

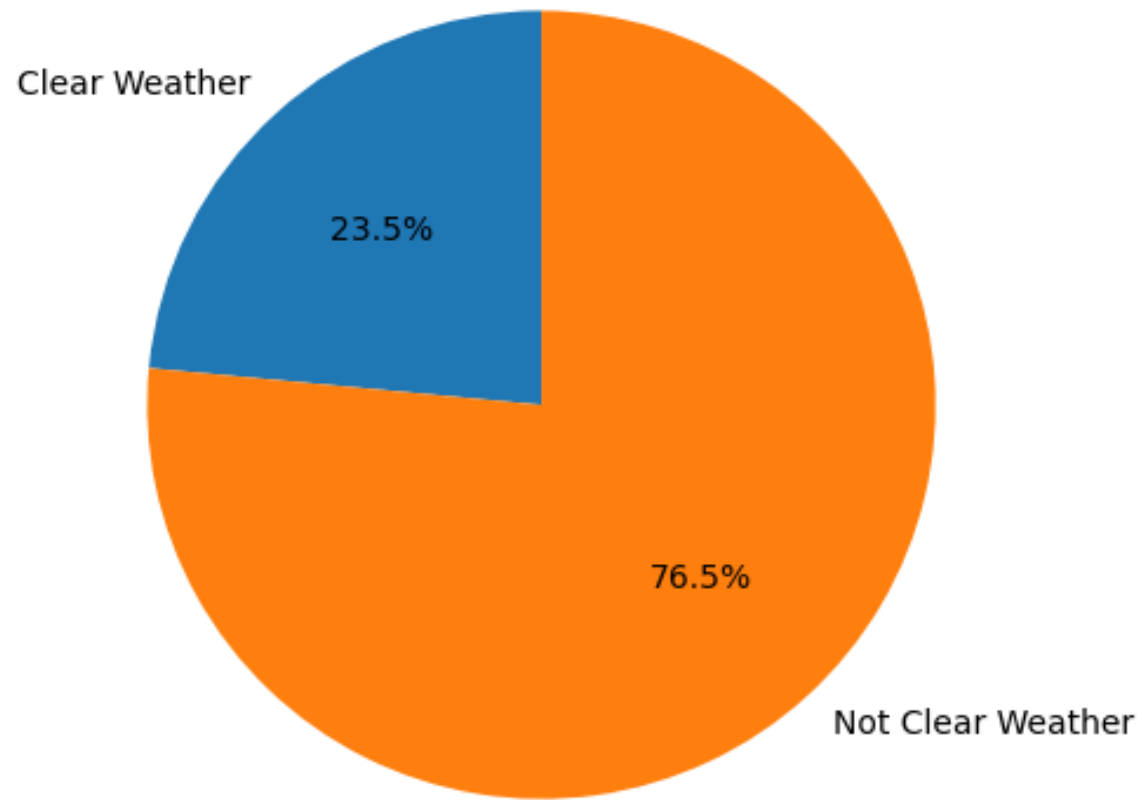
# Car Accident Forecasting

Forecasting car accidents by analyzing their relationships with weather conditions, time of day, and population

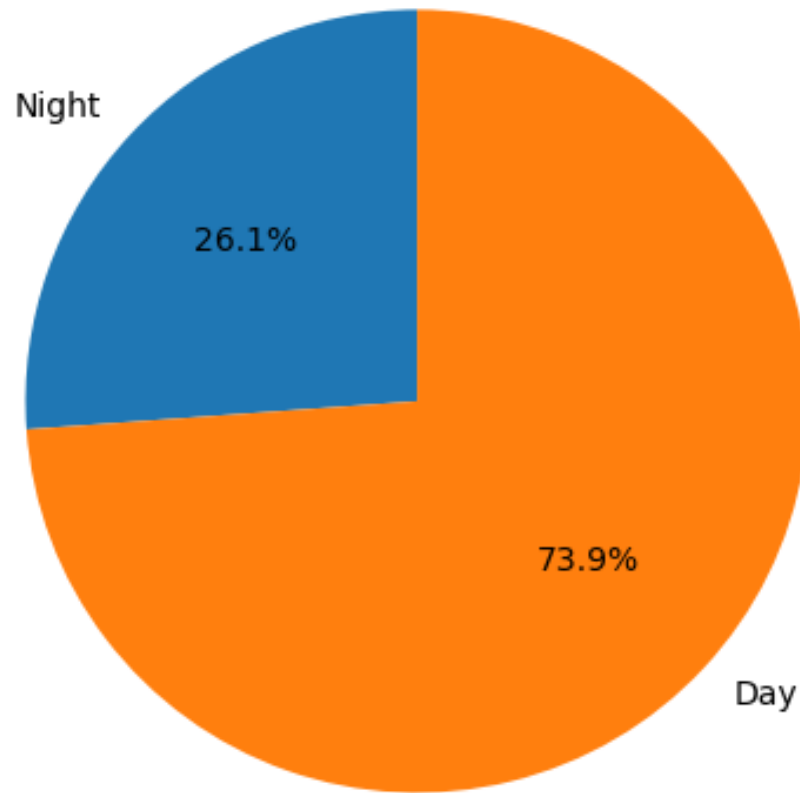
Gregory Happ

# The Datasets

- **US-Accidents: A Countrywide Traffic Accident Dataset**
  - [https://smoosavi.org/datasets/us\\_accidents](https://smoosavi.org/datasets/us_accidents)
- **Census Data**
  - <https://blog.splitwise.com/2013/09/18/the-2010-us-census-population-by-zip-code-totally-free/amp/>



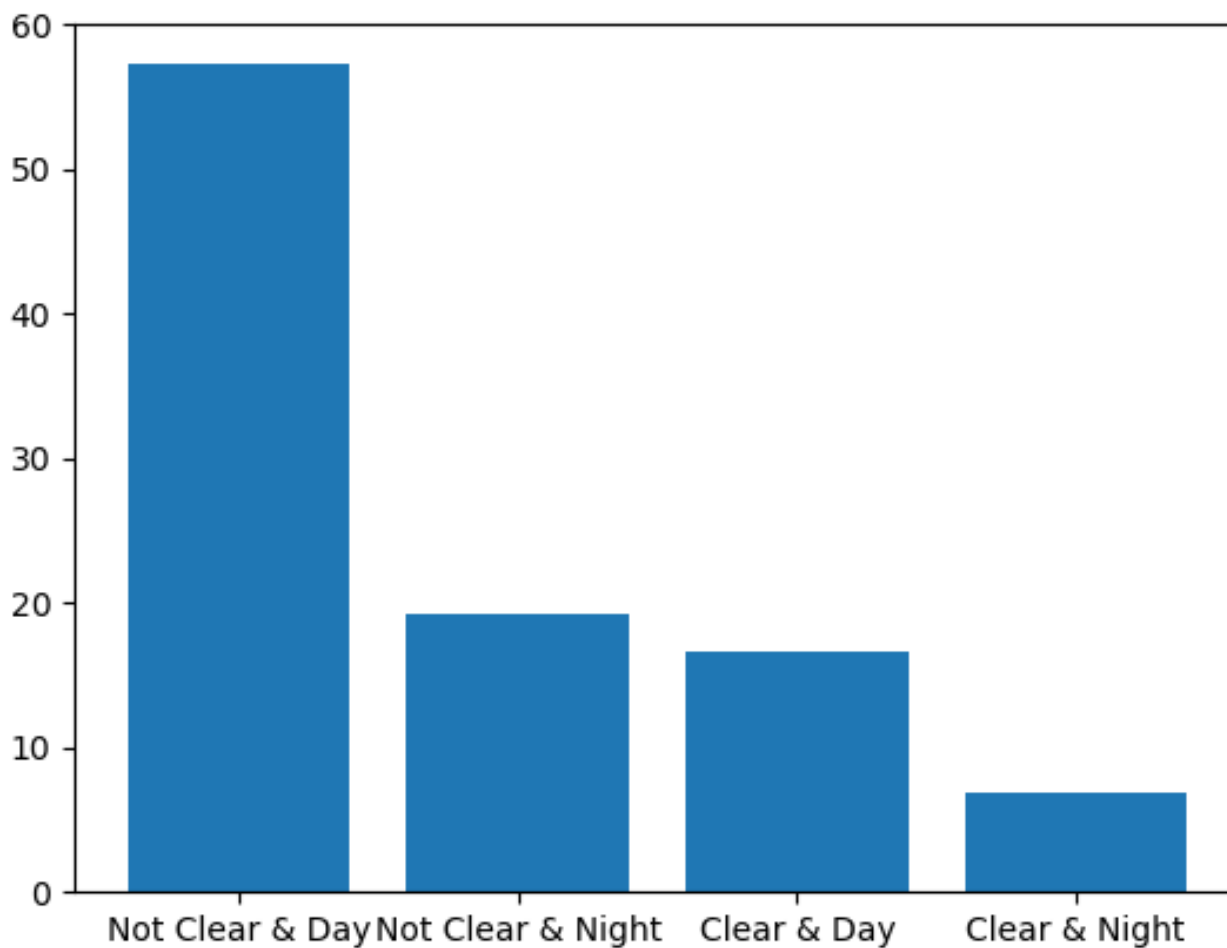
Weather  
Conditions



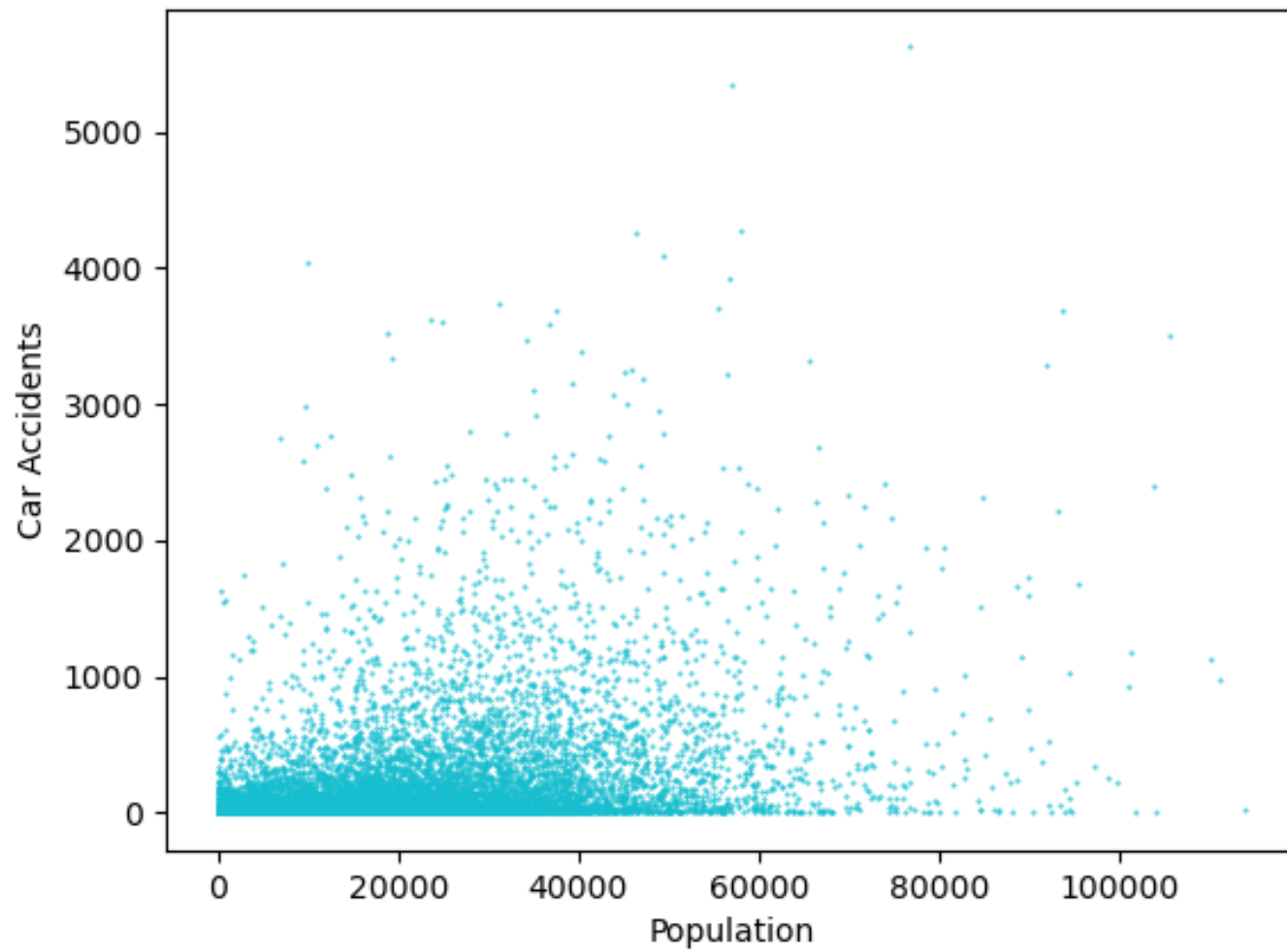
Day or Night

# Weather Conditions and Time

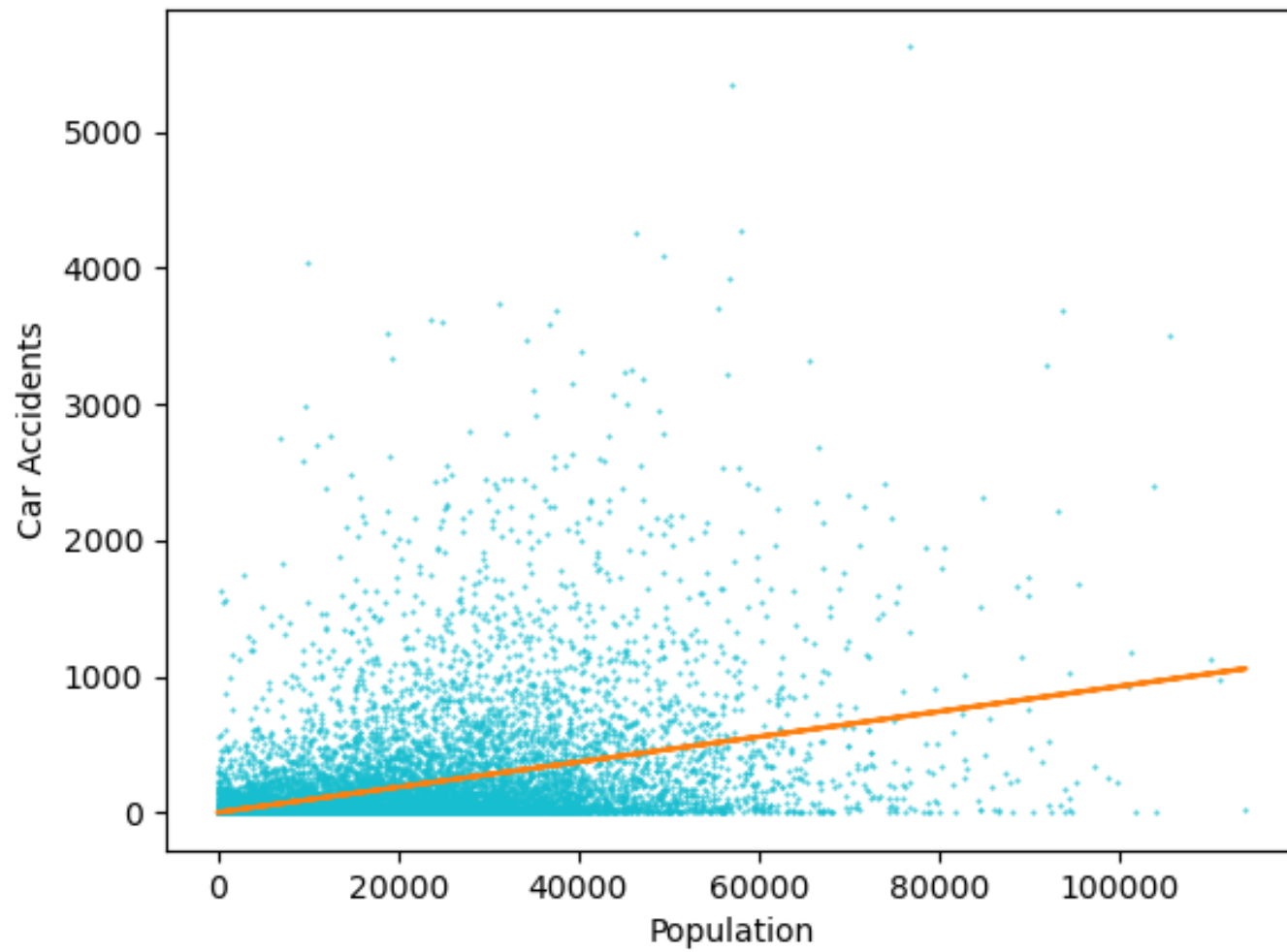
	Day	Night	Total
Weather Not Clear	1966561	662750	2629311
Clear Weather	572087	236090	808177
Total	2538648	898840	3437488



Weather  
Conditions  
and Time



Population



Population



# Correlation: Population and Car Accidents

Pearson's  $r = 0.4278201964372337$

p-value = 0.0

Spearman's  $\rho = 0.5770351900548639$

p-value = 0.0

Kendall's  $\tau = 0.4071861941576106$

p-value = 0.0

	Car Accidents
Population	0.009*** (0.0001)
Constant	3.545 (3.240)
<i>N</i>	17,249
$R^2$	0.183
Adjusted $R^2$	0.183
Residual Std. Error	310.046 (df = 17247)
F Statistic	3,863.937*** (df = 1; 17247)

*Notes:* \*\*\* Significant at the 1 percent level.  
 \*\* Significant at the 5 percent level.  
 \* Significant at the 10 percent level.

OLS:  
Population  
and Car  
Accidents

# Summary

There is evidence that suggests that weather conditions, time of day, and population could be used to forecast the total amount of car accidents for a given area during a given time period.

# Acknowledgments

- Python and R were used for coding the project
  - Python Packages: pandas, matplotlib, scipy, sklearn
    - McKinney, W., & others. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference* (Vol. 445, pp. 51–56).
    - Hunter, J. D. (2007). **Matplotlib**: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90–95.
    - Virtanen, P., Gommers, R., Oliphant, Travis E., Haberland, M., Reddy, T., Cournapeau, D., ... Contributors, SciPy 1. 0. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*.
    - [Scikit-learn: Machine Learning in Python](#), Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011.
  - R Packages: stargazer, dplyr
    - Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.2. <https://CRAN.R-project.org/package=stargazer>
    - Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2018). dplyr: A Grammar of Data Manipulation. R package version 0.7.6. <https://CRAN.R-project.org/package=dplyr>
- 1<sup>st</sup> Dataset:
  - [https://smoosavi.org/datasets/us\\_accidents](https://smoosavi.org/datasets/us_accidents)
  - Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. [“A Countrywide Traffic Accident Dataset.”](#), arXiv preprint arXiv:1906.05409 (2019).
  - Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, Radu Teodorescu, and Rajiv Ramnath. [“Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights.”](#) In proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2019.
- 2<sup>nd</sup> Dataset:
  - <https://blog.splitwise.com/2013/09/18/the-2010-us-census-population-by-zip-code-totally-free/amp/>
  - Jon Bittner, The Splitwise Blog Contributor