Capstone Project (Week 2)
Bank Marketing dataset

Introduction

- In banks it is important to meet the goals that are set
- For meeting those goals banks market their product throgh telemarketing
- For effective marketing, banks have to identify the customers who are very like to take
- By targeting only those customers banks can get huge margin of profit

Data

- Dataset used here is of a portuguese bank

- They recorded the outcome of direct marketing campaign

- Dataset contains 45211 records and 17 attributes

- Publicly available dataset found at UCI Machine Learning Repository

Variables

<u>Age</u> – Numeric attribute

<u>job</u> – categorical attribute having type of job they have

<u>marital</u> – Marital status of customer (Married, single, divorced)

<u>education</u> – categorical attribute (Primary, secondary, tertiary, unknown)

<u>default</u> – categorical attribute (has credit in default or not)

<u>balance</u> – Numeric attribute (average yearly balance in euros)

<u>housing</u> – categorical attribute (has housing loan or not)

<u>loan</u> – categorical attribute (has personal loan or not)

Variables cont..

contact – mode of contact (cellular or telephone)

day – numeric (last contact day of the month)

month – categorical (last contact month of the year)

duration – Numeric (contact duration)

campaign – numeric (number of times contacted during the campaign)

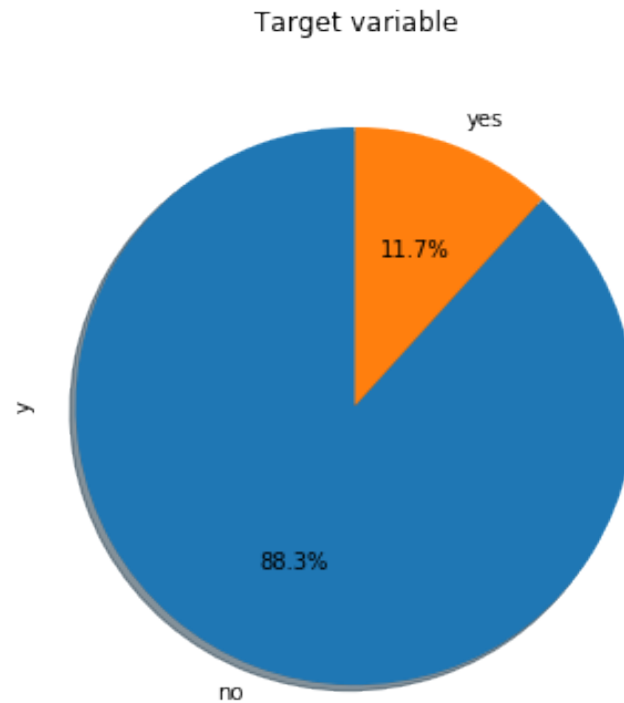**pdays** – numeric (number of days passed by after contacting)

previous – numeric (number of contacts performed before the campaign)

poutcome – categorical (out come of previous campaign)

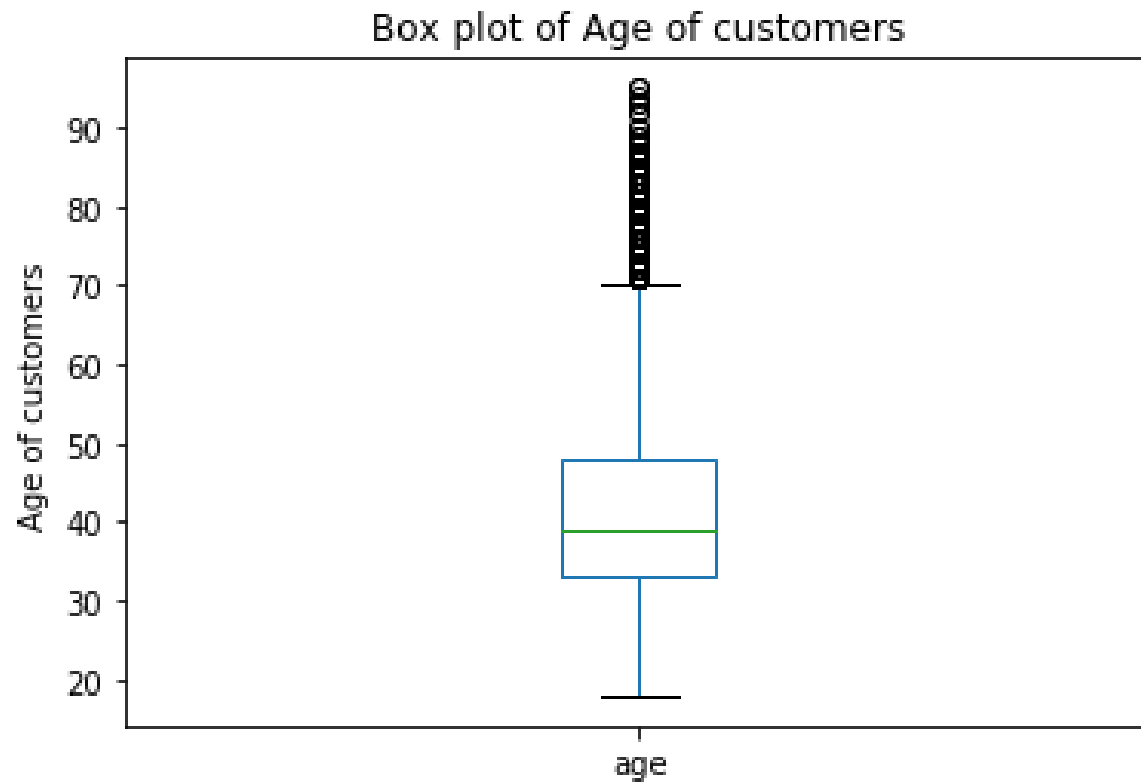**y** – categorical - customer subscribed to product or not **(target variable)**

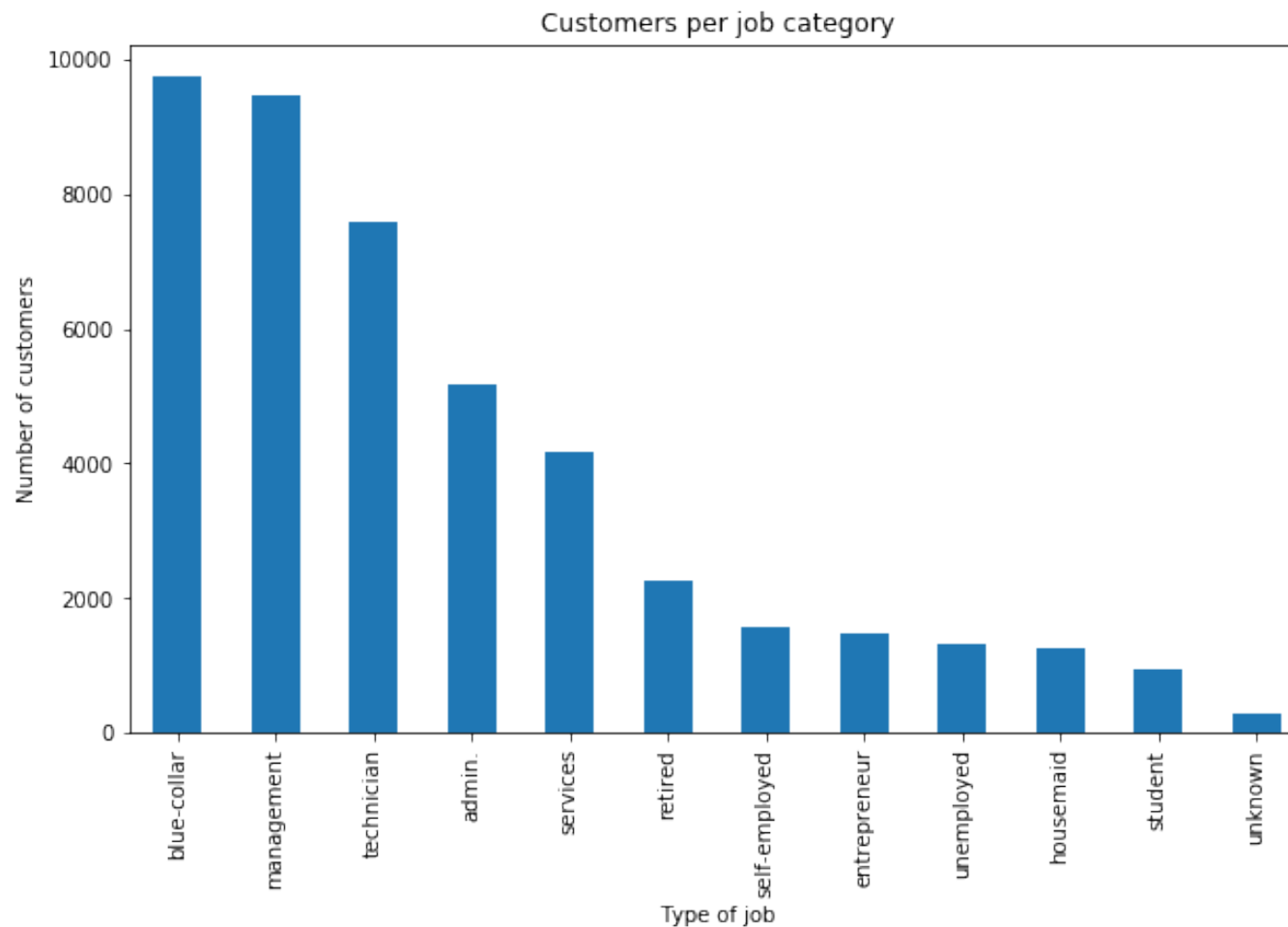# Data Exploration

- Target variable exploration

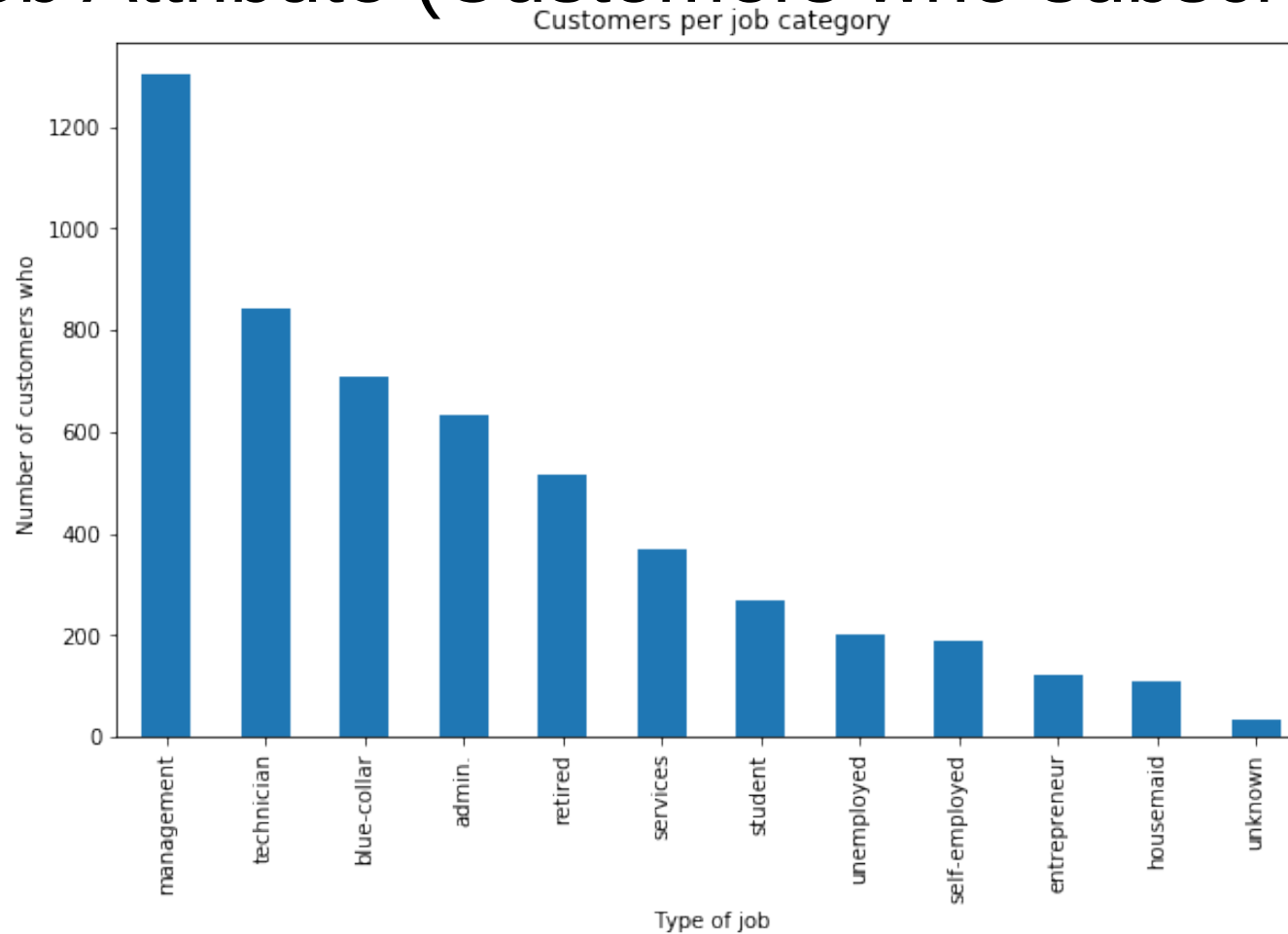# Data Exploration

- Age Attribute (Min: 18, Max: 94)

Box plot of Age of customers

Data Exploration

- Job Attribute (12 categories)

Data Exploration

- Job Attribute (Customers who subscribed)

Customers per job category
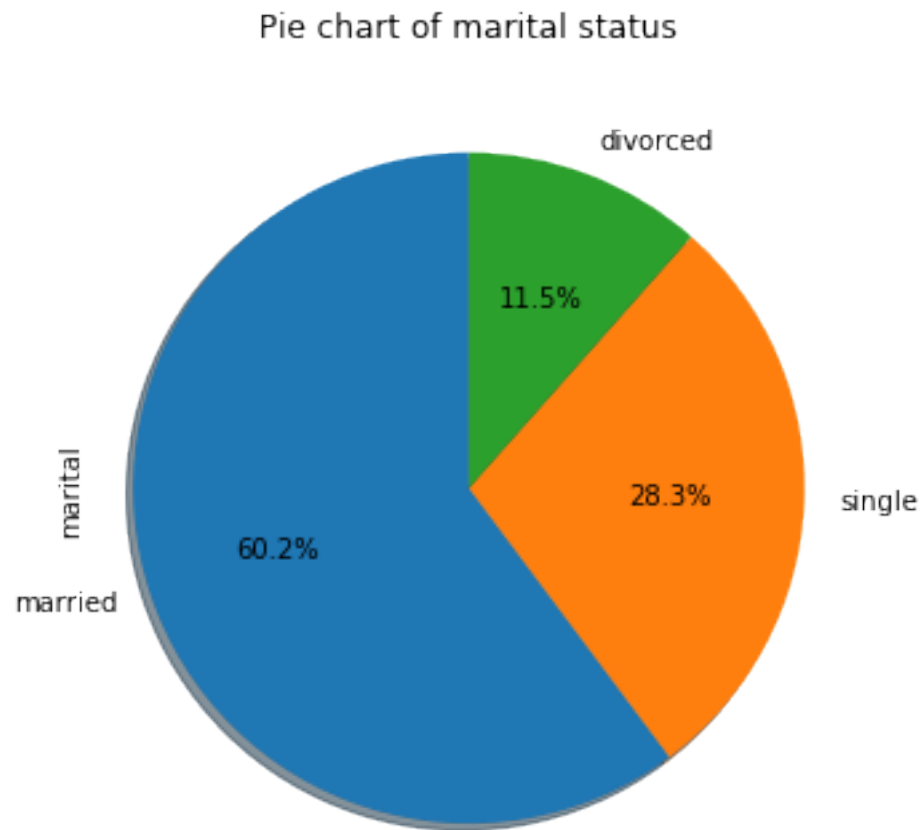
Data Exploration

- Job Attribute (Customers who not subscribed)

Customers per job category
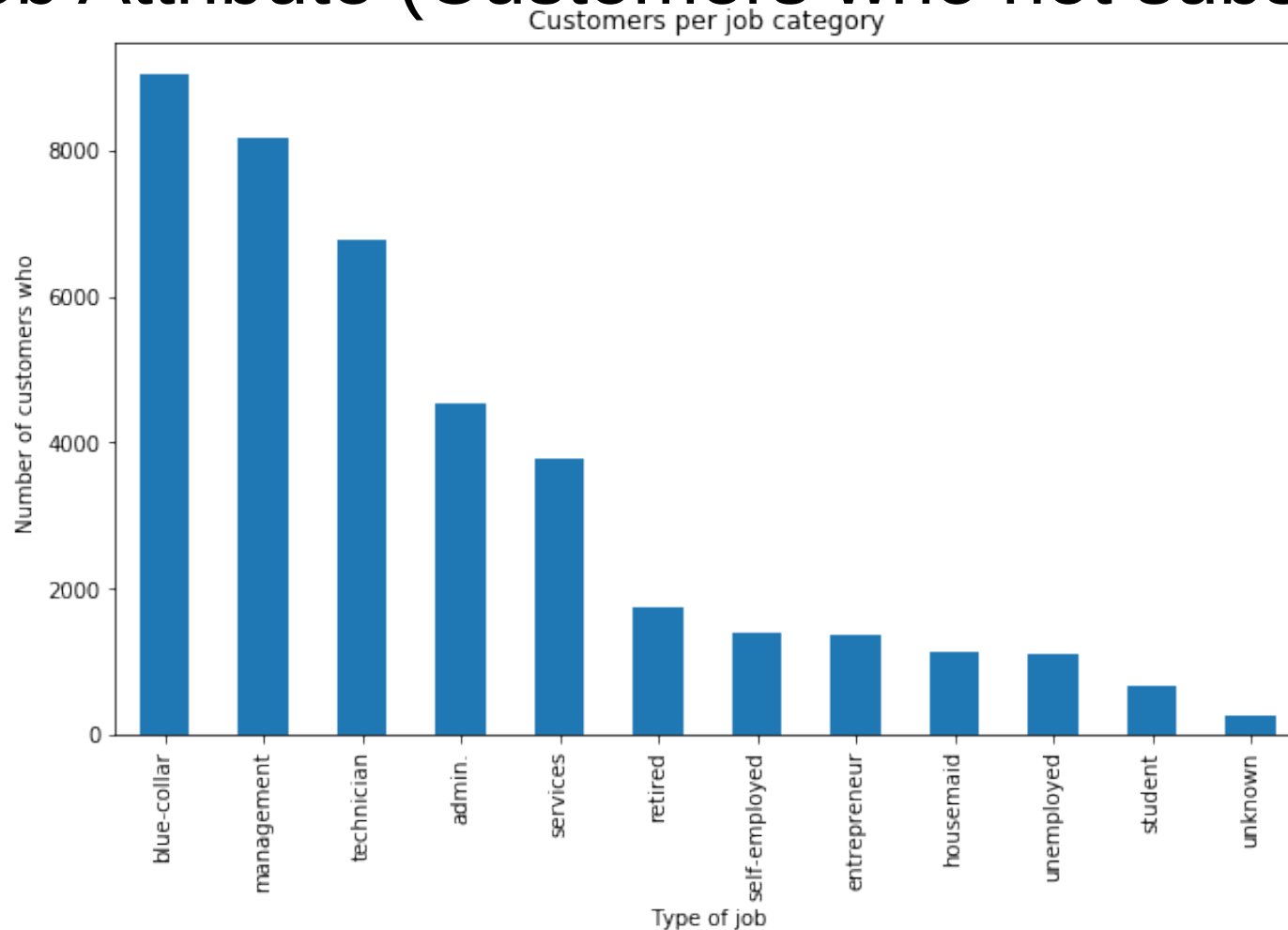
Data Exploration

- Marital status distribution

Pie chart of marital status

Pre-processing

- Dataset is very well formated

- It has no missing values

- Data explored for outliers

- Whole dataset considered for analysis resulting in very good accuracy

Model Building

- 3 classifiers are tested

- Logistic regression gave best results

- Decision tree can also be considered eventhough it gave less accuracy as it gives if-then rules

- If-then rules are easy to interpret and implement

-

Decision tree
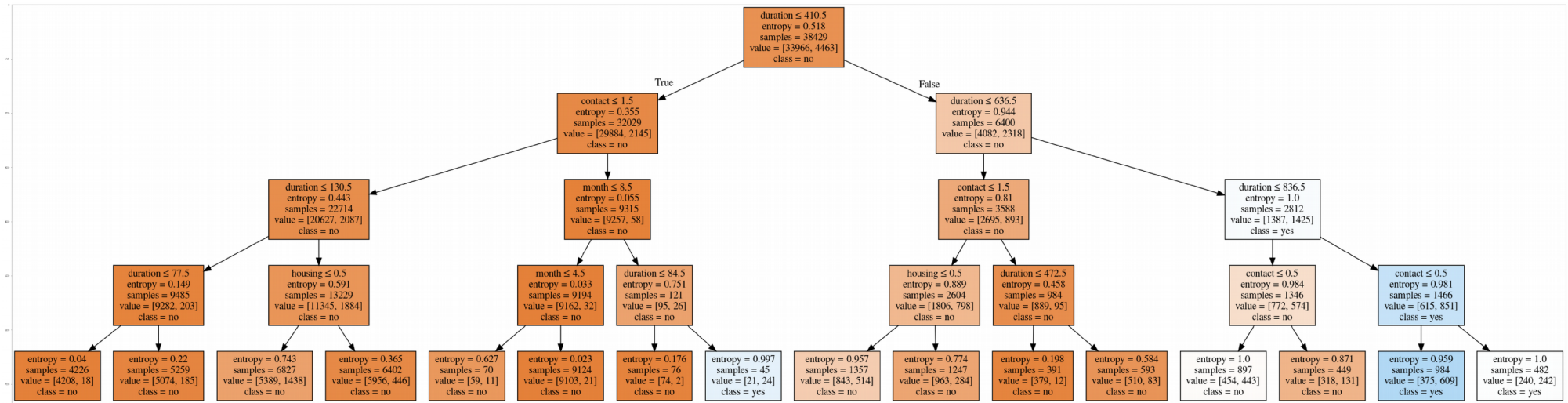
Results and comparision

Algorithm	Accuracy	Jaccard similarity	F1 score	Log loss
Decision Tree	0.89	0.89	0.86	NA
Logistic Regression	1	1	1	0.01
KNN	0.99	0.99	1	NA

Thank you