ASSIGNMENT-6

AIM: Perform different data cleaning operations using R/ python

PROBLEM STATEMENT: Perform following operations using R/ python on the Quality & heart disease dataset

a. Data cleaning
b. Data integration
c. Data transformation
d. Error correction
e. Data model building

THEORY:

1. DATA CLEANING:

It is the process of transforming raw data into consistent data that can be analysed. It is used for improving content of statistical statements.

R has profound functions to perform data cleaning.

Ex:

```
data ← read.csv("ABC.csv", na=" ")
```

Now analyse the dataset & remove unwanted content.

Trimming whitespaces

```
data $ distance ← str_trim(data $ Dist_Taxi)
```

Page No.:
Date:

2. DATA INTEGRATION:
Data from dataset can be encapsulated into a datdframe.

3. DATA TRANSFORMATION:
A no. of reasons can be attributed to when a predictive model compiles.
a. Inadequate data preprocessing
b. Inadequate model validation

4. DATA MODEL BUILDING:
We can build models in the form of linear regression & much more

ex: linearMod ← lm (dataset $ NO(G+), data = datas

print (linearMod)
summary (linearMod)

CONCLUSION:
In this assignment, we learnt basic data cleaning operations on heart disease dataset