Chad Banfield- Amini technical assignment report

During initial data quality assessment of the file, the following characteristics were observed:

- **Total records**: 13,013

- **Completeness issues**: No column contained 13,013 entries, indicating missing values across all fields.

- **Missing value distribution**:

| Column Name | Total Entries | Missing Enties |
|---|---|---|
| facility_id | 9305 | 3708 |
| facility_name | 12989 | 24 |
| facility_type | 11015 | 1998 |
| capacity | 10361 | 2652 |
| region | 12980 | 33 |
| licence_issue_date | 12990 | 23 |
| inspection_date | 13002 | 11 |
| gps_location | 10721 | 2292 |
| remarks | 10080 | 2933 |

- **Data types**: All columns stored as object (string) type.

- **Duplicates**: 1,672 fully duplicated rows detected.

Average capacity per region including missing regions:

| Region | Average | Standard deviation |
|---|---|---|
| christ church | 133 | 161 |
| missing | 136 | 161 |
| st. andrew | 145 | 165 |
| st. george | 144 | 162 |
| st. james | 151 | 162 |
| st. john | 154 | 163 |
| st. joseph | 142 | 165 |
| st. lucy | 146 | 165 |
| st. michael | 143 | 162 |
| st. peter | 157 | 168 |

The substantial standard deviations (ranging from 161 to 168) indicate high variability in capacity values within each region. This variability is reflected in the extreme values observed: while the maximum capacity per region approaches approximately 500, the minimum non-zero value (excluding missing entries coded as 0) is consistently 1 across all regions.

Chad Banfield- Amini technical assignment report

| Region | MIN | MAX |
|---|---|---|
| christ church | 1 | 499 |
| missing | 1 | 497 |
| st. andrew | 1 | 500 |
| st. george | 1 | 500 |
| st. james | 1 | 499 |
| st. john | 1 | 499 |
| st. joseph | 1 | 500 |
| st. lucy | 1 | 500 |
| st. michael | 1 | 500 |
| st. peter | 1 | 500 |

A notable observation is the absence of St. Thomas and St. Philip in the regional data. Three plausible explanations exist:

1. These parishes may be included in the 'missing' region category,

2. They might lack medical facilities of the type captured in this dataset, or

3. Data for these parishes could have been excluded during collection.