```python
#importing the libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import pickle
import warnings
from sklearn import metrics
warnings.filterwarnings('ignore')
```

```python
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import cross_val_score
from sklearn.metrics import classification_report, accuracy_score
```

Load the Exploratory Data Analysis pickle file and build the model

```python
df_eda = pickle.load(open('C:\\Users\\acer\\Desktop\\Vechicle_Insurance_DataSet\\models\\ExploratoryDataAnalysis.pkl','rb'))
```

```python
df_eda.shape
```

(370789, 12)

```python
df_eda.size
```
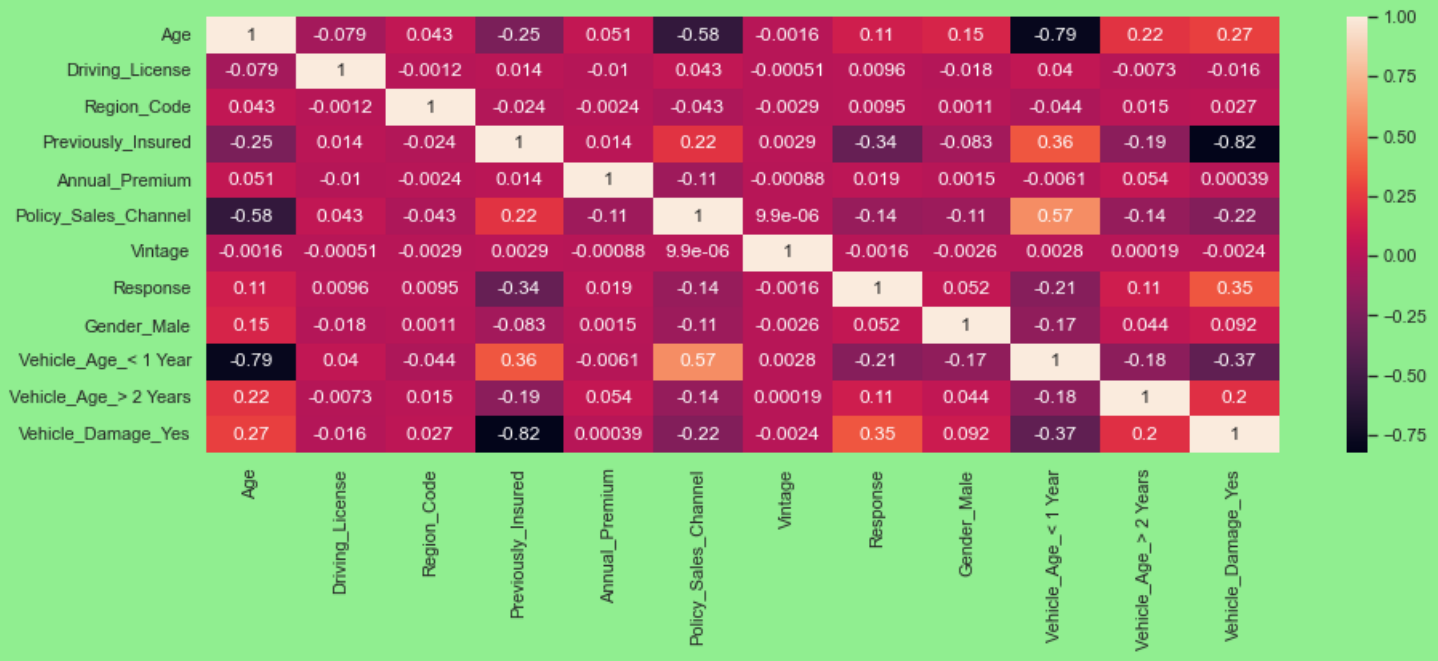
4449468

```python
df_eda.isnull().sum().sum()
```

0

```python
sns.set(rc={'figure.facecolor':'lightgreen'})
plt.figure(figsize=(15,5))
sns.heatmap(df_eda.corr(),annot=True)
```

<AxesSubplot:>

```python
# plt.figure(figsize=(30,5))
# sns.boxplot(data=df_eda)
```

```python
#pd.concat([X,y],axis=1)
```

```python
#independent and dependent variable
X=df_eda.drop(columns=['Response'],axis=1)
y=df_eda['Response']
```

In [48]:

```python
#splitting the data into training and testing data
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.3,random_state=1)
```

In [49]:

```python
#size of the train and test data
X_train.shape,X_test.shape,y_train.shape,y_test.shape
```

Out[49]:

((259552, 11), (111237, 11), (259552,), (111237,))

In [51]:

```python
#building a Random Forest classifier model

rfc=RandomForestClassifier()
rfc.fit(X_train,y_train)
```

Out[51]:

▼ RandomForestClassifier

RandomForestClassifier()

In [55]:

```python
#predicting the outcome value
y_pred=rfc.predict(X_test)
y_pred
```

Out[55]:

array([0, 0, 0, ..., 0, 0, 0], dtype=int64)

In [53]:

```python
#evaluating the performance of the training data
rfc.score(X_train,y_train)
```

Out[53]:

0.9998921218098878

In [56]:

```python
#evaluating the performance of the test data/predicting data
accuracy_score(y_test,y_pred)
```

Out[56]:

0.867472154049462

In [58]:

```python
print(classification_report(y_test,y_pred))
metrics.confusion_matrix
```
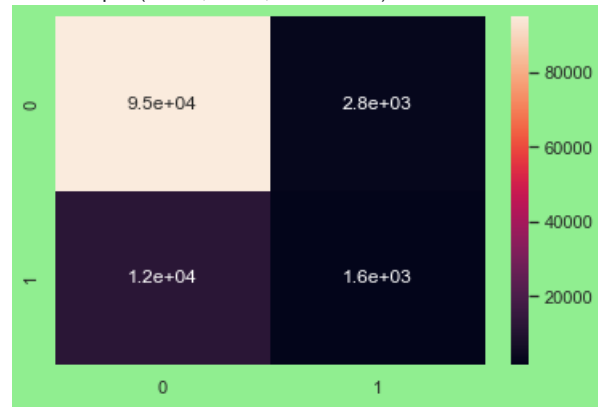
	precision	recall	f1-score	support
0	0.89	0.97	0.93	97751
1	0.36	0.12	0.18	13486
accuracy			0.87	111237
macro avg	0.62	0.54	0.55	111237
weighted avg	0.82	0.87	0.84	111237

In [62]:

```python
print(sns.heatmap(metrics.confusion_matrix(y_test,y_pred),annot=True))
```

AxesSubplot(0.125,0.125;0.62x0.755)

Loading the test model prediction data in pickle file

```
# with open(r'C:\Users\acer\Desktop\Vechicle_Insurance_DataSet\models\test_model.pkl','wb') as file:
#     pickle.dump(rfc,file)
```