Anomalies in a Coal Quality data

A group project by
Soham Das, Arjama Das and Aditya Ghosh,
under the supervision of Prof. Debashis Sengupta

Indian Statistical Institute, Kolkata
August, 2020

1 Introduction

This project is based on a *coal quality data* provided by *Coal India Limited (CIL)*. The grade of non-coking coal samples and the values of various quality parameters, determined independently by CIL's laboratory and by designated Third Parties ideally should be more or less similar. However, the measurements have been reported to vary significantly (worse than that of CIL's values) which reduces the declared grade, resulting in loss of revenue for CIL. The data analysis aims to identify the anomalies and the factors causing them, so as to take appropriate measures. In addition, the data on cases referred to the arbitrators needs to be investigated for differences of Referral measurements with those by CIL and Third Parties. However it is to be noted apriori that Referral Data might be affected by a possible selection bias.

2 Data Description

The data under consideration comprises two data sets: Third Party data and Referral data. The Third Party data has total 1,01,226 coal samples. Each sample has source, seller and buyer details along with calorific value and coal grades. Only a few of these samples (which are considered to have variations in coal grades) were sent to referrals. The Referral data consists of those samples which were sent to referrals. It has 4,605 coal samples. Below we discuss different components of the dataset.

Gross Calorific Value (GCV) and Coal Grades

GCV (Gross Calorific Value) of a coal sample determines the quality and grades are assigned for the samples accordingly. There are three types GCVs and coal grades:
Lab GCV: The GCV determined by Coal India's laboratory by sampling from outgoing

shipments.

Third Party GCV: The GCV determined by the designated Third Party in case of dispute between Coal India and the buyer.

Referral GCV: The GCV assigned by arbitrators in case of further dispute (provided for a much smaller number of cases which were referred).

Grades: Three kinds of grades corresponding to the above three kinds of GCVs are also provided. Some Lab and Third party grades were wrongly allocated, but we could correct them because of the grades having a direct relationship with the GCVs.

Subsidiary, Area, Buyer and Plant

Sub: The subsidiary of CIL from which the sample is being generated, having 8 levels.

Area: The area from which the coal sample is coming from, having 77 levels.

Buyer: The Agency buying the coal from Coal India, having 152 levels.

Plant: The plant buying the coal, having 202 levels.

Siding and Source unit

Siding: The rail siding transporting the coal, having 160 levels.

Source unit: The mining project under which the coal sample is being produced, having 318 levels.

Agency, Sector and Mode of dispatch

Agency: The Third Party Lab testing the coal sample, having 6 levels.

Sector: The sector for which the coal sample is manufactured, having 10 levels.

Mode of dispatch: Mode of transportation of coal.

Figure 1 shows a hierarchy for the main factors involved in this process.

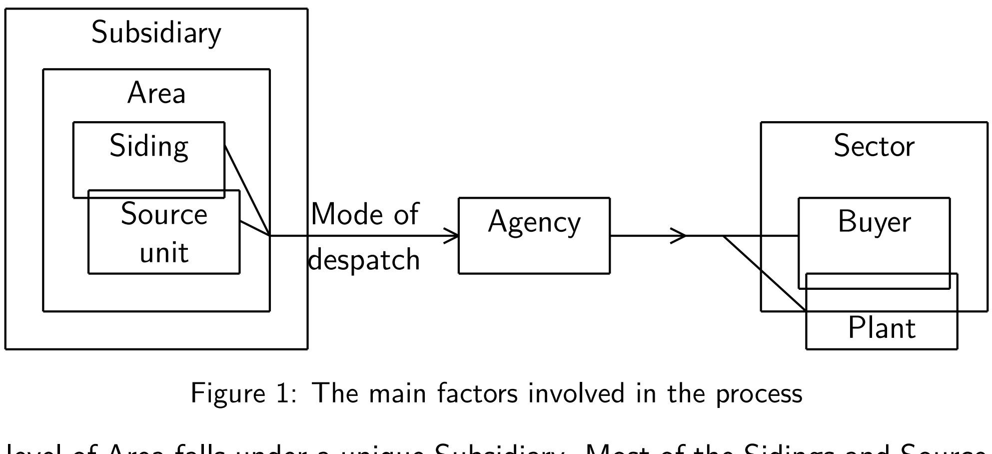


Figure 1: The main factors involved in the process

Each level of Area falls under a unique Subsidiary. Most of the Sidings and Source units come not only under a unique Subsidiary, but also under a unique Area. But this is not the case for Buyer and Plant, they overlap with subsidiaries. Also, there are common buyer for different plants and common plant for different buyers. However, each buyer belongs to a unique sector.

Note: Factors with less number of levels are: Agency (6), Despatch (6), Sub (8), Sector (10), and factors with many levels are: Area (77), Siding (160), Buyer (152), Plant (202), Source unit (318). For Sector and Siding, one level is '–', which possibly accounts for missing data. When we remove this level for Sector and Siding, we lose 6 levels for Area, 6 levels for Plant, 2 levels for Sector, 35 levels of Buyer and 20 levels for Source unit.

Moisture and Ash content

Some other variables like Moisture, Total Moisture and Ash content are also mentioned. It is known that the variable Moisture does not have any relation with the quality for non-coking coal. Hence we can ignore the columns on Moisture for our Analysis.

Regarding Ash content, an interesting observation is that there is a significant correlation between Ash content and GCV values of the samples. For the Third Party data, we found the correlation between Lab Ash and Lab GCV to be -0.92, and the correlation between Third Party Ash and Third Party GCV to be -0.91. Now it is quite understandable that Ash content is directly related to the quality of coal and hence to the GCV values. But since our primary goal is to find anomalies, Ash content won't be a good variable to consider because we can not determine whether the strong association between Ash content and GCV present in the samples is due to the natural cases, or were intentionally manipulated to cover up for the anomalies.

3 Which variables to consider?

Response variable: We will consider difference of GCVs as our primary response variables. A similar analysis was done before using disagreement of grades as response variable. We chose GCVs over Grades because GCVs are more informative than the Grades. Explanatory variables:

- Subsidiary (8 levels)
- Plant (202 levels)
- Siding (160 levels)
- Area (77 levels)
- Buyer (152 levels)
- Source unit (318 levels)

Other variables: Agency (6 levels, CIMFR: 95%), Sector (10 levels, Power: 96%) and Mode of despatch (6 levels, Rail: 83%). For each of these factors, one of its levels carries a huge percentage of the total frequency. Hence for our purpose of detecting anomalous levels, we should leave out these factors.

4 Are disagreements really there?

For the Third Party data, we may consider the difference GCV$_{T-L}$:= (Third Party GCV − Lab GCV) as our response variable. The plot in Figure 2 shows that for a major proportion of the Third Party data, Third Party GCVs are lower than Lab GCVs, indicating a degradation in terms of quality. We further observe that usually for samples with lower (better) Lab Grades, the variable GCV$_{T-L}$ is mostly negative.

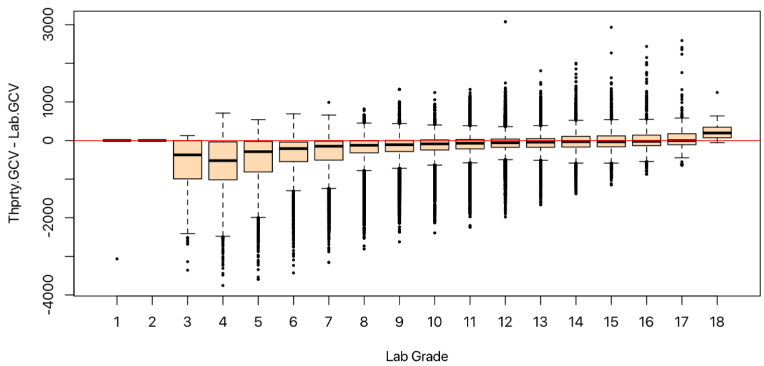


Figure 2: Boxplot of GCV$_{T-L}$ vs Lab Grade for the Third Party data

To assess the disagreement between the Third Party Grades and the Lab Grades, we can use the variable $y = I_{[\text{Lab Grade}<\text{Third Party Grade}]} - I_{[\text{Lab Grade}>\text{Third Party Grade}]}$. Note that $y = 1$ when GCV$_{T-L} > 0$, $y = 0$ when GCV$_{T-L} = 0$ and $y = -1$ when GCV$_{T-L} < 0$. (Recall that higher the GCV, lower the Grade.) We found the following distribution of y from the samples in the Third Party data:

y	-1	0	1
%age	41.77879	47.46113	10.76008

The above table shows that for the Third Party data, a large proportion of samples were found to have Lab Grades lower (better) than Third Party Grades, which agrees with our earlier observation. To support this observation statistically, we performed the one sample t-test on GCV$_{T-L}$ and the Wilcox test on y, and both of these were rejected with p-values less than 2.2×10^{-16}. We also performed these tests separately for each grade (both Third Party and Lab). Except for samples with Third Party Grade 1 or Lab Grades $1, 2$ (for which the two types of GCVs are degenerate and possibly faulty) and Lab Grade 15, the p-values of all of these tests were found to be less than 10^{-5}.

It can thus be said that **either Coal India's lab tends to overrate the coal quality than the Third Parties, or the Third Parties tend to underrate the coal quality. But in either case, it is a loss for Coal India,** because the pricing is done on the basis of Third Party Grades.

The same trend is reflected for Referral Data as well, but the proportion of samples with Lab Grade lower than Third Party Grade increases here:

y	-1	0	1
%age	76.199783	17.090119	6.710098

To avoid confusion, we shall call the variable y for Referral data as y_1. We also define y_2 and y_3 as follows: $y_2 = I_{[\text{Lab Grade}>\text{Ref Grade}]} - I_{[\text{Lab Grade}<\text{Ref Grade}]}$ and $y_3 = I_{[\text{Third Party Grade}>\text{Ref Grade}]} - I_{[\text{Third Party Grade}<\text{Ref Grade}]}$. On the other hand, the variables GCV$_{R-T}$ and GCV$_{R-L}$ are defined just as GCV$_{T-L}$. Figure 3 shows the boxplot of different GCV differences against different Grades for the Referral data.

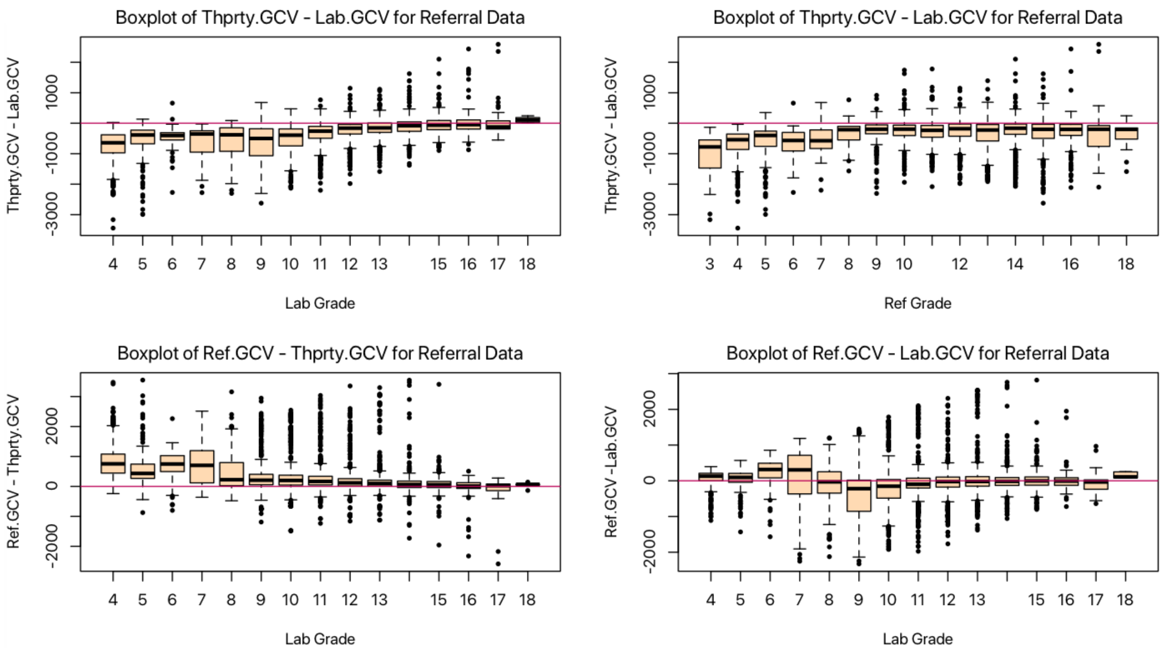


Figure 3: Boxplot of different GCV differences for the Referral data

From these boxplots, we observe the following:

- For samples with lower Lab Grade, the Referral GCV is typically found to be higher than the Third Party GCV, which suggests that the samples were indeed down-graded by the Third Party. This is also supported by one sample t-tests on GCV_{T-R} separately for each Lab Grade.
- The boxplot of GCV_{T-L} for different Referral Grades shows that irrespective of whether the Ref. Grade is lower or higher, the quantity GCV_{T-L} is usually negative. This is understandable, because these are the samples that actually went to arbitration for the dispute.
- The Ref. GCV is typically higher than the Third Party GCV, which shows that the samples going to arbitration was somewhat in favour of CIL, and also suggests that the Third Party Grading might be biases to the negative side.
- For samples with lower (better) Lab Grades, the Referral GCV is usually higher than the Lab grade.

- Each of the above boxplots shows the presence of a large number of outliers.

The above observations are also supported by one sided t-tests on the variables GCV$_{T-L}$, GCV$_{R-L}$, GCV$_{R-T}$, done both for the whole data as well as for each Lab grade. (Note that we defined these 3 variables in such a way that whenever any of them is negative, it is 'bad' for CIL.) Thus, we not only find disagreements between the grading of the Third Party and the Lab, but also between Referral and any of those parties. Therefore we should do a thorough study and find abnormalities present in the data as much as possible.

5 Some visual tools

First we visually assess the difference between Third Party and Lab GCVs by breaking down the Third Party data using different factors, e.g. Sub, Area etc. Figure 4 below shows the disparity between these two kinds of GCVs at the subsidiary level. Here we have sorted the Third Party data according to Lab GCVs, and that index is shown on the x-axis. Then at every index we plot both the Third Party GCV and the Lab GCV, and join the two points using a vertical line of green color if GCV$_{T-L} > 0$ and red color if GCV$_{T-L} < 0$. Obviously, the more it is red, the more it is bad for CIL.

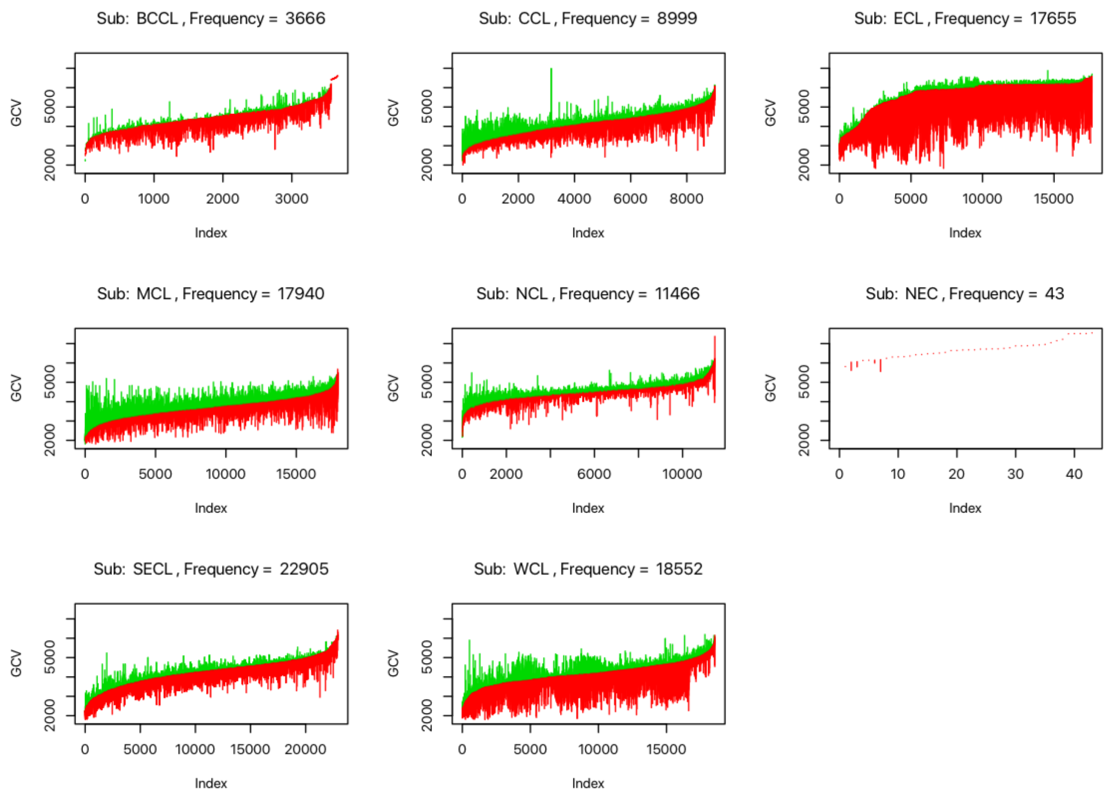


Figure 4: The difference GCV_{T-L} at the Subsidiary level in the Third Party data

We observe from Figure 4 that subsidiaries ECL and WCL have the most red (with considerably higher magnitudes), BCCL and SECL also have more red than green (but with lower magnitude), while for MCL and NCL the green and the red are more or less equally distributed. For NEC we do have all red, but we see many dots here because of the degeneracy in the two types of GCVs.

The picture at subsidiary level becomes more clear with the Referral data, when we do a 3D plot of the three types of GCVs and color the points by using one color for each subsidiaries. This plot is shown in Figure 5. It shows some clustering on the basis of subsidiaries and suggests us to do a separate analysis for each subsidiary, especially for the subsidiaries ECL, WCL and MCL. Since these subsidiaries constitute 95% of the Referral data, it is no surprise that they would stand out from other subsidiaries. But what strikes us most is that the clusters from these 3 subsidiaries are very distinctive.

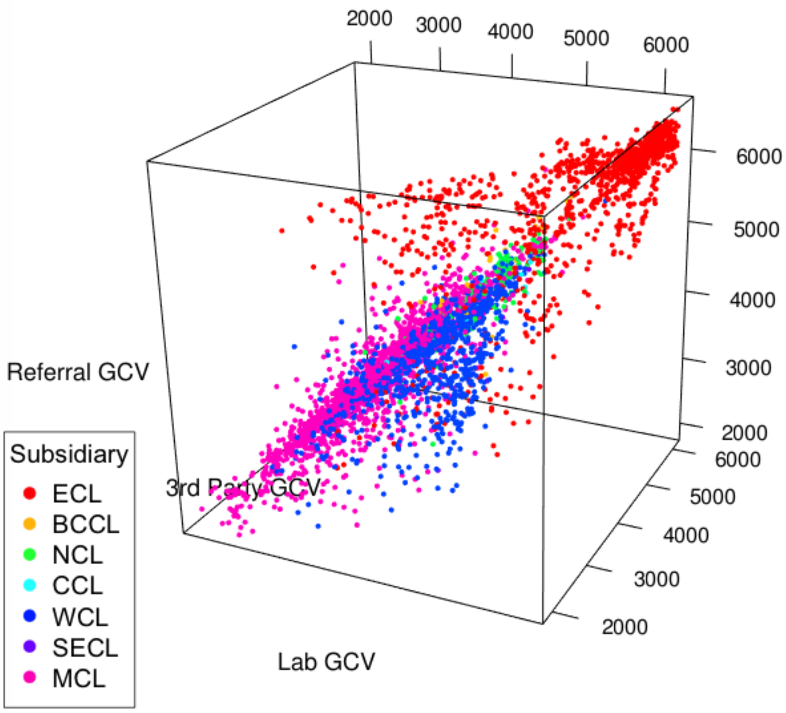


Figure 5: A 3D plot of the 3 types of GCVs in the Referral data. A unique color is used for each Subsidiary which reveals to a clustering.

We next dive into the Area level. Recall that each area falls under a unique subsidiary. In Figure 6 the plot of the average GCV difference (GCV$_{T-L}$) is shown for each area and a unique color is used for each subsidiary. Also, the area of each bubble represents the number of samples from that area. Among the two horizontal dotted lines, the lower one represents the average value of GCV$_{T-L}$ over the whole data. This plot tells us that in the Third Party data, areas from subsidiaries ECL, SECL and WCL typically have the Third Party GCV much lower than the Lab GCV. It also shows that for MCL and NCL, the samples are more or less equally distributed along the horizontal line at level 0.

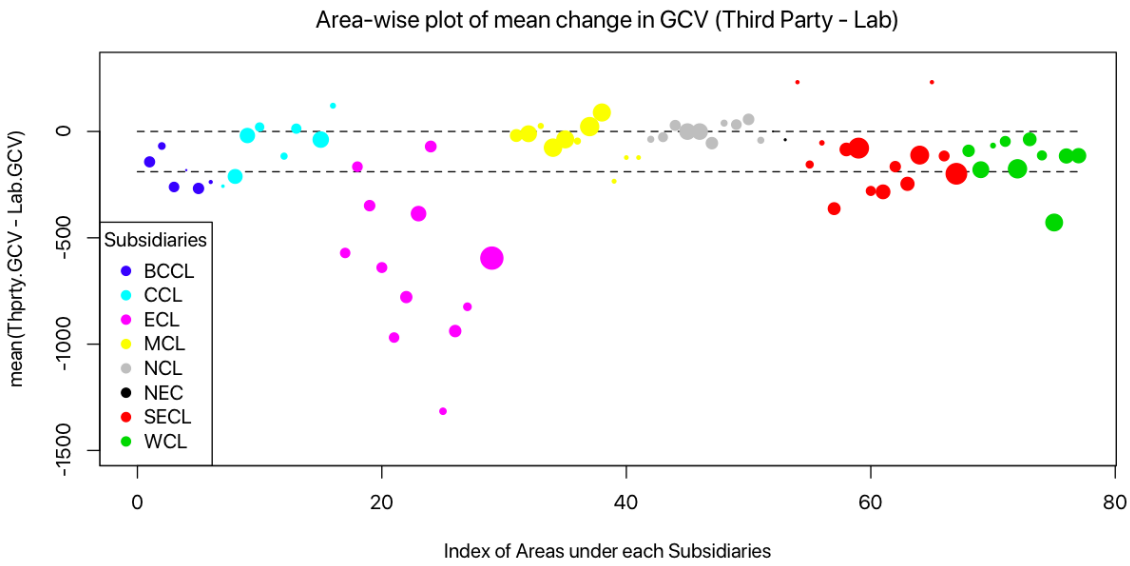


Figure 6: Plot of the average value of GCV_{T-L} for each Area in the Third Party data. Area of the bubbles are proportional to the number of samples from that Area and the colors of the bubbles corresponds to different subsidiaries.

Next we do similar bubble plots for the Referral data, with each of the three types of GCV differences.

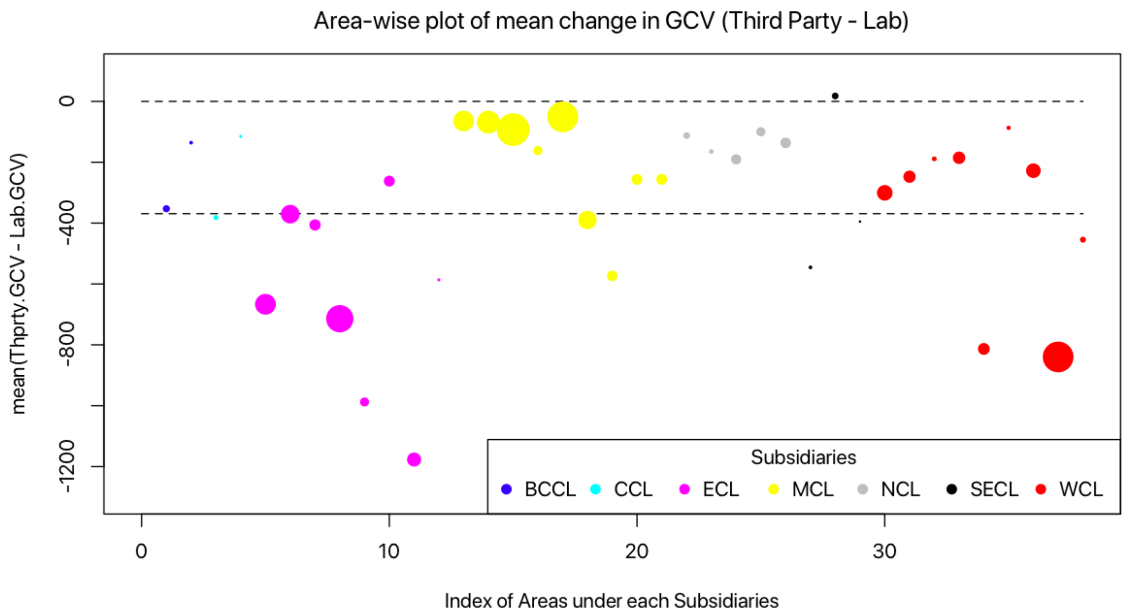


Figure 7: Plot of the average value of GCV_{T-L} for each Area in the Referral data.

Figure 7 shows that except for one area from SECL, the average value of GCV_{T-L} is negative for all the areas. Also note that some areas within the subsidiaries ECL and WCL have this average extremely low. Figures 8 and 9 show the area-wise means of the GCV differences GCV_{R-T} and GCV_{R-L} respectively.

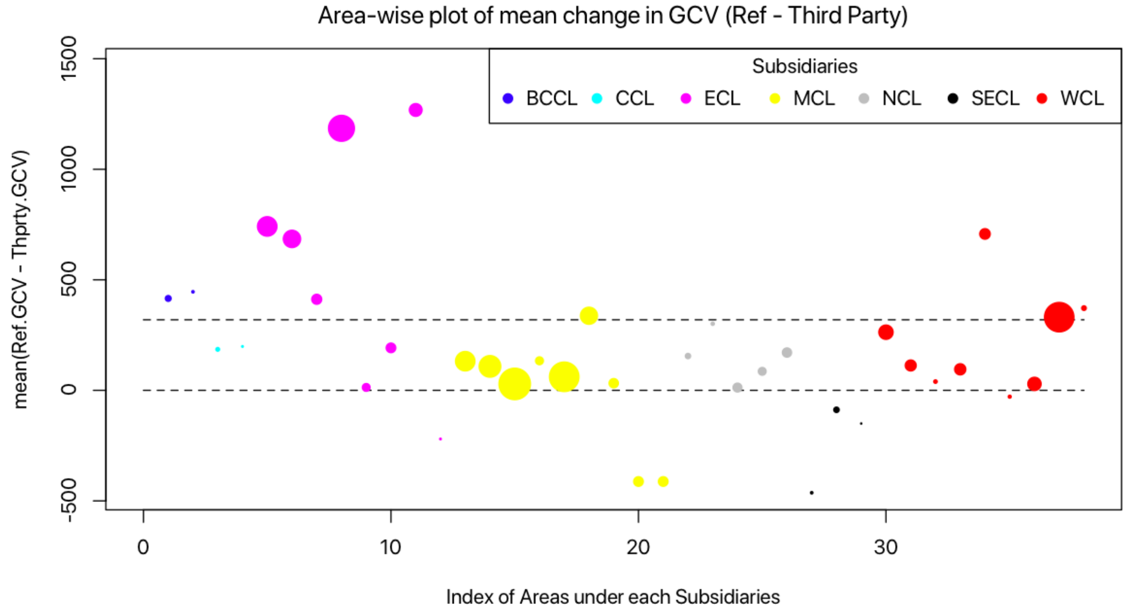


Figure 8: Plot of the average value of GCV_{R-T} for each Area in the Referral data.

Figure 8 shows that for some areas in ECL and WCL, the average GCV_{R-T} is highly positive, which means that those areas were benefited much for the samples going to arbitration.

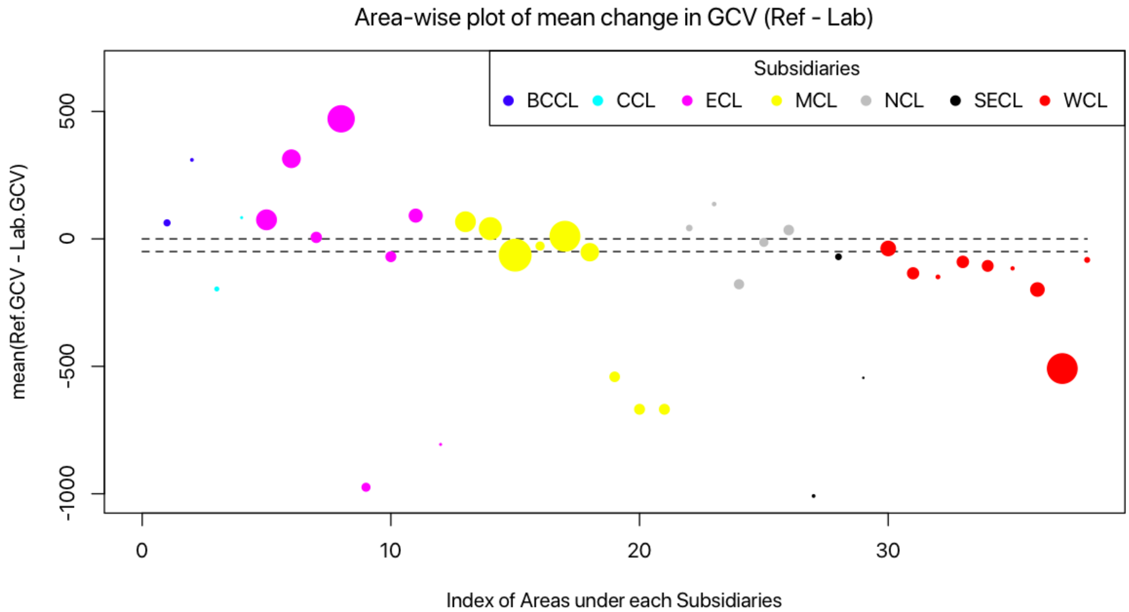
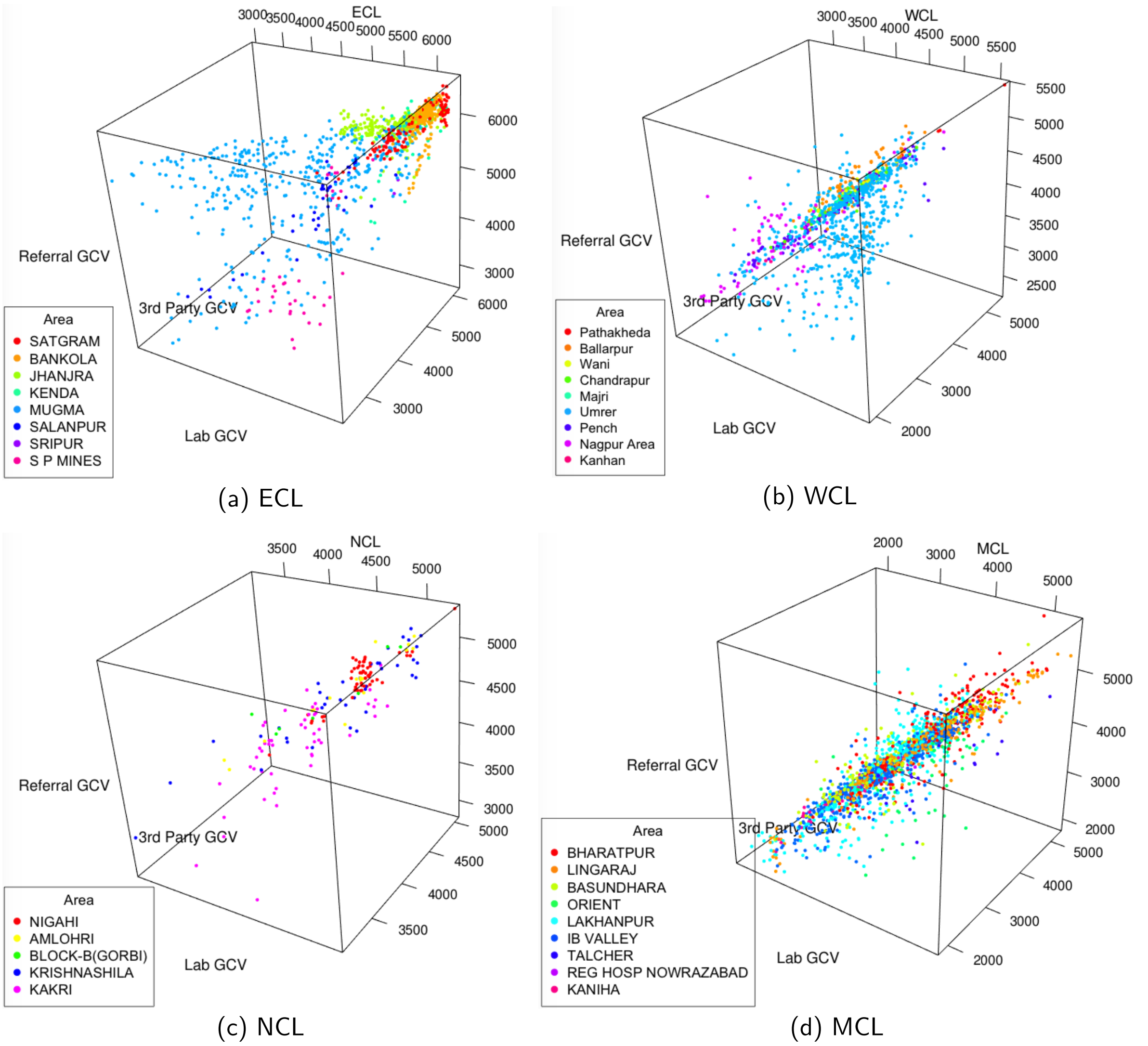


Figure 9: Plot of the average value of GCV_{R-L} for each Area in the Referral data.

In the above figure we see subsidiaries ECL and WCL to behave in a different manner, unlike in Figures 7 and 8. Another observation here is that the two horizontal dotted lines are closer, indicating that the overall sample average of GCV_{R-L} is much lower than that of GCV_{R-T} or GCV_{T-L}.

Next we take up each of the subsidiaries ECL, WCL, MCL, and NCL in the Referral data (remaining subsidiaries constitute just 1.32% of the dataset) and do the 3D plot of the three types of GCVs by coloring them according different levels of Area, which is shown in Figure 10.



(a) ECL (b) WCL

(c) NCL (d) MCL

Figure 10: A plot of 3 types of GCVs for different subsidiaries from the Referral data. The points are colored according to different levels of Area within each subsidiary.

Here we observe some clustering on the basis of Area. We also see that

- For (a) ECL, the samples from area MUGMA are scattered the most, and some of them have very high Referral GCVs despite having moderate values of other two GCVs. Samples from S P MINES and SALANPUR have very low Third Party GCVs but moderately high Lab GCVs and some of them have very high Referral GCVs. Some samples from SATGRAM are seen to have very high Referral GCVs and Lab GCVs but quite lower Third Party GCVs. Most samples from BANKOLA have very high GCVs (of all 3 types).

- For (b) WCL, the samples from area Umrer are scattered the most. Majority of the samples from Umrer and Nagpur Area have quite low Third Party GCVs but relatively higher Lab and Referral GCVs. Samples from other areas seem to be pretty much well-behaved and mostly scattered near the line $x = y = z$ (which represents the ideal case, had there been no significant difference in the three types of GCVs).

- For (c) NCL, the samples from NIGAHI form a small cluster and have high Referral and Lab GCVs but comparatively lower Third Party GCVs. Samples from other areas in this subsidiary are found to be very scattered.

- For (d) MCL, we do not find any particular clustering. It seems that all the areas under this subsidiary behaves in a similar fashion. We also see that they are scattered mainly around the line $x = y = z$.

A pairwise comparison of different GCVs for some of the areas under ECL are shown in Figure 11. This plot is done in the same way as we did for Figure 4. For each sample we plot the two types of GCVs and join them by a vertical line. The color red is used when that GCV differencce causes a loss to CIL, green if it does not.

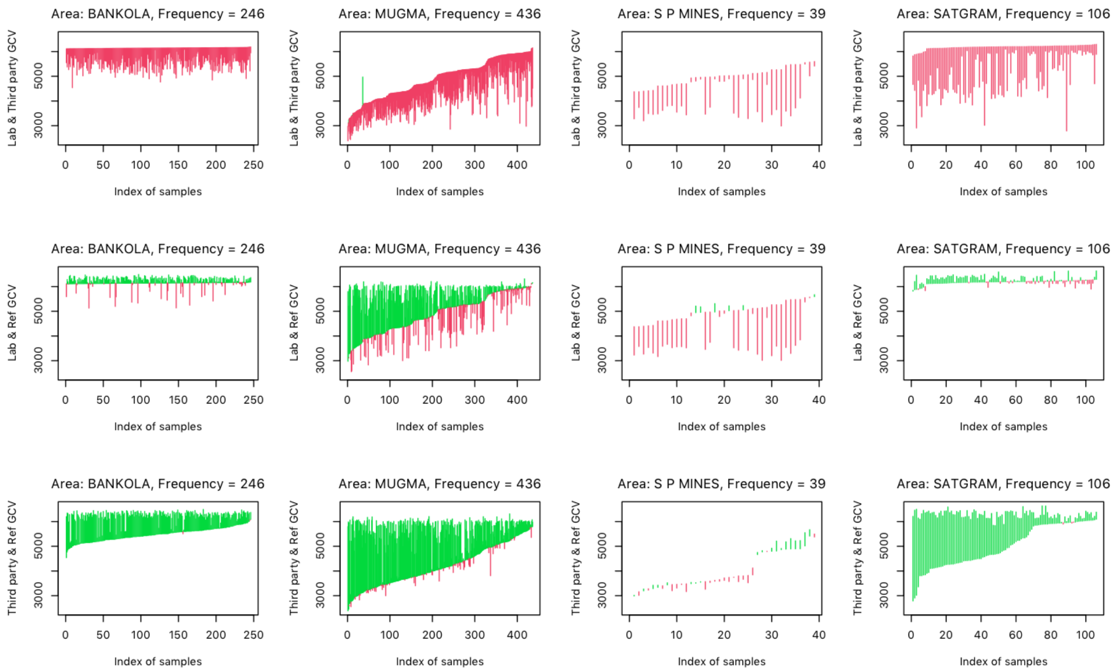


Figure 11: Pairwise GCV plots for selective areas within ECL

Overall, we find that the anomalies within a subsidiary might be rooted from some, if not all, of the areas within that subsidiary. The subsidiaries ECL and WCL seem to contain majority of the anomalies.

Next we look at the GCV difference in different Plants and Buyers. We did not find any clustering for Plant or Buyer levels from the 3D plot as above. We suspect that the anomalies in Plants and Buyers might be rooted at the subsidiary level. To investigate this visually, we did two kinds of plots, which are given below.

First, we plot in Figure 12 the average GCV difference (T-L) for each buyer, after separating the samples subsidiary-wise. We ordered the Buyer levels such that Buyers which do not buy any sample from ECL are on the left side of the vertical dotted line. Also, the area of each bubble is proportional to the GCV difference in that buyer level under that subsidiary. Observe that same buyer has much greater GCV difference when it buys from ECL, compared to other subsidiaries. In Figure 13 we do the same thing for Plants and arrived at the same conclusions.

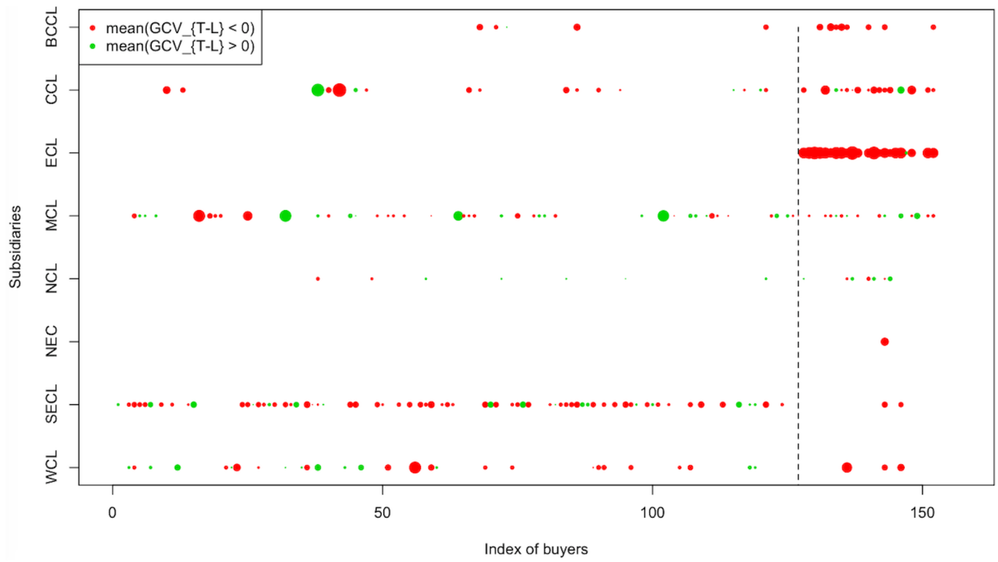


Figure 12: The effect of Subsidiaries on the mean GCV difference for different Buyers.

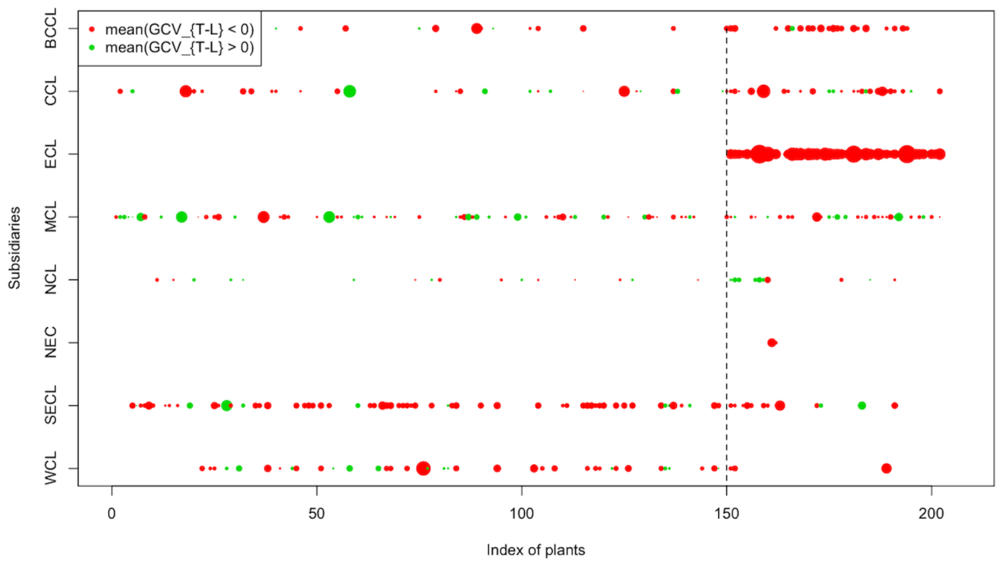
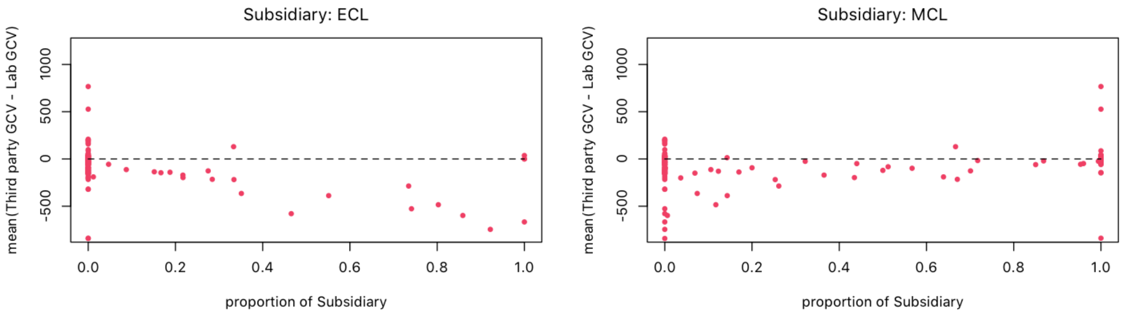


Figure 13: The effect of Subsidiaries on the mean GCV difference for different Plants.

Thus, **the samples corresponding to any Plant or Buyer are found to have more GCV difference (on an average) when they come from ECL.** Why have we shown only samples from ECL? Because otherwise if there is a blank space in the line of ECL, we would not be able to tell whether it is due to no sample or due to very small GCV difference. We did this plot for each of the subsidiaries and the same conclusion (for ECL) was found.

Again, many buyers buy coal from more than one subsidiary and we suspect that the huge difference between different kind of GCVs might be related to the *proportion* of 'bad' subsidiaries from which the buyers buy the coal. This is supported by Figure 14, where we have plotted the mean GCV difference (GCV_{T-L}) for different buyers against the proportion for one fixed subsidiary. Note that for ECL points are moving down from the zero-line as the proportion of ECL increases, while for MCL points are moving close to zero-line.

(a) For Third Party data



(b) For Referral data

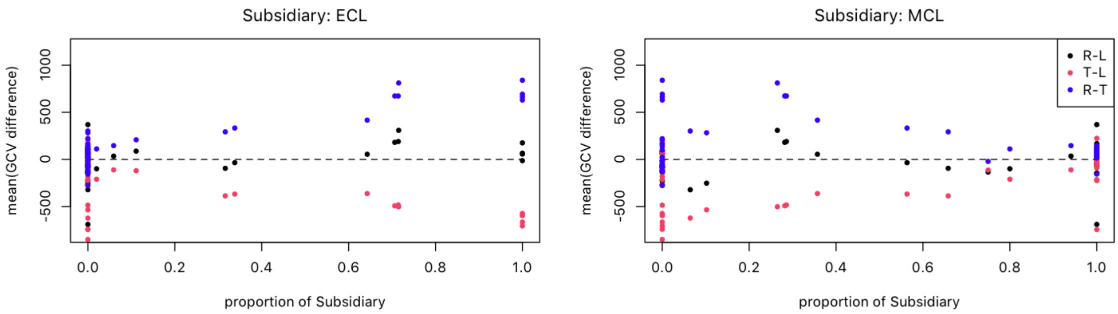


Figure 14: Effect of ECL and MCL on difference in GCV for different buyers (for Third party data and Referral data)

In the lower panel of Figure 14, we observe similar features for the Referral data. In Figure 15, we show a similar plot for Plants in Referral data. There also we get the same pattern: the more proportion of ECL we have, worse is the GCV difference. This observation is also suggested by noticing how the 2nd order raw moments on moving intervals change when proportion of a particular subsidiary increases, which is shown in Table 1 and 2 below (corresponding to Figure 14 (a) and Table 15 (a) respectively).

Prop	(0,0.4)	(0.1,0.5)	(0.2,0.6)	(0.3,0.7)	(0.4,0.8)	(0.5,0.9)	(0.6,1)
ECL	33842	65754	90486	136769	211454	220591	301512
MCL	72831	49518	25780	20173	18810	15128	12575

Table 1: How the 2nd order raw moment changes when the proportion of a subsidiary increases in a certain level of Buyer

In Figure 15 we observe similar characteristics for Plant levels.

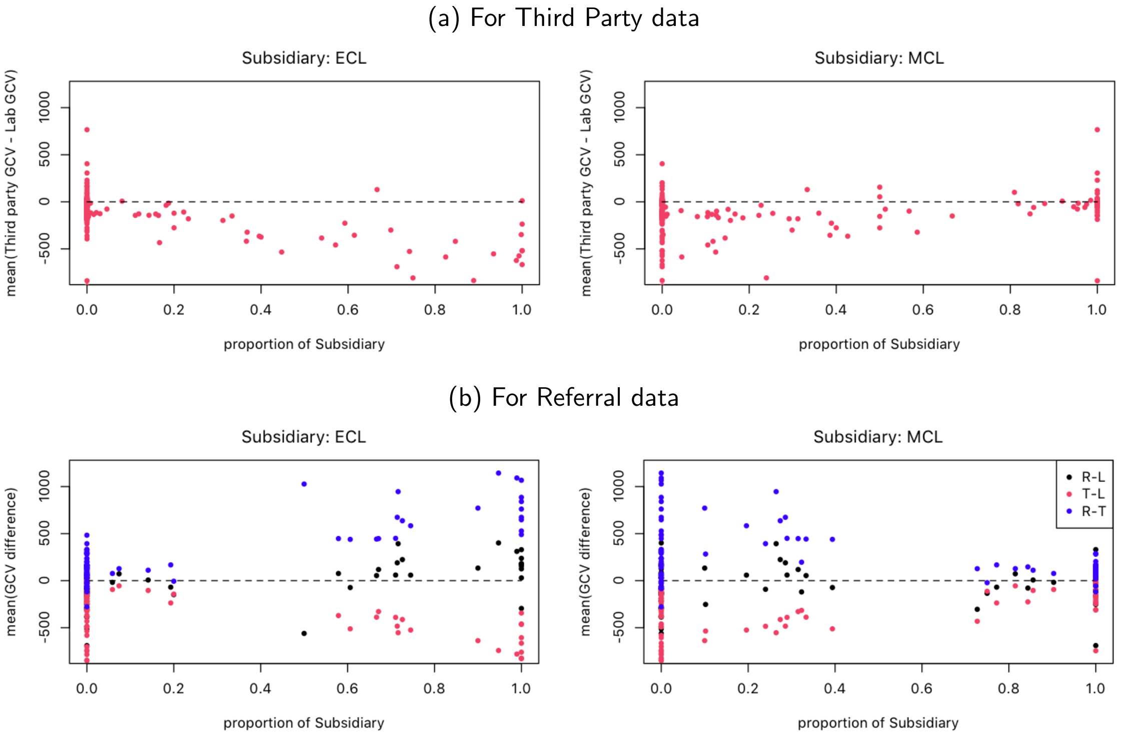


Figure 15: Effect of ECL and MCL on difference in GCV for different plants (for Third party data and Referral data)

Prop	(0,0.4)	(0.1,0.5)	(0.2,0.6)	(0.3,0.7)	(0.4,0.8)	(0.5,0.9)	(0.6,1)
ECL	39483	69500	182400	182859	304721	272415	398609
MCL	123505	84830	75635	48110	44865	19306	5163

Table 2: How the 2nd order raw moment changes when the proportion of a subsidiary increases in a certain level of Plant

What we conclude from all these visual tools is that **the anomaly in Plant of Buyer might not be their intrinsic property, the main problem might be rooted at the subsidiary level or in some particular areas under certain subsidiaries.** Of course in the next sections we shall try to find more statistical evidences in favour of this claim of ours.

6 Detecting anomalous levels by t-score

Here we detect anomalous levels of different factors, e.g. Area, Plant etc. by computing the t-score of different levels of the factor. Note that this method detects levels which might be bad in overall, and does not focus on individual samples from that level. In section 8 later, we shall address the latter issue, i.e., finding proportion of samples in some different levels of a factor that might be anomalous.

We divide the Third Party data using subsidiaries and then for each of the factors Area, Siding, Source unit, Plant and Buyer, we do the following:

- Take $y = \text{GCV}_{T-L}$ as our response and let x be the factor we are considering.
- We fit the model $y_{ij} = \mu + x_i + \epsilon_{ij}$ where μ is the *overall effect of the subsidiary* and x_i is the *extra effect of the i-th level of the factor x.* We assume that the errors ϵ_{ij}'s have mean 0 and are uncorrelated.
- Now we compute the t-score for the x_i's. Whenever its magnitude is greater than 3, we call the level to be anomalous.
- Note that $y < 0$ means bad for CIL and $y > 0$ means good of CIL. So whenever the t-score of a level is less than -3, we label that anomalous level as 'bad' and when it is greater than 3, we label it as 'good'.

In Figure 16 we have shown the t-scores for different levels of different factors. The areas of the bubbles are proportional to the magnitude of its t-score.

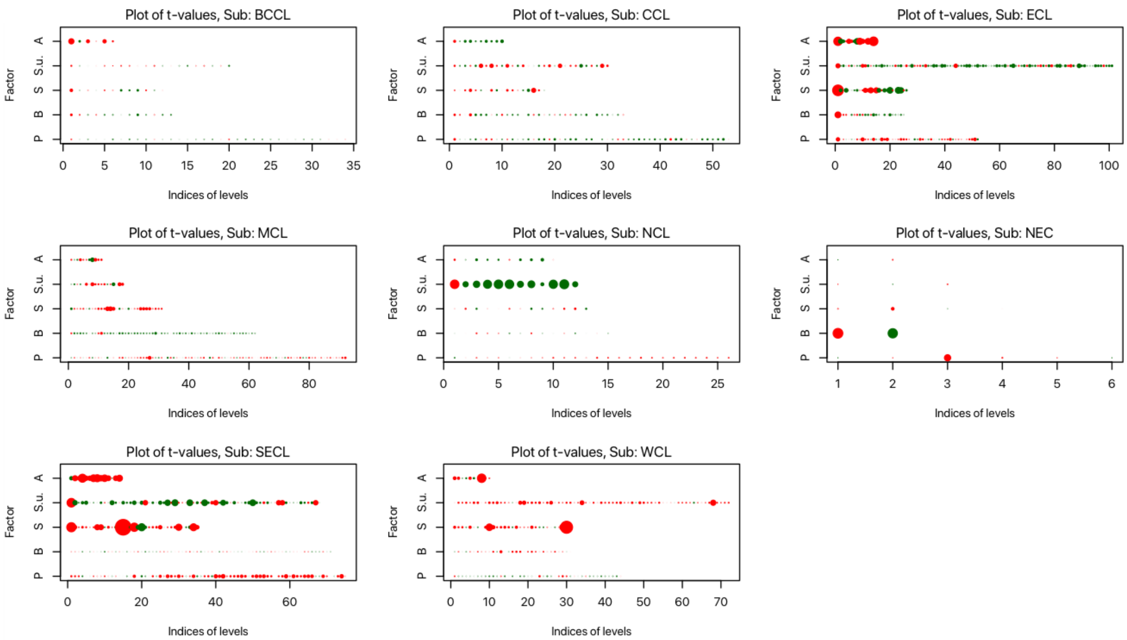


Figure 16: A plot illustrating the t-scores of different levels of the factors Area (A), Siding (S), Source unit (S.u.), Buyer (B) and Plant (P). The areas of the bubbles are proportional to the absolute value of the t-score and the bubbles are coloured red when the t-score is negative and green if it is positive.

In this method we can point out certain anomalous levels for each of the above mentioned factors and also categorise them as 'good' or 'bad' depending on whether the t-score is greater than 3 or less than -3, respectively (which are respectively good or bad for CIL). Table 3 below shows the number of such levels of different factors that were marked as anomalous in this method.

Factor →	Area		Siding		Source unit		Buyer		Plant	
Sub ↓	bad	good	bad	good	bad	good	bad	good	bad	good
BCCL	3	1	0	0	1	2	1	1	0	0
CCL	1	7	9	5	5	2	2	6	2	14
ECL	8	4	16	43	6	10	3	7	15	2
MCL	4	2	7	1	11	3	1	5	4	0
NCL	1	6	1	11	2	3	0	0	0	0
NEC	0	0	0	0	1	0	1	1	1	0
SECL	12	1	10	34	18	6	0	0	38	1
WCL	4	2	27	0	17	0	5	0	2	0

Table 3: The number of bad or good anomalous levels of different factors (in Third Party data), marked on the basis of t-scores of the levels of the factors.

Observe that for ECL and SECL, more anomalous levels of Siding were marked good and for Area it is the opposite. We have to keep in mind that here the anomalous levels are raised after removing the subsidiary effect. A complete list of the anomalous levels based on the t-values is given in Appendix A. We performed similar procedure for the Referral data, which are illustrated in Figures 17 to 19. Since the subsidiaries ECL, MCL and WCL make up 95% of the Referral data, we have clubbed the other subsidiaries and labelled them as "Others".

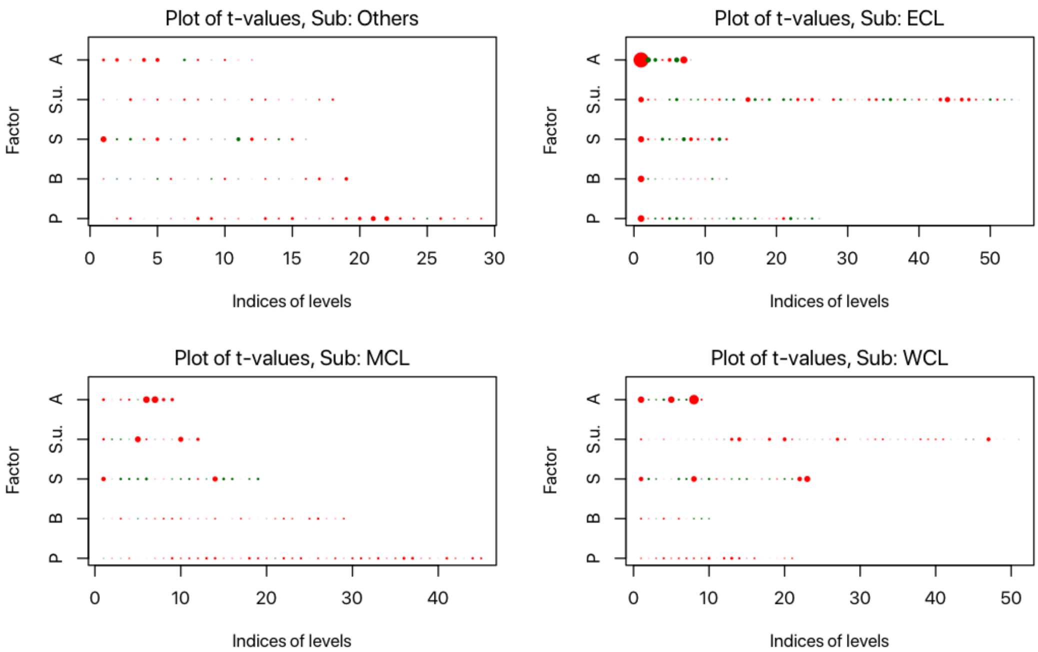


Figure 17: The t-scores of different levels of the factors for the Referral data, with GCV$_{T-L}$ as the response. The areas of the bubbles are proportional to the t-score and it is coloured red when t-score is negative and green if it is positive.

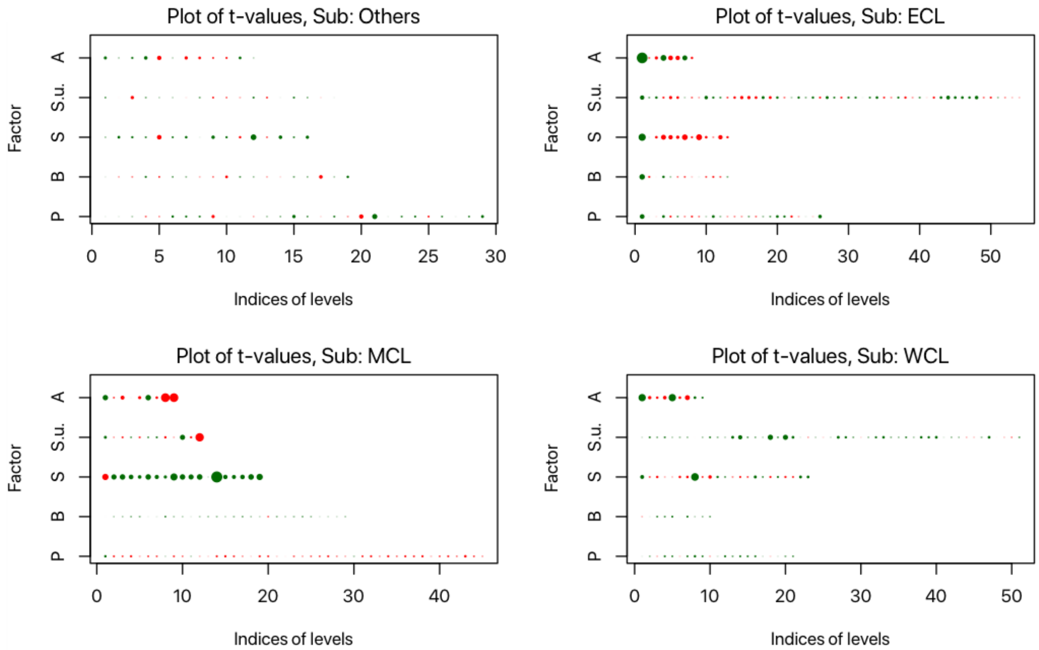


Figure 18: The t-scores of different levels of the factors for the Referral data, with GCV$_{R-T}$ as the response. The areas of the bubbles are proportional to the t-score and it is coloured red when t-score is negative and green if it is positive.

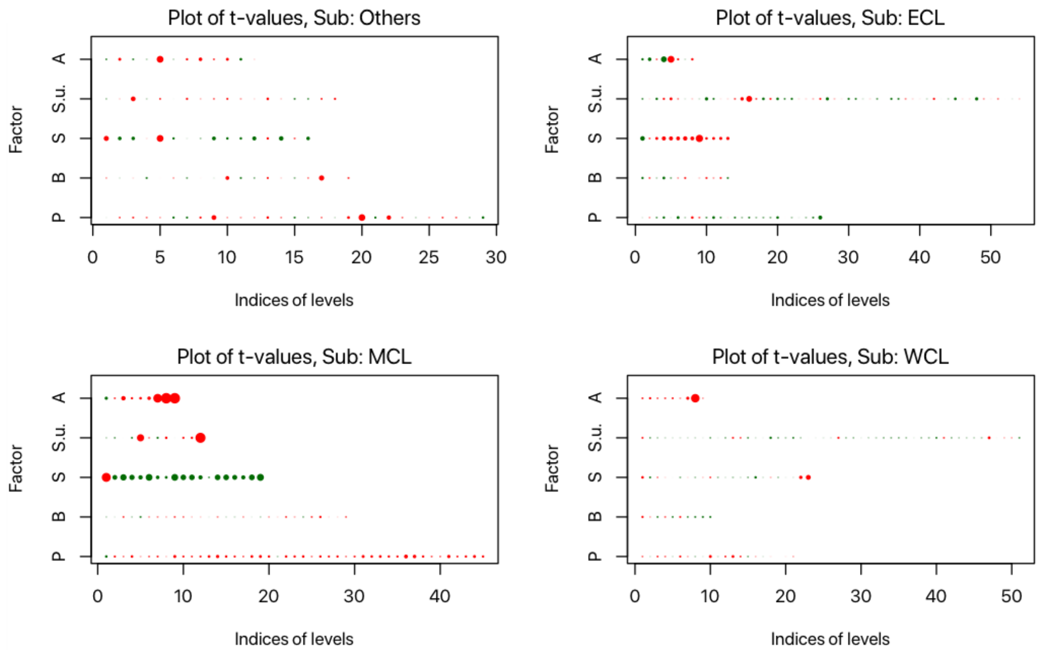


Figure 19: The t-scores of different levels of the factors for the Referral data, with GCV$_{R-L}$ as the response. The areas of the bubbles are proportional to the t-score and it is coloured red when t-score is negative and green if it is positive.

7 Modeling

7.1 Possible linear models

We take the variable GCV$_{T-L}$ as the response variable for the Third Party data and the variables GCV$_{T-L}$, GCV$_{R-T}$, and GCV$_{R-L}$ as the response variable for the Referral data. To choose between several combinations for the set of explanatory variables we use adjusted R^2 and BIC (Bayes Information Criterion). In the following table we have shown those for various choices of the explanatory variables.

Response $\rightarrow$	GCV$_{T-L}$ (Thprty)		GCV$_{T-L}$ (Ref)		GCV$_{R-L}$ (Ref)		GCV$_{R-T}$ (Ref)	
Model $\downarrow$	R^2	BIC	R^2	BIC	R^2	BIC	R^2	BIC
Area	0.35	1473256	0.42	67917	0.31	69142	0.46	69241
Buyer	0.08	1509055	0.18	69650	0.12	70342	0.14	71452
Sub + Buyer	0.23	1490825	0.29	69038	0.16	70149	0.33	70322
Area + Buyer	0.36	1473677	0.44	68117	0.31	69475	0.47	69481
Plant	0.17	1499810	0.29	69233	0.18	70254	0.29	70841
Sub + Plant	0.26	1488206	0.33	68984	0.20	70219	0.36	70405
Area + Plant	0.36	1473930	0.44	68389	0.32	69717	0.47	69767
Area+Buyer+Plant	0.38	1474410	0.44	68553	0.32	69936	0.47	69936
Siding	0.32	1478751	0.41	68225	0.25	69724	0.41	69832
Area + Siding	0.37	1471435	0.43	68164	0.31	69446	0.46	69495
Source.unit	0.38	1471759	0.44	68468	0.32	69771	0.47	69836
Area + Source.unit	0.38	1470603	0.45	68478	0.32	69788	0.47	69850

Table 4: Adjusted R^2 and BIC (rounded) for different combinations of explanatory variables

Observe that whenever we include Area in the model, it becomes better. Also observe that Area + Buyer + Plant is worse than both Area + Plant and Area + Buyer.

7.2 Model fit (diagnostics)

To assess the fitting of the linear model, we look at different plots involving the standardised residuals. Here is an example: consider the linear model GCV$_{T-L}$ = Area + Plant + error, for the Referral data. In Figure 20 we have shown different diagnostic plots for the linear model GCV$_{T-L}$ = Area + Plant + i.i.d. error, for the Referral data.

- The Residuals versus Fitted plot is used to check the linear relationship assumptions. Here we found an almost horizontal line, without distinct patterns, so it is an indication for a linear relationship, which is good.
- The Scale-Location plot is used to check the homogeneity of variance of the residuals (homoscedasticity). The red line deviating much from being horizontal line and not equally spread points indicates that we might have a heteroscedasticity problem here.

- The Normal Q-Q plot shows that the errors might deviate much from Normality. This is also confirmed by tests for normality on the studentised residuals, which was rejected with extremely low p-values.
- The Residuals versus Leverage plot is used to identify influential cases, that is extreme values that might influence the regression results when included or excluded from the analysis. Here we do not find any such cases.
- We also have adjusted R^2 equal to 0.44, so the model does capture a lot of variations present in the data.

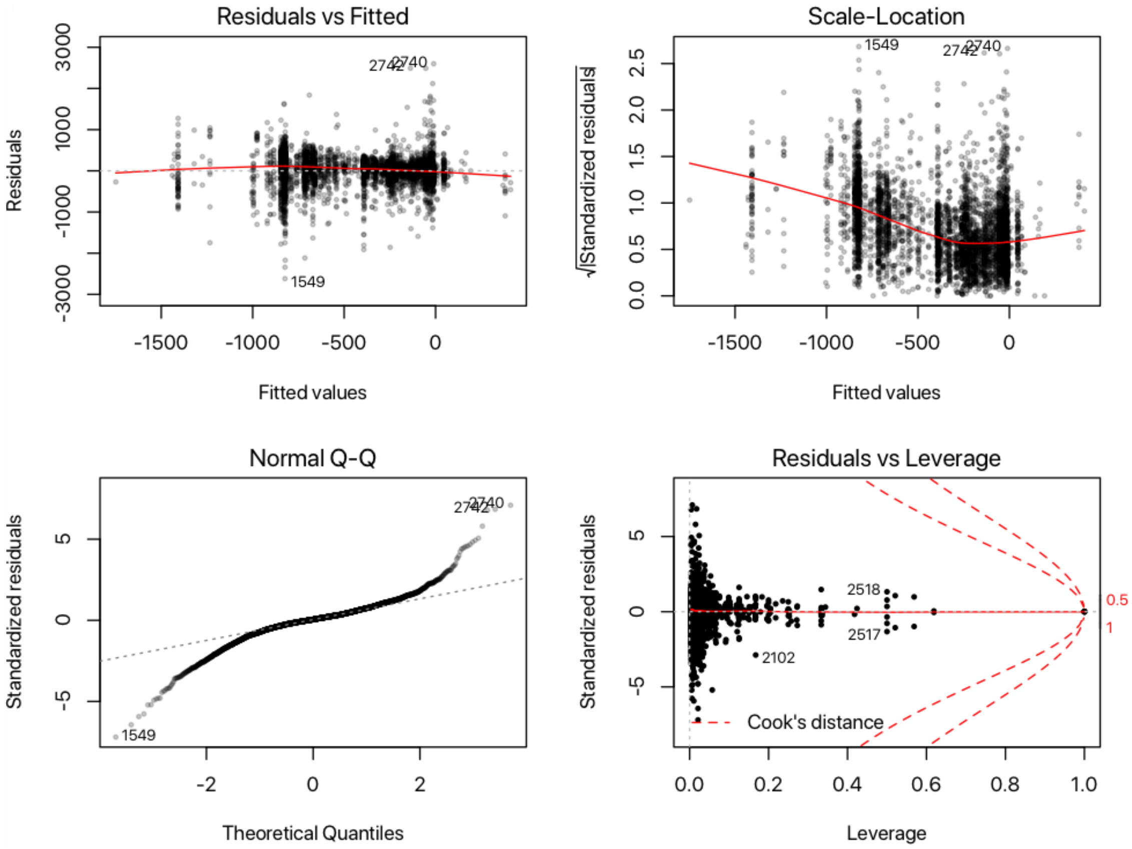


Figure 20: Studentised residuals for the linear model GCV_{T-L} = Area + Plant + error for the Referral data

Overall, the model might not be a good fit to the data, but it does provide a grounding to compare different plants, on the basis of studentised residuals. Because the more a plant produces samples with high standardised residuals, the more it can be thought as anomalous. These models will be used in the next section to flag certain anomalous levels of the factors Area, Plant, Buyer etc.

We got similar results for the other models in case of the Referral data, and some of them deteriorate a bit for the Third Party data (perhaps because of the greater size).

8 Another method to find anomalies

In this section we shall raise flags for levels of different factors which contain major proportion of the anomalous samples.

8.1 Possible flags for Area

First we describe how we raise flags for Area. Our intuition is that we should raise flags for those areas for which the percentage of 'anomalous samples' turns out to be the most. To do this, we propose the following procedure.

- Take $y = \text{GCV}_{T-L}$ from the third party data as our response variable and the factor Area, say x, as the only explanatory variable and fit the model $y_{ij} = x_i + \epsilon_{ij}$ where y_{ij} is the j-th response under the i-th level of x. We assume that ϵ_{ij}'s are iid with mean 0 and uncorrelated.
- A sample (say, corresponding to y_{ij}) is called anomalous if its standardized residual is greater than 3 in magnitude.
- The **anomaly score** of an area is then determined as the proportion of anomalous cases among the samples coming from that area.
- We raise a flag for the i-th level of Area if its score is among the top 5% of all scores. Of course we can replace this 5% with any other plausible fraction.

The result of the last procedure is shown in the following figure, where we have pointed out the possible flags, i.e., the areas that come in top 5% in terms of the anomaly score.

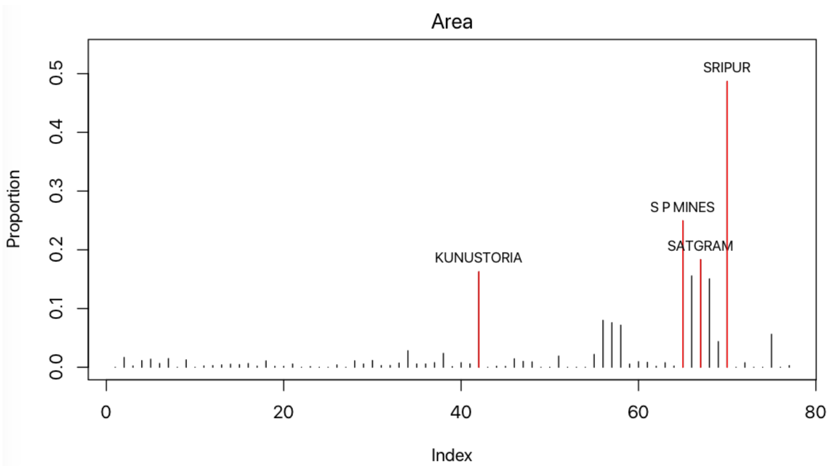


Figure 21: Plot of the percentage of anomalous samples from each area. The areas with top 5% anomaly scores are shown in red.

8.2 Raising flags for the other factors

For each of the factors Plant, Buyer, Siding and Source unit, we perform the following procedure:

- Take $y = \text{GCV}_{T-L}$ from the third party data as our response variable and the factor, say x, as the only explanatory variable *other than Area*, and fit the model $y_{ijk} = x_i + a_j + \epsilon_{ijk}$ where y_{ijk} is the k-th response under the i-th level of x and the j-th level of Area. We assume that ϵ_{ijk}'s are iid with mean 0 and uncorrelated.
- A sample (say, corresponding to y_{ij}) is called anomalous if its standardized residual is greater than 3 in magnitude.
- The **anomaly score** of each level of that factor x is determined as the proportion of anomalous cases among the samples having that level of the factor x.
- We raise a flag for the i-th level of the factor if its score is among the top 5% of all scores.
- For the Referral data, different GCV differences can be used as the response variable y.

Figure 22 shows the result of this procedure applied to the whole Third Party data. We have used red color to show the indices of levels of each of the factors that turn out to be possible flags in our method. The complete list of the flags, along with the percentages of its samples that were pointed as anomalous cases will be given below.

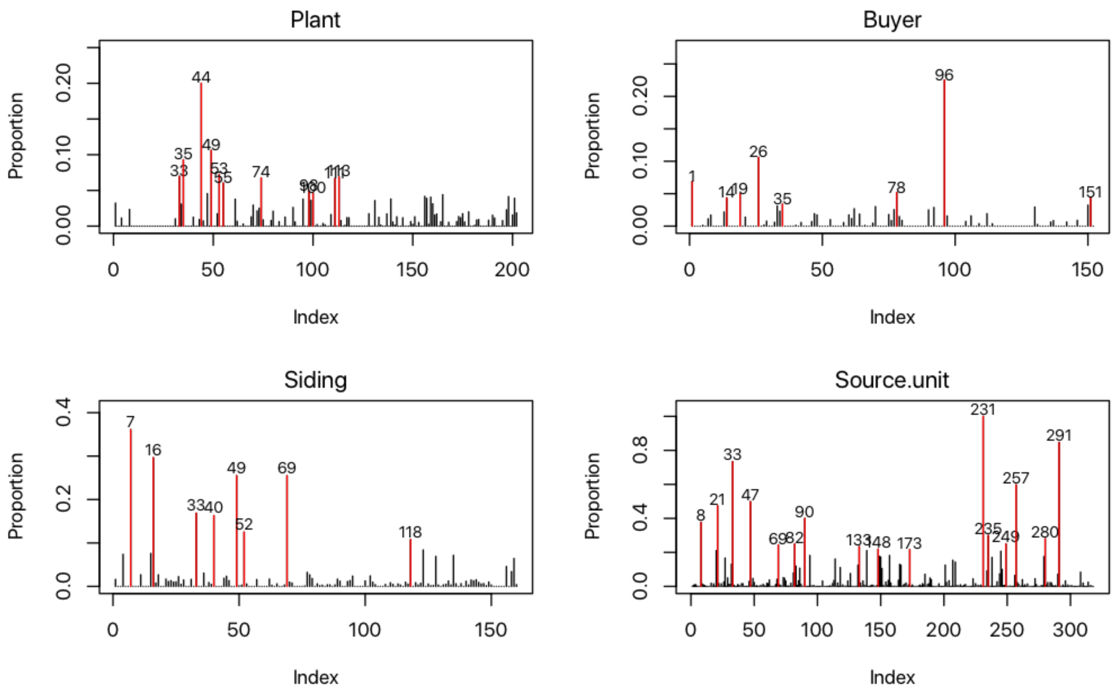


Figure 22: Possible Flags for Plant, Buyer, Siding and Source unit (indices only) in the Third Party data

8.3 Digging further

Since we suspected earlier that a major cause of the anomalies might be just the subsidiaries the samples are coming from, we also list the percentage of each subsidiary shared by the above flags. These are shown in Figures 23 – 25.

	% of anomalous cases	% BCCL	% CCL	% ECL	% MCL	% NCL	% NEC	% SECL	% WCL
KUNUSTORIA	16.25	0	0	100	0	0	0	0	0
S P MINES	24.94	0	0	100	0	0	0	0	0
SATGRAM	18.33	0	0	100	0	0	0	0	0
SRIPUR	48.67	0	0	100	0	0	0	0	0

Figure 23: Percentage share of Subsidiaries in flags for Area

	% of anomalous cases	% BCCL	% CCL	% ECL	% MCL	% NCL	% NEC	% SECL	% WCL
2	36.11	0	0	100	0	0	0	0	0
BELBAID	29.63	0	0	100	0	0	0	0	0
Bonjemihari	16.85	0	0	100	0	0	0	0	0
CHINAKURI 3	16.30	0	0	100	0	0	0	0	0
DN I	25.49	0	0	100	0	0	0	0	0
DUMRI KHURD	12.50	0	0	0	0	0	0	0	100
JAMTARA	25.48	0	0	100	0	0	0	0	0
P/D	10.78	0	0	100	0	0	0	0	0

Figure 24: Percentage share of Subsidiaries in flags for Siding

	% of anomalous cases	% BCCL	% CCL	% ECL	% MCL	% NCL	% NEC	% SECL	% WCL
AMRITNAGAR COLLIERY	37.70	0	0	100	0	0	0	0	0
BANSRA OCP	47.42	0	0	100	0	0	0	0	0
BHANORA WEST OC PATCH	73.39	0	0	100	0	0	0	0	0
BONJEMEHARI COLLIERY	49.85	0	0	100	0	0	0	0	0
CHAPUIKHAS COLLIERY	24.32	0	0	100	0	0	0	0	0
Chitra-A OC	24.90	0	0	100	0	0	0	0	0
DHEMOMAIN INCLINE	40.00	0	0	100	0	0	0	0	0
J.K. NAGAR	23.48	0	0	100	0	0	0	0	0
KALIDASPUR PROJECT	21.90	0	0	100	0	0	0	0	0
KUNUSTORIA COLLIERY	21.67	0	0	100	0	0	0	0	0
NIGHA COLLIERY	100.00	0	0	100	0	0	0	0	0
NINGHA	30.00	0	0	100	0	0	0	0	0
PATMOHANA COLLIERY	25.00	0	0	100	0	0	0	0	0
PURE SEARSOLE COLLIERY	59.70	0	0	100	0	0	0	0	0
SATGRAM PROJECT	28.07	0	0	100	0	0	0	0	0
SRIPUR SEAM INCLINE	84.62	0	0	100	0	0	0	0	0

Figure 25: Percentage share of Subsidiaries in flags for Source unit

Since all the Areas and most Sidings and Source units fall under a unique Subsidiary, it is no surprise that the percentage shares for each of them will be 100% for one subsidiary and 0% for the others. But this uniqueness falls apart for Plants and Buyers – levels of these factors do overlap with Subsidiaries. Let's see next what happens for them.

	% of anomalous cases	% BCCL	% CCL	% ECL	% MCL	% NCL	% NEC	% SECL	% WCL
(TANGENDCO + NLC)-NTPL	6.89	0.00	18.06	74.12	0.00	7.82	0.00	0.0	0.0
Aryan Ispat and Power Pvt	4.35	0.00	0.00	0.00	0.00	0.00	0.00	100.0	0.0
BSL	5.00	0.00	0.00	0.00	100.00	0.00	0.00	0.0	0.0
CESC	10.59	7.81	0.00	92.19	0.00	0.00	0.00	0.0	0.0
DVC	3.39	27.79	29.66	35.11	7.44	0.00	0.00	0.0	0.0
MIGK	5.17	0.00	48.28	46.55	0.00	5.17	0.00	0.0	0.0
NTPC Farakka	22.52	0.00	0.30	85.89	0.60	0.00	11.71	0.3	1.2
kpps	4.35	0.00	34.78	21.74	43.48	0.00	0.00	0.0	0.0

Figure 26: Percentage share of Subsidiaries in flags for Buyer

	% of anomalous cases	% BCCL	% CCL	% ECL	% MCL	% NCL	% NEC	% SECL	% WCL
800	6.98	0.00	0.00	90.70	1.16	2.33	0.00	2.33	3.49
905	9.28	0.63	0.00	99.28	0.00	0.00	0.09	0.00	0.00
Aryan Ispat and Power Pvt 130211	20.00	0.00	0.00	20.00	40.00	0.00	0.00	40.00	0.00
BUDGE-BUDGE (BGB)	10.61	11.14	0.00	88.86	0.00	0.00	0.00	0.00	0.00
Bandel	7.22	1.14	0.00	74.90	23.95	0.00	0.00	0.00	0.00
Barh	6.04	7.82	20.96	71.23	0.00	0.00	0.00	0.00	0.00
Durgapur TPS	6.76	13.06	0.00	82.43	4.50	0.00	0.00	0.00	0.00
Khaparkheda	4.91	0.00	0.00	0.00	11.41	0.00	0.00	6.54	82.05
Koradi	4.60	0.00	0.00	0.00	8.08	0.00	0.00	25.90	66.02
MGPV,Gangavaram	6.67	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00
MIGK	6.89	0.00	18.06	74.12	0.00	7.82	0.00	0.00	0.00

Figure 27: Percentage share of Subsidiaries in flags for Plant

Note that despite having overlaps with more than one subsidiary, majority of the anomalous cases for Plants and Buyers in the Third Party data come from the Subsidiaries ECL, MCL and CCL.

8.4 and further...

We have another suspicion that the anomalous cases in different subsidiaries might be caused by particular areas within that subsidiary (i.e., blaming all areas within a bad subsidiary might not be reasonable). To check this, we consider the areas that hold the top 10% anomaly scores, then for each flag of the other factors (Plant, Buyer, Siding, Source unit), we calculate the percentage share of them with those flagged levels of Area. These are shown in Figures 28 – 31.

	KUNUSTORIA	Nagpur Area	ORIENT	S P MINES	SALANPUR	SATGRAM	SODPUR	SRIPUR
800	90.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00
905	6.72	0.00	0.00	2.96	2.63	5.70	0.00	1.94
Aryan Ispat and Power Pvt 130211	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BUDGE-BUDGE (BGB)	0.00	0.00	0.00	0.00	70.03	0.00	0.00	0.00
Bandel	8.37	0.00	0.00	0.76	3.80	5.32	18.25	0.00
Barh	0.71	0.00	0.00	30.73	0.36	1.42	0.00	0.00
Durgapur TPS	5.86	0.00	0.00	0.00	8.56	17.57	0.00	0.90
Khaparkheda	0.00	74.57	0.16	0.00	0.00	0.00	0.00	0.00
Koradi	0.00	47.72	0.21	0.00	0.00	0.00	0.00	0.00
MGPV,Gangavaram	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MIGK	5.77	0.00	0.00	0.00	18.44	3.72	0.00	0.00

Figure 28: Percentage share of flagged Areas in flagged Plants

	KUNUSTORIA	Nagpur Area	ORIENT	S P MINES	SALANPUR	SATGRAM	SODPUR	SRIPUR
(TANGENDCO + NLC)-NTPL	5.77	0	0.00	0.00	18.44	3.72	0	0.00
Aryan Ispat and Power Pvt	0.00	0	0.00	0.00	0.00	0.00	0	0.00
BSL	0.00	0	0.00	0.00	0.00	0.00	0	0.00
CESC	0.00	0	0.00	0.00	60.04	0.00	0	0.00
DVC	0.46	0	0.12	0.12	8.66	1.75	0	0.56
MIGK	3.45	0	0.00	0.00	8.62	13.79	0	0.00
NTPC Farakka	2.10	0	0.00	0.00	2.10	29.13	0	30.63
kpps	8.70	0	0.00	0.00	0.00	0.00	0	0.00

Figure 29: Percentage share of flagged Areas in flagged Buyers

	KUNUSTORIA	Nagpur Area	ORIENT	S P MINES	SALANPUR	SATGRAM	SODPUR	SRIPUR
2	0	0	0	0	0	50.00	0	50.00
BELBAID	100	0	0	0	0	0.00	0	0.00
Bonjemihari	0	0	0	0	100	0.00	0	0.00
CHINAKURI 3	0	0	0	0	0	0.00	100	0.00
DN I	0	0	0	0	0	96.76	0	3.24
DUMRI KHURD	0	100	0	0	0	0.00	0	0.00
JAMTARA	0	0	0	100	0	0.00	0	0.00
P/D	0	0	0	0	0	100.00	0	0.00

Figure 30: Percentage share of flagged Areas in flagged Sidings

	KUNUSTORIA	Nagpur Area	ORIENT	S P MINES	SALANPUR	SATGRAM	SODPUR	SRIPUR
AMRITNAGAR COLLIERY	100	0	0	0	0	0	0	0
BANSRA OCP	100	0	0	0	0	0	0	0
BHANORA WEST OC PATCH	0	0	0	0	0	0	0	100
BONJEMEHARI COLLIERY	0	0	0	0	100	0	0	0
CHAPUIKHAS COLLIERY	0	0	0	0	0	100	0	0
Chitra-A OC	0	0	0	100	0	0	0	0
DHEMOMAIN INCLINE	0	0	0	0	0	0	100	0
J.K. NAGAR	0	0	0	0	0	100	0	0
KALIDASPUR PROJECT	0	0	0	0	0	100	0	0
KUNUSTORIA COLLIERY	100	0	0	0	0	0	0	0
NIGHA COLLIERY	0	0	0	0	0	0	0	100
NINGHA	0	0	0	0	0	0	0	100
PATMOHANA COLLIERY	0	0	0	0	0	0	100	0
PURE SEARSOLE COLLIERY	0	0	0	0	0	100	0	0
SATGRAM PROJECT	0	0	0	0	0	100	0	0
SRIPUR SEAM INCLINE	0	0	0	0	0	0	0	100

Figure 31: Area share in bad Source units

These tables do suggest that the anomalous samples from different Plants, Buyers, Sidings, or Source units might actually be rooted from be some particular Areas, like SATGRAM, SRIPUR, SALANPUR, etc. Also observe that majority of these 'bad' areas belong to the subsidiary ECL.

8.5 Flags for the Referral data

For the Referral data, we follow the same procedure, just modifying the response variable. Recall that here we have three response variables: GCV_{T-L}, GCV_{R-T}, and GCV_{R-L}. Figure 32 shows the flags for Area for each of these responses. The number of levels of Area being lower for the Referral data, we have shown the top 10% flags according to the anomaly score.

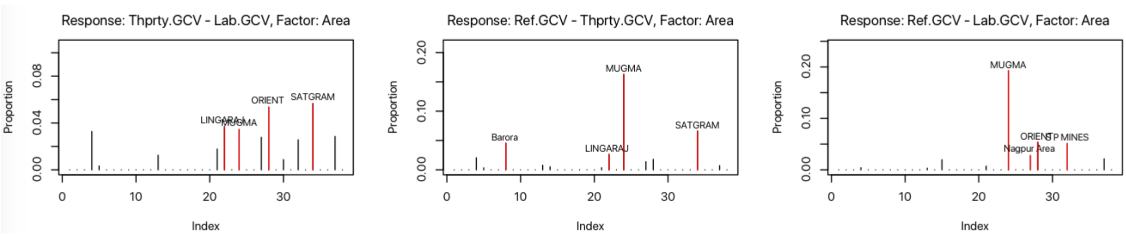


Figure 32: Possible flags for Area for the Referral data

Figure [33] shows the flags for Plants and Buyers for each of the three possible responses. Here we have shown the top 5% flags according to the anomaly score, because the number of levels for these are quite high.

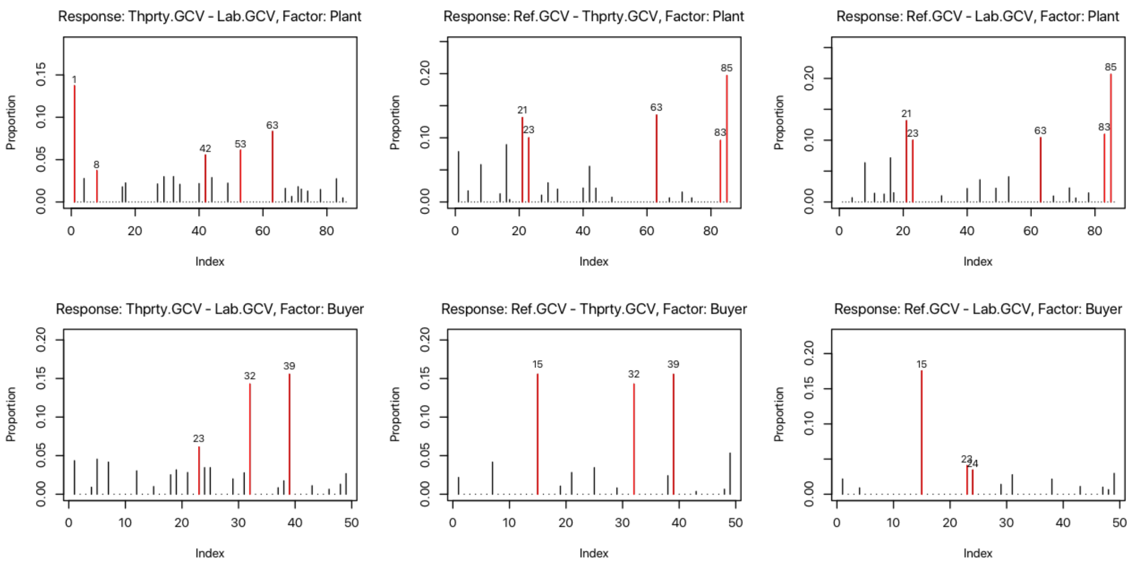


Figure 33: Possible flags for Plants and Buyers for the Referral data

8.6 Shares of Subsidiaries in the flags of Plants and Buyers (for Referral)

	% of anomalous cases	% BCCL	% CCL	% ECL	% MCL	% NCL	% SECL	% WCL
APPDCL, TPS	13.73	0.00	1.96	74.51	19.61	0	3.92	0.0
905	3.70	0.00	0.00	100.00	0.00	0	0.00	0.0
Mettur	5.56	0.00	0.00	0.00	100.00	0	0.00	0.0
Raichur	6.12	0.00	0.00	0.00	10.20	0	0.00	89.8
Sagardighi	8.33	1.04	0.00	98.96	0.00	0	0.00	0.0

(a) With GCV$_{T-L}$ as response

	% of anomalous cases	% BCCL	% CCL	% ECL	% MCL	% NCL	% SECL	% WCL
Dadri	13.16	5.26	0	94.74	0.00	0	0	0
Durgapur TPS	10.00	0.00	0	90.00	10.00	0	0	0
Sagardighi	13.54	1.04	0	98.96	0.00	0	0	0
durgapur steel TPS	9.59	1.37	0	67.12	31.51	0	0	0
mejia	19.71	1.92	0	71.63	26.44	0	0	0

(b) With GCV$_{R-T}$ as response

	% of anomalous cases	% BCCL	% CCL	% ECL	% MCL	% NCL	% SECL	% WCL
Dadri	13.16	5.26	0	94.74	0.00	0	0	0
Durgapur TPS	10.00	0.00	0	90.00	10.00	0	0	0
Sagardighi	10.42	1.04	0	98.96	0.00	0	0	0
durgapur steel TPS	10.96	1.37	0	67.12	31.51	0	0	0
mejia	20.67	1.92	0	71.63	26.44	0	0	0

(c) With GCV$_{R-L}$ as response

Figure 34: Percentage share of Subsidiaries in flags for Plants for Referral data

	% of anomalous cases	% BCCL	% CCL	% ECL	% MCL	% NCL	% SECL	% WCL
KPCL	6.12	0	0	0	10.2	0	0	89.8
MGPV-Hinduja Bros.	14.29	0	0	0	100.0	0	0	0.0
NTPC Farakka	15.56	0	0	100	0.0	0	0	0.0

(a) With GCV$_{T-L}$ as response

	% of anomalous cases	% BCCL	% CCL	% ECL	% MCL	% NCL	% SECL	% WCL
DVC	15.56	1.66	0.33	71.52	26.49	0	0	0
MGPV-Hinduja Bros.	14.29	0.00	0.00	0.00	100.00	0	0	0
NTPC Farakka	15.56	0.00	0.00	100.00	0.00	0	0	0

(b) With GCV$_{R-T}$ as response

	% of anomalous cases	% BCCL	% CCL	% ECL	% MCL	% NCL	% SECL	% WCL
DVC	17.55	1.66	0.33	71.52	26.49	0	0	0.0
KPCL	4.08	0.00	0.00	0.00	10.20	0	0	89.8
KPL	3.45	0.00	0.00	0.00	100.00	0	0	0.0

(c) With GCV$_{R-L}$ as response

Figure 35: Percentage share of Subsidiaries in flags for Buyers for Referral data

From Figures 34 and 35 we can see that the subsidiaries ECL, MCL and WCL share a huge chunk of these anomalous levels of Plants and Buyers.

Comment on testing for anomalous levels: We can in fact do a formal test using the average value of residuals over all observations corresponding to a particular level of any factor and if that measure is significantly different from 0, then it can be said that the corresponding category of that factor gives rise to anomalous cases. However this produces a lot of anomalous levels so we preferred the above mentioned procedure.

9 Conclusions

Following is a summary of our conclusions.

- Our data analysis showed that some anomalies are indeed present in the Coal Quality data, regarding grading of coal samples.
- Using some visual tools we guessed that the primary source of anomalies might be Subsidiary or some areas within a particular subsidiaries.
- We developed a simple procedure to identify the most anomalous levels of each of the main factors: Area, Plant, Buyer, Siding and Source unit.
- We also found that some particular Subsidiary or some particular Area within them carried a lot of the proportion of the anomalous cases from the flags of Plant, Buyer, Siding and Source unit.

A List of anomalous levels based on t-scores

Bad and good levels of different factors for subsidiary BCCL

Bad levels for Area : Barora , Chanch Victoria, Kusunda

Good levels for Area : Block- II

Bad levels for Source.unit :

(empty means no such level was found)

Good levels for Source.unit :

Bad levels for Siding : -

Good levels for Siding : KESSARGARH, KKC-MAIN

Bad levels for Buyer : CESC

Good levels for Buyer : NFL

Bad levels for Plant :

Good levels for Plant :

Bad and good levels of different factors for subsidiary CCL

Bad levels for Area : ARGADA

Good levels for Area : BOKARO & KARGALI , Dhori, KATHARA , KUJU , MAGADH-AMRAPALI , NORTH KARANPURA , Piparwar

Bad levels for Source.unit : AKK OCP, BHURKUNDA OC, BIRSA OC, GIDDI-A COLLIERY, GOVINDPUR PH-II OC, KDH OC, NCL DESK OFFICE KOLKATA, URIMARI OCP, URIMARI UGP

Good levels for Source.unit : AMRAPALI OCP, BARAREE, GIRIDIH COLLIERY, ROHINI OCP, SELECTED DHORI OCM

Bad levels for Siding : CENTRAL SAUNDA, GIDI-A(CHP), JARANGDIH PF-II, KARGALI WASHERY PF-II, SAUNDA

Good levels for Siding : GIRIDIH(CP), RCM

Bad levels for Buyer : (TANGENDCO + NLC)-NTPL, DPCB/PPAP

Good levels for Buyer : DSTP, DVC, HIL, MGTPP-Jhajhar, ROSA, TVNL

Bad levels for Plant : "APPDCL, Sri Damodaram Sanjeevaiah TPS", Sipat

Good levels for Plant : Bokaro, Chandrapura, Durgapur Steel Plant, Hindalco Industries Ltd (40 MW), Jhajhar, Paricha, Ropar, Simhadri, Tanda, Tenughat, West Coast Paper Mills Ltd, Yamunanagar, koderma, mejia

Bad and good levels of different factors for subsidiary ECL

Bad levels for Area : BANKOLA, KUNUSTORIA, MUGMA, S P MINES, SALANPUR, SATGRAM, SODPUR, SRIPUR

Good levels for Area : JHANJRA, KAJORA, PANDAVESWAR, RAJMAHAL

Bad levels for Source.unit : AMRITNAGAR COLLIERY, BHANORA WEST BLOCK UG, BHANORA WEST OC PATCH, CHINAKURI COLLIERY (III), Chitra-A OC, DUBESWARY, GOURANDI BEGUNIA COLLIERY, KALIPAHARI OCP, MOUTHDIH, NARAYANKURI

OC-PATCH, NARSUMDA COLLIERY, NIGHA COLLIERY, NINGHA, PARBELIA COLLIERY, PATMOHANA COLLIERY, SRIPUR SEAM INCLINE

Good levels for Source.unit : BANKOLA COLLIERY, BARMURI OCP, BONBAHAL OCP G6, Bahula G5, CENTRAL KAJORA COLLIERY, CHORA COLLIERY, CHORA INCLINE, DALURBAND OCP, GOURANDI COLLIERY, I & II INCLINE, III & IV INCLINE, JAMBAD COLLIERY, JAMBAD OCP, KAURDIH COLLIERY, KHANDRA COLLIERY, KHASKAJORA COLLIERY, KHOTTADIH COLLIERY, KHOTTADIH OCP, KUMARDIHI (A) COLLIERY, KUMARDIHI (B) COLLIERY, MADHABPUR COLLIERY, MADHABPUR OCP, MADHUSUDANPUR COLLIERY, MAHABIR OC PATCH, MAIN INCLINE, MOIRA COLLIERY, NABAKAJORA COLLIERY, NAKRAKONDA OCP, NEW KENDA COLLIERY, NIMCHA COLLIERY, PANDAVESWAR COLLIERY, PARASCOLE COLLIERY, PARASEA OCP EXT, PURANDIP, RAJMAHAL OCP, RAJPURA OCP, S.B. PROJECT, SHANKARPUR COLLIERY, SHANKARPUR OCP, SHYAMSUNDARPUR COLLIERY, Siduli G4, Siduli JK G5, TILABONI COLLIERY

Bad levels for Siding : -, BELBAID, Bonjemihari, CHINAKURI 3, DN I, JAMTARA

Good levels for Siding : 10, 12, BANKOLA 1, KK/PE, NKC, P/D, POCP I, RJ (I/R), RJ (W/W), UKHRA V

Bad levels for Buyer : (TANGENDCO + NLC)-NTPL, DVC, Jhajjar Power Ltd.

Good levels for Buyer : HPGCL, LPDC/ DCC, MGTPP-Jhajhar, NTECL, NTPC, NTPC FSTP, NTPL,NLC Tamilnadu Power.

Bad levels for Plant : "APPDCL, Sri Damodaram Sanjeevaiah TPS", 800, BUDGEBUDGE (BGB), Bandel, Barauni. Bihar, Barh, Dadri, Durgapur Steel Plant, Kolaghat, MIGK, North Chennai, Simhadri, durgapur steel TPS, koderma, mejia

Good levels for Plant : Kahalgaon, paradeep

Bad and good levels of different factors for subsidiary MCL

Bad levels for Area : IB VALLEY , ORIENT , REG HOSP NOWRAZABAD, TALCHER

Good levels for Area : LAKHANPUR , LINGARAJ

Bad levels for Source.unit : BHUBANESWARI OCP, HIRAKHAND BUNDIA, JAGANNATH OCP, KANHIA OCP, LAJKURA OCP, SAMLESWARI OCP, TALCHER UG

Good levels for Source.unit : LINGARAJ OCP

Bad levels for Siding : Kaniha MGR-I, Kaniha MGR-II, Kanika Siding, LOCM I SIDING, LOCM II SIDING, LOCM III SIDING, Spur-I, Spur-II, Spur-III, Spur-IV, Spur-VI

Good levels for Siding : -, Lakhanpur Road Sale, Lingraj MNKT

Bad levels for Buyer : D.B.POWER

Good levels for Buyer : GMR, HIL, MBMB, OPGC, Shyam Metalics and Energy Ltd

Bad levels for Plant : Bhusan steel, DB POWER, mejia, paradeep

Good levels for Plant :

Bad and good levels of different factors for subsidiary NCL

Bad levels for Area : AMLOHRI

Good levels for Area : BLOCK-B(GORBI), DUDHICHUA, DUDHICHUA PROJECT, KAKRI, KHADIA, KRISHNASHILA

Bad levels for Source.unit : 103700

Good levels for Source.unit : AMLOHRI PROJECT, BINA PROJECT, BLOCK-B PROJECT, DUDHICHUA, DUDHICHUA PROJECT, JAYANT PROJECT, KAKRI PROJECT, KHADIA, KHADIA PROJECT, KRISHNASHILA PROJECT, NIGAHI PROJECT

Bad levels for Siding : NIGAHI MGR SILO, SPUR-1 Good levels for Siding : BINA WHARF WALL, KHADIA MGR CHP, SPUR-2

Bad levels for Buyer :

Good levels for Buyer :

Bad levels for Plant :

Good levels for Plant :

Bad and good levels of different factors for subsidiary NEC

Bad levels for Area :

Good levels for Area :

Bad levels for Source.unit :

Good levels for Source.unit :

Bad levels for Siding : TIRAP

Good levels for Siding :

Bad levels for Buyer : NTPC

Good levels for Buyer : NTPC Farakka

Bad levels for Plant : 903

Good levels for Plant :

Bad and good levels of different factors for subsidiary SECL

Bad levels for Area : Bhatgaon, Bishrampur , Chirimiri, Dipka, Gevra, Hasdeo, J&K, Johilla, Korba, Kusmunda, Raigarh, Sohagpur

Good levels for Area : Baikunthpur

Bad levels for Source.unit : Amadon OCP, Chirimiri UG (Bartunga), Gauri Exp(A) OCM, Kurasia OC, Kuresia, Raj OCM, RaniAtari (Sindurgarh), SASTI OC, Vijay West, West Chirimiri

Good levels for Source.unit : Amera Project, Amlai OC, BALLARPUR 3 & 4 PITS, Bangwar, Barud, Beheraband P/M, Bhatgoan, Bijuri, Bishrampur OC, Chhal OC, Chirimiri OC, Damini, Dhanpuri OC, Dhanpuri UG, Dipka Exp Pro, Gevra Project, Jhilimilli, Kanchan, Katkona 3/4, Khairaha, Kurja, Kusmunda Project, Mahan-II, Manikpur OC, Norozabad, Pali, Pandavpara, Pinora, Piparia, Rajendra UG, Sharda HW, UKNI, Umaria, Vindhya

Bad levels for Siding : -, 180, BIJURI SIDING, Bhatgaon, Burhar, Burhar-A, CSEB OWS, Duman Hill, Govinda, Mahan II Wharf Wall (1), Manikpur II, NCPH, NEW RAJNAGAR SIDING, Nowrozabad (I), Nowrozabad (II), RAJNAGAR RO, Surakachhar, TUNDU

Good levels for Siding : Dipka, Gevra Silo/MGR, Junadih, Katora 2, Kumda, Sangma OWS

Bad levels for Buyer :

Good levels for Buyer :

Bad levels for Plant : "APPDCL, Sri Damodaram Sanjeevaiah TPS", #NAME?, ADANI POWER MAHARASHTRA LTD TIRODA, Aryan Ispat and Power Pvt 130211, Bhilai, Bhusawal, Chandrapur, Chennai, Chhabra, DB POWER, Gandhinagar, Janjgir, Jindal Power Ltd., Tamnnar, Raigarh (CG)., Kalisindh TPS, Khaparkheda, Koradi, Korba-(E)-Exp, Korba-(E)., Korba-STPS, Kota, LANCO, Marwa, Mauda, Nabha power limited , rajpura, Nasik, Paricha, RGTPP(Hissar), Ropar, S.Gandhi/Birsinghpur, Sabarmati-Ahmadabad, Sarni/Satpura, Seoni, Shri Singaji Khandwa, Sipat, Suratgarh, Ukai, WARORA, Wanakbori

Good levels for Plant : AMBUJA CEMENT

Bad and good levels of different factors for subsidiary WCL

Bad levels for Area : Ballarpur , Chandrapur , Nagpur Area , Umrer

Good levels for Area : Majri , Pathakheda

Bad levels for Source.unit : AMBARA, BARKUHI OCM, BHATADI OC, BHOJUDIH WASHERY, CRC, DRC, DURGAPUR OC, GOKUL, GONDEGOAN, HLC-1, HLOC, INDER UG CON TO OC, JUNAD, KAMPTEE OC, MAHAKALI, MAHUDA WASHERY, MKD - 1, MUNGOLI, NAIGOAN, NILJAI II, NILJAI SOUTH, PADMAPUR OC, PATHERDIH WASHERY, PAUNI II & III OC, Pinora, UKNI, UMRER OC

Good levels for Source.unit :

Bad levels for Siding : -, BALLARPUR CHP, BG SIDING, BHATADI OC (COST PLUS), CRC SIDING, DUMRI KHURD , DURGAPUR OC, EDC SIDING, GHUGUS NEW SIDING, GHUGUS OLD SIDING, GHUGUS OLD SIDING , HLC SIDING, PADMAPUR(UTS/ MGR) SIDING, PALACHOURI SIDING, UMRER NEW SIDING, UMRER NEW SIDING , WANI SIDING

Good levels for Siding :

Bad levels for Buyer : HPGCL, KPCL, Lalitpur Power, MAHAGENCO, NTPC

Good levels for Buyer :

Bad levels for Plant : Lalitpur, Panipat

Good levels for Plant :