\* Do the clustering methods generate the same clusters?

YES, the two methods (K-MEANS, DBSCAN) generate the same result.

\* Does scaling effect the clustering?

No

\* Does the clustering produce interesting groupings?

Most vehicles crash happened under three conditions:

1, the weather is below 4 and the road condition is below 4.

2. the illumination is below 5 and the road condition is below 6.

3. the weather is below 4 and the illumination is below 5.

\* Is the relationship significant?

For multivariate linear regression, the equivalence is

y=1.186-0.025∗ILLUMINATION-0.415∗WEATHER-0.041∗ROAD\_CONDITION

There exists a linear relationship, but the relationship is not that strong, because it does not mean there must be some accident crash happens if the weather, road condition and the illumination are getting worse.

\* Are any model assumptions violated?

No.

\* Is there any multi-colinearity in the model?

Three factors, as mentioned above.

\* In the multiple regression models are predictor variables independent of all the other predictor variables?

Based on the single linear regression model, all of the three factors affect results separately.

WEATHER: y=1.958-0.079\*weather

ROAD CONDITION: y=1.896-0.079\*road\_condition

ILLUMINATION: y=1.992-0.089\*illumination

\* In in multiple regression models rank the most significant predictor variables and exclude insignificant ones from the model.

WEATHER > ROAD CONDITION > ILLUMINATION

\* Does the model make sense?

Negative relationship can be calculated.

\* Cross-validate the model. How well did it do?

RMES: 0.57255733319016622

\* Does regularization help with creating models that validate better on out of sample data?

\* Is the relationship significant?

\* Are any model assumptions violated?

No, all of these conditions formed a negative relationship with the number of dead driver.

\* Cross-validate the model. How well did it do?