**Lab 3 – Data Quality Report**

**(i) Missing Values**

HEALTH TYPE, HEALTH DEPS ADULTS, HEALTH DEPS KIDS all contain missing values for customers that do not have a health insurance policy with the company as indicated in the HEALTH INS column.

MOTOR VALUE is blank where the customer does not have a motor insurance policy with the company as indicated in the MOTOR INS field.

OCC is blank for some customers, they may be unemployed. If they are then “unemployed“ should be entered instead of leaving the field blank.

**(ii) Irregular Cardinality**

The customer with id of 5200 has their gender set as 47 and their age set as female. These were clearly entered in error and will have to be remedied before the data is used by a machine learning algorithm.

**(iii) Outliers**

The customer with id of 2749 has a motor value far higher than any other customer.

**(iv) Feature Distribution**

There is a lack of younger people in the ABT. Only 4 out of 21 are under 30.

There is a lack of more expensive motor vehicles. Only 1 has a value greater than 45k.

The rest of the values have a good distribution.