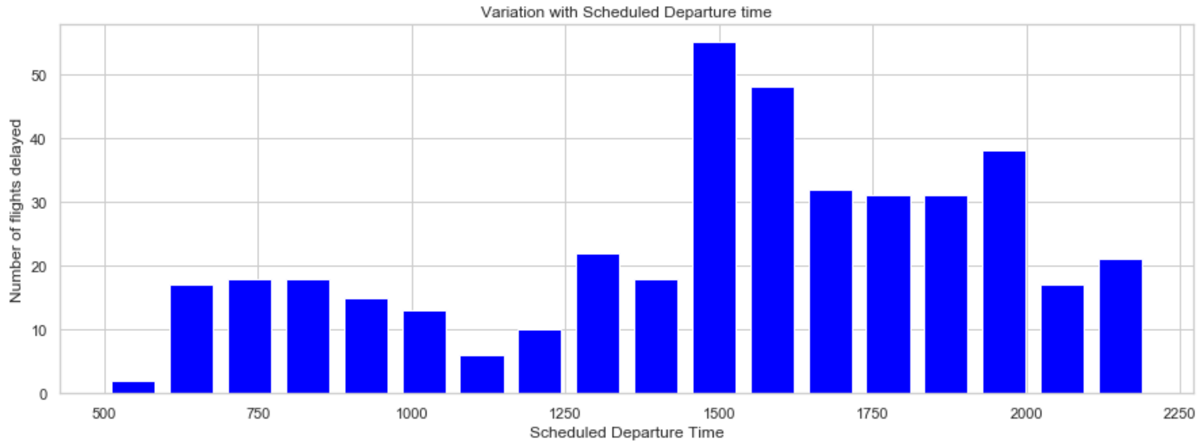
ASSIGNMENT-2

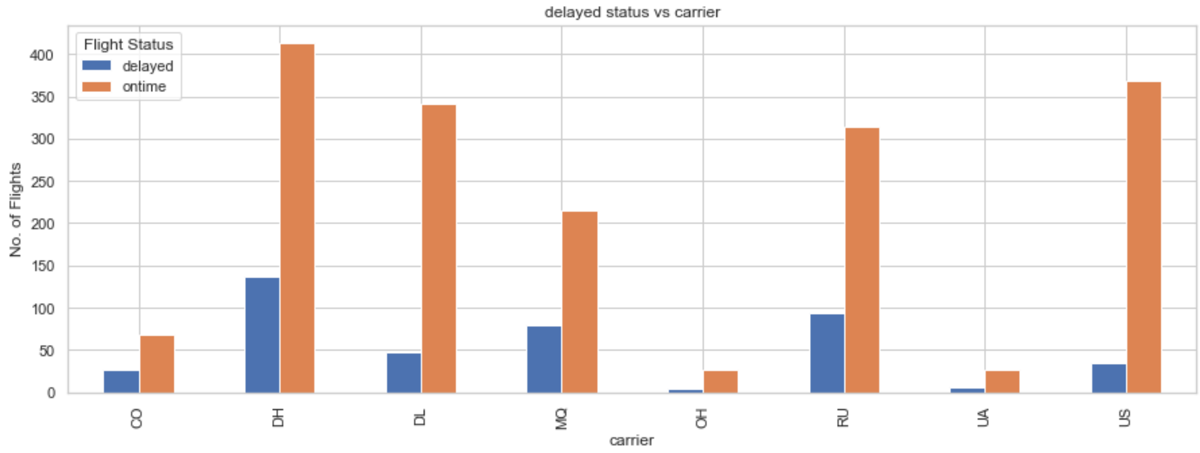
Name: Ruturaj D

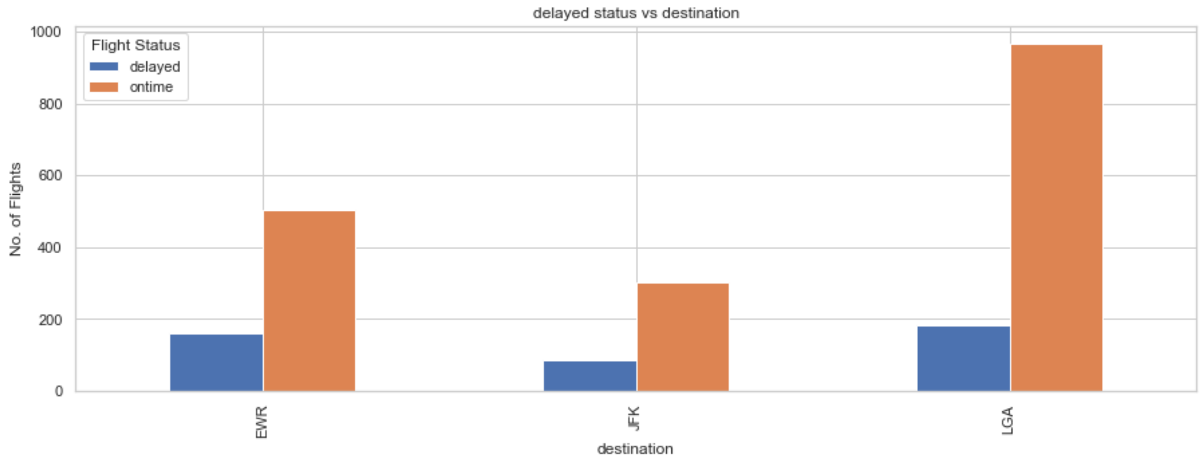
Roll. No.-18B030021

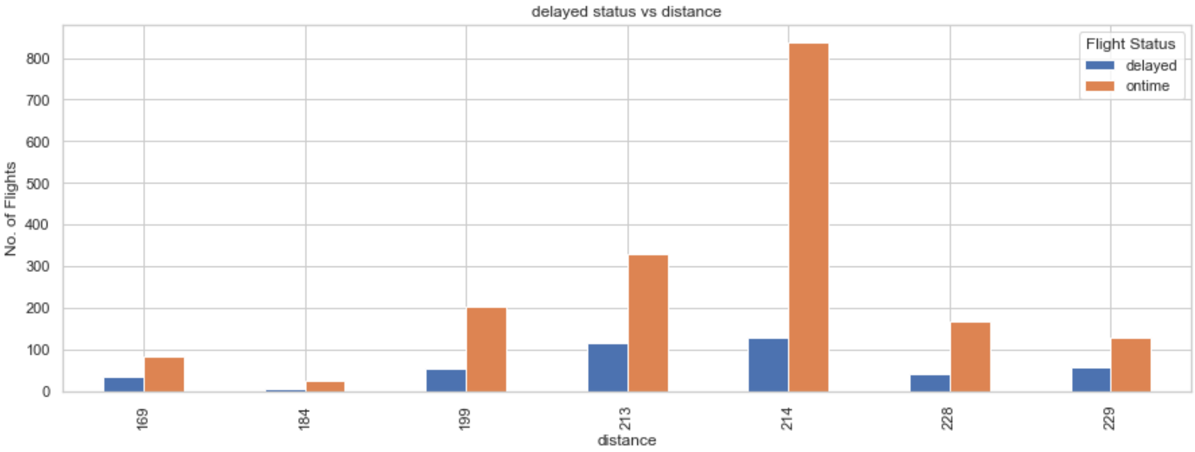
Q.1 Exploratory Data Analysis

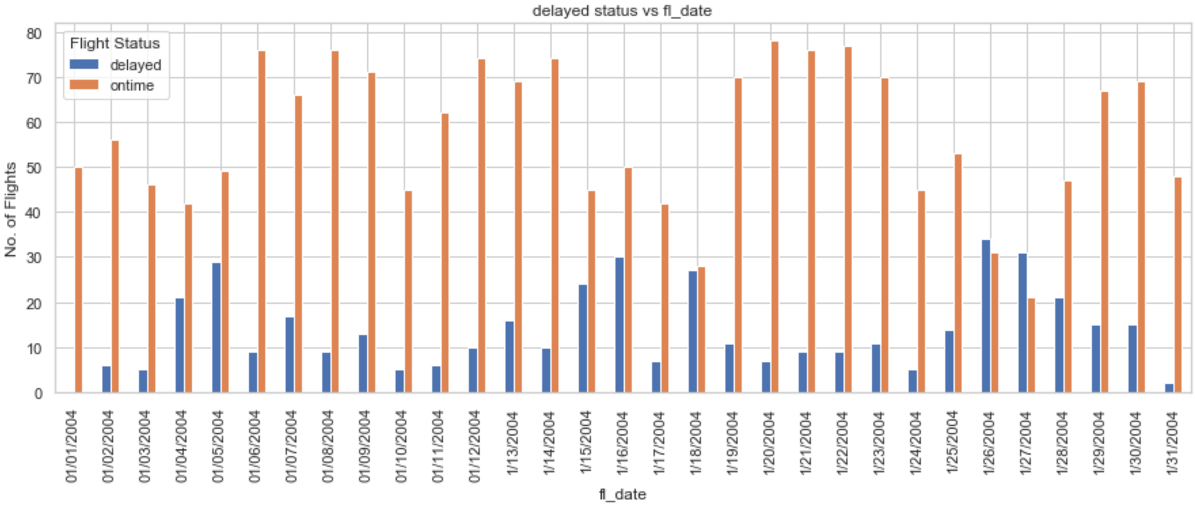
We have shown various graphs for viewing data visually.

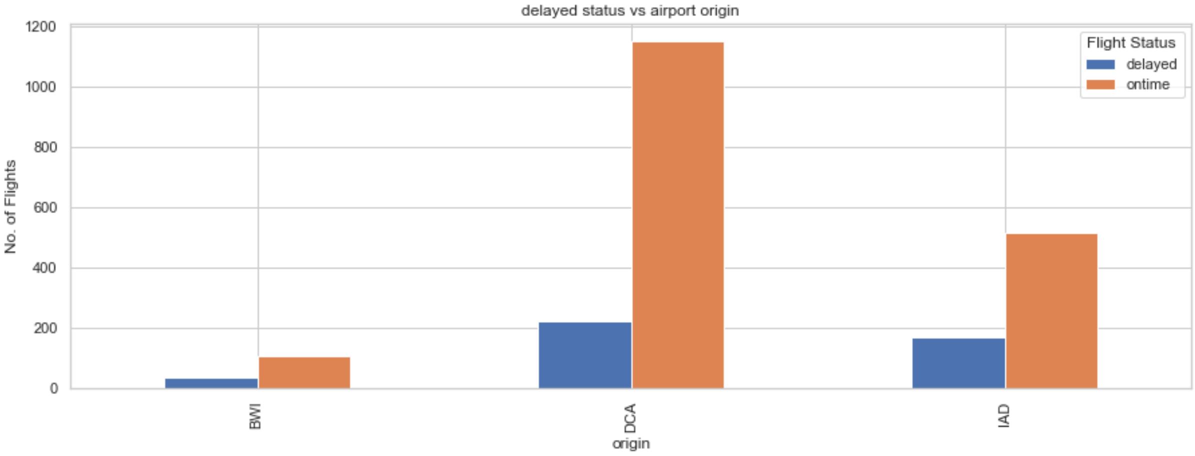


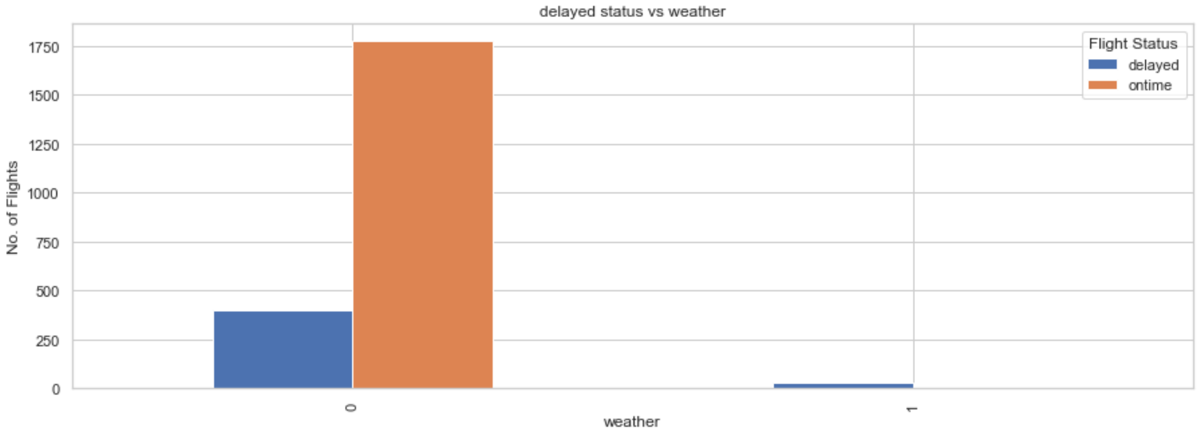


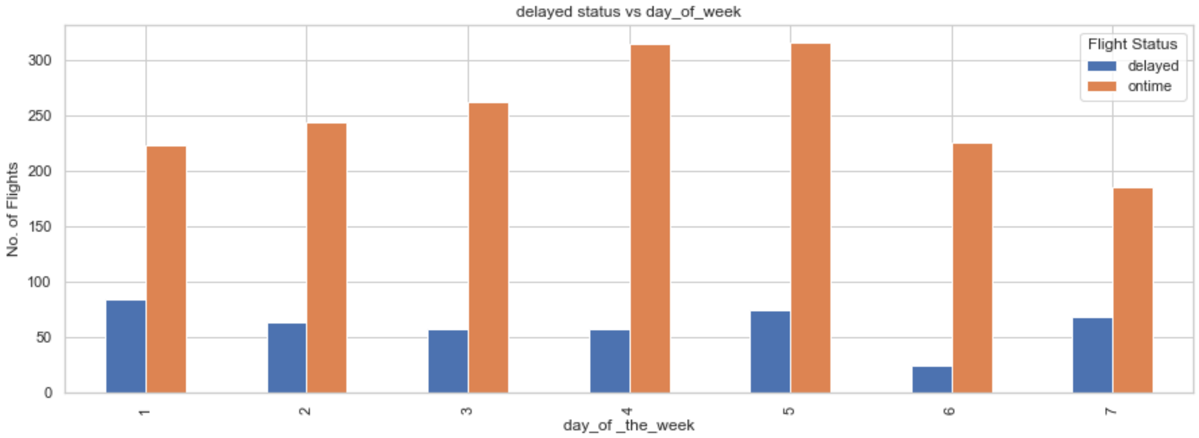


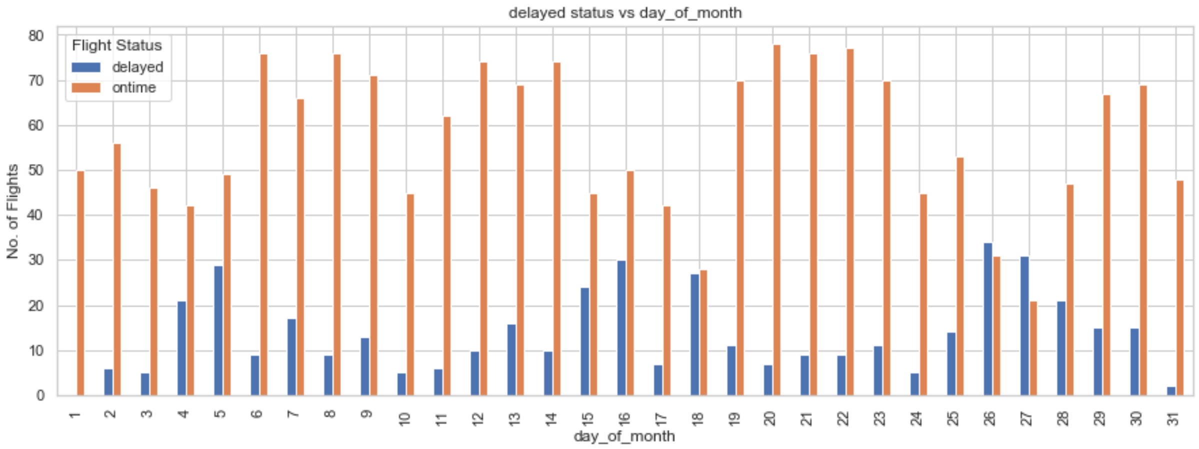

delayed status vs distance
No. of Flights
0
100
200
300
400
500
600
700
800
distance
169
184
199
213
214
228
229
Flight Status
delayed
ontime


delayed status vs fl_date
No. of Flights
0
10
20
30
40
50
60
70
80
fl_date
01/01/2004
01/02/2004
01/03/2004
01/04/2004
01/05/2004
01/06/2004
01/07/2004
01/08/2004
01/09/2004
01/10/2004
01/11/2004
01/12/2004
1/13/2004
1/14/2004
1/15/2004
1/16/2004
1/17/2004
1/18/2004
1/19/2004
1/20/2004
1/21/2004
1/22/2004
1/23/2004
1/24/2004
1/25/2004
1/26/2004
1/27/2004
1/28/2004
1/29/2004
1/30/2004
1/31/2004
Flight Status
delayed
ontime


delayed status vs airport origin
No. of Flights
0
200
400
600
800
1000
1200
origin
BWI
DCA
IAD
Flight Status
delayed
ontime







Q.2&Q.3

There were no missing values.

Logistic regression model without variable selection is used.

After using coorelation function we know that

1)CRS_DEP_TIME & DEP_TIME are almost same

2)FL_DATE & DAY_OF_MONTH are almost same

	Coefficients
INTERCEPT	1.766262
CRS_DEP_TIME	9.726938
CARRIER	-0.206328
DEP_TIME	-10.184732
DEST	0.057600
DISTANCE	0.203350
FL_DATE	-0.074595
FL_NUM	-0.032839
ORIGIN	-0.324501
Weather	-5.109902
DAY_WEEK	-0.101592
DAY_OF_MONTH	-0.074595
TAIL_NUM	-0.171747

Correctly predicted labels: 779

Accuracy: 88.42224744608399 %

Misclassification Rate(Error Rate): 11.577752553916005 %

True Positive Rate(Recall): 87.46928746928747 %

False Positive Rate: 0.0 %

True Negativity Rate(Specificity): 100.0 %

Precision: 100.0 %

Prevalence: 92.3950056753689 %

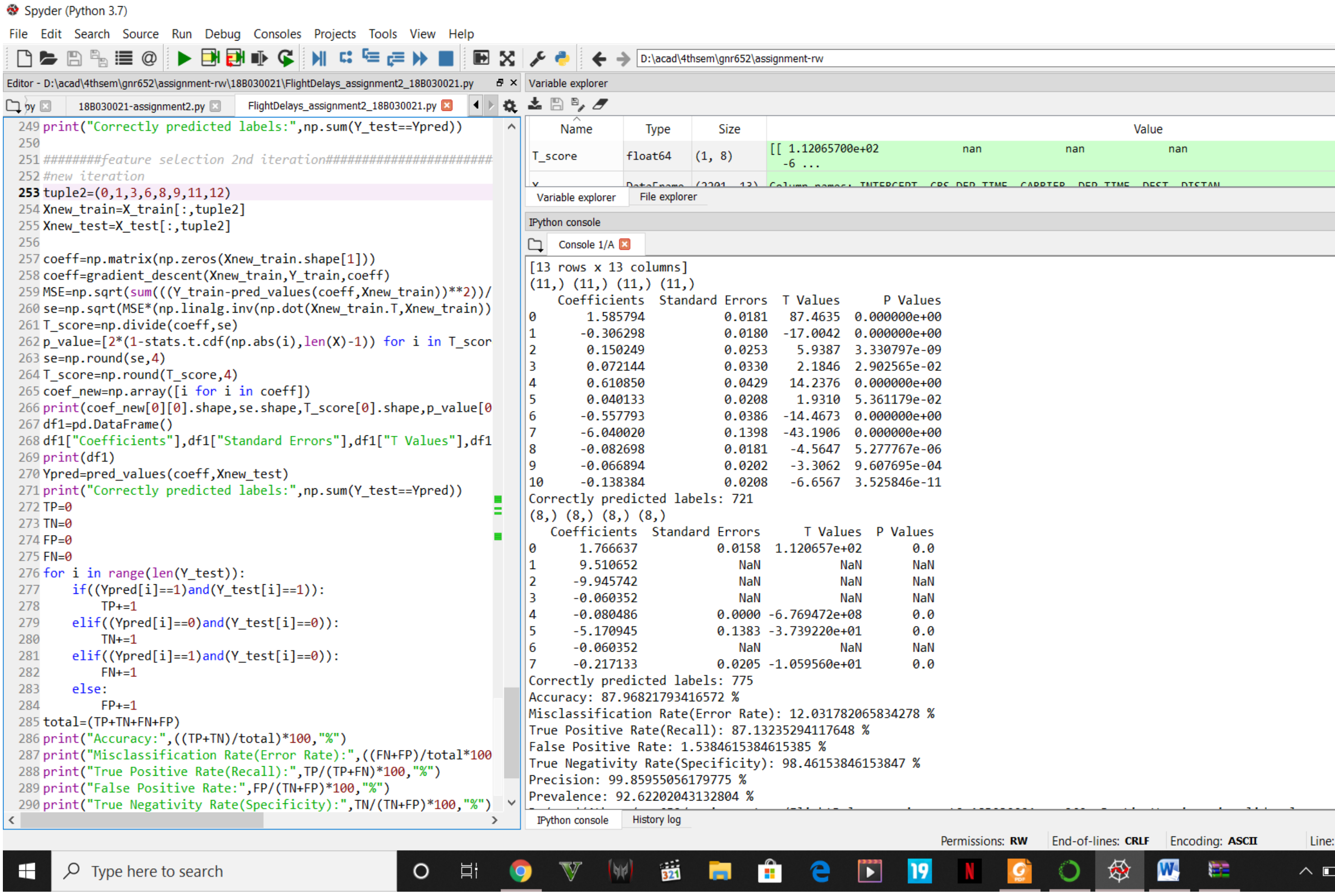
Accuracy of 88.4333 is observed

Q.4 Variable Selection

We drop variables which have coorelation of less than 0.05.

We drop carrier, destination, distance, flight number,and week day in second iteration and perform with remaining variables.

New Accuracy=87.96%



Spyder (Python 3.7)
File Edit Search Source Run Debug Consoles Projects Tools View Help
D:\acad\4thsem\gnr652\assignment-rw
Editor - D:\acad\4thsem\gnr652\assignment-rw\18B030021\FlightDelays_assignment2_18B030021.py
18B030021-assignment2.py
FlightDelays_assignment2_18B030021.py
Variable explorer

Name	Type	Size	Value
T_score	float64	(1, 8)	[[1.12065700e+02 nan nan nan -6 ...

Variable explorer
File explorer
IPython console
Console 1/A

```python
print("Correctly predicted labels:",np.sum(Y_test==Ypred))

#######feature selection 2nd iteration#######################
#new iteration
tuple2=(0,1,3,6,8,9,11,12)
Xnew_train=X_train[:,tuple2]
Xnew_test=X_test[:,tuple2]

coeff=np.matrix(np.zeros(Xnew_train.shape[1]))
coeff=gradient_descent(Xnew_train,Y_train,coeff)
MSE=np.sqrt(sum(((Y_train-pred_values(coeff,Xnew_train))**2))/
se=np.sqrt(MSE*(np.linalg.inv(np.dot(Xnew_train.T,Xnew_train))
T_score=np.divide(coeff,se)
p_value=[2*(1-stats.t.cdf(np.abs(i),len(X)-1)) for i in T_scor
se=np.round(se,4)
T_score=np.round(T_score,4)
coef_new=np.array([i for i in coeff])
print(coef_new[0][0].shape,se.shape,T_score[0].shape,p_value[0
df1=pd.DataFrame()
df1["Coefficients"],df1["Standard Errors"],df1["T Values"],df1
print(df1)
Ypred=pred_values(coeff,Xnew_test)
print("Correctly predicted labels:",np.sum(Y_test==Ypred))
TP=0
TN=0
FP=0
FN=0
for i in range(len(Y_test)):
    if((Ypred[i]==1)and(Y_test[i]==1)):
        TP+=1
    elif((Ypred[i]==0)and(Y_test[i]==0)):
        TN+=1
    elif((Ypred[i]==1)and(Y_test[i]==0)):
        FN+=1
    else:
        FP+=1
total=(TP+TN+FN+FP)
print("Accuracy:",((TP+TN)/total)*100,"%")
print("Misclassification Rate(Error Rate):",((FN+FP)/total*100
print("True Positive Rate(Recall):",TP/(TP+FN)*100,"%")
print("False Positive Rate:",FP/(TN+FP)*100,"%")
print("True Negativity Rate(Specificity):",TN/(TN+FP)*100,"%")
```

```
[13 rows x 13 columns]
(11,) (11,) (11,) (11,)
    Coefficients  Standard Errors  T Values      P Values
0       1.585794           0.0181   87.4635  0.000000e+00
1      -0.306298           0.0180  -17.0042  0.000000e+00
2       0.150249           0.0253    5.9387  3.330797e-09
3       0.072144           0.0330    2.1846  2.902565e-02
4       0.610850           0.0429   14.2376  0.000000e+00
5       0.040133           0.0208    1.9310  5.361179e-02
6      -0.557793           0.0386  -14.4673  0.000000e+00
7      -6.040020           0.1398  -43.1906  0.000000e+00
8      -0.082698           0.0181   -4.5647  5.277767e-06
9      -0.066894           0.0202   -3.3062  9.607695e-04
10     -0.138384           0.0208   -6.6567  3.525846e-11
Correctly predicted labels: 721
(8,) (8,) (8,) (8,)
   Coefficients  Standard Errors      T Values  P Values
0      1.766637           0.0158  1.120657e+02       0.0
1      9.510652              NaN           NaN       NaN
2     -9.945742              NaN           NaN       NaN
3     -0.060352              NaN           NaN       NaN
4     -0.080486           0.0000 -6.769472e+08       0.0
5     -5.170945           0.1383 -3.739220e+01       0.0
6     -0.060352              NaN           NaN       NaN
7     -0.217133           0.0205 -1.059560e+01       0.0
Correctly predicted labels: 775
Accuracy: 87.96821793416572 %
Misclassification Rate(Error Rate): 12.031782065834278 %
True Positive Rate(Recall): 87.13235294117648 %
False Positive Rate: 1.5384615384615385 %
True Negativity Rate(Specificity): 98.46153846153847 %
Precision: 99.85955056179775 %
Prevalence: 92.62202043132804 %
```

Q.5)Ideal conditions for flight from dc to new york

Ans.   Weather should be ideal(no delay due to weather,weather=0)

Dep_Time=21:30

Day= First day of month

Carrier doesn’t matter

Day= First day of month

Q.6)

Q1)Name any AIs made by Tony Stark in the Marvel ………Ans. Ultron

Q4)In Star Wars Universe, name the robotic duo.
Ans. Their name was C-3PO and R2-D2.