**Audit Data Analytics**

**Service Area Grouping and Anomaly Detection**

1. **Grouping Service Areas via Unsupervised Learning**

The 24 service areas can be grouped into four clusters based on their similarities in transaction averages and transaction counts. The scatter plot next page illustrates where each of the service area is in each of the coloured clusters. The table below it gives details about service area names and their associated clusters. The clustering algorithm used is K-means, which is based on Euclidean distances.

Grouping service areas based on similarities in spending behavior has some key benefits for conducting audits.

* Enhancing anomaly detection: Clustering allows auditors to establish a baseline for what constitutes 'normal' behavior for each group of service areas. Transactions that significantly deviate from the norms of their respective clusters can be flagged as potential anomalies. In addition, instead of comparing transactions across all service areas, auditors can focus on deviations within each cluster, making anomaly detection more targeted and efficient.
* Prioritizing audit efforts with risk-based auditing: Clustering can help in prioritizing audit efforts based on the risk profiles of different clusters. Service areas in clusters with large numbers of transaction (e.g., Children's Family Services) and high average transaction amounts (e.g., Customer Support Group) might be audited more frequently. This, when combined with historical control deficiencies and exceptions noted, will help auditors to allocate resources more effectively and focus on areas that are subject to higher risks.
* Strategic insights and reporting: Grouping service areas allows for detailed and structured reporting. Auditors can present findings by cluster, providing a clearer narrative of the auditee, showing their deep understanding of the client business, and solidifying auditors’ positions as trusted partners toward their stakeholders.

A graph with numbers and dots

Description automatically generated with medium confidence

|  |  |  |  |
| --- | --- | --- | --- |
| **Service Area** | **total\_transaction\_count** | **average\_amount** | **Cluster** |
| Adults and Communities | 155 | 153.23 | 0 |
| Assurance | 246 | 70.46 | 0 |
| Commercial | 9 | 304.78 | 0 |
| Children's Service DSG | 231 | 128.39 | 0 |
| Deputy Chief Operating Officer | 113 | 36.89 | 0 |
| Corporate Accounting | 4 | 132.98 | 0 |
| Control Accounts | 8 | 23.84 | 0 |
| Commissioning | 258 | 161.57 | 0 |
| Education | 95 | 134.90 | 0 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Service Area** | **total\_transaction\_count** | **average\_amount** | **Cluster** |
| Street Scene | 39 | 64.93 | 0 |
| Strategic Commissioning Board | 1 | 244.00 | 0 |
| Regional Enterprise | 5 | 375.40 | 0 |
| Public Health | 3 | 0.73 | 0 |
| Parking & Infrastructure | 5 | 156.18 | 0 |
| NSCSO | 3 | 151.83 | 0 |
| Internal Audit & CAFT | 11 | 65.10 | 0 |
| Streetscene | 257 | 172.74 | 0 |
| Children's Family Services | 5318 | 65.55 | 1 |
| Customer Support Group | 96 | 1,548.85 | 2 |
| CSG Managed Budget | 36 | 1,971.84 | 2 |
| Governance | 8 | 1,780.31 | 2 |
| Family Services | 770 | 62.31 | 3 |
| Children's Education & Skills | 642 | 95.41 | 3 |
| Childrens Services | 1297 | 70.63 | 3 |

1. **Anomalous Transactions**

With limited negative transaction amounts, most service areas have transaction amounts follow positively skewed distributions. The skewed sides, however, look like the shape of normal distributions with flattened tails toward the end.

I tried different anomaly detection techniques (z-score of normal distribution, z-score of log-normal distribution, and Isolation Forest) and decided that z-score of normal distribution is the most appropriate for this merged dataset. Thus, transactions with amounts beyond 3 standard deviations are considered anomalous. Since the data is somewhat right skewed, we can increase the threshold of 3 standard deviations to the right side if we need to limit the number of detected anomalies.

In total, 172 anomalous transactions are noted. You can see them in the listing attached below. Please be aware that due to differences in numbers of transactions and underlying patterns among service areas, service areas that have only few transactions do not have anomalies or only have few anomalies representing in the anomaly list (see the scatter plots of “Assurance” and “NSCSO” next page). On the other hand, service areas with large numbers of transactions and great variability within transaction amounts have more anomalous transactions detected (e.g., see the scatter plot of “Children’s Family Services” below).



A graph of numbers and dots

Description automatically generated with medium confidence

A graph of numbers and a line

Description automatically generated with medium confidence

A graph with numbers and a line

Description automatically generated

**Closing Note**

If you find these visualizations and insights useful, we can productionize this data analysis into automated BI solutions for continuous monitoring. Also, let me know if there were areas you would like to analyze further.