

# Global Plastic Waste

```
In [1]: #Importing Libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
%matplotlib inline

warnings.filterwarnings('ignore')
sns.set_style('whitegrid')
```

```
In [2]: '''reading the Dataset'''
#as GDP Per capita Increases Plastic waste Increases
df=pd.read_csv('per-capita-plastic-waste-vs-gdp-per-capita.csv')
```

```
In [3]: '''displaying first five rows'''
df.head()
```

Out[3]:

	Entity	Code	Year	Per capita plastic waste (kg/person/day)	GDP per capita, PPP (constant 2011 international \$)	Total population (Gapminder, HYDE & UN)	Continent
0	Abkhazia	OWID_ABK	2015	NaN	NaN	NaN	Asia
1	Afghanistan	AFG	2002	NaN	1063.635574	22601000.0	NaN
2	Afghanistan	AFG	2003	NaN	1099.194507	23681000.0	NaN
3	Afghanistan	AFG	2004	NaN	1062.249360	24727000.0	NaN
4	Afghanistan	AFG	2005	NaN	1136.123214	25654000.0	NaN

```
In [4]: '''shape of the data'''
df.shape
```

Out[4]: (48168, 7)

```
In [5]: '''checking the null values'''
df.isnull().sum()
```

```
Out[5]: Entity                                0
Code                                           2014
Year                                           0
Per capita plastic waste (kg/person/day)      47982
GDP per capita, PPP (constant 2011 international $) 41761
Total population (Gapminder, HYDE & UN)      1285
Continent                                     47883
dtype: int64
```

```
In [6]: '''checking percentage of null values at each column'''
for column in df.columns:
    print("{} has {:.2f}% null values:".format(column,df[column].isnull().sum()/len(df[column])*100))
```

Entity has 0.00% null values:

-----

Code has 4.18% null values:

-----

Year has 0.00% null values:

-----

Per capita plastic waste (kg/person/day) has 99.61% null values:

-----

GDP per capita, PPP (constant 2011 international \$) has 86.70% null values:

-----

Total population (Gapminder, HYDE & UN) has 2.67% null values:

-----

Continent has 99.41% null values:

-----

```
In [7]: '''checking info of data'''
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48168 entries, 0 to 48167
Data columns (total 7 columns):
 #   Column                                                                 Non-Null Count  Dtype
---  -
 0   Entity                                                                48168 non-null  object
 1   Code                                                                    46154 non-null  object
 2   Year                                                                    48168 non-null  int64
 3   Per capita plastic waste (kg/person/day)                             186 non-null    float64
 4   GDP per capita, PPP (constant 2011 international $)                 6407 non-null   float64
 5   Total population (Gapminder, HYDE & UN)                             46883 non-null  float64
 6   Continent                                                              285 non-null    object
dtypes: float64(3), int64(1), object(3)
memory usage: 2.6+ MB
```

```
In [8]: '''renaming columns in the dataframe'''
df.rename(columns={'GDP per capita, PPP (constant 2011 international $)': 'GDP per capita in PPP',
                  'Total population (Gapminder, HYDE & UN)': 'Total Population',
                  'Per capita plastic waste (kg/person/day)': 'Waste per person'})
#if we place False in place of inplace a new dataframe will get created with changes
#if we use True in place of inplace changes will happen to the existing data frame
```

```
In [9]: df.head()
```

Out[9]:

	Entity	Code	Year	Waste per person(kg/day)	GDP per capita in PPP	Total Population	Continent
0	Abkhazia	OWID_ABK	2015	NaN	NaN	NaN	Asia
1	Afghanistan	AFG	2002	NaN	1063.635574	22601000.0	NaN
2	Afghanistan	AFG	2003	NaN	1099.194507	23681000.0	NaN
3	Afghanistan	AFG	2004	NaN	1062.249360	24727000.0	NaN
4	Afghanistan	AFG	2005	NaN	1136.123214	25654000.0	NaN

```
In [10]: '''removing entities/countries with incomplete or missing data'''
incmp_df_idx=df[(df['Total Population'].isna())&(df['GDP per capita in PPP'].isna())]
df.drop(incmp_df_idx,inplace=True)
```

In [11]: `df.head()`

Out[11]:

	Entity	Code	Year	Waste per person(kg/day)	GDP per capita in PPP	Total Population	Continent
1	Afghanistan	AFG	2002	NaN	1063.635574	22601000.0	NaN
2	Afghanistan	AFG	2003	NaN	1099.194507	23681000.0	NaN
3	Afghanistan	AFG	2004	NaN	1062.249360	24727000.0	NaN
4	Afghanistan	AFG	2005	NaN	1136.123214	25654000.0	NaN
5	Afghanistan	AFG	2006	NaN	1161.124889	26433000.0	NaN

In [12]: `df.shape`

Out[12]: (48113, 7)

In [13]: `'''retrieving rows in the year-2010'''`  
`df_2010=df[df['Year']==2010]`  
`df_2010=df_2010.drop(columns='Continent')`

In [14]: `df_2010.head()`

Out[14]:

	Entity	Code	Year	Waste per person(kg/day)	GDP per capita in PPP	Total Population
9	Afghanistan	AFG	2010	NaN	1614.255001	2.918600e+07
333	Africa	NaN	2010	NaN	NaN	1.039304e+09
344	Albania	ALB	2010	0.069	9927.181841	2.948000e+06
564	Algeria	DZA	2010	0.144	12870.602699	3.597700e+07
844	American Samoa	ASM	2010	NaN	NaN	5.600000e+04

In [15]: `'''retrieving continent name in which year==2015'''`  
`df_2015=df[df['Year']==2015]`  
`df_2010['Continent']=df_2015['Continent'].values`

In [16]: `df_2015.head()`

Out[16]:

	Entity	Code	Year	Waste per person(kg/day)	GDP per capita in PPP	Total Population	Continent
14	Afghanistan	AFG	2015	NaN	1809.016488	3.441400e+07	Asia
338	Africa	NaN	2015	NaN	NaN	1.182439e+09	NaN
369	Albania	ALB	2015	NaN	10970.452245	2.891000e+06	Europe
589	Algeria	DZA	2015	NaN	13724.723853	3.972800e+07	Africa
849	American Samoa	ASM	2015	NaN	NaN	5.600000e+04	Oceania

In [17]: `df_2010.head()`

Out[17]:

	Entity	Code	Year	Waste per person(kg/day)	GDP per capita in PPP	Total Population	Continent
9	Afghanistan	AFG	2010	NaN	1614.255001	2.918600e+07	Asia
333	Africa	NaN	2010	NaN	NaN	1.039304e+09	NaN
344	Albania	ALB	2010	0.069	9927.181841	2.948000e+06	Europe
564	Algeria	DZA	2010	0.144	12870.602699	3.597700e+07	Africa
844	American Samoa	ASM	2010	NaN	NaN	5.600000e+04	Oceania

In [18]: `'''dropping rows with missing Continent values using index'''`  
`missing_idx=df_2010[df_2010['Continent'].isna()].index`  
`df_2010.drop(missing_idx,inplace=True)`

In [19]: `'''dropping rows with missing per person waste generation values'''`  
`df_2010=df_2010[df_2010['Waste per person(kg/day)'].notna()]`  
`wa_g=df_2010.reset_index().drop('index',axis=1)`

In [20]: `df_2010.head()`

Out[20]:

	Entity	Code	Year	Waste per person(kg/day)	GDP per capita in PPP	Total Population	Continent
344	Albania	ALB	2010	0.069	9927.181841	2948000.0	Europe
564	Algeria	DZA	2010	0.144	12870.602699	35977000.0	Africa
1074	Angola	AGO	2010	0.062	5897.682841	23356000.0	Africa
1294	Anguilla	AIA	2010	0.252	NaN	13000.0	North America
1365	Antigua and Barbuda	ATG	2010	0.660	19212.720131	88000.0	North America

In [21]: `wa_g.head()`

Out[21]:

	Entity	Code	Year	Waste per person(kg/day)	GDP per capita in PPP	Total Population	Continent
0	Albania	ALB	2010	0.069	9927.181841	2948000.0	Europe
1	Algeria	DZA	2010	0.144	12870.602699	35977000.0	Africa
2	Angola	AGO	2010	0.062	5897.682841	23356000.0	Africa
3	Anguilla	AIA	2010	0.252	NaN	13000.0	North America
4	Antigua and Barbuda	ATG	2010	0.660	19212.720131	88000.0	North America

In [22]: `'''reading the second file'''`  
`df2=pd.read_csv('per-capita-mismanaged-plastic-waste-vs-gdp-per-capita.csv')`

In [23]: `'''diplaying first 5 rows of df2 dataframe'''`  
`df2.head()`

Out[23]:

	Entity	Code	Year	Per capita mismanaged plastic waste	GDP per capita, PPP (constant 2011 international \$)	Total population (Gapminder, HYDE & UN)	Continent
0	Abkhazia	OWID_ABK	2015	NaN	NaN	NaN	Asia
1	Afghanistan	AFG	2002	NaN	1063.635574	22601000.0	NaN
2	Afghanistan	AFG	2003	NaN	1099.194507	23681000.0	NaN
3	Afghanistan	AFG	2004	NaN	1062.249360	24727000.0	NaN
4	Afghanistan	AFG	2005	NaN	1136.123214	25654000.0	NaN

In [24]: `'''renaming coulumns'''`  
`df2.rename(columns={'Per capita mismanaged plastic waste': 'Mismanaged waste per',`  
`'GDP per capita, PPP (constant 2011 international $)': 'GDP per',`  
`'Total population (Gapminder, HYDE & UN)': 'Total Population'},`

In [25]: `df2.head()`

Out[25]:

	Entity	Code	Year	Mismanaged waste per person(kg/day)	GDP per capita in PPP	Total Population	Continent
0	Abkhazia	OWID_ABK	2015	NaN	NaN	NaN	Asia
1	Afghanistan	AFG	2002	NaN	1063.635574	22601000.0	NaN
2	Afghanistan	AFG	2003	NaN	1099.194507	23681000.0	NaN
3	Afghanistan	AFG	2004	NaN	1062.249360	24727000.0	NaN
4	Afghanistan	AFG	2005	NaN	1136.123214	25654000.0	NaN

```
In [26]: '''dropping continent column'''
df2.drop('Continent',axis=1,inplace=True)
```

```
In [27]: '''retrieving rows in which year is 2010'''
df2_2010=df2[df2.Year==2010]
df2_2010.head()
```

Out[27]:

	Entity	Code	Year	Mismanaged waste per person(kg/day)	GDP per capita in PPP	Total Population
9	Afghanistan	AFG	2010	NaN	1614.255001	2.918600e+07
333	Africa	NaN	2010	NaN	NaN	1.039304e+09
344	Albania	ALB	2010	0.032	9927.181841	2.948000e+06
564	Algeria	DZA	2010	0.086	12870.602699	3.597700e+07
844	American Samoa	ASM	2010	NaN	NaN	5.600000e+04

```
In [28]: '''dropping rows with missing mismanaged values'''
df2_2010=df2_2010[df2_2010['Mismanaged waste per person(kg/day)'].isna()!=True]
```

```
In [29]: df2_2010.head()
```

Out[29]:

	Entity	Code	Year	Mismanaged waste per person(kg/day)	GDP per capita in PPP	Total Population
344	Albania	ALB	2010	0.032	9927.181841	2948000.0
564	Algeria	DZA	2010	0.086	12870.602699	35977000.0
1074	Angola	AGO	2010	0.045	5897.682841	23356000.0
1294	Anguilla	AIA	2010	0.010	NaN	13000.0
1365	Antigua and Barbuda	ATG	2010	0.051	19212.720131	88000.0

```
In [30]: '''reset index'''
w_m=df2_2010.reset_index()
```

In [31]: `w_m.head()`

Out[31]:

	index	Entity	Code	Year	Mismanaged waste per person(kg/day)	GDP per capita in PPP	Total Population
0	344	Albania	ALB	2010	0.032	9927.181841	2948000.0
1	564	Algeria	DZA	2010	0.086	12870.602699	35977000.0
2	1074	Angola	AGO	2010	0.045	5897.682841	23356000.0
3	1294	Anguilla	AIA	2010	0.010	NaN	13000.0
4	1365	Antigua and Barbuda	ATG	2010	0.051	19212.720131	88000.0

In [32]: `'''drop index column'''`  
`w_m=w_m.drop('index',axis=1)`  
`w_m.head()`

Out[32]:

	Entity	Code	Year	Mismanaged waste per person(kg/day)	GDP per capita in PPP	Total Population
0	Albania	ALB	2010	0.032	9927.181841	2948000.0
1	Algeria	DZA	2010	0.086	12870.602699	35977000.0
2	Angola	AGO	2010	0.045	5897.682841	23356000.0
3	Anguilla	AIA	2010	0.010	NaN	13000.0
4	Antigua and Barbuda	ATG	2010	0.051	19212.720131	88000.0

In [33]: `'''merging w_m and w_g'''`  
`df_plastic_waste=pd.merge(wa_g,w_m,how='inner')`

In [34]: `df_plastic_waste.head()`

Out[34]:

	Entity	Code	Year	Waste per person(kg/day)	GDP per capita in PPP	Total Population	Continent	Mismanaged waste per person(kg/day)
0	Albania	ALB	2010	0.069	9927.181841	2948000.0	Europe	0.032
1	Algeria	DZA	2010	0.144	12870.602699	35977000.0	Africa	0.086
2	Angola	AGO	2010	0.062	5897.682841	23356000.0	Africa	0.045
3	Anguilla	AIA	2010	0.252	NaN	13000.0	North America	0.010
4	Antigua and Barbuda	ATG	2010	0.660	19212.720131	88000.0	North America	0.051



```

In [35]: '''converting column names into list'''
df_plastic_waste.columns.tolist()

'''column names'''
col_names=['Entity',
           'Code',
           'Year',
           'Waste per person(kg/day)',
           'GDP per capita in PPP',
           'Total Population',
           'Continent',
           'Mismanaged waste per person(kg/day)']

df_plastic_waste=df_plastic_waste[col_names]

#'''rounding the values per person'''
#df_plastic_waste.iloc[:,3:5]=np.around(df_plastic_waste[['Waste per person(kg/day)',
#                  'GDP per capita in PPP']],2)

'''changing datatype of Total Population'''
df_plastic_waste['Total Population']=df_plastic_waste['Total Population'].astype(int)

'''generating Total waste and Total mismanaged waste by country'''
df_plastic_waste['Total waste(kgs/year)']=((df_plastic_waste['Waste per person(kg/day)']
                                             *df_plastic_waste['Total Population']))

df_plastic_waste['Total waste mismanaged(kgs/year)']=((df_plastic_waste['Mismanaged waste per person(kg/day)']
                                                         *df_plastic_waste['Total Population']))

```

```

In [36]: df_plastic_waste.head()

```

Out[36]:

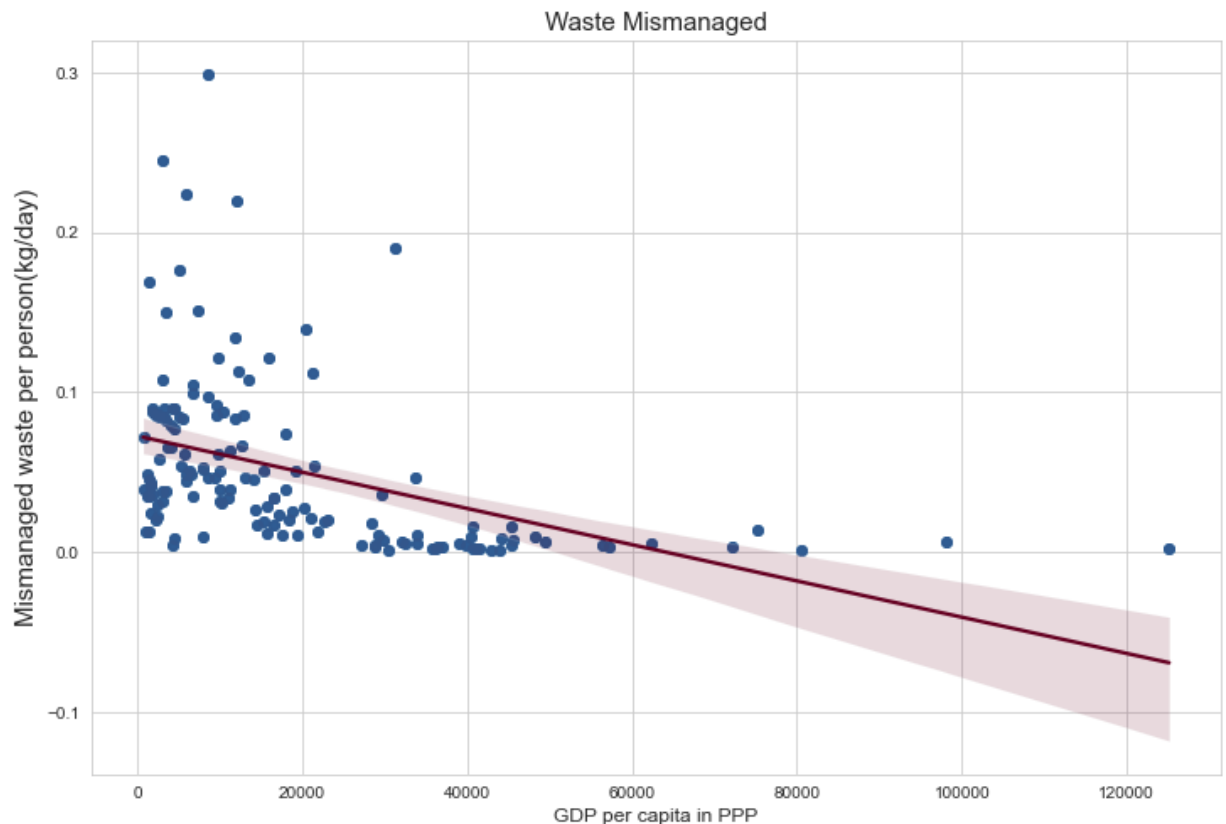
	Entity	Code	Year	Waste per person(kg/day)	GDP per capita in PPP	Total Population	Continent	Mismanaged waste per person(kg/day)	Total waste mismanaged(kgs/year)
0	Albania	ALB	2010	0.069	9927.181841	2948000	Europe	0.032	7.4
1	Algeria	DZA	2010	0.144	12870.602699	35977000	Africa	0.086	1.8
2	Angola	AGO	2010	0.062	5897.682841	23356000	Africa	0.045	5.2
3	Anguilla	AIA	2010	0.252	NaN	13000	North America	0.010	1.3
4	Antigua and Barbuda	ATG	2010	0.660	19212.720131	88000	North America	0.051	2.1

```
In [37]: ''' scatter plot graph '''

plt.figure(1, figsize=(12,8))
plt.scatter(df_plastic_waste['GDP per capita in PPP'], df_plastic_waste['Mismanaged waste per person(kg/day)'])
plt.title('Waste Mismanaged', loc='center', fontsize=15)
plt.ylabel('Mismanaged waste per person(kg/day)', loc='center', fontsize=15)
plt.xlabel('GDP per capita in PPP', fontsize=12)

sns.regplot(x=df_plastic_waste['GDP per capita in PPP'], y=df_plastic_waste['Mismanaged waste per person(kg/day)'],
            scatter_kws={'color': '#34568B'}, line_kws={'color': '#650021'})

plt.show()
```

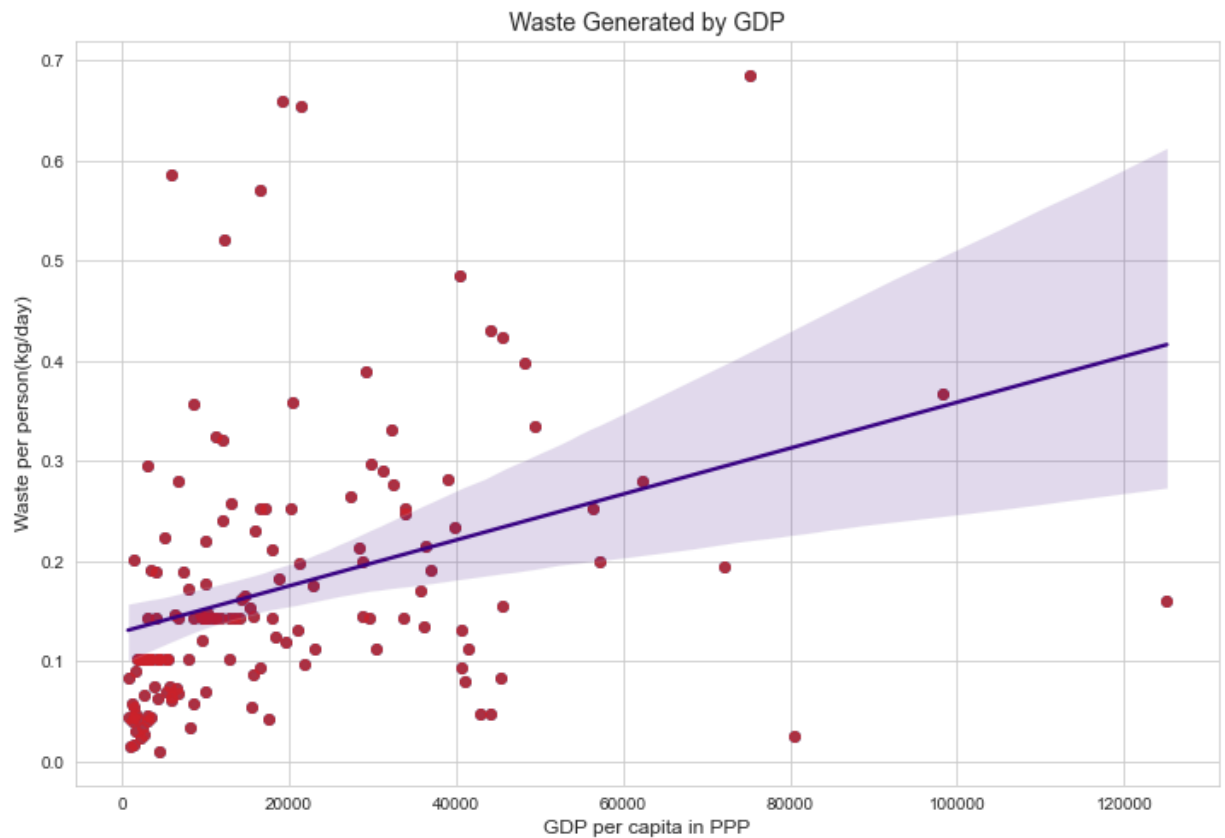


In [38]: `'''scatter plot'''`

```
plt.figure(2, figsize=(12,8))
plt.scatter(df_plastic_waste['GDP per capita in PPP'], df_plastic_waste['Waste per person(kg/day)'])
plt.title('Waste Generated by GDP', loc='center', fontsize=14)
plt.ylabel('Waste per person(kg/day)',loc='center',fontsize=12)
plt.xlabel('GDP per capita in PPP',fontsize=12)

sns.regplot(x=df_plastic_waste['GDP per capita in PPP'], y=df_plastic_waste['Waste per person(kg/day)'],
            scatter_kws={'color':'#CD212A'}, line_kws={'color': '#380282'})

plt.show()
```



In [ ]:

