

WEEK 8 : DELIVERABLES

BANK MARKETING CAMPAIGN

‘DATA SCIENCE’

GROUP NAME: DATA SCIENCE MASTER

NAME : ABHIMANYU GANGANI

EMAIL : Agangani97@gmail.com

COUNTRY : UNITED KINGDOM

COLLEGE : ANGLIA RUSKIN UNIVERSITY

SPECIALIZATION : DATA SCIENCE

**PROBLEM DESCRIPTION :**

ABC Bank wants to sell it's 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.

To achieve this task they have consulted an analytics consultancy to automate the process of classification.

The Analytics company have to come up with an ML model to shortlist the customers whose chances to buy the product is higher. This will lead marketing team to target on the given lead.

**BUSINESS UNDERSTANDING :**

There’s been a revenue decline for the ABC bank and to overcome that they want to come up with the actions needed to be taken. With analysis they came to know that customers are not depositing as frequently as before. Banks make investments from the investment made by customers to make high profits.

Banks also urges customers to buy other products such as insurance and Different kind of deposits. They want to check the customers from existing data they pursue and filter the customers having higher chances of buying any new schemes or products from the bank.

**DATA UNDERSTANDING :**

Data belongs to a banking organisation and corresponds to marketing campaigns. These campaigns are based on phone calls. More than one call to the same client tells whether the bank term deposit (product) was subscribed by client or not.

There are four datasets provided for this classification problem. We are having 2 pairs of test and train datasets.

Bank.csv and Bank\_full.csv are one pair having 16 features and Bank\_additional.csv and Bank\_additional\_full.csv are having 20 features. Bank.csv is the older version of bank\_additional.csv.

Below are the details of all four datasets:

|  |  |  |
| --- | --- | --- |
| File | Dataset Type | Description |
| Bank.csv | Test | 4521  observations(10% of train data) and 16  features |
| Bank\_full.csv | Train | 45211 observations and 16 features |
| Bank\_additional.csv | Test | 4111  observations(10% of train data) and 20  features |
| Bank\_additional\_full.csv | Train | 41118 observations and 20 features |

**Datatype and Description of columns:**

Data columns (total 21 columns):

# Column Dtype Description

--- ------ ----- -----------

1. age int64 Age of Client.
2. job object Type of Job.
3. marital object Marital Status.
4. education object Level of Education.
5. default object Has credit in default?
6. housing object Has housing loan?
7. loan object Has personal loan?
8. contact object How client has been communicated?
9. month object last contacted month.
10. day\_of\_week object last contacted day.
11. duration int64 duration of communication(seconds).
12. campaign int64 number of contacts performed in

Campaign.

1. pdays int64 number of days passed after contact.
2. previous int64 number of total contacts performed.
3. poutcome object outcome of the previous campaign.
4. emp.var.rate float64 Employment variation rate. 16 cons.price.idx float64 Consumer price index.
5. cons.conf.idx float64 Consumer confidence index.
6. euribor3m float64 Euribor 3 months rate.
7. nr.employed float64 number of employees.
8. y object has the client subscribed product. dtypes: float64(5), int64(5), object(11)

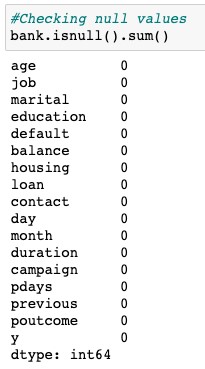
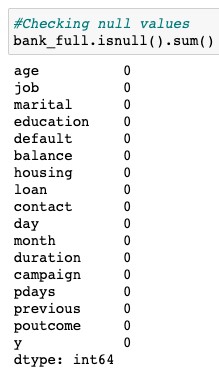
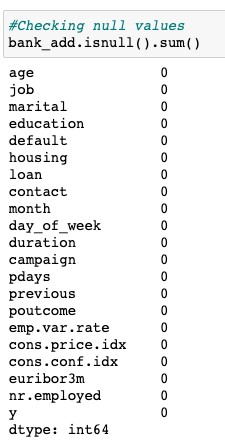
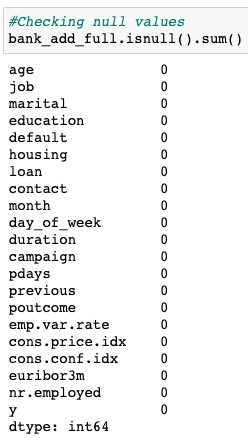
* + First 7 features are the client information.
  + Features 8-11 are last contact information.
  + Features 12-15 are other important details regarding contact.
  + Features 16-20 are economic and social features.
  + The 21st feature is the target variable(dependent).

**Data Problems :**

**Missing Attribute:**

None of the dataset contains any missing value.

**Value Counts :**



Some of the variables consists of value counts as “Unknown” which is significantly high. ***So we assume “Unknown” as another category for these variables.***

admin. 10422 blue-collar 9254 technician 6743 services 3969 management 2924 retired 1720 entrepreneur 1456 self-employed 1421 housemaid 1060 unemployed 1014 student 875 unknown 330

Name: job, dtype: int64

------------------------------ married 24928 single 11568 divorced 4612 unknown 80

Name: marital, dtype: int64

------------------------------ university.degree 12168 high.school 9515 basic.9y 6045 professional.course 5243 basic.4y 4176 basic.6y 2292 unknown 1731 illiterate 18

Name: education, dtype: int64

------------------------------ no 32588 unknown 8597 yes 3

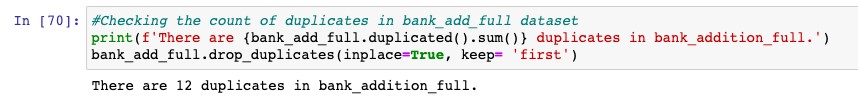
Name: default, dtype: int64

------------------------------ yes 21576 no 18622 unknown 990

Name: housing, dtype: int64

------------------------------ no 33950 yes 6248 unknown 990 Name: loan, dtype: int64

**Duplicate Counts :**



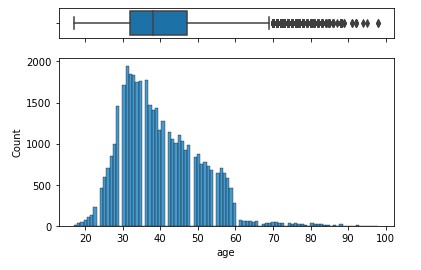
***There are 12 duplicates present in the bank\_additonal\_full dataset, we will remove the duplicates using drop\_duplicates function.***

**Outliers :**

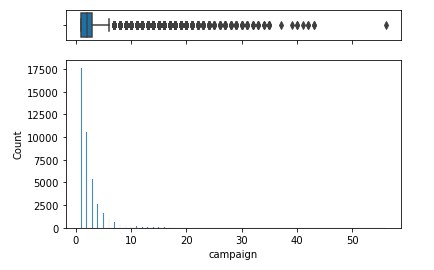
Outliers are the values which lie at above 3 standard deviation distance fr om the other Values in normal distribution.

It might occur due to improper collection of the data. . Outliers can disturb our analysis by changing the mean on normal distribution graph. F ollowing variables consists of significant outliers.

* ‘Age’ Feature :



* ‘Campaign Feature :



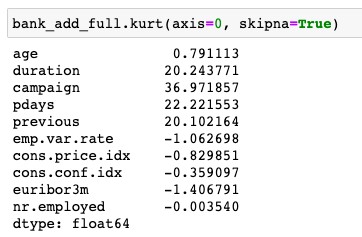
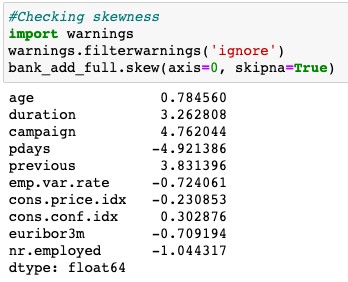
***The maximum value for ‘age’ variable is 98 and that of ‘campaign’ variable is 56 and both are significant values.***

Since model is needed to be generalized to reflect the real world data we are not going to remove these outliers as these seems to be realistic value

.

**Skewness and Kurtosis:**

Skewness is a **measure of symmetry, or more precisely, the lack of symmetry**. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point. Kurtosis is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution.



GITHUB LINK : **https://github.com/AbhimanyuGangani/Week\_7**

**\_Bank\_Marketing/tree/main/Week\_8\_Bank\_Marketing**