***Data Analytics Project***

***bank-direct-marketing-campaigns***

1. **The describe of the DataSet**

This data set contains records relevant to a direct marketing campaign of a Portuguese banking institution. The marketing campaign was executed through phone calls. Often, more than one call needs to be made to a single client before they either decline or agree to a term deposit subscription. The classification goal is to predict if the client will subscribe (yes/no) to the term deposit (variable y)

Input variables:

bank client data:  
1 - age (numeric)  
2 - job : type of job (categorical: 'admin.','blue-collar','entrepreneur','housemaid','management','retired','self-employed','services','student','technician','unemployed','unknown')  
3 - marital : marital status (categorical: 'divorced','married','single','unknown'; note: 'divorced' means divorced or widowed)  
4 - education (categorical: 'basic.4y','basic.6y','basic.9y','high.school','illiterate','professional.course','university.degree','unknown')  
5 - default: has credit in default? (categorical: 'no','yes','unknown')  
6 - housing: has housing loan? (categorical: 'no','yes','unknown')  
7 - loan: has personal loan? (categorical: 'no','yes','unknown')

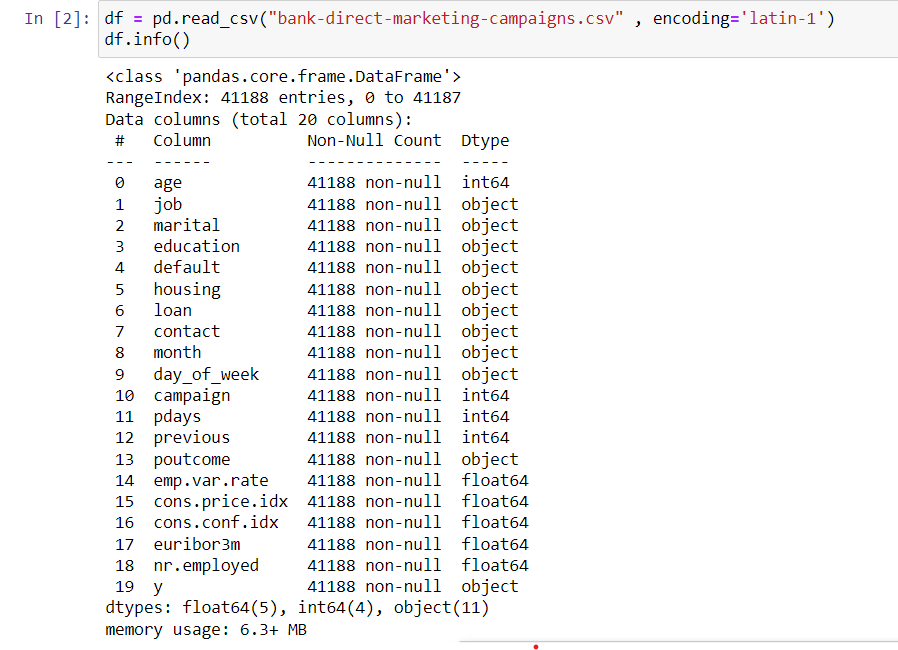
related with the last contact of the current campaign:  
8 - contact: contact communication type (categorical: 'cellular','telephone')  
9 - month: last contact month of year (categorical: 'jan', 'feb', 'mar', …, 'nov', 'dec')  
10 - day\_of\_week: last contact day of the week (categorical: 'mon','tue','wed','thu','fri')

other attributes:  
11 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)  
12 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)  
13 - previous: number of contacts performed before this campaign and for this client (numeric)  
14 - poutcome: outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')

social and economic context attributes:  
15 - emp.var.rate: employment variation rate - quarterly indicator (numeric)  
16 - cons.price.idx: consumer price index - monthly indicator (numeric)  
17 - cons.conf.idx: consumer confidence index - monthly indicator (numeric)  
18 - euribor3m: euribor 3 month rate - daily indicator (numeric)  
19 - nr.employed: number of employees - quarterly indicator (numeric)

Output variable (desired target):

20 - y - has the client subscribed a term deposit? (binary: 'yes','no')



1. **The objectives of the DataSet**

Predictive Modeling: Develop a machine learning model to predict if a client will subscribe to a term deposit based on the available features.

Feature Importance: Identify which features in the data set have the most significant impact on the decision of clients to subscribe to a term deposit.

Campaign Analysis: Analyze the effectiveness of the marketing campaign by assessing the number of contacts made, the outcome of the previous marketing campaign, and the social and economic context attributes.

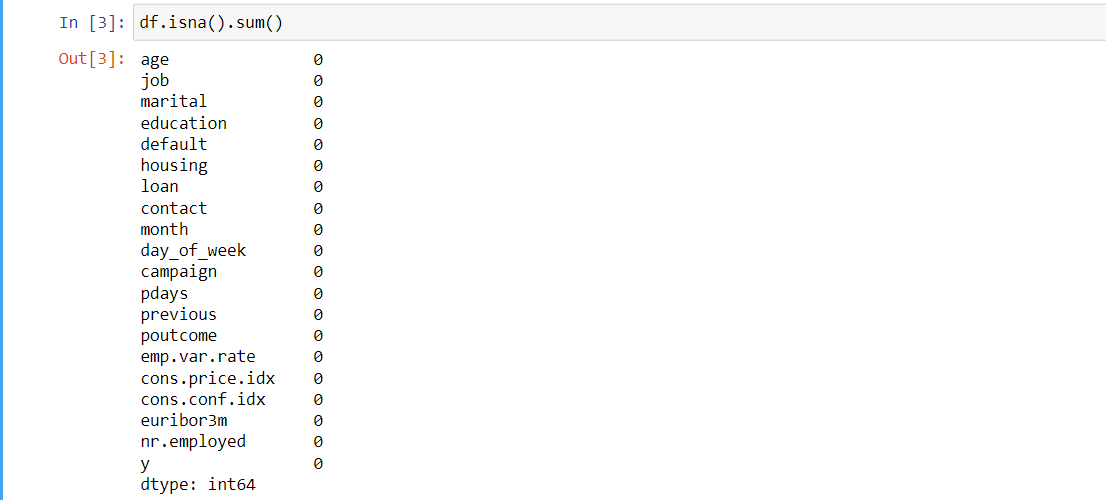
Customer Segmentation: Segment the clients based on their demographic and other characteristics to understand the behavior of different groups regarding term deposit subscriptions.

Feature Engineering: Create new features that may help to improve the performance of the machine learning model in predicting term deposit subscriptions.

Business Insights: Gain insights into factors that influence clients' decisions to subscribe to a te

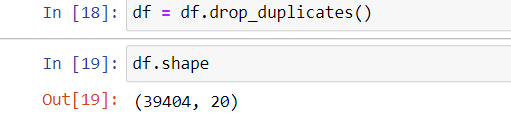
1. **DataSet Preprocessing**

* Remove nulls



-There is no nulls to remove

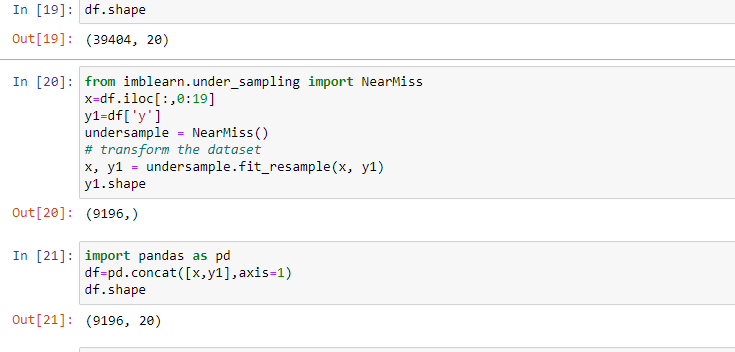
* Remove duplicates



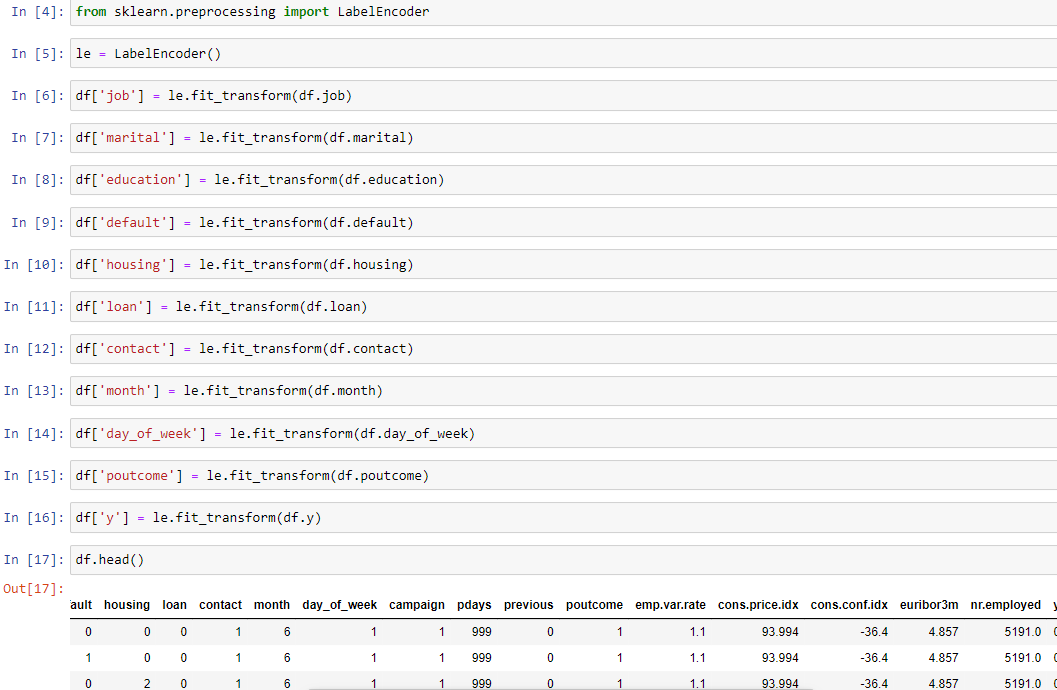
Data reduced from 41188 to 39404

* Undersampling

-data unbalanced so we use undersampling



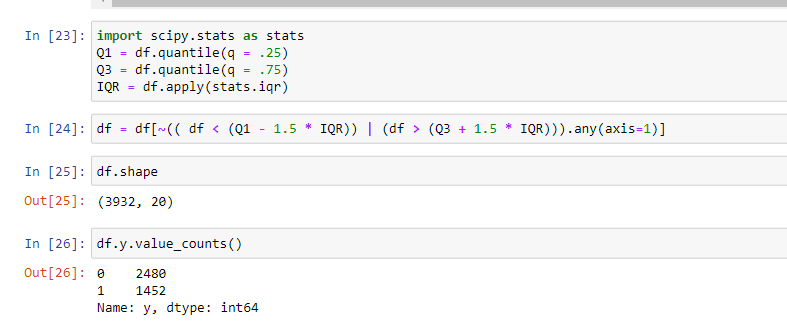
1. **DataSet Transformation**



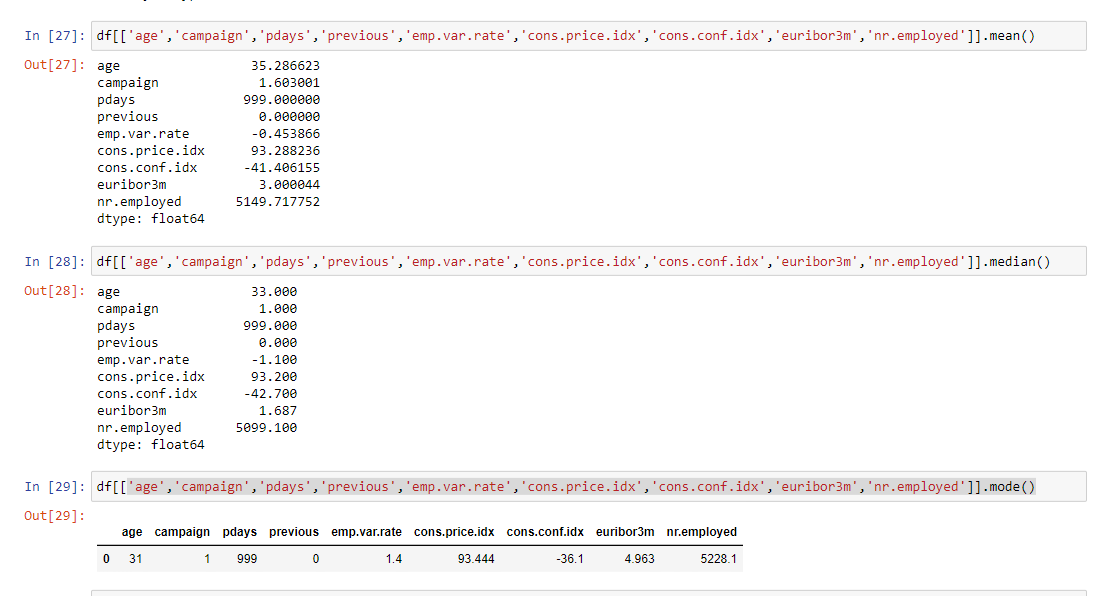
-The data needed to apply transformation and we use label encoder method

1. **Anomaly Detection**

-the data has outliers so we use IQR to remove it

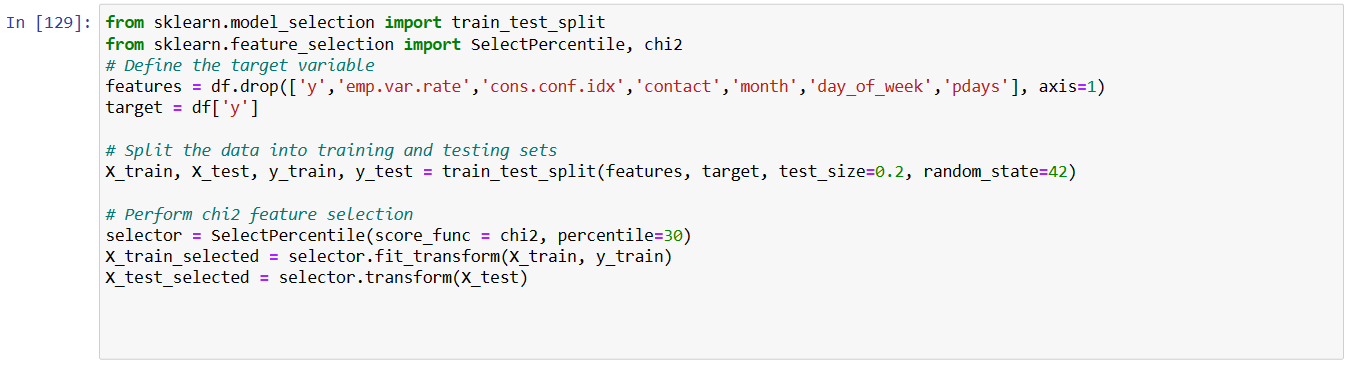


1. **Data Statistics**



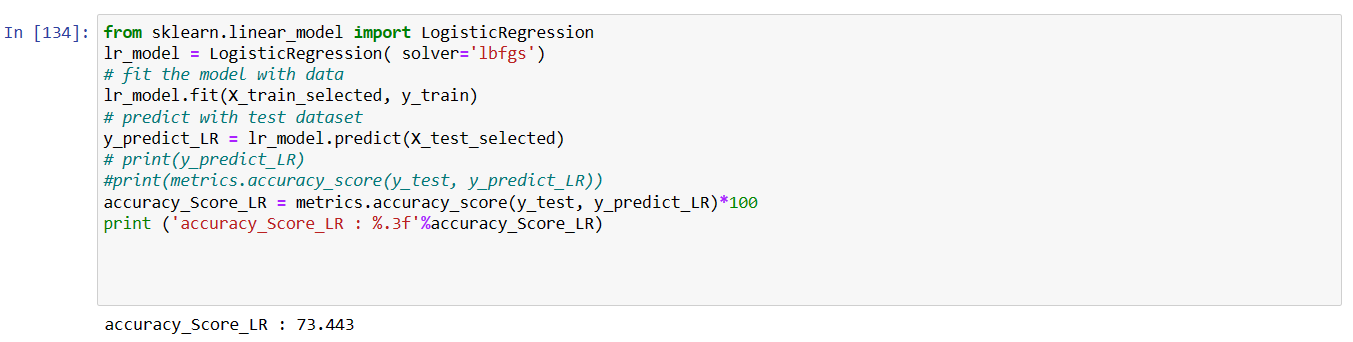


1. **Prepare the data and Split the data into training and testing sets and Perform chi2 feature selection**

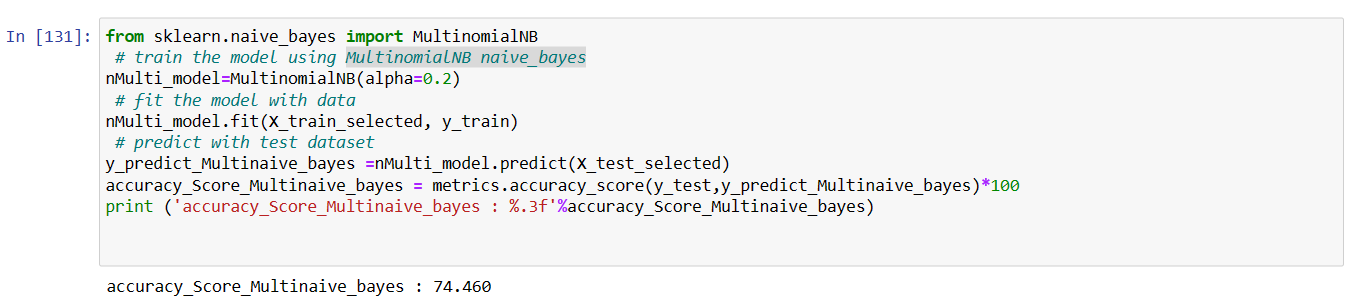


1. **Apply algorithms :**

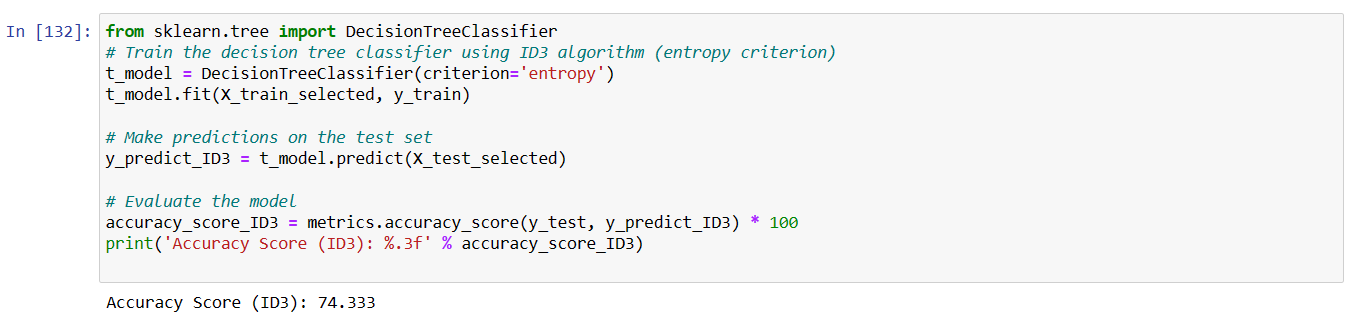
* **Logistic Regression**



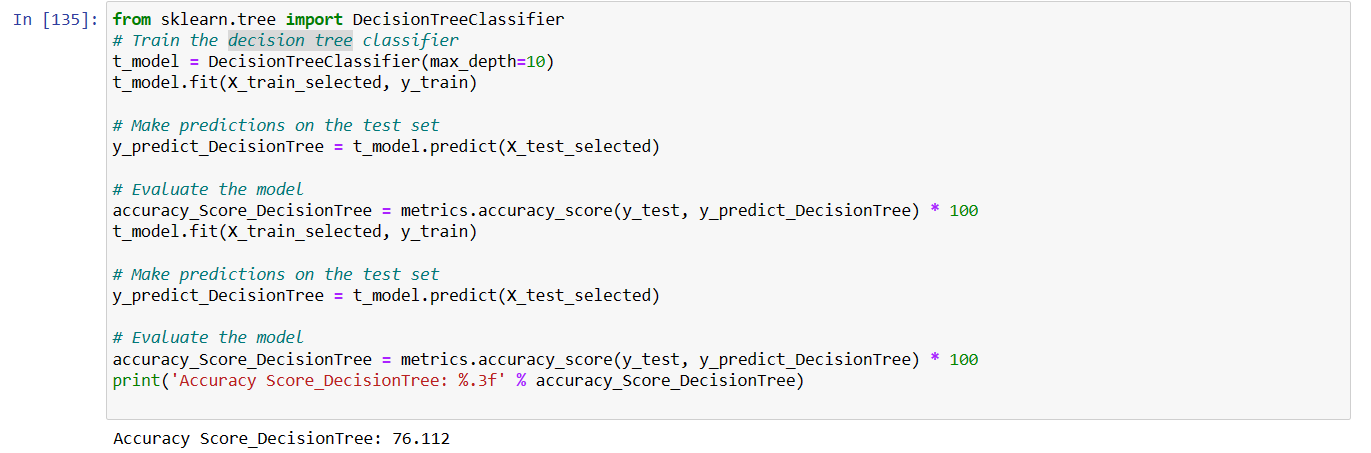
* **MultinomialNB naive\_bayes**

****

* **ID3**

****

* **decision tree**

****