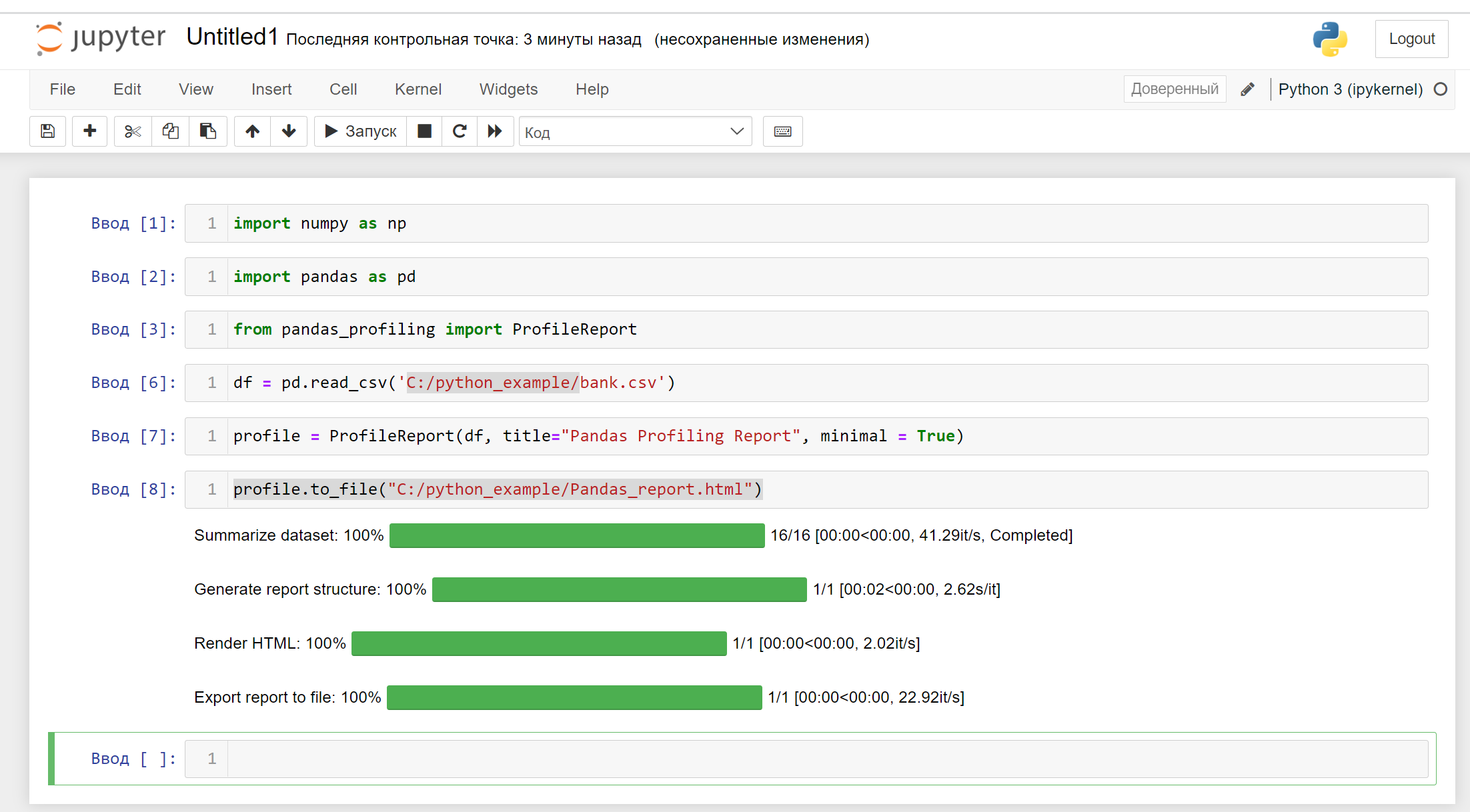
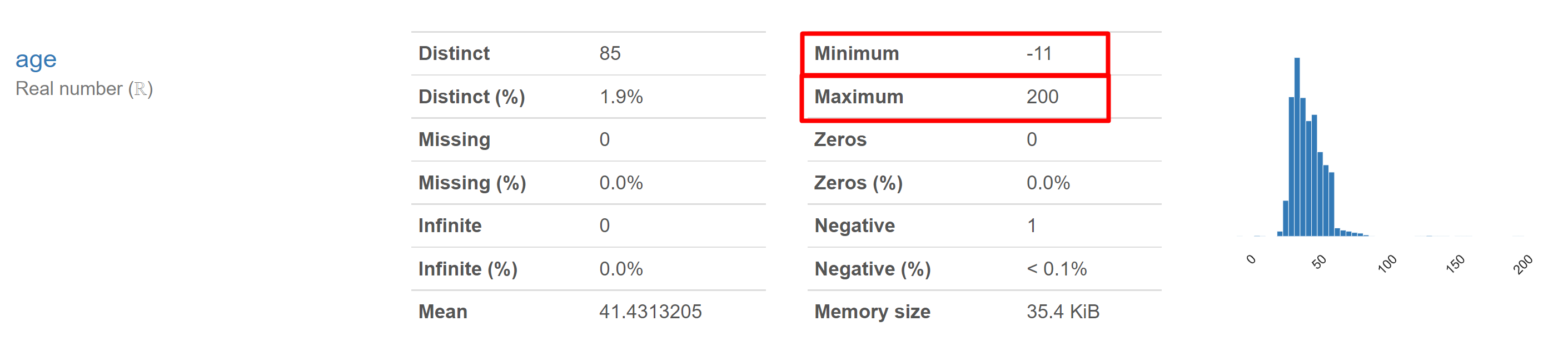
|  |
| --- |
| **Topic\_01. Data Profiling** |
|  |

# PYTHON DATA PROFILING



There are such anomalies as:

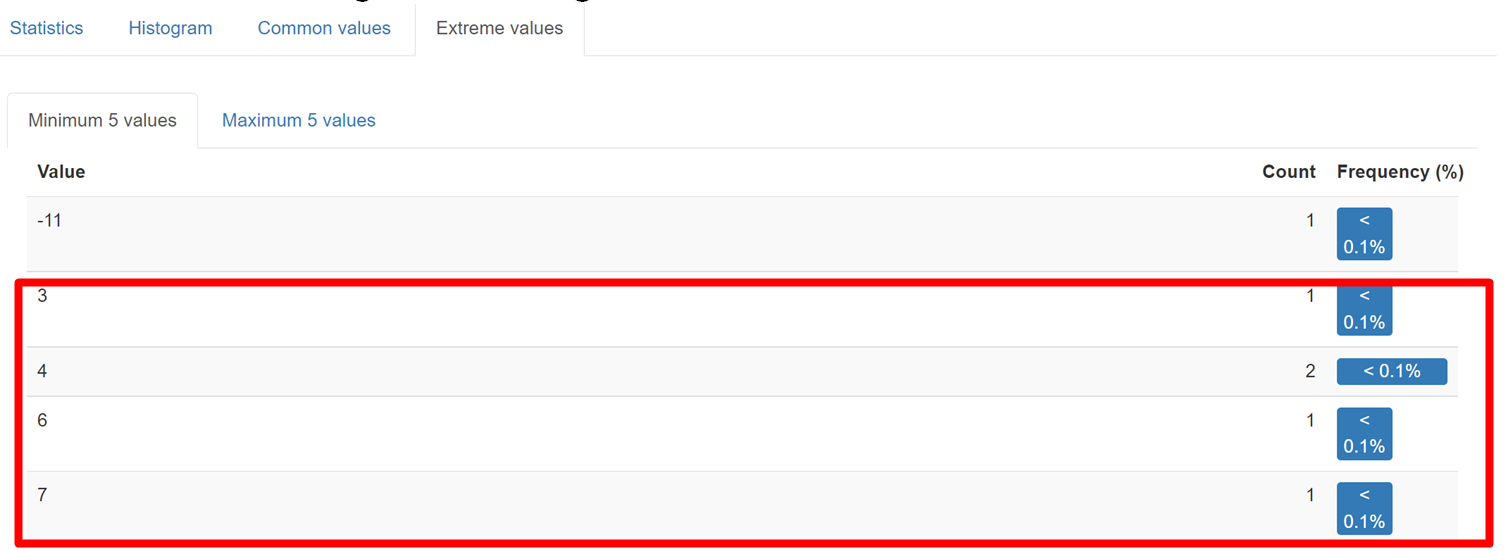
* Negative value of age (-11):



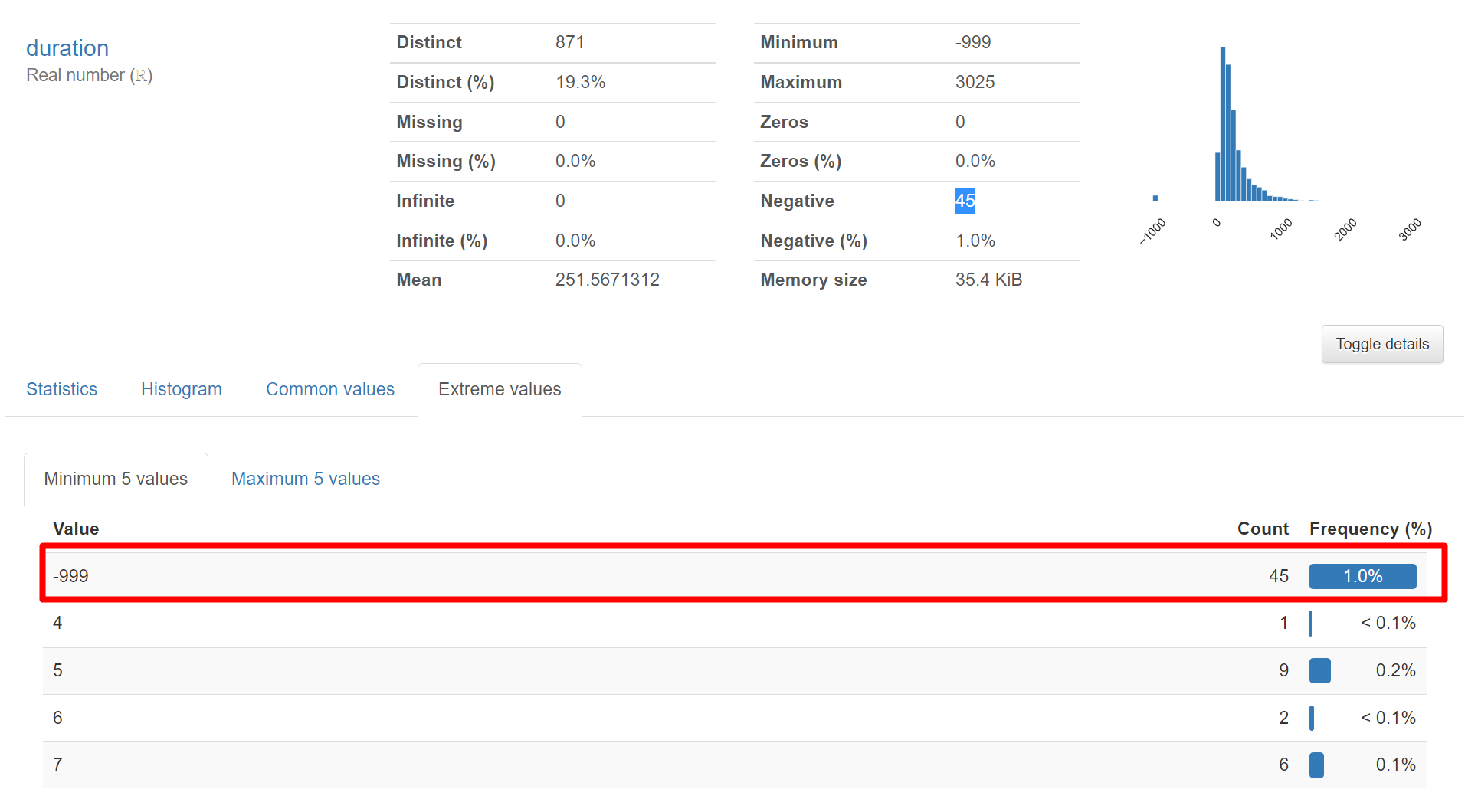
* Maximum age (>100): there are 13 rows with ages more than 100:



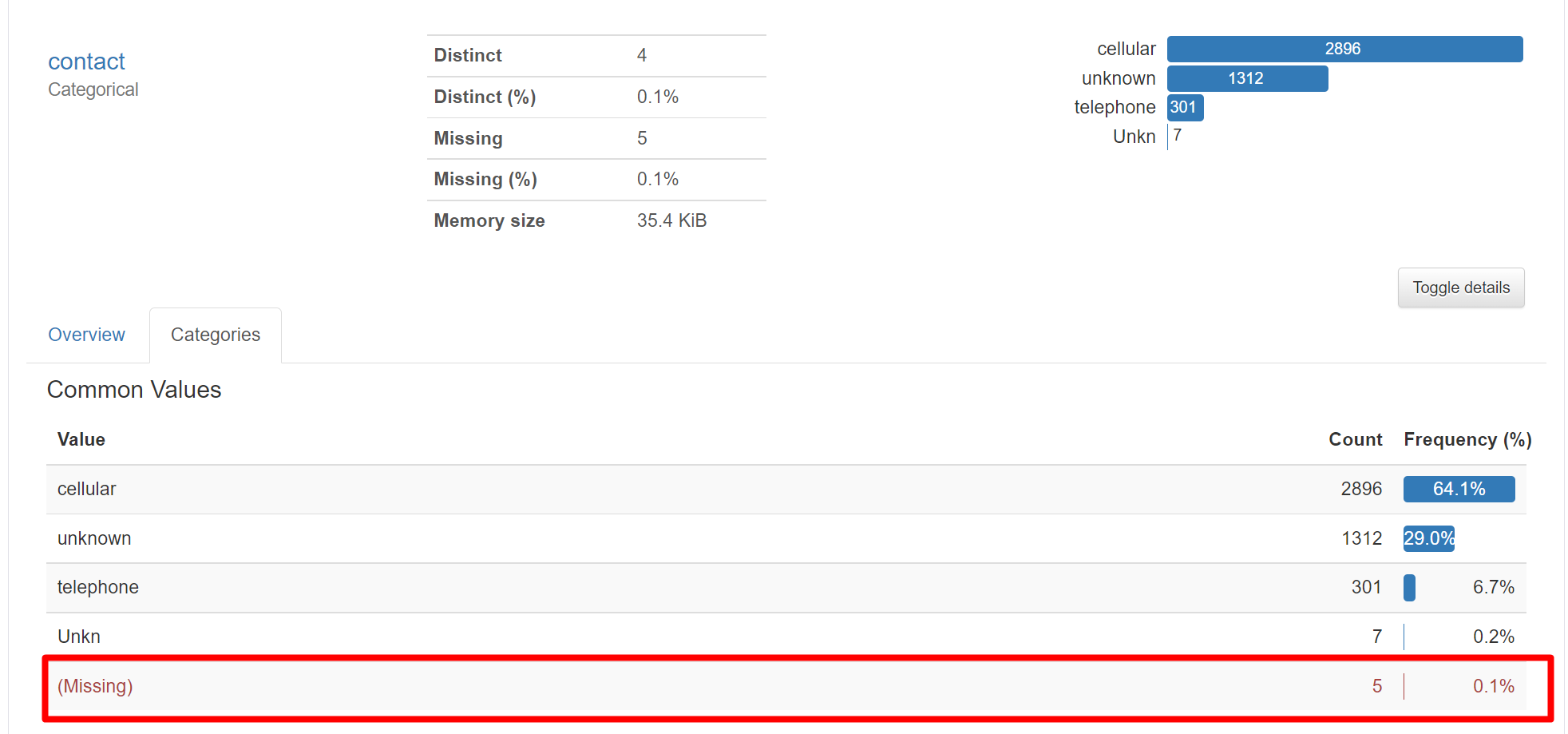
* Too small ages of banking clients:



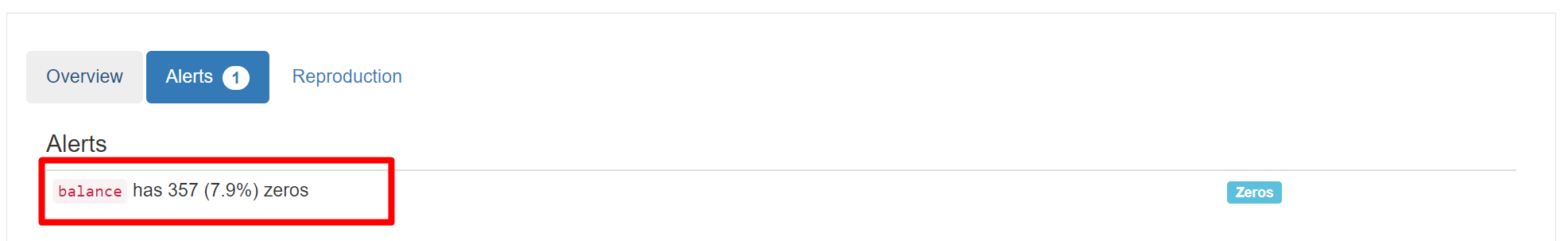
- Negative duration. It confuses because it's not obvious if it's default value (instead of empty field) or real duration. Moreover it can affect the summary duration.



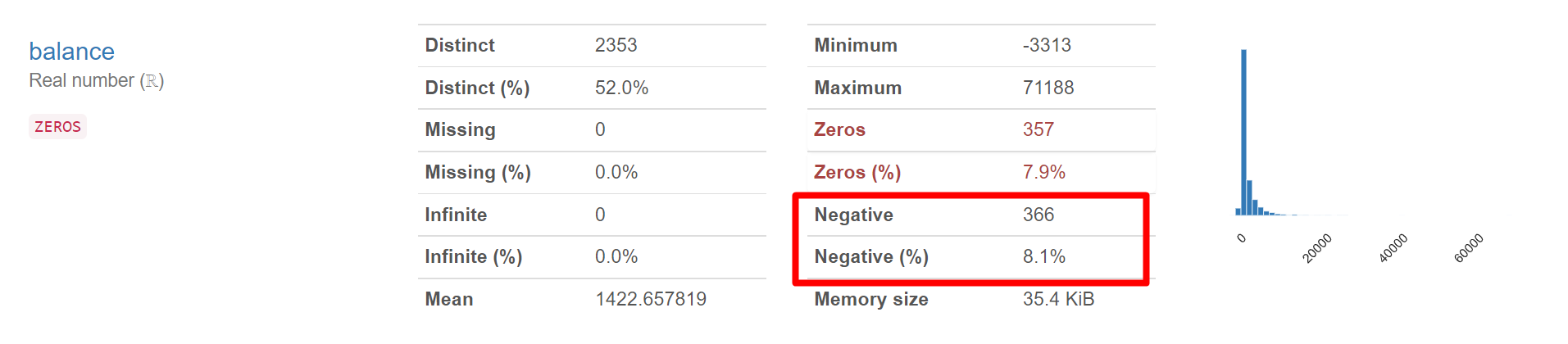
* Nulls can affect the result of query so it’s better try to avoid nulls:



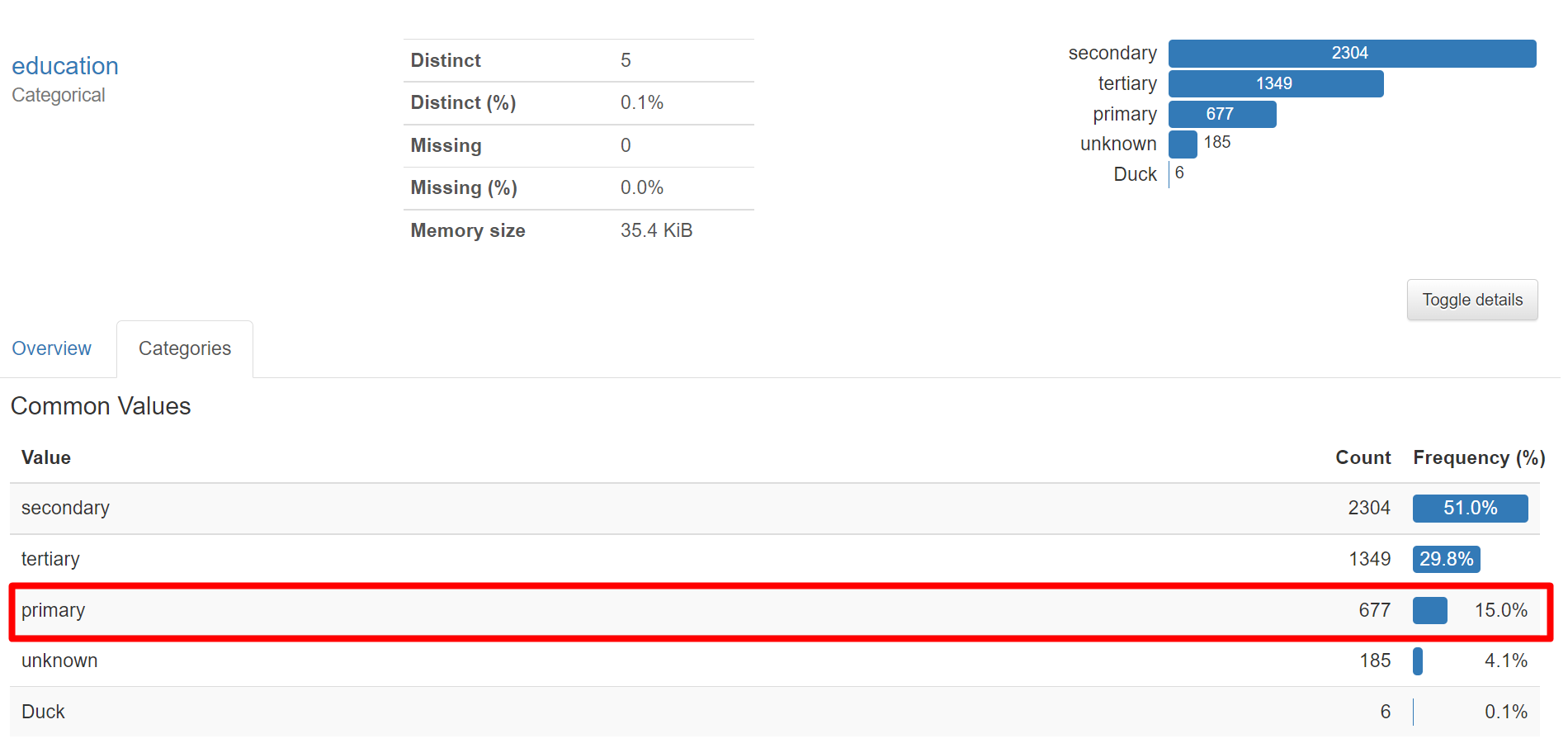
* Balance has 357 (7.9%) zeros. 7.9% is a rather big percent:



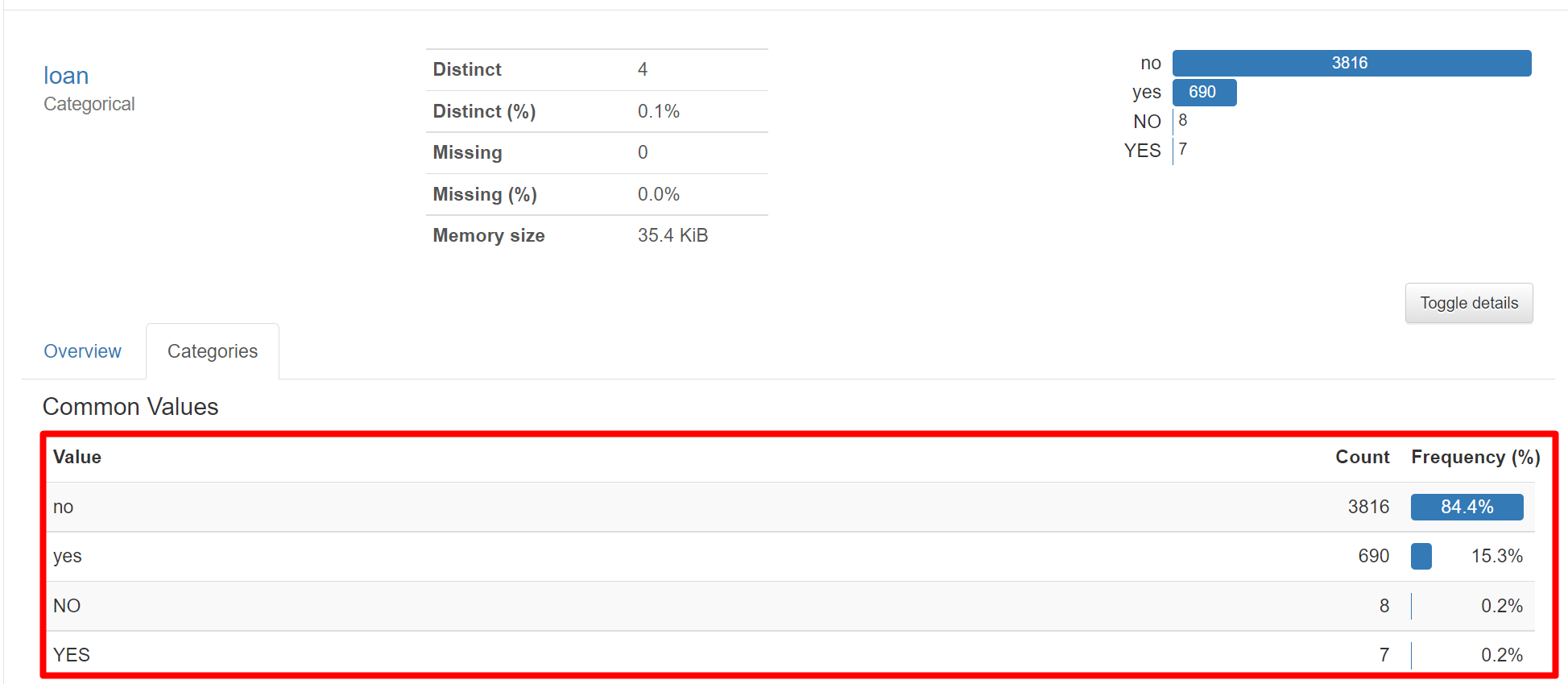
* Balance has 366 negative values:



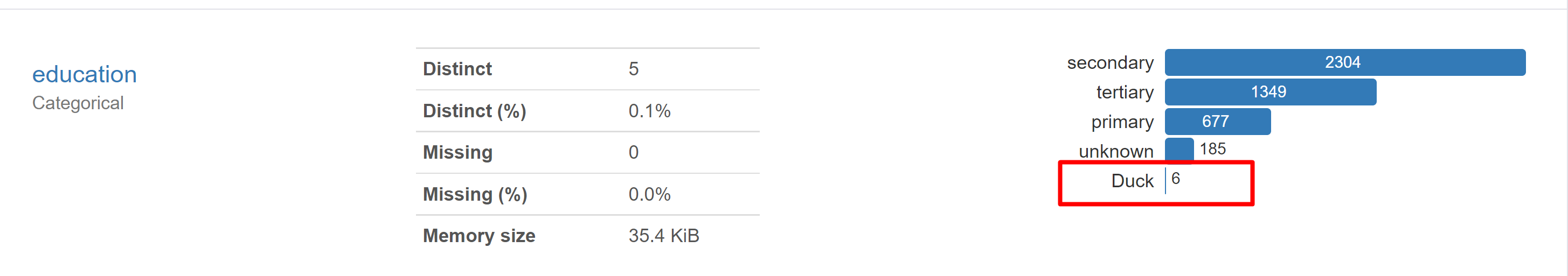
* About 15% of banking clients have **primary** education and at the same time, average value of ages (mean) is 41.4313205 and median = 39:



* There are some fields with duplicates with different cases (upper or lower):



* Strange names in some fields:



* Unobvious name of field and its value:

