

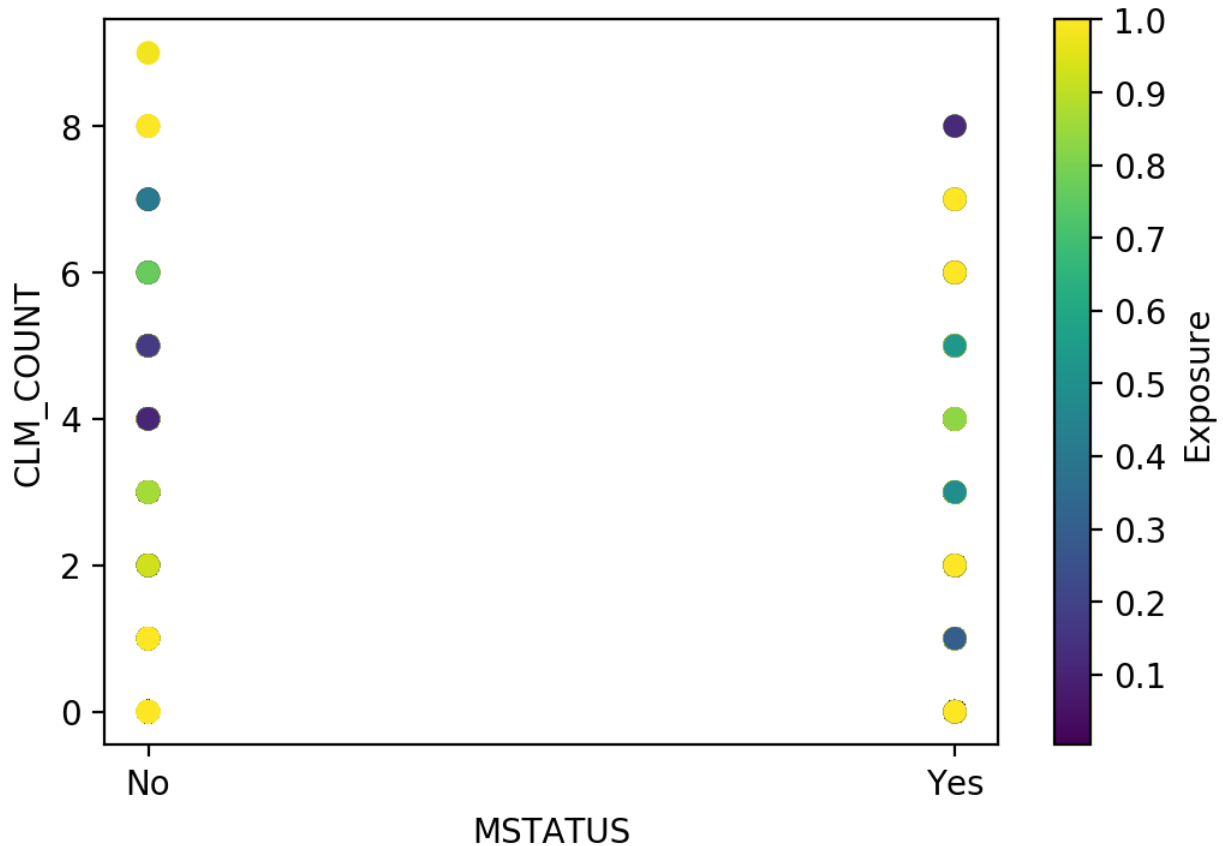
MSCA31010: Linear & Non-Linear Models

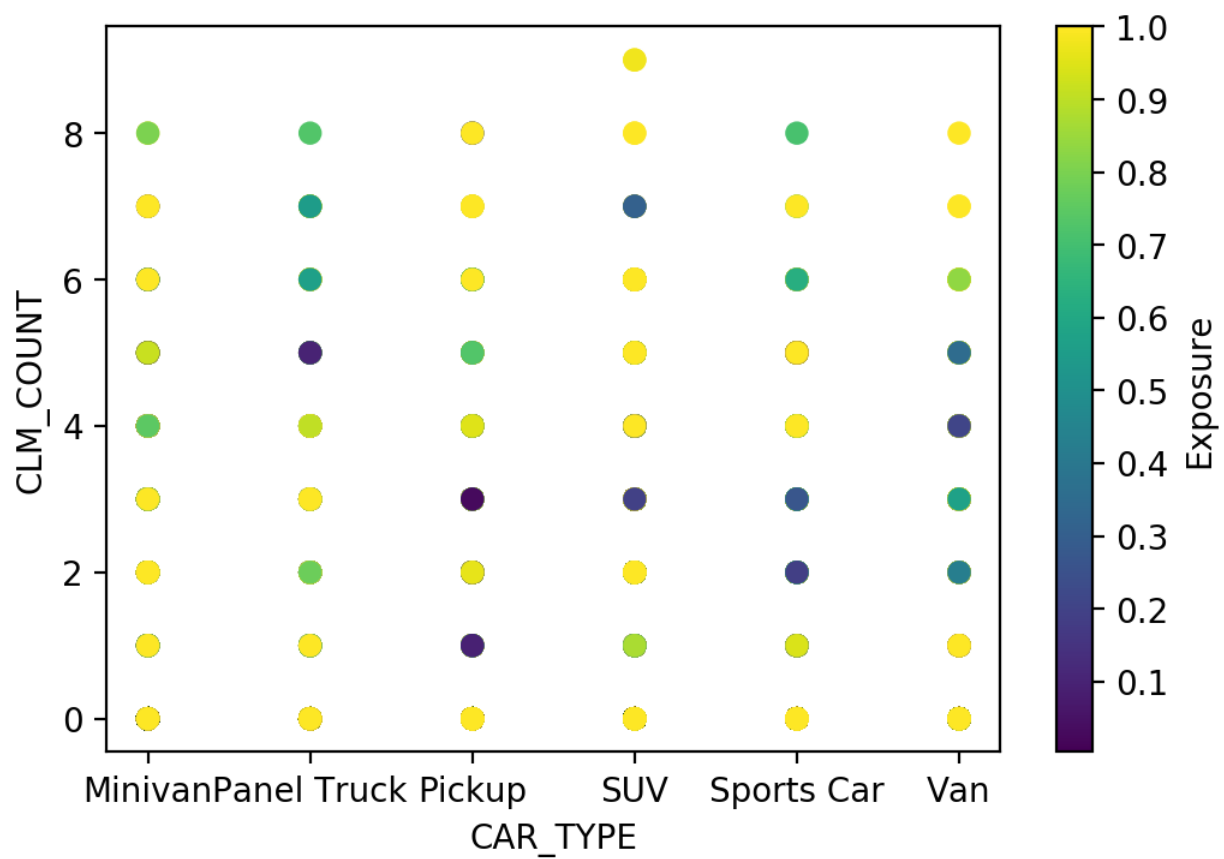
Winter 2022 Assignment 2

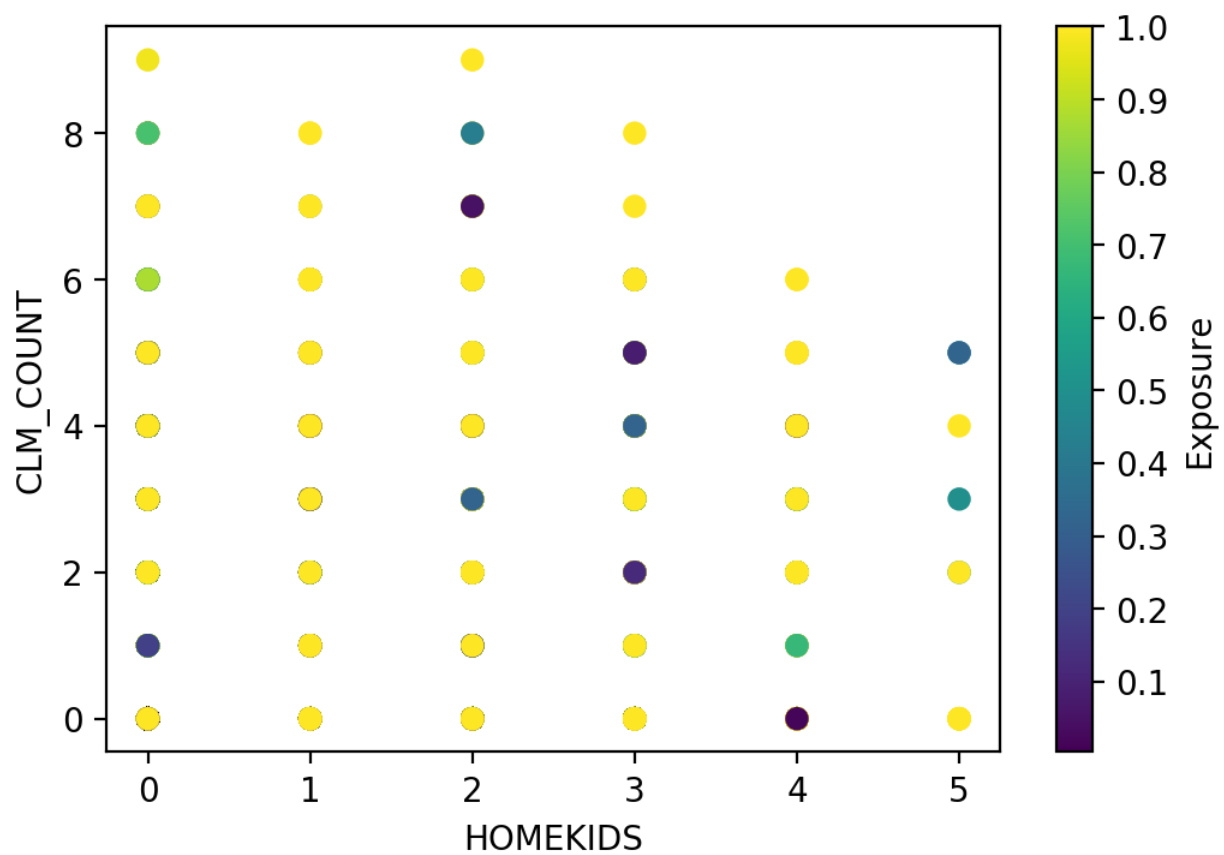
Question 1 (20 points)

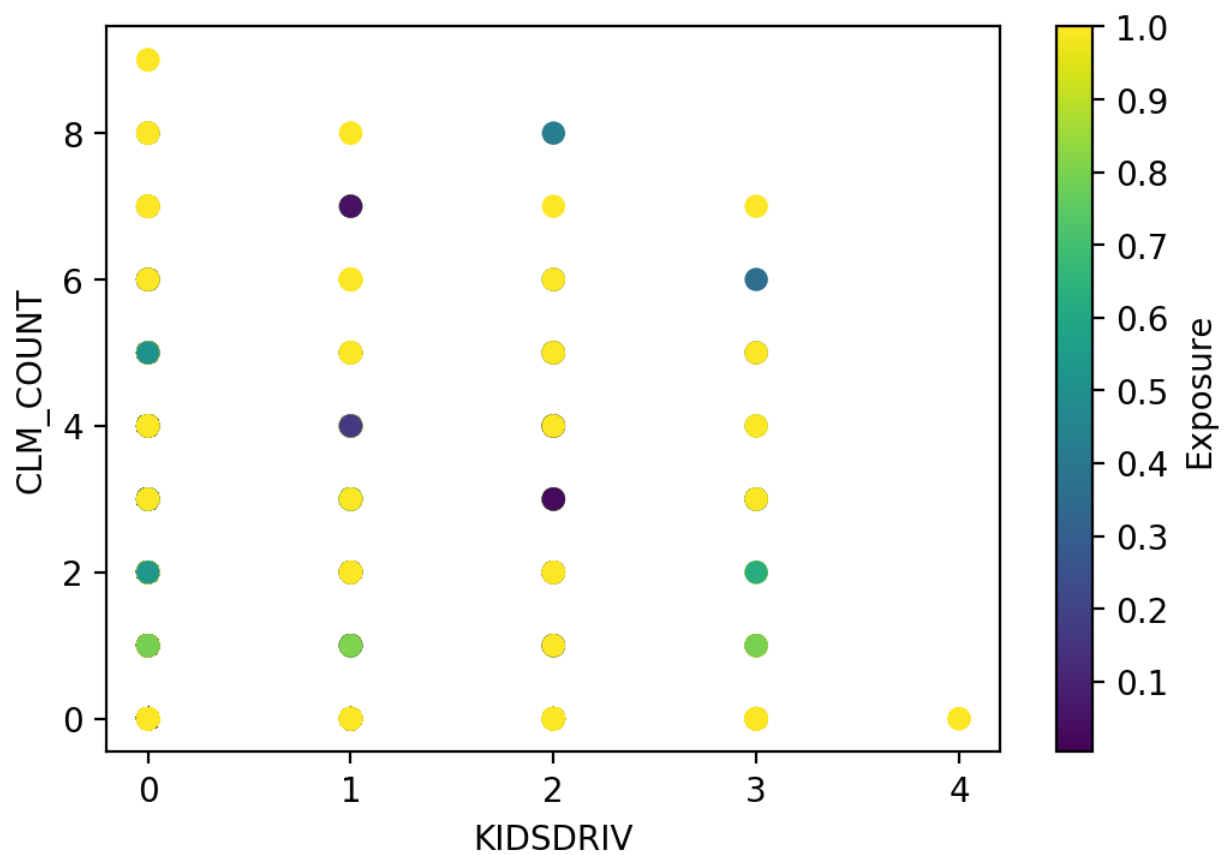
Before you train the model, you want to explore the predictors.

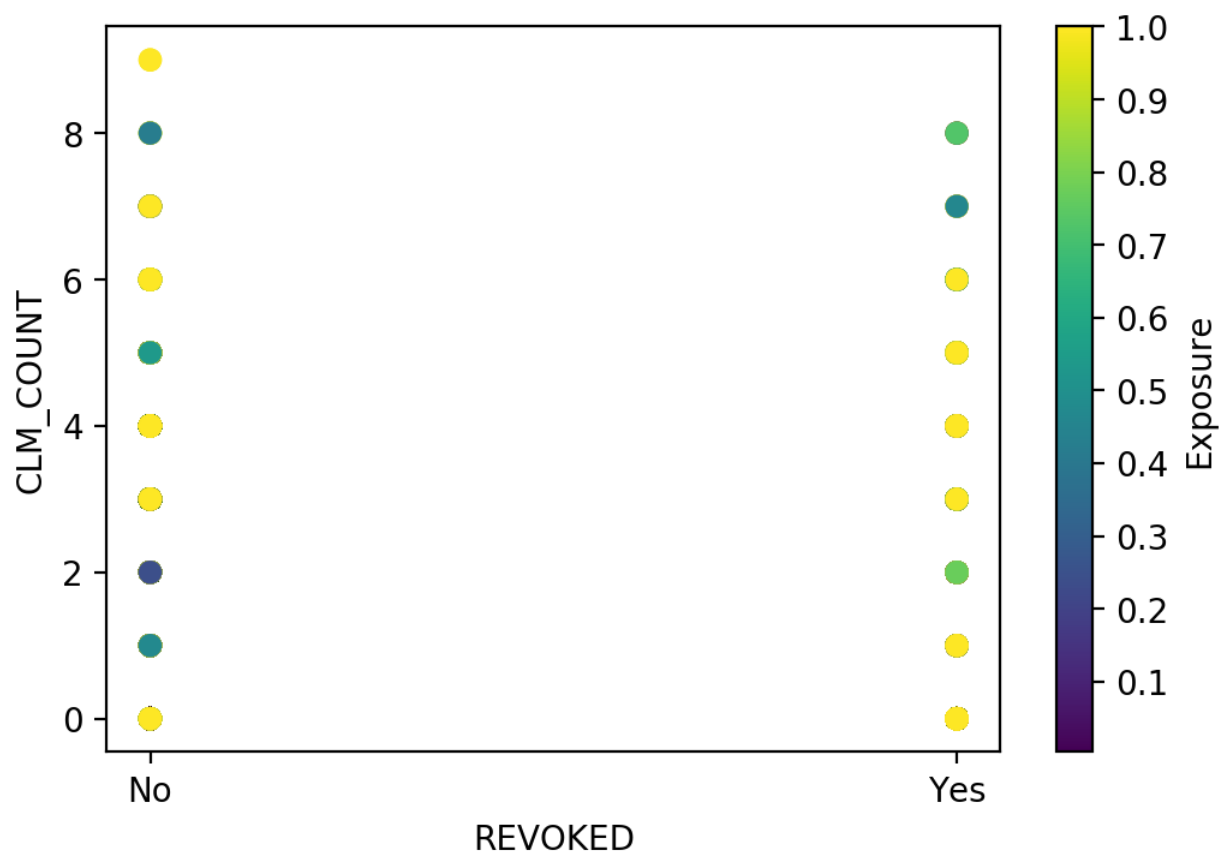
- a) (20 points) For each predictor, generate a scatterplot chart that shows the number of claims by the predictor's values. Also, color-code the markers by the exposure values. Please display the predictor's values are displayed in ascending lexical order.

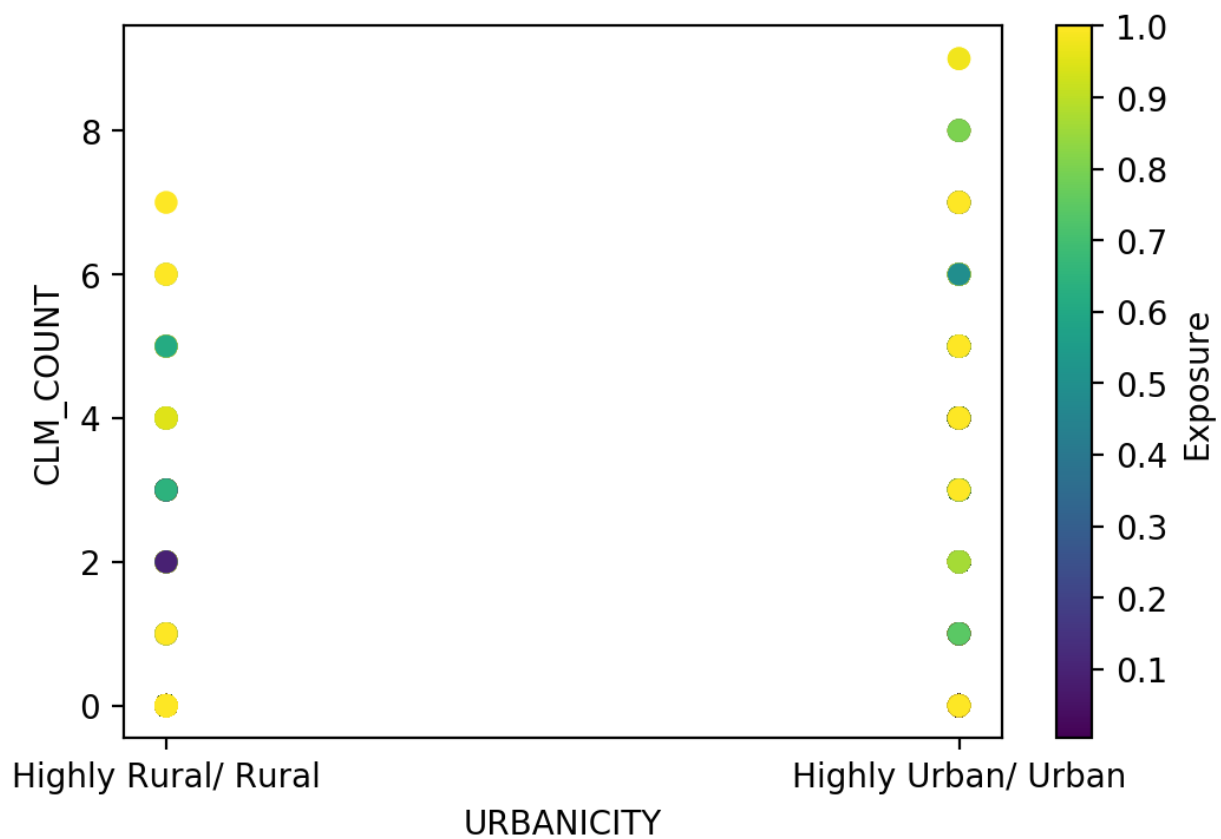


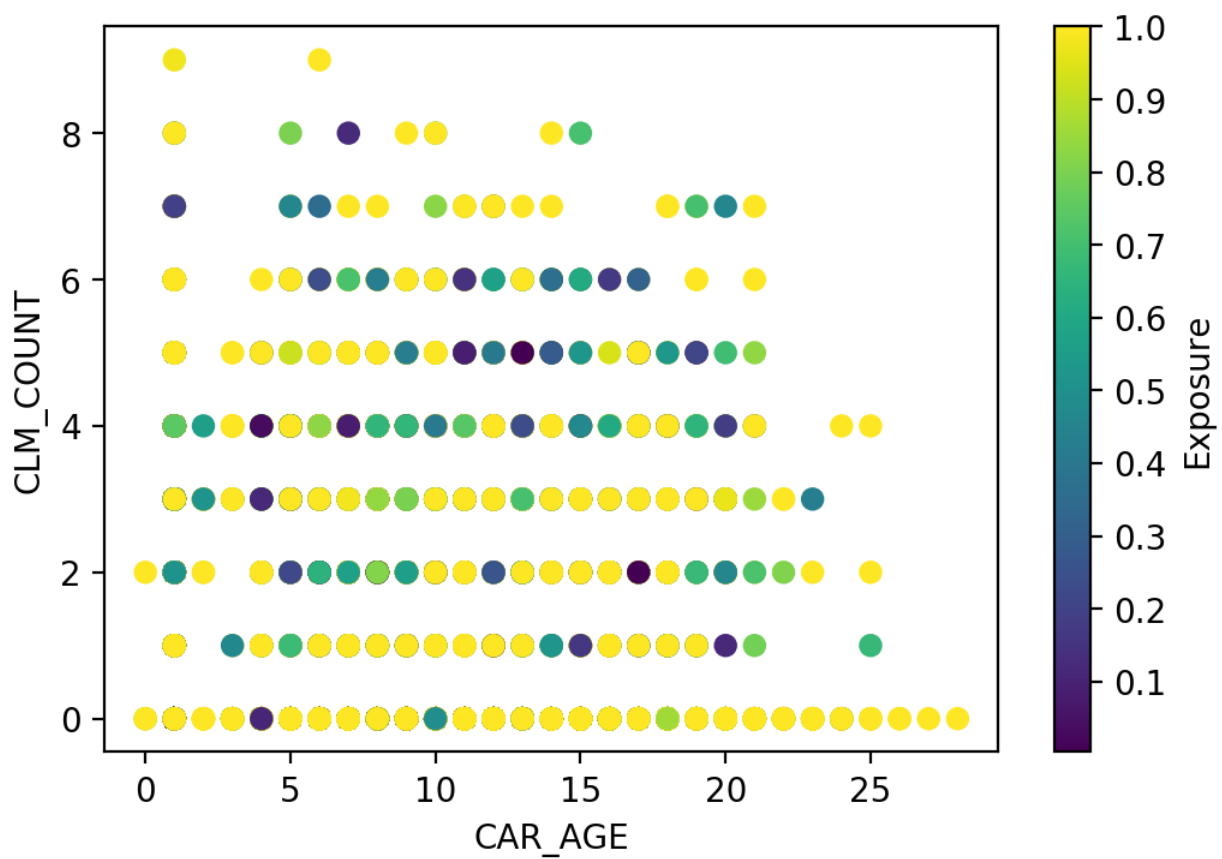


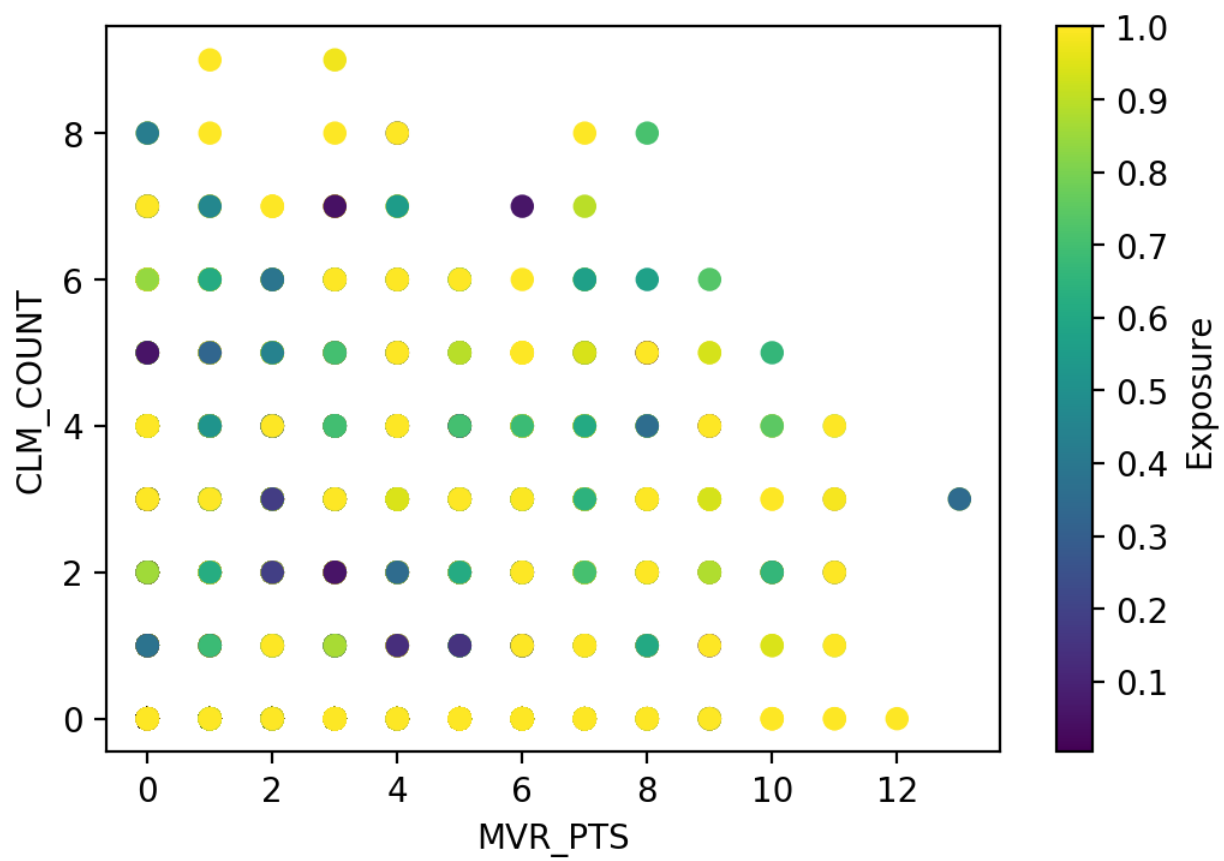


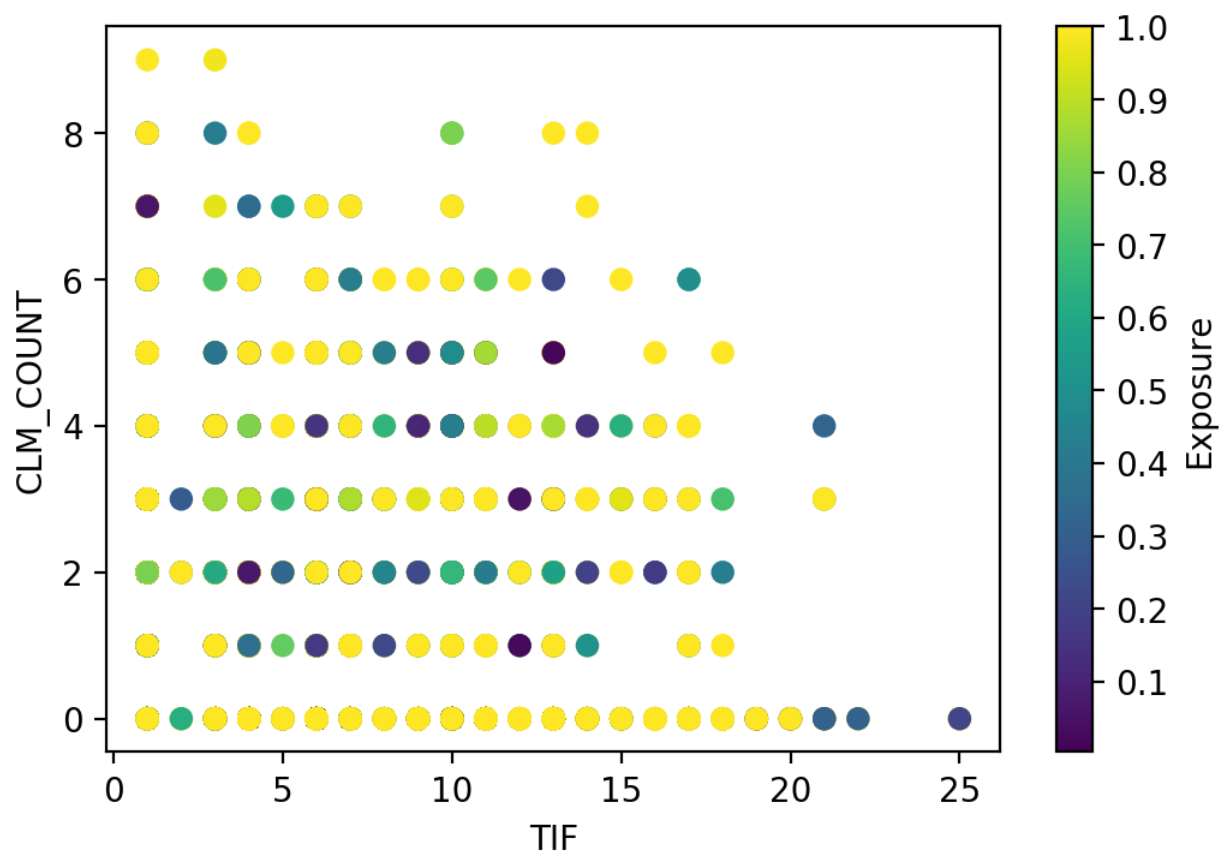


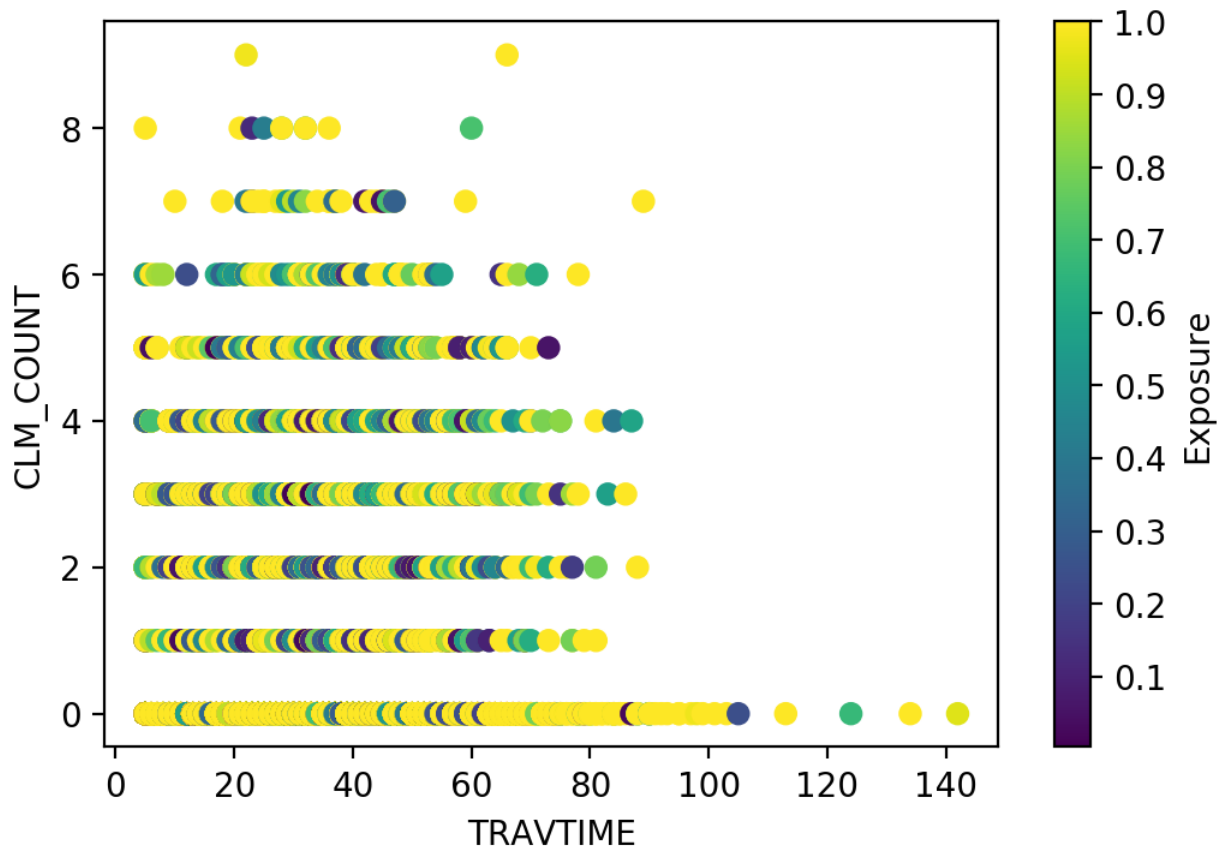












Question 2 (40 points)

Enter the predictors into your model using Forward Selection. The Entry Threshold is 0.05.

- a) (15 points). Please provide a summary report of the Forward Selection. The report should include (1) the step number, (2) the predictor entered, (3) the number of non-aliased parameters in the current model, (4) the log-likelihood value of the current model, (5) the Deviance Chi-squares statistic between the current and the previous models, (6) the corresponding Deviance Degree of Freedom, and (7) the corresponding Chi-square significance.

step		predictors	num of non aliased parameter	log-likelihood	Deviance Chi- Squares	Deviance Df	Chi-square significance
0	0	Intercept	1	-1.7324418e+04	nan	nan	nan
1	1	+ MVR_PTS	2	-1.6816770e+04	1.0152957e+03	1.0000000e+00	8.5025203e-223
2	2	+ MVR_PTS + URBANICITY	3	-1.6132234e+04	1.3690720e+03	1.0000000e+00	1.1046068e-299
3	3	+ MVR_PTS + URBANICITY + CAR_AGE	4	-1.5898009e+04	4.6844946e+02	1.0000000e+00	6.9691259e-104
4	4	+ MVR_PTS + URBANICITY + CAR_AGE + MSTATUS	5	-1.5720053e+04	3.5591160e+02	1.0000000e+00	2.1869339e-79
5	5	+ MVR_PTS + URBANICITY + CAR_AGE + MSTATUS + CAR_TYPE	10	-1.5539157e+04	3.6179176e+02	5.0000000e+00	5.0584841e-76
6	6	+ MVR_PTS + URBANICITY + CAR_AGE + MSTATUS + CAR_TYPE + REVOKED	11	-1.5403153e+04	2.7200815e+02	1.0000000e+00	4.1423638e-61
7	7	+ MVR_PTS + URBANICITY + CAR_AGE + MSTATUS + CAR_TYPE + REVOKED + KIDSDRIV	12	-1.5299053e+04	2.0819974e+02	1.0000000e+00	3.3934421e-47
8	8	+ MVR_PTS + URBANICITY + CAR_AGE + MSTATUS + CAR_TYPE + REVOKED + KIDSDRIV + TRAVTIME	13	-1.5202055e+04	1.9399703e+02	1.0000000e+00	4.2649355e-44
9	9	+ MVR_PTS + URBANICITY + CAR_AGE + MSTATUS + CAR_TYPE + REVOKED + KIDSDRIV + TRAVTIME + TIF	14	-1.5145287e+04	1.1353589e+02	1.0000000e+00	1.6467361e-26
10	10	+ MVR_PTS + URBANICITY + CAR_AGE + MSTATUS + CAR_TYPE + REVOKED + KIDSDRIV + TRAVTIME + TIF + HOMEKIDS	15	-1.5102010e+04	8.6553304e+01	1.0000000e+00	1.3602126e-20

b) (5 points). What predictors does your final model contain?

MVR_PTS , URBANICITY , CAR_AGE , MSTATUS , CAR_TYPE, REVOKED , KIDSDRIV , TRAVTIME , TIF ,
HOMEKIDS

c) (5 points). What are the aliased parameters in your final model? Please list the predictor's name and the aliased categories.

aliased parameters: Urbanicity,Mstatus,CAR_TYPE, REVOKED

aliased categories are: **Highly Urban/ Urban, MSTATUS_YES, SUV, REVOKED_YES**

d) (5 points). How many non-aliased parameters are in your final model?

15 non-aliased parameters in the final model.

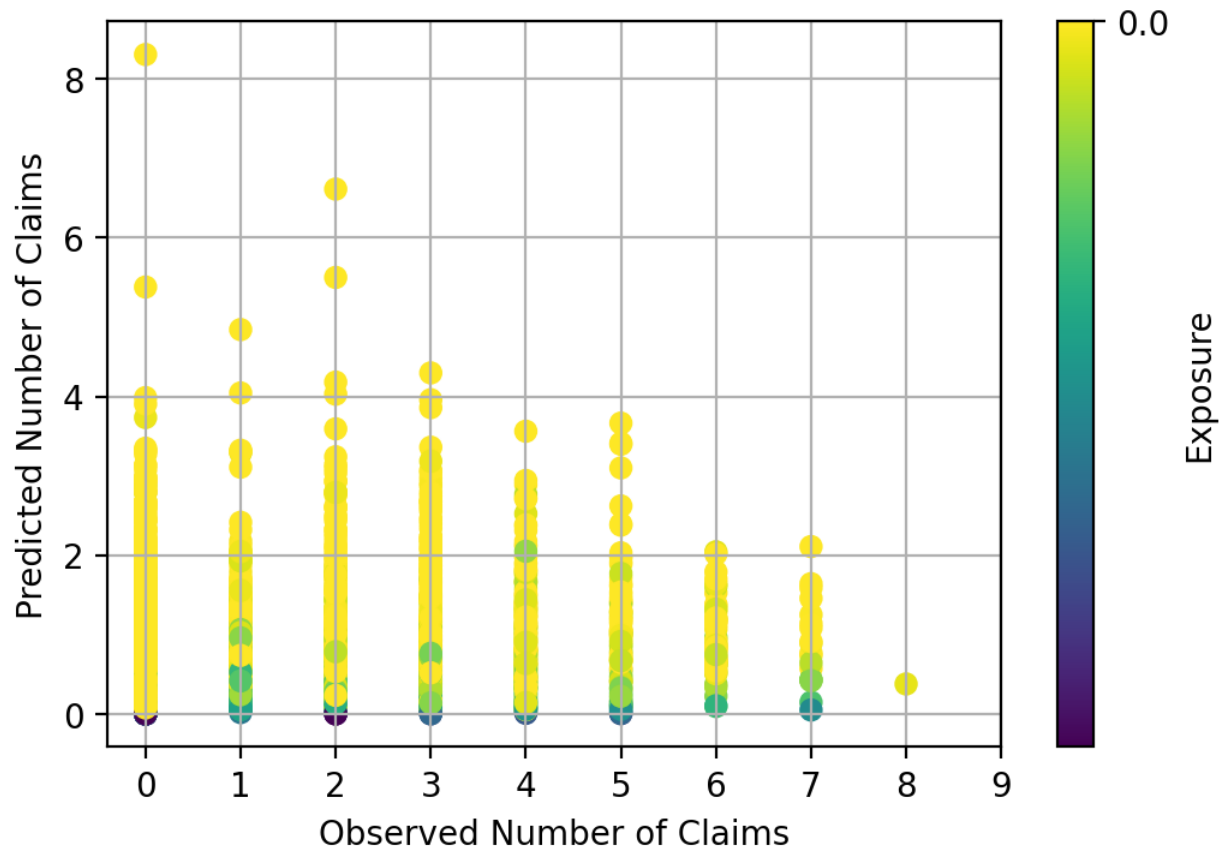
- e) (10 points). Please show a table of the complete set of parameters of your final model (including the aliased parameters). Besides the parameter estimates, please also include the standard errors, and the 95% asymptotic confidence intervals. Conventionally, aliased parameters have missing standard errors and confidence intervals

	Estimate	Standard Error	Lower 95% CI	Upper 95% CI	Exponentiated
Intercept	-2.6322414e-01	4.4018393e-02	-3.4949860e-01	-1.7694967e-01	7.6856961e-01
MVR_PTS	8.7009883e-02	4.4357124e-03	7.8316047e-02	9.5703720e-02	1.0909075e+00
Highly Rural/ Rural	-1.7758313e+00	5.4677121e-02	-1.8829965e+00	-1.6686661e+00	1.6934262e-01
Highly Urban/ Urban	0.0000000e+00	0.0000000e+00	0.0000000e+00	0.0000000e+00	1.0000000e+00
CAR_AGE	-3.9573491e-02	2.1044246e-03	-4.3698088e-02	-3.5448895e-02	9.6119931e-01
MSTATUS_NO	4.5935083e-01	2.2857571e-02	4.1455082e-01	5.0415085e-01	1.5830460e+00
MSTATUS_YES	0.0000000e+00	0.0000000e+00	0.0000000e+00	0.0000000e+00	1.0000000e+00
Panel Truck	3.9330010e-02	4.4299910e-02	-4.7496218e-02	1.2615624e-01	1.0401137e+00
Van	-1.8766485e-02	4.3011406e-02	-1.0306729e-01	6.5534321e-02	9.8140851e-01
Sports Car	1.7762272e-01	3.6036263e-02	1.0699295e-01	2.4825250e-01	1.1943746e+00
Pickup	5.9603276e-02	3.2913905e-02	-4.9067932e-03	1.2411334e-01	1.0614154e+00
Minivan	-4.7059180e-01	3.4284237e-02	-5.3778767e-01	-4.0339593e-01	6.2463250e-01
SUV	0.0000000e+00	0.0000000e+00	0.0000000e+00	0.0000000e+00	1.0000000e+00
REVOKED_NO	4.3719893e-01	2.8261101e-02	3.8180819e-01	4.9258967e-01	1.5483641e+00
REVOKED_YES	0.0000000e+00	0.0000000e+00	0.0000000e+00	0.0000000e+00	1.0000000e+00
KIDSDRIV	1.7564921e-01	2.0486256e-02	1.3549689e-01	2.1580154e-01	1.1920198e+00
TRAVTIME	1.0231442e-02	7.2432061e-04	8.8117994e-03	1.1651084e-02	1.0102840e+00
TIF	-3.1315006e-02	2.9534848e-03	-3.7103730e-02	-2.5526282e-02	9.6917023e-01
HOMEKIDS	1.0284513e-01	1.0790460e-02	8.1696220e-02	1.2399405e-01	1.1083198e+00

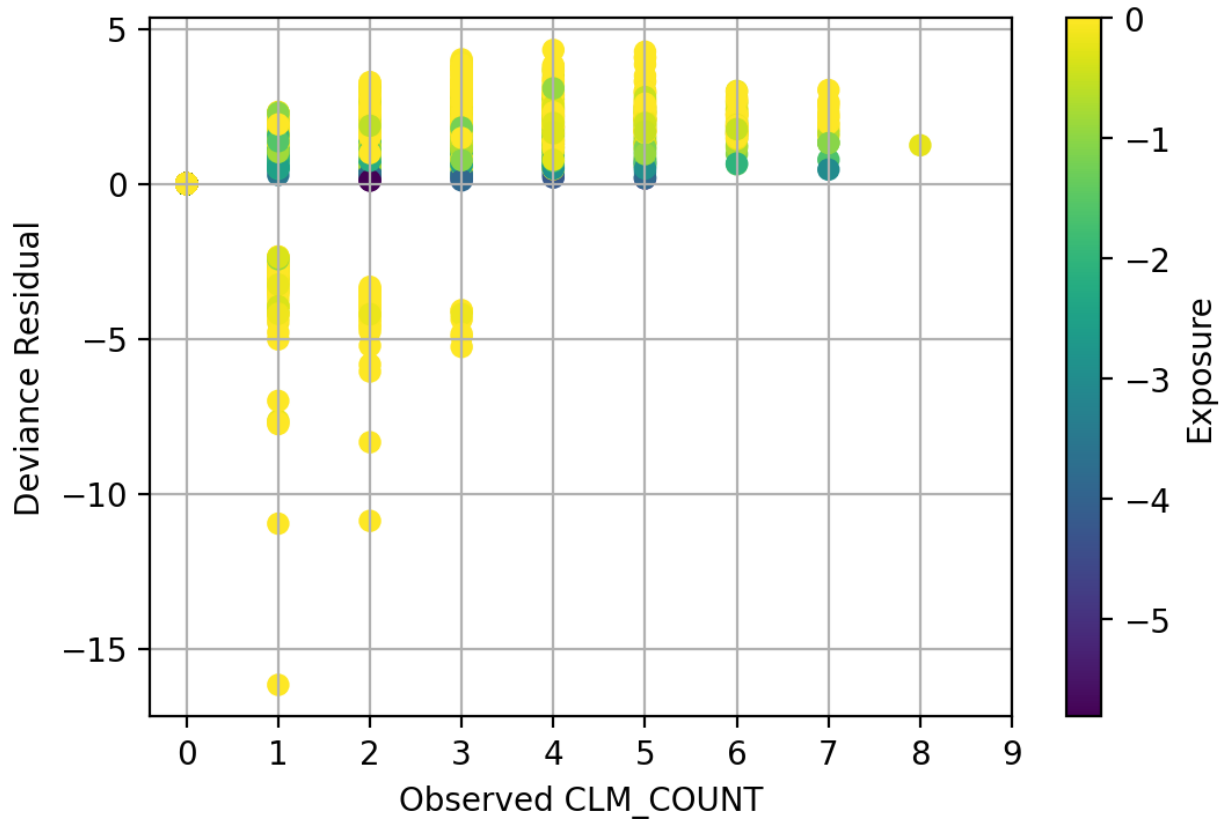
Question 3 (20 points)

You will visually assess your final model in Question 2. Please color-code the markers according to the Exposure value.

- a) (10 points). Please plot the predicted number of claims versus the observed number of claims.



b) (10 points). Please plot the Deviance residuals versus the observed number of claims.



Question 4 (20 points)

You will calculate the Accuracy metric to assess your final model in Question 2.

- a) (10 points). Please calculate the Root Mean Squared Error, the Relative Error, and the R-squared metrics.

Root Mean Squared Error = 1.516457528529213

Relative Error = 0.9947252006025616

Squared Correlation = 0.05899548511957755

- b) (10 points). Please comment on the Final Model based on the above three metrics and the diagnostic charts in Question 3.

The RMSE is 1.52. It is larger than 1. The relative error is 0.99. It is very high. The r-squared is very small. The charts look very random and chaotic with no relationships.

In conclusion, the model performs badly on the prediction.