Experiment 2:
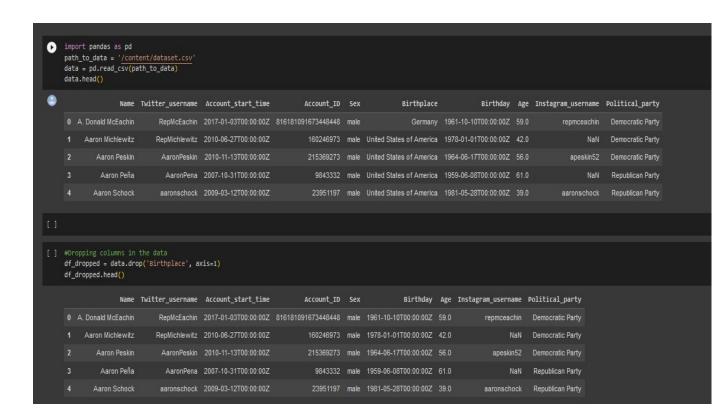
Manipulation of twitter dataset

AIM:

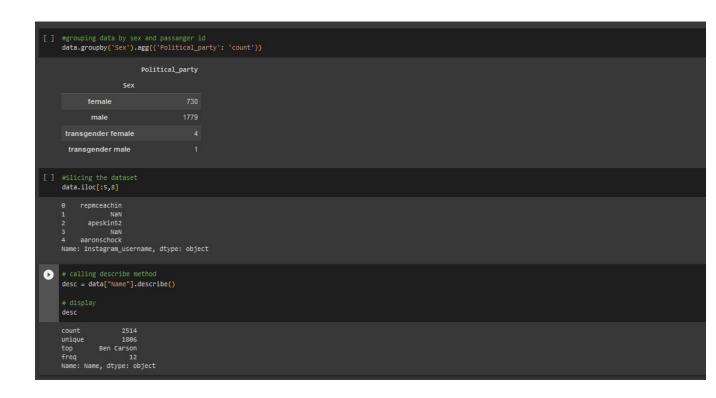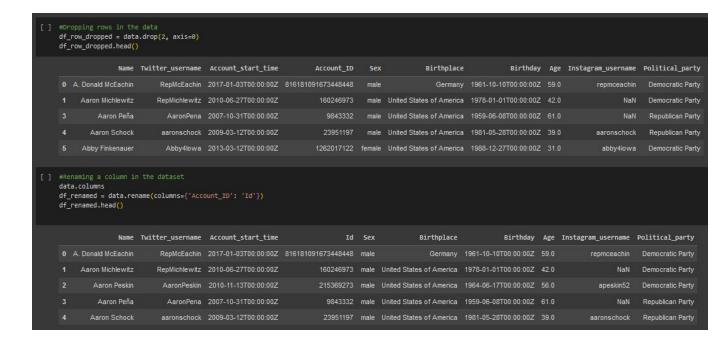   To manipulate a twitter dataset using python.

Algorithm:

    1. Import the required modules, dataset and load the dataset.

    2. Then drop a certain column from the dataset.

    3. Drop a row fro the dataset.

    4. Then rename a column from the dataset.

    5. Then group a data under a certain features.

    6. Slicing the dataset to get required information.

    7. Describe the dataset and print it.

Program:

```python
import pandas as pd
path_to_data = '/content/dataset.csv'
data = pd.read_csv(path_to_data)
data.head()
```

	Name	Twitter_username	Account_start_time	Account_ID	Sex	Birthplace	Birthday	Age	Instagram_username	Political_party
0	A. Donald McEachin	RepMcEachin	2017-01-03T00:00:00Z	816181091673448448	male	Germany	1961-10-10T00:00:00Z	59.0	repmceachin	Democratic Party
1	Aaron Michlewitz	RepMichlewitz	2010-06-27T00:00:00Z	160246973	male	United States of America	1978-01-01T00:00:00Z	42.0	NaN	Democratic Party
2	Aaron Peskin	AaronPeskin	2010-11-13T00:00:00Z	215369273	male	United States of America	1964-06-17T00:00:00Z	56.0	apeskin52	Democratic Party
3	Aaron Peña	AaronPena	2007-10-31T00:00:00Z	9843332	male	United States of America	1959-06-08T00:00:00Z	61.0	NaN	Republican Party
4	Aaron Schock	aaronschock	2009-03-12T00:00:00Z	23951197	male	United States of America	1981-05-28T00:00:00Z	39.0	aaronschock	Republican Party

```python
#Dropping columns in the data
df_dropped = data.drop('Birthplace', axis=1)
df_dropped.head()
```

	Name	Twitter_username	Account_start_time	Account_ID	Sex	Birthday	Age	Instagram_username	Political_party
0	A. Donald McEachin	RepMcEachin	2017-01-03T00:00:00Z	816181091673448448	male	1961-10-10T00:00:00Z	59.0	repmceachin	Democratic Party
1	Aaron Michlewitz	RepMichlewitz	2010-06-27T00:00:00Z	160246973	male	1978-01-01T00:00:00Z	42.0	NaN	Democratic Party
2	Aaron Peskin	AaronPeskin	2010-11-13T00:00:00Z	215369273	male	1964-06-17T00:00:00Z	56.0	apeskin52	Democratic Party
3	Aaron Peña	AaronPena	2007-10-31T00:00:00Z	9843332	male	1959-06-08T00:00:00Z	61.0	NaN	Republican Party
4	Aaron Schock	aaronschock	2009-03-12T00:00:00Z	23951197	male	1981-05-28T00:00:00Z	39.0	aaronschock	Republican Party

```
[ ] #grouping data by sex and passanger id
    data.groupby('Sex').agg({'Political_party': 'count'})
```

	Political_party
Sex	
female	730
male	1779
transgender female	4
transgender male	1

```
[ ] #Slicing the dataset
    data.iloc[:5,8]
```

```
0    repmceachin
1            NaN
2      apeskin52
3            NaN
4    aaronschock
Name: Instagram_username, dtype: object
```

```
# calling describe method
desc = data["Name"].describe()

# display
desc
```

```
count          2514
unique         1806
top      Ben Carson
freq             12
Name: Name, dtype: object
```

```
[ ] #Dropping rows in the data
    df_row_dropped = data.drop(2, axis=0)
    df_row_dropped.head()
```

	Name	Twitter_username	Account_start_time	Account_ID	Sex	Birthplace	Birthday	Age	Instagram_username	Political_party
0	A. Donald McEachin	RepMcEachin	2017-01-03T00:00:00Z	816181091673448448	male	Germany	1961-10-10T00:00:00Z	59.0	repmceachin	Democratic Party
1	Aaron Michlewitz	RepMichlewitz	2010-06-27T00:00:00Z	160246973	male	United States of America	1978-01-01T00:00:00Z	42.0	NaN	Democratic Party
3	Aaron Peña	AaronPena	2007-10-31T00:00:00Z	9843332	male	United States of America	1959-06-08T00:00:00Z	61.0	NaN	Republican Party
4	Aaron Schock	aaronschock	2009-03-12T00:00:00Z	23951197	male	United States of America	1981-05-28T00:00:00Z	39.0	aaronschock	Republican Party
5	Abby Finkenauer	Abby4Iowa	2013-03-12T00:00:00Z	1262017122	female	United States of America	1988-12-27T00:00:00Z	31.0	abby4iowa	Democratic Party

```
[ ] #Renaming a column in the dataset
    data.columns
    df_renamed = data.rename(columns={'Account_ID': 'Id'})
    df_renamed.head()
```

	Name	Twitter_username	Account_start_time	Id	Sex	Birthplace	Birthday	Age	Instagram_username	Political_party
0	A. Donald McEachin	RepMcEachin	2017-01-03T00:00:00Z	816181091673448448	male	Germany	1961-10-10T00:00:00Z	59.0	repmceachin	Democratic Party
1	Aaron Michlewitz	RepMichlewitz	2010-06-27T00:00:00Z	160246973	male	United States of America	1978-01-01T00:00:00Z	42.0	NaN	Democratic Party
2	Aaron Peskin	AaronPeskin	2010-11-13T00:00:00Z	215369273	male	United States of America	1964-06-17T00:00:00Z	56.0	apeskin52	Democratic Party
3	Aaron Peña	AaronPena	2007-10-31T00:00:00Z	9843332	male	United States of America	1959-06-08T00:00:00Z	61.0	NaN	Republican Party
4	Aaron Schock	aaronschock	2009-03-12T00:00:00Z	23951197	male	United States of America	1981-05-28T00:00:00Z	39.0	aaronschock	Republican Party

Result:

Thus the twitter dataset has been manipulated and result is succesfully executed.