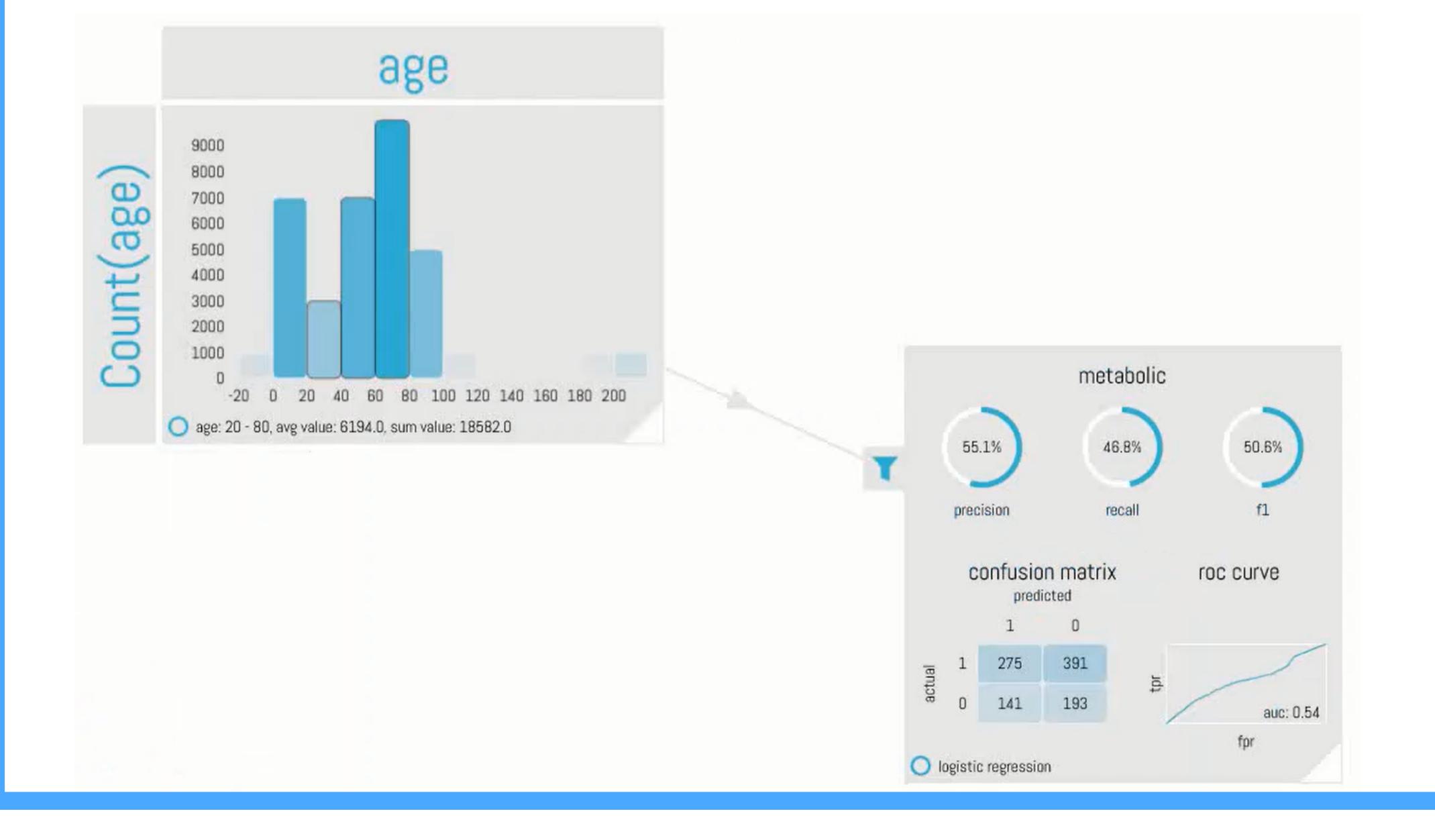
## A Tech Stack for interactive Big Data exploration

Data Sciene does not only include setting up data pipelines and training ML models with them. Often data must be explored and insights must be generated. The long execution time on Big Data can slow down such a process, which is why Prof. Carsten Binnig from the TU Darmstadt is working on new technologies for interactive query processing. This poster is intended to give an insight into which techniques they use and what makes them special.



With Vizdom, data analysts can visually assemble complex workflows of statistics operators on an interactive whiteboard with pen and touch. The goal is to make the analysis of large data accessible to everyone, not just to domain experts with experience in machine learning and statistical reasoning. This interactivity places special demands on the database to ensure that queries deliver results quickly.



O VIZDOM (Visual Analysis) Natural language interfaces allow a natural and concise way of querying data, but understanding these queries is really difficult. DBPal converts natural language queries into parsable SQL queries. It is a deep learning based system that is not based on any rules. Therefore it must need to be trained on a specific data set. This training data set can actually be generated using templates in combination with the data. These SQL/NL pairs are then



used to generate new pairs by removing phrases or translating phrases back and forth in NL. Benchmarks have shown that these combined optimization techniques can significantly outperform other systems.

Both interfaces described above are in direct contact with humans. Research has shown that they become bored and unproductive if the calculation of results takes longer than 500ms. When

working with large data this limit can easily be reached, which is why IDEA was developed. Already in the 90s the idea of using only partial data to accumulate the final result came up. Unfortunately this so called Approximate Query Processing (AQP) is not supported by any modern database. However, IDEA can work as middleware before existing databases and implement AQP. To do so, IDEA



