36104 Data Visualisation and Narratives

Waste Management Analysis

Executive Summary for Randwick City Council

29 May 2022

Contents

Business Context

Randwick Council runs a waste management program for 22 zones across 13 suburbs and is committed to providing a positive impact on the environment through its Recycling Centre. Since the waste disposal costs are high in Australia but range depending on the level of management controls, the Council is interested in analysing current waste generation – garbage and recycling – across Randwick City. This report presents key findings and several opportunities that will benefit the Council and Randwick City community both in the short and long term.

Principal Findings

Waste generation analysis identified several opportunities to benefit the Council and Randwick City community. Based on the real external data sourced from .id community data exporter, it was confirmed that education could be the most prominent factor in preventing the generation of waste per unit in the area under the Randwick Council.

Specific results per street, zone, suburb, and collection day are communicated below. The links to interactive dashboards for each challenge are outlined in Appendix 1.

Average waste generation per week

During the analysis, zones 21, 10 and 16 and suburbs Maroubra and Randwick were found to generate the highest weekly average waste (both garbage and recycling), while zone 2 and La Perouse produced the least. It also became evident that some collection days generate less waste than others. Particularly, schedules

Every Mon, Wed, Fri	Every Mon, Thu	Every Tue, Fri	Every weekday

yield less than 1kg weekly, on average, and are less efficient.

Average garbage and recycling generation – by street, suburb and zone

Being a 13km street with several facilities such as a stadium, racecourse, university and hospitals, Anzac Parade produces a vast amount of garbage. It is both the worst street for generating garbage (average of almost 4,000 kg) and the best street for recycling (average of more than 4,500 kg).

When considering the density of properties per street, Bega Avenue generates the highest amount of average garbage per unit (6.71kg/unit), with Baragoola Avenue yielding the highest average recycled waste (10.42 kg/unit). Key suburbs and zones for garbage and recycling generation are delineated in Figure 1 and Figure 2 below.

	AVERAGE GARBAGE GENERATION	AVERAGE RECYCLING GENERATION
BEST SUBURB	Kensington 1.04 kg/unit	Phillip Bay 6.25 kg/unit
WORST SUBURB	La Perouse 3.96 kg/unit	Little Bay 0.96 kg/unit

Figure 1: Average garbage and recycling generation by suburb

	AVERAGE GARBAGE GENERATION	**AVERAGE RECYCLING GENERATION**
BEST ZONE	Zone 22 0.91 kg/unit	Zone 16 3.12 kg/unit
WORST ZONE	Zone 1 3.33 kg/unit	Zone 22 1.41 kg/unit

Figure 2: Average garbage and recycling generation by zone

Recommendations

Based on the analysis outcomes, the following actions by Randwick Council would lead towards a greener Randwick City:

1. Revise less efficient schedules where average weekly waste generation is below 1kg.
2. Since education is proven to be the main factor preventing waste generation, residents of Phillip Bay and zone 16 should be considered a priority for recycling training.
3. Residents of Anzac Parade could be offered garbage generation awareness materials if the budget is available.

Challenges and Opportunities

Lastly, the main challenge faced during the analysis was the quality of acquired data and the absence of a data dictionary which forced the data team to interpret some attributes independently. Specifically, zones differed in the property list and garbage record datasets, indicating inconsistent data collection. Furthermore, there are more than 30% of invalid values for weight and street attributes caused by typing errors.

Therefore, the main area of improvement for the business would be improving the accuracy and reliability of the raw data by critically evaluating the business logic and data collection process. The team took an opportunity to initiate a data cleaning process documented in Appendix 4 of this report and included as a Jupyter notebook to this submission.

References

Randwick City Council: community profile, viewed Mar 28, 2022, <https://profile.id.com.au/randwick/data-export>.

Appendix 1: Ethical Considerations

Vital ethical considerations of this analysis are depicted below:

- This task is assigned under the confidentiality deed between the Randwick Council, Waste Management Unit and Master of Data Science and Innovation Course (MDSI), UTS. No individual outside of these parties should be given access to the provided data.
- Upon completion of this visualisation task, it is imperative that the MDSI, UTS party will return, destroy or delete the data provided by the Randwick Council.
- The medium of sharing the data should be secure and only accessible by both parties.
- The correlation study for challenge 4 includes factors like English level and income level and, therefore, may involve discrimination. Discrimination is treated very seriously in Australia, which could urge residents from non-English speaking countries to file an official complaint against the Randwick City Council. Thus, it is crucial that the Council discloses this was not the intent.

Appendix 2: Teamwork Reflection

Overall, the group found that the working process on the four data challenges was smooth and efficient. In particular, we succeeded in:

- **Objectives setting.** Team members' skill sets varied significantly, and we managed to utilise each one's strengths when splitting the workload. While we worked in parallel on the tasks requiring more technical expertise, we also ensured we reached more high-level tasks. This approach boosted team discussion around key ideas before progressing to a more async working style and helped us manage better the project.
- **Facilitation and feedback.** The team was open to receiving and providing structured feedback, which facilitated the exchange of acquired beforehand knowledge. It encouraged creativity and innovative solutions.

We have also identified that the area of further improvement would be time management. Next time, from the beginning, we will be clear on the expected duration of the communication and stick to the meeting agenda and schedule.

Appendix 3: Interactive dashboards

Please ensure you have downloaded the zip file provided as part of the submission before attempting to open the links below.

Challenge 1: [Average waste generation per week](#)

Challenge 2: [Average garbage and recycling generation by street](#)

Challenge 3: [Average garbage and recycling generation by suburb](#)

[Average garbage and recycling generation by zone](#)

Challenge 4: [Main factors contributing/preventing waste generation](#)

Appendix 4: Data Cleaning and Description

This appendix provides a descriptive summary of data cleaning steps. Further details and the coding process is outlined in the jupyter notebook submitted along with this executive summary.

Cleaned data file name: all_data_2405.csv

Note: "XX_1" and "XX_3" represent the answer to question 1 and question 2, respectively.

Description of garbage record data

There are nine csv files including historical records about the waste collection and covering periods Feb 2 - Feb 15, May 1 - May 15, Nov 15 - Nov 30 from 2013 to 2015:

File name	Number of Records	Number of Attributes
15 to 30 Nov 2015.csv	37769	14
15 to 30 Nov 2014.csv	73348	14
15 to 30 Nov 2013.csv	44276	14
1 to 15 May 2013.csv	8438	14
1 May to 15 May 2015.csv	78650	14
1 May to 15 May 2014.csv	47563	14
1 to 15 Feb 2015.csv	31374	14
1 to 15 Feb 2014.csv	36173	14
1 to 15 Feb 2013.csv	89138	14

Initial abnormal findings:

- 1 to 15 May 2013.csv contains a larger number of missing data (only 8438 rows)
- The number of rows varies a lot across provided nine files despite sharing a similar timeframe of 15 days

Data dictionary assumptions

After checking the structure of the CSV file, all datasets were combined to ease with data cleaning and visualisation. As the data dictionary was not provided by the Council, the following assumptions were made based on common knowledge factoring in general relevant data guidance provided by the Randwick City Council:

- date & time: date and time each garbage bin was collected
- rfid: RFID tags used for tracking each garbage bin
- serial: ignored based on the general relevant descriptions
- bin type: types of garbage bin
- schedule: schedule of garbage collection

- weight: weight of the bin
- truck: truck number
- zone: zone
- reference: ignored based on the general relevant descriptions
- billing: ignored as there are no values
- unit: unit number of garbage bins allocated (if apartment)
- house: street number of garbage bins allocated (if house)
- street: street of garbage bin allocated
- suburb: suburb of garbage bin allocated

Data cleaning – record data

Basic changes
- Drop duplicates (~ 0.37% of data)
- Check if data is valid in the file:
  - all records in the CSVs are in the correct range of file time
  - weight not null and <= 75 are considered as valid values to keep (36.38% of data is dropped up to this step)
- Remove leading and trailing spaces, change to lower case
- Convert "date & time" to DateTime format
- Add features "year", "month" from "date & time"
- Add "week":
  - assume day 1-7 is one week, day 8-15 is another week, day 15-21 is the following week, day 22-30 is the one after
  - there are 18 weeks in the dataset
  - Note: there is no data in week 4. This could be the underlying reason why 1 to 15 May 2013.csv has a significantly lower number of records compared to others.
- Standardise bin_type 'ga' to 'g', 'ra' to 'r'
- Drop bin size if not equal 240 and 140
- Standardised house numbers, for example, 'jan' to '1'…

Zones
The Randwick City Council covers 18 suburbs in the Eastern suburbs of Sydney (source: https://www.paulsrubbish.com.au/council-rubbish-collection/randwick-city-council/).

Since zones in record data and property list data do not align, they were changed based on this map: https://www.randwick.nsw.gov.au/__data/assets/pdf_file/0012/26103/Waste-clean-up-zones.pdf.

According to the completed research and provided zone map, suburbs and zones listed in the record data are assumed to be valid:

- Centennial Park (shared with City of Sydney Council): not in zone map
- Chifley: 3, 19
- Clovelly (shared with Waverley Council): 15, 16
- Coogee: 10, 11, 14, 16, 15
- Coogee Beach: not in zone map

- Kensington: 22, 13
- Kingsford: 13, 21, 12
- La Perouse: 1
- Little Bay: 1, 19
- Malabar: 20, 2
- Maroubra: 4, 6, 9, 8, 7, 5
- Maroubra Junction, not in zone map
- Matraville (minor part located in Bayside Council): 3, 20, 4, 19
- Phillip Bay: 1
- Port Botany: not in zone map
- Randwick: 18, 21, 7, (future subdivision), 11, 14, 16, 17
- South Coogee: 9, 8, 11, 10
- University of New South Wales: not in zone map

List of abnormal zones with value higher than 1 % (grouped by suburb):

- 'Coogee':
- 8: 493 (1.36%)
- 9: 1,301 (3.6%)
- 'Little Bay':
- 18: 627 (9.1%)

Zones that only appeared in both suburb and zone map are kept.

Streets
- Standardise street, for example 'st' to 'street'
- Search valid streets based on open street map API and get location information
- Manually edit streets not found by open street API, such as type error
- Number of records for each street should be higher than 9 as there are 18 weeks and the garbage bins are assumed to be collected at least fortnightly. Street with records lower than 9 are dropped.
- Streets are checked if included in the property list dataset. Only four addresses are not included. They were dropped when joining data.
- 37.0 % of data was dropped compared to raw data. The new number of rows: 279,980.

Data cleaning – property list data
- The property list is also used to get the total number of units per suburb, street and zone.
- The total number of units for each property should be at least 1. Change the property with total_unit is zero to one.
- "zone" in property list data and "zone" in record data are not aligned and are updated based on the map (see Zones section).
- Total number of units per suburb and total number of units per zone are calculated based on property list data.

Data cleaning – external data

The real external data was sourced and transformed for challenge 4.

- Data includes such factors as income, English proficiency, education, and voluntary work for each suburb.
- Data is used as categorical variables in label encoding and calculated as the total score of each suburb based on the percentage of people located in that group. Level is labelled and transformed to min-max scaled score.

Other notes

- Average waste is calculated as the sum of waste per week divided by 18 (18 is a total number of weeks).
- Avg function is not used as there is waste collected fortnightly which may lead to errors.
- Errors may exist in the average waste as there is a significant amount of missing data and some suburbs have no waste recorded for some weeks.
- Average waste per unit is calculated as the total average waste in total divided by the number of units.

Checklist for plots

Challenge 1:

- 3 bar charts based on new data. Check the below column used for Q1.

Challenge 2:

- Use the 'latitude', 'longitude' in all_data_2405.csv
- Make sure no points at the same place
- When extracting data, keep only one langitute and longitude for each street (there are multiple locations for one street).

Challenge 2 and Challenge 3:

- Map by ranking showing value at tooltips
- Map size and color should be consistent: green for recycling, red for garbage; can be dark red to yellow for garbage, and dark green to yellow for recycling
- Tooltip only showing associated info
- Include maps for per unit
- Consistent HTML name for maps, for example:
  - 'garbage_zone_3': challenge 3 average garbage zone
  - 'recycle_suburb_unit_2': challenge 2 average recycle per unit
- 4 maps in 1 HTML file:
  - Html should be saved with all lengths and heights as the challenge 2 map draft was
  - Similarly to challenge 3, put the position to show all Randwick
  - For each combined HTML maps, all maps included must locate in the same folder
  - Follow 'combine_draft' as a template, and the notes inside

Notes for each challenge

Challenge 1 (3 plots)

Columns used:

- 'schedule'
- 'zone'
- 'suburb'
- 'weight_waste_schedule_1'
- 'weight_waste_zone_1'
- 'weight_waste_suburb_1'

Challenge 2 (4 plots)

Columns used:

- 'street'
- 'latitude', 'longitude'
- 'weight_garbage_street_2'
- 'rank_garbage_street_2'
- 'weight_recycle_street_2',
- 'rank_recycle_street_2'
- 'weight_garbage_street_unit_2'
- 'rank_garbage_street_unit_2'
- 'weight_recycle_street_unit_2'
- 'rank_recycle_street_unit_2'

Challenge 2 (4 plots)

Columns used:

- 'zone'
- 'suburb'
- 'weight_garbage_zone_3'
- 'rank_garbage_zone_3'
- 'weight_garbage_suburb_3'
- 'rank_garbage_suburb_3'
- 'weight_recycle_zone_3'
- 'rank_recycle_zone_3'
- 'weight_recycle_suburb_3'
- 'rank_recycle_suburb_3'
- 'weight_garbage_zone_unit_3'
- 'rank_garbage_zone_unit_3'
- 'weight_recycle_zone_unit_3'
- 'rank_recycle_zone_unit_3'
- 'weight_garbage_suburb_unit_3'
- 'rank_garbage_suburb_unit_3'
- 'weight_recycle_suburb_unit_3'
- 'rank_recycle_suburb_unit_3'