Metacrate

Organize and Analyze Millions of Data Profiles

Data science is mostly not about doing data science…

Data scientists spend about **80% of their time on data preparation** (while perceiving this as the least enjoyable activity) [1]. **Collecting**, **organizing**, and **cleaning** are particularly time-consuming. Metacrate aims to facilitate these tasks in **three steps**…

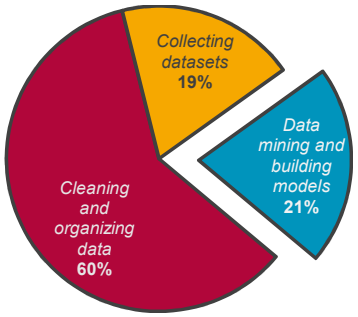
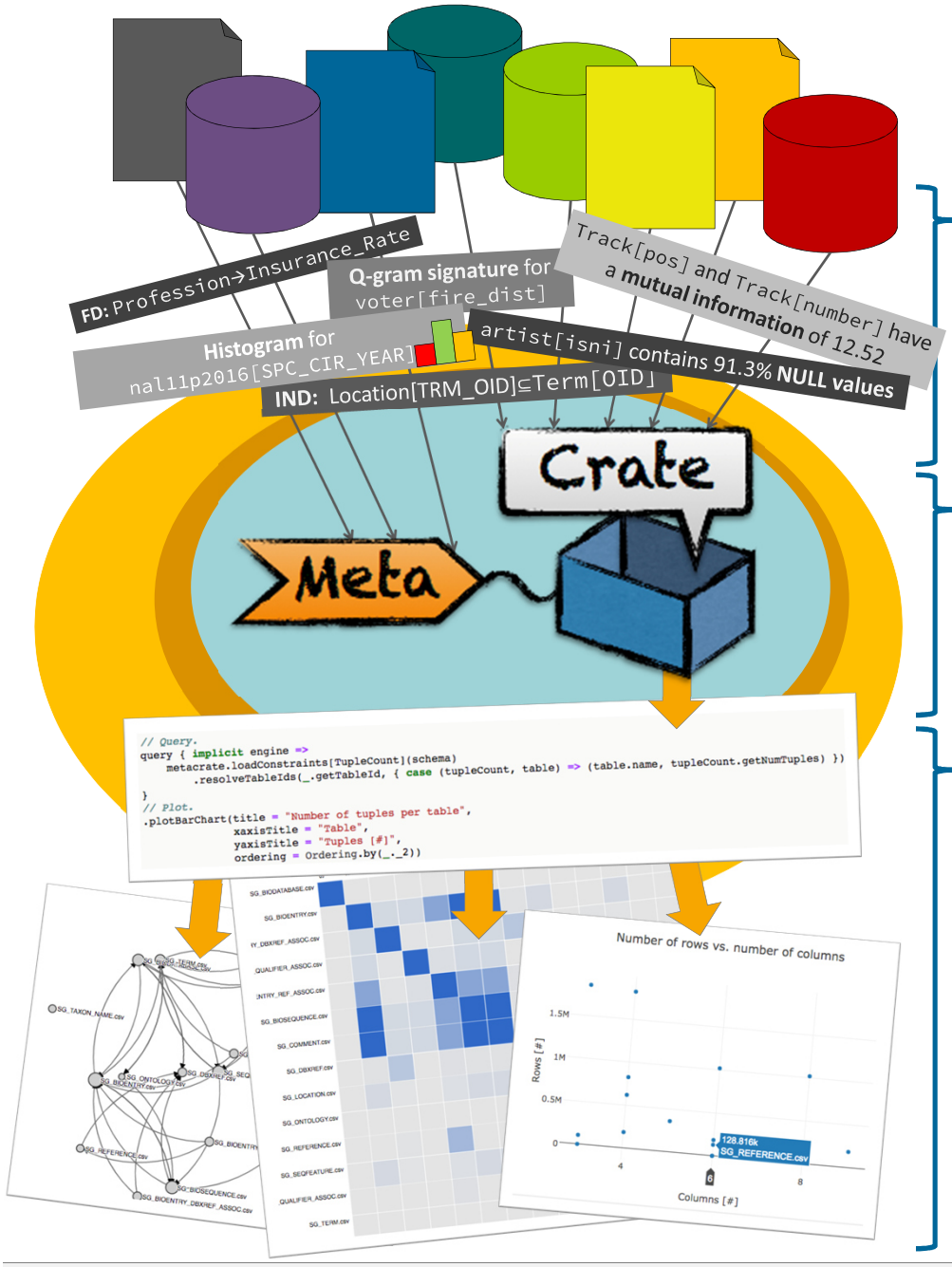


Chart: *Average time spent on tasks throughout data science projects [1].*



Profile.

Data profiling tools, such as Metanome [2], can **extract** a wide array of data profiles; particularly, data **dependencies** and data **summaries**.

Organize.

Metacrate's flexible data model **organizes** the various data profile types. It supports several storage backends, thereby providing **scalability**.

Analyze.

Metacrate allows to query, integrate, and analyze data profiles. Besides a set of common **visualizations**, Metacrate provides a metadata analytics **library**. This combination facilitates data **anamnesis**, data **discovery**, and data **cleaning**.

[1] Gil Press. Cleaning Big Data: Most time-consuming, least enjoyable, data scientists say. *Forbes.com*, 2016.

[2] Thorsten Papenbrock et al. Data Profiling with Metanome. *PVLDB*, 8(12):1860-1863, 2015.

Sebastian Kruse[1], David Hahn[2], Marius Walter[2], Felix Naumann[1]

Information Systems Group
Hasso Plattner Institute, University of Potsdam, Germany
[1]*firstname.lastname@hpi.de* [2]*firstname.lastname@student.hpi.de*

Crate Meta

Metanome

https://hpi.de/naumann/

HPI Hasso Plattner Institut