Data Analysis Of Singapore's Waste Management

Yadanar Aung (P2214621)
DAAA/FT/1B/02
ST1510 PDAS

Table of Contents

Nature of Dataset 1

2003_2017_waste.csv

https://www.kaggle.com/datasets/mahaanand/singapore-waster-mgmt?select=2003_2017_waste.csv

First 5 Rows:

	waste_type	waste_disposed_of_tonne	total_waste_recycled_tonne	total_waste_generated_tonne	recycling_rate	year
0	Food	679900	111100.0	791000	0.14	2016
1	Paper/Cardboard	576000	607100.0	1183100	0.51	2016
2	Plastics	762700	59500.0	822200	0.07	2016
3	C&D	9700	1585700.0	1595400	0.99	2016
4	Horticultural waste	111500	209000.0	320500	0.65	2016

Last 5 Rows:

	waste_type	waste_disposed_of_tonne	total_waste_recycled_tonne	total_waste_generated_tonne	recycling_rate	year
220	Ash and sludge	214800	28600.0	243400	0.12	2017
221	Plastic	763400	51800.0	815200	0.06	2017
222	Textile/Leather	141200	9600.0	150800	0.06	2017
223	Others (stones, ceramic, rubber, etc.)	319300	7100.0	326400	0.02	2017
224	Total	2980000	4724300.0	7704300	0.61	2017

Information of Dataset 1:

```
**** Singapore Waste Management from 2013 to 2017 ****

Information of dataset 1
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 225 entries, 0 to 224
Data columns (total 6 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   waste_type                  225 non-null    object
 1   waste_disposed_of_tonne     225 non-null    int64
 2   total_waste_recycled_tonne  225 non-null    float64
 3   total_waste_generated_tonne 225 non-null    int64
 4   recycling_rate              225 non-null    float64
 5   year                        225 non-null    int64
dtypes: float64(2), int64(3), object(1)
```

Nature of Dataset 2

2018_2020_waste.csv

https://www.kaggle.com/datasets/mahaanand/singapore-waster-mgmt?select=2018_2020_waste.csv

First 5 Rows:

	Waste Type	Total Generated ('000 tonnes)	Total Recycled ('000 tonnes)	Year
0	Construction& Demolition	1624	1618	2018
1	Ferrous Metal	1269	1260	2018
2	Paper/Cardboard	1054	586	2018
3	Plastics	949	41	2018
4	Food	763	126	2018

Last 5 Rows:

	Waste Type	Total Generated ('000 tonnes)	Total Recycled ('000 tonnes)	Year
40	Non-ferrous metal	75	73	2020
41	Glass	66	7	2020
42	Scrap tyres	23	22	2020
43	Others (stones, ceramics, etc.)	193	21	2020
44	Overall	5880	3040	2020

Information of Dataset 2:

```
**** Singapore Waste Management from 2018 to 2020 ****

Information of dataset 2
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 45 entries, 0 to 44
Data columns (total 4 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Waste Type                      45 non-null     object
 1   Total Generated ('000 tonnes)   45 non-null     int64
 2   Total Recycled ('000 tonnes)    45 non-null     int64
 3   Year                            45 non-null     int64
dtypes: int64(3), object(1)
```

Nature of Dataset 3

ghg-emissions.csv

https://www.climatewatchdata.org/ghg-emissions?breakBy=gas&end_year=2019&gases=all-ghg&regions=SGP&sectors=waste&start_year=1990

	Gas	unit	1990	1991	1992	1993	1994	1995	1996	1997	...	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
0	CH4	MtCO2e	1.15	1.18	1.22	1.26	1.30	1.33	1.37	1.42	...	2.45	2.54	2.62	2.71	2.79	2.88	2.96	3.03	3.11	3.19
1	CO2	MtCO2e	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	F-Gas	MtCO2e	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	N2O	MtCO2e	0.05	0.05	0.05	0.05	0.06	0.06	0.06	0.06	...	0.09	0.09	0.10	0.10	0.10	0.10	0.11	0.11	0.11	0.11

Information of Dataset 3:

```
**** Gases Produced from Singapore Waste from 1990 to 2019

Information of dataset 3
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4 entries, 0 to 3
Data columns (total 32 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Gas     4 non-null      object
 1   unit    4 non-null      object
 2   1990    2 non-null      float64
 3   1991    2 non-null      float64
 4   1992    2 non-null      float64
 5   1993    2 non-null      float64
 6   1994    2 non-null      float64
 7   1995    2 non-null      float64
 8   1996    2 non-null      float64
 9   1997    2 non-null      float64
 10  1998    2 non-null      float64
 11  1999    2 non-null      float64
 12  2000    2 non-null      float64
 13  2001    2 non-null      float64
 14  2002    2 non-null      float64
 15  2003    2 non-null      float64
 16  2004    2 non-null      float64
...
 30  2018    2 non-null      float64
 31  2019    2 non-null      float64
dtypes: float64(30), object(2)
```

Nature of Dataset 4

solid-waste-management-total-waste-incinerated.csv

https://data.gov.sg/dataset/solid-waste-management-total-waste-incinerated-annual?view_id=882ae208-bd25-4b99-8cc9-a3b7b24d063a&resource_id=e4c8461f-e7de-4fc3-ad25-cf068ae09509

year	total_waste_incinerated
2000	2.44
2001	2.55
2002	2.42
2003	2.31
2004	2.26
2005	2.28
2006	2.33
2007	2.38
2008	2.45
2009	2.48
2010	2.59
2011	2.66
2012	2.73
2013	2.82
2014	2.87
2015	2.83

Information of Dataset 4:

```
**** Total Waste Incinerated in Singapore from 2000 to 2015 ****

Information of dataset 4
<class 'pandas.core.frame.DataFrame'>
Int64Index: 16 entries, 2000 to 2015
Data columns (total 1 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   total_waste_incinerated  16 non-null     float64
dtypes: float64(1)
```

Nature of Dataset 5

solid-waste-management-total-waste-landfilled.csv

hhttps://data.gov.sg/dataset/solid-waste-management-total-waste-landfilled-annual?view_id=c89734bc-5081-4730-be0e-83f2f59ab735&resource_id=81292d12-57a5-4e76-a65b-effacc6806b7

year	total_waste_landfilled
2000	0.36
2001	0.25
2002	0.20
2003	0.19
2004	0.22
2005	0.27
2006	0.23
2007	0.19
2008	0.18
2009	0.15
2010	0.17
2011	0.20
2012	0.20
2013	0.20
2014	0.17
2015	0.19

Information of Dataset 5:

```
**** Total Waste Landfilled in Singapore from 2000 to 2015 ****

Information of dataset 5
<class 'pandas.core.frame.DataFrame'>
Int64Index: 16 entries, 2000 to 2015
Data columns (total 1 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   total_waste_landfilled  16 non-null    float64
dtypes: float64(1)
```

Datasets 1 & 2 Cleaning

Dataset 1 :

	waste_type	waste_disposed_of_tonne	total_waste_recycled_tonne	total_waste_generated_tonne	recycling_rate	year
0	Food	679900	111100.0	791000	0.14	2016
1	Paper/Cardboard	576000	607100.0	1183100	0.51	2016
2	Plastics	762700	59500.0	822200	0.07	2016
3	C&D	9700	1585700.0	1595400	0.99	2016
4	Horticultural waste	111500	209000.0	320500	0.65	2016

Dataset 2 :

	Waste Type	Total Generated ('000 tonnes)	Total Recycled ('000 tonnes)	Year
0	Construction& Demolition	1624	1618	2018
1	Ferrous Metal	1269	1260	2018
2	Paper/Cardboard	1054	586	2018
3	Plastics	949	41	2018
4	Food	763	126	2018

- Datasets 1 & 2 are related to Waste Management across the years & have common columns.
- Hence, we will be cleaning the datasets with the intention to concatenate them afterwards

Check for Missing Values

Pandas function used:

- **isnull().sum()**

```
** Detecting Missing Values in Dataset 1 **
waste_type                      0
waste_disposed_of_tonne         0
total_waste_recycled_tonne      0
total_waste_generated_tonne     0
recycling_rate                  0
year                            0
dtype: int64

** Detecting Missing Values in Dataset 2 **
Waste Type                        0
Total Generated ('000 tonnes)     0
Total Recycled ('000 tonnes)      0
Year                              0
dtype: int64
```

There are no missing values, but we will check again after concatenating the datasets

Check if Values are Represented Consistently

Dataset 1
:

	waste_type	waste_disposed_of_tonne	total_waste_recycled_tonne	total_waste_generated_tonne	recycling_rate	year
0	Food	679900	111100.0	791000	0.14	2016
1	Paper/Cardboard	576000	607100.0	1183100	0.51	2016
2	Plastics	762700	59500.0	822200	0.07	2016
3	C&D	9700	1585700.0	1595400	0.99	2016
4	Horticultural waste	111500	209000.0	320500	0.65	2016

Dataset 2
:

	Waste Type	Total Generated ('000 tonnes)	Total Recycled ('000 tonnes)	Year
0	Construction& Demolition	1624	1618	2018
1	Ferrous Metal	1269	1260	2018
2	Paper/Cardboard	1054	586	2018
3	Plastics	949	41	2018
4	Food	763	126	2018

Correct Values Under ‘Total Generated’ & ‘Total Recycled’ In Dataset 2

```
df2["Total Generated ('000 tonnes)"] = df2["Total Generated ('000 tonnes)"] * 1000
df2["Total Recycled ('000 tonnes)"] = df2["Total Recycled ('000 tonnes)"] * 1000
```

	Waste Type	Total Generated ('000 tonnes)	Total Recycled ('000 tonnes)	Year
0	Construction& Demolition	1624000	1618000	2018
1	Ferrous Metal	1269000	1260000	2018
2	Paper/Cardboard	1054000	586000	2018
3	Plastics	949000	41000	2018
4	Food	763000	126000	2018

Standardise Column Names

```
** Columns in Dataset 1 **
Index(['waste_type', 'waste_disposed_of_tonne', 'total_waste_recycled_tonne',
       'total_waste_generated_tonne', 'recycling_rate', 'year'],
      dtype='object')

** Columns in Dataset 2 **
Index(['Waste Type', 'Total Generated ('000 tonnes)',
       'Total Recycled ('000 tonnes)', 'Year'],
      dtype='object')
```

Pandas function used:

- **.rename()**

```
# Standardise Column Names
df1 = df1.rename(columns = {"waste_type":"Waste Type", "waste_disposed_of_tonne":
                            "total_waste_recycled_tonne":"Total Waste Recycled",
                            "total_waste_generated_tonne":"Total Waste Generated",
                            "recycling_rate":"Recycling Rate", "year":"Year"})

df2 = df2.rename(columns = {"Total Generated ('000 tonnes)":"Total Waste Generate
                            "Total Recycled ('000 tonnes)":"Total Waste Recycled"})
```

Check Unique Column Values In Each Dataset

```
** Unique Values in Dataset 1 **
 Total No.: 32
{'Ash & Sludge',
 'Ash and sludge',
 'C&D',
 'Construction Debris',
 'Construction debris',
 'Ferrous Metal',
 'Ferrous Metals',
 'Ferrous metal',
 'Food',
 'Food waste',
 'Glass',
 'Horticultural Waste',
 'Horticultural waste',
 'Non-ferrous Metals',
 'Non-ferrous metal',
 'Non-ferrous metals',
 'Others',
 'Others (stones, ceramic, rubber, etc.)',
 'Others (stones, ceramics & rubber etc)',
 'Others (stones, ceramics & rubber etc.)',
 'Paper/Cardboard',
 'Plastic',
 'Plastics',
 'Scrap Tyres',
 'Scrap tyres',
 'Sludge',
 'Textile/Leather',
 'Total',
 'Used Slag',
 'Used slag',
 'Wood',
 'Wood/Timber'}
```

```
** Unique Values in Dataset 2 **
 Total No.: 22
{'Ash & Sludge',
 'Ash & sludge',
 'Construction & Demolition',
 'Construction& Demolition',
 'Ferrous Metal',
 'Ferrous metal',
 'Food',
 'Glass',
 'Horticultural',
 'Non-Ferrous Metal',
 'Non-ferrous metal',
 'Others (stones, ceramic, rubber, ect)',
 'Others (stones, ceramics, etc.)',
 'Overall',
 'Paper/Cardboard',
 'Plastics',
 'Scrap Tyres',
 'Scrap tyres',
 'Textile/Leather',
 'Used Slag',
 'Used slag',
 'Wood'}
```

Pandas function used:

- **.unique()**

Expect No. of Unique Values:

```
** Latest Naming Convention **
Total No.: 15
{'Ash & sludge',
 'Construction & Demolition',
 'Ferrous metal',
 'Food',
 'Glass',
 'Horticultural',
 'Non-ferrous metal',
 'Others (stones, ceramics, etc.)',
 'Overall',
 'Paper/Cardboard',
 'Plastics',
 'Scrap tyres',
 'Textile/Leather',
 'Used slag',
 'Wood'}
```

Standardise Column Values Under ‘Waste Type’

```
# 1. Ash & sludge
df1 = df1.replace('Ash & Sludge', 'Ash & sludge')
df1 = df1.replace('Sludge', 'Ash & sludge')
df1 = df1.replace('Ash and sludge', 'Ash & sludge')
df2 = df2.replace('Ash & Sludge', 'Ash & sludge')
```

Pandas function used:

- **.replace()**

After Standardising:

```
** Unique Values in Dataset 1 **
 Total No.: 15

 ['Food' 'Paper/Cardboard' 'Plastics' 'Construction & Demolition'
 'Horticultural' 'Wood' 'Ferrous metal' 'Non-ferrous metal' 'Used slag'
 'Ash & sludge' 'Glass' 'Textile/Leather' 'Scrap tyres'
 'Others (stones, ceramics, etc.)' 'Overall']


** Unique Values in Dataset 2 **
 Total No.: 15

 ['Construction & Demolition' 'Ferrous metal' 'Paper/Cardboard' 'Plastics'
 'Food' 'Wood' 'Horticultural' 'Ash & sludge' 'Textile/Leather'
 'Used slag' 'Non-ferrous metal' 'Glass' 'Scrap tyres'
 'Others (stones, ceramics, etc.)' 'Overall']
```

Concatenate Dataset 1 & 2

Dataset 1 :

	Waste Type	Total Waste Disposed	Total Waste Recycled	Total Waste Generated	Recycling Rate	Year
0	Food	679900	111100	791000	0.14	2016
1	Paper/Cardboard	576000	607100	1183100	0.51	2016
2	Plastics	762700	59500	822200	0.07	2016
3	Construction & Demolition	9700	1585700	1595400	0.99	2016
4	Horticultural	111500	209000	320500	0.65	2016

```
# concatenate datasets & reset index
df_concat = pd.concat([df1, df2], ignore_index=True)
```

Dataset 2 :

	Waste Type	Total Waste Generated	Total Waste Recycled	Year
0	Construction & Demolition	1624000	1618000	2018
1	Ferrous metal	1269000	126000	2018
2	Paper/Cardboard	1054000	586000	2018
3	Plastics	949000	41000	2018
4	Food	763000	126000	2018

After concatenating:

	Waste Type	Total Waste Disposed	Total Waste Recycled	Total Waste Generated	Recycling Rate	Year
0	Horticultural	185300.0	119300.0	304600	0.39	2003
1	Paper/Cardboard	618500.0	466200.0	1084700	0.43	2003
2	Plastics	540800.0	39100.0	579900	0.07	2003
3	Construction & Demolition	24600.0	398300.0	422900	0.94	2003
4	Wood	172600.0	40800.0	213400	0.19	2003
...	...	...	...	...	...	...
265	Plastics	NaN	36000.0	868000	NaN	2020
266	Ferrous metal	NaN	930000.0	934000	NaN	2020
267	Paper/Cardboard	NaN	432000.0	1144000	NaN	2020
268	Wood	NaN	195000.0	304000	NaN	2020
269	Overall	NaN	3040000.0	5880000	NaN	2020

270 rows × 6 columns

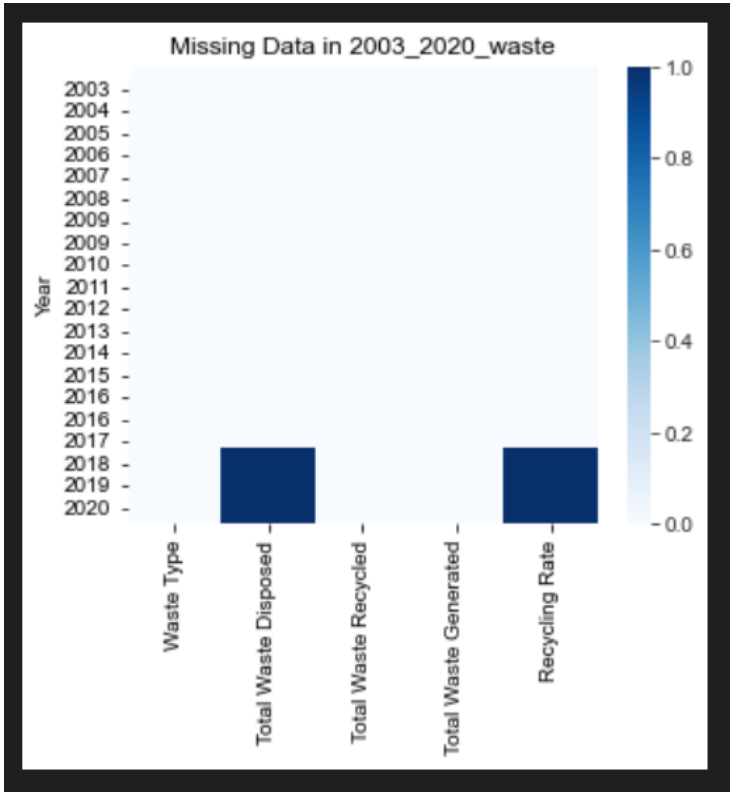
Check for Missing Value

Pandas function used:

- **.isnull().sum()**

```
Waste Type                0
Total Waste Disposed     45
Total Waste Recycled      0
Total Waste Generated     0
Recycling Rate           45
dtype: int64
```

	Waste Type	Total Waste Disposed	Total Waste Recycled	Total Waste Generated	Recycling Rate	Year
225	Construction & Demolition	NaN	1618000	1624000	NaN	2017
226	Glass	NaN	12000	64000	NaN	2018
227	Non-ferrous metal	NaN	170000	171000	NaN	2018
228	Overall	NaN	4726000	7695000	NaN	2018
229	Others (stones, ceramics, etc.)	NaN	11000	286000	NaN	2018
230	Used slag	NaN	179000	181000	NaN	2018
231	Textile/Leather	NaN	14000	220000	NaN	2018
232	Scrap tyres	NaN	29000	32000	NaN	2018
233	Horticultural	NaN	227000	320000	NaN	2018



Handle Missing Value

```
# Handle Missing Values under "Total Waste Disposed"
df_concat.loc[225:, "Total Waste Disposed"] = df_concat.loc[225:, "Total Waste Generated"] - df_concat.loc[225:, "Total Waste Recycled"]

# Handle Missing Values under "Recycling Rate", precision to 2dp
df_concat.loc[225:, "Recycling Rate"] = (df_concat.loc[225:, "Total Waste Recycled"] / df_concat.loc[225:, "Total Waste Generated"]).apply(lambda x: x - x % 0.01)
```

	Waste Type	Total Waste Disposed	Total Waste Recycled	Total Waste Generated	Recycling Rate	Year
0	Horticultural	185300	119300	304600	0.39	2003
1	Paper/Cardboard	618500	466200	1084700	0.43	2003
2	Plastics	540800	39100	579900	0.07	2003
3	Construction & Demolition	24600	398300	422900	0.94	2003
4	Wood	172600	40800	213400	0.19	2003
...	...	...	...	...	...	...
265	Plastics	832000	36000	868000	0.04	2020
266	Ferrous metal	4000	930000	934000	0.99	2020
267	Paper/Cardboard	712000	432000	1144000	0.37	2020
268	Wood	109000	195000	304000	0.64	2020
269	Overall	2840000	3040000	5880000	0.51	2020

```
Waste Type              0
Total Waste Disposed    0
Total Waste Recycled    0
Total Waste Generated   0
Recycling Rate          0
Year                    0
dtype: int64
```

Final Cleaned Dataset 1 & 2

```
df_concat.set_index(["Year"], inplace=True)
```

Year	Waste Type	Total Waste Disposed	Total Waste Recycled	Total Waste Generated	Recycling Rate
2003	Horticultural	185300	119300	304600	0.39
2003	Paper/Cardboard	618500	466200	1084700	0.43
2003	Plastics	540800	39100	579900	0.07
2003	Construction & Demolition	24600	398300	422900	0.94
2003	Wood	172600	40800	213400	0.19
...	...	...	...	...	...
2020	Plastics	832000	36000	868000	0.04
2020	Ferrous metal	4000	930000	934000	0.99
2020	Paper/Cardboard	712000	432000	1144000	0.37
2020	Wood	109000	195000	304000	0.64
2020	Overall	2840000	3040000	5880000	0.51

Save to New File:

```
# Save Cleaned DataFrame to a new File
df_concat.to_csv('cleaned_2003_2020_waste.csv')
```

Dataset 3 Cleaning

Before:

	Gas	unit	1990	1991	1992	1993	1994	1995	1996	1997	...	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
0	CH4	MtCO2e	1.15	1.18	1.22	1.26	1.30	1.33	1.37	1.42	...	2.45	2.54	2.62	2.71	2.79	2.88	2.96	3.03	3.11	3.19
1	CO2	MtCO2e	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	F-Gas	MtCO2e	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	N2O	MtCO2e	0.05	0.05	0.05	0.05	0.06	0.06	0.06	0.06	...	0.09	0.09	0.10	0.10	0.10	0.10	0.11	0.11	0.11	0.11

Reshaping Dataset:

```
# Drop "unit" column
df3_v1 = df3.drop(['unit'], axis=1)
# Transpose to make gas & the years the rows
df3_v1 = df3_v1.T

# set column names to values under Gas row at index 0
df3_v1.columns = df3_v1.iloc[0]
# remove that row, which is now set as the column headers
df3_v1 = df3_v1[1:]
# set index name to Year
df3_v1.index.name = "Year"
```

After reshaping:

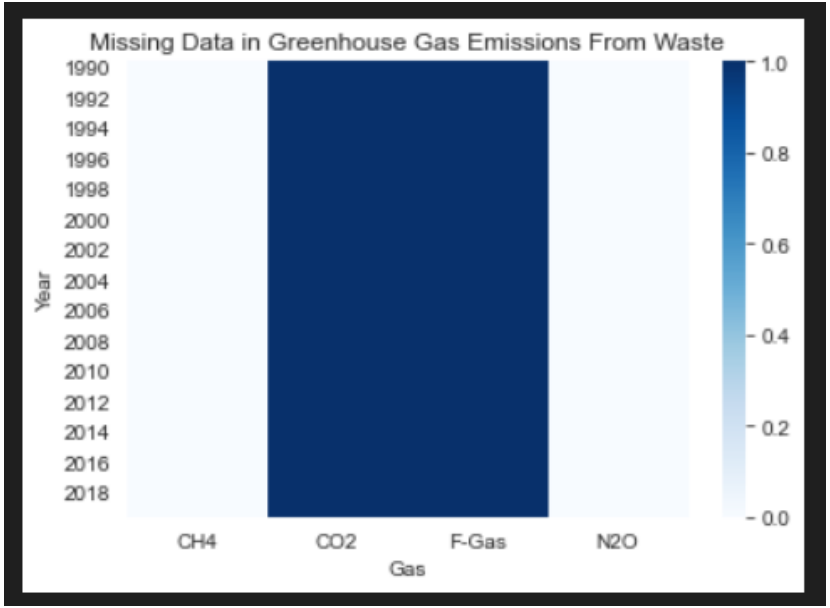
Gas	CH4	CO2	F-Gas	N2O
Year				
1990	1.15	NaN	NaN	0.05
1991	1.18	NaN	NaN	0.05
1992	1.22	NaN	NaN	0.05
1993	1.26	NaN	NaN	0.05
1994	1.3	NaN	NaN	0.06
1995	1.33	NaN	NaN	0.06
1996	1.37	NaN	NaN	0.06
1997	1.42	NaN	NaN	0.06
1998	1.46	NaN	NaN	0.07
1999	1.5	NaN	NaN	0.07
2000	1.54	NaN	NaN	0.07
2001	1.58	NaN	NaN	0.07
2002	1.61	NaN	NaN	0.08
2003	1.64	NaN	NaN	0.08
2004	1.68	NaN	NaN	0.08

Check for Missing Values

Pandas function used:

- **isnull().sum()**

```
** Detecting Missing Values in Dataset 3 **
Gas
CH4       0
CO2      30
F-Gas    30
N2O       0
dtype: int64
```



Handle Missing Value

According to Health New York State:

Methane and carbon dioxide make up 90 to 98% of landfill gas. The remaining 2 to 10% includes nitrogen, oxygen, ammonia, sulfides, hydrogen and various other gases. Landfill gases are produced when bacteria break down organic waste.

https://www.health.ny.gov › environmental › outdoors › air

Important Things to Know About Landfill Gas

- **Drop columns 'CO2' & 'F-Gas'**
- **Methane(CH4) & Nitrous Oxide(N20) are the more significant greenhouse gases produced from waste**

```
df3_v2 = df3_v1.drop(['CO2', 'F-Gas'], axis = 1)
```

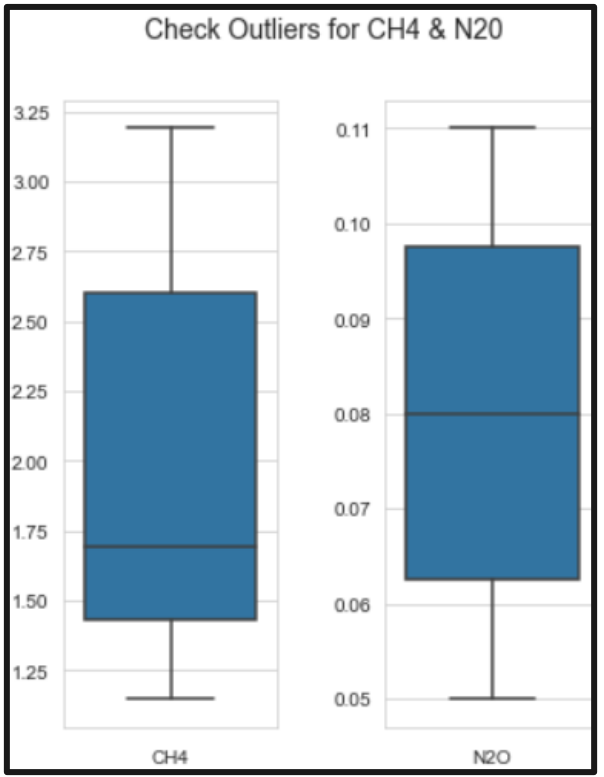
After dropping the columns:

Year	Gas	CH4	N2O
1990		1.15	0.05
1991		1.18	0.05
1992		1.22	0.05
1993		1.26	0.05
1994		1.3	0.06
1995		1.33	0.06
1996		1.37	0.06
1997		1.42	0.06
1998		1.46	0.07
1999		1.5	0.07
2000		1.54	0.07
2001		1.58	0.07
2002		1.61	0.08
2003		1.64	0.08

Check for Outliers Using BoxPlot

```
# 1st subplot
sns.boxplot(data=df3_v2['CH4'], ax=ax[0])
ax[0].set_xlabel('CH4')

# 2nd subplot
sns.boxplot(data=df3_v2['N2O'], ax=ax[1])
ax[1].set_xlabel('N2O')
```



Final Cleaned Dataset 3

Gas	CH4	N2O
Year		
1990	1.15	0.05
1991	1.18	0.05
1992	1.22	0.05
1993	1.26	0.05
1994	1.30	0.06
1995	1.33	0.06
1996	1.37	0.06
1997	1.42	0.06
1998	1.46	0.07
1999	1.50	0.07
2000	1.54	0.07
2001	1.58	0.07
2002	1.61	0.08
2003	1.64	0.08
2004	1.68	0.08
2005	1.71	0.08
2006	1.86	0.08
2007	2.01	0.09
2008	2.15	0.09
2009	2.30	0.09

Nature of Dataset:

```
<class 'pandas.core.frame.DataFrame'>
Index: 30 entries, 1990 to 2019
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   CH4     30 non-null     float64
 1   N2O     30 non-null     float64
dtypes: float64(2)
```

Save to New File:

```
df3_v2.to_csv('cleaned_ghg-emissions.csv')
```

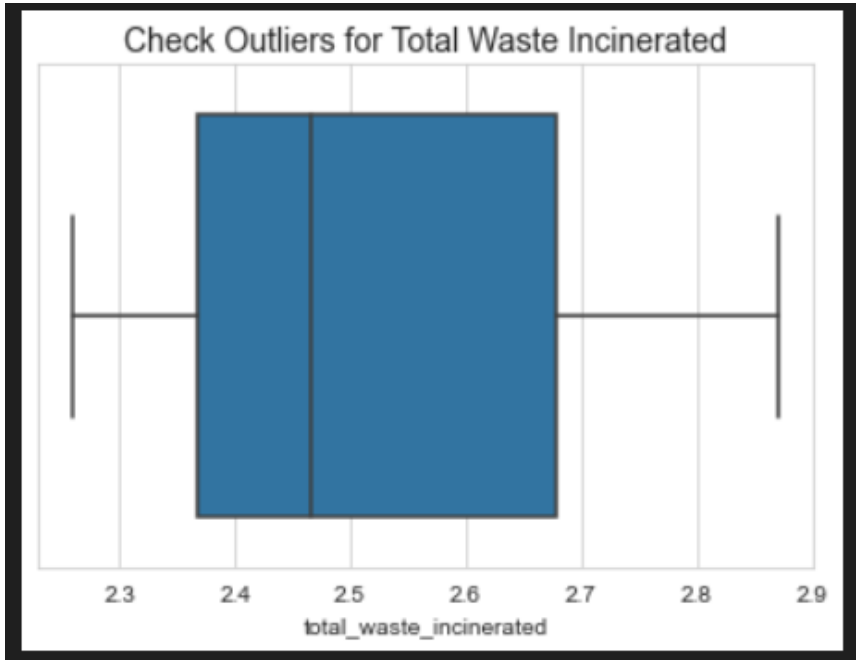
Dataset 4 Cleaning

Checking for Missing Values:

- **isnull().sum()**

```
** Detecting Missing Values in Dataset 5 **
total_waste_incinerated    0
dtype: int64
```

Checking for Outliers:



- **No outlier**

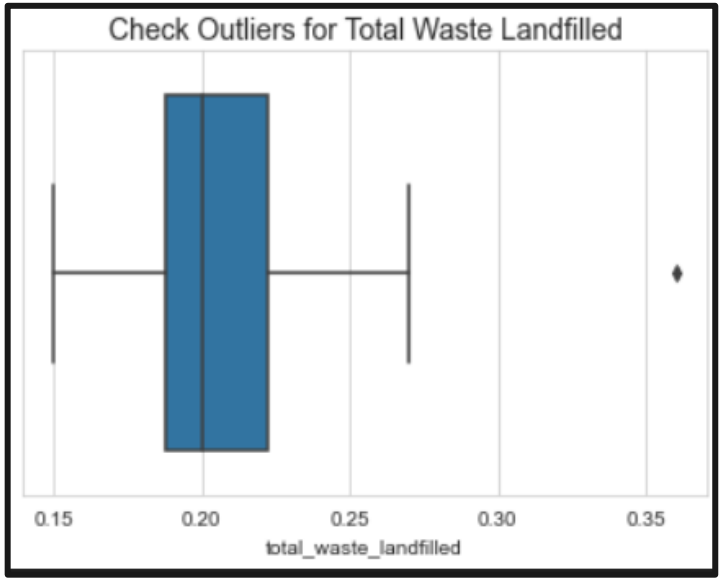
Dataset 5 Cleaning

Checking for Missing Values:

- **isnull().sum()**

```
** Detecting Missing Values in Dataset 5 **
total_waste_landfilled    0
dtype: int64
```

Checking for Outliers:



- **There's 1 outlier, but we will ignore it as we want to see how Singapore is managing waste sent to the landfill**

Concatenate Datasets 4 & 5

- As both datasets 4 & 5 are related to Waste Disposal, we will concatenate them together

```
# Concatenate Dataset 4 & 5 **
df4_5concat = pd.concat([df4, df5], axis=1)
```

Final Dataset:

year	total_waste_incinerated	total_waste_landfilled
2000	2.44	0.36
2001	2.55	0.25
2002	2.42	0.20
2003	2.31	0.19
2004	2.26	0.22
2005	2.28	0.27
2006	2.33	0.23
2007	2.38	0.19
2008	2.45	0.18
2009	2.48	0.15
2010	2.59	0.17
2011	2.66	0.20
2012	2.73	0.20
2013	2.82	0.20
2014	2.87	0.17
2015	2.83	0.19

Save to New File:

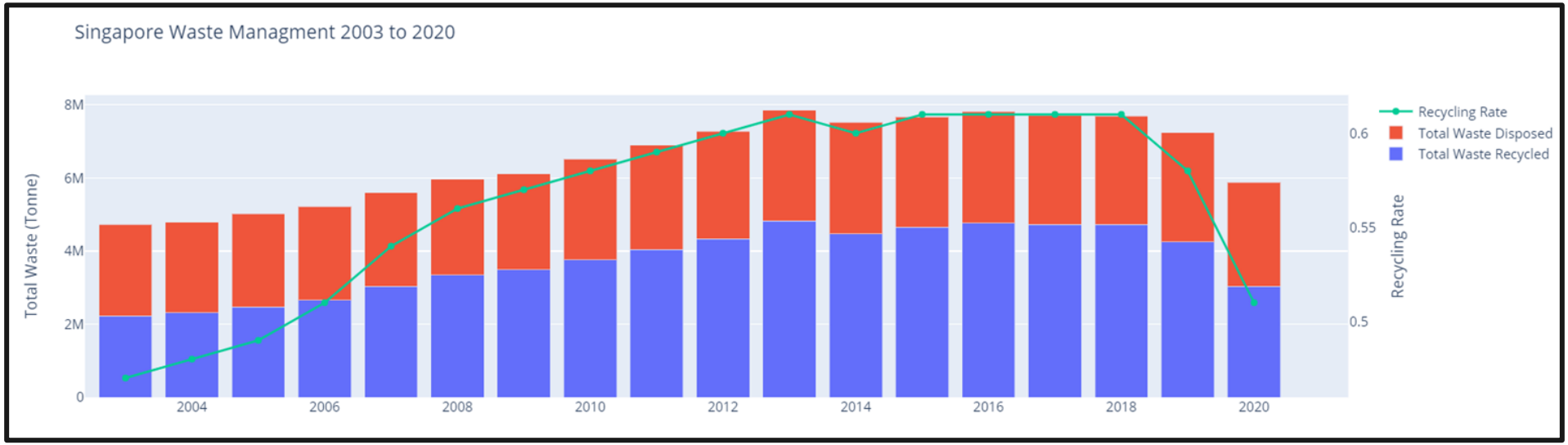
```
df4_5concat.to_csv('cleaned_solid-waste-management-total-waste-landfilled-and-incinerated.csv')
```

Final Datasets

Final No. of Datasets: 3

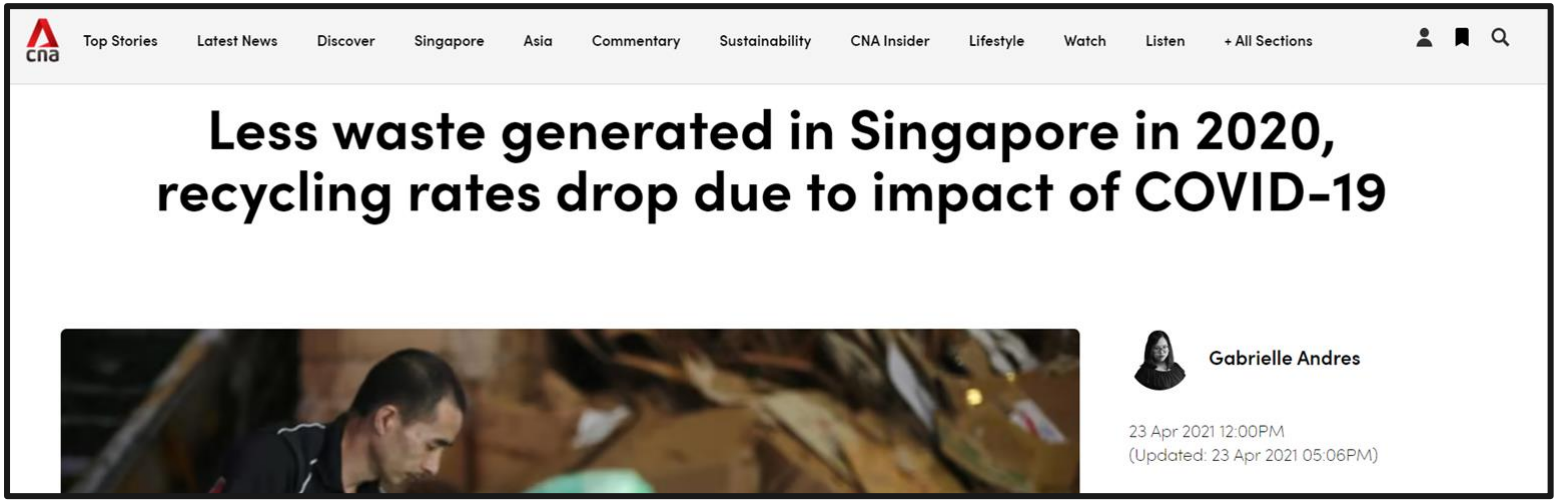
1. cleaned_2003_2020_waste.csv
2. cleaned_ghg-emissions.csv
3. cleaned_solid-waste-management-total-waste-landfilled-and-incinerated.csv

Graph 1: Total Singapore Waste 2003 to 2020



- There is generally an increase in the total waste generated due to rapid urbanisation. But we do see a drop in 2019 when COVID-19 first started.
- The recycling rate also increases over the years, showing great efforts by the government to raise awareness about the importance of recycling, however we do see a drop in 2019 & 2020, when COVID-19 started.

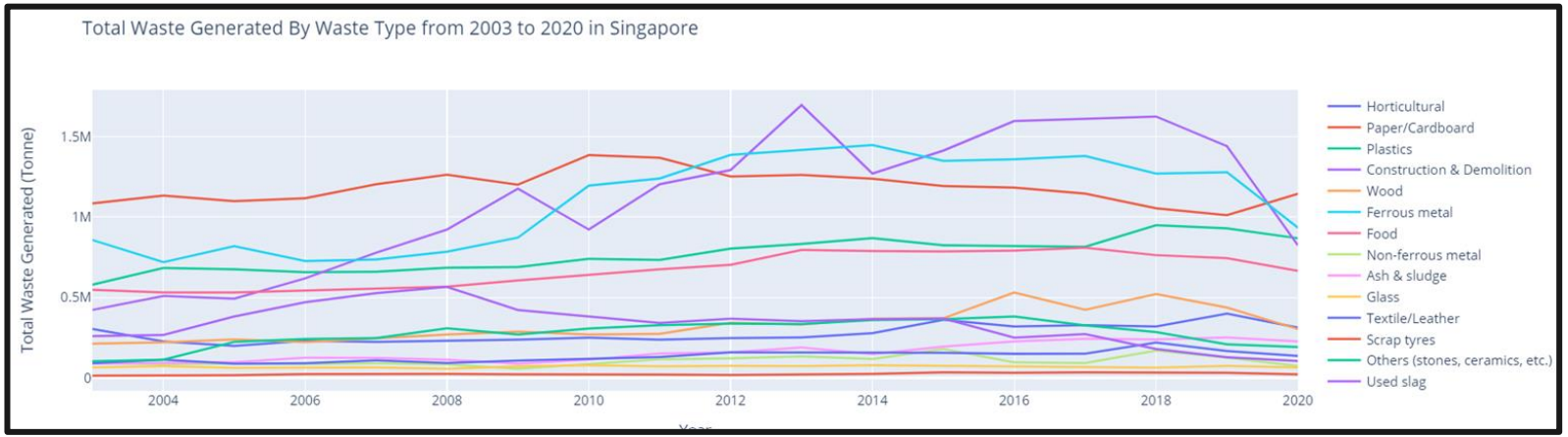
According to CNA:



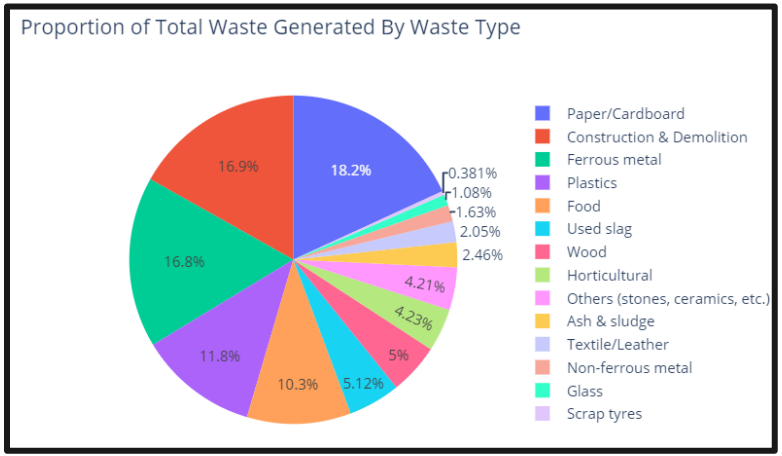
- **Waste Collection Process were either halted or delayed due to the 'circuit breaker' which decreased the recycling rates**

SINGAPORE: Overall waste generation in Singapore declined in 2020 for the fourth consecutive year, with less waste being sent to the Semakau Landfill, but overall recycling rates also dropped as the COVID-19 pandemic impact industries and collection process.

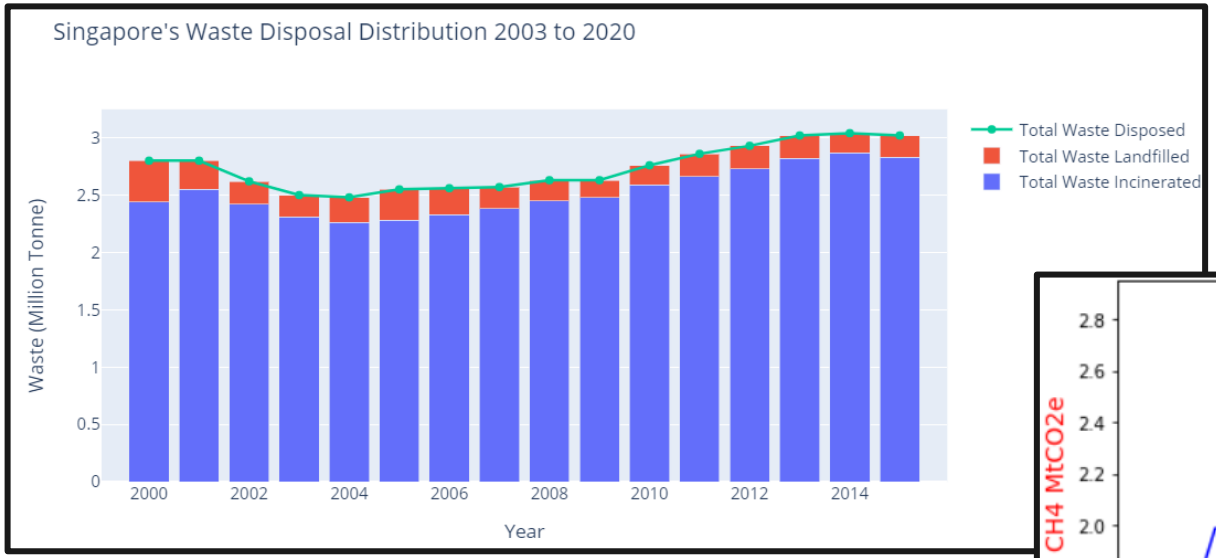
Graph 2/3: Total Waste Generated by Waste Type 2003 to 2020



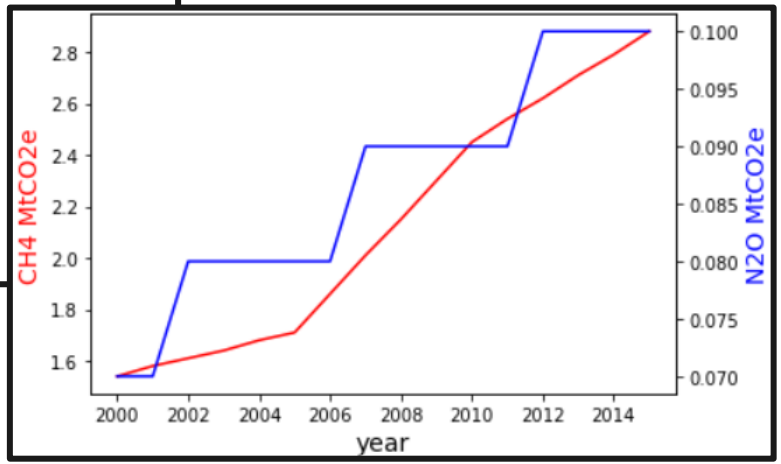
- **There are 3 prominent Waste Types generating the most waste which are 'Paper/Cardboard', 'Construction & Demolition' & 'Ferrous Metal',**
- **However, there is a big drop in "Construction & Demolition" & "Ferrous Metal" due to the COVID-19 pandemic as construction work have to be paused & delayed due to the 'Circuit Breaker" & "Health Measures"**



Graph 4/5: Total Waste Disposed & Greenhouse Gas Emissions From Waste



- **Moreover, Singapore is closely managing the disposal of waste as it emits the greenhouse gases CH4 & N2O , which warms the earth.**



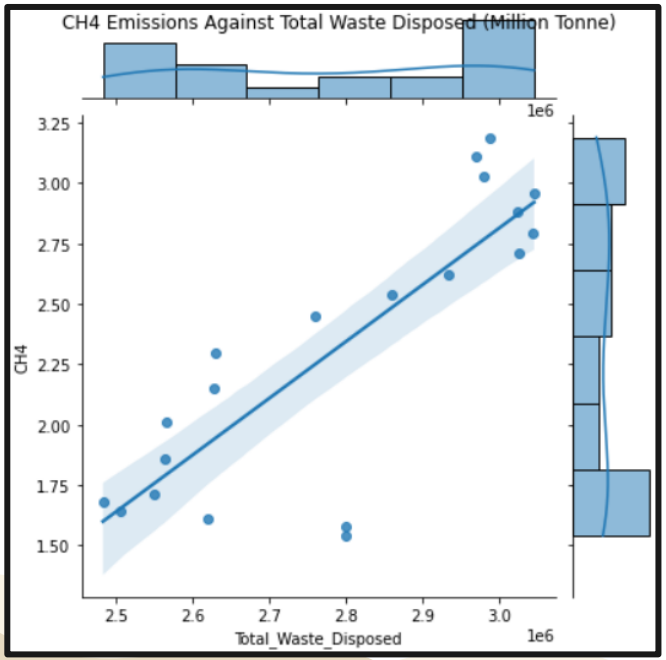
Insights:
Over the years, there is a gentle increase in the total amount of waste disposed. Most of the waste is incinerated, while the rest is landfilled to control the burning of waste to control the air quality & emissions of air pollutants like CO2, which contributes to climate change

Moreover, according to World Economic Forum:

- **The waste sector is responsible for 20% of global methane emissions. And these methane emissions have more than 80 times the warming power of CO_2, hence it is critical to monitor how waste is disposed**
- **Thus, next we will be fitting a simple linear model to see if we can predict CH_4 emissions based on the total waste disposed**

Graph 6: Simple Linear Model for CH4 Emissions vs Total Waste Disposed

There is a positive correlation between Total Waste Disposed & CH4 emissions from it



There's a R*2 value of 0.683, indicating that:
Linear model is a moderate fit
There's moderate correlation between Total Waste Disposed & CH4 Emissions

```
Intercept              -4.232462
Total_Waste_Disposed    0.000002
dtype: float64
                            OLS Regression Results
==============================================================================
Dep. Variable:                    CH4   R-squared:                       0.683
Model:                            OLS   Adj. R-squared:                  0.666
Method:                 Least Squares   F-statistic:                     38.83
Date:                Mon, 06 Feb 2023   Prob (F-statistic):           7.05e-06
Time:                        03:09:52   Log-Likelihood:                -5.4534
No. Observations:                  20   AIC:                             14.91
Df Residuals:                      18   BIC:                             16.90
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      -4.2325      1.054     -4.016      0.001      -6.447      -2.018
Total_Waste_Disposed  2.349e-06   3.77e-07      6.231      0.000   1.56e-06    3.14e-06
==============================================================================
Omnibus:                        8.878   Durbin-Watson:                   0.263
Prob(Omnibus):                  0.012   Jarque-Bera (JB):                6.245
Skew:                          -1.244   Prob(JB):                       0.0441
Kurtosis:                       4.143   Cond. No.                     3.93e+07
==============================================================================
...
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.93e+07. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Fitting Linear Model

Graph 7: Total Waste Sent to Semakau

Insights:

- **Big drop of 30% in waste sent to Semakau Landfill from 2000 to 2001. After this big drop, the amount of waste sent continuously decreases & remains low**
- **This is because, the Semakau Landfill will run out of space by 2035 at the current rate of waste growth**
- **Hence, the government is controlling the amount of waste sent by gradually decreasing the rate, in order to have increase the lifespan of Semakau by 5 to 10 more years**

According to Zero Waste Nation.gov:

Semakau Landfill

Semakau Landfill is Singapore's one and only landfill. It receives more than 2,000 tonnes of WTE incineration ash and non- incinerable waste daily.

At our current rate of waste growth, **Semakau Landfill will run out of space by 2035.**

Conclusion

- **Waste Management plays a big part in mitigating climate change & all the scarce materials on earth by recycling & being responsible about our consumption of resources**
- **By ensuring responsible consumption of resources, it reduces greenhouse gases emissions from waste, which contributes to climate change & the warming of earth's surface**
- **It is also especially important in Singapore where there is limited land & the only landfill, Semakau, we have is slowly running low on its lifespan over the next few decades**

World Health Organisation Sustainable Goal 12:

Goal 12:
Ensure sustainable consumption and production patterns

References

- https://www.channelnewsasia.com/singapore/less-waste-generated-singapore-recycling-rates-drop-239601
- https://www.health.ny.gov/environmental/outdoors/air/landfill_gas.htm#:~:text=Methane%20and%20carbon%20dioxide%20make,bacteria%20break%20down%20organic%20waste.
- https://www.weforum.org/agenda/2022/11/waste-emissions-methane-cities/#:~:text=The%20waste%20sector%20is%20responsible,for%20Incinerator%20Alternatives%20(GAIA).
- https://www.towardszerowaste.gov.sg/zero-waste-nation/#:~:text=Semakau%20Landfill%20is%20Singapore's%20one,out%20of%20space%20by%202035.

Datasets:

- https://www.kaggle.com/datasets/mahaanand/singapore-waster-mgmt?select=2003_2017_waste.csv
- https://www.kaggle.com/datasets/mahaanand/singapore-waster-mgmt?select=2018_2020_waste.csv
- https://www.climatewatchdata.org/ghg-emissions?breakBy=gas&end_year=2019&gases=all-ghg&regions=SGP&sectors=waste&start_year=1990
- https://data.gov.sg/dataset/solid-waste-management-total-waste-incinerated-annual?view_id=882ae208-bd25-4b99-8cc9-a3b7b24d063a&resource_id=e4c8461f-e7de-4fc3-ad25-cf068ae09509
- https://data.gov.sg/dataset/solid-waste-management-total-waste-landfilled-annual?view_id=c89734bc-5081-4730-be0e-83f2f59ab735&resource_id=81292d12-57a5-4e76-a65b-effacc6806b7