10/17/2020

ADVERTISING PROJECT

USING

BY
FAHEEM MOHAMMED ABDUL

UNDER GUIDANCE
PROF. VIJAY KUMAR

METRO COLLEGE
OF TECHNOLOGY

INDEX

EXECUTIVE SUMMARY

Advertising Project:

Introduction:

In this project, we will be working with advertising data set, indicating whether a particular internet user clicked on an Advertisement on a company website. We will try to create a model that will predict whether they will click on an ad based off the features of that user.

The dataset contains 1000 data points collected from an Advertising Train Dataset and 200 from Advertising Test Dataset all together it has 1200 observations.

Data Set Variables Information:

- **Daily Time Spent on Site': consumer time on site in minutes**
- **'Age': customer age in years**
- **'Area Income': Avg. Income of geographical area of consumer**
- **'Daily Internet Usage': Avg. minutes a day consumer is on the internet**
- **'Ad Topic Line': Headline of the advertisement**
- **'City': City of consumer**
- **'Male': Whether or not consumer was male**
- **'Country': Country of consumer**
- **'Timestamp': Time at which consumer clicked on Ad or closed window**
- **Clicked on Ad': 0 or 1 indicated clicking on Ad**

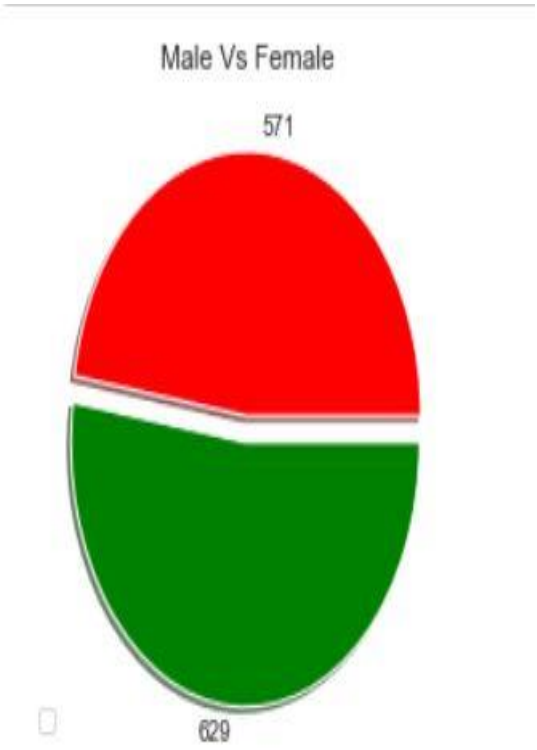
```
ad_data.info()
```

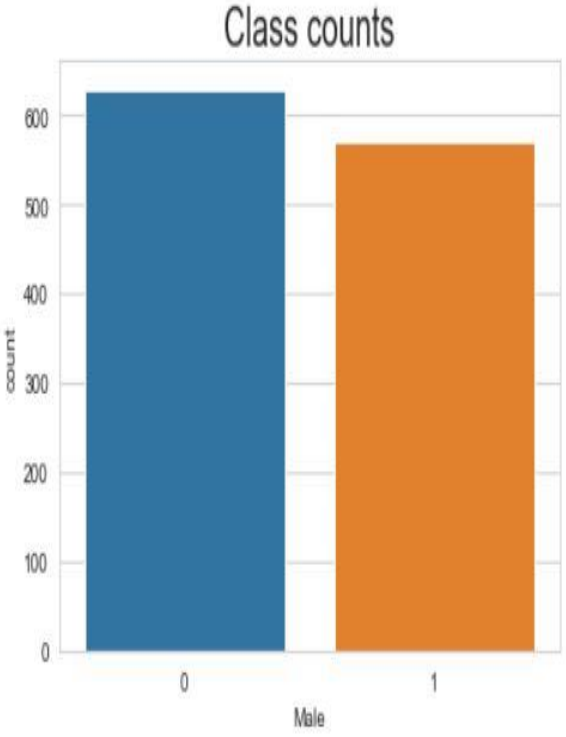
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1200 entries, 0 to 1199
Data columns (total 11 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   Ad Topic Line             1200 non-null   object 
 1   Age                       1200 non-null   int64  
 2   Area Income               1200 non-null   float64
 3   City                      1200 non-null   object 
 4   Clicked on Ad             1000 non-null   float64
 5   Country                   1200 non-null   object 
 6   Daily Internet Usage      1200 non-null   float64
 7   Daily Time Spent on Site  1200 non-null   float64
 8   Male                      1200 non-null   int64  
 9   Timestamp                 1200 non-null   object 
 10  source                    1200 non-null   object 
dtypes: float64(4), int64(2), object(5)
memory usage: 103.2+ KB
```

OBJECTIVE OF THE PROJECT:

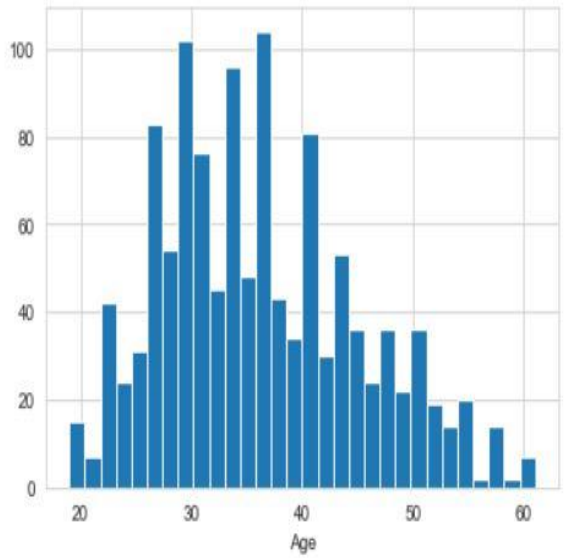
- **Analyzing Advertising Dataset.**
- **Analyzing different model in Classification such as Logistic Regression, Random Forest Model, Support Vector Machine, KNN & Decision Tree.**
- **Splitting the dataset into Training set and Test sets, Scaling the dataset**
- **Training and model on Test sets.**
- **Evaluating Accuracy, Confusion Matrix, and Micro Recall on classification mode.**
- **Performing K-Fold on Logistic Regression models.**

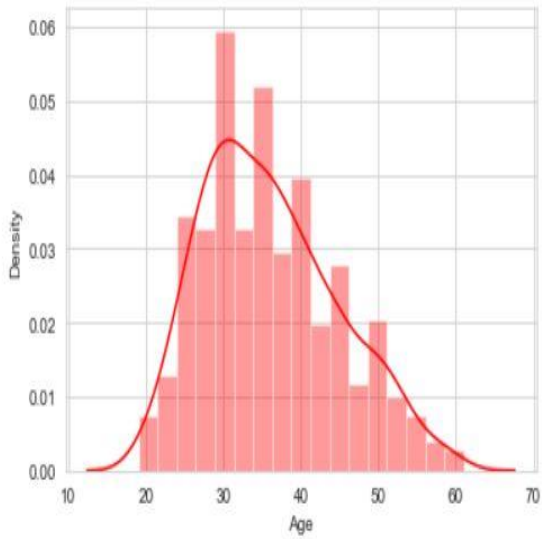
ANALYSIS OF GENDER:

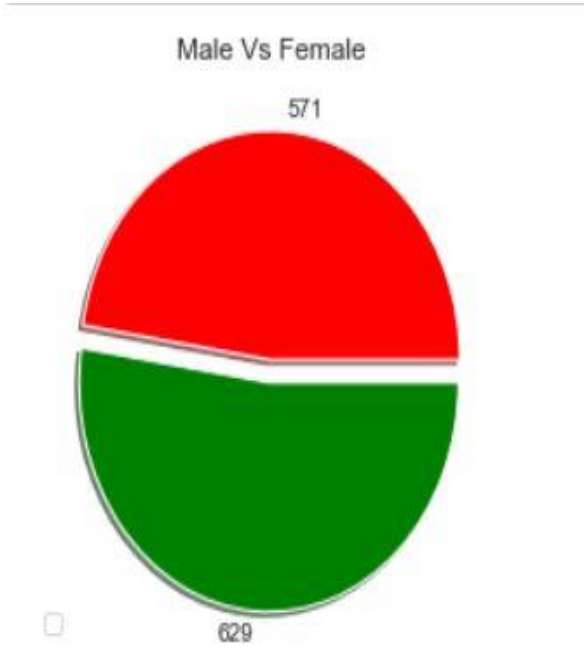


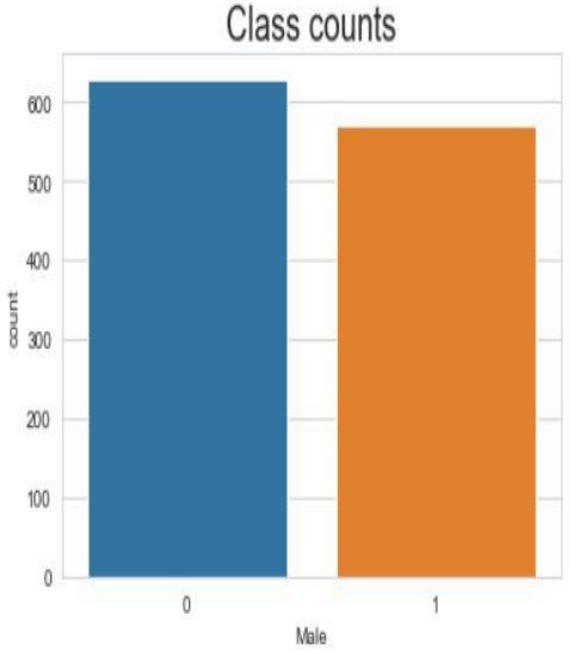


ANALYSIS OF AGE:

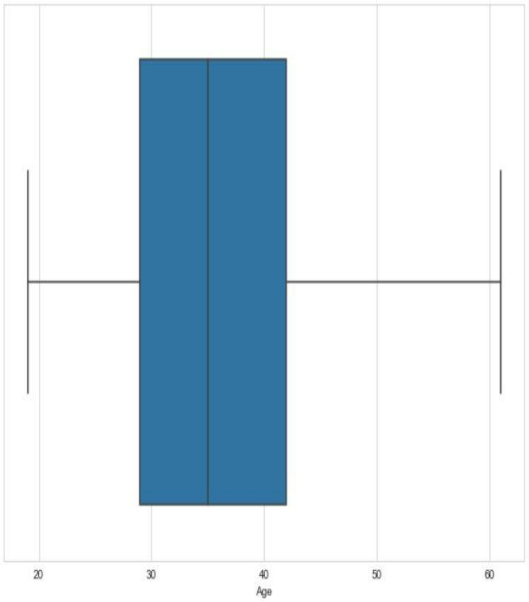


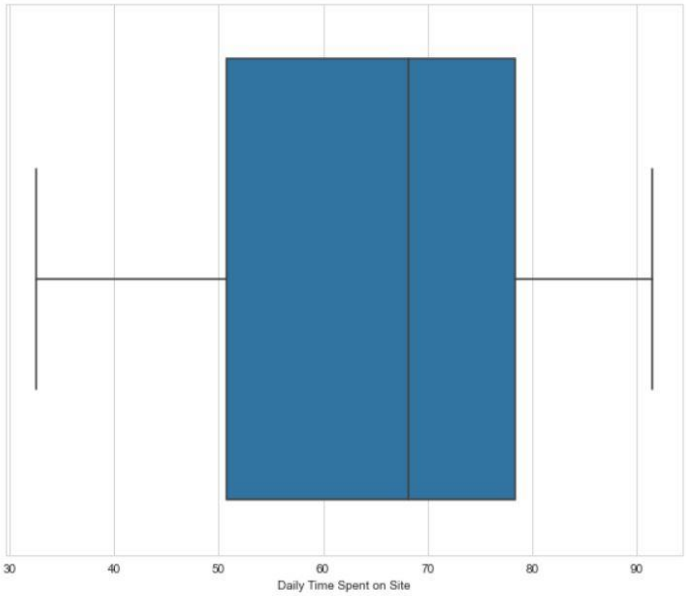


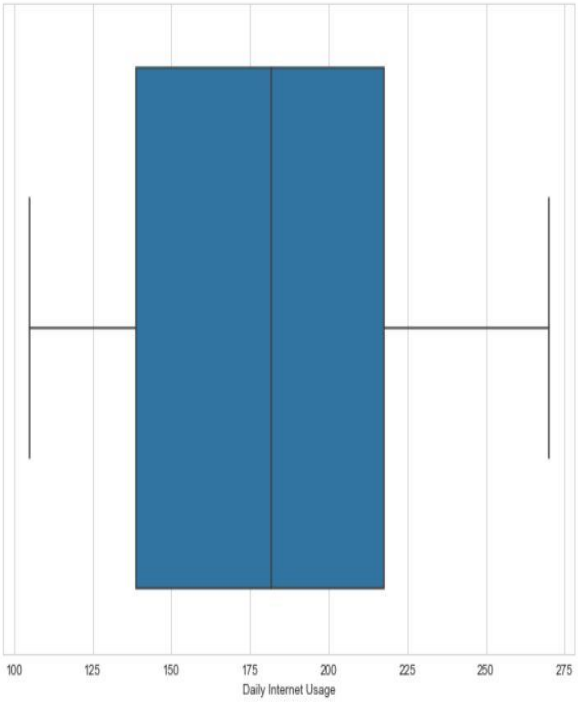


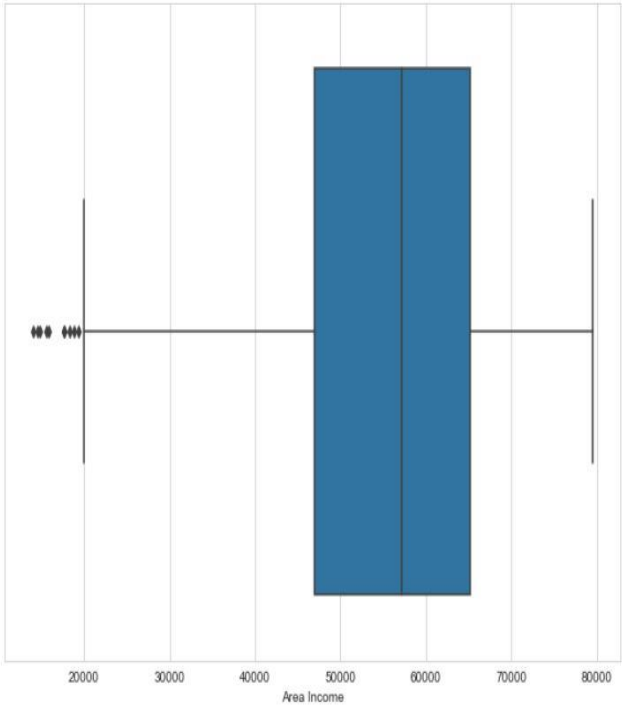


OUTLIER ANALYSIS

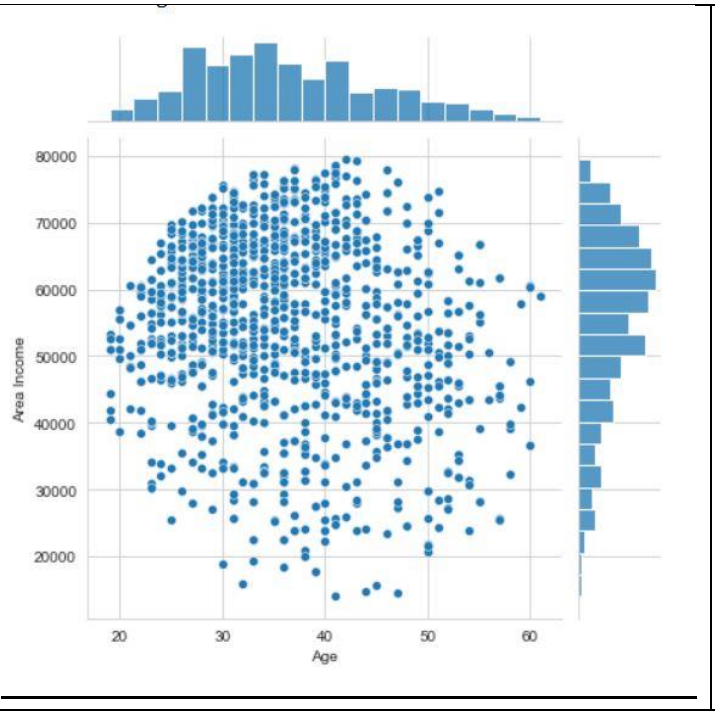


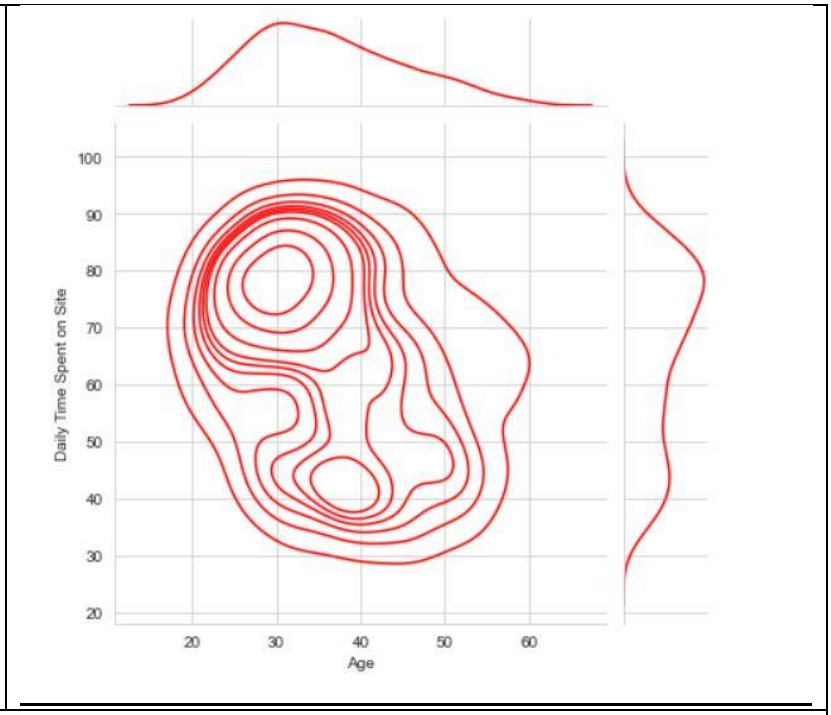


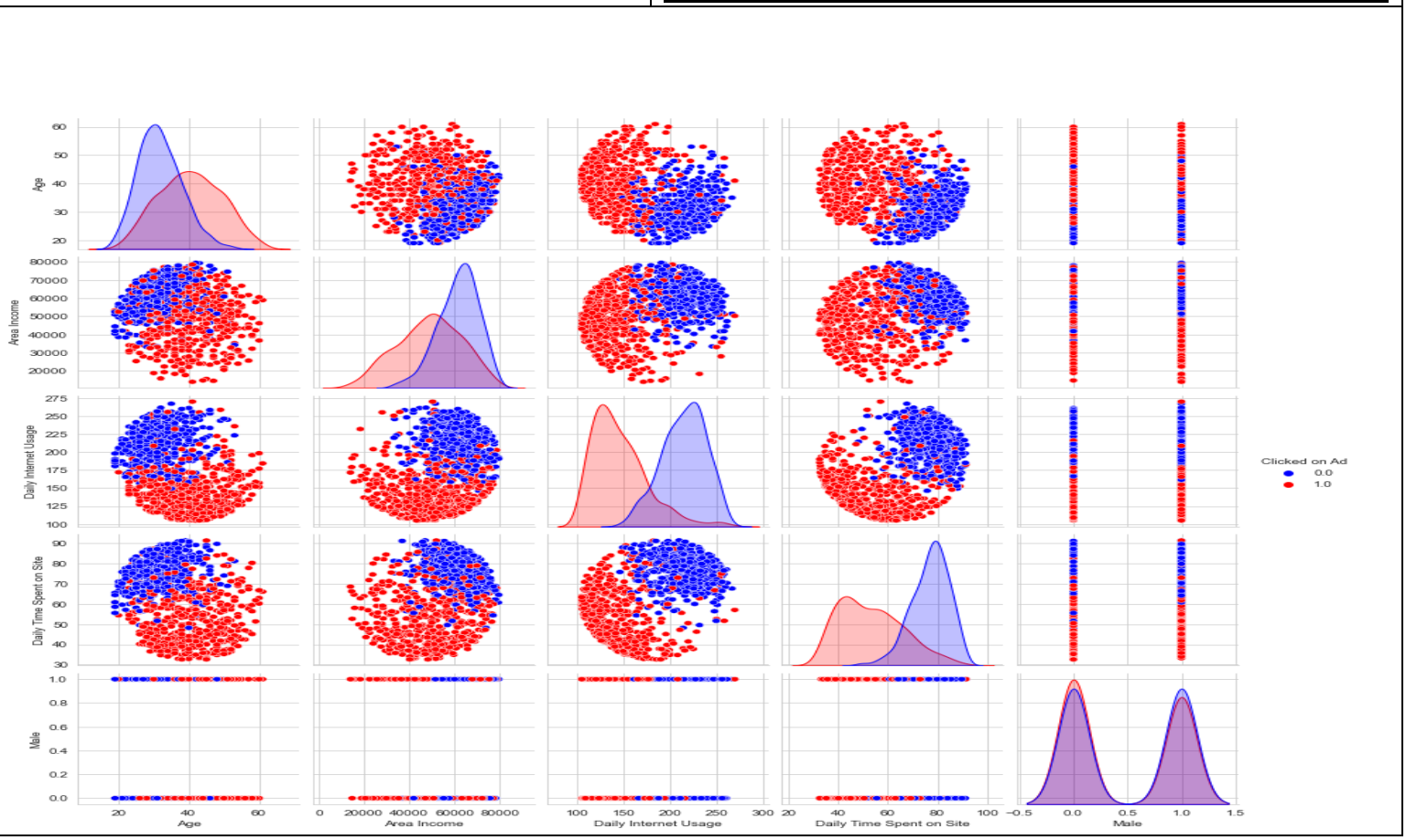




BIVARIATE ANALYSIS







CLASSIFICATION:

ACCURACY SCORE OF LOGISTIC REGRESSION MODEL

Create a classification report for the model.

```
from sklearn.metrics import classification_report, confusion_matrix
```

```
print(classification_report(y_test,y_pred))
```

```
              precision    recall  f1-score   support

           0       0.97      0.97      0.97       164
           1       0.96      0.96      0.96       136

    accuracy                           0.97       300
   macro avg       0.97      0.97      0.97       300
weighted avg       0.97      0.97      0.97       300
```

```
from sklearn.metrics import accuracy_score
print(accuracy_score(y_test, y_pred))
```

```
0.9666666666666667
```

ACCURACY SCORE OF KNN

Classification Report

```
from sklearn.metrics import classification_report
print (classification_report(y_test,pred))
```

```
              precision    recall  f1-score   support

           0       0.69      0.68      0.69       164
           1       0.62      0.64      0.63       136

    accuracy                           0.66       300
   macro avg       0.66      0.66      0.66       300
weighted avg       0.66      0.66      0.66       300
```

Accuracy

```
accuracy_score(y_test, pred)
```

```
0.66
```

ACCURACY SCORE OF RANDOM FOREST

Classification Report

```
from sklearn.metrics import classification_report
print (classification_report(y_test,rfc_pred))
```

	precision	recall	f1-score	support
0	0.96	0.95	0.95	164
1	0.94	0.95	0.95	136
accuracy			0.95	300
macro avg	0.95	0.95	0.95	300
weighted avg	0.95	0.95	0.95	300

Accuracy

```
accuracy_score(y_test, rfc_pred)
```

```
0.95
```

ACCURACY SCORE OF SVM

Classification Report

```
print(classification_report(y_test,predictions))
```

	precision	recall	f1-score	support
0	0.97	0.96	0.97	164
1	0.96	0.96	0.96	136
accuracy			0.96	300
macro avg	0.96	0.96	0.96	300
weighted avg	0.96	0.96	0.96	300

Accuracy

```
accuracy_score(y_test, predictions)
```

```
0.9633333333333334
```

ACCURACY SCORE OF DECISION TREE

Classification Report

```
print(classification_report(y_test,predictions_dtree))
```

	precision	recall	f1-score	support
0	0.95	0.90	0.92	164
1	0.88	0.95	0.91	136
accuracy			0.92	300
macro avg	0.92	0.92	0.92	300
weighted avg	0.92	0.92	0.92	300

Accuracy

```
accuracy_score(y_test, predictions_dtree)
```

```
0.92
```

CONCLUSION:

- **During our analysis, we noticed that most of the Advertisements are clicked by Internet users who ages are between 25 – 45**
- **Most the internet user who clicked the advertisements were females approximately 55%**
- **We have built five different model whose accuracy score are as follows:**

  Logistic Regressions : 97%

  K Nearest Neighbors : 66%

  Random Forest : 95%

  SVM : 96.33%

  Decision Tree : 92%

- **We conclude that the Logistic Regression is the best model followed by SVM.**

THANK YOU...