**NANYANG POLYTECHNIC**

**IT1394 VISUAL ANALYTICS**

**DATA PREPARATION & MODELLING**

**Cheng Sheh Nee | 234745J**

**Table of Contents**

1. Scope of Datasets
2. Data Preparation
3. Data Visualisation

3.1 Dashboard Prototype

3.2 Dashboard Iteration and Feedback

3.3 Final Dashboard and Analysis

1. Statistical Modelling
2. **Scope of Datasets**

This section focuses on the scope of datasets underpinning the **Sustainable Living** Pillar of the Singapore Green Plan 2030 (Green Plan).

|  |  |  |
| --- | --- | --- |
| **S/N** | **Dataset** | **Reason** |
| 1 | **Recycling Waste by Waste Type**  <https://beta.data.gov.sg/datasets/d_9740df787da2b59a0b5bd76a6c33453d/view> | The dataset is **included** for further analysis.  *(not shared)*   * gain insights into waste management efficiency * identify the most and least recycled materials * *all fields in the dataset will be used* |
| 2 | **Volume of waste recycled in Singapore 2013-2022**  <https://www.statista.com/statistics/628048/waste-recycled-in-singapore/> | The dataset is **included** for further analysis.  *(not shared)*   * evaluating the effectiveness of recycling programs over time * understanding trends in resource recovery * *all fields in the dataset will be used* |
| 3 | **Volume of waste disposed in Singapore from 2013 to 2022**  <https://www.statista.com/statistics/628000/waste-disposed-in-singapore/> | The dataset is **included** for further analysis.  *(not shared)*   * critical insights into the consumption patterns and efficiency of waste reduction efforts * *all fields in the dataset will be used* |
| 4 | **Waste Statistics and Overall Recycling Table**  <https://www.nea.gov.sg/our-services/waste-management/waste-statistics-and-overall-recycling> | The dataset is **included** for further analysis.  *(not shared)*   * understanding the full scope of waste management and recycling efficacy * focus only on the year 2022 to get the latest data * *all fields in the dataset will be used except the ‘overall’ row* |
| 5 | **Reverse Vending Machine**  <https://www.nea.gov.sg/our-services/waste-management/reverse-vending-machines> | The dataset is **included** for further analysis.  *(not shared)*   * evaluating the infrastructure's role in promoting recycling * offering insights into accessibility for consumers to participate in sustainable practices * *all fields in the dataset will be used with one address for each location for latitude and longitude* |
| 6 | **About NEA’s 2023 Survey on Household Recycling**  <https://www.nea.gov.sg/media/news/news/index/72-per-cent-of-households-recycle-in-2023#:~:text=3%20NEA's%202023%20survey%20on,the%20recycling%20bins%20and%20chutes>. | The dataset is **included** for further analysis.  *(not shared)*   * focusing analysis on 2023 responses only to get the latest data * offering valuable insights into the evolving attitudes and behaviours of citizens towards recycling * *selected fields in the dataset will be used (2023 datasets)* |
| 7 | **Frequency of recycling single-use plastic bottles in Singapore as of May 2023**  <https://www.statista.com/statistics/1389977/singapore-frequency-of-recycling-plastic-bottles/> | The dataset is **NOT included** for further analysis.   * limited scope * focusing on a single point in time |
| 8 | **The Total Volume of Waste Paper generated in Singapore from 2013 to 2022**  <https://www.statista.com/statistics/962178/waste-paper-generated-volume-singapore/> | The dataset is **NOT included** for further analysis.   * narrow focus on one type of waste * might not reflect broader waste management and recycling efforts across multiple materials |
| 9 | **Consumers’ Perception towards Food Packaging in Singapore in 2020**  <https://www.statista.com/statistics/1208386/singapore-perception-towards-food-packaging/> | The dataset is **NOT included** for further analysis.   * might not provide substantial, actionable insights due to its specificity to a particular year and subject * limited value in assessing overall progress |
| 10 | **Frequency of using Reusable Bags when shopping in Singapore as of June 2021**  <https://www.statista.com/statistics/1312708/singapore-frequency-of-using-reusable-shopping-bags/> | The dataset is **NOT included** for further analysis.   * focus on a specific behaviour at a single point in time * may not effectively represent ongoing trends |
| 11 | **Preference for reusable packaging among consumers in Singapore as of March 2022**  <https://www.statista.com/statistics/1314610/singapore-preference-for-reusable-packaging-among-consumers/> | The dataset is **NOT included** for further analysis.   * narrow focus on consumer preferences at a specific moment * might not adequately reflect the ongoing and comprehensive changes in consumer behavior |
| 12 | **Average number of passengers per day using public transport in Singapore from 2015 to 2022**  <https://www.statista.com/statistics/1006174/singapore-daily-public-transport-ridership/> | The dataset is **NOT included** for further analysis.   * falls outside the primary focus of the pillar * avoids overlap with my teammate's work, ensures a more organized and collaborative approach |
| 13 | **Recycling rate for food waste generated in Singapore from 2013 to 2022**  <https://www.statista.com/statistics/962933/food-waste-recycling-rate-singapore/> | The dataset is **NOT included** for further analysis.   * maintaining a clear focus within the pillar * avoids overlap with my teammate's work |

1. **Data Preparation**

This section details the processes undertaken to ensure the dataset’s integrity and relevance of analysis. This includes data cleaning and transformation and aligning it with the objectives of the analysis and the overarching hypothesis of the study.

Data cleaning and transformation:

1. **Rename tables** for readability
2. **Rename columns** when necessary
3. **Remove unnecessary columns and rows**
4. **Handle with missing values**, whether to remove rows, replace value or leave them
5. **Use the first row as the header**
6. **Data type conversion,** ensure each columns having the correct data type
7. **Add descriptions** for the tables
8. **Create date table**

Screen capture on the process of data cleaning and transformation:

*(not all steps are presented in the screen capture)*

|  |  |
| --- | --- |
| * Remove unnecessary rows and use the first row as the header * Data type conversion | |
| Before cleaning | After cleaning |
|  |  |

|  |  |
| --- | --- |
| * Remove unnecessary rows and use the first row as the header * Data type conversion | |
| Before cleaning | After cleaning |
|  |  |

|  |  |
| --- | --- |
| * Data type conversion * Handle missing value (Replace Error)   + *The original source mentioned that the recycling rate for the "Others" category is not meaningful as it is an aggregation of miscellaneous waste types that are managed and recycled differently. Hence, error is replaced from N.A.1 to 0%.* * Remove *Overall* row to avoid misleading representation in the visualisation | |
| Before cleaning | After cleaning |
|  |  |
| Before cleaning | After cleaning |
|  |  |

|  |  |
| --- | --- |
| * Create new column to convert recycling rate into decimals * Data type conversion | |
| Before cleaning | After cleaning |
|  |  |

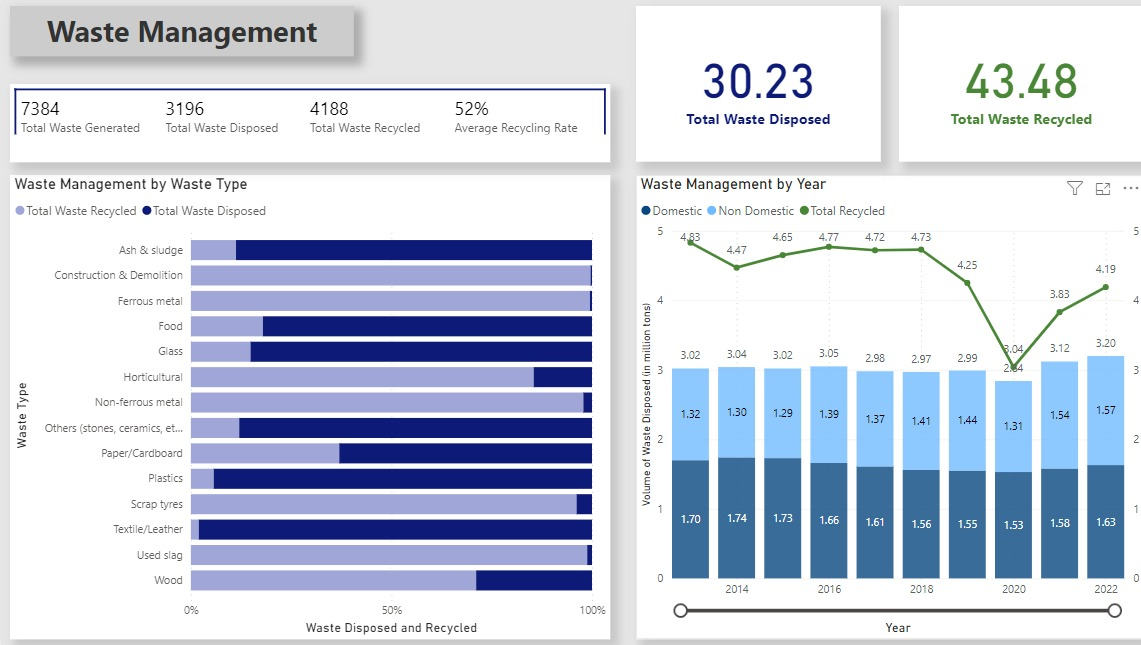
|  |  |
| --- | --- |
| * Create new column to calculate the total waste disposed | |
| Before cleaning | After cleaning |
|  |  |

The datasets are related to the hypothesis of “**The Sustainable Living pillar of the Green Plan is committed to building an eco-friendly country**”.

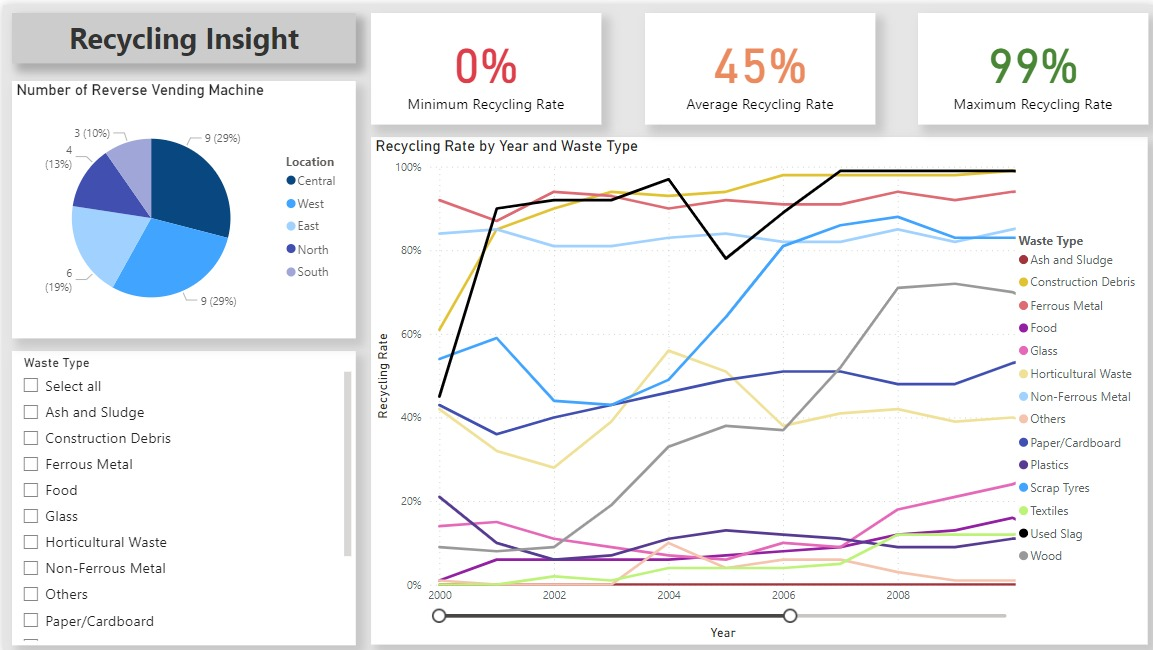
1. The datasets chosen for analysis are highly relevant to the hypothesis, as they comprise reliable and comprehensive environmental data sourced from authoritative and credible sources such as the National Environment Agency (NEA) and Statista.
2. This data provides us with valuable insights into key environmental metrics and trends, directly addressing the hypothesis and enabling a well-founded conclusion.
3. The integrity of the sources is known for their rigorous data collection and validation methods, ensure that the analysis is based on accurate and trustworthy information.
4. **Data Visualisation**

**3.1 Dashboard Prototype**

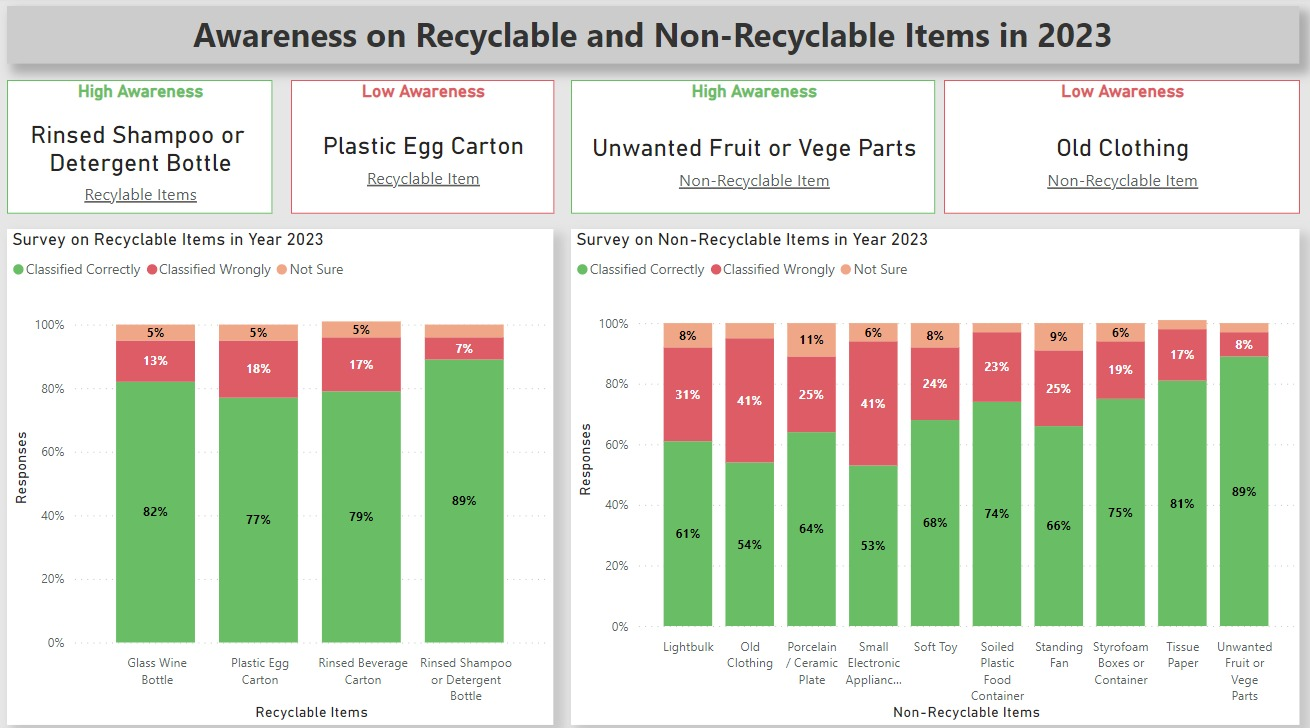
Dashboard 1: Waste Management



Dashboard 2: Recycling Insight



Dashboard 3: Awareness on Recyclable and Non-Recyclable Items in 2023



**3.2 Dashboard Iteration and Feedback**

This section presents the iterative development of the dashboards through visual screenshots and accompanying feedback logs. It's an interactive process that involves soliciting critiques from team members, using their insights to iteratively refine and enhance the dashboards’ design and functionality.

**Dashboard 1: Waste Management**

|  |  |
| --- | --- |
| 1st Iteration | |
| Before | After |
|  |  |
| Feedback | Changes |
| Title(top left)   * Background colour too dark, blending with the dashboard background with no contrast | Title (top left)   * Change background colour to make the title stands out |
| Waste Management by Waste Type (Left)   * Multi-row Card   + Prefix “Total” is not necessary for the category label   + Round corner will differentiate card with other visualisation (graph) * 100% Stacked Bar Chart   + Missing unit for the data   + Standardise legend with the card   + Total Waste Generated is fine as it summarise the waste disposed and waste recycled | Waste Management by Waste Type (Left)   * Multi-row Card   + Remove prefix “Total”   + Round corner for the visual * 100% Stacked Bar Chart   + Add unit for the data in the title   + Change legend to better suit the visualisation |
| Waste Management by Year (Right)   * Card   + Redundant prefix ‘Total’ for Total Waste Recycled   + Total Waste Disposed is fine as it consisted Domestic Waste and Non Domestic Waste   + Round corner will differentiate card with other visualisation (graph) * Line and Stacked Column Chart   + Misleading y-axis as both are showing the same value   + Misleading y-axis title for both column and line | Waste Management by Year (Right)   * Card   + Remove prefix ‘Total’ for Total Waste Recycled   + Round corner for the visual * Line and Stacked Column Chart   + Remove secondary y-axis   + Change y-axis title to fit both column and line   + Change the legend to better suit the column and line as the y-axis title changed   + Add unit to the title of the chart to standardise with the 100% Stacked Bar Chart |

|  |  |
| --- | --- |
| 2nd Iteration | |
| Before | After |
|  |  |
| Feedback | Changes |
| Title (top left)   * White spaces can be filled | Title (top left)   * Expand title to fill up the white space |
| Waste Management by Waste Type (Left)   * 100% Stacked Bar Chart   + Missing timeline for title | Waste Management by Waste Type (Left)   * 100% Stacked Bar Chart   + Add timeline (year 2022) for title |
| Waste Management by Year (Right)   * Cards   + ‘Waste Recycled’ will cause misinterpretation that it is the subset of Total Waste Recycled * Line and Stacked Column Chart   + Missing timeline for title   + Standardise legend with the cards | Waste Management by Year (Right)   * Cards   + Add ‘Total” for Waste Recycled * Line and Stacked Column Chart   + Add timeline (year 2013 – 2022) for title   + Change legend to better suit the columns |

**Dashboard 2: Recycling Insight**

|  |  |
| --- | --- |
| 1st Iteration | |
| Before | After |
|  |  |
| Feedback | Changes |
| Title(top left)   * Background colour too dark, blending with the dashboard background with no contrast | Title (top left)   * Change background colour to make the title stands out |
| Number of Reverse Vending Machine   * Pie Chart   + Unpleasant pie chart colours | Number of Reverse Vending Machine   * Pie Chart   + Change pie chart colours |
| Recycling Rate by Year and Waste Type   * Card   + Round corner will differentiate card with other visualisation (graph) * Line Chart   + Great analysis on recycling rate   + Title of y-axis is fine to have no unit (data consisted of unit) | Recycling Rate by Year and Waste Type   * Card   + Round corner for the visual |

|  |  |
| --- | --- |
| 2nd Iteration | |
| Before | After |
|  |  |
| Feedback | Changes |
| Title (top left)   * White spaces can be filled | Title (top left)   * Expand title to fill up the white space |
| Number of Reverse Vending Machine   * Pie Chart   + Requires further analysis on number of reverse vending machine in Singapore   + Map visualisation would be a great choice | Number of Reverse Vending Machine   * Pie Chart   + Move Pie Chart to the top right   + Further analysis using Map visualisation (bottom right) * Map   + Add map visualisation |
| Recycling Rate by Year and Waste Type   * Cards   + Misinterpretation on the minimum and maximum recycling rate when selecting more than one waste type | Recycling Rate by Year and Waste Type   * Cards   + Remove cards * Slicer   + Move slicer to the left * Line Chart   + Move Line Chart to the middle |

**Dashboard 3: Awareness on Recyclable and Non-Recyclable Items in 2023**

|  |  |
| --- | --- |
| 1st Iteration | |
| Before | After |
|  |  |
| Feedback | Changes |
| Title(top left)   * Background colour too dark, blending with the dashboard background with no contrast | Title (top left)   * Change background colour to make the title stands out |
| Survey on Recyclable Items in Year 2023   * Card   + Round corner will differentiate card with other visualisation (graph)   + Top of the card looks cramped * Stacked Column Chart   + Top of the title looks cramped | Survey on Recyclable Items in Year 2023   * Cards   + Round corner for the visual   + Add 5px padding to the top of the cards   Stacked Column Chart   * + Add 5px padding to the top of the chart |
| Survey on Non-Recyclable Items in Year 2023   * Card   + Round corner will differentiate card with other visualisation (graph)   + Top of the card looks cramped * Stacked Column Chart   + Top of the title looks cramped | Survey on Non-Recyclable Items in Year 2023   * Cards   + Round corner for the visual   + Add 5px padding to the top of the cards   Stacked Column Chart   * + Add 5px padding to the top of the chart |

|  |  |
| --- | --- |
| 2nd Iteration | |
| Before | After |
|  |  |
| Feedback | Changes |
| Overall   * No context on the survey * Misinterpretation for common knowledge where items like old clothing and soft toys could be recycled but it is under the category of non-recyclable item   Card   * Underlined category labels look like hyperlinks | Overall   * Add context such as the survey question and the image of the blue recycling bin/chute   Card   * Italic category labels |

**3.2 Final Dashboard and Analysis**

1. **Waste Management**

A screenshot of a data analysis

Description automatically generated

The dashboard presents a comprehensive overview of waste management insights. The left-side visual, a 100% stacked column chart, offers insights into Singapore's 2022 waste management, displaying the breakdown of recycled versus disposed waste by type. It demonstrates waste processing efficiency, with each category's combined recycling and disposal equalling the total waste generated. This provides a snapshot of Singapore's recycling effectiveness and identifies potential areas for improvement.

With a 52% average recycling rate, the chart evaluates the success of the Green Plan's Sustainable Living pillar. High recycling rates indicate effective waste management policies, while lower rates suggest areas needing further development. The multi-row card at the top concisely summarizes the key figures, offering an at-a-glance view of the waste volumes managed.

The right-side chart on the dashboard delineates the yearly trend from 2013 to 2022, detailing the volume of domestic and non-domestic waste disposed of, alongside recycled waste. Domestic waste primarily includes household items, whereas non-domestic waste comprises industrial byproducts requiring specialized disposal.

The data reveals that recycled waste volumes consistently surpass disposed waste each year, underscoring Singapore's commitment to recycling and the goals of the Sustainable Living pillar in the Green Plan. Despite a dip in recycling rates in 2020, likely due to the COVID-19 pandemic, which impacted waste management systems worldwide, there's a visible recovery, indicative of Singapore's resilience and ongoing progress towards its eco-friendly aspirations.

1. **Recycling Insight**

A screen shot of a graph

Description automatically generated

The dashboard titled "Recycling Insight" is designed to critically analyse and communicate the progress of Singapore's recycling initiatives over time. The line chart reflects recycling rates by waste type across multiple years, showcasing the effectiveness of policies and public participation in recycling programs. An interactive slicer gives the viewer control over which waste categories to display, allowing for a comparative analysis of trends. The inclusion of a zoom slider empowers users to focus on specific time frames, enhancing the granularity of the examination of recycling patterns.

The pie chart and map on the dashboard serve as visual tools to assess the Singapore government's endeavours to promote recycling among its citizens. The pie chart categorizes the number of machines in each region, while the map offers a geographical perspective on their distribution, showing not only the quantity but also the extent of the area each cluster serves. By illustrating the allocation of reverse vending machines across different regions, these visuals highlight the government's strategic placement of recycling facilities to ensure they are accessible to the public. This approach aligns with the broader vision of the Green Plan, which aims to transform Singapore into a model of sustainable living.

1. **Awareness on Recyclable and Non-Recyclable Items in 2023**

A screenshot of a graph

Description automatically generated

The dashboard effectively visualizes the results of the NEA’s 2023 household recycling survey, illuminating public knowledge about recyclable and non-recyclable items. The side-by-side stacked column charts provide a detailed breakdown of survey responses, indicating whether items were correctly identified as suitable for blue recycling bins, incorrectly classified, or if there was uncertainty. Accompanying the charts, the survey question and an image of a blue recycling bin offer additional context, enhancing the dashboard’s communicative impact.

On the left, strategically placed cards employ a Top N analysis to spotlight the most and least recognized items from the survey. These key insights reveal which recyclable item is most familiar to the public and which items are most frequently misclassified, underscoring the need for focused public education in those areas. The distinct colour coding across the charts corresponds to the legend, allowing for quick interpretation of the data.

This dashboard showcases the effectiveness of Singapore's efforts towards sustainable living, aligning with the Green Plan's Sustainable Living pillar. The insights gained here can help to assess the government's progress in fostering an eco-conscious nation and pinpoint where additional resources may be needed to ensure public understanding and participation in recycling initiatives align with national sustainability goals.

1. **Statistical Modelling**

A graph with a line

Description automatically generated

The statistical model presented in the dashboard uses linear regression to analyse the relationship between the total waste generated by type and the waste recycled. The linear regression line, defined by the equation [276.92+0.84x], provides a predictive model where 'x' represents the waste recycled and 'y' represents the total waste generated. The R-squared value of 0.64 indicates that approximately 64% of the variability in total waste generated can be explained by the amount of waste recycled.

The positive slope of the regression line suggests that as recycling increases, the total waste generated also tends to increase, which may reflect the overall waste management practices and consumption patterns.