Deliverable Week 10

Group Name: Destined Data Team
Specialization: Data Science

Team members:

1. **Name:** Praneetha Rajupalepu
Email: praneetha.724@gmail.com
Country: Canada
Company: Modest Tree
Specialization: Data Science

2. **Name:** Selaelo Kholofelo Ramokgopa
Email: sly.kholo@gmail.com
Country: South Africa
College: University of Johannesburg
Specialization: Chemical Engineering

3. **Name:** Surya Chandra
Email: ksuryachandra619@gmail.com
Country: Germany
College: Otto von Guericke University
Specialization: Electrical Engineering and Information Technology

4. Name: Cagla Yucel
Email: caglasipahi@gmail.com
Country: Switzerland
College: Georgia Institute of Technology
Specialization: Analytics

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 will help them in understanding whether a particular costumer will buy their product or not (based on customer's past interaction with bank and other Financial Institutions).

Data Cleaning and Transformation

The dataset was checked for missing values, duplicates, outliers, skewness. Many machine learning algorithms do not understand categorical data. Hence, they must be converted into integer values. The unknown values are handled and columns with categorical data are changed to Boolean or integer values. The dataset is cleaned and any ML model can use it for training and prediction.

GitHub repo link: https://github.com/PraneethaRajupalepu/Bank-DataScience-Project

Results and approaches

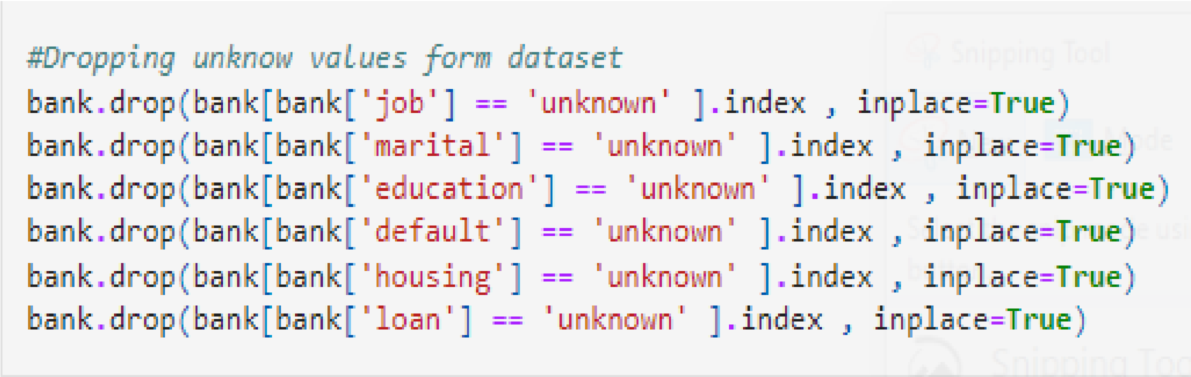
- EDA was performed for continuous and categorical variables

- Methods for data cleaning and transformation

    - Step 1: Removed unknown values from the Dataset

      Rows (Job, Marital, Education, Default, Housing, and Loan) with “unknown” values are removed from the Dataset using drop function as shown in Figure 1.

Figure 1 – Python Code – Unknown Values

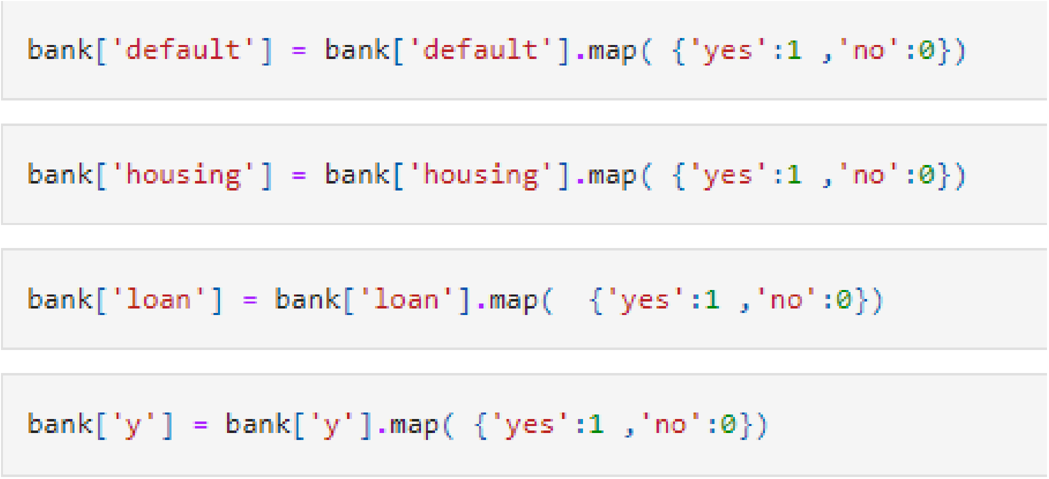


```
#Dropping unknow values form dataset
bank.drop(bank[bank['job'] == 'unknown' ].index , inplace=True)
bank.drop(bank[bank['marital'] == 'unknown' ].index , inplace=True)
bank.drop(bank[bank['education'] == 'unknown' ].index , inplace=True)
bank.drop(bank[bank['default'] == 'unknown' ].index , inplace=True)
bank.drop(bank[bank['housing'] == 'unknown' ].index , inplace=True)
bank.drop(bank[bank['loan'] == 'unknown' ].index , inplace=True)
```

- Step 2: Binning of outliers

  Rows (Default, Housing, y, and Loan) which has only two values are replaced with 1s and 0s using map function as shown in figure 2.

Figure 2 – Python Code – Mapping



```
bank['default'] = bank['default'].map( {'yes':1 ,'no':0})
```

```
bank['housing'] = bank['housing'].map( {'yes':1 ,'no':0})
```

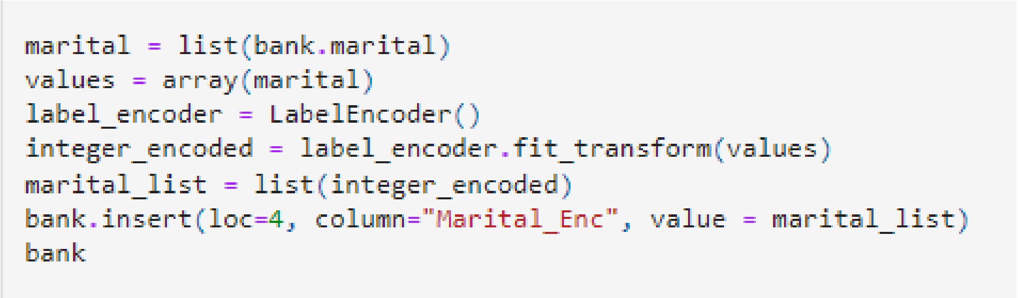
```
bank['loan'] = bank['loan'].map(  {'yes':1 ,'no':0})
```

```
bank['y'] = bank['y'].map( {'yes':1 ,'no':0})
```

- Step 3: One hot encoding

  Categorical values are mapped to integer values using one hot encoding technique using scikit learn as shown in figure 3.

Figure 3 – Python Code – One Hot Encoding



```
marital = list(bank.marital)
values = array(marital)
label_encoder = LabelEncoder()
integer_encoded = label_encoder.fit_transform(values)
marital_list = list(integer_encoded)
bank.insert(loc=4, column="Marital_Enc", value = marital_list)
bank
```

The dataset can now be used to train machine learning models and predict whether a particular costumer will buy their product or not.