Sentiment of

Blood Donation - Pipeline

Importing the Libraries

↓

Reading the Data from Dt. Set

↓

Data Exploration

↓

Feature Engineering

↓

Feature Transformation

↓

Models

↓

Analysis, Comparison, Results.

Dataset Description

Test					Train					
1st column ?	no. since last	no. of donat	Vol.	no. since first	1st column ?	no. since last	no. of don	Vol	no. since first	did donate in March 2007

syntaxes

loc
/
location
name
locator

iloc
/
location
number
locator.

— has already sorted list

faster than loc

- Use both to target rows

iloc [:, :3]

all the columns until the third one.

- loc is faster for selecting specific ~~rows~~ columns
-
iloc faster for rows.

---

loc [,]

rows columns

loc [0, :]

all columns

row zero

loc [[0,1,2], :]

columns all.

rows 0,1,2

loc[0:2, :]

↓ slicing rows 0 through 2

↓ loc is inclusive on both sides 0, 1, 2 all.

→ all columns.

* If you leave the column thing blank, pandas just assumes it.

↓ Bad Programming Practise.

iloc → filtering rows and selecting columns.

↳ with respect to the integer position.

iloc[:, [0, 3]]

↓ colums in bet^n position 0 & 3

- Columns are selected index wise
  - index is integer wise.

iloc [: , 0:4]

↓

exclusive of 4(index)

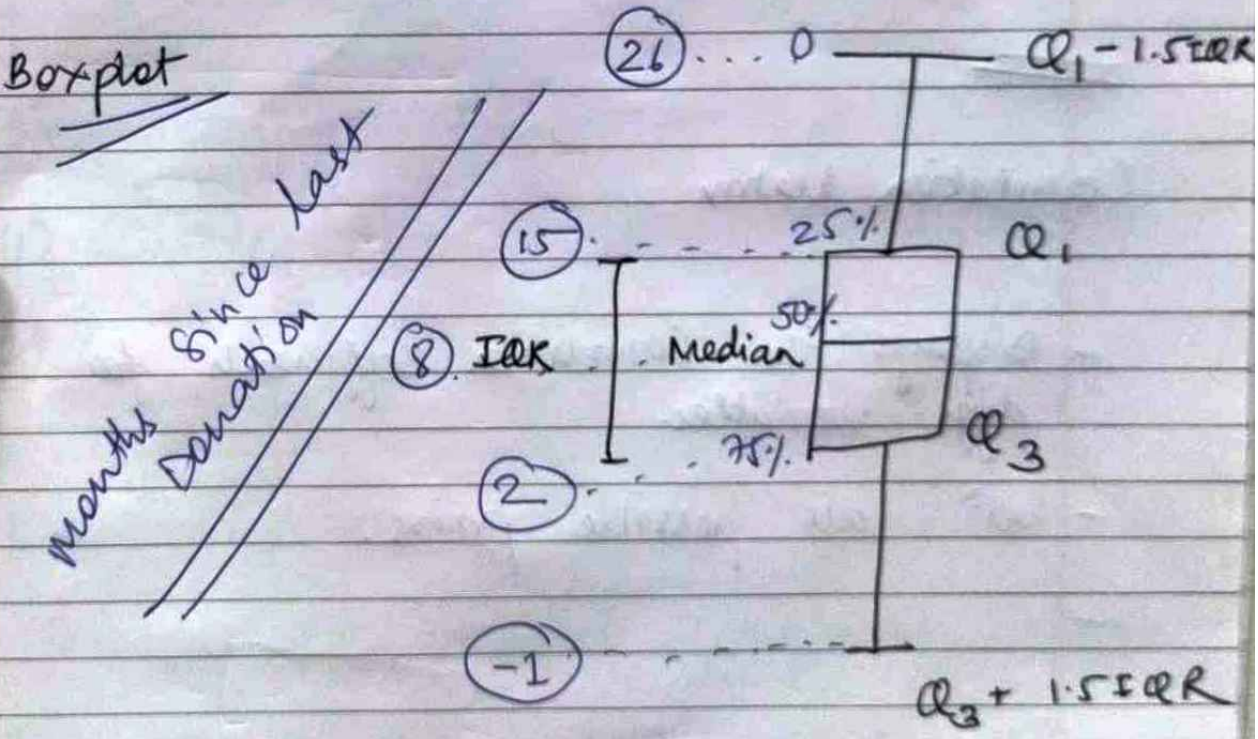
inclusive of 0.

loc → labels → ranges → inclusive both sides

iloc → integer positions → inclusive: start
exclusive: end.

~~~~~

Boxplot



The Circled values depict the Months since last donation.
~~~~~

Outliers lie at 32, 39, 70, 71.

Minimum → the first quartile Q1

Median

Third Quartile, Q3

Maximum

- The Boxplot helped us visualise where and at which position how the dataset is distributed.

Correlation Matrix:

- Displays the Correlation coefficients for diff variables.
- betⁿ all possible pairs.
-

classmate
Date
Page

* We use the correlation map & the Boxplot to figure out the relations among the variables in our dataset.

SKLEARN :
- RandomForest Classifier
- LogisticRegression
- DecisionTree Classifier
- MLP Classifier
- SVC

We compare them through:

1) accuracy_score
2) roc_auc_score.

---

1) accuracy_score:

- This is the Accuracy Classification score.
- It computes the subset accuracy.

- The set of labels predicted for a sample must exactly match the corresponding set of labels in y_true.

② roc_auc_score:

- Compute Area under the Receiver Operating Characteristic Curve.

ROC AUC — from pred^n scores.