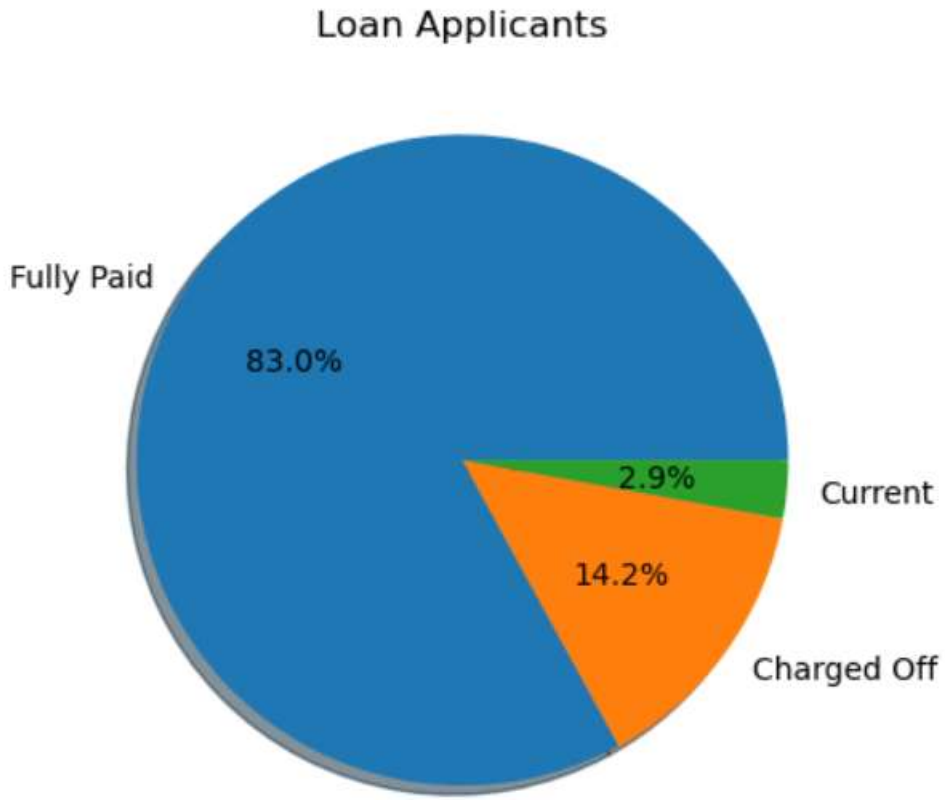
Lending Club Case Study

The aim is to identify patterns which indicate if a loan applicant is likely to default.

Univariate analysis is a form of analysis that only involves a single variable.

The “loan_status” is the variable analysed here and shown below.



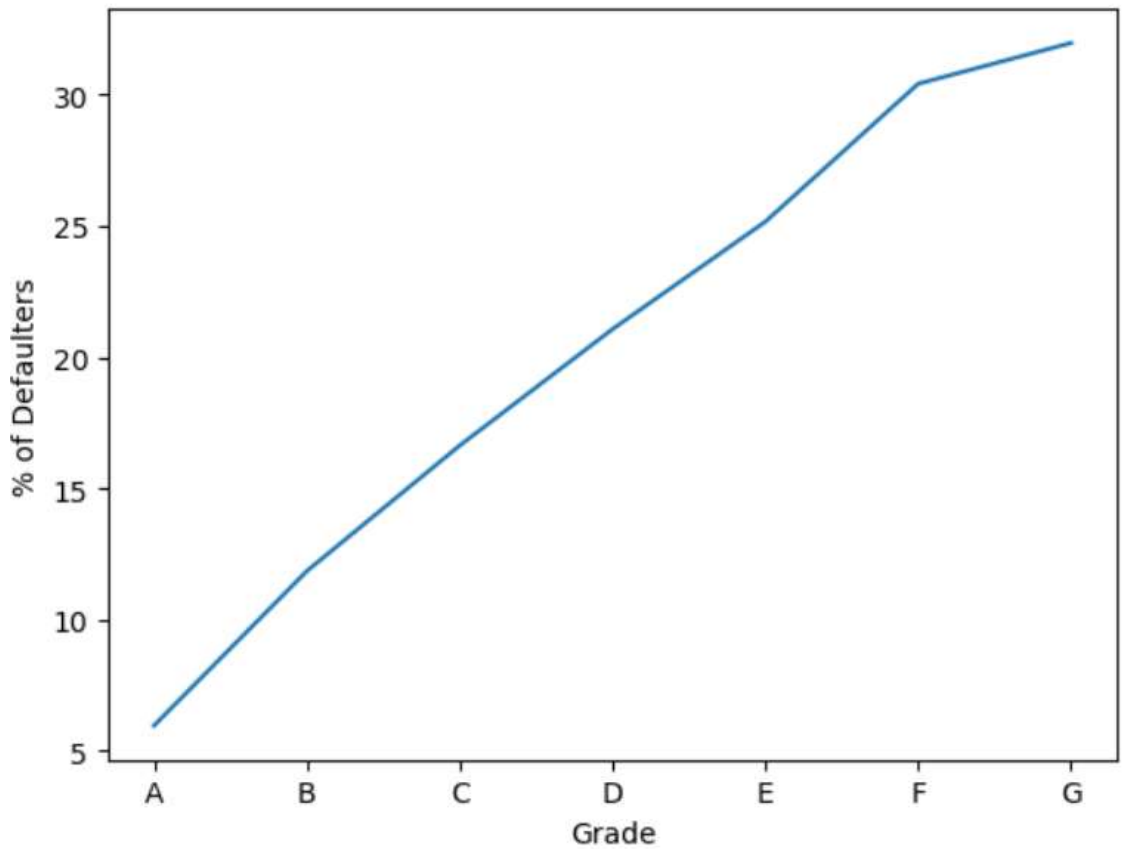
We find that 14.2% out of total loan applicants have difficulties in repaying loans.

A **derived metric** is a calculation based on the data included in the report definition

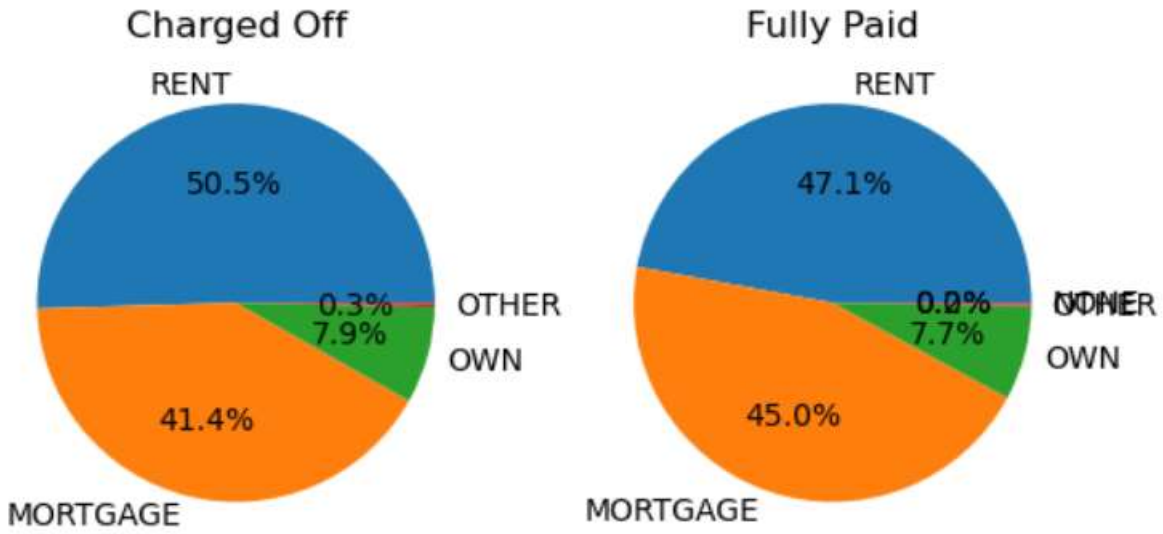
A column “Defaulter” has been added which is 1 if the loan is charged off and 0 if it is fully paid.

Bivariate analysis involves the analysis of two variables (often denoted as X, Y), for the purpose of determining the empirical relationship between them.

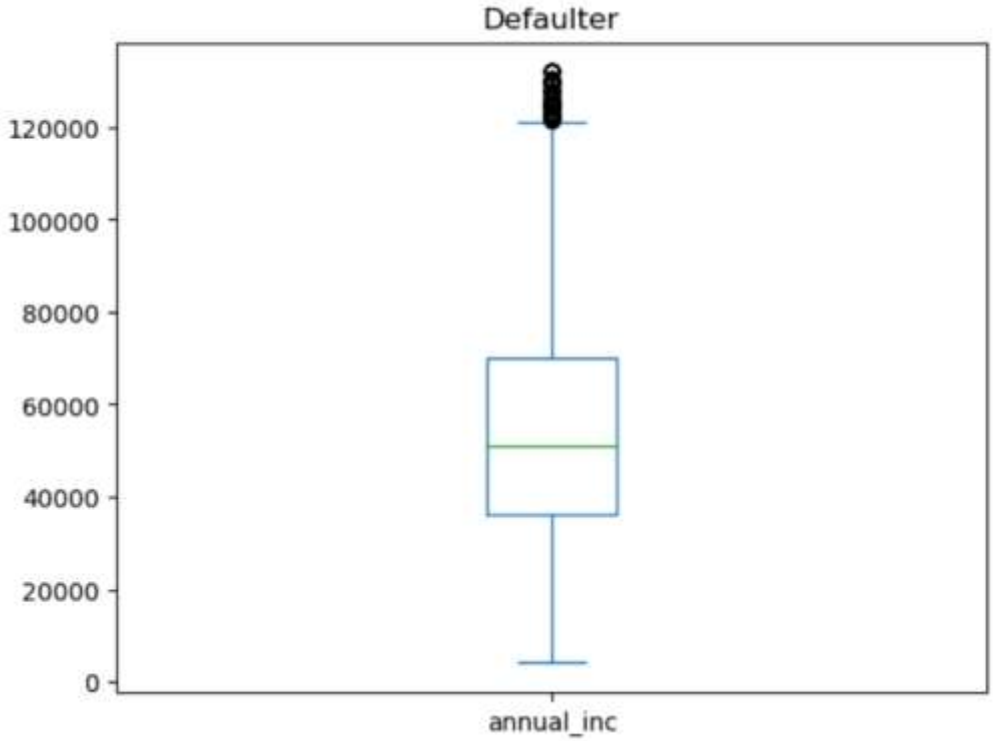
Relationship between the column “grade” and “Defaulter” is depicted below

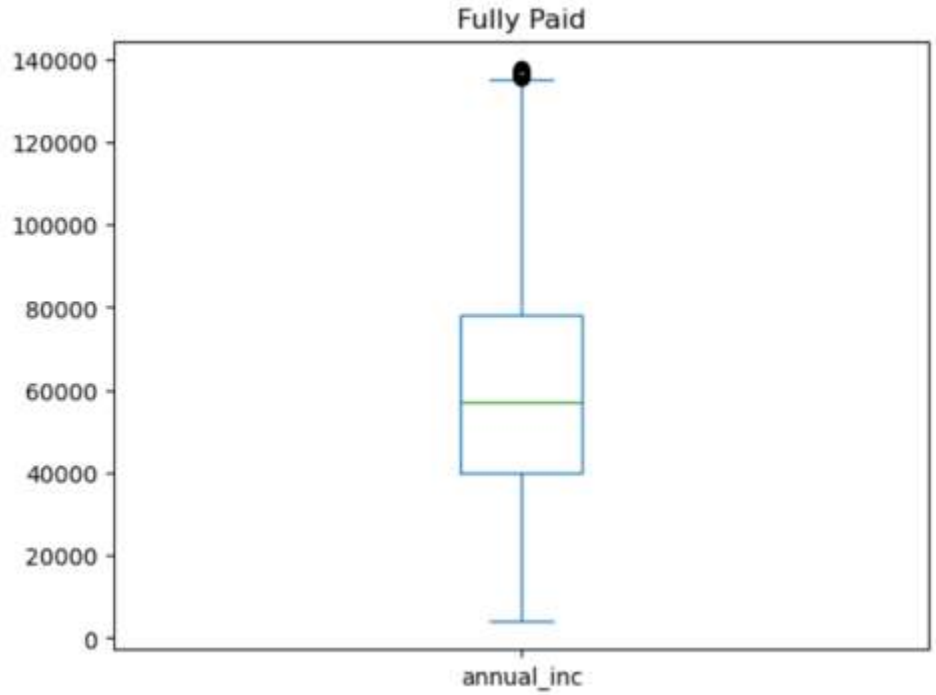

30
25
20
15
10
5
% of Defaulters
A
B
C
D
E
F
G
Grade

Home ownership



Income distribution





Conclusions

- The distribution of income amongst charged off customers and fully paid is very similar.
- 14.2% out of total loan applicants have difficulties in repaying loans.
- The percentage of Defaulters is directly related to the grade of the applicant.
- The chance of a loan applicant defaulting on his loan payment is not dependent on the home ownership.