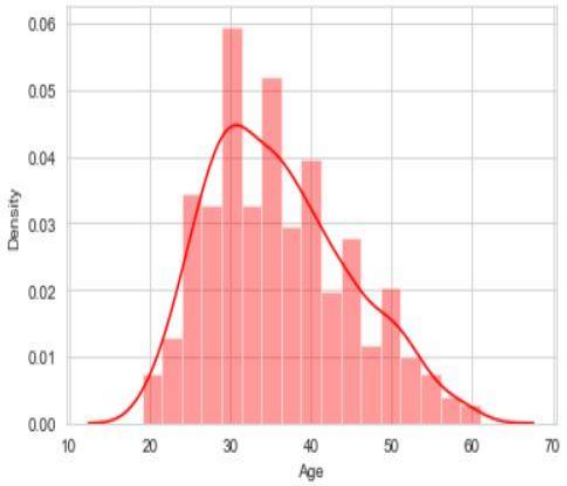
Introduction:

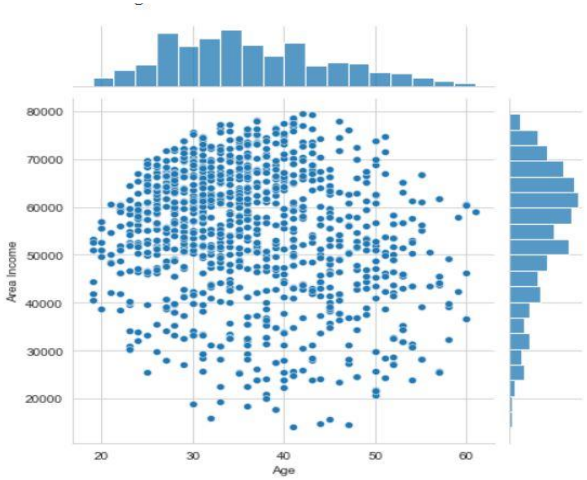
- We have two datasets called Advertising Train and Advertising Test with 1000 and 200 observation respectively.
- Concatenated to Ad_data, which contains 1200 observation and 10 columns.
- Excluded some variables because their analysis is not useful for our objective.
- Goal is to predict the whether or not a particular internet user clicked on an Advertisement on a company website.
- Presentation using Jupyter NoteBook

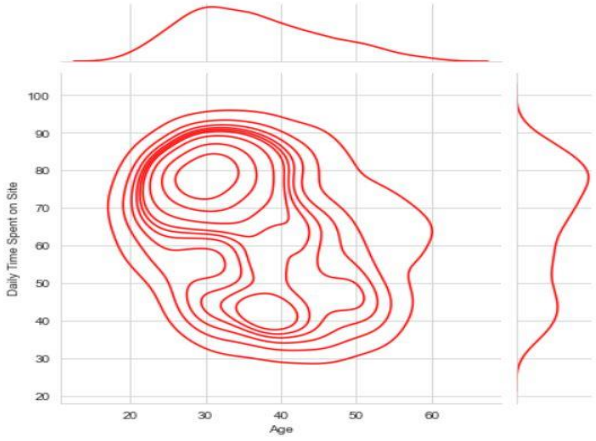
Dataset Attributes:

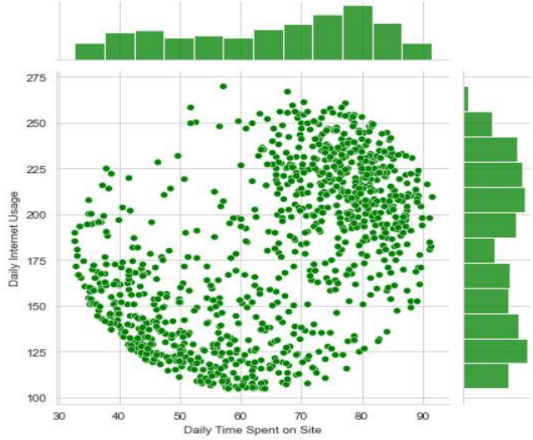
```
ad_data.info()
```

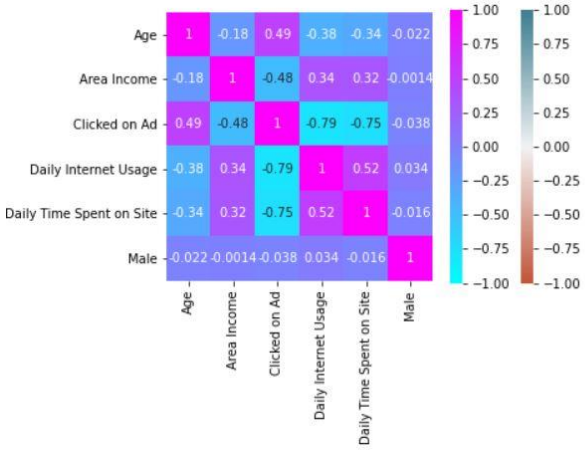
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1200 entries, 0 to 1199
Data columns (total 11 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Ad Topic Line             1200 non-null   object
 1   Age                       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
```











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.

PRESENTED BY: FAHEEM MOHAMMED