Group Name: Data Science Group (Boshra and Omer) - Data Glacier

Name:

1- Boshra Abdualrahman Eisa

2- Omer Salih Dawood Omer

Email: bush.eisa9@gmail.com , omercomail@gmail.com

Country: Saudi Arabia

College/Company: Prince Sattam Bin Abdulaziz University

Specialization: Data Science

GitHub Link: https://github.com/boshraEisa/BankMarketing

Problem Description

ABC Bank wants to sell its term deposit product to customers and before launching the product they want to develop a model which help them in understanding whether a particular customer will buy their product or not.

Data Cleansing

For data cleansing, we started off by ensuring that there's no null values in the data and we handled outliers and skewness by using the square root function.

```
dataset.isnull().sum()
```

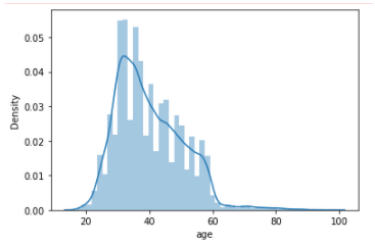
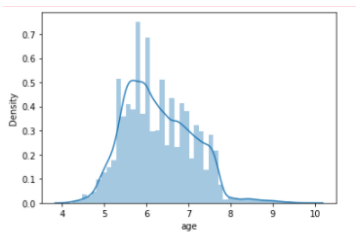
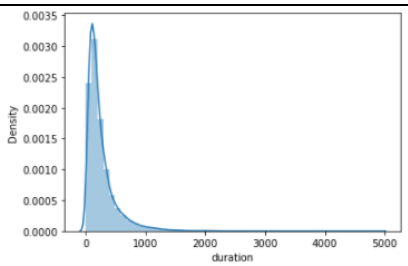
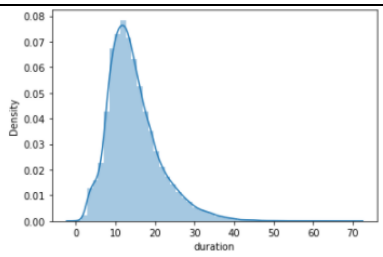
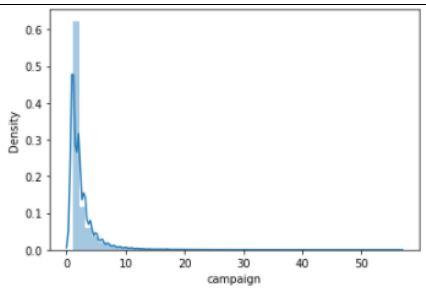
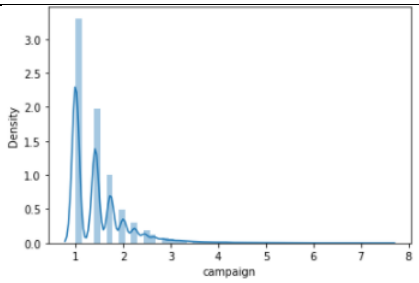
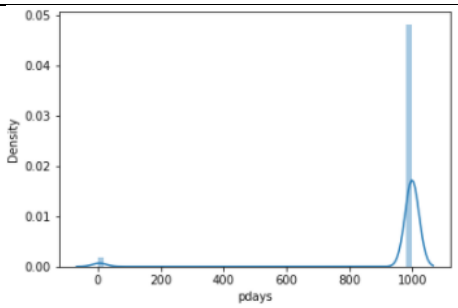
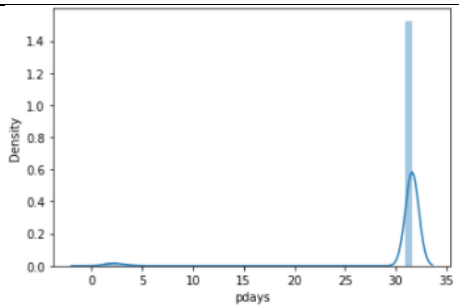
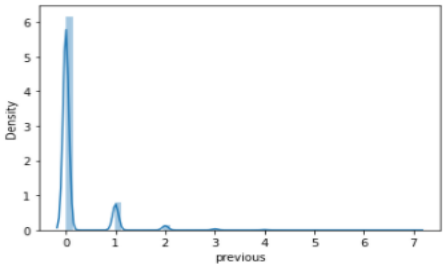
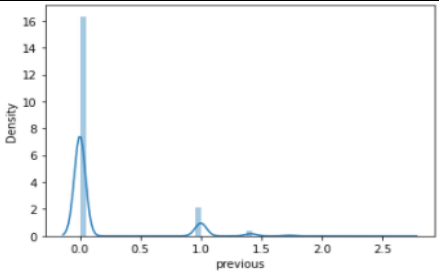
```
age               0
job               0
marital           0
education         0
default           0
housing           0
loan              0
contact           0
month             0
day_of_week       0
duration          0
campaign          0
pdays             0
previous          0
poutcome          0
emp.var.rate      0
cons.price.idx    0
cons.conf.idx     0
euribor3m         0
nr.employed       0
y                 0
dtype: int64
```

Dealing with the skewness

```
sqrAge = np.sqrt(subsetData['age'])
```

```
skew(sqrAge)
```

```
0.440619036249743
```

	Before Square root	After square root
Age	 0.7846682380932293	 0.440619036249743
Duration	 3.2630224157610432	 1.1995245640956793
campaign	 4.762333252560967	 2.1685028501394124
pdays	 -4.922010656450046	 -4.927171203655262
previous	 3.8319026847007036	 2.3886653743053796

After that, we removed any unknown values in the data frame since it will only hinder the analysis

```
# removal of unknowns values function
def removal_unknowns(dataset, column):
    col_values = dataset[column].values
    dataset[column] = np.where(col_values=='unknown',
                               dataset[column].mode(), col_values)
    return dataset
```

We also changed the names of the columns to make sure they're readable and easy to understand

```
mapping = {dataset.columns[15]:'employment_variation_rate',
           dataset.columns[16]: 'consumer_price_index',
           dataset.columns[17]:'consumer_confidence_index',
           dataset.columns[19]: 'number_of_employees',
           dataset.columns[20]: 'subscribed' }
dataset = dataset.rename(columns=mapping)
```