OLUWATOBI JOHN OLUWAFEMI – DATA SCIENTIST – 006191 – ANALYSIS REPORT

TASK I – Identifying firms to be giving most attention in resource allocation.

Problem: There are scarce resources available for allocation to existing firms.

Aim: To prioritize firms for resource allocation based on three factors:

firm size,

business profile (yearly substantial change)

and outliers,

through adequate data analysis.

Domain Knowledge

The available metrics are studied to understand their meaning and relevance, therefore obtaining sufficient domain knowledge to deal with the task. Python programming and Microsoft excel sheet would be used in this analysis.
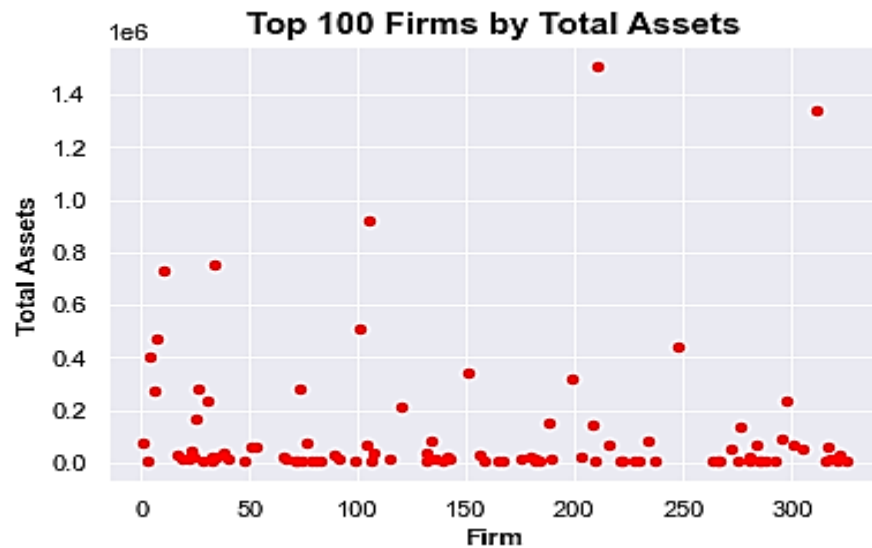
Dataset

The technical data given contains a total of **325 firms** with over several metrics including equity, total assets, and Net Written Premium (NWP) for **a period of five years** (2016 – 2020).

This technical data is utilized to observe the metrics available and determine which would be most relevant in achieving the aim stated. Each preceding factor stated in the aim is prioritized against the next.

A joined data set containing NWP, GWP, total assets, gross claims incurred, net combined ratio, equity, and SCR coverage ratio metrics for the five-year period on each of the 325 firms is created. These metrics have been selected for use in this analysis.

Factor 1 - Determining largest 100 firms by asset size

The metric '**assets**', for the 5-year period, is used an indicator for firm size and the top 100 is selected. Firms with missing data are highlighted in the analysis excel sheet. From year 2016 to 2020, the assets are added up to give a total asset value which is plotted against the firms.

Figure 1 Total Assets scatter plot for top 100 largest firms

It is observed that most firms fall within a certain range of total assets as seen by clusters at the bottom, however some firms happen to be way beyond the typical range. These outliers as seen in Figure 1 would be filtered out later in this analysis.

Factor 2 – Changing Business Profile

In a bid to determine if these 100 firms are growing substantially or not, the Net Written Premium (NWP) metric for the five-year period is utilised. If the NWP rises over time, it suggests that the firm is growing, and could indicate more policyholders within the firm.

To achieve this, the difference between NWP values of each year and the preceding year is calculated and added up. If this total score is positive, it indicates an overall/substantial growth in the firm meanwhile negative values indicate an overall decline and zero values mean no substantial growth.
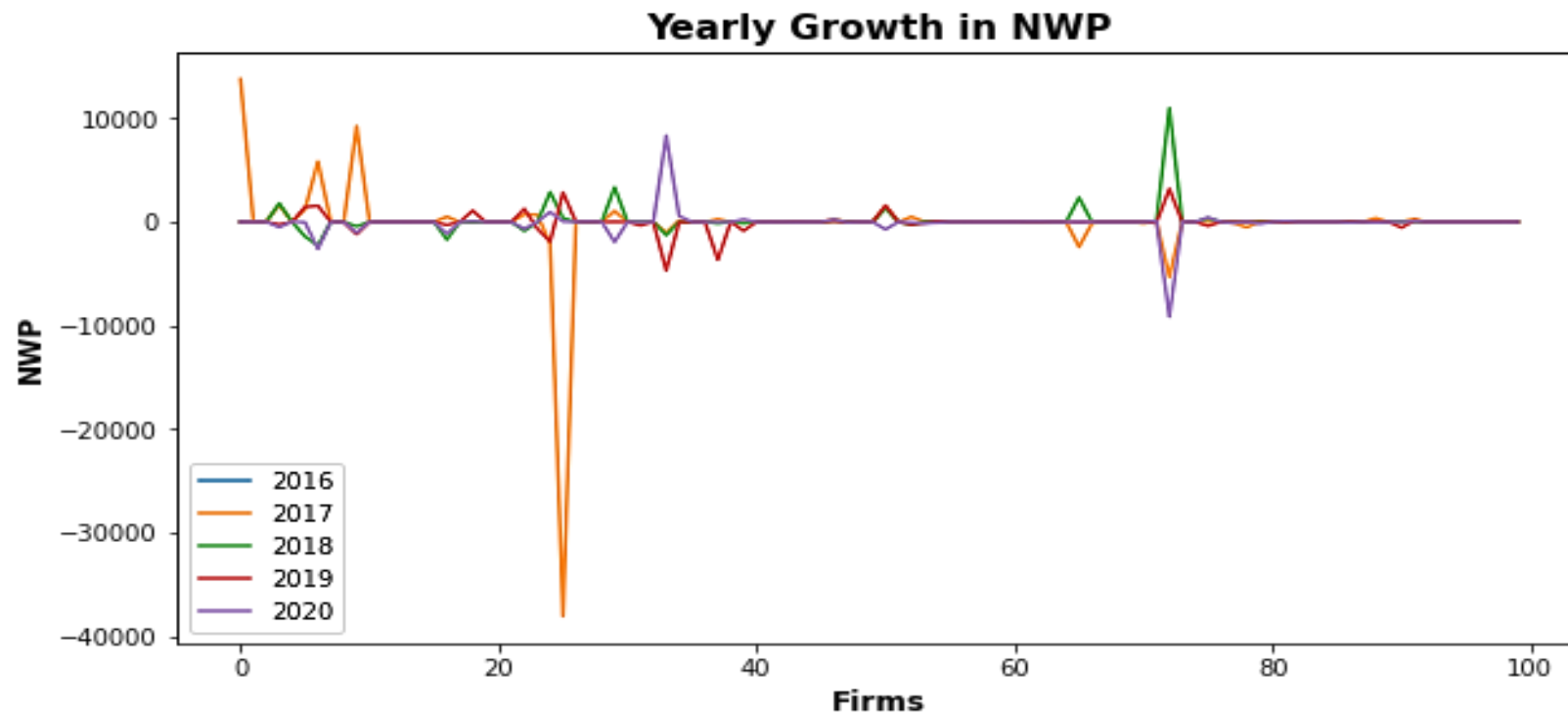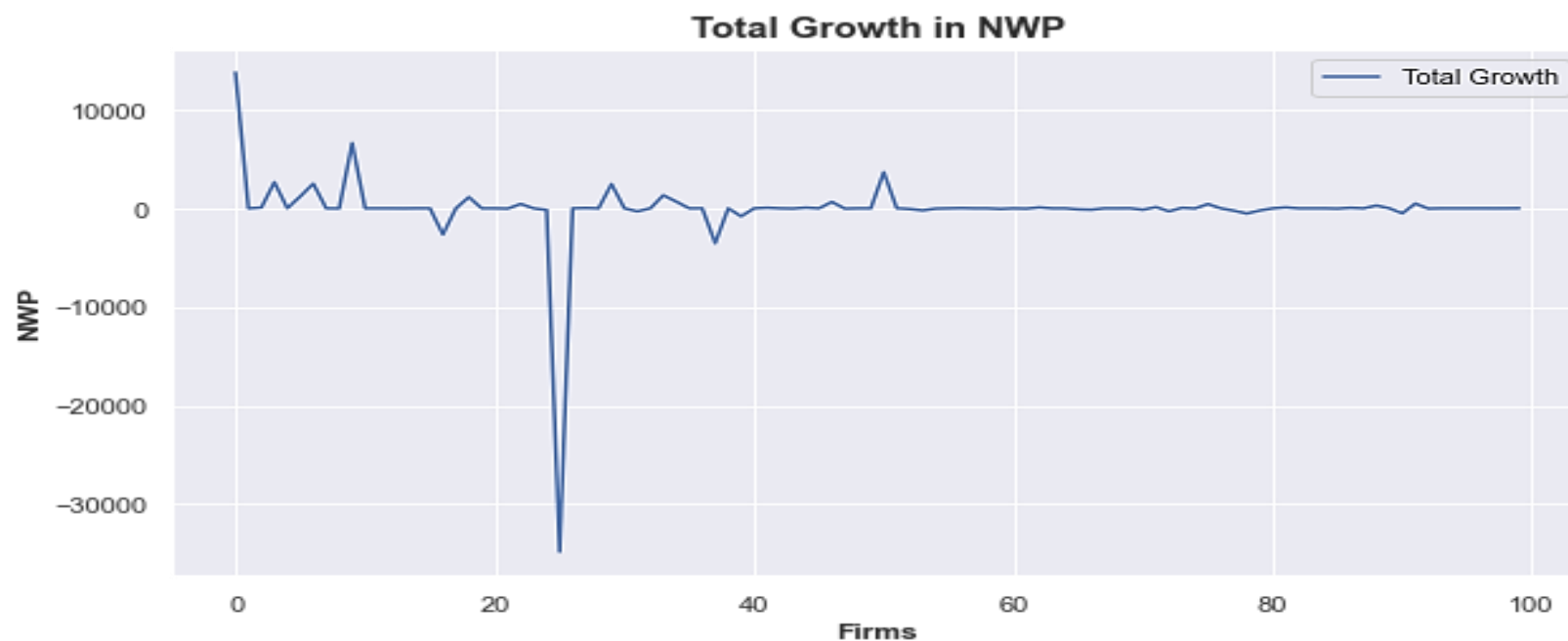

Figure 2 Yearly NWP Growth trend for top 100 largest firms

Figure 3 Total NWP Growth trend for top 100 largest firms

In both charts as shown in Figure 2 and Figure 3, the peaks and troughs indicate growth and decline respectively, however, very high peaks and troughs are likely to be outliers or errors (due to missing values) such as the first firm- firm 1. These are further shown in Table 2 in the Missing Data and Errors Section. The firms with a positive total growth value are filtered out for further analysis.

Factor 3 – Significant deviation for the 2020-year period (Outliers and errors)

The firms with positive total growth filtered out total 41 and they are observed to all meet PRA requirement of SCR coverage ratio greater than 1 for the year 2020. Two of these firms are observed to have missing metrics data for year 2020 and where dropped, leaving 39 firms. To understand the distribution of the NWP and Asset metrics used in the preceding factors, the following plots in Figure 4 and Figure 5 where made.
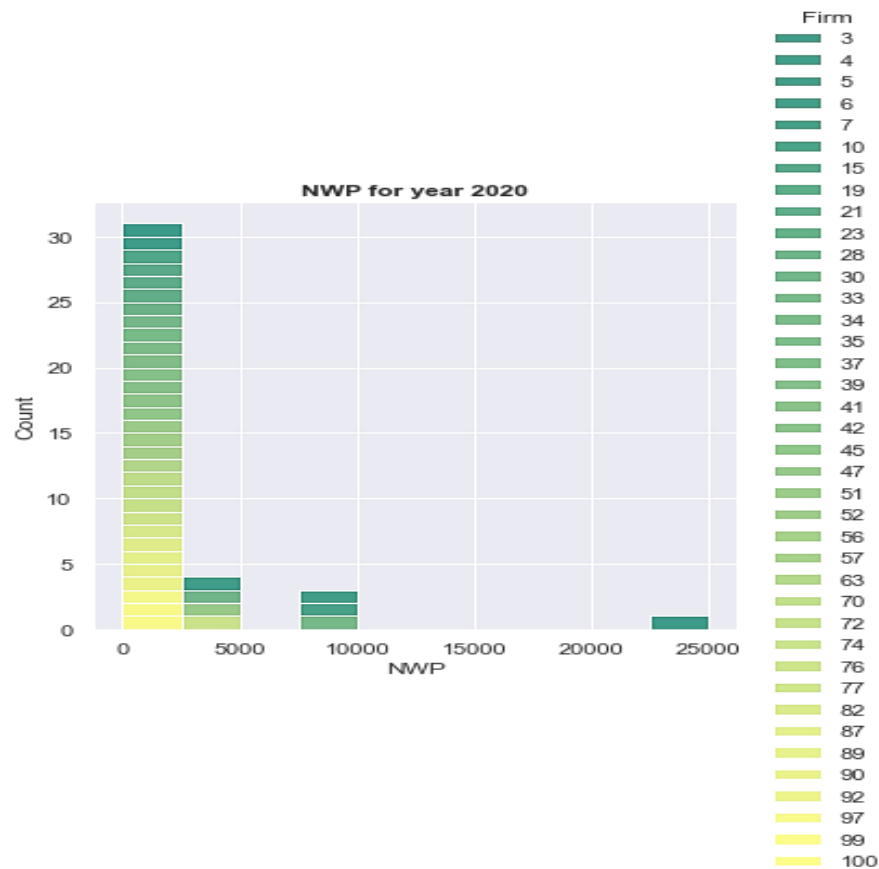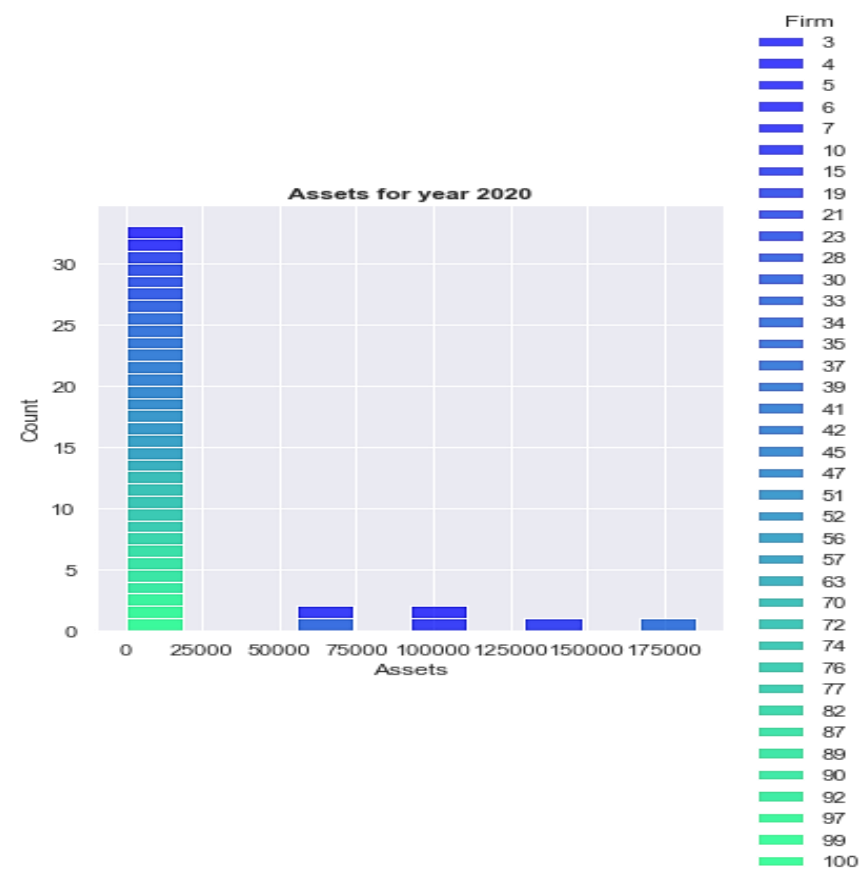

Figure 4 NWP Distribution for year 2020

Figure 5 Asset Distribution for year 2020

It is observed from the plots Figure 4 and Figure 5, that over 30 firms out of 41, seem to fall into one bin of values both for the NWP and Assets. To identify the deviations and outliers observed, the z-scores are calculated by the mean and standard deviation. Firms with Z-scores less than one for both NWP and Assets are labelled priority else they are labelled non-priority. The total z-scores are also calculated, which is taken as the sum of the NWP and Asset z scores.
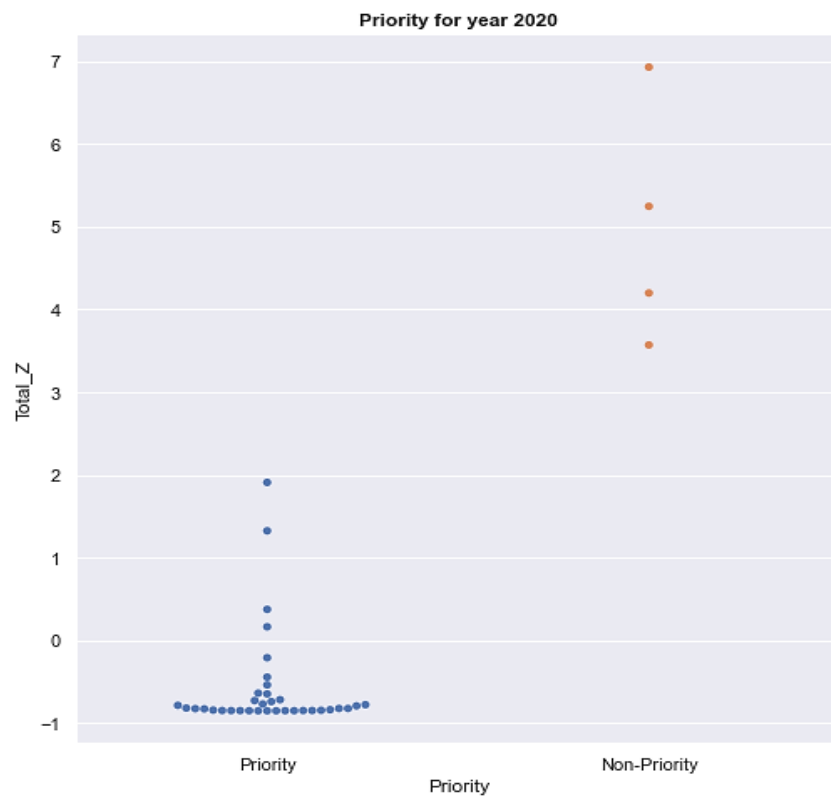

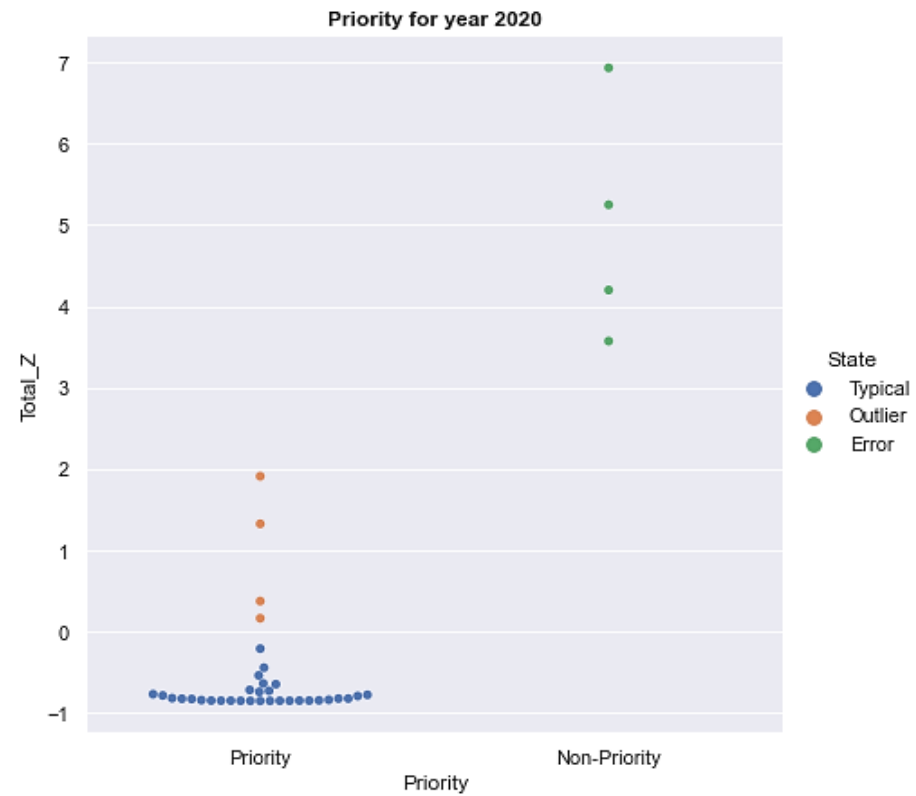*Figure 6 Priority plotted against total z score.*

Figure 7 Priority plotted against total z score with state shown

From the plots in Figure 6 and Figure 7, a bowl-like group of blue dots is observed indicating the "typical" range for the total z values. Based on the plot, points above two are classified as errors, points below 2 but above zero are classified as genuine outliers.

The firms who are not classified as errors total 35 and are finally filtered out to make up the **final list of firms** that should be giving the most attention in allocating scarce resources, highlighted in blue in Table 1.

Table 1 Final list of firms that should be giving the most attention.

Firm	NWP_Z	Assets_Z	Total_Z	Priority	State	Firm	NWP_Z	Assets_Z	Total_Z	Priority	State
100	-0.402442423	-0.433993903	-0.836436326	Priority	Typical	57	-0.402753085	-0.437925691	-0.840678776	Priority	Typical
21	-0.409764166	-0.438588981	-0.848353148	Priority	Typical	70	-0.40514815	-0.438267095	-0.843415245	Priority	Typical
3	-0.390526214	-0.399844541	-0.790370756	Priority	Typical	82	-0.310035401	-0.405229248	-0.715264649	Priority	Typical
45	-0.39000943	-0.426761572	-0.816771002	Priority	Typical	52	-0.407604425	-0.439550453	-0.847154878	Priority	Typical
28	-0.397845556	-0.425697867	-0.823543424	Priority	Typical	5	-0.391527677	-0.434884685	-0.826412363	Priority	Typical
19	-0.15302509	-0.289433648	-0.442458737	Priority	Typical	74	-0.359886186	-0.41604579	-0.775931976	Priority	Typical
15	-0.407526318	-0.439667858	-0.847194176	Priority	Typical	33	-0.409656799	-0.438600572	-0.848257371	Priority	Typical

99	-0.37153432	-0.412290968	-0.783825288	Priority	Typical	92	-0.301128123	-0.437382487	-0.738510611	Priority	Typical
6	0.272931116	1.052682299	1.325613415	Priority	Outlier	37	-0.409484593	-0.440009875	-0.849494468	Priority	Typical
76	0.182403571	-0.016876223	0.165527348	Priority	Outlier	51	0.412061826	-0.036155696	0.375906129	Priority	Outlier
97	-0.407774276	-0.438555912	-0.846330188	Priority	Typical	72	-0.318087735	-0.408029129	-0.726116863	Priority	Typical
30	0.69996113	1.210398194	1.910359323	Priority	Outlier	39	-0.407469711	-0.438699781	-0.846169492	Priority	Typical
42	-0.396169052	-0.425047623	-0.821216675	Priority	Typical	41	-0.408375559	-0.440148593	-0.848524152	Priority	Typical
63	-0.388392958	-0.432872889	-0.821265847	Priority	Typical	89	-0.344626755	-0.291560492	-0.636187248	Priority	Typical
77	-0.409656035	-0.43979214	-0.849448175	Priority	Typical	23	0.014931945	-0.223118269	-0.208186324	Priority	Typical
87	-0.332876092	-0.434890232	-0.767766323	Priority	Typical	47	-0.263297652	-0.382343703	-0.645641354	Priority	Typical
35	-0.212869036	-0.324091352	-0.536960388	Priority	Typical	56	-0.409768219	-0.440124528	-0.849892747	Priority	Typical
90	-0.409767521	-0.436300489	-0.846068011	Priority	Typical						

Task II: Using Machine Learning Techniques for further insights

Firms with complete data on the following metrics: NWP, GWP, Assets, equity, SCR, SCR coverage ratio, Net combined ratio, and Gross claims incurred are filtered out. Their total z-score (NWP z-score + Asset z-score) is calculated as it reflects their state (typical, error or outlier). An ML feature selection technique is then used to identify the correlation between these metrics and their state. From the chart. NWP, GWP, Assets, Equity and SCR are strong indicators of the priority and state of a firm as seen in Figure 8 and Figure 9.
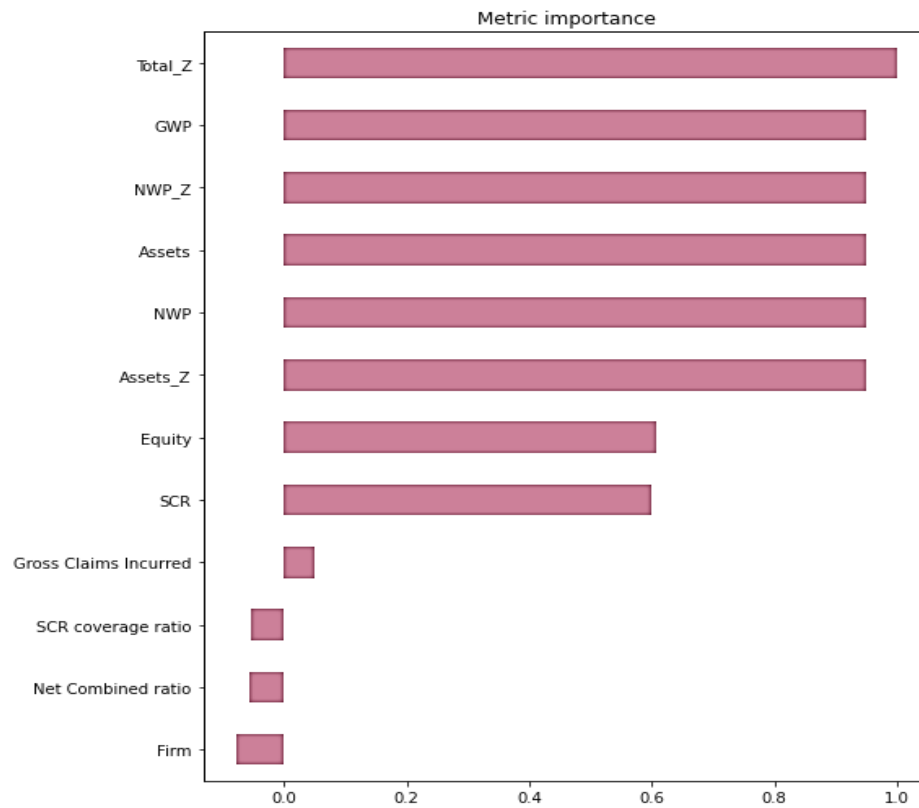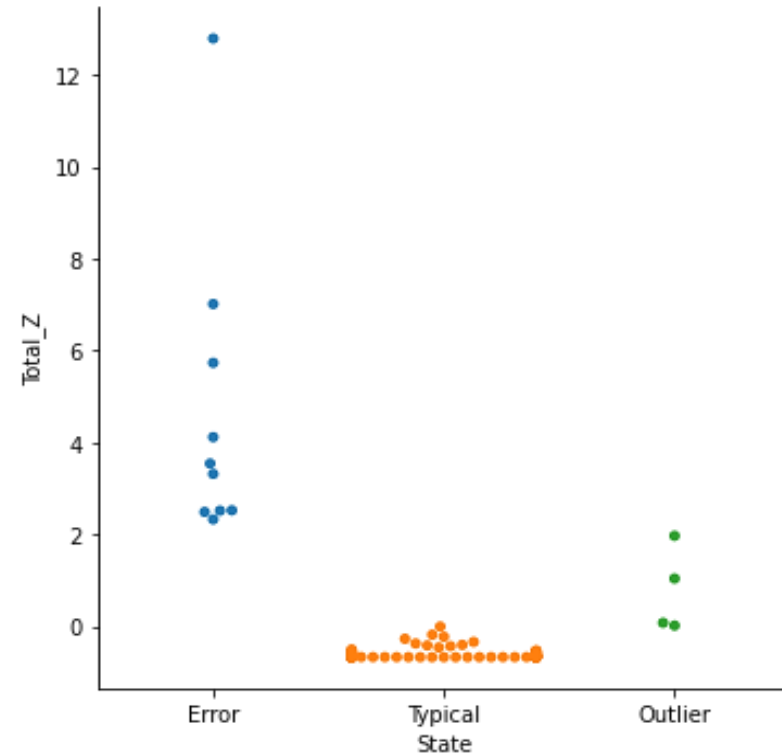

Figure 8 Metrics correlation to z-score   *Figure 9 State distribution plotted against total z-score*

Regression plots between these metrics are then made. GWP should have a linear relationship with NWP as an indicator of changing business profile as used in task one. This is supported by the regression plots in Figure 10. Equity and SCR are seen to have the same relationship as observed in figure 11.
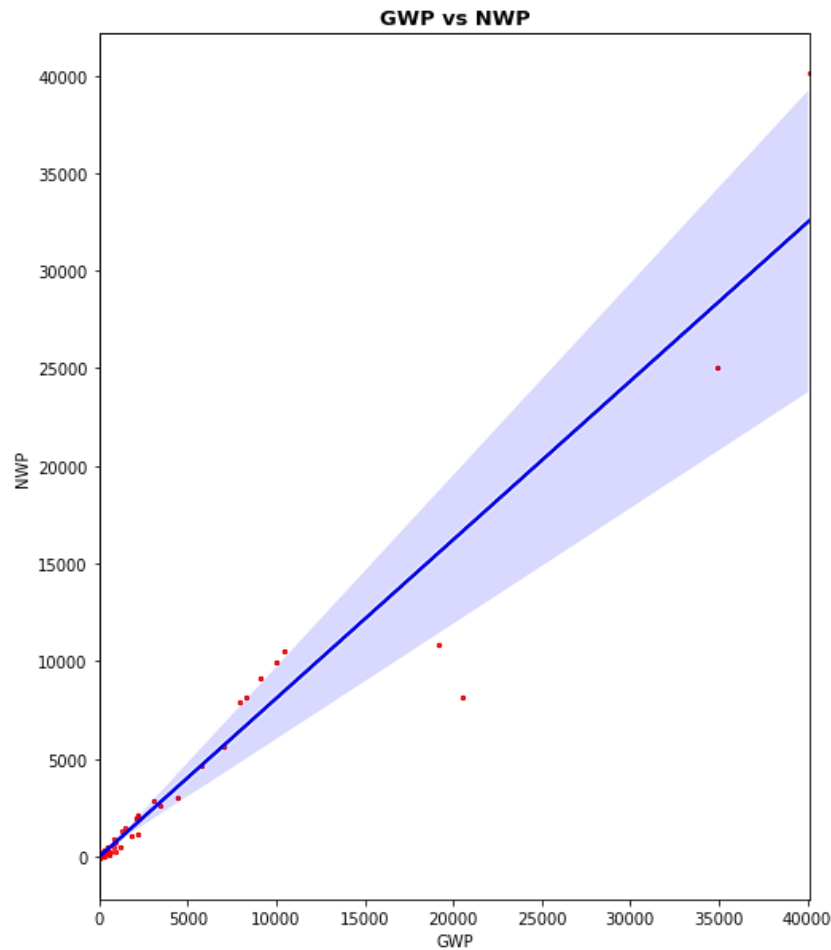

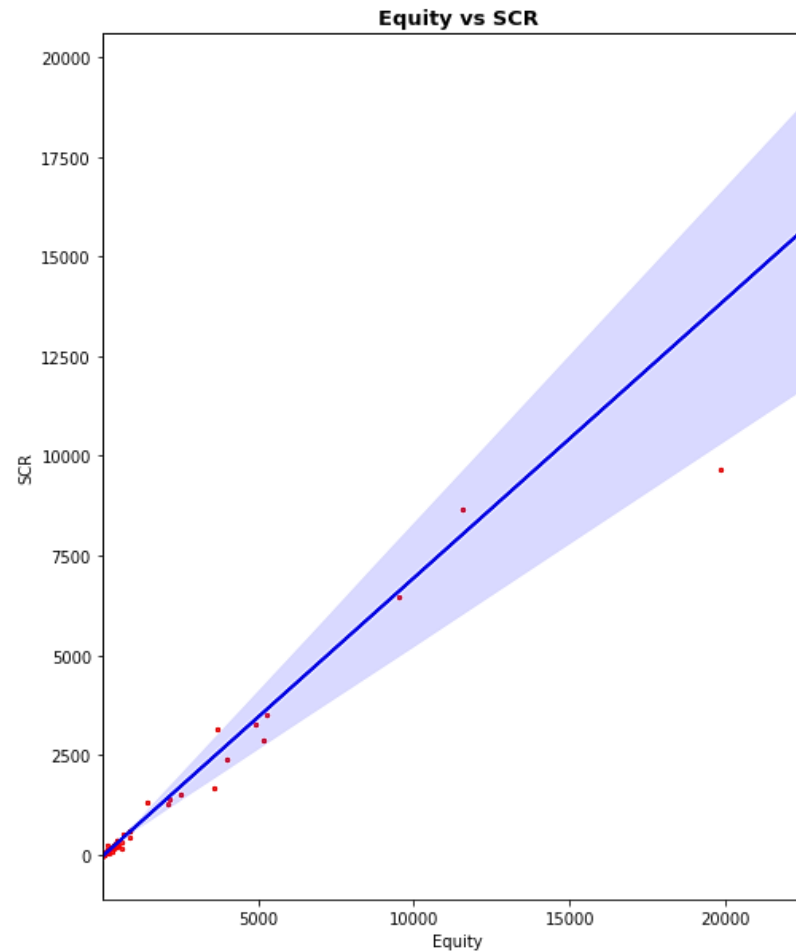*Figure 10 Gross Written Premium plotted against Net Written Premium*   *Figure 11 Regression plot of Equity against SCR for year 2020*

Missing Data and Errors

Table 2 NWP Total growth data for top 100 firms with missing values and errors highlighted

	Firm	Total Growth	Firm	Total Growth	Firm	Total Growth	Firm	Total Growth	Firm	Total Growth	Firm	Total Growth
0	Firm 1	13779.81563	Firm 20	0.004001764	Firm 39	-0.022544628	Firm 58	0	Firm 77	0.207106509	Firm 96	0.010365273
1	Firm 2	-6.459500483	Firm 21	0.018057899	Firm 40	-801.3542272	Firm 59	-0.852431171	Firm 78	-232.6841688	Firm 97	1.886248095
2	Firm 3	85.73582975	Firm 22	-23.29027587	Firm 41	0.005740542	Firm 60	-53.43026542	Firm 79	-523.2276713	Firm 98	0
3	Firm 4	2651.821119	Firm 23	446.1372075	Firm 42	60.59326265	Firm 61	-1.272386643	Firm 80	-212.9463469	Firm 99	1.185311955
4	Firm 5	13.07265926	Firm 24	0	Firm 43	0	Firm 62	-28.40511992	Firm 81	-0.371571721	Firm 100	10.79477415
5	Firm 6	1215.078074	Firm 25	-163.0884286	Firm 44	-20.20547769	Firm 63	95.24089649	Firm 82	97.35116948		
6	Firm 7	2504.732468	Firm 26	-34820.59062	Firm 45	72.83993718	Firm 64	-0.196064146	Firm 83	0		
7	Firm 8	0	Firm 27	0.008825217	Firm 46	-1.584292619	Firm 65	0	Firm 84	0		
8	Firm 9	-7.728284818	Firm 28	32.03350658	Firm 47	652.62304	Firm 66	100.3278789	Firm 85	0		
9	Firm 10	6619.114437	Firm 29	0	Firm 48	-27.37162443	Firm 67	-132.1569754	Firm 86	-24.64860034		
10	Firm 11	0	Firm 30	2464.193297	Firm 49	0.105358072	Firm 68	0	Firm 87	55.09597236		
11	Firm 12	0	Firm 31	-0.000809057	Firm 50	0	Firm 69	-0.031732247	Firm 88	-0.681007124		

12	Firm 13	-1.208639031	Firm 32	-300.9337023	Firm 51	3661.795209	Firm 70	4.850486798	Firm 89	290.2482105		
13	Firm 14	-6.833741197	Firm 33	4.94169012	Firm 52	9.641130182	Firm 71	-134.8933556	Firm 90	0.003109596		
14	Firm 15	4.504908554	Firm 34	1328.218082	Firm 53	-63.13786934	Firm 72	124.2266544	Firm 91	-499.6972498		
15	Firm 16	0	Firm 35	679.9020227	Firm 54	-203.5467961	Firm 73	-296.2719855	Firm 92	484.0633226		
16	Firm 17	-2642.301221	Firm 36	-0.001552334	Firm 55	-25.55645144	Firm 74	52.09917532	Firm 93	-26.00802361		
17	Firm 18	0	Firm 37	1.263742437	Firm 56	0.035850143	Firm 75	-11.09119587	Firm 94	0		
18	Firm 19	1143.960077	Firm 38	-3527.89263	Firm 57	20.9029895	Firm 76	436.2788578	Firm 95	0		

In Table 2, data highlighted in blue are missing values and those highlighted in yellow are examples of errors as they fall far above the average values. This could have also been due to missing NWP data for certain year periods.

The Net combined ratio which indicates profit as observed from the plot in Figure 12 to fall within a certain range irrespective of asset size. Therefore, red points seen far above this range are likely errors.
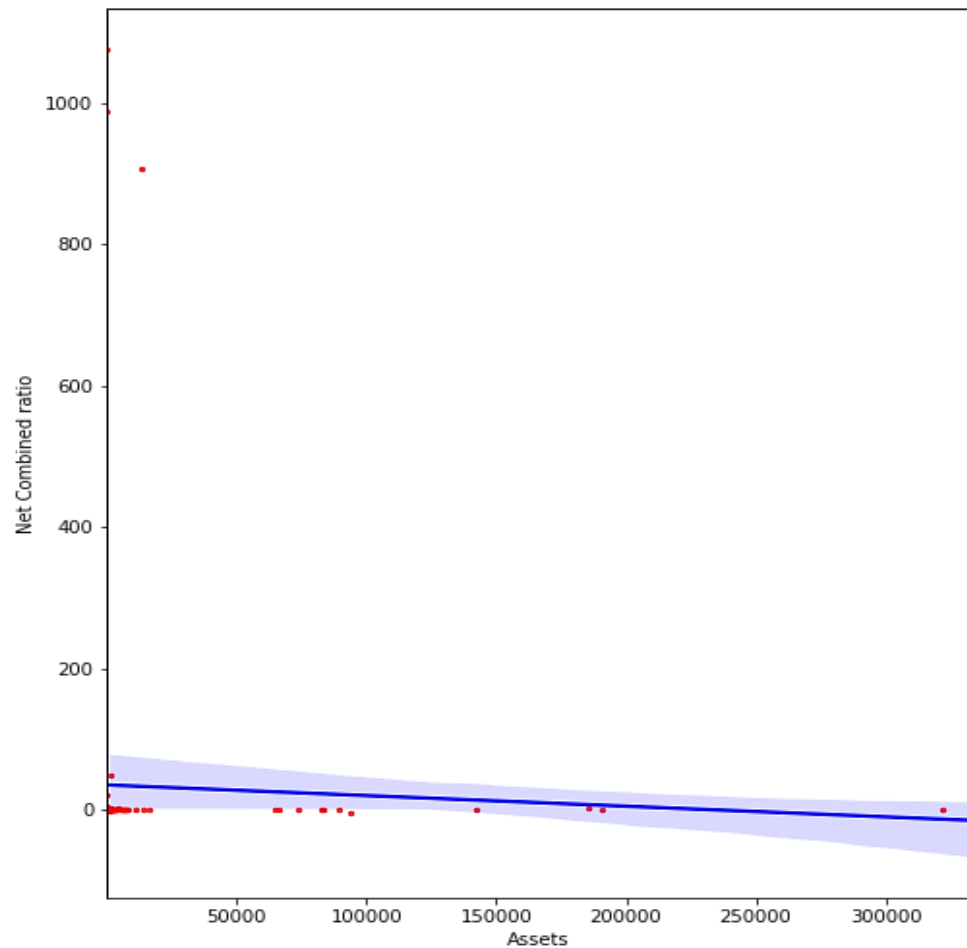

Figure 12 Net Combined ratio plotted against Assets for year 2020.