Smaira Pandita

23112314

Loading the dataset as a pandas DataFrame object

In [132…
```
import pandas as pd
import plotly.express as px
import matplotlib.pyplot as plt
import numpy as np
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import LogisticRegression
import warnings as w
w.filterwarnings('ignore')
```

In [133…
```
df=pd.read_csv("election-data-states.csv")
```

In [134…
```
df
```

Out[134]:

	st_name	year	ac_no	ac_name	ac_type	cand_name	cand_sex	partyname	partyabb
0	Andhra Pradesh	1978.0	1	Ichapuram	GEN	Manabala Ramarao	M	Independent	IN
1	Andhra Pradesh	1978.0	1	Ichapuram	GEN	Appadu Sahu	M	Independent	IN
2	Andhra Pradesh	1978.0	1	Ichapuram	GEN	Uppada Rangababu	M	Indian National Congress	IN
3	Andhra Pradesh	1978.0	1	Ichapuram	GEN	Kalla Balarama Swamy	M	Indian National Congress (I)	INC
4	Andhra Pradesh	1978.0	1	Ichapuram	GEN	Bendalam Venkatesam Sarma	M	Janata Party	JN
...	...	...	...	...	...	...	...	...	
327289	West Bengal	2011.0	294	Murarai	GEN	Ajrail Md	M	IND	IN
327290	West Bengal	2011.0	294	Murarai	GEN	Shiblal Sardar	M	BSP	B
327291	West Bengal	2011.0	294	Murarai	GEN	Sudhiranjan Das Goswami	M	BJP	B
327292	West Bengal	2011.0	294	Murarai	GEN	Elahi Kamre Dr	M	CPM	CP
327293	West Bengal	2011.0	294	Murarai	GEN	Nure Alam Chowdhury	M	AITC	AI

327294 rows × 11 columns

In [135…

```
df['st_name']
```

Out[135]:

```
0         Andhra Pradesh
1         Andhra Pradesh
2         Andhra Pradesh
3         Andhra Pradesh
4         Andhra Pradesh
              ...
327289       West Bengal
327290       West Bengal
327291       West Bengal
327292       West Bengal
327293       West Bengal
Name: st_name, Length: 327294, dtype: object
```

Removing the unwanted columns like - ac_no, ac_name and partyabbre

In [136…
```
df=df.drop(columns=['ac_no','ac_name','partyabbre'])
```

In [137…
```
df
```

Out[137]:

	st_name	year	ac_type	cand_name	cand_sex	partyname	totvotpoll	electors
0	Andhra Pradesh	1978.0	GEN	Manabala Ramarao	M	Independent	813.0	83247
1	Andhra Pradesh	1978.0	GEN	Appadu Sahu	M	Independent	1743.0	83247
2	Andhra Pradesh	1978.0	GEN	Uppada Rangababu	M	Indian National Congress	4427.0	83247
3	Andhra Pradesh	1978.0	GEN	Kalla Balarama Swamy	M	Indian National Congress (I)	19805.0	83247
4	Andhra Pradesh	1978.0	GEN	Bendalam Venkatesam Sarma	M	Janata Party	34251.0	83247
...	...	...	...	...	...	...	...	...
327289	West Bengal	2011.0	GEN	Ajrail Md	M	IND	2471.0	189320
327290	West Bengal	2011.0	GEN	Shiblal Sardar	M	BSP	3074.0	189320
327291	West Bengal	2011.0	GEN	Sudhiranjan Das Goswami	M	BJP	4961.0	189320
327292	West Bengal	2011.0	GEN	Elahi Kamre Dr	M	CPM	73414.0	189320
327293	West Bengal	2011.0	GEN	Nure Alam Chowdhury	M	AITC	77817.0	189320

327294 rows × 8 columns

Renaming for better undertanding of data

In [138…
```
df=df.rename(columns={'st_name':'state','cand_name':'candidate','totvotpoll':'votes
```

In [139…
```
df
```

Out[139]:

	state	year	ac_type	candidate	cand_sex	partyname	votes	electors
0	Andhra Pradesh	1978.0	GEN	Manabala Ramarao	M	Independent	813.0	83247
1	Andhra Pradesh	1978.0	GEN	Appadu Sahu	M	Independent	1743.0	83247
2	Andhra Pradesh	1978.0	GEN	Uppada Rangababu	M	Indian National Congress	4427.0	83247
3	Andhra Pradesh	1978.0	GEN	Kalla Balarama Swamy	M	Indian National Congress (I)	19805.0	83247
4	Andhra Pradesh	1978.0	GEN	Bendalam Venkatesam Sarma	M	Janata Party	34251.0	83247
...	...	...	...	...	...	...	...	...
327289	West Bengal	2011.0	GEN	Ajrail Md	M	IND	2471.0	189320
327290	West Bengal	2011.0	GEN	Shiblal Sardar	M	BSP	3074.0	189320
327291	West Bengal	2011.0	GEN	Sudhiranjan Das Goswami	M	BJP	4961.0	189320
327292	West Bengal	2011.0	GEN	Elahi Kamre Dr	M	CPM	73414.0	189320
327293	West Bengal	2011.0	GEN	Nure Alam Chowdhury	M	AITC	77817.0	189320

327294 rows × 8 columns

reducing the dataset to 5 states

In [140…
```
df_1=df[df['state']=='Jammu & Kashmir']
df_2=df[df['state']=='Uttar Pradesh']
df_3=df[df['state']=='Himachal Pradesh ']
df_4=df[df['state']=='Meghalaya ']
df_5=df[df['state']=='Puducherry']
```

In [141…
```
df=pd.concat([df_1,df_2,df_3,df_4,df_5])
```

In [142…
```
df
```

Out[142]:

	state	year	ac_type	candidate	cand_sex	partyname	votes	electors
100427	Jammu & Kashmir	1977.0	GEN	Mohd. Yassain	M	Independents	818.0	41360
100428	Jammu & Kashmir	1977.0	GEN	Ghulam Rasool Malik	M	Janata Party	5868.0	41360
100429	Jammu & Kashmir	1977.0	GEN	Mohd. Yunus	M	Indian National Congress	7641.0	41360
100430	Jammu & Kashmir	1977.0	GEN	Ghulam Qadir Mir	M	Jammu & Kashmir National Conference	12357.0	41360
100431	Jammu & Kashmir	1977.0	GEN	Ghulam Rasool War	M	Indian National Congress	571.0	40387
...	...	...	...	...	...	...	...	...
204706	Puducherry	2011.0	GEN	Mellam Srinivasa Rao	M	IND	134.0	30936
204707	Puducherry	2011.0	GEN	Kalla Venkata Ratnam	M	IND	153.0	30936
204708	Puducherry	2011.0	GEN	Nalla Naga Raju	M	IND	154.0	30936
204709	Puducherry	2011.0	GEN	Manchala Satya Sai Kumar	M	ADMK	4867.0	30936
204710	Puducherry	2011.0	GEN	Malladi Krishna Rao	M	INC	23985.0	30936

65371 rows × 8 columns

Dropping null values

In [143…

```
df.dropna()
```

Out[143]:

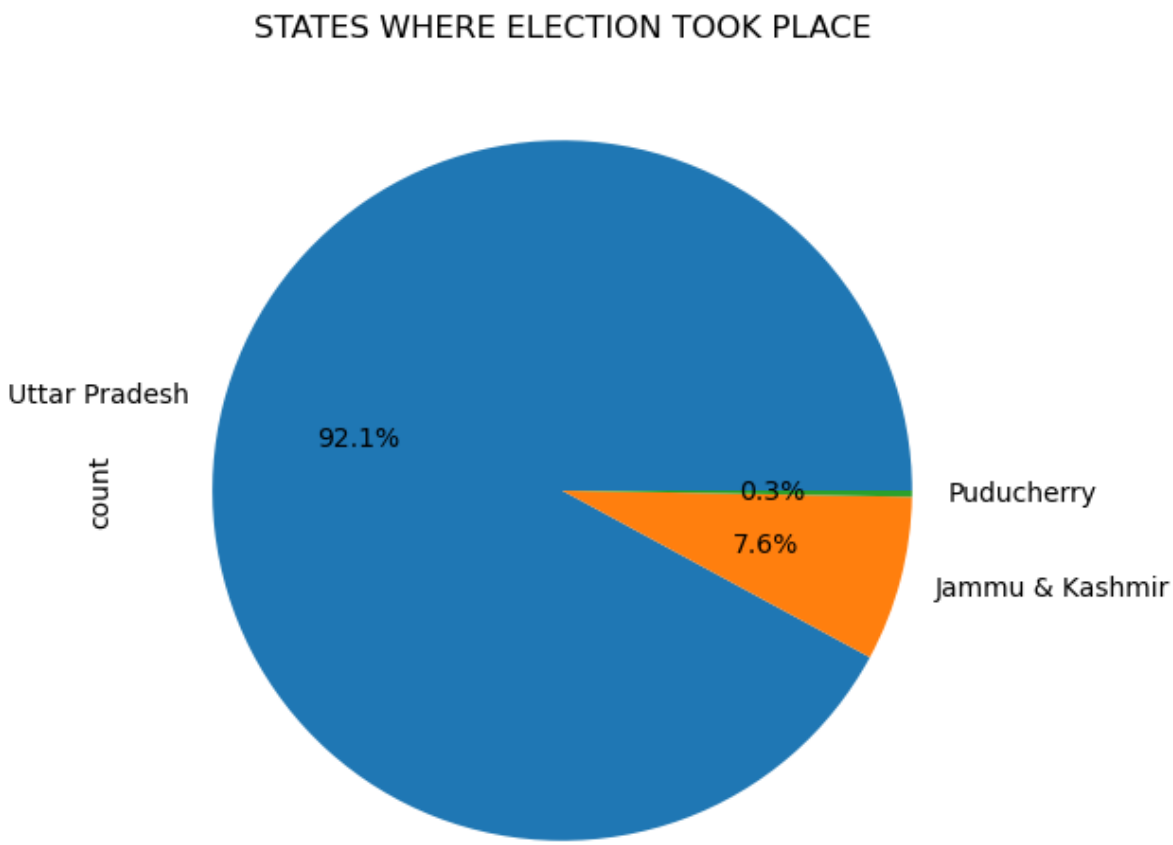
	state	year	ac_type	candidate	cand_sex	partyname	votes	electors
100427	Jammu & Kashmir	1977.0	GEN	Mohd. Yassain	M	Independents	818.0	41360
100428	Jammu & Kashmir	1977.0	GEN	Ghulam Rasool Malik	M	Janata Party	5868.0	41360
100429	Jammu & Kashmir	1977.0	GEN	Mohd. Yunus	M	Indian National Congress	7641.0	41360
100430	Jammu & Kashmir	1977.0	GEN	Ghulam Qadir Mir	M	Jammu & Kashmir National Conference	12357.0	41360
100431	Jammu & Kashmir	1977.0	GEN	Ghulam Rasool War	M	Indian National Congress	571.0	40387
...	...	...	...	...	...	...	...	...
204706	Puducherry	2011.0	GEN	Mellam Srinivasa Rao	M	IND	134.0	30936
204707	Puducherry	2011.0	GEN	Kalla Venkata Ratnam	M	IND	153.0	30936
204708	Puducherry	2011.0	GEN	Nalla Naga Raju	M	IND	154.0	30936
204709	Puducherry	2011.0	GEN	Manchala Satya Sai Kumar	M	ADMK	4867.0	30936
204710	Puducherry	2011.0	GEN	Malladi Krishna Rao	M	INC	23985.0	30936

65284 rows × 8 columns

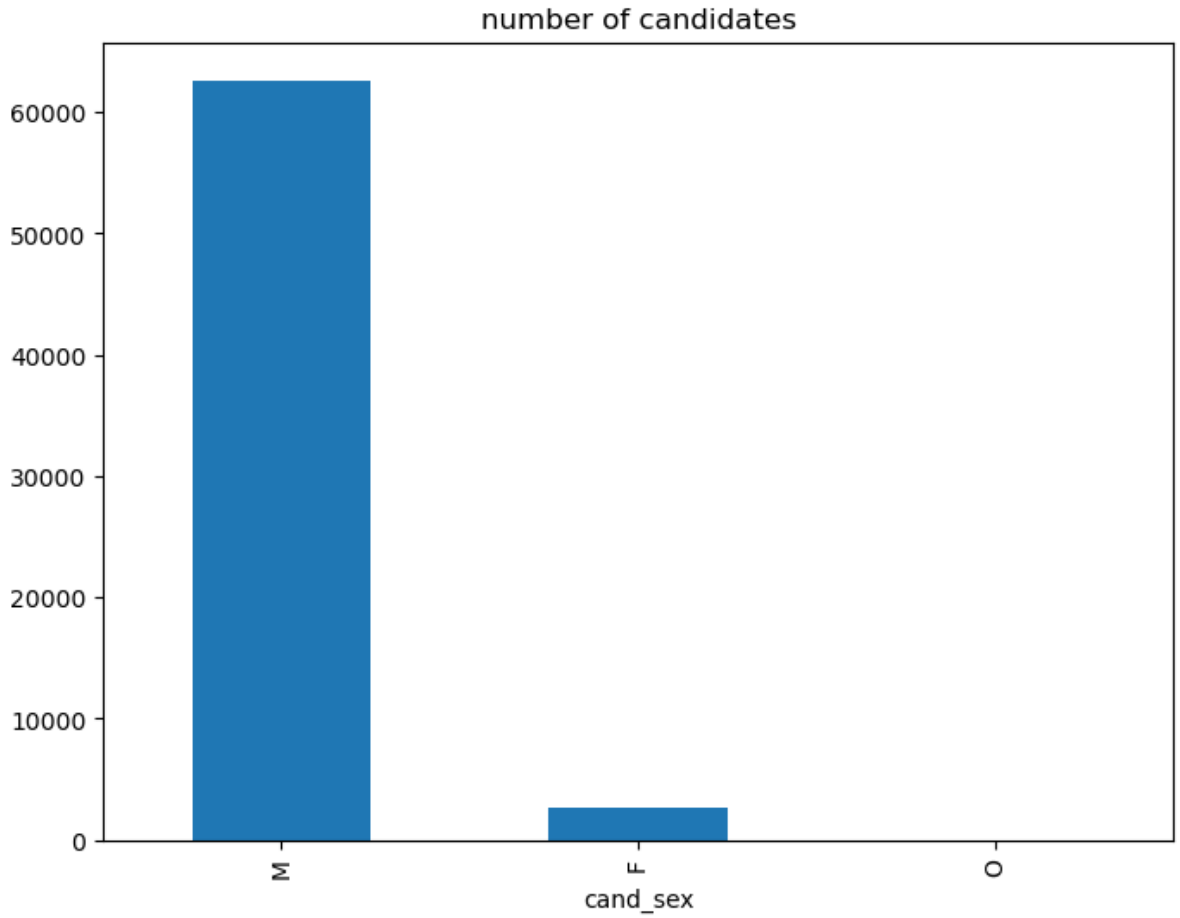
Data Visualisation

In [144…

```
plt.figure(figsize=(8,6))
df['state'].value_counts().plot(kind='pie',autopct=('%1.1f%%'))
plt.title('STATES WHERE ELECTION TOOK PLACE')
plt.show()
```

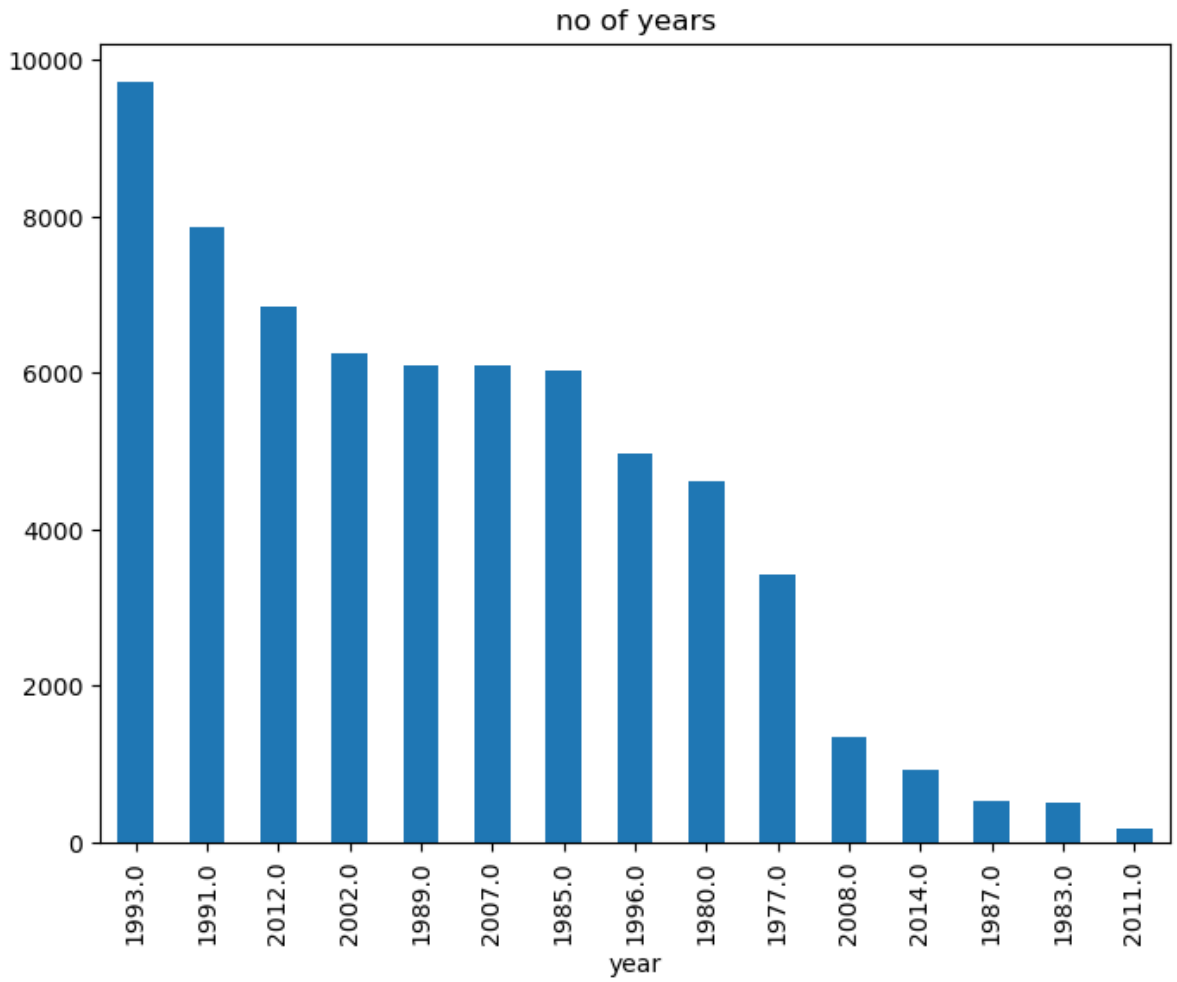


In [145…
```
plt.figure(figsize=(8,6))
df['cand_sex'].value_counts().plot(kind='bar')
plt.title('number of candidates')
plt.show()
```



In [146…

```
plt.figure(figsize=(8,6))
df['year'].value_counts().plot(kind='bar')
plt.title('no of years')
plt.show()
```



Label Encoding

In [147…
```
df['state'].value_counts()
```

Out[147]:
```
state
Uttar Pradesh      60206
Jammu & Kashmir     4978
Puducherry           187
Name: count, dtype: int64
```

In [148…
```
Label_encoder=LabelEncoder()
df['state_encoded']=Label_encoder.fit_transform(df['state'])
df[['state','state_encoded']]
```

Out[148]:

	state	state_encoded
100427	Jammu & Kashmir	0
100428	Jammu & Kashmir	0
100429	Jammu & Kashmir	0
100430	Jammu & Kashmir	0
100431	Jammu & Kashmir	0
...	...	...
204706	Puducherry	1
204707	Puducherry	1
204708	Puducherry	1
204709	Puducherry	1
204710	Puducherry	1

65371 rows × 2 columns

In [149…

```
Label_encoder=LabelEncoder()
df['year_encoded']=Label_encoder.fit_transform(df['year'])
df[['year','year_encoded']]
```

Out[149]:

	year	year_encoded
100427	1977.0	0
100428	1977.0	0
100429	1977.0	0
100430	1977.0	0
100431	1977.0	0
...	...	...
204706	2011.0	12
204707	2011.0	12
204708	2011.0	12
204709	2011.0	12
204710	2011.0	12

65371 rows × 2 columns

In [150…

```
Label_encoder=LabelEncoder()
df['partyname_encoded']=Label_encoder.fit_transform(df['partyname'])
df[['partyname','partyname_encoded']]
```

Out[150]:

	partyname	partyname_encoded
100427	Independents	239
100428	Janata Party	313
100429	Indian National Congress	250
100430	Jammu & Kashmir National Conference	300
100431	Indian National Congress	250
...	...	...
204706	IND	231
204707	IND	231
204708	IND	231
204709	ADMK	10
204710	INC	230

65371 rows × 2 columns

In [151…

```
Label_encoder=LabelEncoder()
df['votes_encoded']=Label_encoder.fit_transform(df['votes'])
df[['votes','votes_encoded']]
```

Out[151]:

	votes	votes_encoded
100427	818.0	818
100428	5868.0	4899
100429	7641.0	5773
100430	12357.0	7584
100431	571.0	571
...	...	...
204706	134.0	134
204707	153.0	153
204708	154.0	154
204709	4867.0	4339
204710	23985.0	11388

65371 rows × 2 columns

In [186…

```
df
```

Out[186]:

	state	year	ac_type	candidate	cand_sex	partyname	votes	electors
100427	Jammu & Kashmir	1977.0	GEN	Mohd. Yassain	M	Independents	818.0	41360
100428	Jammu & Kashmir	1977.0	GEN	Ghulam Rasool Malik	M	Janata Party	5868.0	41360
100429	Jammu & Kashmir	1977.0	GEN	Mohd. Yunus	M	Indian National Congress	7641.0	41360
100430	Jammu & Kashmir	1977.0	GEN	Ghulam Qadir Mir	M	Jammu & Kashmir National Conference	12357.0	41360
100431	Jammu & Kashmir	1977.0	GEN	Ghulam Rasool War	M	Indian National Congress	571.0	40387
...	...	...	...	...	...	...	...	...
204706	Puducherry	2011.0	GEN	Mellam Srinivasa Rao	M	IND	134.0	30936
204707	Puducherry	2011.0	GEN	Kalla Venkata Ratnam	M	IND	153.0	30936
204708	Puducherry	2011.0	GEN	Nalla Naga Raju	M	IND	154.0	30936
204709	Puducherry	2011.0	GEN	Manchala Satya Sai Kumar	M	ADMK	4867.0	30936
204710	Puducherry	2011.0	GEN	Malladi Krishna Rao	M	INC	23985.0	30936

65371 rows × 8 columns

Logistic Regression

In [190…

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report


label_encoder = LabelEncoder()
df['state_encoded'] = label_encoder.fit_transform(df['state'])
df['partyname_encoded'] = label_encoder.fit_transform(df['partyname'])
df['cand_sex_encoded'] = label_encoder.fit_transform(df['cand_sex'])

X = df[['state_encoded', 'year', 'cand_sex_encoded', 'partyname_encoded', 'votes']]
y = (df['votes'] > df['votes'].mean()).astype(int)


X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_sta

model = LogisticRegression()
model.fit(X_train, y_train)

y_pred = model.predict(X_test)

accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)
print(classification_report(y_test, y_pred))
```

```
Accuracy: 0.9997960432388334
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     15207
           1       1.00      1.00      1.00      4405

    accuracy                           1.00     19612
   macro avg       1.00      1.00      1.00     19612
weighted avg       1.00      1.00      1.00     19612
```

In [ ]:

In [ ]: