BUSINESS REPORT

MACHINE LEARNING

Harsh Alkesh Pandya
[PGPDSBA FEB_2021_A]

MACHINE LEARNING

PROJECT 1 (Exit-Poll Prediction)

You are hired by one of the leading news channel CNBE who wants to analyze recent elections. This survey was conducted on 1525 voters with 9 variables. You have to build a model, to predict which party a voter will vote for on the basis of the given information, to create an exit poll that will help in predicting overall win and seats covered by a particular party. Dataset for Problem: Election_Data.xlsx.

PROJECT 2 (Speeches of Presidents)

TEXT ANALYSIS

In this particular project, we are going to work on the inaugural corpora from the nltk in Python. We will be looking at the following speeches of the Presidents of the United States of America:

1. President Franklin D. Roosevelt in 1941
2. President John F. Kennedy in 1961
3. President Richard Nixon in 1973

TABLE OF CONTENTS

LIST OF FIGURE AND TABLE

LIST OF Table

LIST OF FIGURES

Problem 1:

You are hired by one of the leading news channel CNBE who wants to analyze recent elections. This survey was conducted on 1525 voters with 9 variables. You have to build a model, to predict which party a voter will vote for on the basis of the given information, to create an exit poll that will help in predicting overall win and seats covered by a particular party.

Dataset for Problem: Election_Data.xlsx Data Ingestion:

1.1) Read the dataset. Do the descriptive statistics and do the null value condition check. Write an inference on it.

We have read the dataset Election_Data.xlsx from pandas read_excel function

- **Let's Check the Head of the dataset**

	Unnamed: 0	vote	age	economic.cond.national	economic.cond.household	Blair	Hague	Europe	political.knowledge	gender
0	1	Labour	43	3	3	4	1	2	2	female
1	2	Labour	36	4	4	4	4	5	2	male
2	3	Labour	35	4	4	5	2	3	2	male
3	4	Labour	24	4	2	2	1	4	0	female
4	5	Labour	41	2	2	1	1	6	2	male

(Table no: 1 Data set)

We found that there are two categorical Variables.

- **Description of the dataset**

	count	mean	std	min	25%	50%	75%	max
age	1525.0	54.182295	15.711209	24.0	41.0	53.0	67.0	93.0
economic.cond.national	1525.0	3.245902	0.880969	1.0	3.0	3.0	4.0	5.0
economic.cond.household	1525.0	3.140328	0.929951	1.0	3.0	3.0	4.0	5.0
Blair	1525.0	3.334426	1.174824	1.0	2.0	4.0	4.0	5.0
Hague	1525.0	2.746885	1.230703	1.0	2.0	2.0	4.0	5.0
Europe	1525.0	6.728525	3.297538	1.0	4.0	6.0	10.0	11.0
political.knowledge	1525.0	1.542295	1.083315	0.0	0.0	2.0	2.0	3.0

(Table no: 2 Describe)

We can see that most values are ranging between 0 to 11 except Age so to put it in the same range We will try Binning in subsequent Steps.

- **Check for Null values**

```
vote                       0
age                        0
economic.cond.national     0
economic.cond.household    0
Blair                      0
Hague                      0
Europe                     0
political.knowledge        0
gender                     0
dtype: int64
```

From panda's null value check function, we found Above results which says there are not any null value present in the dataset

- **Checking Duplicate Records.**

From Pandas duplicate function we found 8 duplicate values so we have dropped them.

```
dups.sum()
```

```
8
```

- **Skewness**

```
age                          0.139800
economic.cond.national      -0.238474
economic.cond.household     -0.144148
Blair                       -0.539514
Hague                        0.146191
Europe                      -0.141891
political.knowledge         -0.422928
```

As we can see from distribution plot that most of the data points are negatively/Left skewed except age & Hague. skewness value is given below.

- **Shape**

```
no. of rows:  1525
no. of columns:  9
```

As we show in the Data frame there is total 1525 no of Rows and Columns.

- **Printing Categorical Values with their Counts**

```
vote   No of Levels: 2
Labour          1063
Conservative     462
Name: vote, dtype: int64
```

```
gender   No of Levels: 2
female    812
male      713
Name: gender, dtype: int64
```

A we have shown in the Above result, we found there are two Categorical Variables For Male, Female, Labour, Conservative.

1.2) Perform Univariate and Bivariate Analysis. Do exploratory data analysis. Check for Outliers.

We have done Null Value check and found that there are No Null Values in the Dataset.

- **Shape of the Dataset**

(1525,10) – Dataset is Having 1525 rows and 10 Columns

- **Info of the dataset**

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1525 entries, 0 to 1524
Data columns (total 10 columns):
 #   Column                   Non-Null Count  Dtype 
---  ------                   --------------  ----- 
 0   Unnamed: 0               1525 non-null   int64 
 1   vote                     1525 non-null   object
 2   age                      1525 non-null   int64 
 3   economic.cond.national   1525 non-null   int64 
 4   economic.cond.household  1525 non-null   int64 
 5   Blair                    1525 non-null   int64 
 6   Hague                    1525 non-null   int64 
 7   Europe                   1525 non-null   int64 
 8   political.knowledge      1525 non-null   int64 
 9   gender                   1525 non-null   object
dtypes: int64(8), object(2)
memory usage: 119.3+ KB
```

(Fig: 1 Information)

Data set contains 2 Categorical columns Vote and gender, all other columns are of Integer type.

- **Unique Value Counts for all the object data types**

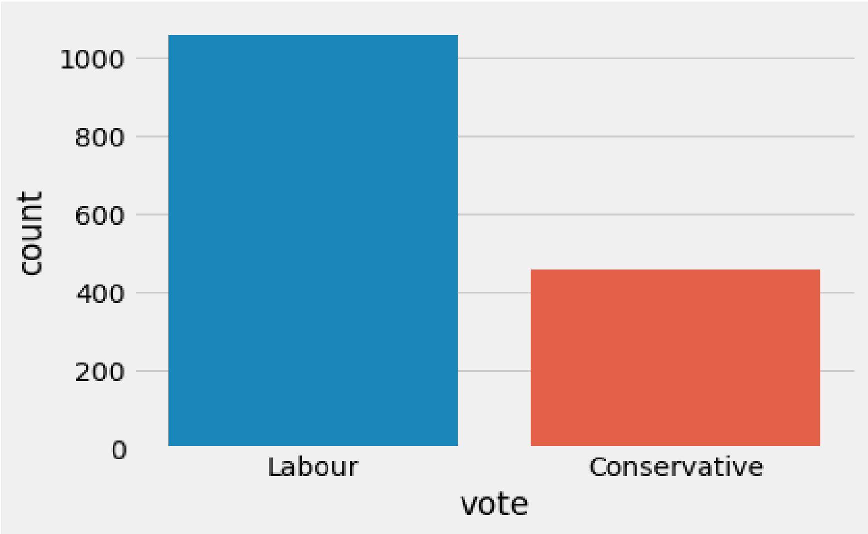
```
vote   No of Levels: 2
Labour          1063
Conservative     462
Name: vote, dtype: int64



gender   No of Levels: 2
female    812
male      713
Name: gender, dtype: int64
```

(Fig: 2 Unique Values)

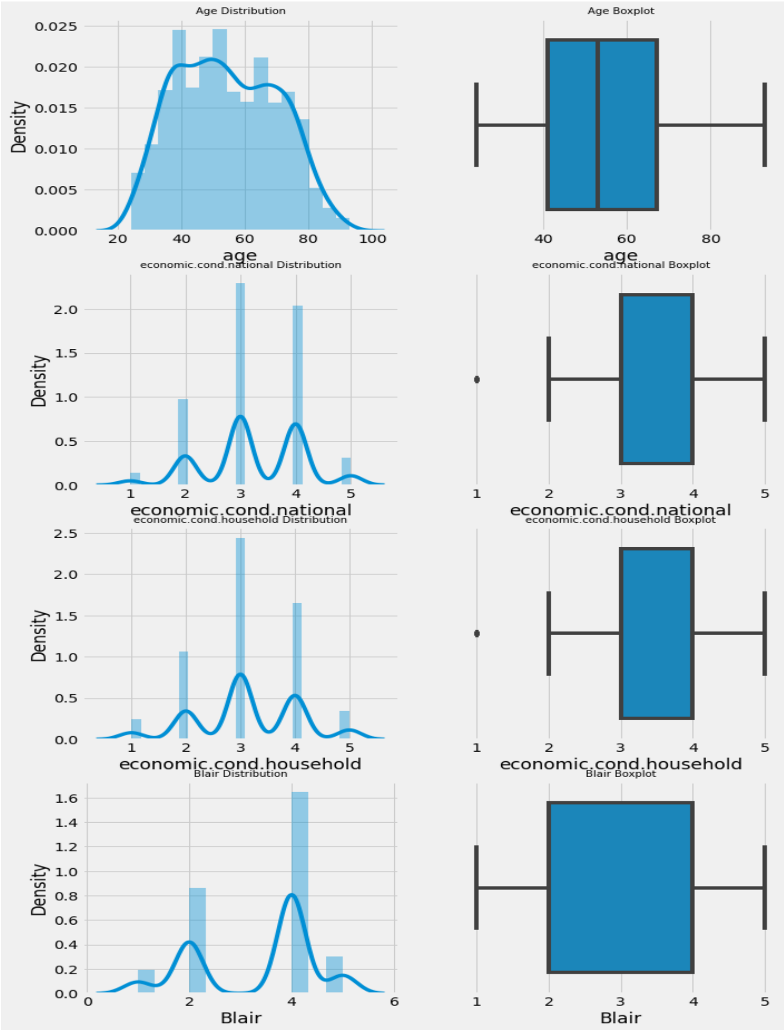
EDA- Univariate Analysis

- **How Many Votes Each Party have Got :- Shown using Countplot**

- `Labour          1057`
- `Conservative     460`



(Fig: 3 Vote vs. Count Plot)



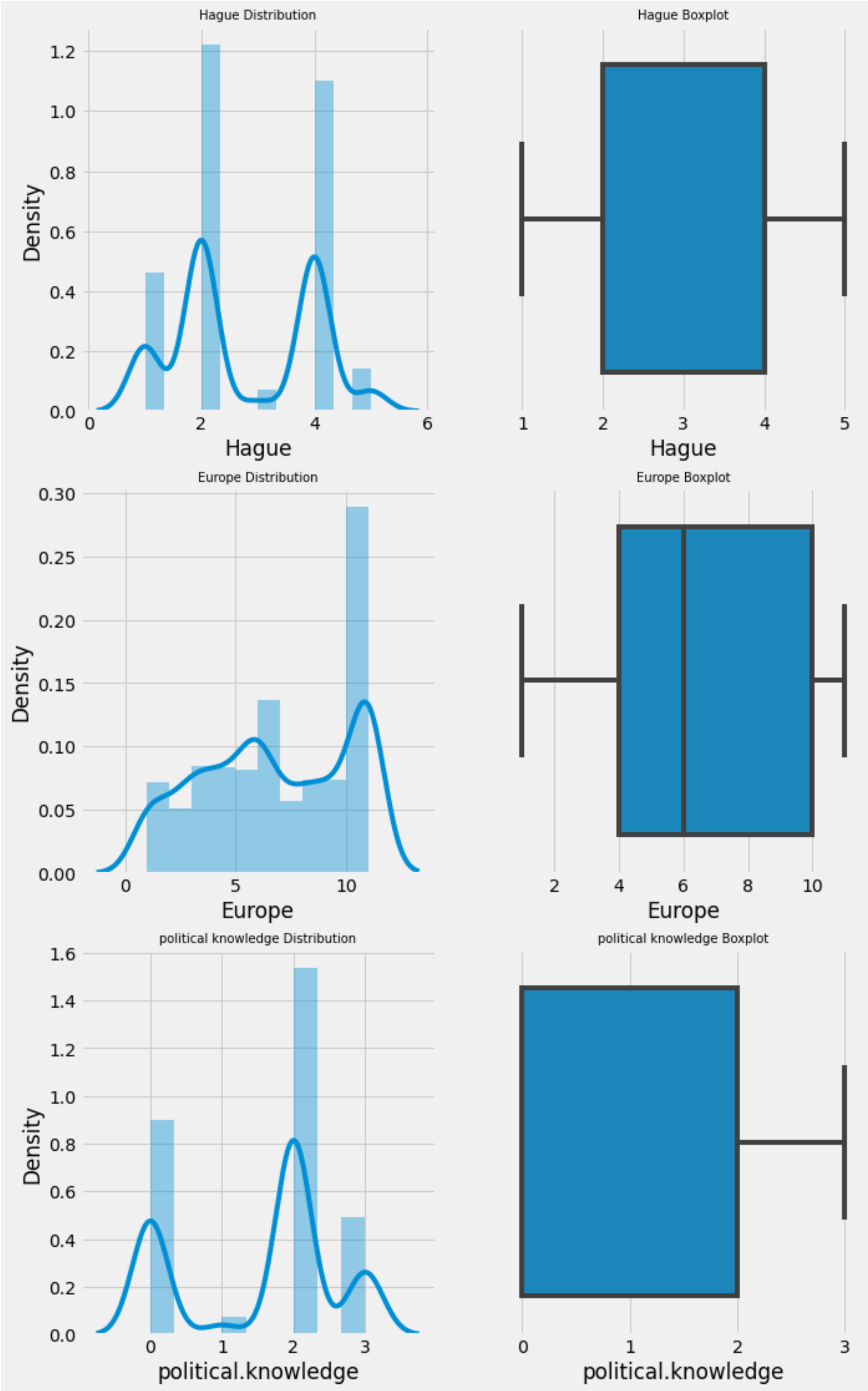
(Fig: 4 Distribution Plot)

As we can see from distribution plot that most of the data points are negatively/Left skewe d except age & Hague. skewness value is given below.

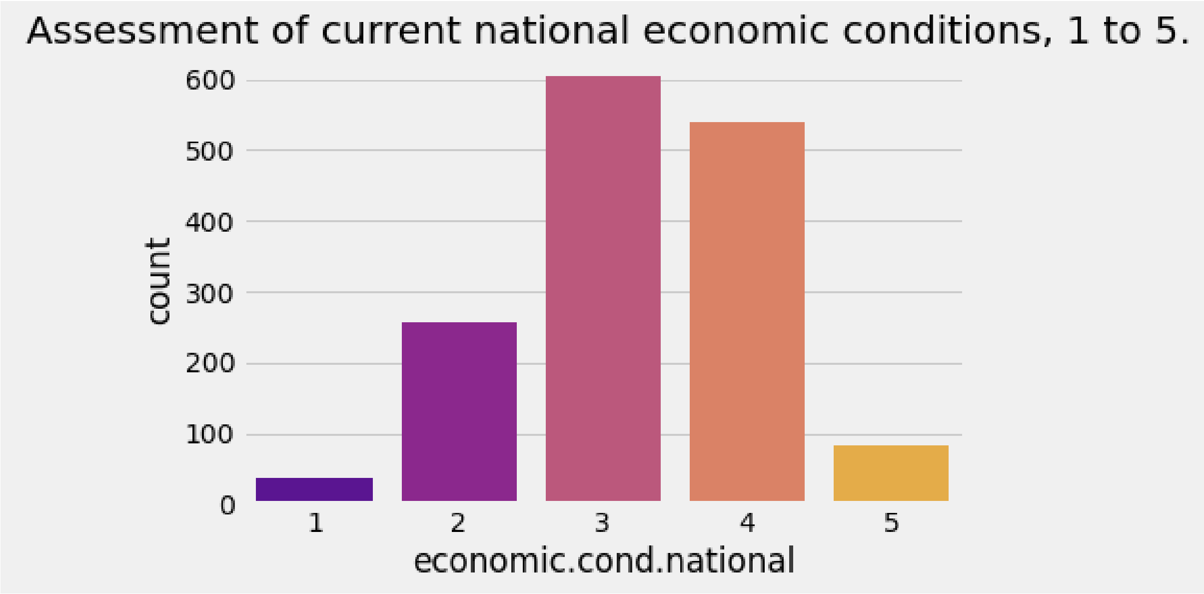
```
age                        0.139800
economic.cond.national    -0.238474
economic.cond.household   -0.144148
Blair                     -0.539514
Hague                      0.146191
Europe                    -0.141891
political.knowledge       -0.422928
```

(Fig: 5 Skewness)

Also, from Boxplot we have found that there are outliers present on economic condition national & household attributes, but we hold on here for a minute and looked that ratings can be 1 (' Assessment of current household/National economic conditions, 1 to 5.'). So, all the values are correct and so from a business perspective it will not be a wise decision to treat these Outliers.

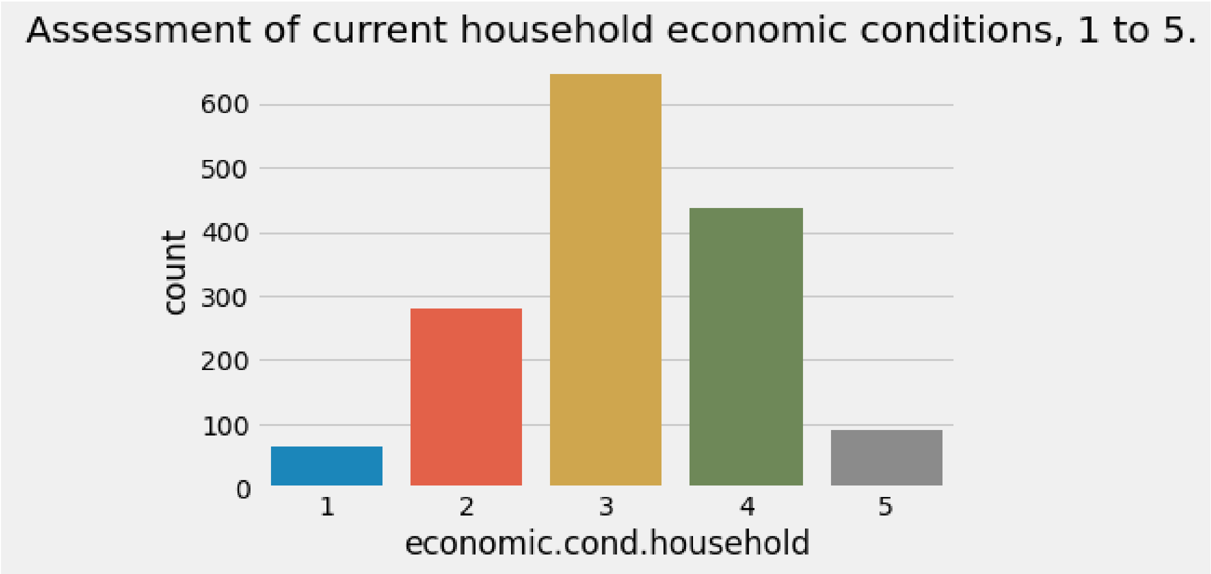


(Fig: 6 Distribution Plot2)



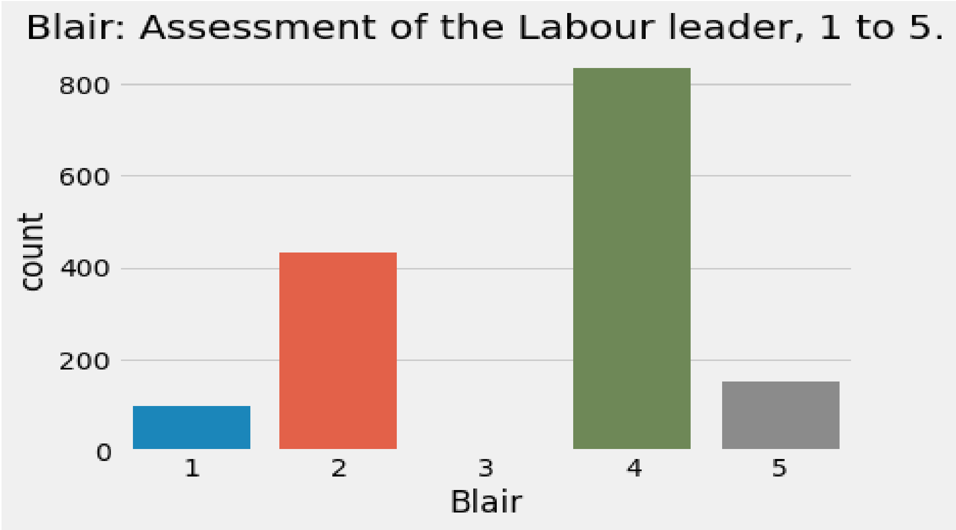
(Fig: 7 Economic Count Plot)

Most frequent national Current economic condition rating is 3 and least frequent condition Is 1. we can say that average rating is lying between 2 & 3.



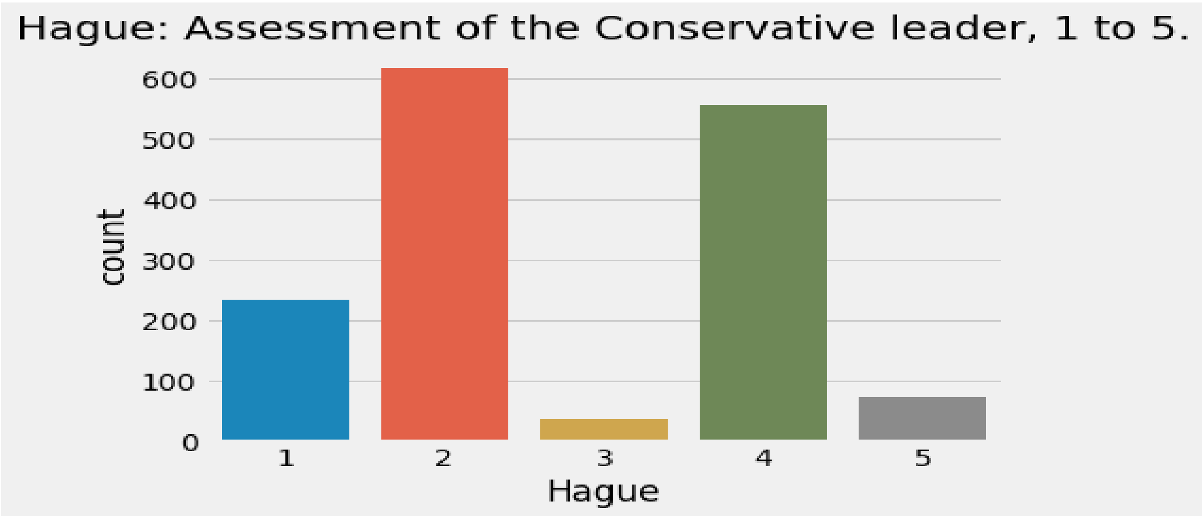
(Fig: 8 Economic Household Plot)

Most frequent national Current household condition rating is 3 and least frequent condition Is 1. We can say that average rating is lying between 2 & 3.



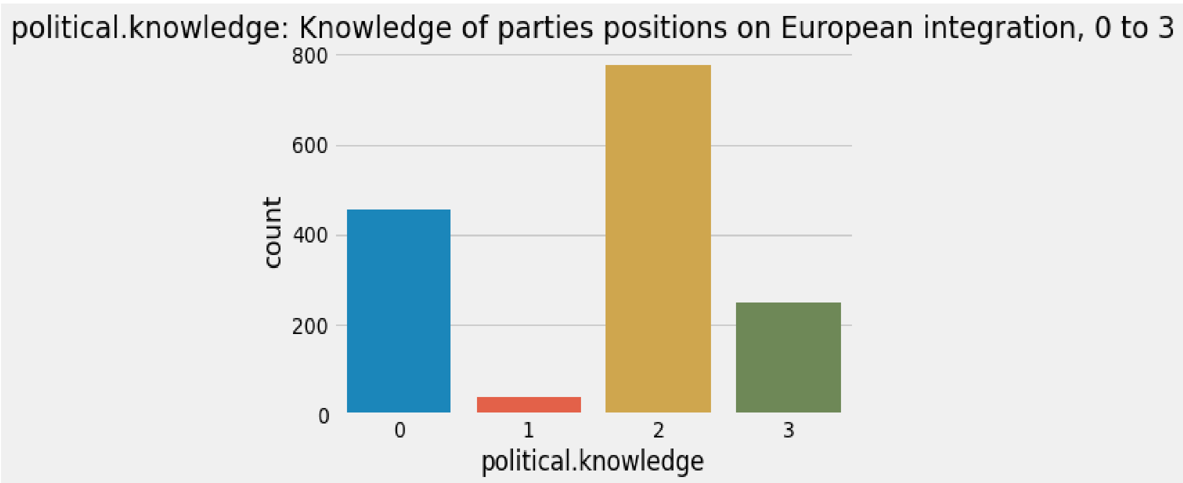
(Fig: 9 Blair vs. Count Plot)

Labor leader has got most frequent rating of 4 in the surveys.



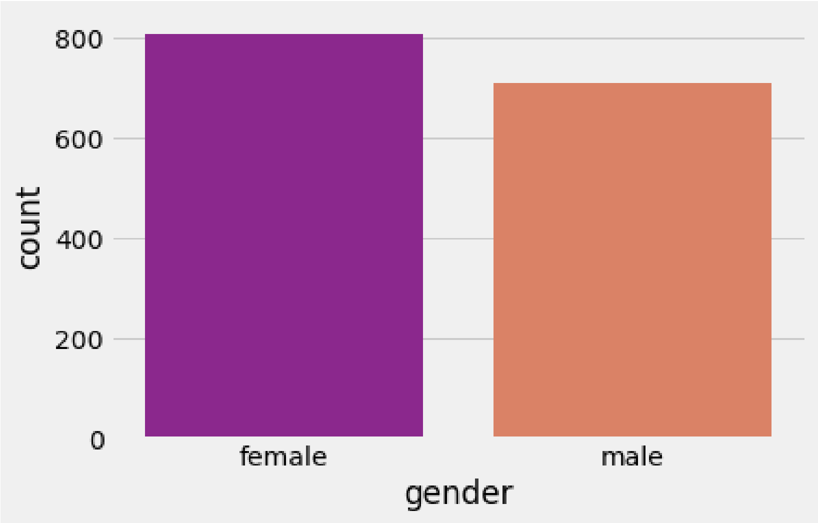
(Fig: 10 Huge vs. Count Plot)

Conservative leader have got most frequent rating of 2.



(Fig: 11 Political. Knowledge vs. Count Plot)

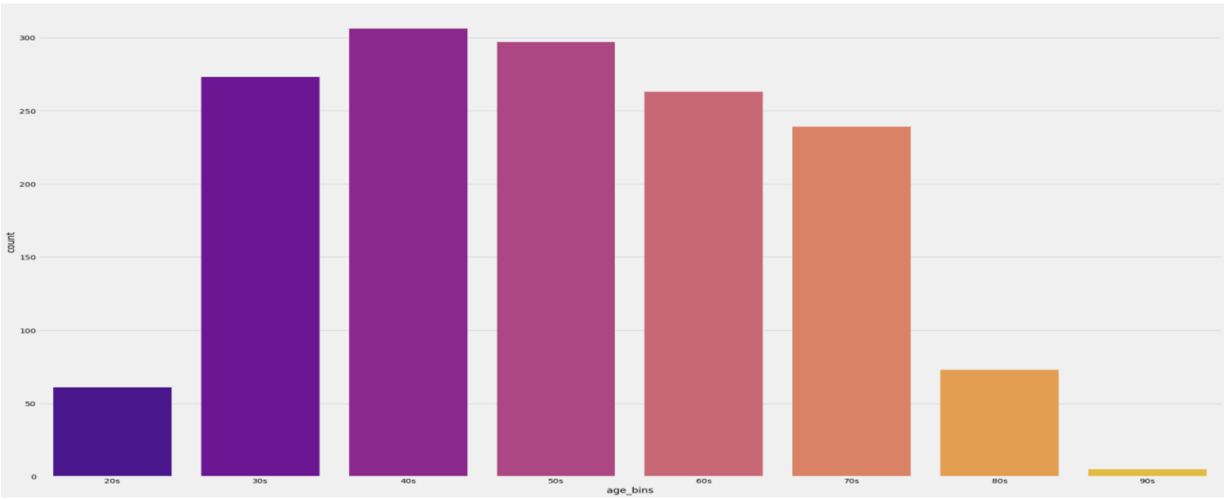
Most frequent scale for political knowledge is 2.



(Fig: 12 Gender Count Plot)

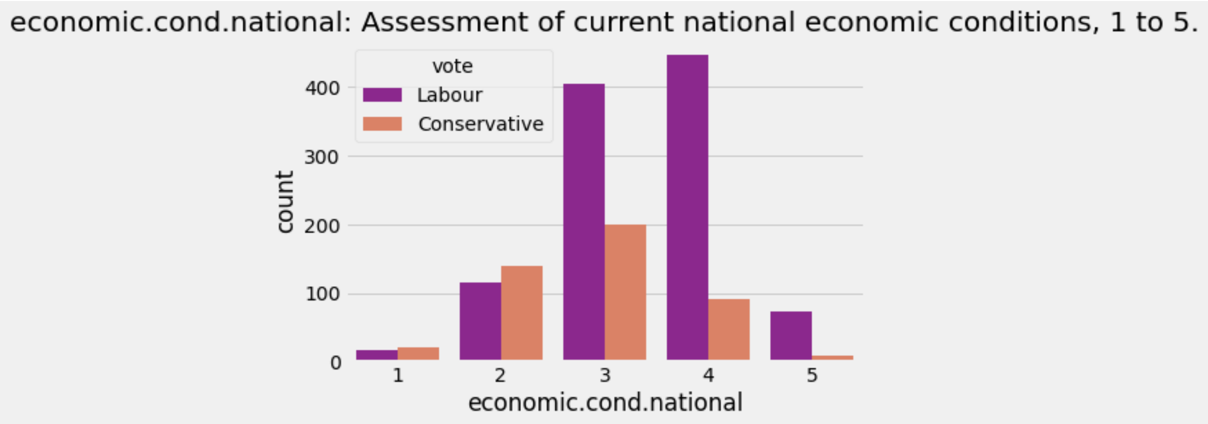
From the Graph there are 812 No. of females & 713 are males

- **Age Bins Count: -**



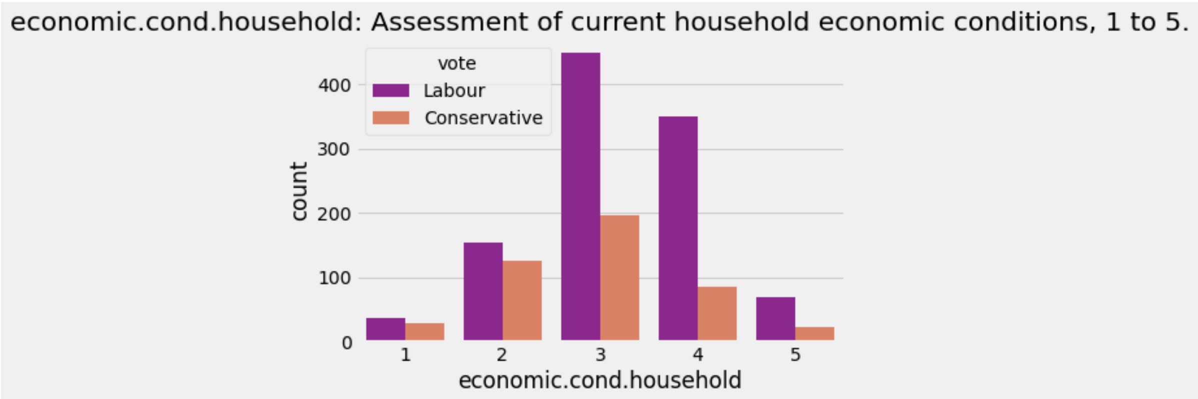
(Fig: 13 Age Bins Plot)

From the above Graph we can found out that most of the respondents are in between their 40s & 50s.



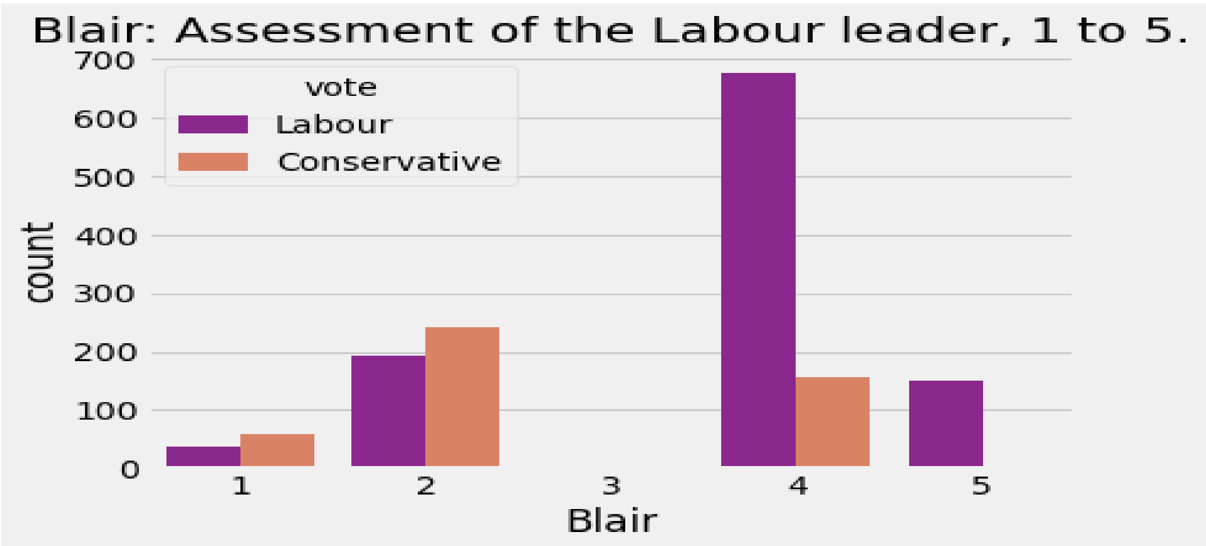
(Fig: 14 National vs. Count Plot)

As per above graph Labor party have got most of the great ratings



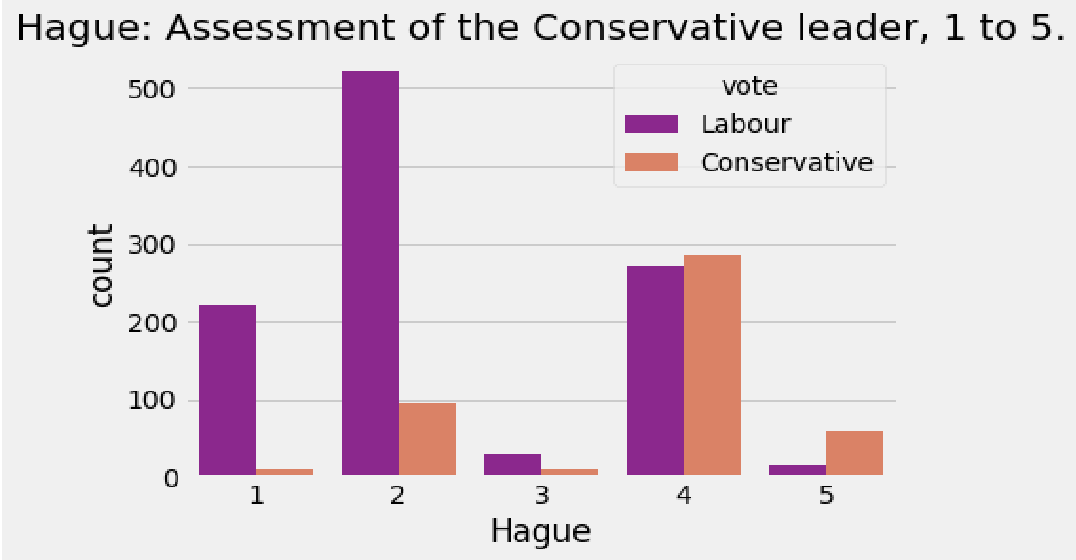
(Fig: 15 Current Household Plot)

As per above graph Conservative party have got less ratings in comparison to labor party.



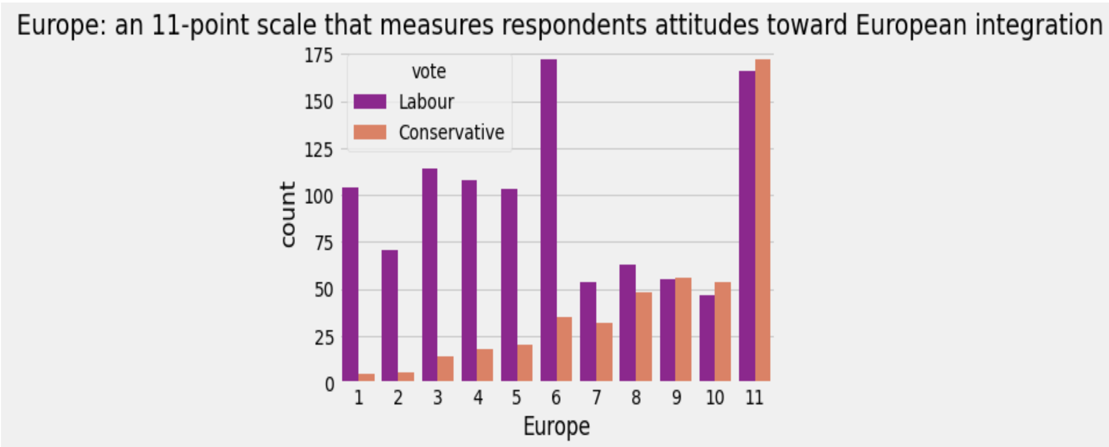
(Fig: 16 Blair vs. Count Plot)

Labor leader have got 4 ratings in their assessment.



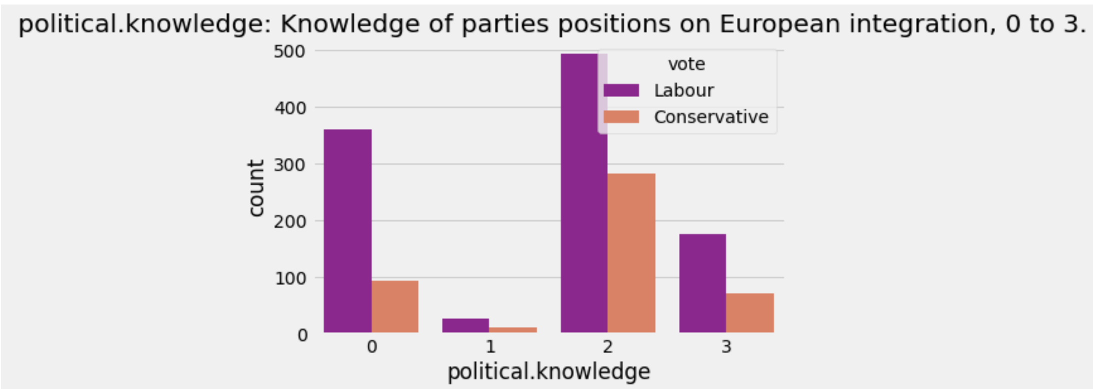
(Fig: 17 Huge Count Plot)

Conservative leader which are given good ratings as per the graph most of the time. but Vote will be given to Labor party leader.



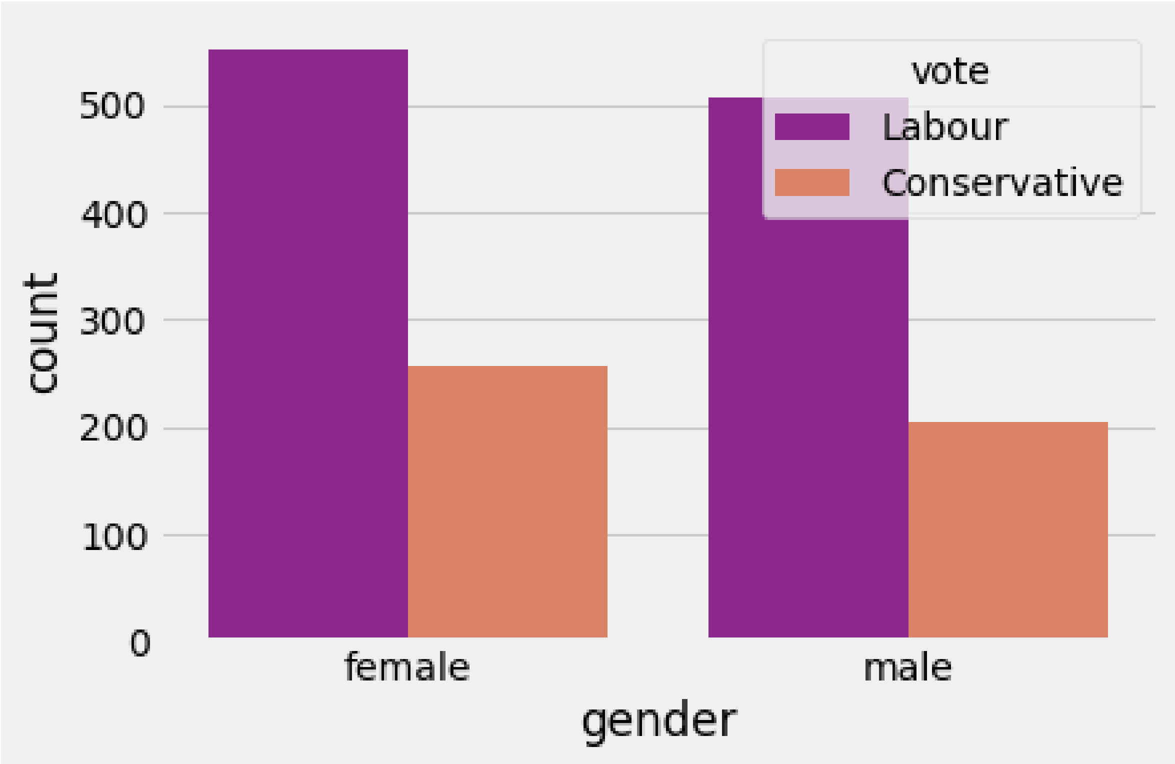
(Fig: 18 Europe vs. Count Plot)

Above graph shows that most of the Voters have Eurosceptic attitude towards European integrations of Conservative Party



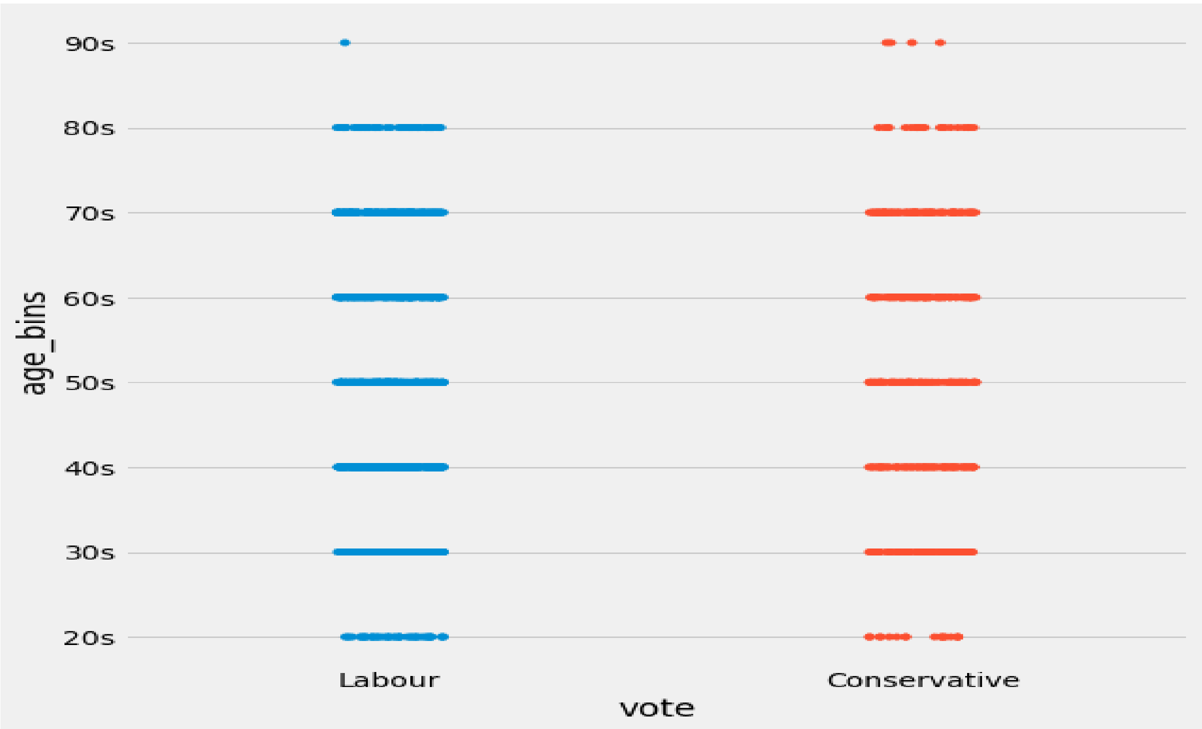
(Fig: 19 Knowledge vs. Count Plot)

Labor part is having less Europeans integration, so they have got most of the votes.



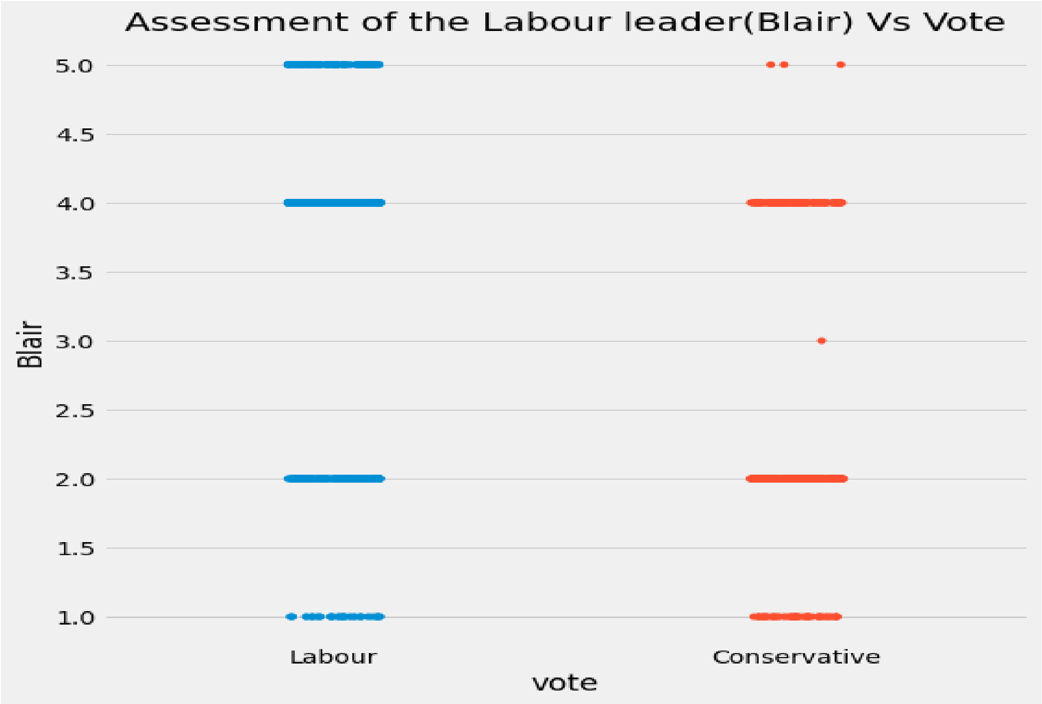
(Fig: 20 Gender vs. Count Plot)

More of the males and females have given vote to the Labor party.



(Fig: 21 Vote vs. Age bins Plot)

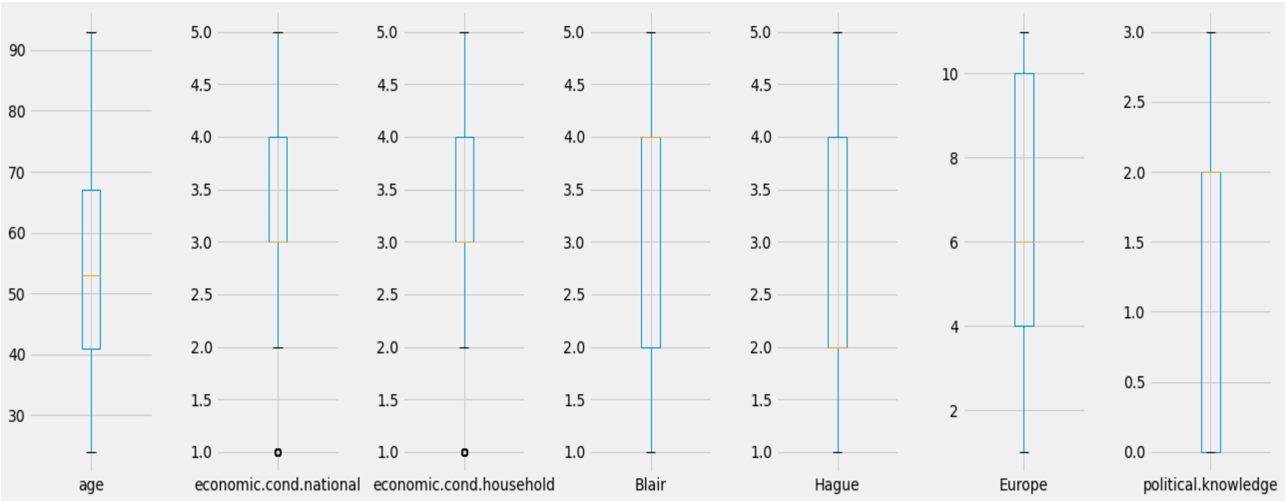
In 20s There are very few peoples which have given Vote to the Conservative party.



(Fig: 22 Vote vs. Blair Plot)

Amongst labor party leaders which are given average ratings of 4, Conservative party leaders have win there.

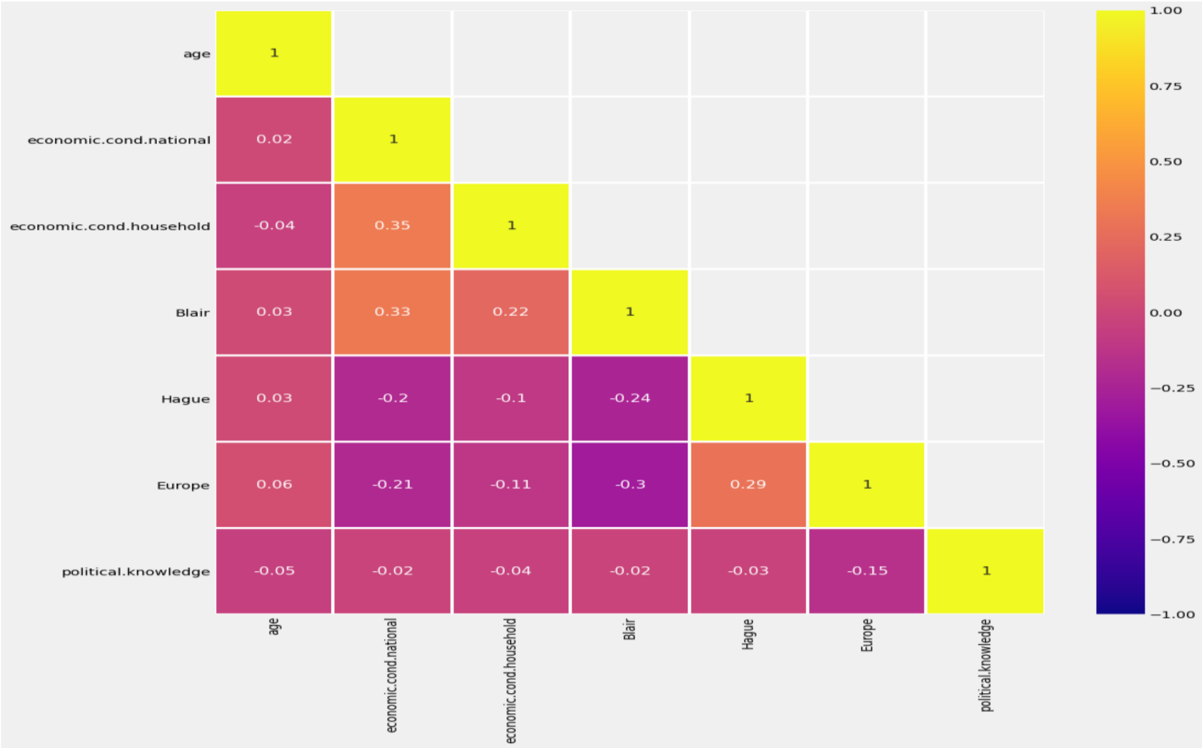
- **Boxplot to check the Outliers**



(Fig: 23 Box Plot)

Also, from Boxplot we have found that there are outliers present on economic condition national & household attributes, but we hold on here for a minute and looked that ratings can be 1 (' Assessment of current household/National economic conditions, 1 to 5.'). In real world Outliers are the values which are mistakenly captures in the data, so all the values are correct and so from a business perspective it will not be a wise decision to treat these Outliers.

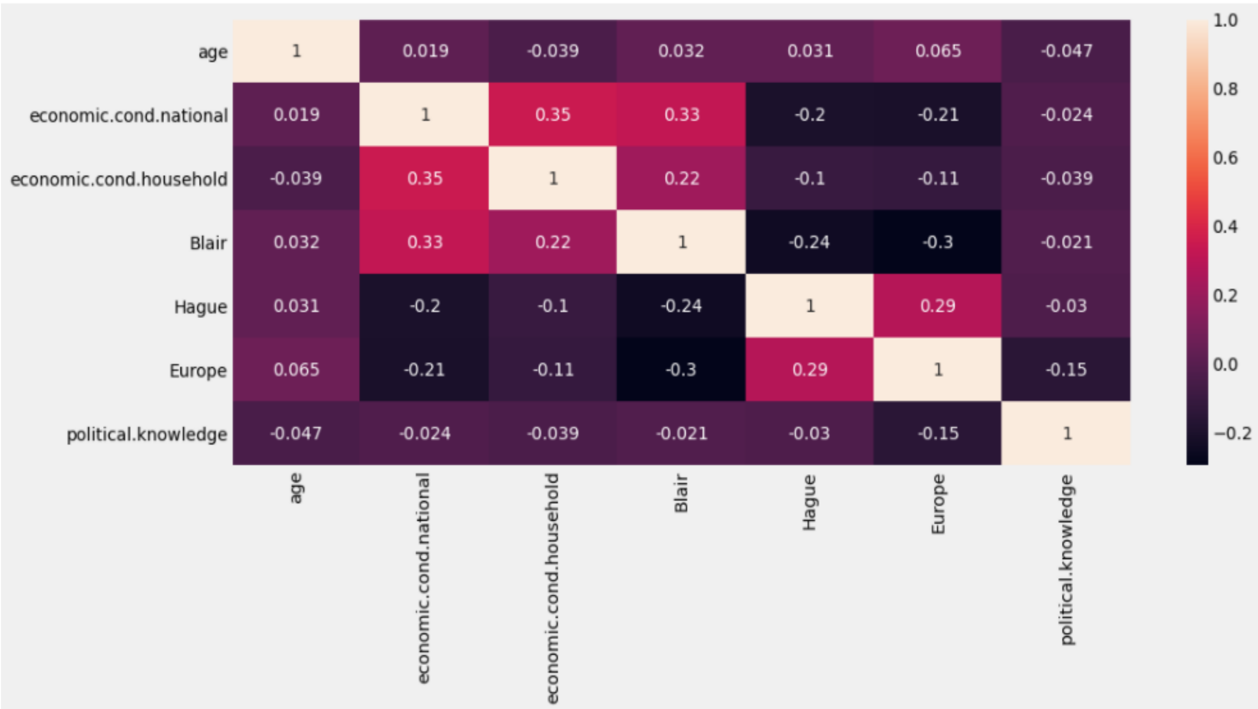
- **Checking for Correlations using Heat map**



(Fig: 24 Correlation Plot)

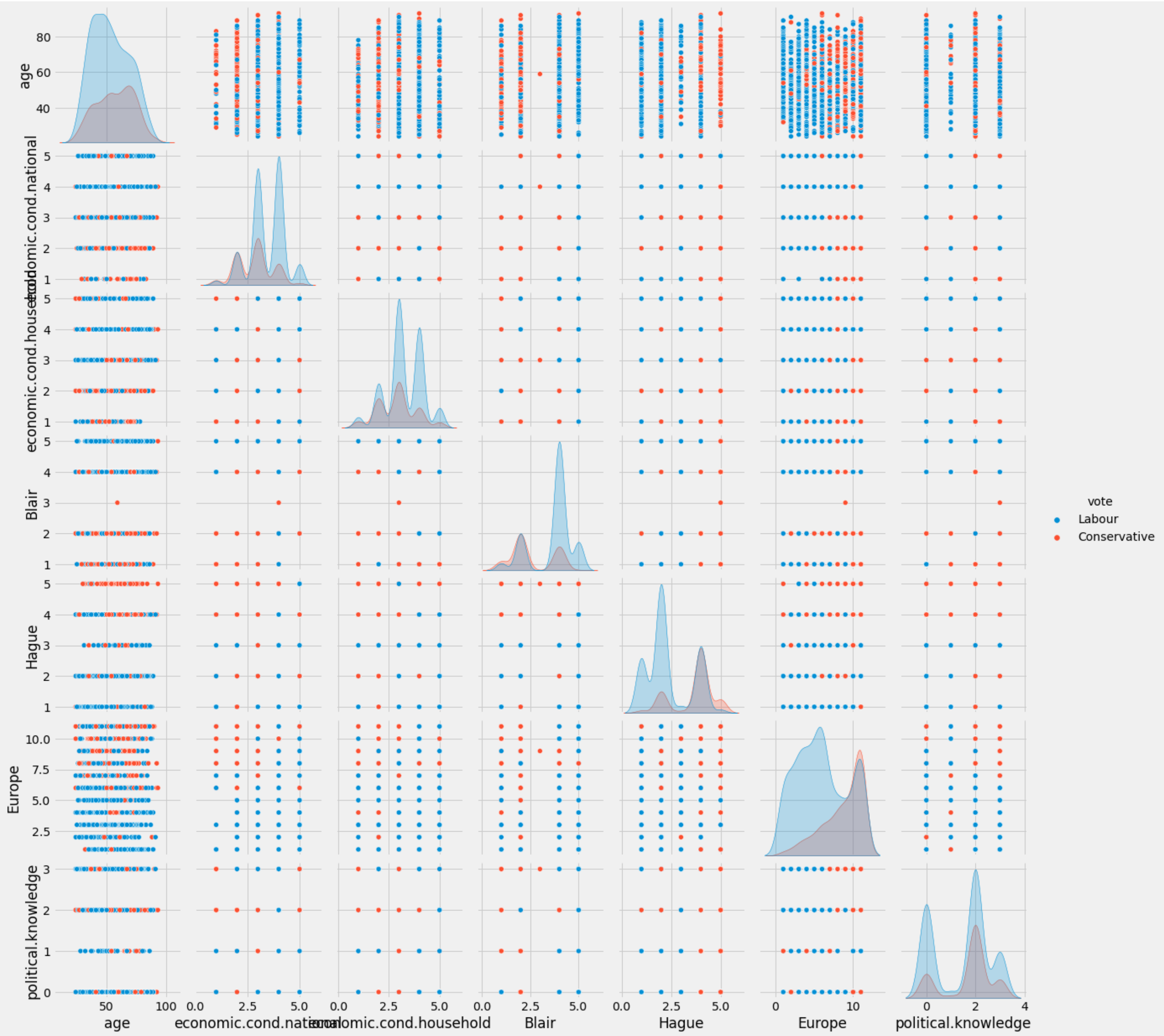
There is not a very strong Correlation amongst any of the variables of the dataset

- **Heat Map**



(Fig: 25 Heat map)

- **Pair plot**



(Fig: 26 Pair Plot)

1.3) Encode the data (having string values) for Modelling. Is Scaling necessary here or not? Data Split: Split the data into train and test (70:30).

- **Encode the data (having string values)**

As we know that we are having 2 Categorical Variables ‘Gender’ & Vote so we have done one hot encoding with dropping of first column to avoid multicollinearity (Refer Jupyter file for more details)

df =pd.get_dummies(df, columns=['gender'],drop_first=True)

As we know that Vote is Our Target Column, we have to classify Whether a person have Voted for Conservative Or labour party so for the sake of our better interpretation of our model we are not doing any encoding there.

Also we know that Age Group ranges from 20 to 100 and all other variables m ost of them are Ordinal Variables like rating ['vote', 'economic.cond.national', ' economic.cond.household,'Blair', 'Hague', 'Europe', 'political.knowledge', 'gend er'], So for better understanding and interpretation of the Model we are Doing Binning of Age Column as below .

df['age_bins'] = pd.cut(x=df['age'], bins=[20, 29, 39, 49,59,69,79,89,99],labels=['20s', '30s', '40s','50s','60s','70s','80s','90s'])

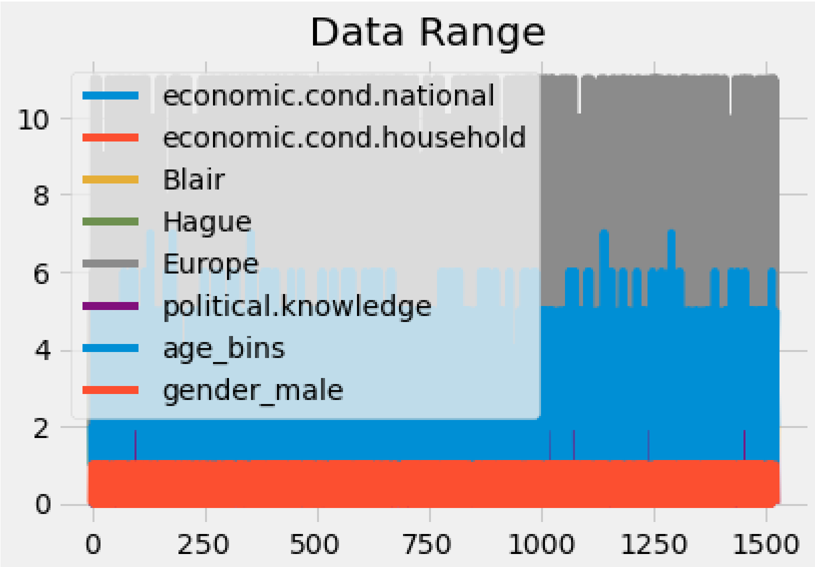
```
[20s, 30s, 40s, 50s, 60s, 70s, 80s, 90s]
Categories (8, object): [20s < 30s < 40s < 50s < 60s < 70s <
80s < 90s]
```

As you can see that We have put peoples which are more than 20 years of age we make it as 20 s, and persons who are more than 30 years of age up to 39, we have put them all in 30s.

Now the above Values are converted into category, we have performed Ordinal encoding for Changing it to numerical values for modelling in our Data.

- **Is Scaling necessary here or not?**

  As we can refer below Data range graph and can Notice that data ranges are lying between 1 to 11 & most of them are ordinal so it has no meaning to scale the ordinal variables, so we are not doing scaling in this case.



(Fig: 27 Scaling Plot)

- **Data Split: Split the data into train and test (70:30)**

  First we have separated our target variable (Vate) from the data &we had split the data in into train test of (70:30) ratio by using python train_test_split function by passing below parameters

  For every model we have to first train that model and then test that model so we have split the data into train and test set by passing below parameters into train test split function

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.30, random_state=1)

Below are the split results

```
X_train:  (1061, 8)
X_test:  (456, 8)
y_train:  (1061,)
y_test:  (456,)
```

1.4) Apply Logistic Regression and LDA (linear discriminant analysis).

- **Apply Logistic Regression**

We have applied logistic regression by passing following parameters

model=LogisticRegression(solver='newtoncg',max_iter=10000,penalty='none', verbose=True, n_jobs=2)model.fit(X_train, y_train)

We have used Newton cg – Newton conjugate gradient method construction of the Model & we find below classification report by passing above parameters

- LR-Train Data classification report

	precision	recall	f1-score	support
Conservative	0.74	0.64	0.68	307
Labour	0.86	0.91	0.88	754
accuracy			0.83	1061
macro avg	0.80	0.77	0.78	1061
weighted avg	0.83	0.83	0.83	1061

(Table: 3 LR- Train Data)

- LR-Test Data Classification Report

	precision	recall	f1-score	support
Conservative	0.76	0.73	0.74	153
Labour	0.86	0.88	0.87	303
accuracy			0.83	456
macro avg	0.81	0.80	0.81	456
weighted avg	0.83	0.83	0.83	456

(Table: 4 LR- Test Data)

the above model is giving comparatively low accuracy, so we have used Grid Search CV for fi ne-tuning the model. Next step will be shown in further Questions

- **Apply LDA (Linear Discriminant Analysis)**

We have applied LDA Function by passing below parameters

LDA_model=LinearDiscriminantAnalysis()LDA_model.fit(X_train, y_train))

Following are default parameters for applying LDA

```
solver='svd',

    shrinkage=None,

    priors=None,

    n_components=None,

    store_covariance=False,

    tol=0.0001,

    covariance_estimator=None,
```

(Fig: 28 LDA)

Since every model is having its best parameters as default and model building is an iterative process, so we will construct model by default values and tune the model by various iterate ones We found out below Classification Report by performing LDA Model

- **LDA Train Classification report**

	precision	recall	f1-score	support
Conservative	0.74	0.65	0.69	307
Labour	0.86	0.91	0.89	754
accuracy			0.83	1061
macro avg	0.80	0.78	0.79	1061
weighted avg	0.83	0.83	0.83	1061

(Table: 5 LDA Train Data)

- **LDA test Classification report**

	precision	recall	f1-score	support
Conservative	0.78	0.73	0.75	153
Labour	0.87	0.89	0.88	303
accuracy			0.84	456
macro avg	0.82	0.81	0.81	456
weighted avg	0.84	0.84	0.84	456

(Table: 6 LDA Test Data)

looking at recall & precision, Accuracy is comparable in training and test data set for so we will fine tune our model by iterating various cut off probabilities and also used grid research for fine-tuning the model.

After Various iterations we came to know about the following cut-off probabilities.

Threshold	Accuracy	F-1 Score	Recall
0.1	0.761	0.854	0.987
0.2	0.791	0.868	0.968
0.3	0.812	0.878	0.952
0.4	0.832	0.888	0.939
0.5	0.833	0.885	0.907
0.6	0.826	0.876	0.87
0.7	0.833	0.877	0.838
0.8	0.79	0.835	0.751
0.9	0.697	0.737	0.599

(Table: 7 Cutoff Possibility)

As we can see that from above table that on 0.4 cut-off probability Our accuracy & F1 Scores Are Comparatively High So we will take 0.4 as our cut-off probabilities in the model.

- **Classification Report of the default cut-off test data:**

```
              precision    recall  f1-score   support

Conservative       0.78      0.73      0.75       153
      Labour       0.87      0.89      0.88       303

    accuracy                           0.84       456
   macro avg       0.82      0.81      0.81       456
weighted avg       0.84      0.84      0.84       456
```

(Table: 8 Default cutoff test data)

- **Classification Report of the custom cut-off test data:**

```
              precision    recall  f1-score   support

Conservative       0.88      0.59      0.73       153
      Labour       0.72      0.98      0.83       303

    accuracy                           0.84       456
   macro avg       0.80      0.61      0.81       456
weighted avg       0.77      0.73      0.84       456
```

(Table: 9 Custom cutoff test data)

as we can clearly see that accuracy has a fraction of increment which is not good so we will perform Grid search CV for further tuning.

	Train Recall	Test Recall	Accuracy Train	Accuracy Test
LR	0.910	0.884	0.830	0.831
LDA	0.907	0.894	0.833	0.838

(Table: 10 Grid search data)

As we can see that taking labor as our positive class recall and Accuracy scores are slightly better in case of LDA.

1.5) Apply KNN Model and Naïve Bayes Model. Interpret the results.

- **Apply KNN Model**

We have applied KNN Model by passing following parameters

KNN_model=KNeighborsClassifier(n_neighbors=7)
KNN_model.fit(X_train,y_train)

Since every model is having its best parameters as default and model building is an iterative process, so we will construct model by default values and tune the model by various iterations.

we find below classification report by passing above parameters

- **KNN -Train Data classification report**

	precision	recall	f1-score	support
Conservative	0.78	0.70	0.74	307
Labour	0.88	0.92	0.90	754
accuracy			0.86	1061
macro avg	0.83	0.81	0.82	1061
weighted avg	0.85	0.86	0.85	1061

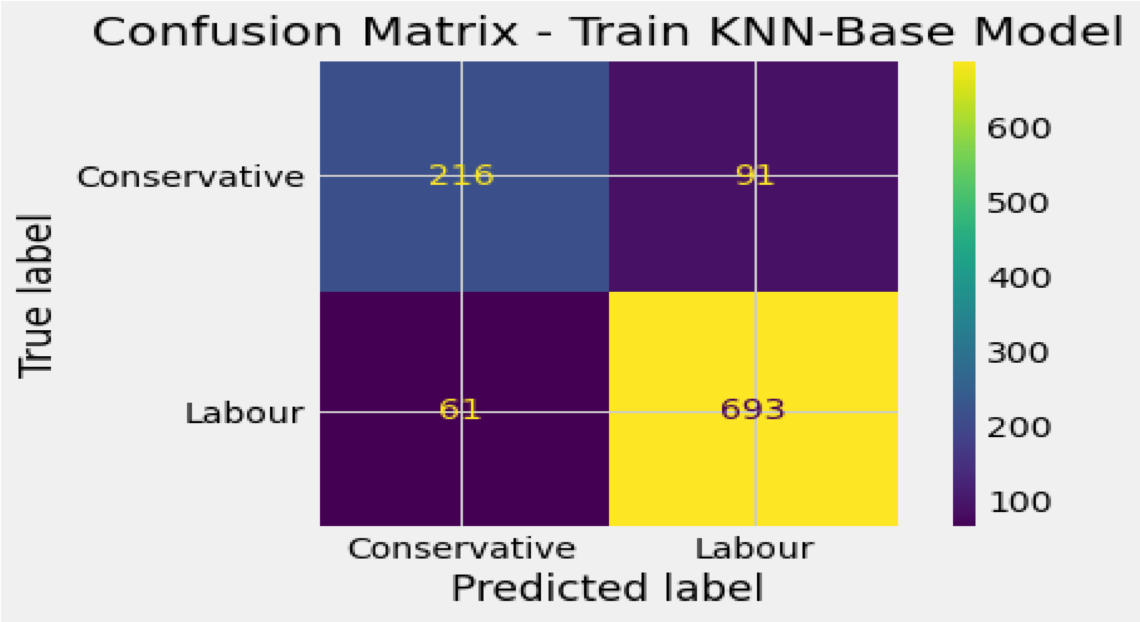
(Table: 11 KNN Train data)

- **KNN-Test Data Classification Report**

	precision	recall	f1-score	support
Conservative	0.74	0.64	0.69	153
Labour	0.83	0.89	0.86	303
accuracy			0.80	456
macro avg	0.79	0.76	0.77	456
weighted avg	0.80	0.80	0.80	456

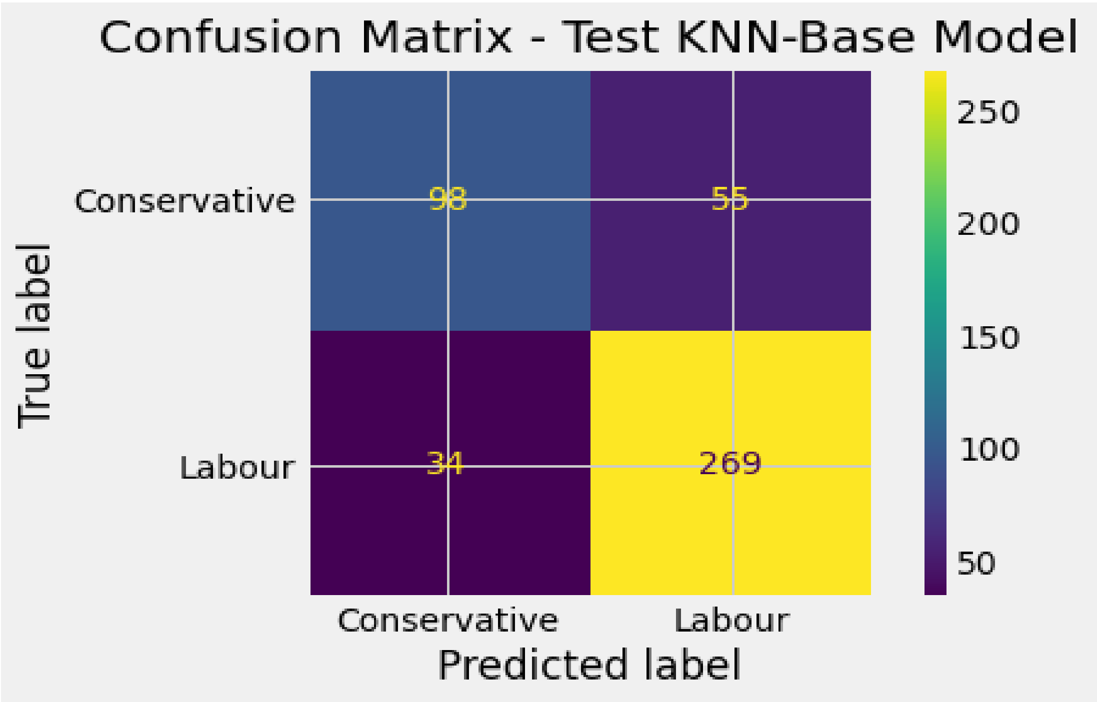
(Table: 12 KNN Test data)

the above model We have used 7 Nearest Neighbors and model is giving comparatively low accuracy so we will change the value of K nearest neighbors to finetuning the model. Tu Ning step will be shown in further Questions



(Fig: 29 KNN Train Metrix)

We have Correctly predicted 216 votes for conservative party and 693 votes for Labor party and 152 predictions are wrong in Train set



(Fig: 30 KNN Test Metrix)

We have Correctly predicted 98 votes for conservative party and 269 votes for Labor party and 89 predictions are wrong in Test set.

- **Apply Naïve Bayes Model.**

We have applied Naïve Bayes Model. by passing following parameters

NB_model = GaussianNB()NB_model.fit(X_train, y_train)

we find below classification report by passing above parameters, we have used default para meters to Construct Naïve bays model

- NB -Train Data classification report

	precision	recall	f1-score	support
Conservative	0.73	0.69	0.71	307
Labour	0.88	0.89	0.89	754
accuracy			0.84	1061
macro avg	0.80	0.79	0.80	1061
weighted avg	0.83	0.84	0.83	1061

(Table: 13 NB Train Data)

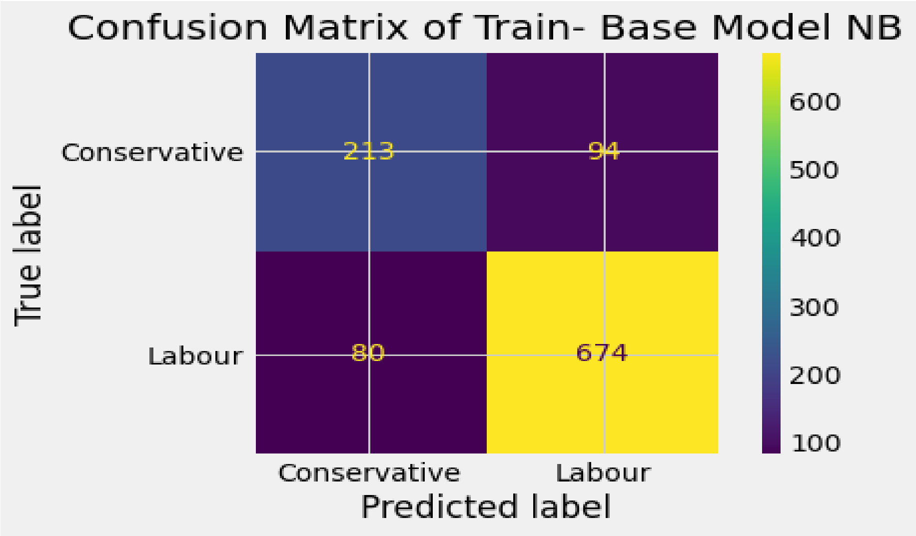
- NB -Test Data Classification Report

	Precision	recall	f1-score	support
Conservative	0.74	0.73	0.73	153
Labour	0.86	0.87	0.87	303
accuracy			0.82	456
macro avg	0.80	0.80	0.80	456
weighted avg	0.82	0.82	0.82	456

(Table: 14 NB Test Data)

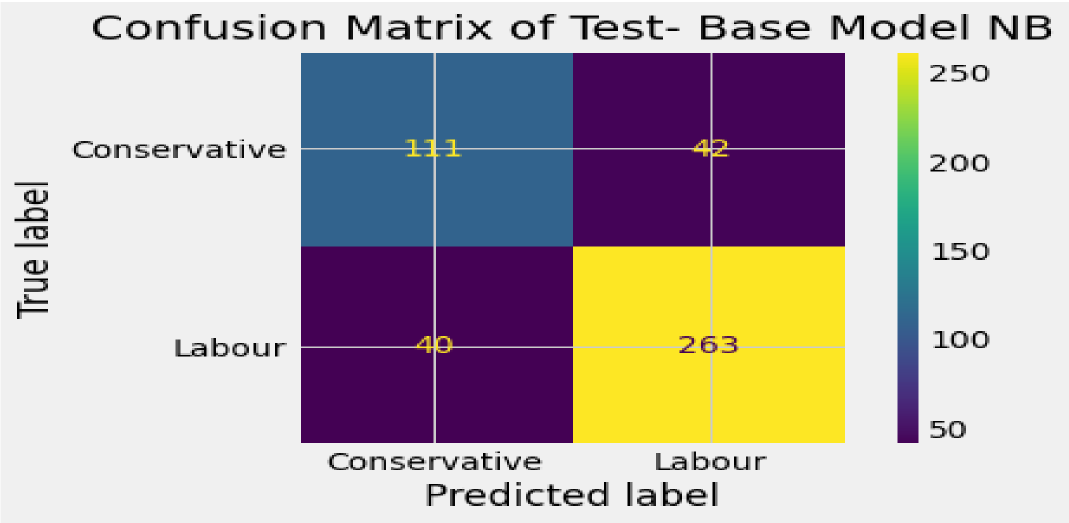
in the above naïve Bays model, we have used default parameters for Naïve bays model is giving comparatively low accuracy so we will change the value of default parameters to finetuning the model

- **Confusion Matrices of Naïve Bays Model**



(Fig: 31 NB Train Metrix)

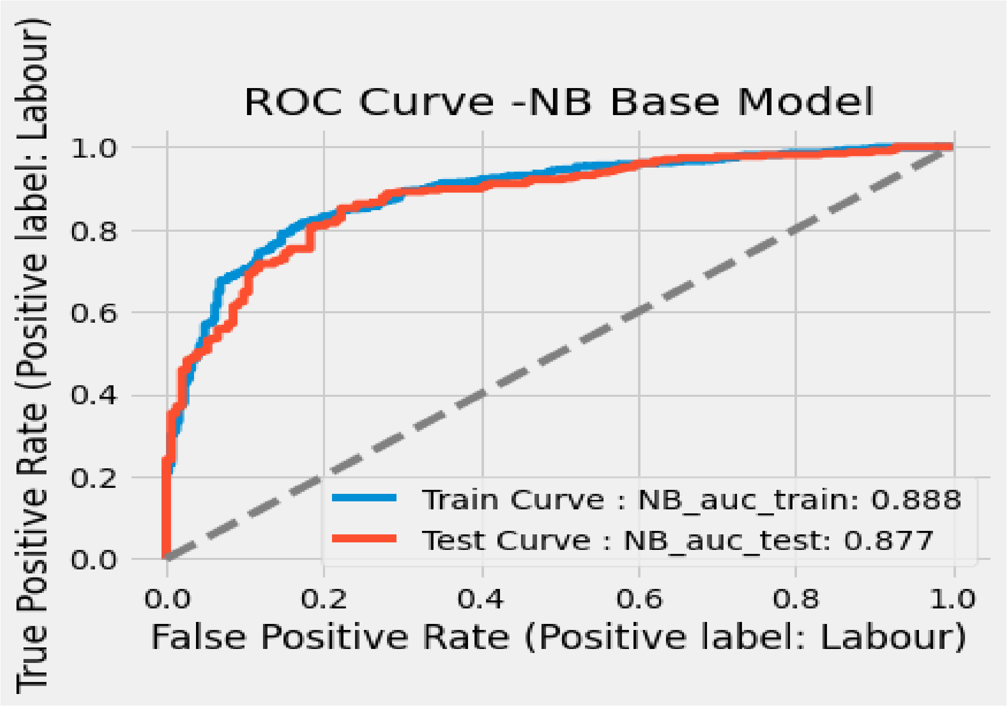
We have Correctly predicted 213 votes for conservative party and 674 votes for Labor part y and 174 predictions are wrong in Train set



(Fig: 32 NB Test Metrix)

We have Correctly predicted 116 votes for conservative party and 263 votes for Labour part y and 82 predictions are wrong in Test set.

- **ROC Curve for Naïve Bays Base model**



(Fig: 33 ROC Curve NB Base Model)

As we can see that Area under the Curve & ROC curve is having .88 & .87 in training and test set respectively so there is not much difference in the AUC score.

1.6) Model Tuning, Bagging (Random Forest should be applied for Bagging), and Boosting.

Model Tuning, Bagging and Boosting.

- **Model Tuning: - Logistic Regression.**

We have tuned the logistic regression Model by applying Grid search CV by passing followin g parameters

{'penalty':['l2','l1'], 'solver':['sag','lbfgs'], 'tol':[0.0001,0.00001]}

L1 – Lasso regression penalty & Lbfgs --Stands for Limited-memory Broyden–Fletcher–Goldf arb–Shanno. It approximates the second derivative matrix updates with gradient evaluation
s. It stores only the last few updates

```
GridSearchCV(cv=3, estimator=LogisticRegression(max_iter=10000, n_jobs=2),n_jobs=-1,param_grid={'penalty': ['l2', 'l1'], 'solver': ['sag', 'lbfgs'],'tol': [0.0001, 1e-05]} scoring='accuracy')
```

We have found below best parameters by using Grid search CV for Logistic Regression

```
{'penalty': 'l2', 'solver': 'sag', 'tol': 0.0001, max_iter=10000}
```

Our best parameters found are L2 – Ridge regression is applied When the issue of multicollin earity occurs penalty & solver is sag (Stochastic Average Gradient) in Stochastic Gradient D escent, a few samples are selected randomly instead of the whole data set for each iteration & tolerance is 0.001. (Refer Jupiter Notebook file for more details)

- **Model Tuning: - Linear Discriminant Analysis.**

We have tuned the LDA Model by applying Grid search CV by passing following

parameters grid={'solver':['lsqr','eigen'], 'n_components':[1,7,2]}

```
GridSearchCV(cv=5, estimator=LinearDiscriminantAnalysis(), n_jobs=-1,
             param grid={'n components': [1, 7, 2],
                         'solver': ['lsqr', 'eigen']})
```

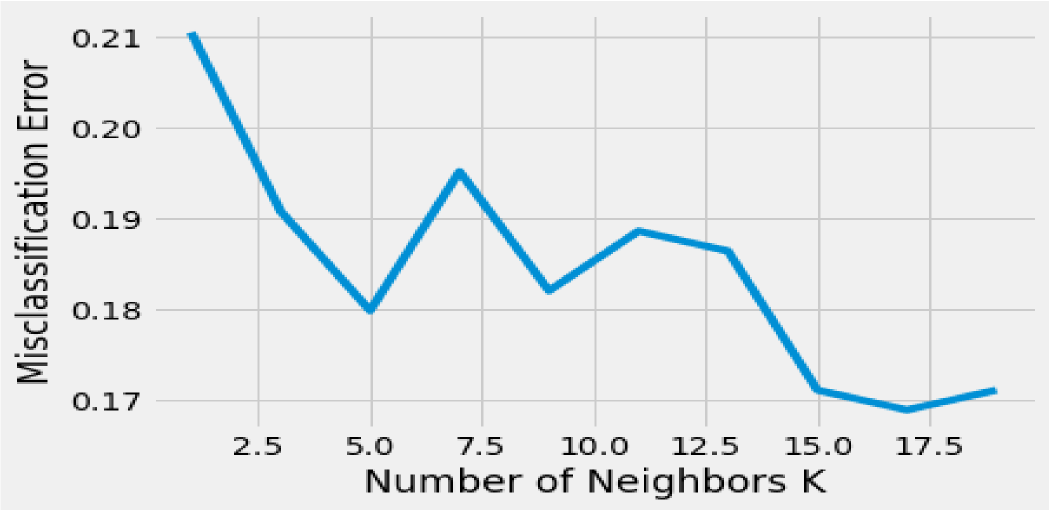
We have found below best parameters by using Grid search CV for LDA

```
(n_components=1, solver='lsqr')
```

Lsqr :- It is the solve which uses least square method for discriminate between two classes. (Refer Jupiter Notebook file for more details)

- **KNN Tuning**

We have performed KNN model in previous questions by choosing the best K- value where misclassification error should be Minimal So we have calculated errors for various k values and plotted them as below.



(Fig: 34 KNN Tuning)

Formula for MCE

Misclassification error (MCE) = 1 - Test accuracy score.

As we have seen from above plot misclassification Error is minimal on K=17 value . This mea ns we will find our best accuracy by considering 17 nearest neighbours.

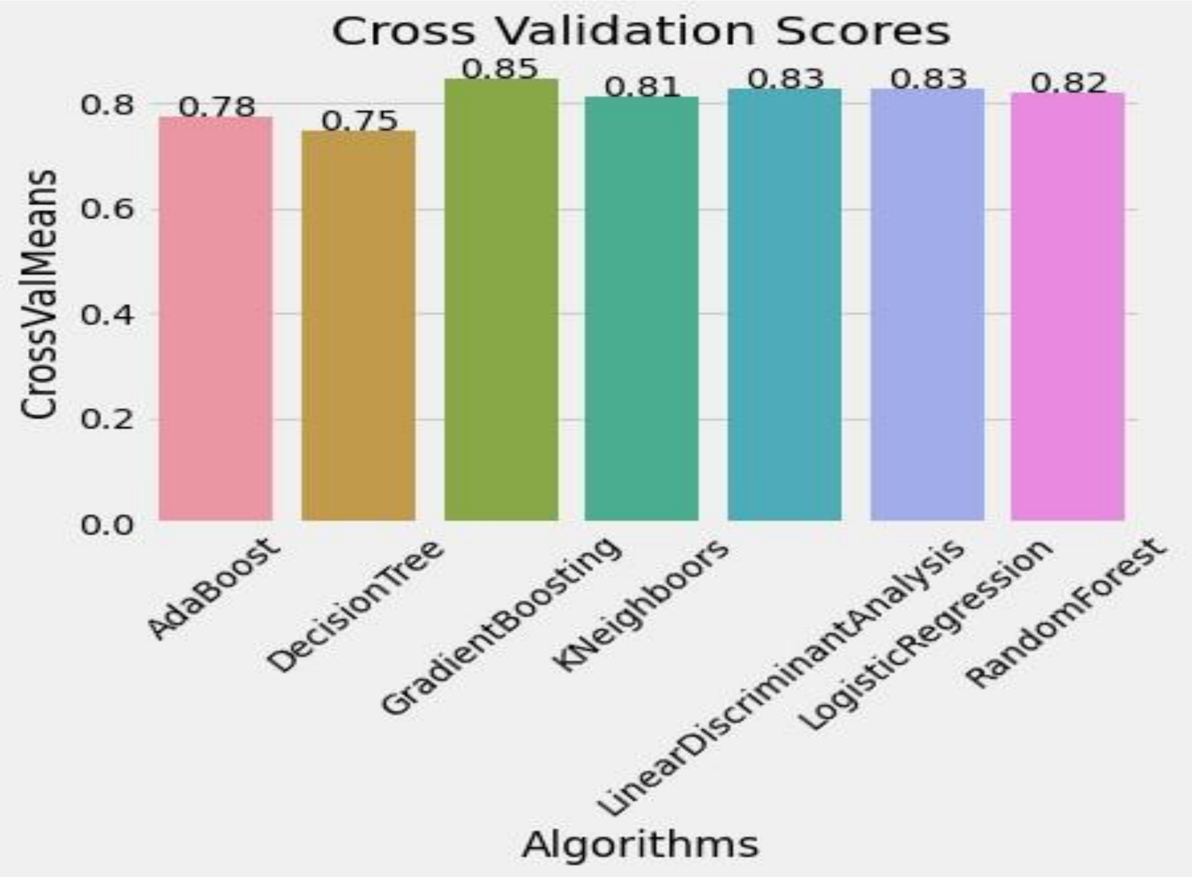
- ## Naïve Bays Model Tuning

We have tuned Naïve Bays model by bagging & cross validation as below

```
rec_scores =
cross_val_score(NB_SM_model,X_train_res,y_train_res,cv=10,scoring='accuracy')
```

after performing all the steps in jupyter notebook we found 83% accuracy.

We have also used Cross validation technique for model tuning and found below accuracy result in terms of Bar Graph



(Fig: 35 Cross Validation Scores)

We have also performed SMOTE on Naïve bays and KNN & Found below results

	Accuracy Train	Accuracy Test
Naive-Bayes SMOTE	0.8342	0.803
KNN SMOTE	0.8886	0.805

We have seen there is no significant effect of smote on any model and there is significant Difference between Accuracy for KNN as we know that SMOTE as a technique is generally applied if minority class is below 5%. But here are 27 & 73 percent.

- **Bagging (Random Forest should be applied for Bagging)**

We have used Random forest Model for applying Bootstrapped Aggregating by passing below parameters

```
(base_estimator=RandomForestClassifier(random_state=1),
 (n_estimators=100, random_state=1)
```

- **Bagging RF- Train Data Classification Report**

	precision	recall	f1-score	support
Conservative	0.98	0.89	0.93	307
Labour	0.96	0.99	0.97	754
accuracy			0.96	1061
macro avg	0.97	0.94	0.95	1061
weighted avg	0.96	0.96	0.96	1061

(Table: 15 RF Train Data)

- **Bagging RF- Test Data Classification Report**

	precision	recall	f1-score	support
Conservative	0.79	0.67	0.73	153
Labour	0.85	0.91	0.88	303
accuracy			0.83	456
macro avg	0.82	0.79	0.80	456
weighted avg	0.83	0.83	0.83	456

(Table: 16 RF Test Data)

With random forest Bagging we found out that Train data Accuracy is 96% and test data accuracy is 83% so there will be overfitting. Because our model is performing better Train set but comparatively less performing in Test Data.

- **Boosting: -**

We have Used Ada-Boost and Gradient Boost for modelling, AdaBoost Classifier for Adaboost and Gradient Boost Classifier for Gradient boost.

ADB_model = AdaBoostClassifier(n_estimators=100,random_state=1)

ADB_model.fit(X_train,y_train)

gbcl = GradientBoostingClassifier(random_state=1)

gbcl = gbcl.fit(X_train, y_train)

After performing all the above techniques, we found out below results.

	Train Recall	Test Recall	Accuracy Train	Accuracy Test
Naive-Bayes	0.894	0.868	0.836	0.820
LR	0.910	0.884	0.830	0.831
LDA	0.907	0.894	0.833	0.838
ADABoost	0.906	0.888	0.841	0.822
GradientBoost	0.934	0.904	0.887	0.831
KNN	0.902	0.908	0.839	0.831
Bagging	0.992	0.908	0.961	0.829

(Table: 17 Boosting Data Classification)

From above table, we found that Boosting & Bagging with Random forest and Naïve Bays accuracy, also we can Make inference that bagging model is overfitted model and Gradient AdaBoost is performing well in this case.

1.7) Performance Metrics: Check the performance of Predictions on Train and Test sets using Accuracy, Confusion Matrix, Plot ROC curve and get ROC_AUC score for each model. Final Model: Compare the models and write inference which model is best/optimized.

Since from Question both classes are important to us so we cannot use recall for checking the performance of the model, we will consider Accuracy as our performance parameters
Performance Metrics: Check the performance of Predictions on Train and Test sets using Accuracy, Confusion Matrix, Plot ROC curve and get ROC_AUC score for each model.

Logistic Regression Performance Matrices

- LR-Train Data classification report of tuned model

	precision	recall	f1-score	support
Conservative	0.74	0.64	0.68	307
Labour	0.86	0.91	0.88	754
accuracy			0.83	1061
macro avg	0.80	0.77	0.78	1061
weighted avg	0.83	0.83	0.83	1061

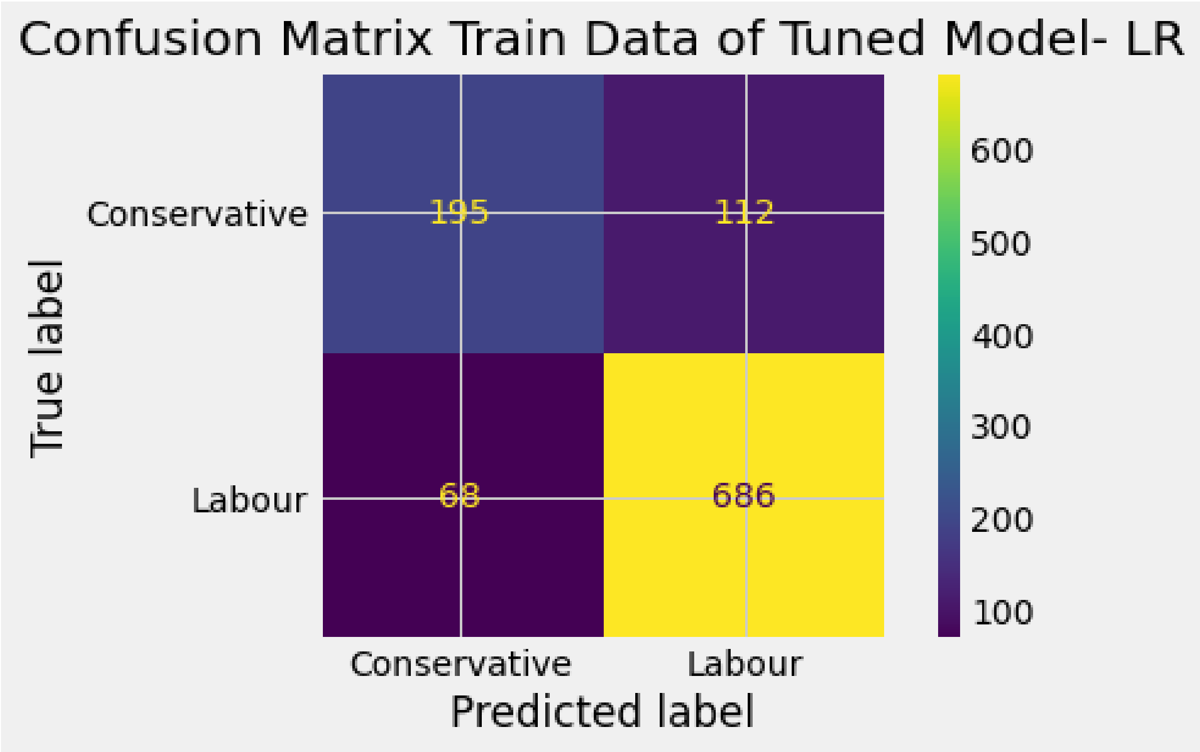
(Table: 18 LR Train Data Classification)

- LR-Test Data Classification Report of tuned model

	precision	recall	f1-score	support
Conservative	0.76	0.73	0.74	153
Labour	0.86	0.88	0.87	303
accuracy			0.83	456
macro avg	0.81	0.80	0.81	456
weighted avg	0.83	0.83	0.83	456

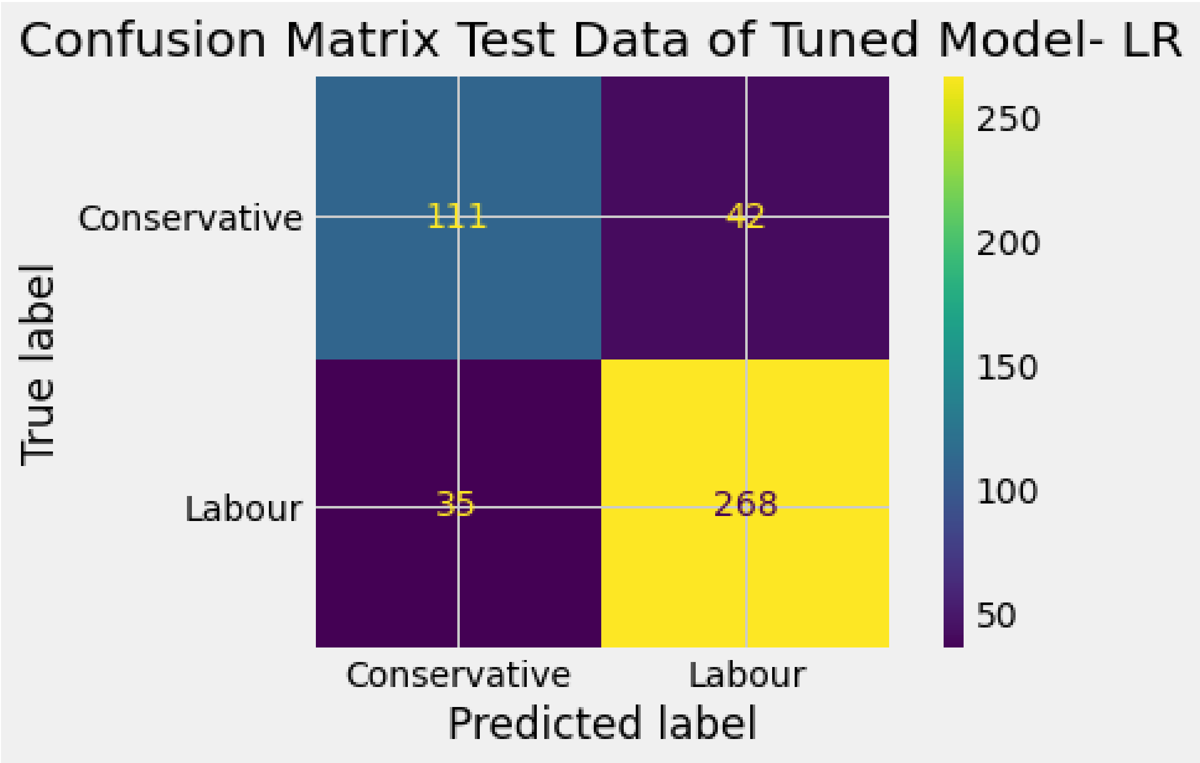
(Table: 19 LR Test Data Classification)

As we can see that Logistic Regression Model accuracy is comparable i.e 83% for both Train ng and test Data, so this model can be a good model.



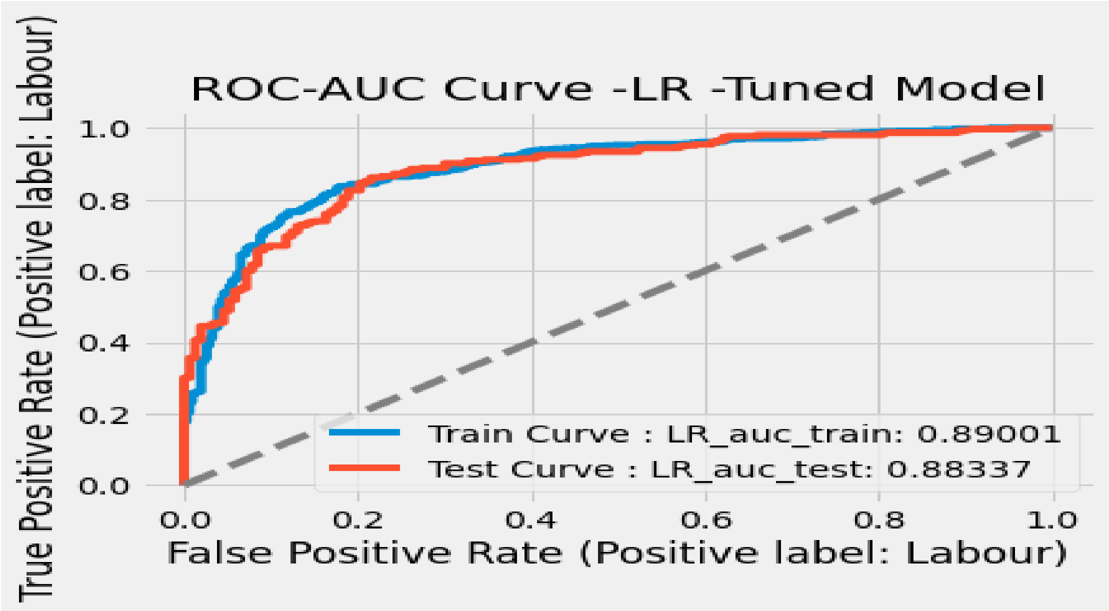
(Fig: 36 LR Train Data Matrix)

We have Correctly predicted 195 votes for conservative party and 686 votes for Labour part y and 180 predictions are wrong in Train set



(Fig: 37 LR Test Data Matrix)

We have Correctly predicted 111 votes for conservative party and 268 votes for Labour part y and 77 predictions are wrong in Test set.



(Fig: 38 Roc Auc Curve LR Tune Model)

From above ROC AUC Curve, we found that AUC for train is 89% and for test it is 88 % so we can say that 88% of the time model is performing good in test data.

• Linear Discriminant Analysis Performance metrics

After fitting the LDA Model and Tuning of the same we have found below performance matrices.

- LDA-Train Data classification report of tuned model

	precision	recall	f1-score	support
Conservative	0.74	0.65	0.69	307
Labour	0.86	0.91	0.89	754
accuracy			0.83	1061
macro avg	0.80	0.78	0.79	1061
weighted avg	0.83	0.83	0.83	1061

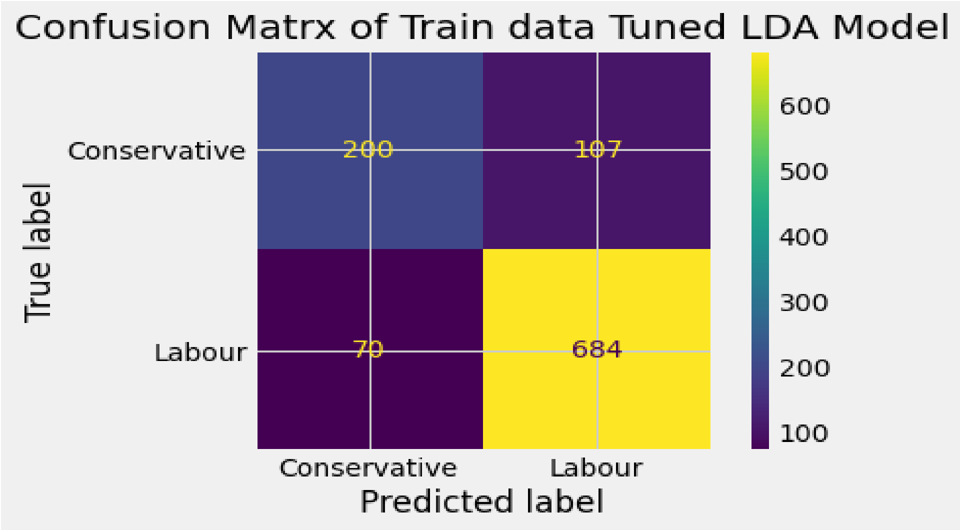
(Table: 20 LDA Train Data Classification)

- LDA-Test Data Classification Report of tuned model

	precision	recall	f1-score	support
Conservative	0.78	0.73	0.75	153
Labour	0.87	0.89	0.88	303
accuracy			0.84	456
macro avg	0.82	0.81	0.81	456
weighted avg	0.84	0.84	0.84	456

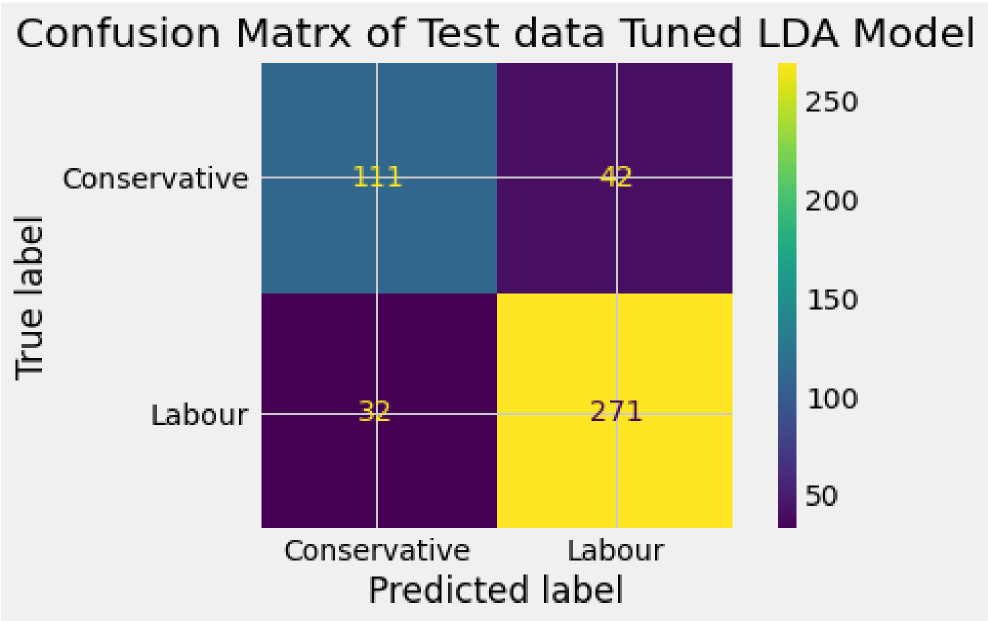
(Table: 21 LDA Test Data Classification)

As we can see that LDA Model accuracy is comparable i.e. 83% for Training and 84 % test Da ta, so we can further improve this by iterative methods.



(Fig: 39 LDA Train Data Matrix)

We have Correctly predicted 200 votes for conservative party and 684 votes for Labour part y and 177 predictions are wrong in Train set.

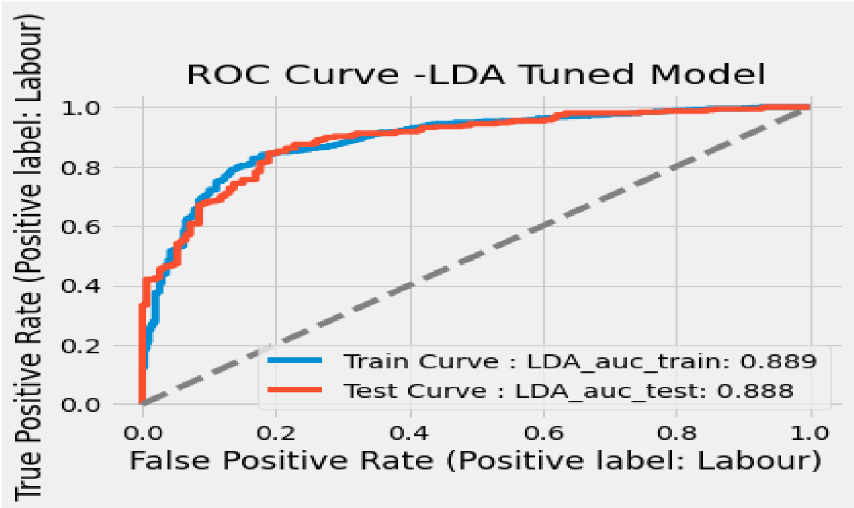


(Fig: 40 LDA Test Data Matrix)

We have Correctly predicted 111 votes for conservative party and 271 votes for Labour part y and 74 vote predictions are wrong in Test set.

- **ROC AUC Curve for Tuned Model**

From below ROC AUC curve, we noticed that AUC Score for Train and test Data is 88% thus we can say that model is performing good.



(Fig: 41 ROC AUC Curve LDA Model)

- KNN Classifier -Test Data Classification Report of tuned model

	precision	recall	f1-score	support
Conservative	0.79	0.68	0.73	153
Labour	0.85	0.91	0.88	303
accuracy			0.83	456
macro avg	0.82	0.79	0.80	456
weighted avg	0.83	0.83	0.83	456

(Table: 22 KNN Classifier Test Data)

- KNN Classifier -Train Data classification report of tuned model

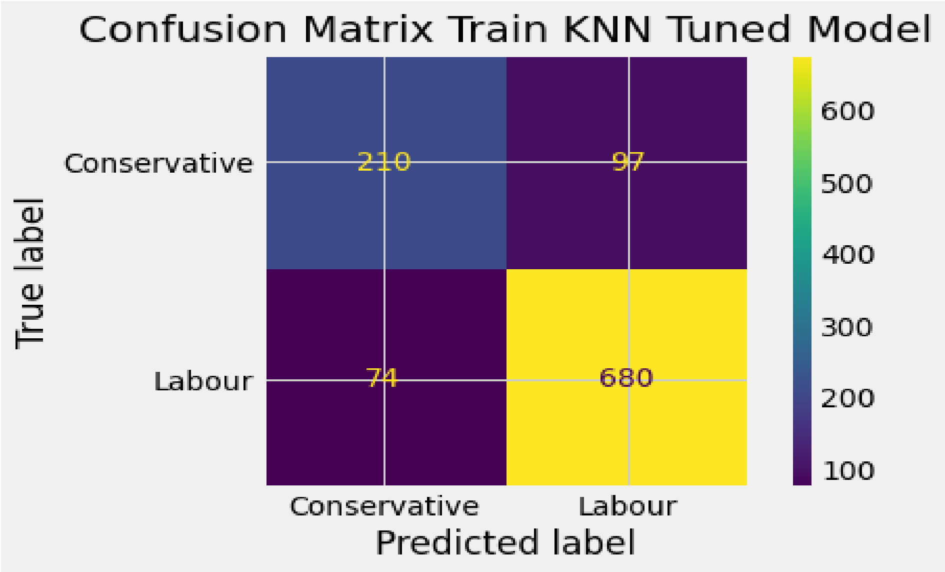
	precision	recall	f1-score	support
Conservative	0.74	0.68	0.71	307
Labour	0.88	0.90	0.89	754
accuracy			0.84	1061
macro avg	0.81	0.79	0.80	1061
weighted avg	0.84	0.84	0.84	1061

(Table: 23 KNN Classifier Train Data)

As we can see that KNN Model accuracy is comparable i.e. 84% for Training and 83 % test Da ta, so we can further improve this by iterative methods.

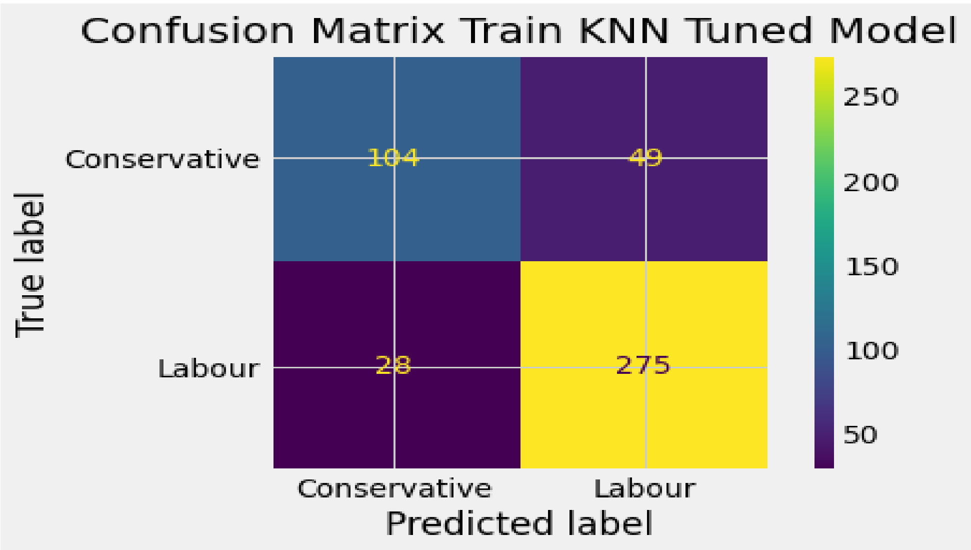
- **K- Nearest Neighbor (KNN Classifier) Performance Matrices**

After building KNN model we found Out below Performance metrics.



(Fig: 42 KNN Train Model Matrix)

We have Correctly predicted 210 votes for conservative party and 680 votes for Labour part y and 171 vote predictions are wrong in Train set.

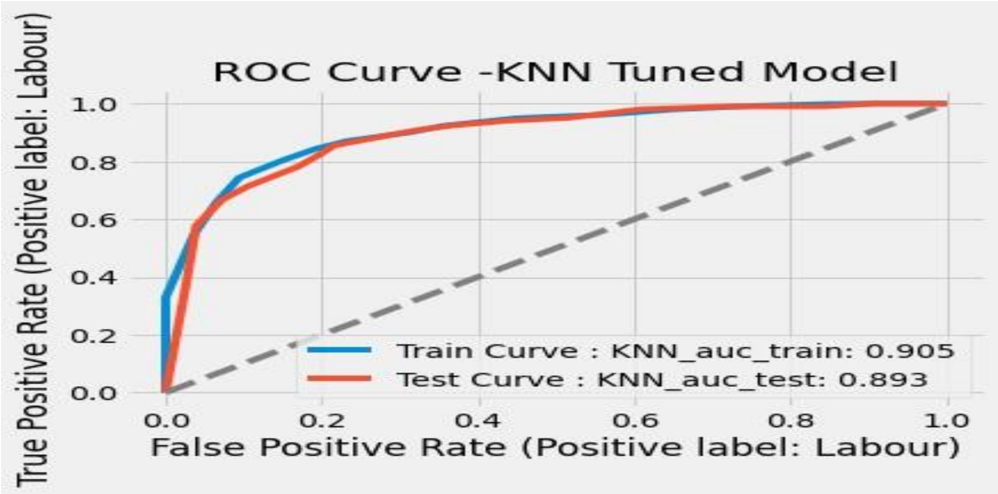


(Fig: 43 KNN Test Model Matrix)

We have Correctly predicted 104 votes for conservative party and 275 votes for Labour part y and 77 vote predictions are wrong in Test set.

- **ROC AUC Curve for Tuned Model**

From below ROC AUC curve, we noticed that AUC Score for Train is 90% and test Data is 89% which represent the degree of separability at various threshold settings thus we can say that model is 89 %. Capable of distinguishing between classes.



(Fig: 44 Roc Auc Curve KNN)

- **Naïve Bays Performance Matrices**

After building Naïve Bays model we found Out below Performance metrics.

- Naïve Bays -Train Data classification report of tuned model

	precision	recall	f1-score	support
Conservative	0.73	0.69	0.71	307
Labour	0.88	0.90	0.89	754
accuracy			0.84	1061
macro avg	0.80	0.79	0.80	1061
weighted avg	0.83	0.84	0.83	1061

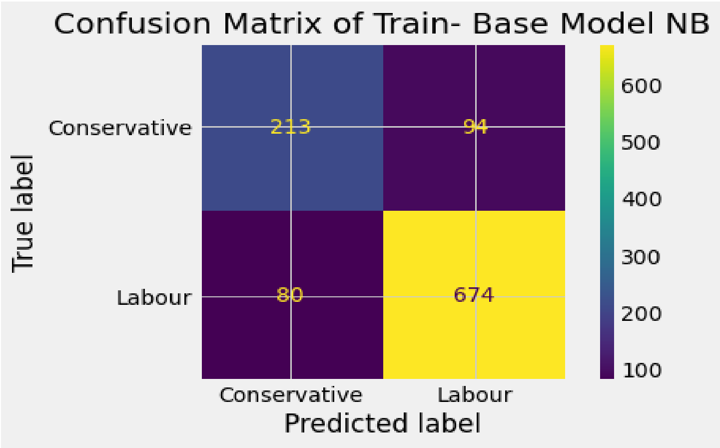
(Table: 24 NB Train Data Classification)

- Naïve Bays Classifier -Test Data Classification Report of tuned model

	precision	recall	f1-score	support
Conservative	0.74	0.73	0.73	153
Labour	0.86	0.87	0.87	303
accuracy			0.82	456
macro avg	0.80	0.80	0.80	456
weighted avg	0.82	0.82	0.82	456

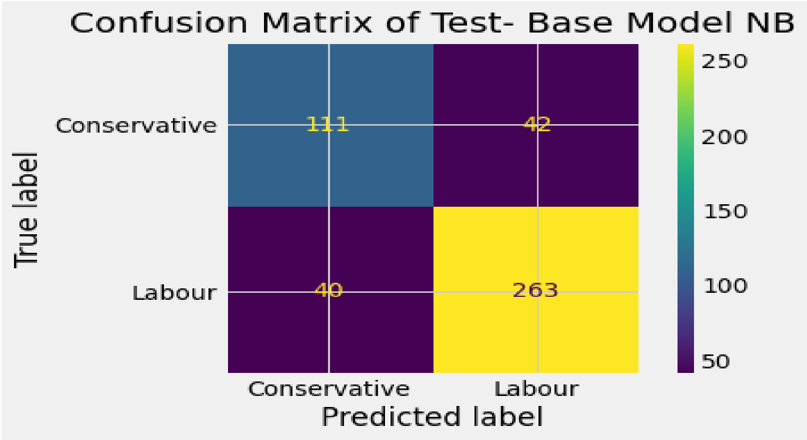
(Table: 25 NB Test Data Classification)

As we can see that Naïve Bays Model accuracy is comparable i.e. 84% for Training and 82 % t est Data, so we can this means model is giving 82% accurate results on testing.



(Fig: 45 NB Train Data Matrix)

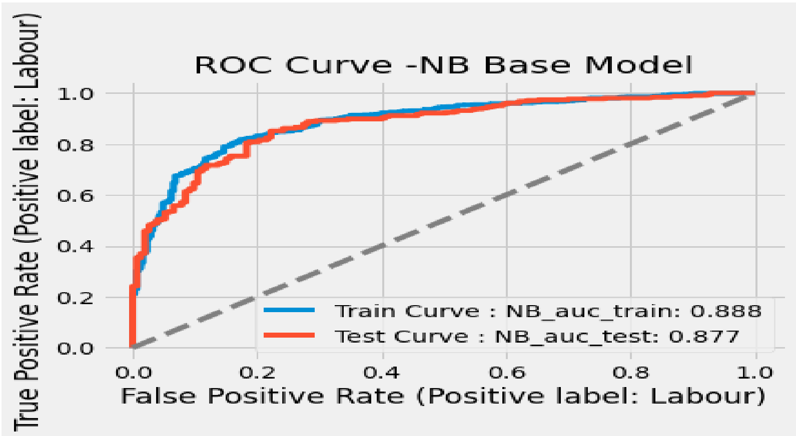
We have Correctly predicted 213 votes for conservative party and 674 votes for Labour part y and 174 vote predictions are wrong in Train set.



(Fig: 46 NB Test Data Matrix)

We have Correctly predicted 111 votes for conservative party and 263 votes for Labour part y and 72 vote predictions are wrong in Test set.

- **ROC AUC Curve for Tuned Model**

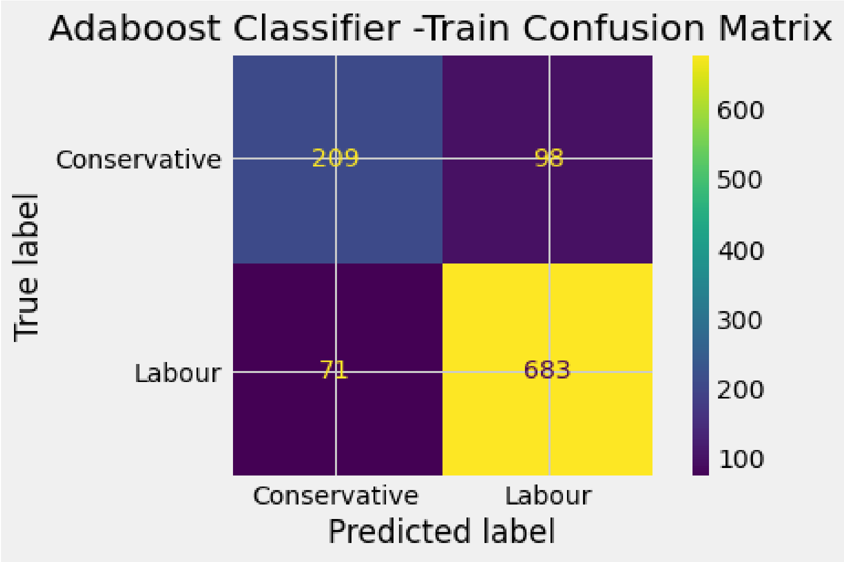


(Fig: 47 Roc Auc Curve NB)

From below ROC AUC curve, we noticed that AUC Score for Train is 88% and test Data is 87% which represent the degree of separability at various threshold settings thus we can say that model is 87 %. Capable of distinguishing between classes.

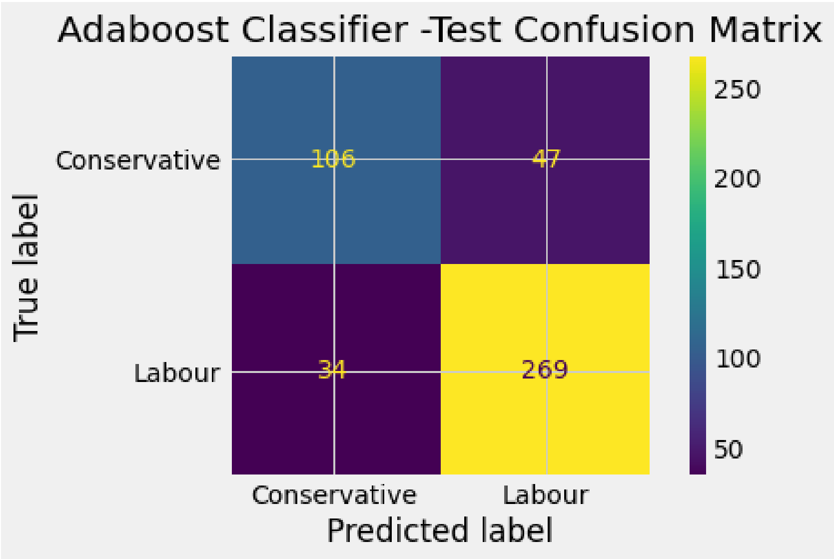
- **AdaBoost Performance Matrices**

After building Adaboost we found Out below Performance metrics.



(Fig: 48 AdaBoost Train Data Matrix)

We have Correctly predicted 209 votes for conservative party and 873 votes for Labour part y and 169 vote predictions are wrong in Train set.



(Fig: 49 AdaBoost Test Data Matrix)

We have Correctly predicted 106 votes for conservative party and 269 votes for Labour party and 81 vote predictions are wrong in Test set.

- **AdaBoost -Train Data classification report**

	precision	recall	f1-score	support
Conservative	0.75	0.68	0.71	307
Labour	0.87	0.91	0.89	754
accuracy			0.84	1061
macro avg	0.81	0.79	0.80	1061
weighted avg	0.84	0.84	0.84	1061

(Table: 26 AdaBoost Train Data Classification)

- **AdaBoost -Test Data Classification Report**

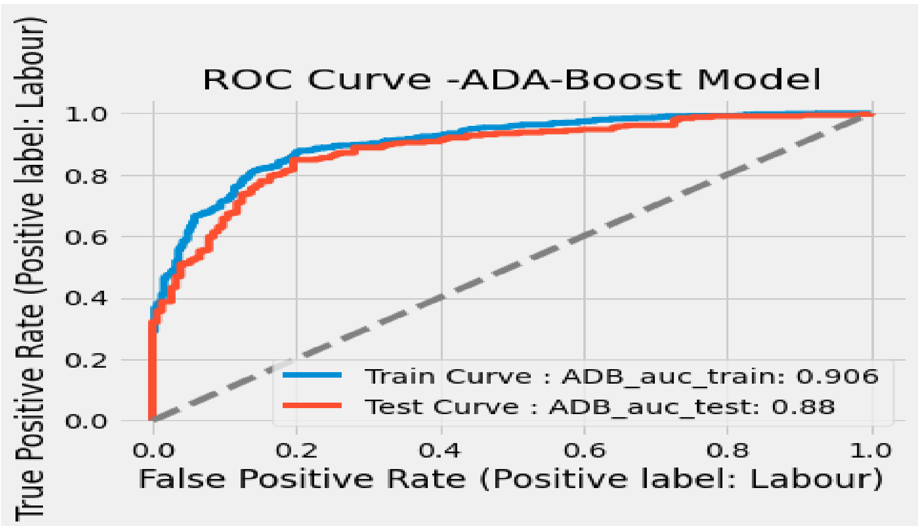
	precision	recall	f1-score	support
Conservative	0.76	0.69	0.72	153
Labour	0.85	0.89	0.87	303
accuracy			0.82	456
macro avg	0.80	0.79	0.80	456
weighted avg	0.82	0.82	0.82	456

(Table: 27 AdaBoost Test Data Classification)

As we can see that AdaBoost accuracy is comparable i.e. 84% for Training and 82 % test Data, so we can this means model is giving 82% accurate results on testing.

- **ROC AUC Curve for Tuned Model**

From below ROC AUC curve, we noticed that AUC Score for Train is 90% and test Data is 88% which represent the degree of separability at various threshold settings thus we can say that model is 88 %. Capable of distinguishing between classes.



(Fig: 50 Roc Auc Curve AdaBoost)

- **Bagging Performance Matrices**

After building Bagging we found Out below Performance metrics.

- **Bagging RF- Train Data Classification Report**

	precision	recall	f1-score	support
Conservative	0.98	0.89	0.93	307
Labour	0.96	0.99	0.97	754
accuracy			0.96	1061
macro avg	0.97	0.94	0.95	1061
weighted avg	0.96	0.96	0.96	1061

(Table: 28 RF Train Data Classification)

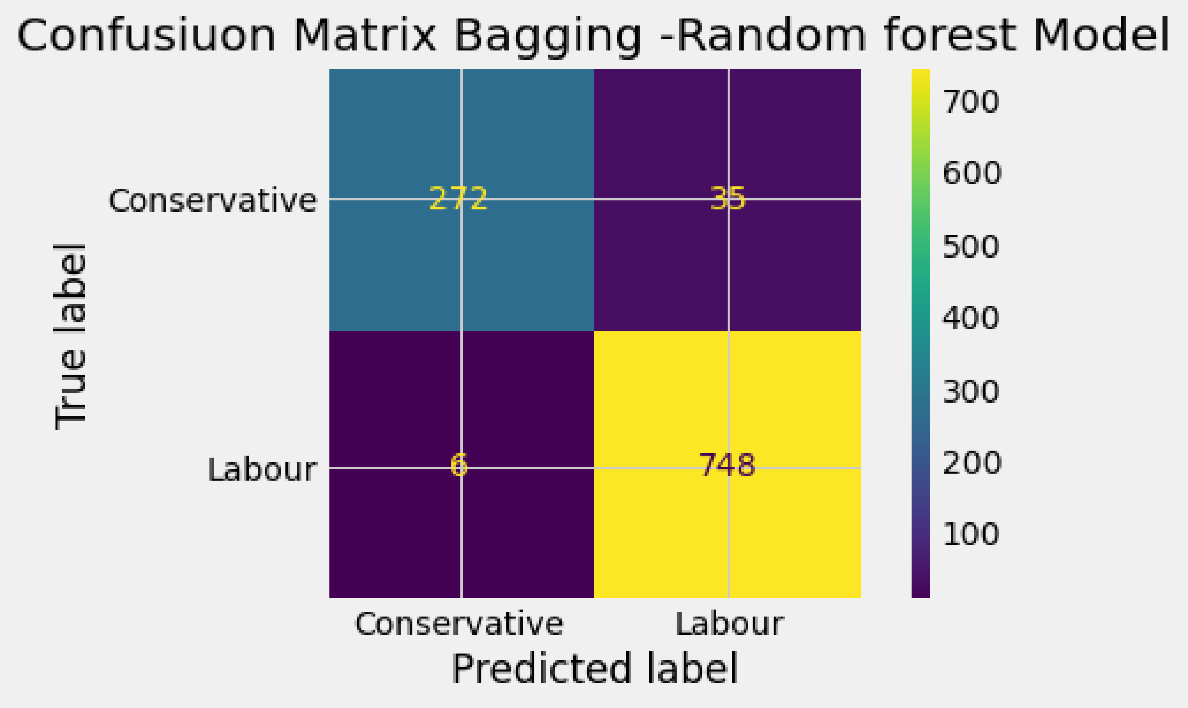
- **Bagging RF- Test Data Classification Report**

	Precision	recall	f1-score	support
Conservative	0.79	0.67	0.73	153
Labour	0.85	0.91	0.88	303
accuracy			0.83	456
macro avg	0.82	0.79	0.80	456
weighted avg	0.83	0.83	0.83	456

(Table: 29 RF Test Data Classification)

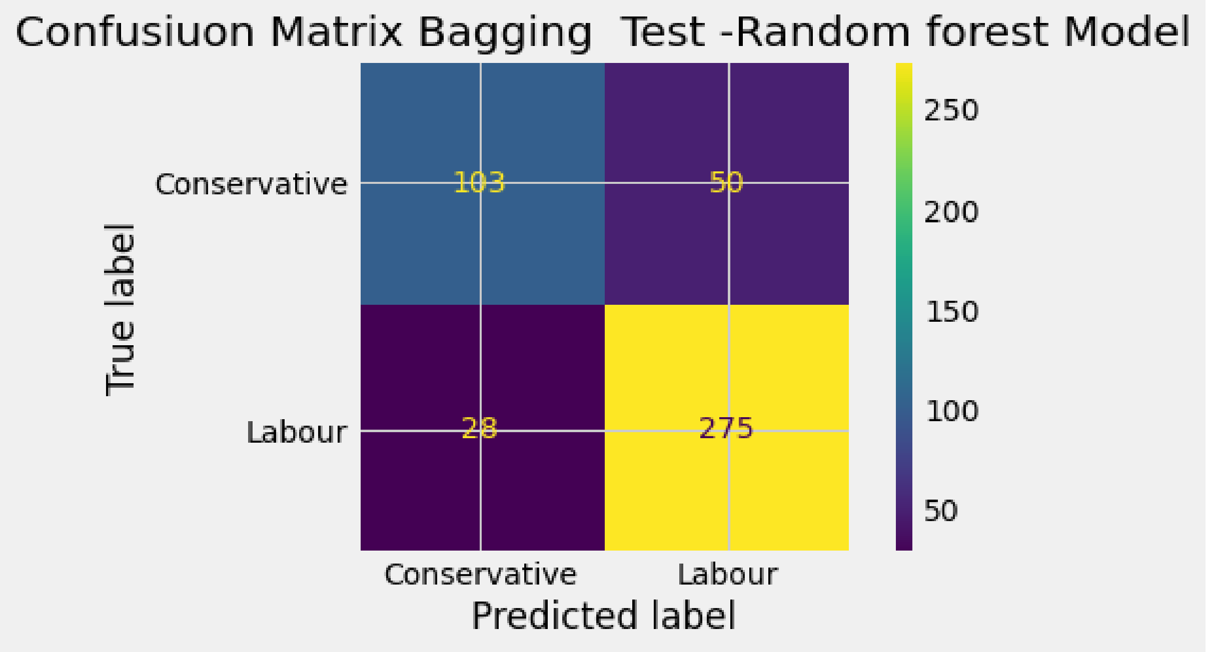
With random forest Bagging we found out that Train data Accuracy is 96% and test data accuracy is 83% so there might be overfitting. Because our model is performing better Train set but comparatively less performing in Test data.

As we can see that Bagging with Random Forest accuracy is comparable i.e. 84% for Training and 82 % test Data, so we can this means model is giving 83% accurate results on testing.



(Fig: 51 RF Train Data Matrix)

We have Correctly predicted 272 votes for conservative party and 748 votes for Labour part y and 41 vote predictions are wrong in Train set.

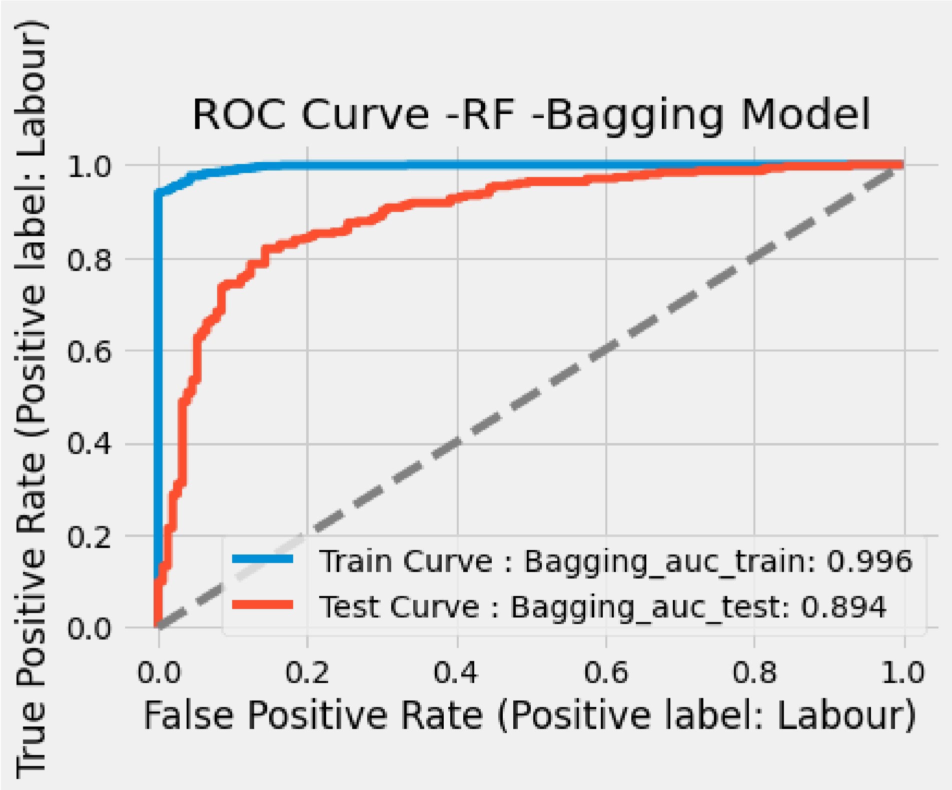


(Fig: 52 RF Test Data Matrix)

We have Correctly predicted 103 votes for conservative party and 275 votes for Labour part y and 78 vote predictions are wrong in Test set.

- **ROC AUC Curve for Tuned Model**

From below ROC AUC curve we noticed that AUC Score for Train is 99% and test Data is 89% which represent the degree of seperability at various threshold settings thus we can say that model is 89 %. Capable of distinguishing between classes.



(Fig: 53 Roc Auc Curve RF)

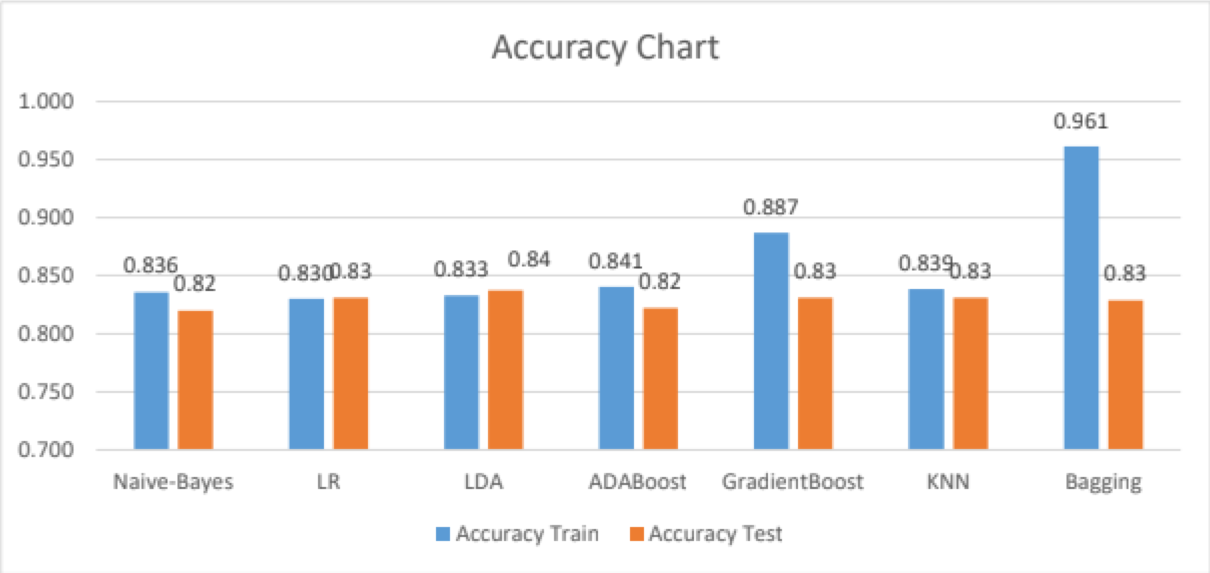
Final Model - Compare all models based on the performance metrics in a structured tabular manner. Describe on which mod el is best/optimized.

After building all the Models we have found below Table for performance metrices.

	Train Recall	Test Recall	Accuracy Train	Accuracy Test
Naive-Bayes	0.894	0.868	0.836	0.82
LR	0.910	0.884	0.830	0.83
LDA	0.907	0.894	0.843	0.84
ADABoost	0.906	0.888	0.841	0.82
GradientBoost	0.934	0.904	0.887	0.83
KNN	0.902	0.908	0.839	0.83
Bagging	0.992	0.908	0.961	0.83

(Table: 30 Performance Matrix table)

As we can notice that Naïve Bays Test accuracy is 82% and LR train& Test accuracy is 83% LD A Train and test accuracy is 84% ADA Boost Test accuracy 82%, Gradient boost, KNN & Bagging accuracy is 83%.



(Fig: 54 Accuracy Chart)

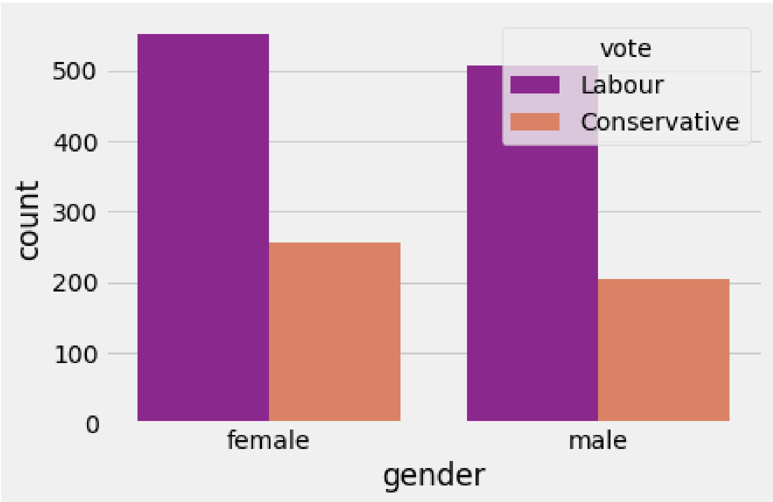
from above Bar Graph LDA model is giving comparatively Higher test Accuracy & Training accuracy so we can say that LDA Model is best optimized Model.

1.8) Based on these predictions, what are the insights?

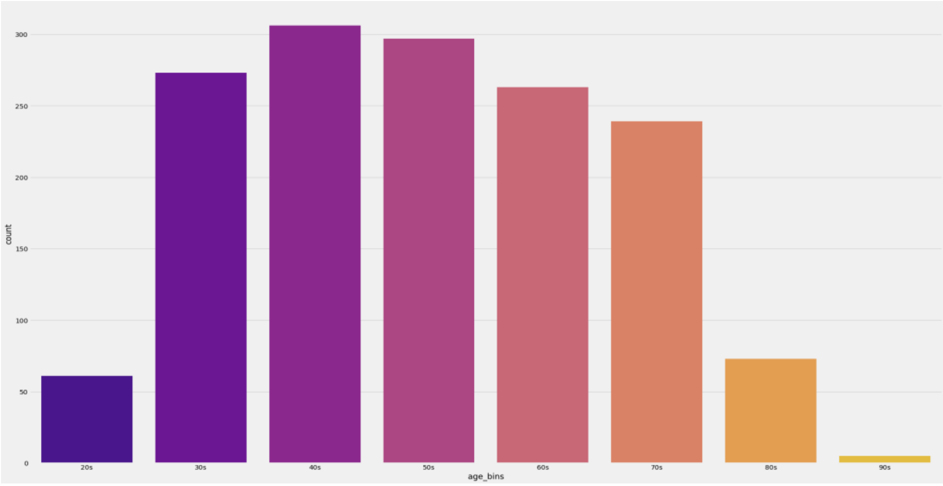
As per our business problem we must predict which party a voter will vote for on the basis of the given information, to solve this problem we have made Logistic Regression, Linear Discriminant Analysis Model, k- Nearest neighbor & Naïve models in terms of accuracy. Bays Model to find out the best results LDA is performing slightly better than other, as we have seen from heat map that there is very less correlation between the variables which is good for the model.

Referring to below graphs we can notice the following:

- Voters between Age group of 30s to 70s are voting more.
- Voters in their 20s & 80s, 90 are voting significantly very less
- Significantly More no. of females have voted for Labour party
- Most of the Peoples, who have Eurosceptic attitudes given vote to the Labor Party.
- Labor party has got more votes than Conservative party.



(Fig: 55 Gender Vs. Count Plot)



(Fig: 56 Age_bins Vs. Count Plot)

Recommendations: -

- Collect more data like ratings on their previous leadership qualities (How they have performed previously), Religion of the respondent etc. to gain more insight.
- CNBE can take Online surveys so that it can reduce their actual cost on surveys in res ult they can collect more data.
- CNBE can also give free eBooks or online Coupons to the voters if they participate in surveys.
- Company can collect the ratings on the attitude of leader towards Current issues

Problem 2:

In this particular project, we are going to work on the inaugural corpora from the nltk in Python. We will be looking at the following speeches of the Presidents of the United States of America:

1. President Franklin D. Roosevelt in 1941
2. President John F. Kennedy in 1961
3. President Richard Nixon in 1973

- Find the number of characters, words and sentences for the mentioned documents.

(Hint: use .words(), .raw(), .sent() for extracting counts)

Code Snippet to extract the three speeches:

"

```
import nltk
nltk.download('inaugural')

from nltk.corpus import inaugural
inaugural.fileids()
inaugural.raw('1941-Roosevelt.txt')
inaugural.raw('1961-Kennedy.txt')
inaugural.raw('1973-Nixon.txt')
```

"

2.1) Find the number of characters, words, and sentences for the mentioned documents.

(Hint: use. words (), .raw(), .sent() for extracting counts)

- **Number of Characters in each text file of speeches of the Presidents of the United States of America**

By using Len function, we have counted No. of character in each text file as below result

Number of Characters in Roosevelt Speech: 7571

Number of Characters in Kennedy Speech: 7618

Number of Characters in Nixon Speech: 9991

- **Number of Words in speeches of the Presidents of the United States of America.**

By using inaugural. Words function we found below result

Number of Words in Roosevelt Speech: 1536
Number of Words in Kennedy Speech: 1546
Number of Words in Nixon Speech: 2028

- **Number of Sentences in the speeches of the Presidents of the United States of America.**

BY using sent for .word count method we found bellow result.

Number of Sentences in Roosevelt Speech: 68
Number of Sentences in Kennedy Speech: 52
Number of Sentences in Nixon Speech: 69

2.2) Remove all the stop words from the three speeches.

- Removal of Stop words from 1941-Roosevelt Speech

Roosevelt Speech Initial Look

	Text	count_stop
0	On each national day of inauguration since 1782.	654

In Roosevelt speech 654 words comes under stop words, We have added punctuation & '--' as a stop word and these needs to be removed from Speech. They are given below

['each', 'of', 'the', 'have', 'their', 'of', 'to', 'the', 'the', 'of',

'the', 'was', 'to', 'and', 'a', 'the', 'of', 'the', 'was', 'to', 'that'

, 'from', 'from', 'this', 'the', 'of', 'the', 'is', 'to', 'that', 'and'

, 'its', 'from', 'from', 'us', 'there', 'has', 'a', 'in', 'the', 'of',

'to', 'for', 'a', 'and', '--', 'to', 'what', 'our', 'in', 'has', 'and',

'to', 'what', 'we', 'are', 'and', 'what', 'we', 'we', 'do', 'we', 'the'

, 'of', 'of', 'are', 'not', 'by', 'the', 'of', 'but', 'by', 'the', 'of'

, 'the', 'of', 'a', 'is', 'and', 'a', 'a', 'of', 'a', 'is', 'the', 'of'

, 'the', 'of', 'its', 'will', 'to', 'are', 'who', 'are', 'who', 'that',

'as', 'a', 'of', 'and', 'a', 'of', 'is', 'or', 'by', 'a', 'of', 'and',

'for', 'some', 'and', 'have', 'the', 'of', 'the', '--', 'and', 'that',

'is', 'an', 'we', 'that', 'this', 'is', 'not', 'when', 'the', 'of', 'th

is', 'by', 'a', 'we', 'that', 'this', 'is', 'not', 'were', 'in', 'the',

'of', '--', 'but', 'we', 'have', 'been', '--', 'for', 'the', 'of', 'thi

s', 'they', 'have', 'to', 'us', 'a', 'that', 'are', 'to', 'be', 'in', '

other', 'than', 'to', 'our', 'and', 'our', 'is', 'this', 'of', 'a', 'wh

ich', 'at', 'on', 'through', 'it', 'the', 'of', 'its', 'has', 'been', '

the', 'of', 'the', 'of', 'the', 'of', 'the', 'to', 'of', 'of', 'is', 'o

f', 'the', 'of', 'have', 'their', 'to', 'is', 'not', 'it', 'because', '

we', 'have', 'it', 'it', '--', 'because', 'it', 'is', 'on', 'the', 'of'

, 'and', 'in', 'a', '--', 'an', 'and', 'through', 'by', 'the', 'of', 'a

', 'it', 'because', 'of', 'all', 'of', 'the', 'of', 'it', 'because', 'h

as', 'an', 'of', 'in', 'the', 'of', 'it', 'if', 'we', 'below', 'the', '

we', 'it', 'on', '--', 'for', 'it', 'is', 'the', 'most', 'the', 'most',

'and', 'in', 'the', 'the', 'most', 'of', 'all', 'of', 'a', 'has', 'a',

'that', 'the', 'and', 'the', 'of', 'its', '--', 'all', 'the', 'other',
'that', 'the', 'of', 'the', 'a', 'a', 'has', 'more', 'than', 'the','of
', 'all', 'its', 'is', 'that', 'which', 'most', 'to', 'its', '--','whi
ch', 'the', 'most', 'of', 'its', 'is', 'a', 'for', 'which', 'we','it',
'--', '--', 'to', 'a', 'we', 'all', 'what', 'it', 'is', '--', 'the','-
-', 'the', 'of', 'is', 'the', 'of', 'was', 'in', 'the', 'of', 'those',
'who', 'from', '--', 'some', 'of', 'but', 'who', 'and', 'to', 'more','
is', 'no', 'in', 'is', 'the', 'of', 'in', 'the', 'was', 'in', 'the','i
ts', 'has', 'been', 'has', 'been', 'the', 'in', 'all', 'to', 'all','no
t', 'because', 'this', 'was', 'a', 'but', 'because', 'all', 'those','w
ho', 'here', 'they', 'this', 'a', '--', 'a', 'that', 'should', 'be','i
n', 'was', 'into', 'our', 'own', 'into', 'the', 'of', 'into', 'the','o
f', 'the', 'into', 'the', 'who', 'here', 'to', 'out', 'the', 'of','the
ir', 'and', 'the', 'who', 'and', 'the', 'that', 'from', 'them', '--','
all', 'have', 'and', 'an', 'which', 'in', 'itself', 'has', 'and','with
', 'each', 'of', 'the', 'or', 'that', 'we', 'have', 'to', 'that','we',
'more', 'the', 'and', 'the', 'and', 'the', 'of', 'in', 'the', 'by','th
e', 'and', 'the', 'of', 'the', 'it', 'is', 'not', 'to', 'these','is',
'not', 'to', 'and', 'the', 'of', 'this', 'and', 'and', 'its','there',
'is', 'the', 'of', 'the', 'the', 'is', 'the', 'the', 'and', 'the','as'
, 'all', 'the', 'not', 'if', 'the', 'of', 'were', 'the', 'and', 'in','
an', 'the', 'we', 'have', '--', 'that', '--', 'to', 'us', 'in', 'our',
'in', 'because', 'they', 'so', 'to', 'us', 'here', 'in', 'the', 'of','
the', 'to', 'us', 'through', 'the', 'of', 'in', 'the', 'of', 'to','us'
, 'in', 'our', 'in', 'our', 'in', 'our', 'and', 'in', 'our', 'to','us'
, 'from', 'the', 'other', 'of', 'the', 'and', 'from', 'those', 'the','
--', 'the', 'as', 'as', 'the', 'we', 'to', 'or', 'these', 'of','becaus
e', 'to', 'us', 'the', 'of', 'our', 'is', 'such', 'an', 'of', 'was','i
n', 'of', 'by', 'our', 'in', 'his', 'in', '--', 'it', 'to', 'this','of
', 'of', 'the', 'of', 'and', 'the', 'of', 'the', 'of', 'are', 'on','th
e', 'to', 'the', 'of', 'the', 'we', 'that', 'we', 'let', 'it', 'be','w
ith', 'and', '--', 'then', 'we', 'the', 'which', 'so', 'and', 'so','to
', 'of', 'the', 'and', 'of', 'the', 'and', 'the', 'for', 'that', 'we',
'in', 'the', 'of', 'the', 'of', 'before', 'our', 'is', 'to', 'and','to
', 'the', 'of', 'this', 'we', 'the', 'of', 'and', 'the', 'of', 'do','n
ot', 'are', 'not', 'to', 'we', 'in', 'the', 'of', 'our', 'by', 'the', ' will', 'of']

- **Roosevelt Speech after removal of stop words**

'national day inauguration since people renewed sense dedication united states washingtons day task people create weld together nation lincolns day task people preserve nation disruption wi thin day task people save nation institutions disruption without us come time midst swift happ enings pause moment take stock recall place history rediscover may risk real peril inaction li ves nations determined count years lifetime human spirit life man threescore years ten little.

little less life nation fullness measure live men doubt men believe democracy form government frame life limited measured kind mystical artificial fate unexplained reason tyranny slavery b ecome surging wave future freedom ebbing tide americans know true eight years ago life republi c seemed frozen fatalistic terror proved true midst shock acted acted quickly boldly decisivel y later years living years fruitful years people democracy brought us greater security hope be tter understanding lifes ideals measured material things vital present future experience democ racy successfully survived crisis home put away many evil things built new structures enduring lines maintained fact democracy action taken within threeway framework constitution united sta tes coordinate branches government continue freely function bill rights remains inviolate free dom elections wholly maintained prophets downfall american democracy seen dire predictions com e naught democracy dying know seen reviveand grow know cannot die built unhampered initiative individual men women joined together common enterprise enterprise undertaken carried free expr ession free majority know democracy alone forms government enlists full force mens enlightened know democracy alone constructed unlimited civilization capable infinite progress improvement human life know look surface sense still spreading every continent humane advanced end unconqu erable forms human society nation like person bodya body must fed clothed housed invigorated r ested manner measures objectives time nation like person mind mind must kept informed alert mu st know understands hopes needs neighbors nations live within narrowing circle world nation li ke person something deeper something permanent something larger sum parts something matters fu ture calls forth sacred guarding present thing find difficult even impossible hit upon single simple word yet understand spirit faith america product centuries born multitudes came many la nds high degree mostly plain people sought early late find freedom freely democratic aspiratio n mere recent phase human history human history permeated ancient life early peoples blazed an ew middle ages written magna charta americas impact irresistible america new world tongues peo ples continent newfound land came believed could create upon continent new life life new freed om vitality written mayflower compact declaration independence constitution united states gett ysburg address first came carry longings spirit millions followed stock sprang moved forward c onstantly consistently toward ideal gained stature clarity generation hopes republic cannot fo rever tolerate either undeserved poverty selfserving wealth know still far go must greatly bui ld security opportunity knowledge every citizen measure justified resources capacity land enou gh achieve purposes alone enough clothe feed body nation instruct inform mind also spirit thre e greatest spirit without body mind men know nation could live spirit america killed even thou gh nations body mind constricted alien world lived america know would perished spirit faith sp eaks us daily lives ways often unnoticed seem obvious speaks us capital nation speaks us proce sses governing sovereignties states speaks us counties cities towns villages speaks us nations hemisphere across seas enslaved well free sometimes fail hear heed voices freedom us privilege freedom old old story destiny america proclaimed words prophecy spoken first president first i naugural words almost directed would seem year preservation sacred fire liberty destiny republ ican model government justly considered deeply finally staked experiment intrusted hands ameri can people lose sacred fireif let smothered doubt fear shall reject destiny washington strove valiantly triumphantly establish preservation spirit faith nation furnish highest justificatio n every sacrifice may make cause national defense face great perils never encountered strong p urpose protect perpetuate integrity democracy muster spirit america faith america retreat cont ent stand still americans go forward service country god' We have also performed stemming on each speech, since stemming is a process of converting any word to its root form.

- **President Kennedy Speech Stop word removal process**

Text	count_stop
Vice President Johnson, Mr. Speaker, Mr. Chief...	642

From above table we can notice that President Kennedy Speech is having 642 stopwords in it , we have also added Punctuation as well as '--' to the stopwords list.

Below is the array of stop words in president Kennedy Inaugural Speech.

['we', 'not', 'a', 'of', 'but', 'a', 'of', '--', 'an', 'as', 'as', 'a', '--', 'as', 'as', 'have', 'before', 'you', 'and', 'the', 'same', 'our', 'a', 'and', 'is', 'very', 'in', 'his', 'the', 'to', 'all', 'of', 'and', 'all', 'of', 'the', 'same', 'for', 'which', 'our', 'are', 'at', 'the', '--', 'the', 'that', 'the', 'of', 'not', 'from', 'the', 'of', 'the', 'but', 'from', 'the', 'of', 'not', 'that', 'we', 'are', 'the', 'of', 'that', 'the', 'from', 'this',

'and', 'to', 'and', 'that', 'the', 'has', 'been', 'to', 'a', 'of', '--', 'in', 'this', 'by', 'by', 'a', 'and', 'of', 'our', '--', 'and', 'to', 'or', 'the', 'of', 'those', 'to', 'which', 'this', 'has', 'been', 'and', 'to', 'which', 'we', 'are', 'at', 'and', 'the', 'it', 'us', 'or', 'that', 'we', 'any', 'any', 'any', 'any', 'any', 'in', 'to', 'the', 'and', 'the', 'of', 'we', '--', 'and', 'those', 'and', 'we', 'we', 'the', 'of', 'there', 'is', 'we', 'do', 'in', 'a', 'of', 'there', 'is', 'we', 'can', 'do', '--', 'for', 'we', 'not', 'a', 'at', 'and', 'those', 'whom', 'we', 'to', 'the', 'of', 'the', 'we', 'our', 'that', 'of', 'not', 'have', 'to', 'be', 'by', 'a', 'more', 'not', 'to', 'them', 'our', 'we', 'to', 'them', 'their', 'own', '--', 'and', 'to', 'in', 'the', 'those', 'who', 'by', 'the', 'of', 'the', 'up', 'those', 'in', 'the', 'and', 'the', 'to', 'the', 'of', 'we', 'our', 'to', 'them', 'for', 'is', '--', 'not', 'because', 'the', 'be', 'doing', 'not', 'because', 'we', 'their', 'but', 'because', 'it', 'is', 'a', 'the', 'who', 'are', 'it', 'the', 'few', 'who', 'are', 'our', 'of', 'our', 'we', 'a', '--', 'to', 'our', 'into', '--', 'in', 'a', 'for', '--', 'to', 'and', 'in', 'off', 'the', 'of', 'this', 'of', 'the', 'of', 'all', 'our', 'that', 'we', 'with', 'them', 'to', 'or', 'in', 'the', 'let', 'other', 'that', 'this', 'to', 'the', 'of', 'its', 'own', 'that', 'of', 'the', 'our', 'in', 'an', 'where', 'the', 'of', 'have', 'the', 'of', 'we', 'our', 'of', 'it', 'from', 'a', 'for', '--', 'to', 'its', 'of', 'the', 'and', 'the', '--', 'and', 'to', 'the', 'in', 'which', 'its', 'to', 'those', 'who', 'themselves', 'our', 'we', 'not', 'a', 'but', 'a', 'that', 'both', 'the', 'for', 'before', 'the', 'of', 'by', 'all', 'in', 'or', 'not', 'them', 'with', 'only', 'when', 'our', 'are', 'can', 'we', 'be', 'that', 'they', 'will', 'be', 'can', 'and', 'of', 'from', 'our', '--', 'both', 'by', 'the', 'of', 'both', 'by', 'the', 'of', 'the', 'both', 'to', 'that', 'of', 'that', 'the', 'of', 'let', 'us', '--', 'on', 'both', 'that', 'is', 'not', 'a', 'of', 'and', 'is', 'to', 'us', 'out', 'of', 'let', 'us', 'to', 'both',

'what', 'us', 'of', 'those', 'which', 'both', 'for', 'the', 'and', 'for', 'the', 'and', 'of', '--', 'and', 'the', 'to', 'other', 'under', 'the', 'of', 'all', 'both', 'to', 'the', 'of', 'of', 'its', 'let', 'us', 'the', 'the', 'the', 'and', 'the', 'and', 'both', 'to', 'in', 'all', 'of', 'the', 'the', 'of', '--', 'to', 'the', 'and', 'to', 'let', 'the', 'if', 'a', 'of', 'the', 'of', 'let', 'both', 'in', 'a', 'not', 'a', 'of', 'but', 'a', 'of', 'where', 'the', 'are', 'just', 'and', 'the', 'and', 'the', 'this', 'will', 'not', 'be', 'in', 'the', 'will', 'it', 'be', 'in', 'the', 'nor', 'in', 'the', 'of', 'this', 'nor', 'in', 'our', 'on', 'this', 'let', 'us', 'your', 'my', 'more', 'than', 'in', 'will', 'the', 'or', 'of', 'our', 'this', 'was', 'each', 'of', 'has', 'been', 'to', 'to', 'its', 'of', 'who', 'the', 'to', 'the', 'the', 'us', 'again', '--', 'not', 'as', 'a', 'to', 'we', 'not', 'as', 'a', 'to', 'we', 'are', '--', 'but', 'a', 'to', 'the', 'of', 'a', 'in', 'and', 'in', 'in', '--', 'a', 'against', 'the', 'of', 'and', 'we', 'against', 'these', 'a', 'and', 'and', 'and', 'that', 'can', 'a', 'more', 'for', 'all', 'you', 'in', 'that', 'the', 'of', 'the', 'only', 'a', 'few', 'have', 'been', 'the', 'of', 'in', 'its', 'of', 'do', 'not', 'from', 'this', '--', 'do', 'not', 'that', 'any', 'of', 'us', 'with', 'any', 'other', 'or', 'any', 'other', 'the', 'the', 'which', 'we', 'to', 'this', 'will', 'our', 'and', 'all', 'who', 'it', '--', 'and', 'the', 'from', 'that', 'can', 'the', 'my', 'not', 'what', 'your', 'can', 'do', 'for', 'you', '--', 'what', 'you', 'can', 'do', 'for', 'your', 'of', 'the', 'not', 'what', 'will', 'do', 'for', 'but', 'what', 'we', 'can', 'do', 'for', 'the', 'of', 'you', 'are', 'of', 'or', 'of', 'the', 'of', 'us', 'the', 'same', 'of', 'and', 'which', 'we', 'of', 'a', 'our', 'only', 'with', 'the', 'of', 'our', 'let', 'us', 'to', 'the', 'we', 'and', 'but', 'that', 'here', 'on', 'be', 'our']

- **President Kennedy Speech after removal of stop words**

Below are the Snapshot of President Kennedy Speech after removal of stop words

'vice president johnson mr speaker mr chief justice president eisenhower vice presi dent nixon president truman reverend clergy fellow citizens observe today victory p arty celebration freedom symbolizing end well beginning signifying renewal well cha nge sworn almighty god solemn oath forebears l prescribed nearly century three quar ters ago world different man holds mortal hands power abolish forms human poverty f orms human life yet revolutionary beliefs forebears fought still issue around globe belief rights man come generosity state hand god dare forget today heirs first revo lution let word go forth time place friend foe alike torch passed new generation am ericans born century tempered war disciplined hard bitter peace proud ancient herit age unwilling witness permit slow undoing human rights nation always committed comm itted today home around world let every nation know whether wishes us well ill shal l pay price bear burden meet hardship support friend oppose foe order assure surviv al success liberty much pledge old allies whose cultural spiritual origins share pl edge loyalty faithful friends united little cannot host cooperative ventures divide d little dare meet powerful challenge odds split asunder new states welcome ranks free pledge word one form colonial control shall passed away merely replaced far iron tyranny shall always expect find supporting view shall always hope findstrongly supporting freedom remember past foolishly sought power riding back tiger ended ins ide peoples huts villages across globe struggling break bonds mass misery pledge be st efforts help help whatever period required communists may seek votes right free society cannot help many poor cannot save rich sister republics south border offer special pledge convert good words good deeds new alliance progress assist free men free governments casting chains poverty peaceful revolution hope cannot become prey hostile powers let neighbors know shall join oppose aggression subversion anywhere americas let every power know hemisphere intends remain master house world assembly sovereign states united nations last best hope age instruments war far outpaced ins truments peace renew pledge supportto prevent becoming merely forum invective stren gthen shield new weak enlarge area writ may run finally nations would make adversar y offer pledge request sides begin anew quest peace dark powers destruction unleash ed science engulf humanity planned accidental selfdestruction dare tempt weakness a rms sufficient beyond doubt certain beyond doubt never employed neither two great p owerful groups nations take comfort present course sides overburdened cost modern w eapons rightly alarmed steady spread deadly atom yet racing alter uncertain balance terror stays hand mankinds final war let us begin anew remembering sides civility s ign weakness sincerity always subject proof let us never negotiate fear let us neve r fear negotiate let sides explore problems unite us instead belaboring problems di vide us let sides first time formulate serious precise proposals inspection control arms bring absolute power destroy nations absolute control nations let sides seek in voke wonders science instead terrors together let us explore stars conquer deserts eradicate disease tap ocean depths encourage arts commerce let sides unite heed corners earth command is a iahundo heavy burdens let oppressed go free beachhead cooperation may push back jungle suspicion let sides join creating new endeavor new balan ce power new world law strong weak secure peace preserved finished first days finis hed first days life administration even perhaps lifetime planet let us begin hands fellow citizens mine rest final success failure course since country founded genera tion americans summoned give testimony national loyalty graves young americans answ ered call service surround globe trumpet summons us call bear arms though arms need call battle though embattled call bear burden long twilight struggle year year rejo icing hope patient tribulation struggle common enemies man tyranny poverty disease war forge enemies grand global alliance north south east west assure fruitful life mankind join historic effort long history world generations granted role defending freedom hour maximum danger shrink responsibility welcome believe us would exchange places people generation energy faith devotion bring endeavor light country serve g low fire truly light world fellow americans ask country ask country fellow citizens world ask america together freedom man finally whether citizens america citizens wo rld ask us high standards strength sacrifice ask good conscience sure reward histor y final judge deeds let us go forth lead land love asking blessing help knowing ear th gods work must truly'

- **President Nixon Speech Stop word Removal process**

Text	word_count	count_stop
Mr. Vice President, Mr. Speaker, Mr. Chief Jus...	1819	950

From above table we can notice that President Nixon Speech is having 954 stopwords in it , we have also added Punctuation as well as '--' to the stop words list.

Below is the array of stopwords found in president Nixon Inaugural Speech.

['and', 'my', 'of', 'this', 'and', 'we', 'we', 'here', 'was', 'in', 'by', 'the', 'of', 'and', 'of', 'at', 'we', 'here', 'we', 'on', 'the', 'of', 'a', 'of', 'in', 'the', 'bef
ore', 'us', 'we', 'that', 'us', 'that', 'this', 'we', 'are', 'about', 'to', 'will', 'not', 'be', 'what', 'other', 'have', 'so', 'a', 'of', 'and', 'that', 'to', 'at', 'and',
'us', 'that', 'this', 'will', 'be', 'what', 'it', 'can', 'a', 'of', 'in', 'which', 'we', 'the', 'and', 'the', 'of', 'as', 'we', 'our', 'as', 'a', 'from', 'our', 'for', 'to', 'o
ur', 'and', 'by', 'our', 'to', 'and', 'to', 'we', 'were', 'to', 'the', 'for', 'a', 'and', 'more', 'of', 'the', 'of', 'the', 'of', 'will', 'be', 'as', 'the', 'of', 'the', 'the'
, 'of', 'a', 'in', 'the', 'we', 'in', 'the', 'is', 'not', 'the', 'which', 'is', 'an', 'between', 'but', 'a', 'which', 'can', 'for', 'to', 'is', 'that', 'we', 'both', 'the', 'a
nd', 'the', 'of', 'in', 'that', 'we', 'in', 'to', 'the', 'there', 'will', 'be', 'no', 'we', 'in', 'to', 'there', 'will', 'be', 'no', 'let', 'us', 'the', 'of', 'as', 'a', 'of', 'th
e', 'we', 'have', 'over', 'these', 'our', 'the', 'that', 'no', 'has', 'the', 'to', 'its', 'will', 'or', 'on', 'by', 'in', 'this', 'of', 'to', 'for', 'the', 'of', 'and', 'to', 't
he', 'of', 'between', 'the', 'do', 'our', 'in', 'and', 'in', 'the', 'we', 'to', 'do', 'their', 'has', 'when', 'will', 'other', 'our', 'or', 'other', 'our', 'or', 'to', 't
he', 'of', 'other', 'how', 'to', 'their', 'own', 'as', 'we', 'the', 'of', 'each', 'to', 'its', 'own', 'we', 'the', 'of', 'each', 'to', 'its', 'own', 'as', 'is', 'in', 'the',
'so', 'is', 'each', 'in', 'its', 'own', 'with', 'the', 'of', 'the', 'let', 'us', 'to', 'from', 'the', 'we', 'have', 'us', 'to', 'down', 'the', 'of', 'which', 'have', 'the',
'for', 'too', 'and', 'to', 'in', 'their', 'of', '--', 'so', 'that', 'between', 'of', 'the', 'of', 'the', 'can', 'be', 'us', 'a', 'of', 'in', 'the', 'in', 'which', 'the', 'are', '
as', 'as', 'the', '--', 'in', 'which', 'each', 'the', 'of', 'the', 'other', 'to', 'by', 'a', '--', 'in', 'which', 'those', 'who', 'will', 'do', 'so', 'by', 'the', 'of', 'their'
, 'and', 'not', 'by', 'the', 'of', 'their', 'us', 'that', 'not', 'as', 'a', 'but', '--', 'because', 'the', 'to', 'such', 'a', 'is', 'the', 'in', 'which', 'a', 'can', 'because
', 'only', 'if', 'we', 'in', 'our', 'will', 'we', 'a', 'and', 'only', 'if', 'we', 'a', 'will', 'we', 'in', 'our', 'at', 'have', 'the', 'to', 'do', 'more', 'than', 'before', 'in
', 'our', 'to', 'in', '--', 'to', 'a', '--', 'to', 'for', 'to', 'our', 'more', '--', 'and', 'to', 'the', 'of', 'to', 'and', 'the', 'of', 'our', 'is', 'so', '--', 'because', 'the', 'o
f', 'our', 'is', 'so', '--', 'let', 'us', 'be', 'in', 'our', 'to', 'those', 'in', 'as', 'a', 'of', 'has', 'from', 'that', 'so', 'a', 'of', 'at', 'from', 'that', 'have', 'the', 'fro
m', 'to', 'has', 'not', 'been', 'a', 'from', 'our', 'but', 'a', 'to', 'at', 'the', 'from', 'to', 'will', 'not', 'be', 'a', 'from', 'our', 'but', 'a', 'to', 'and', 'at', 'the'
, 'to', 'those', 'in', 'the', 'and', 'the', 'of', 'have', 'too', 'with', 'the', 'of', 'to', 'all', 'and', 'in', 'and', 'at', 'the', 'has', 'to', 'from', 'the', 'of', '--', 'of',
'can', 'be', 'to', 'only', 'if', 'he', 'has', 'is', 'let', 'us', 'at', 'and', 'to', 'do', 'more', 'for', 'to', 'more', 'for', 'us', 'in', 'more', 'us', 'what', 'we', 'will', '
do', 'for', 'by', 'what', 'they', 'will', 'do', 'for', 'is', 'why', 'no', 'of', 'a', 'for', 'have', 'too', 'with', 'that', 'too', 'in', 'we', 'have', 'of', 'it', 'more', 'th
an', 'it', 'can', 'only', 'to', 'to', 'and', 'to', 'a', 'and', 'that', 'both', 'in', 'what', 'can', 'do', 'and', 'in', 'what', 'can', 'to', 'from', 'so', 'that', 'an', 'do',
'more', 'for', 'us', 'that', 'was', 'not', 'by', 'but', 'by', '--', 'not', 'by', 'but', 'by', '--', 'not', 'by', 'but', 'by', 'our', 'own', 'let', 'each', 'of', 'us', '--', 'n
ot', 'just', 'what', 'will', 'do', 'for', 'but', 'what', 'can', 'do', 'for', 'the', 'we', 'let', 'each', 'of', 'us', '--', 'not', 'just', 'how', 'can', 'but', 'how', 'can',
'has', 'a', 'and', 'to', 'to', 'you', 'that', 'where', 'this', 'should', 'we', 'will', 'and', 'we', 'will', 'just', 'as', 'is', 'the', 'that', 'each', 'and', 'of', 'us', 'as'
, 'an', 'and', 'as', 'a', 'of', 'his', 'own', 'this', 'let', 'each', 'of', 'us', 'a', 'in', 'his', 'own', 'to',

'his', 'to', 'do', 'his', 'to', 'his', '--', 'so', 'that', 'we', 'can
', 'the', 'of', 'a', 'of', 'for', 'and', 'as', 'we', 'our', 'as', 'a', 'we', 'can', 'do', 'so', 'in', 'the', 'of', 'our', 'to', 'ourselves', 'and', 'to', 'the', 'and', 'most',
'to', 'an', 'let', 'us', 'again', 'to', 'our', 'with', 'and', 'let', 'each', 'of', 'us', 'out', 'for', 'that', '--', 'a', 'of', 'for', 'the', 'and', 'of', 'a', 'of', 'for', 'the', '
which', 'is', 'the', 'of', 'all', 'the', 'has', 'for', 'us', 'to', 'our', 'in', 'ourselves', 'and', 'in', 'that', 'has', 'been', 'have', 'been', 'to', 'be', 'of', 'their', '
of', 'their', 'of', 'at', 'and', 'of', 'its', 'in', 'the', 'we', 'have', 'been', 'by', 'those', 'who', 'with', 'and', 'that', 'is', 'am', 'that', 'this', 'will', 'not', 'be'
, 'the', 'of', 'on', 'these', 'in', 'which', 'we', 'are', 'to', 'in', 'this', 'has', 'been', 'in', 'the', 'for', 'its', 'for', 'its', 'for', 'its', 'and', 'for', 'its', 'us', 'be', '
that', 'our', 'has', 'and', 'more', 'and', 'more', 'more', 'than', 'any', 'other', 'in', 'the', 'of', 'the', 'us', 'be', 'that', 'in', 'each', 'of', 'the', 'in', 'whic
h', 'we', 'have', 'been', 'in', 'this', 'the', 'we', 'are', 'now', 'to', 'an', 'we', 'have', 'not', 'for', 'our', 'but', 'to', 'us', 'be', 'that', 'by', 'our', 'and', 'by
', 'our', 'for', 'with', 'we', 'have', 'a', 'in', 'the', 'what', 'the', 'has', 'not', 'before', '--', 'a', 'of', 'that', 'can', 'not', 'for', 'our', 'but', 'for', 'to', 'are',
'here', 'on', 'an', 'that', 'as', 'those', 'any', 'or', 'any', 'has', 'to', 'to', 'and', 'to', 'our', 'for', 'the', 'in', 'which', 'we', 'these', 'in', 'this', 'so', 'by', '
of', 'who', 'have', 'here', 'before', 'of', 'the', 'they', 'had', 'for', 'and', 'of', 'how', 'each', 'that', 'he', 'himself', 'in', 'to', 'those', 'your', 'that', 'in',
'the', 'have', 'in', 'that', 'are', 'for', 'and', 'for', 'your', 'so', 'that', 'we', 'be', 'of', 'our', 'us', 'to', 'these', 'the', 'in', 'so', 'that', 'on', 'its', 'will', 'be'
, 'as', 'and', 'as', 'as', 'when', 'it', 'and', 'as', 'a', 'of', 'for', 'all', 'the', 'us', 'from', 'here', 'in', 'in', 'our', 'in', 'by', 'our', 'in', 'who', 'and', 'to']

- **Glance On president Nixon inaugural Speech after Removal of Stopwords**

Below is the speech By president Nixon Inagural speech after Removal of stopwords

'mr vice president mr speaker mr chief justice senator cook mrs eisenhower fellow citizens gre at good country share together met four years ago america bleak spirit depressed prospect seem ingly endless war abroad destructive conflict home meet today stand threshold new era peace wo rld central question shall use peace resolve era enter postwar periods often time retreat isol ation leads stagnation home invites new danger abroad resolve become time great responsibiliti es greatly borne renew spirit promise america enter third century nation past year saw farreac hing results new policies peace continuing revitalize traditional friendships missions peking moscow able establish base new durable pattern relationships among nations world americas bold initiatives long remembered year greatest progress since end world war ii toward lasting peace world peace seek world flimsy peace merely interlude wars peace endure generations come import ant understand necessity limitations americas role maintaining peace unless america work prese rve peace peace unless america work preserve freedom freedom clearly understand new nature ame ricas role result new policies adopted past four years shall respect treaty commitments shall support vigorously principle country right impose rule another force shall continue era negoti ation work limitation nuclear arms reduce danger confrontation great powers shall share defend ing peace freedom world shall expect others share time passed america make every nations confl ict make every nations future responsibility presume tell people nations manage affairs respec t right nation determine future also recognize responsibility nation secure future americas ro le indispensable preserving worlds peace nations role indispensable preserving peace together rest world resolve move forward beginnings made continue bring walls hostility divided world long build place bridges understanding despite profound differences systems government people world friends build structure peace world weak safe strong respects right live different system would influence others strength ideas force arms accept high responsibility burden gladly gladly chance build peace noblest endeavor nation engage gladly also act greatly meeting responsib ilities abroad remain great nation remain great nation act greatly meeting challenges home cha nce today ever history make life better america ensure better education better health better h ousing better transportation cleaner environment restore respect law make communities livable insure godgiven right every american full equal opportunity range needs great reach opportunit ies great bold determination meet needs new ways building structure peace abroad required turn ing away old policies failed building new era progress home requires turning away old policies failed abroad shift old policies new retreat responsibilities better way peace home shift old policies new retreat responsibilities better way progress abroad home key new responsibilities lies placing division responsibility lived long consequences attempting gather power responsibility washington abroad home time come turn away condescending policies paternalism washington knows best person expected act responsibly responsibility human nature encourage individuals home nations abroad decide locate responsibility places measure others today offer promise purely governmental solution every problem lived long false promise trusting much government asked deliver leads inflated expectations reduced individual effort disappointment frustration erode confidence government people government must learn take less people people remember america built government people welfare work shirking responsibility seeking responsibility lives ask government challenges face together ask government help help national government great vital role play pledge government act act boldly lead boldly important role every one must play individid ual member community day forward make solemn commitment heart bear responsibility part live id eals together see dawn new age progress america together celebrate th anniversary nation proud fulfillment promise world americas longest difficult war comes end learn debate differences ci vility decency reach one precious quality government cannot provide new level respect rights f eelings one another new level respect individual human dignity cherished birthright every amer ican else time come renew faith america recent years faith challenged children taught ashamed country ashamed parents ashamed americas record home role world every turn beset find everythi ng wrong america little right confident judgment history remarkable times privileged live amer icas record century unparalleled worlds history responsibility generosity creativity progress proud system produced provided freedom abundance widely shared system history world proud four wars engaged century including one bringing end fought selfish advantage help others resist ag gression proud bold new initiatives steadfastness peace honor made breakthrough toward creatin g world world known structure peace last merely time generations come embarking today era pres ents challenges great nation generation ever faced shall answer god history conscience way use years stand place hallowed history think others stood think dreams america think recognized ne eded help far beyond order make dreams come true today ask prayers years ahead may gods help m aking decisions right america pray help together may worthy challenge pledge together make nex t four years best four years americas history th birthday america young vital began bright bea con hope world go forward confident hope strong faith one another sustained faith god created striving always serve purpose'

Comment on Stemming & Lower Case conversion.
We have used porter stemmer for converting word to its root word & also convert whole text into lowercase

2.3) Which word occurs the most number of times in his inaugural address for each president? Mention the top three words. (after removing the stop words)

Let’s first understand

What is Word Cloud?
A word cloud (also known as a tag cloud or text cloud) is a visual representation of a text, in which the words appear bigger the more often they are mentioned. Word clouds are great for visualizing unstructured text data and getting insights on trends and patterns Source;-anlyticdimag.com

We have also added (let & us) which was most occurring words on President John F. Kennedy speech in 1961 to stop words & remove them because those will not make any sense there.

We have found most common words along with frequency using fdist most Common words function & results are shown in above table , As we can see that Nation has occurred 17 times & know have occurred 10 times and people have occurred 9 times in Roosevelt speech .

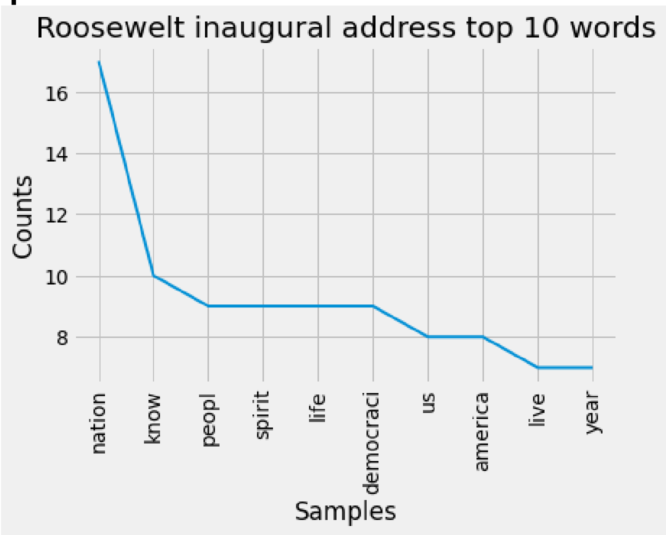
In President Kennedy’s speech power, world & nation are most common words with respective frequency of 9,8,8.

In President Richard Nixon speech We can see that peace , world & new are the 3 most common words with 19,16 & 15 as their respective frequencies .

Let’s see Most common words Visually.

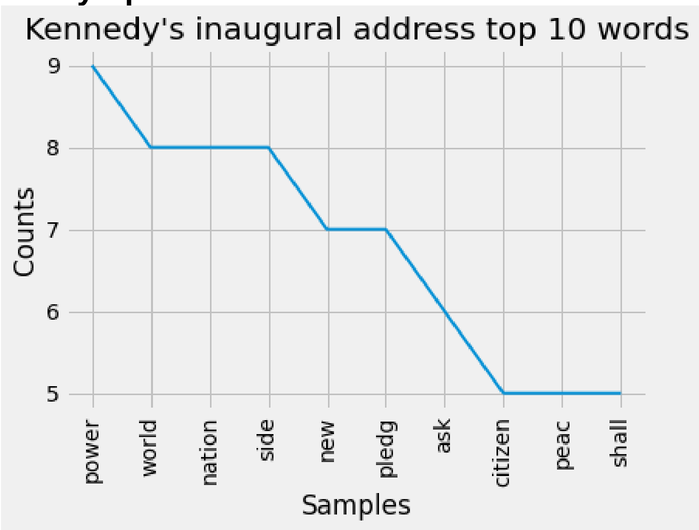
Top 3 Most Occuring words in all the Three Inagural Speech		
President Franklin D. Roosevelt in 1941	President John F. Kennedy in 1961	President Richard Nixon in 1973
('nation', 17), ('know', 10), ('people', 9),	('power', 9), ('world', 8), ('nation', 8)	('peace', 19), ('world', 16), ('new', 15),

- **Roosevelt Speech most common word**



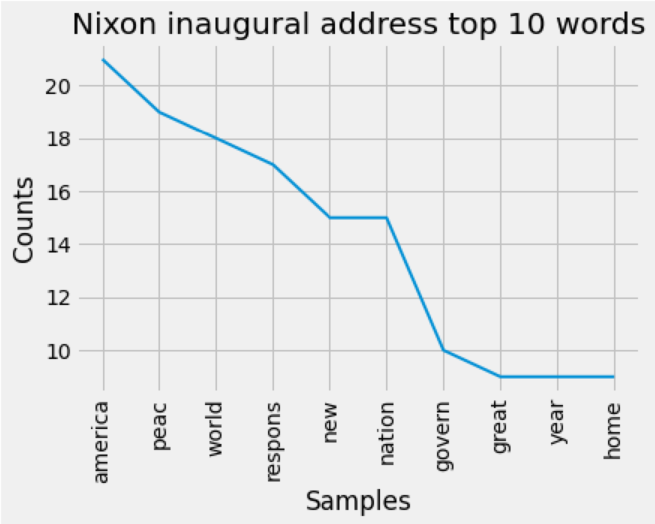
(Fig: 57 Roosevelt Speech Common Word)

- **President Kennedy Speech most common word**



(Fig: 58 Kennedy Speech Common Word)

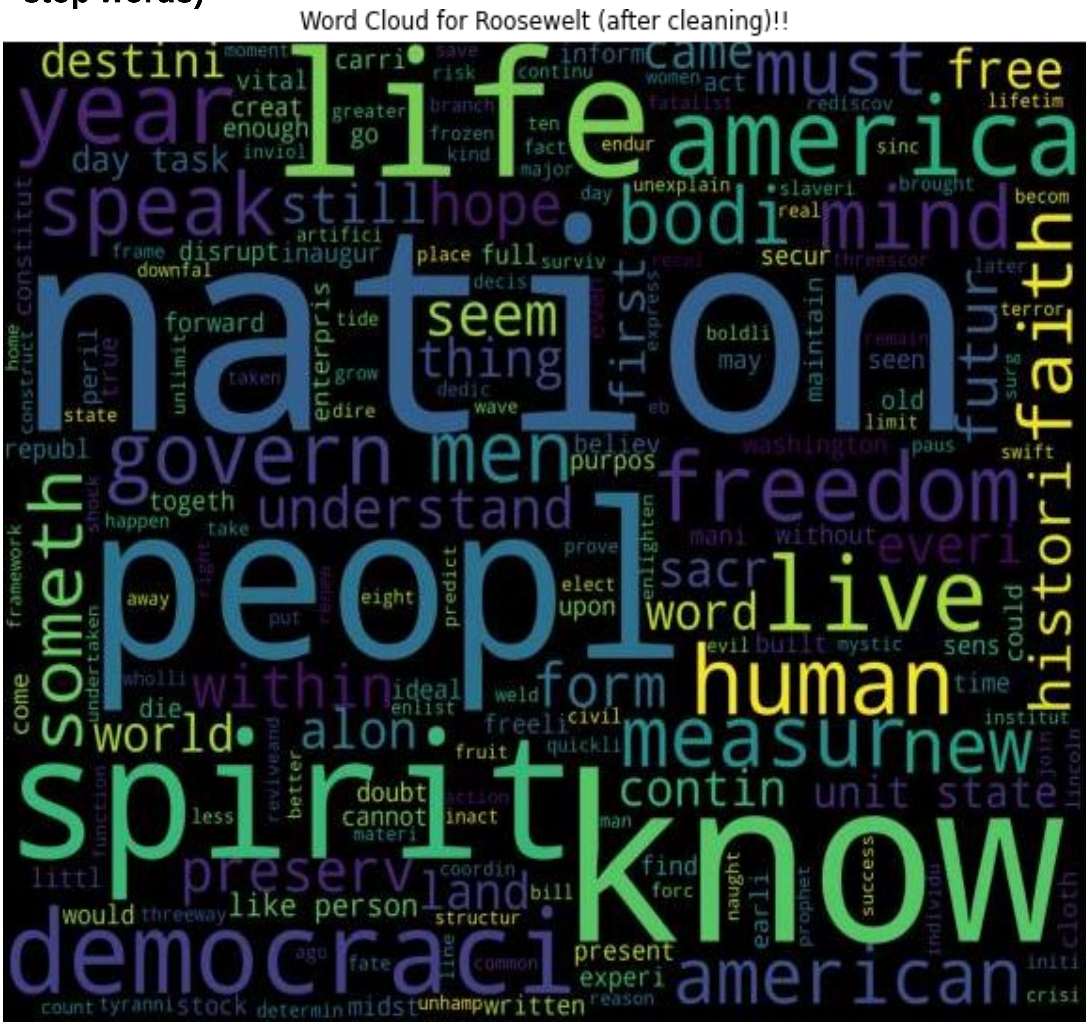
- **President Richard Nixon Speech most common word**



(Fig: 59 Richard Nixon Speech Common Word)

2.4) Plot the word cloud of each of the three speeches. (after removing the stop words)

- **Word Cloud for President Roosevelt Speech (after cleaning & removal of stop words)**



(Fig: 60 Roosevelt Word Plot)

As we know that word Nation has occurred 17 times & know have occurred 10 times and people have occurred 9 times in Roosevelt speech so we can see that text size of Nation, Know, people is large as compared to others.

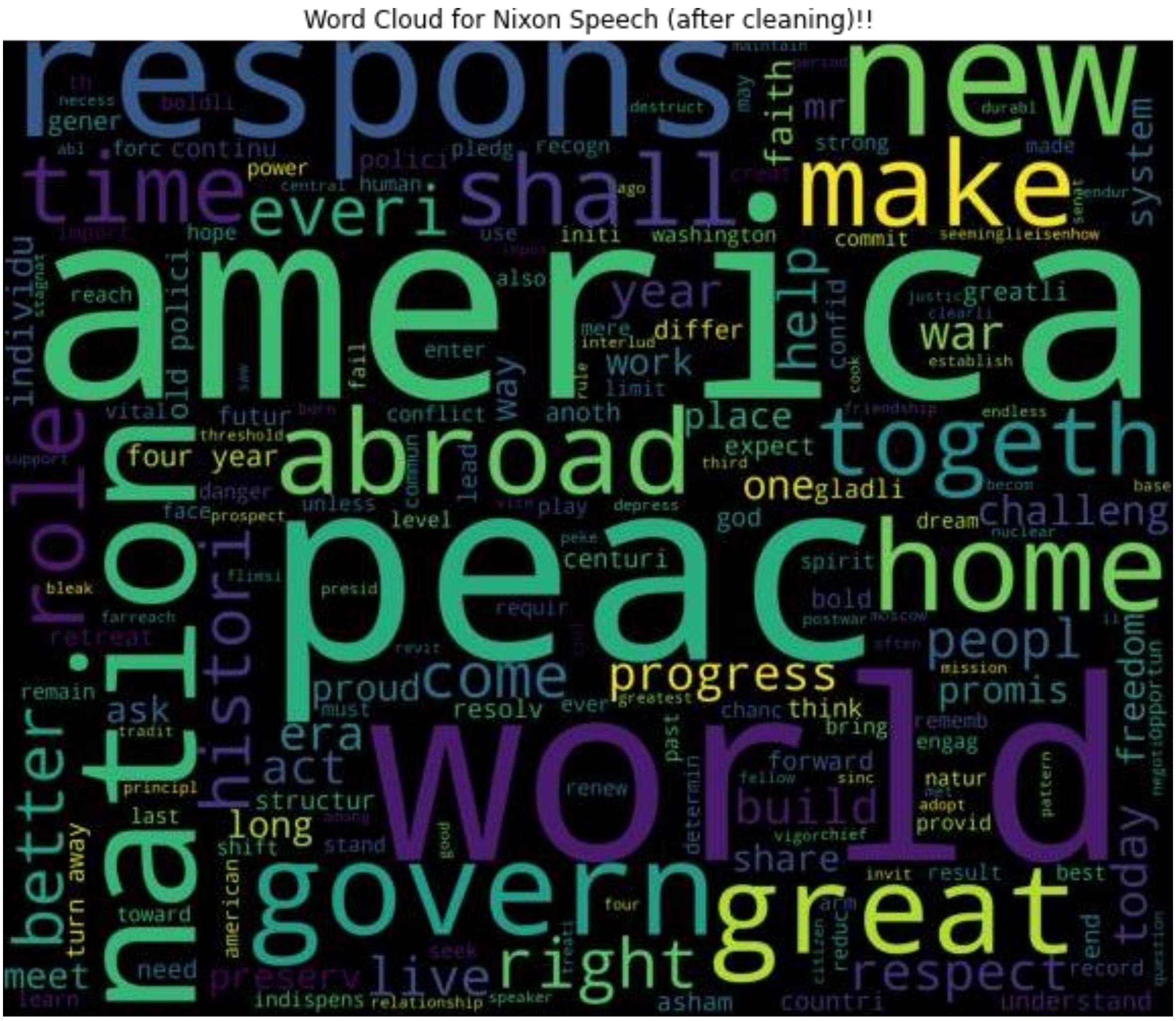
- **Word Cloud for President John F. Kennedy (after cleaning & removal of stopwords)**

Word Cloud for Kennedy Speech (after cleaning)!!

(Fig: 61 Jhon F Kennedy Word Plot)

In President Kennedy's speech power, world & nation are most common words with respective frequency of 9,8,8. so all are shown in bold & big size as compared to other words, means that President Kennedy is mostly talking about power of a in the whole world.

- **Word Cloud for President Richard Nixon speech (after cleaning & removal of stopwords)**



(Fig: 62 Richard Nixon Word Plot)

In President Richard Nixon speech, we can see that peace, world & new are the 3 most common words with 19,16 & 15 as their respective frequencies so all are shown in bold & big size as compared to other words, this means that President Nixon is mostly talking about peace of the world etc.

Thank You…!