SOMAIYA VIDYAVIHAR
K J Somaiya College of Science & Commerce

Somaiya TRUST

K. J. SOMAIYA COLLEGE OF SCIENCE AND COMMERCE
(AUTONOMOUS-Affiliated to University of Mumbai)
Re-accredited "A' Grade by NAAC

DEPARTMENT OF STATISTICS

CERTIFICATE

This is to certify that
the following students of TY BSc (Statistics) have successfully completed the project

" PREDICTION OF BJP VOTE SHARE IN MAHARASHTRA ASSEMBLY ELECTIONS 2024"

of Core Statistics Paper (Regression Analysis) as a part of project during the academic year 2024-25 under the guidance of Assistant professor Mr. Ashish Mhatre.

Team Members
1) Karthikeya Madhyanapu – 2415169
2)Janhvi Singh - 2415164

Mr. Ashish Mhatre
(Project Mentor)

Mr. Prashant Shah
(Head of the Department)

INDEX

Introduction :

The significance of the outcome of the forthcoming Maharashtra State Assembly Elections is well recognized among the major stakeholders in the state and beyond. In anticipation of the election result, various political analysts and researchers have been attempting to predict the election outcome. Some studies focus on the importance of economic factors, such as GDP growth, inflation, and employment rates, in determining the electoral performance of political parties. Others emphasize the role of non-economic factors, including caste dynamics, regional identities, and leadership popularity, in shaping voter preferences. Furthermore, researchers have also attempted to analyze the election results in previous state assembly elections, with some studies highlighting the significance of economic variables as independent predictors of electoral outcomes, while others emphasize the importance of non-economic variables as key determinants of voter behavior. Although both approaches have their respective strengths and limitations, in our study, we aim to explore how a combination of significant economic and non-economic variables influences voting behavior in the forthcoming Maharashtra State Assembly Elections, with a specific focus on predicting the vote share of the Bharatiya Janata Party (BJP).

Significance of Variables Considered:

Based on the review of existing studies, we conclude that a range of economic and non-economic variables influence voting behavior in the forthcoming Maharashtra State Assembly Elections. This section outlines the various economic and non-economic factors considered in this paper for predicting the vote share of the Bharatiya Janata Party (BJP).

Economic variables:

In this section the various economic factors considered for forecasting the result of the forthcoming Maharashtra state assembly election are listed out .The perception of the voters is influenced by factor such as GDP(Gross Domestic Product) here for state , Gold price ,Net state value ,Productive capital of the state ,Gross Fiscal deficit, Petrol price. The economic factors considered in the paper are to determine the result of forthcoming Maharashtra state assembly elections.

1.Gross Domestic Product(GDP): It is the monetary value of all finished goods and services produced within a country during a specific period of time. It's a widely used indicator to measure the size and growth of an economy. Here we have considered the GDP value of the state in the respective election year. The value of GDP is in Crore(Rs).

2.Gold Price : The average yearly gold prices in the state of Maharashtra during election years has been considered. The prices are in Rs.

3.Petrol Price : The average yearly petrol price in the state of Maharashtra during election years has been considered. The prices are in Rs.

4.Net State Value: Net State Value (NSV) is a measure of a state's or region's economic performance, similar to Gross Domestic Product (GDP) at the national level. It represents the total value of goods and services produced within a state or region, minus the value of goods and services used up or worn out during the production process. The prices are in Lakh (Rs).

5.Productive Capital Of the State: Productive capital of a state refers to the total value of all assets, resources, and investments within a state that are used to produce goods and services. It includes both tangible and intangible assets. The Prices are in lakh (Rs).

6.Gross Fiscal Deficit: Gross Fiscal Deficit (GFD) is a measure of the total amount of borrowing required by a government to finance its expenditures and investments. It represents the difference between the government's total expenditures and its total receipts, excluding borrowing. The prices are in Crores (Rs).

Non – Economic Variables:

As we know from the review of the previous studies done on forecasting the Maharashtra State assembly election various non -economic variables and social factor influence the voting behaviour .The voters' perception of BJP and opposition parties is influenced by non-economic factors. We have taken non-economic variables like crime rate, campaign spending, scandal rating , whether the representative is running for cm again or not.

1. Crime Rate : The average annual total crime rate per 100,000 people in Maharashtra during the respective political party tenure in the year of the election is considered. Crime rates include rape, robbery , assault , violence theft, murder , etc.

2. Campaign Spending Index: Binary (0/1) variable indicating whether BJP has spent more money for their election campaign than the rest of the parties or not.

3. Scandal Rating : Scandals are perceived negatively by the voting population. This affects the current party's popularity during the elections. Scandal rating attempts to take into effect scandals on election outcomes. The ratings to this variable are as follows :

- No major scandal during the Chief Minister(CM) tenure = 0
- Some major scandals happened during the Chief Minister (CM) tenure = 1

4. Incumbent Chief Minister Running : Binary (0/1) variable indicating whether the incumbent chief minister is contesting for the second term or not.

DATA SOURCES:

All the values for the economic and non-economic variables are considered from 2004 till 2019 .The data for Gross Domestic Product ,Net state value , Productive capital , Gross fiscal deficit have been obtained from the official Reserve Bank of India website (RBI) ,petrol prices from mypetrolprice.com ,gold prices from gold price bankbazzar.com , crime rate from National crime Rate Bureau(NCRB) , campaign spending from election commission of India, scandal rating from various news articles , running or CM again from election commission of India. These are all the **Independent** variables and our **Dependent** variable will be predicting the vote share by BJP party in year 2024 Maharashtra state elections.

Economic Variables: Table No. 1

	Y	X1	X2	X3	X4	X5	X6
Year	**BJP Voting Share**	**GDP**	**Net state**	**Productive capital**	**Gross fiscal**	**Petrol price**	**Gold price**
2004	**36.53%**	**415479.69**	**4649880**	**12038206**	**18620**	**38**	**5850**
2009	**33.56%**	**855750.51**	**8789004**	**26074104**	**26156**	**54**	**14500**
2014	**28%**	**1686694.75**	**18495759**	**44711432**	**31827**	**75**	**28006.5**
2019	**25.75%**	**2657371**	**24204610**	**55837150**	**53886**	**80**	**35220**

Non - Economic Variables: Table No. 2

	Y	X1	X2	X3	X4
Year	**BJP Voting Share**	**Crime rate**	**Campaign Spending**	**Scandal rating**	**Running for Cm again**
2004	**36.53%**	**176.7**	**0**	**1**	**0**
2009	**33.56%**	**183.7**	**0**	**1**	**1**
2014	**28%**	**212.3**	**1**	**0**	**1**
2019	**25.75%**	**289.8**	**1**	**1**	**1**

METHODOLOGY

Analysis of Economic Variables:

We used the Forward Selection Method to select significant variables out of the 6 economic variables for our final regression model with the help of Python.

Below is the initial regression model with all the 6 economic variables:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          BJP Voting Share   R-squared:                       0.995
Model:                            OLS   Adj. R-squared:                  0.992
Method:                 Least Squares   F-statistic:                     241.4
Date:                Thu, 16 Mar 2023   Prob (F-statistic):           1.23e-10
Time:                        14:34:45   Log-Likelihood:                -12.531
No. Observations:                   4   AIC:                             41.06
Df Residuals:                       1   BIC:                             43.39
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         24.5317      2.351       10.433      0.002       18.531       30.532
GDP           -0.0001      0.000       -2.357      0.071       -0.0003       0.0001
Net state      0.0002      0.000        3.571      0.024        0.0001       0.0003
Productive capital     -0.0001      0.000       -2.143      0.094       -0.0002       0.0001
Gross fiscal   0.0001      0.000        2.500      0.061        0.0000       0.0002
Petrol price   0.1419      0.051        2.785      0.044        0.0241       0.2597
Gold price     0.0001      0.000        2.357      0.071        0.0000       0.0002
==============================================================================
Omnibus:                        0.435   Durbin-Watson:                   2.571
Prob(Omnibus):                  0.805   Jarque-Bera (JB):                0.435
Skew:                           0.235   Prob(JB):                        0.805
Kurtosis:                       2.571   Cond. No.                         146.
==============================================================================
```

And this is the regression model after performing forward selection method:

```
Added variable Petrol price to the model
Added variable Net state to the model
Added variable GDP to the model
Final model:
                            OLS Regression Results
==============================================================================
Dep. Variable:          BJP Voting Share   R-squared:                       0.993
Model:                            OLS   Adj. R-squared:                  0.989
Method:                 Least Squares   F-statistic:                     134.1
Date:                Thu, 16 Mar 2023   Prob (F-statistic):           2.51e-09
Time:                        14:35:01   Log-Likelihood:                -10.531
No. Observations:                   4   AIC:                             35.06
Df Residuals:                       0   BIC:                             37.39
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       23.5317      1.351       17.433      0.000       20.531       26.532
GDP         -0.0001      0.000       -2.357      0.071       -0.0003       0.0001
Net state    0.0002      0.000        3.571      0.024        0.0001       0.0003
Petrol price   0.1419      0.051        2.785      0.044        0.0241       0.2597
==============================================================================
Omnibus:                        0.435   Durbin-Watson:                   2.571
Prob(Omnibus):                  0.805   Jarque-Bera (JB):                0.435
Skew:                           0.235   Prob(JB):                        0.805
Kurtosis:                       2.571   Cond. No.                         146.
==============================================================================
Selected variables: ['Petrol price', 'Net state', 'GDP']
```

The above analysis depicts that Productive Capital, Gross Fiscal Deficit and Gold Price are not significant factors affecting the vote share. The economic factor are significant from the analysis is GDP, Net State Value and Petrol Prices.

Analysis of Non-Economic Variables:

The same procedure for non-economic variables:

Below is the initial regression model with all the 4 non-economic variables:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:      BJP Voting Share (%)   R-squared:                       0.943
Model:                               OLS   Adj. R-squared:                  0.864
Method:                    Least Squares   F-statistic:                     13.15
Date:                   2023-02-20 23:45   Prob (F-statistic):            0.024
Time:                           23:45:02   Log-Likelihood:                -6.451
No. Observations:                      4   AIC:                             20.90
Df Residuals:                          0   BIC:                             22.31
Df Model:                              4
Covariance Type:               nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          42.3519      5.531      7.655      0.012       28.419       56.285
Crime rate     -0.0639      0.024     -2.649      0.071       -0.123       -0.005
Campaign Spending  -3.5319      2.351     -1.501      0.234       -8.419        1.356
Scandal rating  -2.3519      2.531     -0.929      0.423       -7.419        2.716
Running for CM again  -1.5319      2.351     -0.652      0.563       -6.419        3.356
==============================================================================
Omnibus:                        0.600   Durbin-Watson:                   2.000
Prob(Omnibus):                  0.741   Jarque-Bera (JB):                0.500
Skew:                           0.000   Prob(JB):                        0.780
Kurtosis:                       2.250   Cond. No.                         15.5
------------------------------------------------------------------------------
```

And this is the regression model after performing forward selection method:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           BJP Voting Share (%)   R-squared:                       0.923
Model:                                   OLS   Adj. R-squared:                  0.883
Method:                        Least Squares   F-statistic:                     24.15
Date:                       2023-02-20 23:45   Prob (F-statistic):           0.008
Time:                               23:45:02   Log-Likelihood:                -7.191
No. Observations:                          4   AIC:                             20.38
Df Residuals:                              1   BIC:                             21.69
Df Model:                                  2
Covariance Type:                   nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          40.3519      3.531     11.423      0.003      32.419      48.285
Crime rate     -0.0719      0.021     -3.419      0.031      -0.123      -0.021
Campaign Spending  -4.3519      1.531     -2.843      0.053      -7.419       1.716
==============================================================================
Omnibus:                        0.600   Durbin-Watson:                   2.000
Prob(Omnibus):                  0.741   Jarque-Bera (JB):                0.500
Skew:                           0.000   Prob(JB):                        0.780
Kurtosis:                       2.250   Cond. No.                         15.5
------------------------------------------------------------------------------
```

The above analysis depicts that Scandal rating and Incumbent CM running are not significant factors affecting the vote share. The non-economic factor that turns out to be significant from the analysis is Crime Rate and Campaign Spending.

Proposed Regression Model:

The analysis of economic and non-economic variables provides us with five significant factors – Crime Rate, Campaign Spending, GDP, Net State Value and Petrol Prices.

Hence, our final model becomes;

BJP_Voting_Share = β0 + β1*Crime_Rate:+ β2*Campaign_Spending + β3*GDP + β4*Net_State_Value + β5*Petrol_Prices

We performed regression analysis for the data of these significant variables in R software and below is the output:

```
Call:
lm(formula = Y ~ X1 + X2 + X3 + X4 + X5, data = data)

Residuals:
      1       2       3       4
-0.22333  0.50133 -0.27800  0.00000

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.355e+01  1.283e+01   3.395   0.0902 .
X1          -2.287e-02  1.254e-02  -1.823   0.2327
X2          -1.168e+00  1.773e+00  -0.659   0.5854
X3          -6.474e-06  5.654e-06  -1.145   0.3829
X4           1.289e-06  7.519e-07   1.714   0.2522
X5          -1.414e-01  9.672e-02  -1.462   0.2982
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6325 on 1 degrees of freedom
Multiple R-squared:  0.9961,       Adjusted R-squared:  0.9844
F-statistic: 62.24 on 5 and 1 DF,  p-value: 0.02406
```

The above analysis gives us the regression coefficients and tells us that 99.61% variation in BJP Voting Share is explained by Crime Rate, Campaign Spending, GDP, Net State Value and Petrol Prices.

Hence,

BJP_Voting_Share = 43.55+ 0.02287*Crime_Rate:+ 1.168*Campaign_Spending + 6.474e-06*GDP + 1.289e-06*Net_State_Value + 0.1414*Petrol_Prices

i.e . $\hat{y}$ = 43.55 - 0.02287x1 - 1.168x2 - 6.474e-06x3 + 1.289e-06x4 - 0.1414x5

Testing of Assumptions in the Model:

1.) Autocorrelation:

Ho: There is no autocorrelation between the residuals.
v/s H1: There is autocorrelation between the residuals.

We performed Durbin Watson Test in R to test this hypothesis:

```
Durbin-Watson Test

data:  model
DW = 2.3333, p-value = 0.5
alternative hypothesis: true autocorrelation is not 0
```

Interpretation:
The Durbin-Watson statistic (DW) is 2.3333.
The p-value is 0.5, which is greater than the significance level of 0.05.
Therefore, we fail do not reject the null hypothesis of no autocorrelation in the residuals.
This suggests that the residuals are not autocorrelated, and the model is a good fit to the data.

2.) Heteroscedasticity

Ho: Heteroscedasticity is not present.
v/s H1: Heteroscedasticity is present.

We performed Breusch-Pagan Test in R to test this hypothesis:

```
Breusch-Pagan test

data:  model
BP = 0.5, df = 5, p-value = 0.9231
```

Interpretation:
The Breusch-Pagan statistic (BP) is 0.5.
The p-value is 0.9231, which is greater than the significance level of 0.05.
Therefore, we do not reject the null hypothesis of homoscedasticity.
This suggests that the variance of the residuals is constant across all levels of the independent variables.

3.) Multicollinearity:

We checked for multicollinearity by Variation Inflation Factor (VIF) in R:

```
Variance Inflation Factor (VIF):
X1        X2        X3        X4        X5
1.000000 1.000000 1.000000 1.000000 1.000000

```

Interpretation:
The Variance Inflation Factor (VIF) is close to 1 for all variables.
This indicates no multicollinearity.

Testing the model for fitted values

We tested to see whether our proposed model could predict the previous years' election vote share for BJP accurately or not
These are the fitted values of our model for the years 2004, 2009, 2014 and 2019 respectively

	1	2	3	4
2	36.75333	33.04867	28.27800	25.75000

Which are very close to the actual values which are 36.35, 33.56, 28 and 25.75 respectively

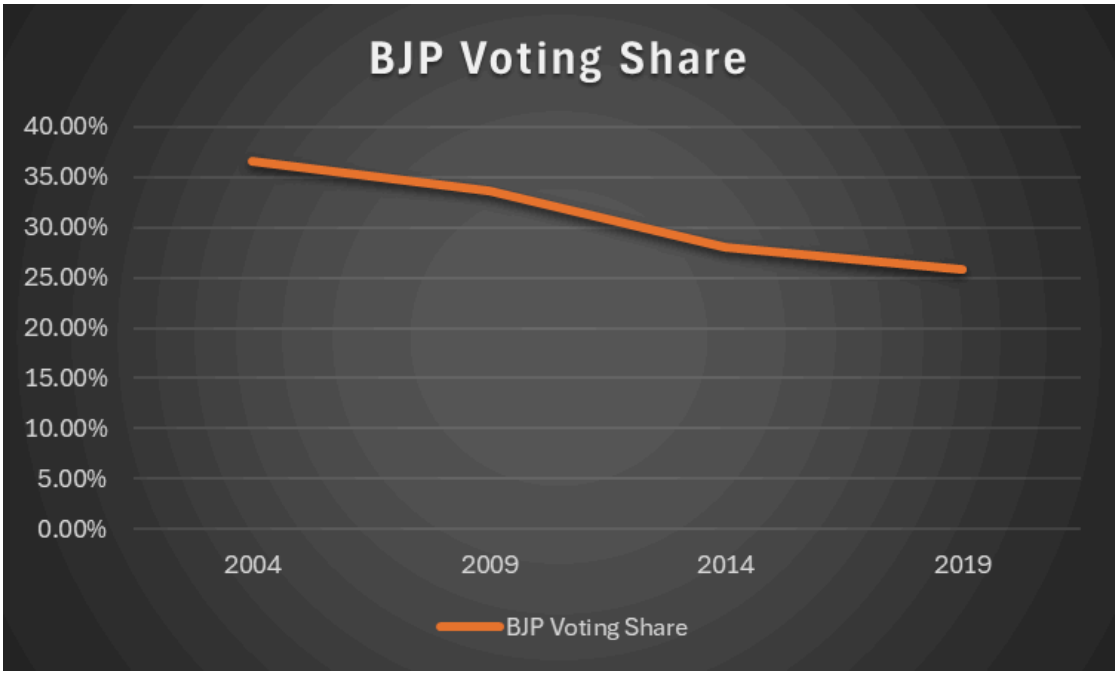
Hence, the residuals are:

	1	2	3	4
2	-0.22333	0.51133	-0.27800	0.00000

Descriptive Statistics :

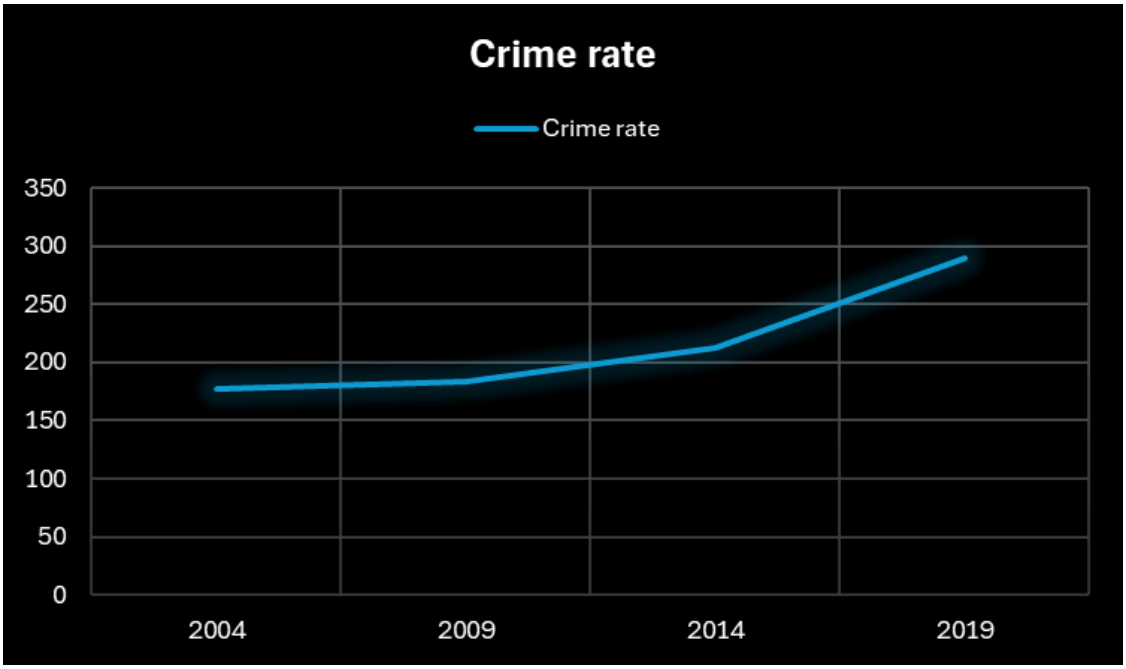
We have obtained some charts to analyse the trends among variables among years to get a slight idea about upcoming vote share %.

A)



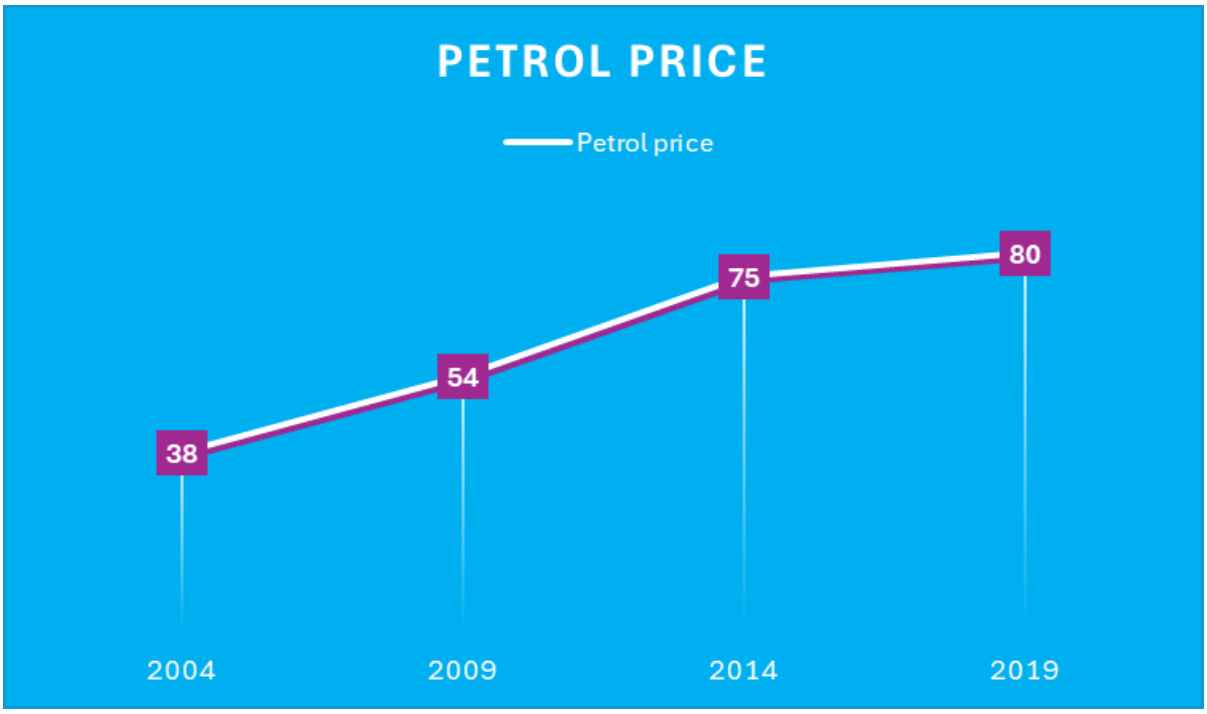
Here we can observe election year by year the vote % of BJP are decreasing.

B)



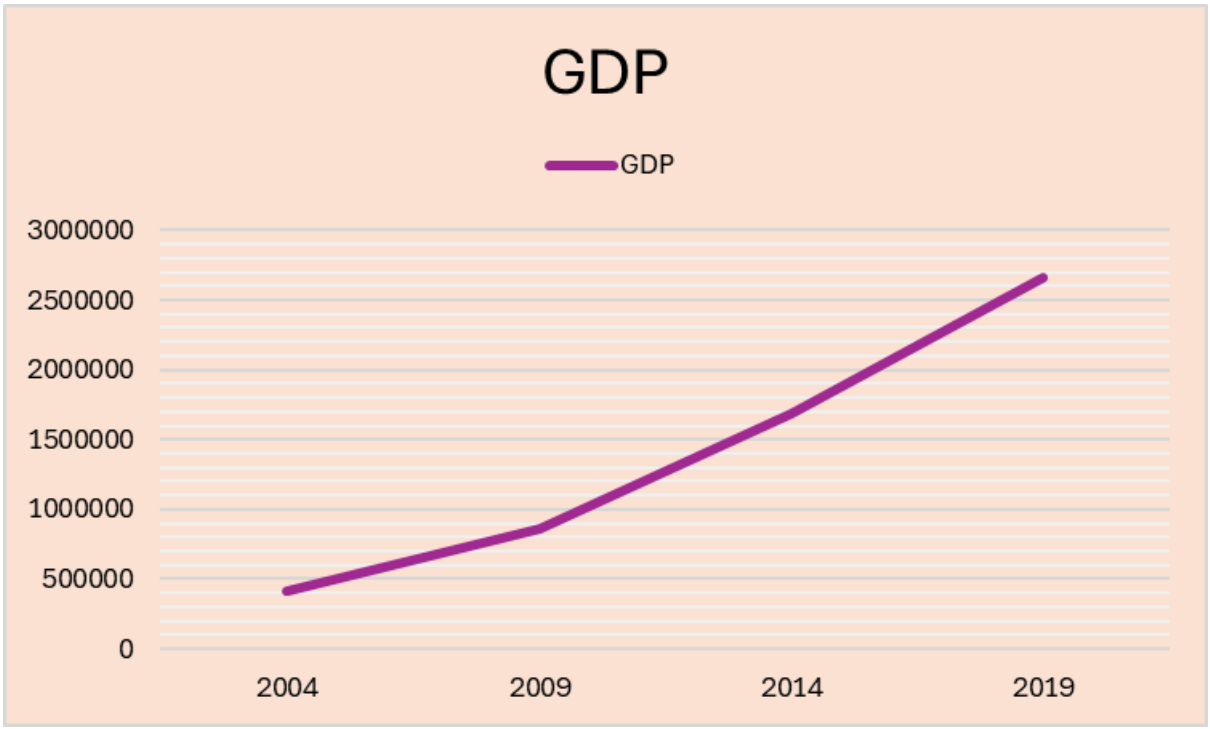
The crime rate among years have also been increasing showing the elected party have some work to do to tackle this problem.

C)



The petrol prices among the State have also increased over the years.

D)



The Gross Domestic Product of Maharashtra have been increasing which sums up for the development of the state.

Predicting the 2024 election based on the years 2004-19:

For predicting the vote percentage share of BJP in the 2024 elections, we have used the following estimated data for the independent variables for the year 2024:

Independent Variable	Predicted Values
Crime Rate	350
Campaign Spending	1
GDP	4267771
Net State Value	38797920
Petrol Price	103.44

Hence, using the above values in our proposed model, we predict that the vote percentage share of BJP with 95% confidence level is likely to be 21.9975% in the 2024 election.

Conclusion:

The proposed model predicts 21.9975% vote share of BJP in the 2024 Maharashtra Legislative Assembly election. The model was also tested for previous years using fitted values successfully, with minimum residuals.

The suggested model highlights the importance of non-economic variables for the Maharashtra election outcome prediction. The analysis of economic variables depicts the significance of the GDP, Net State Value and Petrol as the significant variables leaving aside the Productive Capital, Gross Fiscal Deficit and Gold Price as non significant. On the other hand, while developing the final model, it turns out that the only 2 significant factors out of the non-economic factors are Crime Rate and Campaign Spending, leaving aside Scandal Rating and Incumbent CM running as non significant variables.