Data Issues:

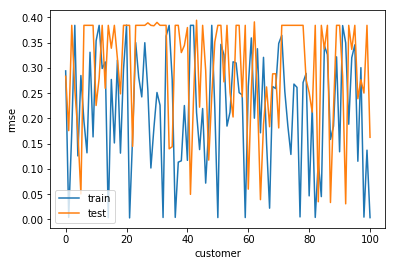
Dates in fv\_cost column

Always problem with this row, Do need to check before giving as input to the file

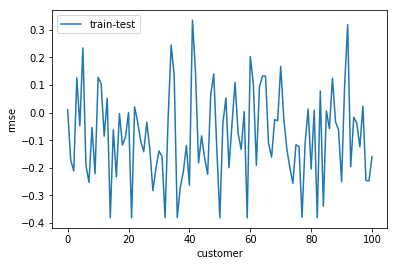




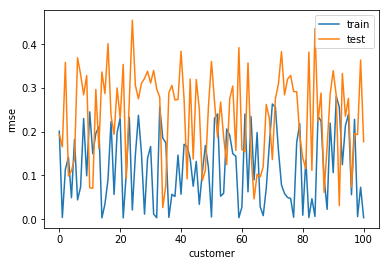
Observations:



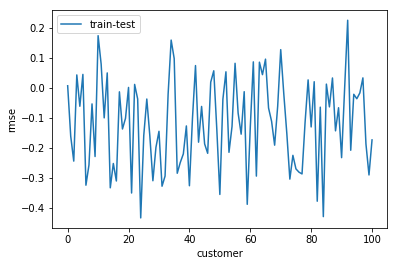
Train-Test rmse of model



## When model is created for each customer differently

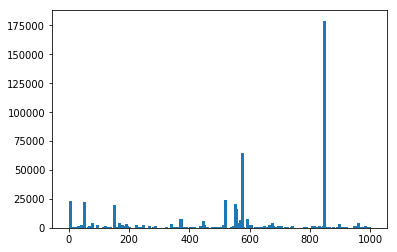


Train\_rmse – test\_rmse ,-ve-> overfitting for train data



# timeseries model for each customer

Test rmse values over 1000 customers



Sorted rmse values in descending order for same data



Count of number of customers in different bins

Actual Values test rmse bin values count of customers in bins Numbers:

(0, 100] 760

(100, 200] 61

(200, 300] 39

(400, 500] 21

(300, 400] 18

(500, 600] 12

(700, 800] 11

(800, 900] 8

(1400, 1500] 6

(600, 700] 4

(1000, 1100] 4

(1100, 1200] 4

(2000, 2100] 4

(1300, 1400] 4

(1500, 1600] 3

(2400, 2500] 3

(3900, 4000] 3

(900, 1000] 3

(1600, 1700] 2

(3100, 3200] 2

(1200, 1300] 2

(7800, 7900] 1

(4300, 4400] 1

(4600, 4700] 1

(5700, 5800] 1

(7000, 7100] 1

(7600, 7700] 1

(3700, 3800] 1

(65000, 65100] 1

(23200, 23300] 1

...

(131400, 131500] 0

(134200, 134300] 0

(131500, 131600] 0

(131600, 131700] 0

(131700, 131800] 0

(131800, 131900] 0

(131900, 132000] 0

(132000, 132100] 0

(132100, 132200] 0

(132200, 132300] 0

(132300, 132400] 0

(132400, 132500] 0

(132500, 132600] 0

(132600, 132700] 0

(132700, 132800] 0

(132800, 132900] 0

(132900, 133000] 0

(133000, 133100] 0

(133100, 133200] 0

(133200, 133300] 0

(133300, 133400] 0

(133400, 133500] 0

(133500, 133600] 0

(133600, 133700] 0

(133700, 133800] 0

(133800, 133900] 0

(133900, 134000] 0

(134000, 134100] 0

(134100, 134200] 0

(100000, 100100] 0

Scaled sorted Test rmse values for 1000 customers



Scaled test rmse bin values and count of customers

binedge\_1 binedge\_2 count

0 0.000003 0.056908 414.0

1 0.056908 0.113814 188.0

2 0.113814 0.170720 229.0

3 0.170720 0.227625 135.0

4 0.227625 0.284531 31.0

5 0.284531 0.341437 1.0

6 0.341437 0.398342 0.0

7 0.398342 0.455248 2.0

Customer who has test error ~ 175000 is 1681085



Predictions for this customer

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | original\_fv\_scaled | prediction\_scaled | prediction\_actual | fv\_actual |
| 0 | -0.53797703 | -0.5550777 | -17.809702 | -6.2491455 |
| 1 | -0.512276477 | -0.53423595 | -17.34842 | -5.681421 |
| 2 | -0.486572489 | -0.5127758 | -16.873447 | -5.1136206 |
| 3 | -0.460865065 | -0.4888636 | -16.344208 | -4.5457443 |
| 4 | -0.435154204 | -0.46729577 | -15.866854 | -3.9777921 |
| 5 | -0.409439906 | -0.44604802 | -15.3965845 | -3.4097639 |
| 6 | -0.38372217 | -0.42317438 | -14.890329 | -2.8416598 |
| 7 | -0.871771776 | -0.39857876 | -14.3459635 | -13.622663 |
| 8 | -0.846115846 | -0.8642999 | -24.653616 | -13.055924 |
| 9 | -0.820456485 | -0.8320621 | -23.940107 | -12.489109 |
| 10 | -0.794793695 | -0.8059194 | -23.3615 | -11.922219 |
| 11 | -0.769127474 | -0.77628255 | -22.705557 | -11.355253 |
| 12 | -0.743457821 | -0.7526567 | -22.182653 | -10.788211 |
| 13 | -0.717784738 | -0.7278898 | -21.634495 | -10.221093 |
| 14 | -0.692108222 | -0.7053344 | -21.135284 | -9.6538994 |
| 15 | -0.666428274 | -0.67806804 | -20.531807 | -9.0866301 |
| 16 | -0.640744893 | -0.6495638 | -19.900932 | -8.5192849 |
| 17 | -0.717289935 | -0.6217407 | -19.285133 | -10.210163 |
| 18 | -0.665874983 | -0.705199 | -21.132288 | -9.0744079 |
| 19 | -0.614453158 | -0.6495764 | -19.901213 | -7.9385012 |
| 20 | -0.563024459 | -0.597208 | -18.742159 | -6.8024426 |
| 21 | -0.511588885 | -0.55441606 | -17.79506 | -5.6662321 |
| 22 | -0.460146434 | -0.5117959 | -16.85176 | -4.5298698 |
| 23 | -0.408697107 | -0.46638322 | -15.846657 | -3.3933555 |
| 24 | -0.357240902 | -0.42205906 | -14.865644 | -2.2566893 |
| 25 | -0.305777819 | -0.36991775 | -13.711618 | -1.1198712 |
| 26 | -0.254307855 | -0.32137907 | -12.637327 | 0.01709894 |
| 27 | 0.323588936 | -0.27163386 | -11.536333 | 12.7828236 |
| 28 | 0.398705542 | 0.26314726 | 0.29980132 | 14.4421474 |
| 29 | 0.47383219 | 0.34467438 | 2.1042142 | 16.1016931 |
| 30 | 0.54896888 | 0.4066977 | 3.4769557 | 17.7614605 |
| 31 | 0.624115614 | 0.5105787 | 5.7761197 | 19.4214499 |
| 32 | 0.699272395 | 0.60120535 | 7.7819295 | 21.0816612 |
| 33 | 0.774439222 | 0.6815326 | 9.559787 | 22.7420944 |
| 34 | 0.849616098 | 0.7528973 | 11.139278 | 24.4027495 |
| 35 | 0.924803023 | 0.8223609 | 12.676693 | 26.0636267 |
| 36 | 1 | 0.8716454 | 13.76749 | 27.7247259 |
| 37 | -0.255081915 | 0.91584885 | 14.7458315 | 0 |
| 38 | -0.255081915 | -0.30336213 | -12.2385645 | 0 |
| 39 | -0.255081915 | -0.29389024 | -12.028926 | 0 |
| 40 | -0.255081915 | -0.27482378 | -11.606935 | 0 |
| 41 | -0.255081915 | -0.27482367 | -11.606932 | 0 |
| 42 | -0.255081915 | -0.27482367 | -11.606932 | 0 |
| 43 | -1 | -0.27482367 | -11.606932 | -16.455221 |
| 44 | -0.974361211 | -0.9522339 | -26.59983 | -15.88886 |
| 45 | -0.948718994 | -0.95157003 | -26.585136 | -15.322425 |
| 46 | -0.92307335 | -0.9520259 | -26.595226 | -14.755913 |
| 47 | 0.5468936 | -0.9032035 | -25.514656 | 17.7156177 |
| 48 | -0.126899899 | 0.52349234 | 6.061933 | 2.83153728 |
| 49 | -0.051843515 | -0.15784287 | -9.017836 | 4.4895308 |
| 50 | 0.023222902 | -0.11579627 | -8.087232 | 6.14774596 |
| 51 | 0.098299355 | -0.042918026 | -6.4742413 | 7.80618279 |
| 52 | 0.173385844 | 0.03023696 | -4.855126 | 9.46484132 |
| 53 | 0.248482371 | 0.12073773 | -2.852102 | 11.1237216 |
| 54 | -0.202831012 | 0.18718463 | -1.3814547 | 1.15422104 |
| 55 | -0.151347287 | -0.23756051 | -10.7822 | 2.29149516 |
| 56 | -0.563674147 | -0.18678576 | -9.658419 | -6.8167942 |
| 57 | -0.53797703 | -0.5576898 | -17.867516 | -6.2491455 |
| 58 | -0.512276477 | -0.5367069 | -17.403109 | -5.681421 |
| 59 | -0.486572489 | -0.5127758 | -16.873447 | -5.1136206 |
| 60 | -0.460865065 | -0.4888636 | -16.344208 | -4.5457443 |
| 61 | -0.255081915 | -0.46729577 | -15.866854 | 0 |
| 62 | -0.255081915 | -0.26806188 | -11.457275 | 0 |
| 63 | -0.255081915 | -0.27123535 | -11.5275135 | 0 |

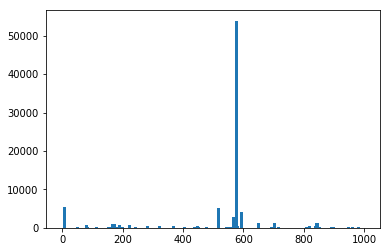
# All Above results became obsolete after changing the scaling objects in the code

Now the Max test rmse is around 50000 for 7902786 customer out of 100 customers



Since in Train dataset customer was always in credit and in test set he was in debt, Model was not performing well for this customer.

For 1000 customer actual test rmse



Scaled test rmse for 1000 customers



Model on cluster 6 data has 483 customers ,without merging and when test data Is just 2017 dec data

Actual test rmse distribution

