Data profiling is the process of analyzing data and producing an output with statistical summaries. The size of data rapidly increases and it is more difficult to process all data in a reasonable time. All data can not be stored in RAM memory, so it is not possible to run exact single-pass algorithms without using slower computer storage. The diploma thesis focuses on the implementation, comparison, and selection of suitable algorithms for data profiling of large input data. Usage of approximate algorithms brings a possibility to limit memory for computation, do the whole process in RAM memory and the duration of data profiling should be reduced. The tool can compute frequency analysis, cardinality, quantiles, histograms, and other single-column statistics in a short time with a relative error lower than one percent.