Author - Mayank Singh

Data Analyst|MySQL|Python|Web Crawling and Scraping |EDA|Visualization|CBAP|Agile Cert

Creation Date - 09-10-2022

Importing Liabrary

```
In [1]: import pandas as pd
```

Reading File

```
In [2]: file = pd.read_csv("C:\\Users\\mayank.singh27\\PycharmProjects\\pythonProject\\TestCode\\AAP_Donation_Data_fy20-21.csv"
```

```
In [3]: file.head()
```

Out[3]:

	Name	Address	Pan	Amount ?	Contribution Mode	Financial
0	1.? Bharatha Swamukti Samsthe	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	Rs. 1,20,00,000.00 �1�Cror+	Cheque State Bank of India Mahalakshmi Layout ...	2020-21
1	NaN	NaN	NaN	NaN	NaN	NaN
2	2.? Prudent Electoral Trust	G-15 Hans Bhawan 1 Bahadur Shah Zafar Marg New...	Y	Rs. 1,00,00,000.00 �1�Cror+	Cheque HDFC Bank Kailash Building K.G. Marg Ne...	2020-21
3	NaN	NaN	NaN	NaN	NaN	NaN
4	3.? Bharatha Swamukti Samsthe	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	Rs. 99,00,000.00 �99�Lacs+	Cheque State Bank of India Mahalakshmi Layout ...	2020-21

Cleaning Data

```
In [4]: file.dropna(inplace = True) # droping blank/Null value columns
        file['Name'] = file['Name'].str.replace('\d+','') # replacing digits form Name Column
        file['Name'] = file['Name'].str.replace('.','') # replacing . from Name column
        file['Name'] = file['Name'].str.replace('?','') # replacing special char ? from Name Column
        file['Amount ?'] = file['Amount ?'].str.replace('�','') # replacing special char � from Name Column
        file['Amount ?'] = file['Amount ?'].str.replace('Rs','')  # replacing Rs from Amount column
        file['Amount ?'] = file['Amount ?'].str.replace(',','')
        file['Amount ?'] = file['Amount ?'].str.lstrip('.') # replacing . from Amount column
        file['Amount ?'] = file['Amount ?'].str.rstrip('+') # replacing + from Amount column
        file['Amount ?'] = file['Amount ?'].str.rstrip('cror') # replacing cror from Amount column
        file['Amount ?'] = file['Amount ?'].str.rstrip('C') # replacing C from Amount column
        file['Amount ?'] = file['Amount ?'].str.rstrip('Lacs') # replacing Lacs from Amount column
        file['Amount ?'] = file['Amount ?'].str.split().str[0] # replacing unwanted digits after space from Amount column
```

```
C:\Users\mayank.singh27\AppData\Local\Temp\ipykernel_29944\2697003470.py:2: FutureWarning: The default value of regex w
ill change from True to False in a future version.
  file['Name'] = file['Name'].str.replace('\d+','') # replacing digits form Name Column
C:\Users\mayank.singh27\AppData\Local\Temp\ipykernel_29944\2697003470.py:3: FutureWarning: The default value of regex w
ill change from True to False in a future version. In addition, single character regular expressions will *not* be trea
ted as literal strings when regex=True.
  file['Name'] = file['Name'].str.replace('.','') # replacing . from Name column
C:\Users\mayank.singh27\AppData\Local\Temp\ipykernel_29944\2697003470.py:4: FutureWarning: The default value of regex w
ill change from True to False in a future version. In addition, single character regular expressions will *not* be trea
ted as literal strings when regex=True.
  file['Name'] = file['Name'].str.replace('?','') # replacing special char ? from Name Column
```

Cleaned data Set

```
In [5]: file.head()
```

Out[5]:

	Name	Address	Pan	Amount ?	Contribution Mode	Financial
0	Bharatha Swamukti Samsthe	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	12000000.00	Cheque State Bank of India Mahalakshmi Layout ...	2020-21
2	Prudent Electoral Trust	G-15 Hans Bhawan 1 Bahadur Shah Zafar Marg New...	Y	10000000.00	Cheque HDFC Bank Kailash Building K.G. Marg Ne...	2020-21
4	Bharatha Swamukti Samsthe	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	9900000.00	Cheque State Bank of India Mahalakshmi Layout ...	2020-21
6	Bharatha Swamukti Samsthe	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	9500000.00	Cheque State Bank of India Mahalakshmi Layout ...	2020-21
8	Bharatha Swamukti Samsthe	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	9100000.00	Cheque State Bank of India Mahalakshmi Layout ...	2020-21

Cleaning Headers - 'Amount ?'

In [6]:
```python
file.columns = file.columns.str.replace ('?', '')
```

```
C:\Users\mayank.singh27\AppData\Local\Temp\ipykernel_29944\434693188.py:1: FutureWarning: The default value of regex wi
ll change from True to False in a future version. In addition, single character regular expressions will *not* be treat
ed as literal strings when regex=True.
  file.columns = file.columns.str.replace ('?', '')
```

Removing extra spaces before and after the columns' Data

In [7]:
```python
file['Name'] = file['Name'].str.strip()
file['Address'] = file['Address'].str.strip()
file['Pan'] = file['Pan'].str.strip()
file['Amount '] = file['Amount '].str.strip()
file['Contribution Mode'] = file['Contribution Mode'].str.strip()
file['Financial'] = file['Financial'].str.strip()
file.columns = file.columns.str.strip()
```

Final Cleaned Data Set

```
In [8]: file
```

Out[8]:

	Name	Address	Pan	Amount	Contribution Mode	Financial
0	Bharatha Swamukti Samsthe	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	12000000.00	Cheque State Bank of India Mahalakshmi Layout ...	2020-21
2	Prudent Electoral Trust	G-15 Hans Bhawan 1 Bahadur Shah Zafar Marg New...	Y	10000000.00	Cheque HDFC Bank Kailash Building K.G. Marg Ne...	2020-21
4	Bharatha Swamukti Samsthe	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	9900000.00	Cheque State Bank of India Mahalakshmi Layout ...	2020-21
6	Bharatha Swamukti Samsthe	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	9500000.00	Cheque State Bank of India Mahalakshmi Layout ...	2020-21
8	Bharatha Swamukti Samsthe	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	9100000.00	Cheque State Bank of India Mahalakshmi Layout ...	2020-21
...	...	...	...	...	...	...
3438	Ramesh Bellamkonda	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	10.00	Cash	2020-21
3440	Ramesh Bellamkonda	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	10.00	Cash	2020-21
3442	Ramesh Bellamkonda	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	10.00	Cash	2020-21
3444	Ramesh Bellamkonda	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	10.00	Cash	2020-21
3446	Ramesh Bellamkonda	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	10.00	Cash	2020-21

1724 rows × 6 columns

Saving to New file And Reading New Data Set

```
In [9]: # file.to_csv("C:\\Users\\mayank.singh27\\PycharmProjects\\pythonProject\\TestCode\\AAP_Donation_Data_fy20-21_Cleaned_Ne
```

```
In [10]: df = pd.read_csv("C:\\Users\\mayank.singh27\\PycharmProjects\\pythonProject\\TestCode\\AAP_Donation_Data_fy20-21_Cleane
```

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

Out[11]:

	Unnamed: 0	Name	Address	Pan	Amount	Contribution Mode	Financial
0	0	Bharatha Swamukti Samsthe	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	12000000.0	Cheque State Bank of India Mahalakshmi Layout ...	2020-21
1	2	Prudent Electoral Trust	G-15 Hans Bhawan 1 Bahadur Shah Zafar Marg New...	Y	10000000.0	Cheque HDFC Bank Kailash Building K.G. Marg Ne...	2020-21
2	4	Bharatha Swamukti Samsthe	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	9900000.0	Cheque State Bank of India Mahalakshmi Layout ...	2020-21
3	6	Bharatha Swamukti Samsthe	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	9500000.0	Cheque State Bank of India Mahalakshmi Layout ...	2020-21
4	8	Bharatha Swamukti Samsthe	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	9100000.0	Cheque State Bank of India Mahalakshmi Layout ...	2020-21

Getting Started with EDA

```
In [12]:  # 1. Basic Info about Data
          df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1724 entries, 0 to 1723
Data columns (total 7 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Unnamed: 0         1724 non-null   int64
 1   Name               1724 non-null   object
 2   Address            1724 non-null   object
 3   Pan                1724 non-null   object
 4   Amount             1724 non-null   float64
 5   Contribution Mode  1724 non-null   object
 6   Financial          1724 non-null   object
dtypes: float64(1), int64(1), object(5)
memory usage: 94.4+ KB
```

info() provides us basic overview of the data:

1. Total columns = 7

2. There is no - null values i.e. we have values available in every column

3. Data type of each coulmn, one (Amount) is in float, one is in int (Unnamed 0) and all others are object/string

4. Total entries = 1724

In [13]: `## with .info() and .head(), we can see that there is a non useful column --Unnamed 0--`

In [14]: `df = df.drop('Unnamed: 0', axis = 1) # droping the non useful column completely`

In [15]: `df.head()`

Out[15]:

	Name	Address	Pan	Amount	Contribution Mode	Financial
0	Bharatha Swamukti Samsthe	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	12000000.0	Cheque State Bank of India Mahalakshmi Layout ...	2020-21
1	Prudent Electoral Trust	G-15 Hans Bhawan 1 Bahadur Shah Zafar Marg New...	Y	10000000.0	Cheque HDFC Bank Kailash Building K.G. Marg Ne...	2020-21
2	Bharatha Swamukti Samsthe	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	9900000.0	Cheque State Bank of India Mahalakshmi Layout ...	2020-21
3	Bharatha Swamukti Samsthe	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	9500000.0	Cheque State Bank of India Mahalakshmi Layout ...	2020-21
4	Bharatha Swamukti Samsthe	108 Sarang of Unitech Parkway Apts 1st Main Ro...	Y	9100000.0	Cheque State Bank of India Mahalakshmi Layout ...	2020-21

In [16]: `Toatl_Donation_Amount = df['Amount'].sum()`

In [17]: `Toatl_Donation_Amount`

Out[17]: `113287398.0`

AAP received a total of 11 cror 32 lacs 87 thousand 398 INR Donation in FY - 20-21

Converting different contribution_mode to uniform Payment_mode using keywords 'Bank, Cheque, Online, Cash'

```python
In [18]:   df['Payment_mode'] = df['Contribution Mode'].str.split().str[0]  #step 1 spliting the sentence into individual
                                                                             #word seperated by space and extrcating word at index [6
                                                                             #step 2 Replacing non uniform words with uniform keyword
                                                                             #from sentence
           df['Payment_mode'] = df['Payment_mode'].str.replace('Cheaue', 'Cheque')
           df['Payment_mode'] = df['Payment_mode'].str.replace('Cheuqe', 'Cheque')
           df['Payment_mode'] = df['Payment_mode'].str.replace('HDFC', 'Bank')
           df['Payment_mode'] = df['Payment_mode'].str.replace('J', 'Bank')
           df['Payment_mode'] = df['Payment_mode'].str.replace('SVC', 'Bank')
           df['Payment_mode'] = df['Payment_mode'].str.replace('State', 'Bank')
```

```python
In [19]:   Amt_by_diff_modes = df.groupby('Payment_mode').sum()
```

```python
In [20]:   Amt_by_diff_modes
```

Out[20]:

	Amount
Payment_mode	
Bank	2135000.0
Cash	110.0
Cheque	94853272.0
Online	16299016.0

Total amount transferred/contributed by different payment modes. Highest donation was given via ##Cheque i.e. approx = 9.5 crores

```python
In [21]:   PAN_NoPAN_contri = df.groupby('Pan').sum()
```

```python
In [22]:   PAN_NoPAN_contri
```

Out[22]:

	Amount
Pan	
N	305500.0
Y	112981898.0

Out of Total donation of appro 11.35 crores donation of INR 3,05,500.00 was given without declaring PAN number.

```
In [23]: Persons = df.groupby('Name').sum()
```

```
In [24]: Persons.sort_values(by = 'Amount',ascending=False )
```

Out[24]:

Name	Amount
Bharatha Swamukti Samsthe	51700000.0
Prudent Electoral Trust	17000000.0
Ramesh Bellamkonda	6400100.0
Nohar Chand & Sons Steels (P) Ltd	5100000.0
Kama Realty (Delhi) Limited	5000000.0
...	...
Dr Ashok Kumar K Lodha	7000.0
Prabhat Kumar Sharma	6000.0
Ramanujam Krishnakumar	5000.0
M Shihabudheen	1500.0
Shantala	10.0

291 rows × 1 columns

Highest Donation was given by "Bharatha Swamukti Samsthe" - which is INR - 5,17,00,000, i.e. more than 5 crores similarly we can see top 3 donators after Bharatha Swamukti Samsthe are:

1. Prudent Electoral Trust
2. Ramesh Bellamkonda
3. Nohar Chand & Sons Steels (P) Ltd

```
In [25]: Person_mode = df.groupby(['Name', 'Payment_mode']).sum()
```

```
In [26]: Person_mode.sort_values(by = 'Amount',ascending=False )
```

		Amount
Name	**Payment_mode**	
Bharatha Swamukti Samsthe	**Cheque**	51700000.0
Prudent Electoral Trust	**Cheque**	15000000.0
Ramesh Bellamkonda	**Cheque**	6400000.0
Nohar Chand & Sons Steels (P) Ltd	**Cheque**	5100000.0
Kama Realty (Delhi) Limited	**Cheque**	5000000.0
...	...	...
Prithvi Reddy	**Online**	2021.0
M Shihabudheen	**Online**	1500.0
Ramesh Bellamkonda	**Cash**	100.0
Prithvi Reddy	**Cash**	10.0
Shantala	**Online**	10.0

311 rows × 1 columns

All the Top Donators had contributed via Cheque

```
Person_mode
```

Out[27]:

Name	Payment_mode	Amount
Aakash Jain	Online	44800.0
Aakash Sharma	Cheque	15000.0
	Online	5800.0
Aareen Healthcare Private Limited	Cheque	200000.0
Abbas Kaiyum Sutarwala	Online	23580.0
...	...	...
Vikram Fernandes	Online	35000.0
Vinodkumar A Pillai	Online	86000.0
Yogeshbhai Valjibhai Makwana	Cheque	21000.0
	Online	6900.0
Zaheer Khan	Online	26942.0

311 rows × 1 columns

There are people who have donated via more than one mode.

In [ ]: