For the frequency and before removing the outliers

| | | | Fre | quency | | | Frequ | iency | | Frequency | | | | |
|------------------|--|----------|----------|-------------|-----------|----------|----------|-----------|----------|-----------------|----------|-----------|----------|-------------------------|
| | | | | st dataset | | _ | on train | | | on all the data | | | | |
| | GLM before removing the outliers | MSE | RMSE | cor_kendall | cor | MSE | RMSE | cor_kenda | cor | MSE | RMSE | cor_kenda | cor | - |
| GLM | GLIVI Defore removing the outliers | - | - | - | - | - | - | - | - | 1.464265 | 1.210068 | 0.095883 | 0.152997 | |
| | Regression Tree before doing anything minsplit=5,cp=0 | 1.394636 | 1.180947 | 0.0600927 | 0.0538421 | 1.472254 | 1.213365 | 0.122743 | 0.217336 | 1.456726 | 1.206949 | 0.105736 | 0.182014 | |
| | Regression Tree after applying the best cp= 0.001477572 on the | 1.395869 | 1.181469 | 0.07498071 | 0.054052 | 1.475831 | 1.214838 | 0.129625 | 0.211896 | 1.459834 | 1.208236 | 0.114134 | 0.176846 | |
| egression | Regression Tree for minsplit=2,cp=0 | 1.386841 | 1.177642 | 0.06557614 | 0.05937 | 1.467036 | 1.211213 | 0.123529 | 0.225141 | 1.450992 | 1.204571 | 0.107191 | 0.190861 | |
| Tree | Regression Tree after applying the best cp=0.001391098 on the | 1.388601 | 1.178389 | 0.07524748 | 0.0588928 | 1.471148 | 1.212909 | 0.129644 | 0.219046 | 1.454633 | 1.206082 | 0.114205 | 0.184997 | |
| | Regression Tree after applying the minsplit=21 and cp=0.00026983 | 1.378074 | 1.173914 | 0.08388137 | 0.0694546 | 1.486956 | 1.219408 | 0.126044 | 0.194118 | 1.465173 | 1.210443 | 0.112147 | 0.165706 | |
| | Regression Tree after applying the minsplit=21 and the best | 1.369126 | 1.170097 | 0.08464738 | 0.0717499 | 1.491612 | 1.221316 | 0.118681 | 0.186195 | 1.467107 | 1.211242 | 0.106509 | 0.160813 | |
| | Boosting for cv.folds = 5 and n.trees = 1500 | 1.991979 | 1.411375 | 0.08600527 | 0.2483779 | 3.091508 | 1.758268 | 0.115837 | 0.236685 | 2.871531 | 1.694559 | 0.104722 | 0.23392 | Out-o |
| Boosting | Boosting for cv.folds = 5 and n.trees = 1000 | 1.983884 | 1.408504 | 0.08590221 | 0.2431216 | 3.080313 | 1.755082 | 0.118003 | 0.243601 | 2.860957 | 1.691436 | 0.105923 | 0.238505 | estima of ro |
| boosting | Boosting for cv.folds = 5 and n.trees = 2500 | 1.98328 | 1.40829 | 0.09045978 | 0.2454939 | 3.078401 | 1.754537 | 0.117309 | 0.244399 | 2.859307 | 1.690948 | 0.106176 | 0.239751 | mean square error |
| | Boosting for cv.folds = 5 and n.trees =5000 | 1.979979 | 1.407117 | 0.08562488 | 0.2473607 | 3.076882 | 1.754104 | 0.119409 | 0.241234 | 2.857431 | 1.690394 | 0.107105 | 0.237177 | |
| | Bagging for nbagg=200 | 1.307622 | 1.143513 | 0.1026546 | 0.289364 | 1.276701 | 1.129912 | 0.116586 | 0.417661 | 1.282888 | 1.132646 | 0.10906 | 0.385531 | 1.141 |
| Bagging | Bagging for nbagg=100 | 1.308101 | 1.143722 | 0.1057791 | 0.2895884 | 1.277785 | 1.130391 | 0.109463 | 0.416619 | 1.28385 | 1.133071 | 0.10377 | 0.38489 | 1.139 |
| Dugging | Bagging for nbagg=150 | 1.300662 | 1.140466 | 0.1012679 | 0.2885136 | 1.274221 | 1.128814 | 0.117511 | 0.419809 | 1.279511 | 1.131155 | 0.10918 | 0.388035 | 1.139 |
| | Bagging without anything | 1.310218 | 1.144648 | 0.1102993 | 0.2792626 | 1.277517 | 1.130273 | 0.112328 | 0.416911 | 1.28406 | 1.133164 | 0.106751 | 0.384327 | 1.137 |
| Random Forest | RF without doing anything | 1.222906 | 1.105851 | 0.08982141 | 0.3302648 | 1.164898 | 1.079304 | 0.221136 | 0.518324 | 1.176503 | 1.084667 | 0.191272 | 0.481074 | |
| | min | 1.222906 | 1.105851 | 0.0600927 | 0.0538421 | 1.164898 | 1.079304 | 0.109463 | 0.186195 | 1.176503 | 1.084667 | 0.095883 | 0.152997 | |
| | max | 1.991979 | 1.411375 | 0.1102993 | 0.3302648 | 3.091508 | 1.758268 | 0.221136 | 0.518324 | 2.871531 | 1.694559 | 0.191272 | 0.481074 | |

For the frequency and after removing the outliers

| | | | | quency | | | Frequ | | | Frequency on all the data | | | | |
|--------------------|--|----------|----------|------------|-----------|-----------|-----------|----------------------|----------|---------------------------|----------|----------------------|----------|--------------------|
| | | MSE | on tes | st dataset | cor | MSE | on train | dataset cor kenda | cor | MSE | on all t | ne data cor kenda | cor | |
| GLM | GLM after removing the outliers | - | - | - | - | - | - | - | - | 0.706502 | 0.840537 | 0.106876 | 0.184172 | |
| | Regression Tree before doing anything minsplit=5,cp=0 | 0.58386 | 0.764107 | 0.0718982 | 0.1094855 | 0.7374066 | 0.8587238 | 0.132813 | 0.220681 | 0.706694 | 0.840651 | 0.11417 | 0.197537 | |
| | Regression Tree after applying the best cp on the minsplit=5 | 0.584687 | 0.764648 | 0.07786129 | 0.1053745 | 0.7380955 | 0.8591248 | 0.140562 | 0.218635 | 0.70741 | 0.841077 | 0.120701 | 0.195174 | |
| Regression Tree | Regression Tree for minsplit=2,cp=0 | 0.583131 | 0.76363 | 0.07371094 | 0.1102565 | 0.7372217 | 0.8586162 | 0.133089 | 0.221227 | 0.7064 | 0.840476 | 0.114775 | 0.198383 | |
| | Regression Tree after applying the best cp on the minsplit=2 | 0.586956 | 0.766131 | 0.07834573 | 0.1023012 | 0.7392458 | 0.8597941 | 0.139004 | 0.215209 | 0.708784 | 0.841893 | 0.119469 | 0.190831 | |
| | Regression Tree after applying the minsplit=21 and cp=0.00026983 | 0.587735 | 0.766639 | 0.07768302 | 0.099048 | 0.7396185 | 0.8600108 | 0.138865 | 0.214088 | 0.709238 | 0.842163 | 0.11926 | 0.189333 | |
| | Boosting for cv.folds = 5 and n.trees = 1500 | 1.283892 | 1.133089 | 0.09546982 | 0.1356776 | 2.156165 | 1.468388 | 0.115024 | 0.194329 | 1.98169 | 1.407725 | 0.104176 | 0.179234 | Out-of- bag |
| Boosting | Boosting for cv.folds = 5 and n.trees = 1000 | 1.280659 | 1.131662 | 0.0920743 | 0.1347186 | 2.150828 | 1.46657 | 0.116256 | 0.194554 | 1.976774 | 1.405978 | 0.104613 | 0.179521 | estimat of root |
| boosting | Boosting for cv.folds = 5 and n.trees = 2500 | 1.284428 | 1.133326 | 0.09747634 | 0.1362288 | 2.153021 | 1.467318 | 0.1146 | 0.194011 | 1.979283 | 1.40687 | 0.104978 | 0.180157 | mean square |
| | Boosting for cv.folds = 5 and n.trees =5000 | 1.280782 | 1.131717 | 0.09581657 | 0.1366142 | 2.150673 | 1.466517 | 0.116203 | 0.194492 | 1.976675 | 1.405943 | 0.105304 | 0.179851 | error |
| | Bagging for nbagg=200 | 0.570341 | 0.755209 | 0.09649546 | 0.1065296 | 0.7565427 | 0.8697946 | 0.116489 | 0.154924 | 0.719298 | 0.848114 | 0.111541 | 0.144837 | 0.8702 |
| Bagging | Bagging for nbagg=100 | 0.569156 | 0.754424 | 0.07540802 | 0.1068434 | 0.7563938 | 0.869709 | 0.112941 | 0.155577 | 0.718942 | 0.847905 | 0.102415 | 0.145412 | 0.8701 |
| 20888 | Bagging for nbagg=150 | 0.57054 | 0.755341 | 0.07540802 | 0.1067641 | 0.7563663 | 0.8696932 | 0.112941 | 0.15567 | 0.719197 | 0.848055 | 0.102415 | 0.145464 | 0.8702 |
| | Bagging without anything | 0.571038 | 0.755671 | 0.09649546 | 0.1070048 | 0.7564406 | 0.8697359 | 0.116489 | 0.155401 | 0.719356 | 0.848149 | 0.111541 | 0.145316 | 0.8702 |
| Random Forest | RF without doing anything | 0.5704 | 0.755248 | 0.09059181 | 0.1198872 | 0.743051 | 0.8620041 | 0.125447 | 0.21238 | 0.708517 | 0.841735 | 0.111432 | 0.192745 | |
| | min | 0.569156 | 0.754424 | 0.0718982 | 0.099048 | 0.7372217 | 0.8586162 | 0.112941 | 0.154924 | 0.7064 | 0.840476 | 0.102415 | 0.144837 | |
| | max | 1.284428 | 1.133326 | 0.09747634 | 0.1366142 | 2.156165 | 1.468388 | 0.140562 | 0.221227 | 1.98169 | 1.407725 | 0.120701 | 0.198383 | |

For the severity and before removing the outliers

| | | | Sev | erity | | | Seve | erity | | | | | | |
|------------------|---|-----------|----------|-----------|-----------|-----------|----------|-----------|----------|-----------|----------|-----------|----------|------------------|
| | | | on test | dataset | | | on train | dataset | | | | | | |
| | | MSE | RMSE | cor_kenda | cor | MSE | RMSE | cor_kenda | cor | MSE | RMSE | cor_kenda | cor | |
| GLM | GLM before removing the outliers | - | - | - | - | - | - | - | - | 286061616 | 16913.36 | 0.367428 | 0.304166 | |
| | Regression Tree before doing anything minsplit=5,cp=0 | 145588789 | 12066.02 | 0.205405 | 0.1005823 | 224106061 | 14970.17 | 0.492513 | 0.621318 | 208397553 | 14435.98 | 0.435212 | 0.582513 | |
| | Regression Tree after applying the best cp= | - | - | - | - | - | - | - | - | - | - | - | - | |
| Regressi | Regression Tree for minsplit=2,cp=0 | 164676262 | 12832.62 | 0.19149 | 0.0888317 | 187077822 | 13677.64 | 0.538887 | 0.698197 | 182596068 | 13512.81 | 0.465464 | 0.650031 | |
| on Tree | Regression Tree after applying the best | - | - | - | - | - | - | - | - | - | - | - | - | |
| | Regression Tree after applying the minsplit | 110185442 | 10496.93 | 0.214537 | 0.2219333 | 245927630 | 15682.08 | 0.284176 | 0.571185 | 218770455 | 14790.89 | 0.27028 | 0.552047 | |
| | Regression Tree after applying the minsplit | 112346964 | 10599.39 | 0.215379 | 0.2115415 | 288877922 | 16996.41 | 0.284542 | 0.456712 | 253560368 | 15923.58 | 0.270682 | 0.440774 | |
| | Boosting for cv.folds = 5 and n.trees = 1500 | 219959822 | 14831.04 | 0.277106 | 0.0945383 | 284481452 | 16866.58 | 0.376981 | 0.47001 | 271572973 | 16479.47 | 0.359428 | 0.392282 | Out-of- bag |
| Boosting | Boosting for cv.folds = 5 and n.trees = 1000 | 203290994 | 14258.02 | 0.276774 | 0.1022091 | 286128850 | 16915.34 | 0.373467 | 0.46618 | 269555947 | 16418.16 | 0.356779 | 0.39218 | estimate of root |
| boosting | Boosting for cv.folds = 5 and n.trees = 2500 | 204135421 | 14287.6 | 0.281073 | 0.0988806 | 286795097 | 16935.03 | 0.375504 | 0.464666 | 270257841 | 16439.52 | 0.359143 | 0.388779 | mean squared |
| | Boosting for cv.folds = 5 and n.trees =5000 | 211834321 | 14554.53 | 0.276168 | 0.0948335 | 286457526 | 16925.06 | 0.373288 | 0.464296 | 271528082 | 16478.11 | 0.356504 | 0.389549 | error |
| | Bagging for nbagg=200 | 170111718 | 13042.69 | 0.238171 | 0.09061 | 274647002 | 16572.48 | 0.33146 | 0.504862 | 253733217 | 15929.01 | 0.31485 | 0.440334 | 18147.55 |
| Bagging | Bagging for nbagg=100 | 200819579 | 14171.08 | 0.236024 | 0.0708895 | 274232991 | 16559.98 | 0.330479 | 0.503896 | 259545584 | 16110.42 | 0.313672 | 0.423455 | 18291.45 |
| 24888 | Bagging for nbagg=150 | 170469706 | 13056.4 | 0.235959 | 0.089481 | 274823305 | 16577.8 | 0.329517 | 0.502999 | 253945868 | 15935.68 | 0.312887 | 0.439865 | 18115.16 |
| | Bagging without anything | 173851963 | 13185.29 | 0.235269 | 0.0882576 | 272231768 | 16499.45 | 0.326124 | 0.506988 | 252549475 | 15891.81 | 0.309771 | 0.446388 | 18151.95 |
| Random Forest | RF without doing anything | 159249284 | 12619.4 | 0.278287 | 0.1249686 | 136893548 | 11700.15 | 0.592574 | 0.849788 | 141366134 | 11889.75 | 0.532152 | 0.772363 | |
| | min | 110185442 | 10496.93 | 0.19149 | 0.0708895 | 136893548 | 11700.15 | 0.284176 | 0.456712 | 141366134 | 11889.75 | 0.27028 | 0.304166 | |
| | max | 219959822 | 14831.04 | 0.281073 | 0.2219333 | 288877922 | 16996.41 | 0.592574 | 0.849788 | 286061616 | 16913.36 | 0.532152 | 0.772363 | |

For the severity and after removing the outliers

| | | | | erity | | | | erity | | | | | | |
|------------------|---|----------|----------|-----------|-----------|----------|----------|-----------|----------|----------|----------|-----------|----------|---|
| | | | on test | | | | dataset | | | | | | | |
| | | MSE | RMSE | cor_kenda | cor | MSE | RMSE | cor_kenda | cor | MSE | RMSE | cor_kenda | cor | |
| GLM | GLM after removing the outliers | - | - | - | - | - | - | - | - | 72258049 | 8500.473 | 0.416373 | 0.490916 | |
| | Regression Tree before doing anything minsplit=5,cp=0 | 52874131 | 7271.46 | 0.272738 | 0.2779398 | 37630376 | 6134.36 | 0.538635 | 0.659053 | 40680176 | 6378.101 | 0.487916 | 0.597603 | |
| Regressi | Regression Tree after applying the best cp= | 41367240 | 6431.737 | 0.316068 | 0.3903493 | 49667485 | 7047.516 | 0.428713 | 0.5034 | 48006865 | 6928.699 | 0.409662 | 0.486022 | |
| _ | Regression Tree for minsplit=2,cp=0 | 59350892 | 7703.953 | 0.260143 | 0.2471022 | 35020053 | 5917.774 | 0.573571 | 0.688177 | 39887896 | 6315.686 | 0.511387 | 0.611473 | |
| | Regression Tree after applying the best cp on the | 43176484 | 6570.882 | 0.336572 | 0.3566808 | 45837628 | 6770.349 | 0.444038 | 0.557657 | 45305216 | 6730.915 | 0.427387 | 0.528622 | |
| Boosting | Boosting for cv.folds = 5 and n.trees = 1500 | 39736243 | 6303.669 | 0.314056 | 0.4181923 | 50102434 | 7078.307 | 0.412332 | 0.496887 | 48028482 | 6930.258 | 0.398003 | 0.485319 | Out-of- bag |
| | Boosting for cv.folds = 5 and n.trees = 1000 | 39788592 | 6307.82 | 0.312977 | 0.4175727 | 50122936 | 7079.755 | 0.411876 | 0.496557 | 48055356 | 6932.197 | 0.397083 | 0.484933 | estimate of root mean squared error |
| | Boosting for cv.folds = 5 and n.trees = 2500 | 39833241 | 6311.358 | 0.312137 | 0.4160701 | 50116247 | 7079.283 | 0.41266 | 0.49669 | 48058938 | 6932.455 | 0.39755 | 0.484815 | |
| | Boosting for cv.folds = 5 and n.trees =5000 | 39833303 | 6311.363 | 0.313758 | 0.4159867 | 50105915 | 7078.553 | 0.413547 | 0.496852 | 48050686 | 6931.86 | 0.398714 | 0.484965 | |
| | Bagging for nbagg=200 | 41467257 | 6439.507 | 0.311964 | 0.3815496 | 49848426 | 7060.342 | 0.42168 | 0.501405 | 48171615 | 6940.577 | 0.403283 | 0.483044 | 7167.048 |
| Bagging | Bagging for nbagg=100 | 41624610 | 6451.714 | 0.309787 | 0.3792241 | 49748673 | 7053.274 | 0.421385 | 0.502902 | 48123301 | 6937.096 | 0.402638 | 0.483813 | 7161.661 |
| Dugging | Bagging for nbagg=150 | 41373853 | 6432.251 | 0.318118 | 0.3837749 | 49829089 | 7058.972 | 0.422884 | 0.501597 | 48137460 | 6938.116 | 0.406084 | 0.483556 | 7166.858 |
| | Bagging without anything | 41307907 | 6427.123 | 0.315103 | 0.3846659 | 49943231 | 7067.052 | 0.429441 | 0.500223 | 48215571 | 6943.743 | 0.410046 | 0.482372 | 7189.949 |
| Random Forest | RF without doing anything | 41488283 | 6441.14 | 0.328869 | 0.3803677 | 41390263 | 6433.527 | 0.511783 | 0.622385 | 41409874 | 6435.05 | 0.481945 | 0.588058 | |
| | min | 39736243 | 6303.669 | 0.260143 | 0.2471022 | 35020053 | 5917.774 | 0.411876 | 0.496557 | 39887896 | 6315.686 | 0.397083 | 0.482372 | |
| | max | 59350892 | 7703.953 | 0.336572 | 0.4181923 | 50122936 | 7079.755 | 0.573571 | 0.688177 | 72258049 | 8500.473 | 0.511387 | 0.611473 | |