**Annotated Bibliography**

Abdel-Aty, M., Keller, J., & Brady, P. A. (2005). Analysis of Types of Crashes at Signalized Intersections by Using Complete Crash Data and Tree-Based Regression. *Transportation Research Record*, *1908*(1), 37–45. https://doi.org/10.1177/0361198105190800105

There are a lot of factors that contributes to crashes at intersections. This paper analyzes these factors contributing to traffic crash at signalized intersection, using 'tree-based regression methodology'. The authors used this methodology so as to 'cope with multicollinearity between variables, missing observations, and the fact that the true model form was unknown.' The result of this paper shows that there is huge variations regarding the factors that influence the various 'collision types'.

This paper will provide vital information on the additional regression methodology for my course project work.

Jonathan, A.-V., Wu, K.-F. (Ken), & Donnell, E. T. (2016). A multivariate spatial crash frequency model for identifying sites with promise based on crash types. *Accident Analysis & Prevention*, *87*, 8–16. https://doi.org/10.1016/j.aap.2015.11.006

The authors created a model to improve the precision of 'crash frequency models, used as an input in identifying 'sites with promise (SWiPs)'. The authors used 'multivariate spatial model for modeling crash frequencies for different crash types.' The paper reports that 'the model that considers both multivariate and spatial correlation has the best fit'. This paper will inform the variables I will use to create my regression mdoel.

Qin, X., Ng, M., & Reyes, P. E. (2010). Identifying crash-prone locations with quantile regression. *Accident Analysis & Prevention*, *42*(6), 1531–1537. https://doi.org/10.1016/j.aap.2010.03.009

The authors identified crash prone areas using regression model at the quantile level instead of changes at the mean level. According to the authors, this methodology is useful due to the 'heterogeneity of crash data'. The result shows that 'quantile regression' produced a 'refined subset of risk-prone locations' when compared with other methods. This paper will help me understand the various methods available for analyzing crash data.

Rezapour, M., & Ksaibati, K. (2018). Application of multinomial and ordinal logistic regression to model injury severity of truck crashes, using violation and crash data. *Journal of Modern Transportation*, *26*(4), 268–277. https://doi.org/10.1007/s40534-018-0166-x

The authors aimed to identify factors contributing to 'severe truck crashes in conjunction with violation data'. The authors used ordinal logistic regression to identify the factors that 'increased the odds of severe single-truck and multiple-vehicle crashes.' The result shows that factors such as 'non-normal conditions at the time of crash, driving on dry-road condition and having a distraction in the cabin' increased the odds of severe single-truck crashes'. This article will help me identify independent variables useful for my regression model.