.0259	0.001	-27.898	0.000	-0.028	-0.024

Omnibus:	14.284	Durbin-Watson:	0.238
Prob(Omnibus):	0.001	Jarque-Bera (JB):	15.071
Skew:	-1.329	Prob(JB):	0.000534
Kurtosis:	4.808	Cond. No.	3.97e+05

Q & A