Waste management in Melbourne CBD

1. Research question: Can Degraves Street Recycling Facility meet the demand of waste disposal?

Melbourne CBD is a high traffic area with 40000 residents, numerous shops and cafeterias which produce a large amount of waste. To confront this, Degraves Street Recycling Facility was established in 2013 in the hope to simultaneously decrease waste sent to landfill while increasing the rate of recycling. We will analyze if the amount of waste recycled is high enough and suggest the best pathway to help Melbourne CBD improve its waste management and sustainability.

2. Data processing and analysis
a. Overview of dataset

For this project, 5 datasets will be used to evaluate the performance of the recycling center. The first 4 datasets contain essential information about Degraves Street Recycling Facility – records of the amounts of 4 collected waste categories (organic, cardboard, commingled and polystyrene waste) throughout 2016 while the last one provides waste collection services of Melbourne from 2011 to 2016.

As 4 datasets at Degraves recycling centre are collected in different time periods, their sizes vary greatly, in which the first 2 datasets are collected daily (366 rows) while the last 2 are gathered weekly. Despite such a difference in size and variables, for the purpose of this project, only 2 columns are required from each dataset for further analysis, which are the date and the net collection columns.

	A	B	C	D	E	F	G	H	I
1	Day	Date	Number o	Time bin 1	Time bin 2	Time bin 3	Time bin 4	Time bin 5	Daily Estimated Net Weight (kg)
2	Friday	1/1/2016 0:00	3	9:10	10:30	16:00			180.939
3	Saturday	1/2/2016 0:00	1	8:20					60.313
4	Sunday	1/3/2016 0:00	2	10:00	15:30				120.626
5	Monday	1/4/2016 0:00	4	6:20	12:10	16:30	17:00		241.252
6	Tuesday	1/5/2016 0:00	3	6:10	14:15	18:00			180.939
7	Wednesd;	1/6/2016 0:00	4	6:00	12:00	16:10	18:00		241.252
8	Thursday	1/7/2016 0:00	3	6:00	13:15	17:30			180.939

Commingled.csv (366x9)

Similarly, the *'master'* dataset also provides data of waste collection services (for the entire City of Melbourne), detailing the weight (in tonnes) of waste collected in 7 different categories including the four previously mentioned categories.

Although the number of features and rows in each dataset are different, they all share a common column – *date* column, on which the corresponding net weight column of each dataset is joint. which will be further illustrated in data wrangling.

b. *Data wrangling*

In this project, the majority of work involved with data cleansing is dealing with missing values. In particular, there are up to 23 NAN values (since September no observation is recorded) in polystyrene.csv, 7 missing values in organic.csv, 1 missing value in commingle.csv. Therefore, missing values are imputed with the mean value, rather than backward fill or forward fill to ensure the stability of the dataset

In order to merge datasets using the *date* column, **data binning** is necessary to scale all time stamps to the same period. Hence, for further data analysis and time series plots, frequency resampling to weekly/monthly is applied to showcase waste trends and recycling center's strengths and weaknesses . In addition, the data of Melbourne's collected waste is also carried out on a monthly basis; therefore, timestamps will need to be scaled monthly for the first 4 datasets again to evaluate the performance of this center versus the entire city of Melbourne.

Data preparation

– Degraves Street Recycling Center:

a. Aggregating and merging daily datasets to create commingle_organic dataset

2 datasets that are merged are commingle.csv and organic.csv. After being merged, 2 columns months and week will be appended at the end to perform data aggregation (grouping by month) to merge with cardboard.csv and polystyrene dataset to create the center's recycling rate with regard to City of Melbourne

This dataset is used to generate Figure 1

	Date	Amounts of commingle waste	Amounts of organic waste	month	week
0	2016-01-01	180.939	34.0	1	53
1	2016-01-02	60.313	35.7	1	53
2	2016-01-03	120.626	23.6	1	53
3	2016-01-04	241.252	169.5	1	1
4	2016-01-05	180.939	102.0	1	1
...	...	...	...	...	...

b. Aggregating and Merging weekly dataset to create board_poly dataset

2 datasets that are merged are cardboard.csv and polystyrene.csv. The same process (merging data aggregation – groupby().sum()) is applied to merge with the dataset in part a to create part c. This dataset is used to create Figure 1

month	Estimated Polystyrene Total (kg)	Estimated weight of cardboard (kg)
1	53.226667	7250
2	83.850000	8250
3	58.863333	7000
4	25.800000	7250
5	91.113333	8000

c. Combining the city of Melbourne dataset in 2016 with datasets of Degraves Center to show the center's recycling rate with regard to City of Melbourne

Remarks:
– Scaling between units (kg in center's dataset vs tons in master dataset)
– Green waste accounts for 7 /(7+7+35) = 14.286% organic waste (WASTE AND RECYCLING IN MELBOURNE)
– By using month, data of the City of Melbourne in 2016 is extracted and merged with the monthly processed waste of the recycling center (new columns are commingled_recycling, green_waste, cardboard). Then 3 other columns (%commingle, %cardboard, %organic) are appended showing the contribution of the recycling center of these 3 categories to the total amount of waste processed in Melbourne.
– Polystyrene is not merged because of a large number of missing values and the master dataset does not record any observations about polystyrene waste
– This is used to generate 3 bar charts: Figure 5, 6, 7

c. Machine learning
Building a machine learning model for predicting the recycling rate in a given year

month	Amounts of commingle waste	Amounts of organic waste	Total_cardboard_2016	commingled_recycling	green_waste	cardboard	%commingle	%cardboard	%organic
1	5669.422	3157.6	7250	715.00	3.60	176.38	0.792926	4.110443	12.530159
2	5005.979	3180.3	8250	708.31	5.56	196.96	0.706750	4.188668	8.171377
3	5548.796	3471.8	7000	749.20	1.68	201.34	0.740629	3.476706	29.522109
4	5729.735	2994.5	7250	672.57	5.28	189.00	0.851917	3.835979	8.102002
5	5367.857	2787.3	8000	702.07	3.08	199.68	0.764576	4.006410	12.928108
6	5005.979	2773.9	6250	668.04	3.32	201.04	0.749353	3.108834	11.935886

As an illustration, we are going to predict the center's recycling rate for **commingled waste** in 2017. In order to achieve this, there are 2 tasks to be completed:

- Task 1: predict the total amount of commingled waste in the City of Melbourne in 2017 from a given time
- Task 2: predict the recycling rate based on the previously calculated total amount of commingled waste

Task 1: Input (beginning of the month – 1/10/2017) –– Output: Total amount of commingled waste in Melbourne

Preprocessing: Create a data pipeline (feature selection, standard scaling, imputation) to perform data transformation on datasets.

Training: A sketch of data points reveals that the data is not exactly linear; and to test the validity of this assumption, I have trained 3 different algorithms: LinearRegression, Support Vector Regression, and RandomForestRegressor (an ensemble learning algorithm) with default parameters.
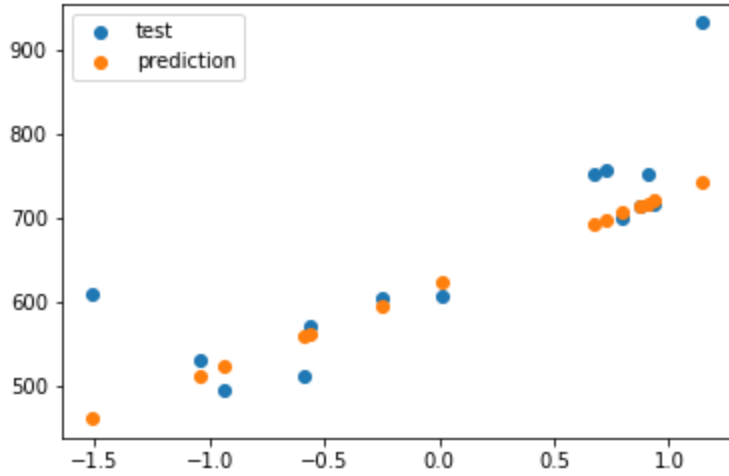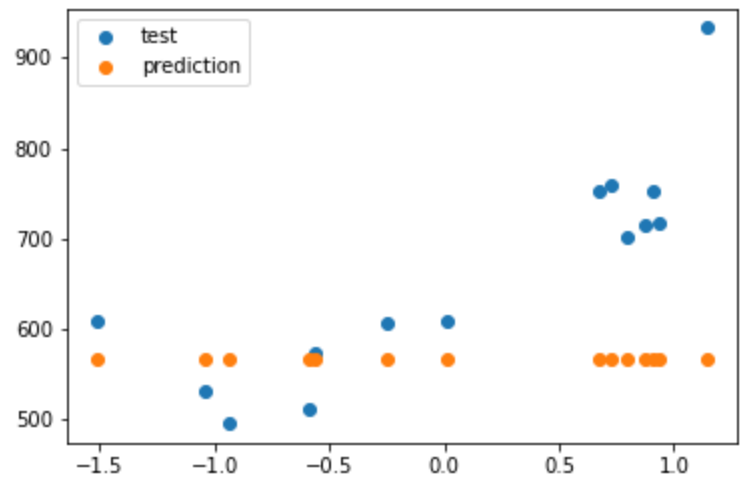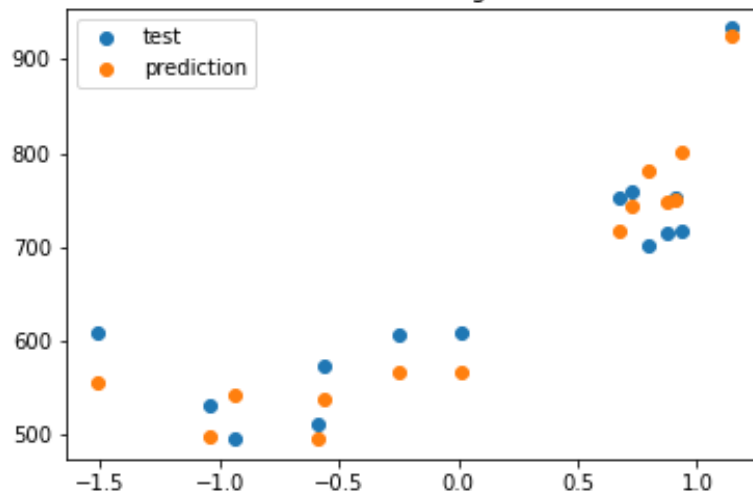
Evaluation: By using negative mean squared error as an evaluation metric, RandomForestRegressor is the most suitable algorithm for this model

Tuning hyperparameters: By setting a range of values on 2 hyperparameters and put the model under cross-validation, it was found that the optimal values for this model is:

max_depth = 6 and n_estimators = 15

Randomforest Regression on test set (adjusted parameters)
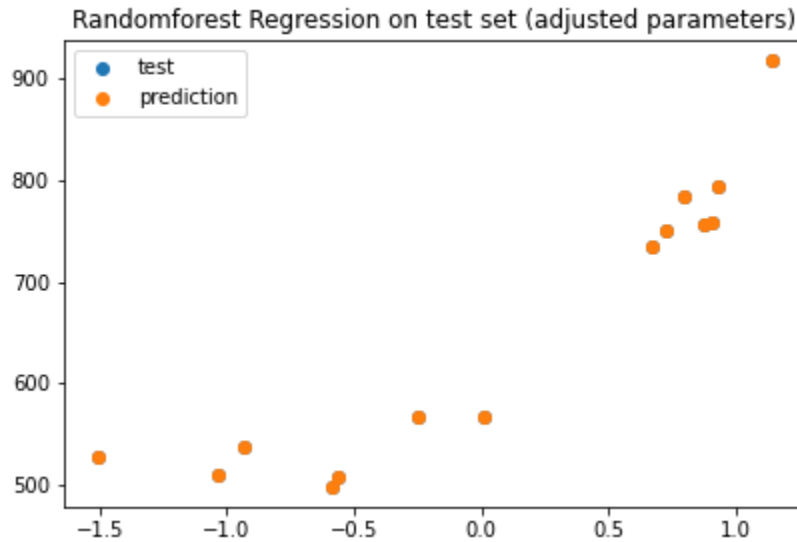
Limitations: Because there is only a small amount of data (roughly 132 instances), there might be a chance of overfitting. Besides, it does not scale very well on the far future due to the small dataset.

2. Task 2: predict the recycling rate based on the previously calculated total amount of commingled waste

–   Assumptions: There are only 12 instances for this recycling center in 2016; therefore, to avoid overfitting, we are going to use a simple model: Linear Regression, without splitting into train and test set

–   Limitations: The amount of data is too small for training and testing, which makes it difficult to extrapolate in the far future (2050). Therefore, we are going to predict the recycling rate in October 2017, (reasonably close enough to 2016). If more data is collected for this center monthly in a larger interval of time, it can predict more precisely in a further future.

–   Result: The predicted recycling rate is 0.7045%, which suggests that the center's contribution to recycling commingled waste is likely to decrease in the future.
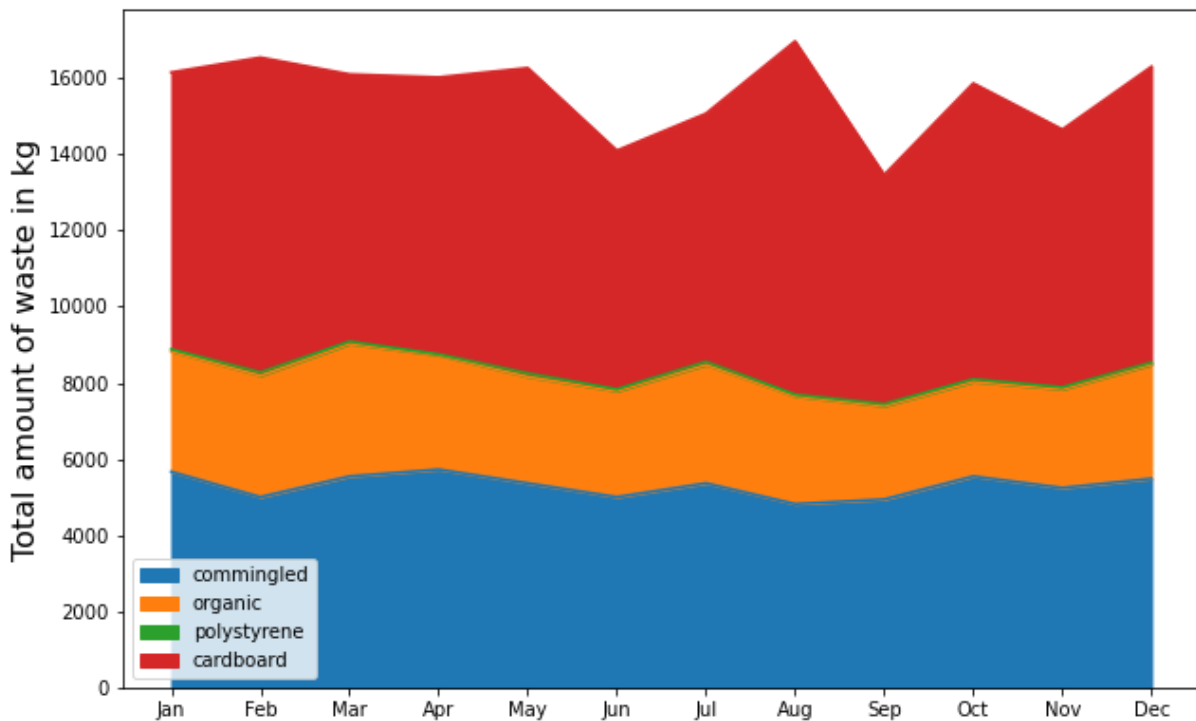
## 3. Analysis/Key result

Waste disposal throughout the year – Figure 1
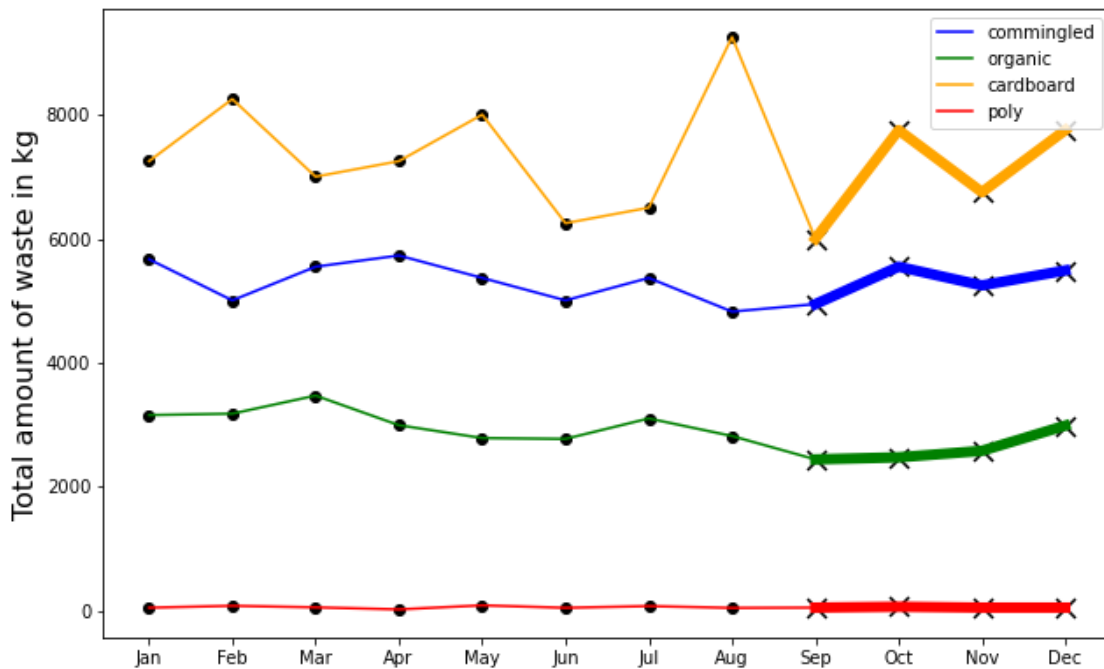
- From the graph, there is an increase in the amount of waste in all 4 categories, which is highly supported by green waste and commingled waste (Figure 2-5). The significant growth in the amount of waste circulating in the city from September onward is expected annually since it is the time of the year where Melbourne reaches its seasonal tourism. As the city gets more populated it will be more polluted. All waste materials show a significant uptrend in mass from September to the remainder of the year. The general trend of all waste materials was similar with respect to month of the year except for the steep incline in mass of cardboard waste recycled in the month of August.
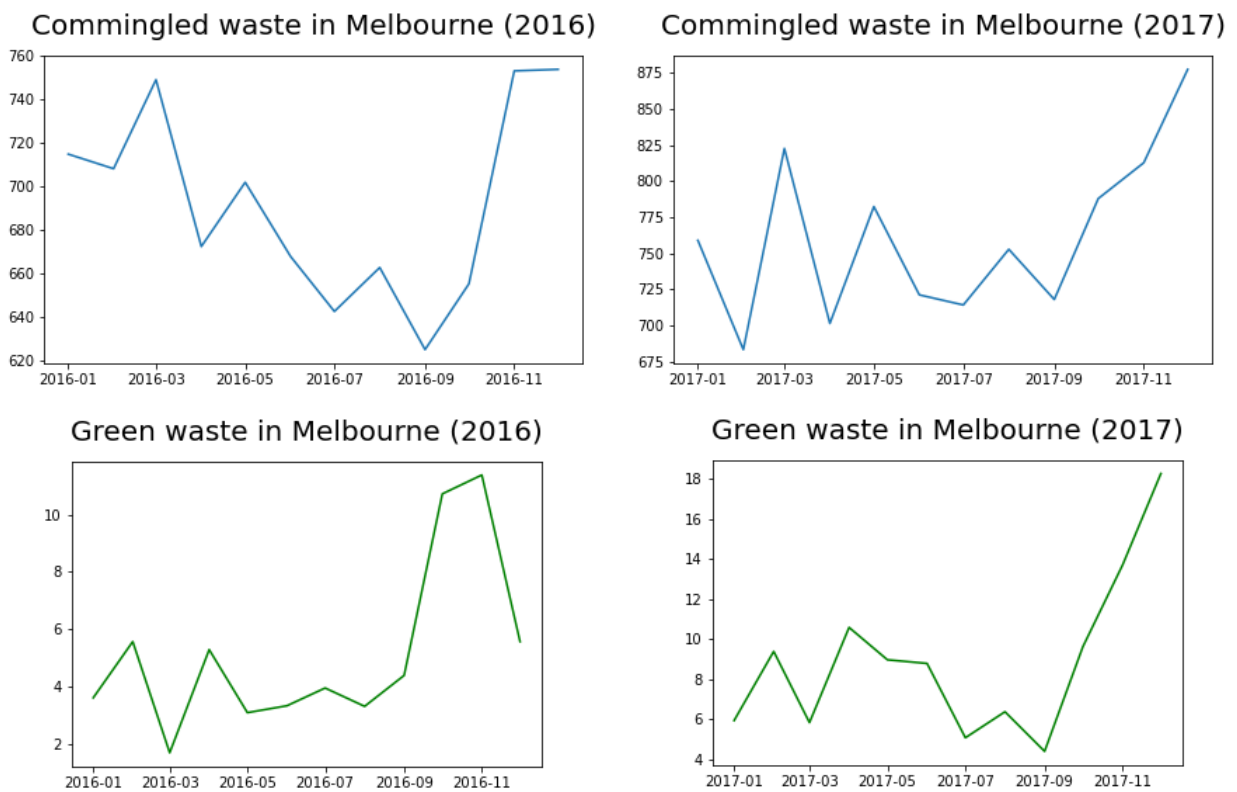

*Holidays, festival, summer time, tourist season and waste trend*

Volatility of waste (2016)
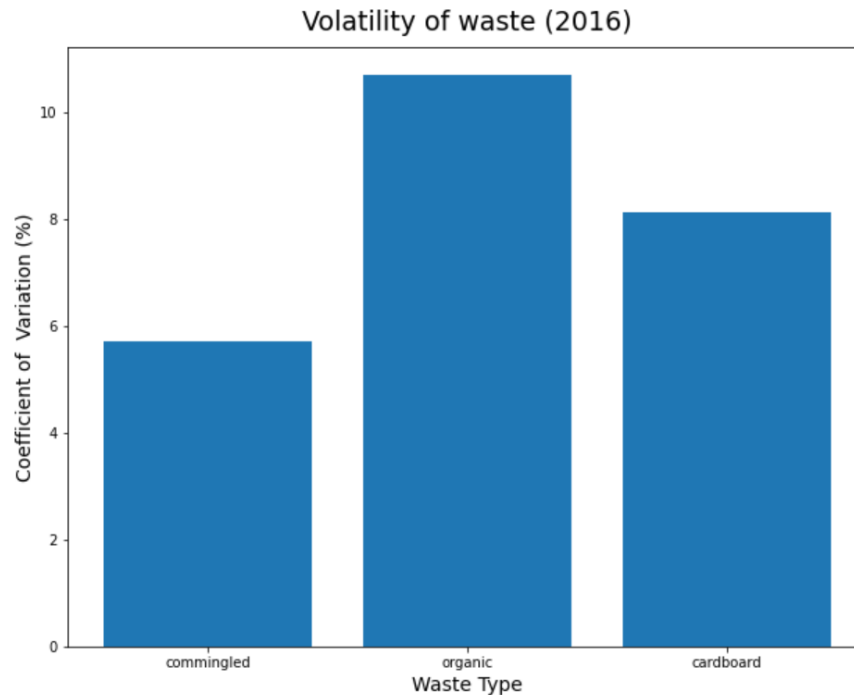
– Despite the amount of cardboard and commingled waste can fluctuate to 25% in 2016, they can be considered "normal day waste" as the amount of organic and polystyrene waste can vary to 37% and 57% throughout the year. We assume this is due to special times like holidays and festivals where larger numbers of people travel to the CBD and produce larger amounts of food and food container waste.

Recycling rate– To truly understand the impact that Degraves Street Recycling Center has on Melbourne, we first have to first walk through the "position" of the city in relation to recycling measures. The given situation as well as main reason why DSRC was launched in the first place was mostly due to the poorly managed waste bin and also no recycling methods were recorded amongst businesses in the precinct operate area of DSRC.
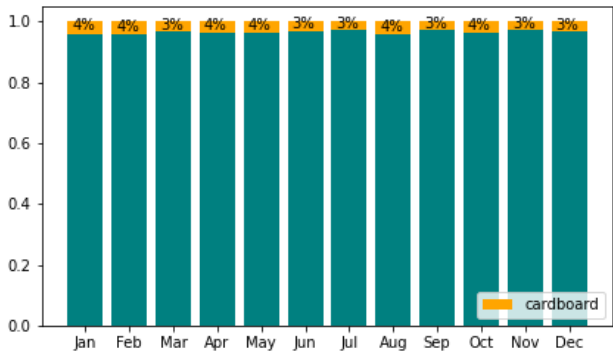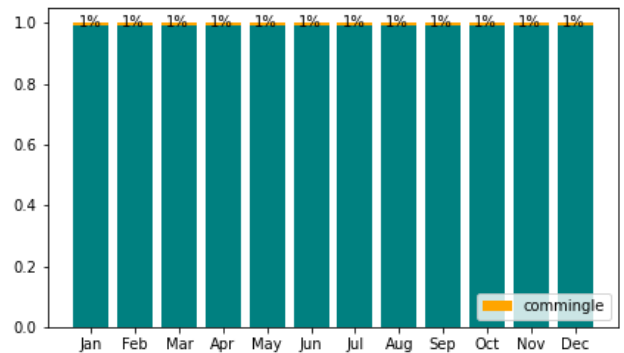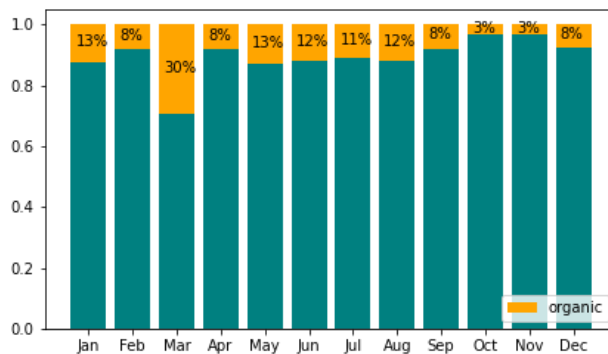
(Figure 5, Figure 6, Figure 7)

The produced bar charts that may seem somewhat blizzard at first actually has a lot to do with the location of DSRC. Since the facilities are located in the CBD, it is no surprise that there are large amounts of food– related waste compared to other parts of the city. As a result DSRC thrives in the amount of waste it is able to collect in this specific waste category. However, the reason why commingled waste is so low is because commingled waste comprises a mixture of waste including metal. Therefore, recycling centers in industrial zones can potentially create a large amount of commingled waste, which reduces the contribution of DSRC in this waste category.

In summary cardboard waste was responsible for 46.99% of the total waste in Degraves recycling centre, followed by commingled waste at 34.04%, organic waste at 18.57% and polystyrene at 0.39% in the year of 2016. The data gives us an insight on the relatively high percentage of cardboard waste being recycled at the Degraves recycling centre and low percentage of polystyrene waste being recycled.

Proporion of recycling of each waste category in 2016

commingled waste

34.04% (63751)

Organic waste

18.57% (34777)

0.39% (727)

Polystyrene

46.99% (88000)

Cardboard
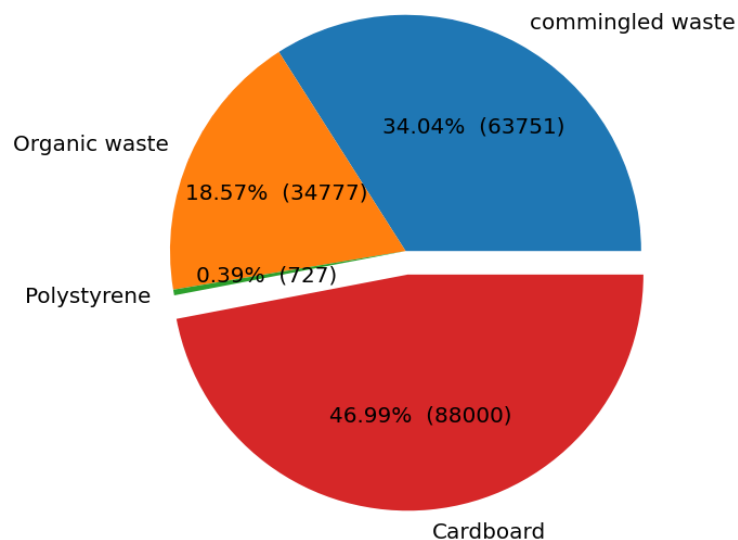
4. Significance of results

Data visualization helps curate our raw data, which is raw statistics and dates, into an easier to interpret form and from there (waste trends and the center's distribution of waste are easily obtained). Therefore, if waste trends are recognized (an influx of trash from September to December), the center should be warned and more infrastructure and machines need to be installed to adapt to this substantial amount of trash.

Besides, the amount of waste processed by DSRC is enormous, taking up to approximately a third of the city of Melbourne's organic waste in March, which is because there are many restaurants, cafes and bars in the Central Business District releasing a huge amount of green waste. Therefore, more public trash bins should be installed publicly (on the streets and in apartments) to facilitate the center's organic recycling process.

By contrast, the amount of commingled waste is so low because commingled waste consists of metals and glass, which is highly detrimental to human's health when processed. Therefore, an assumption of the new plan in the future is that all of these harmful wastes are transferred to other facilities in the suburb of Melbourne to minimize

the environmental and health risks, which confirms the result of the machine learning model that there is a drop in the amount of commingled waste recycled by DSRC next year.

5. Limitations

Challenges:

– A large number of missing values especially in polystyrene.csv, which might lead to error in analysis

– An assumption is that green waste accounts for 14.286% organic waste; however this number is approximation and might be different in each area ([WASTE AND RECYCLING IN MELBOURNE](#))

– The amount of data for DSRC is collected only in 2016, which is very small, making it difficult for further analysis in other years and prediction in the far future (machine learning)

– If more data is collected for this center in other years (2015,2017,...), waste trends can be better analyzed and the machine learning model can generalize better and predict in a more distant future with higher accuracy.

Risks:

– Upsetting the local cafe for too much rule about bargain collecting

– New plan might reduce the efficiency of recycling with regard to other waste categories