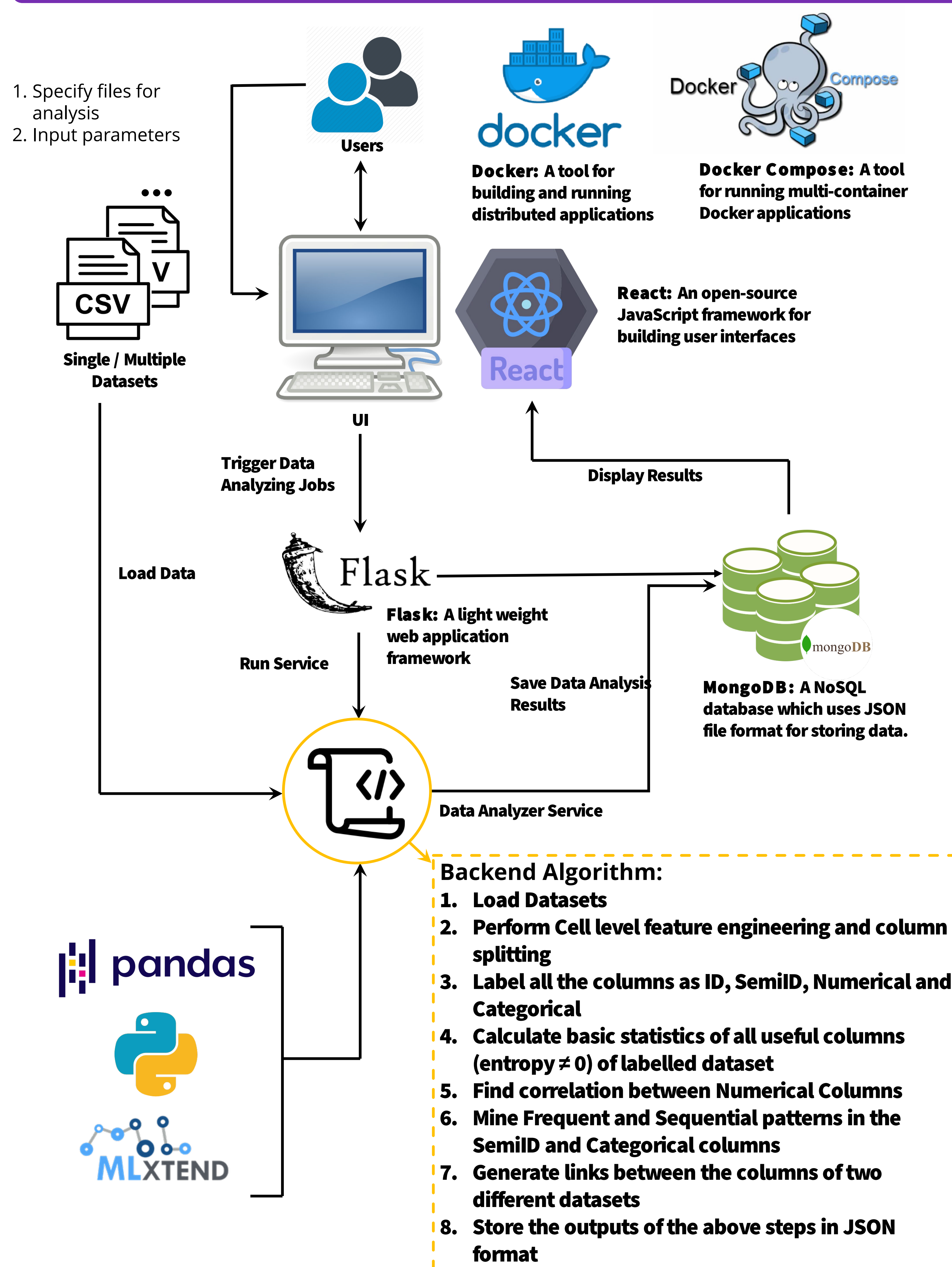


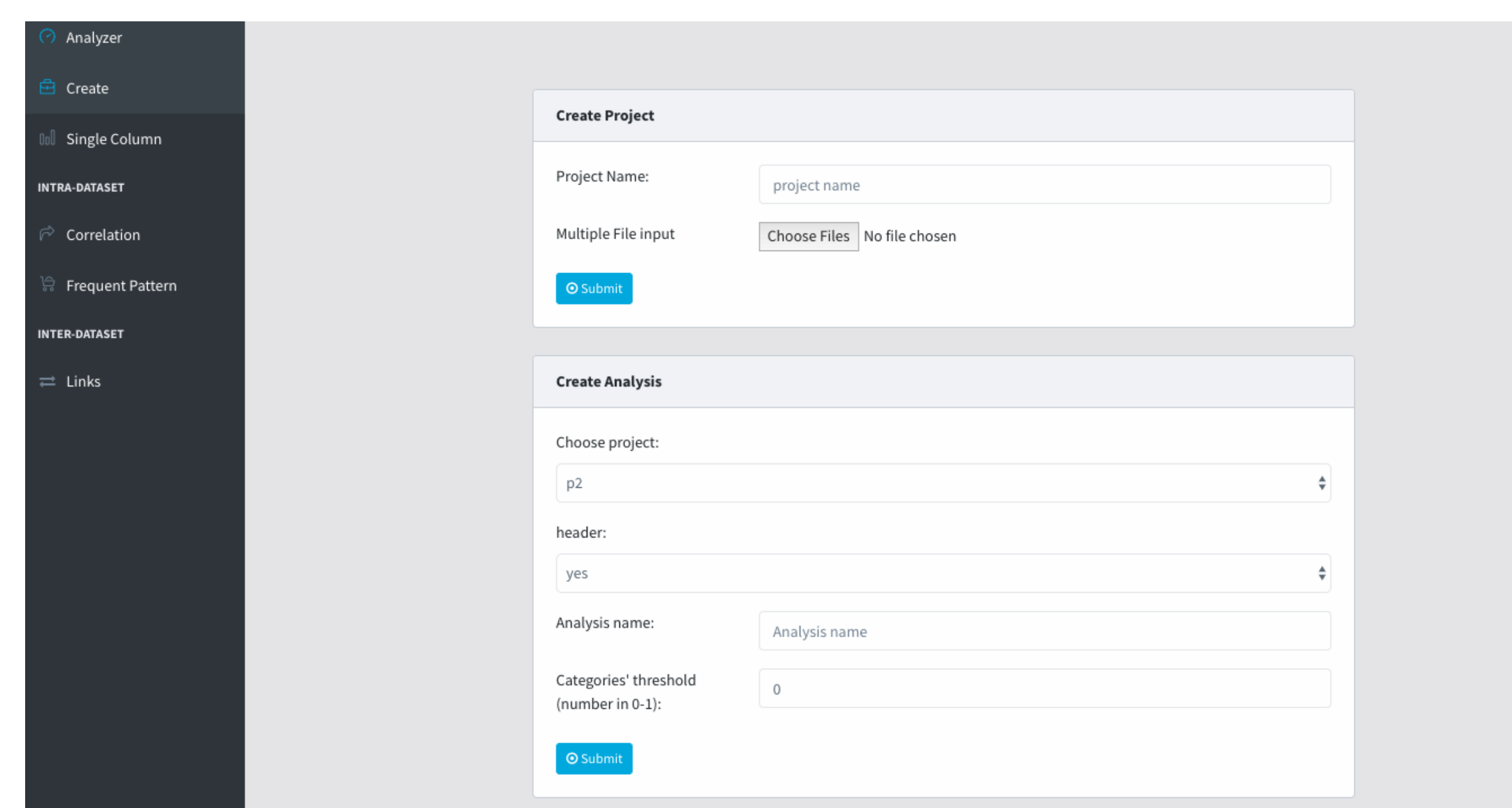
Motivation

Analyzing data from telecom operators becomes a time-consuming process for every project and majority of the data analysis work requires R&D team working manually to dig out useful information from given datasets. Thus, to facilitate data analysis process, we have created a tool which automatically provides basic knowledge of the data, such as statistics, correlations, patterns and relations, which directs data scientists for deeper analysis.

Technology and Workflow

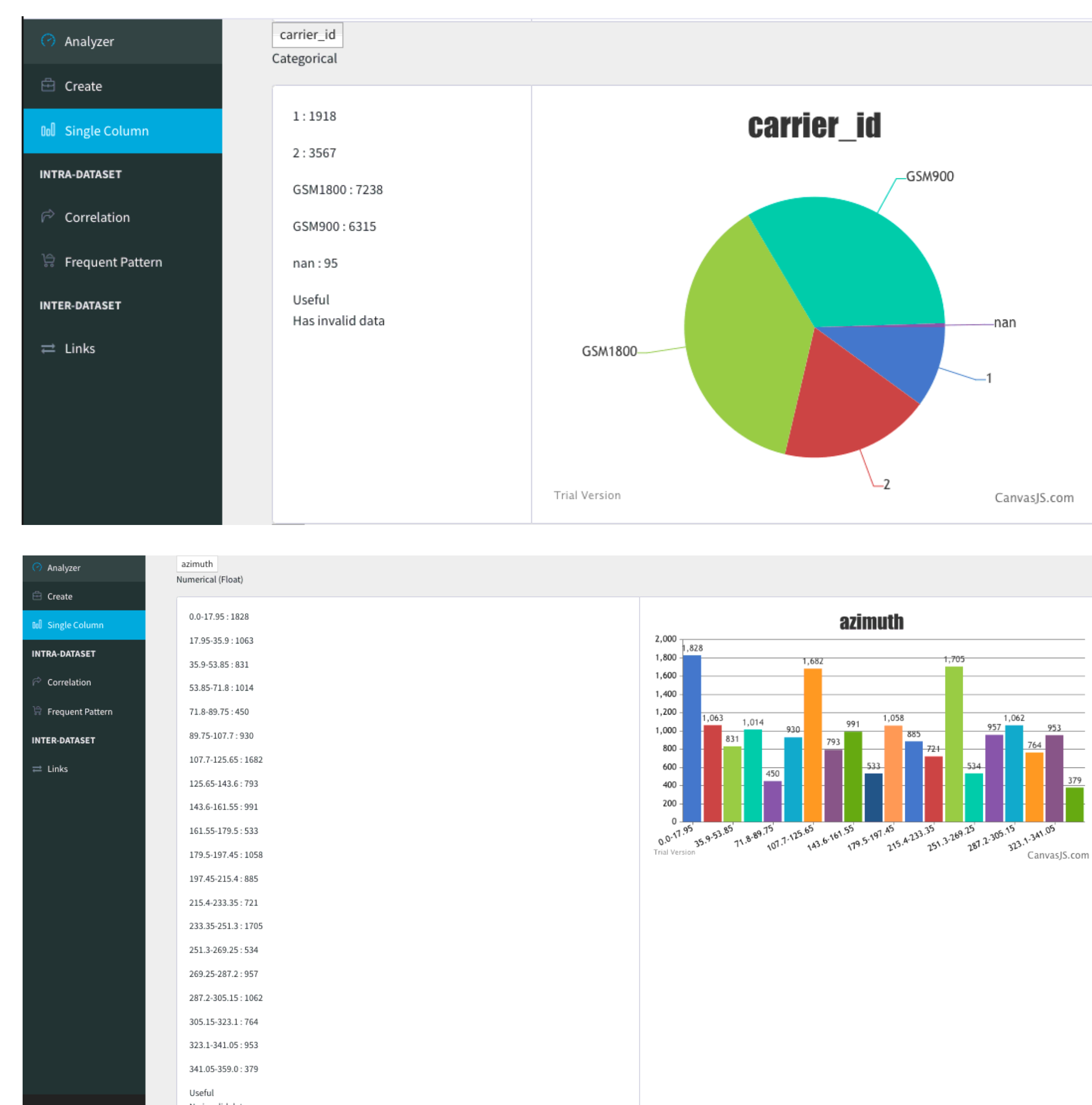


User Interface



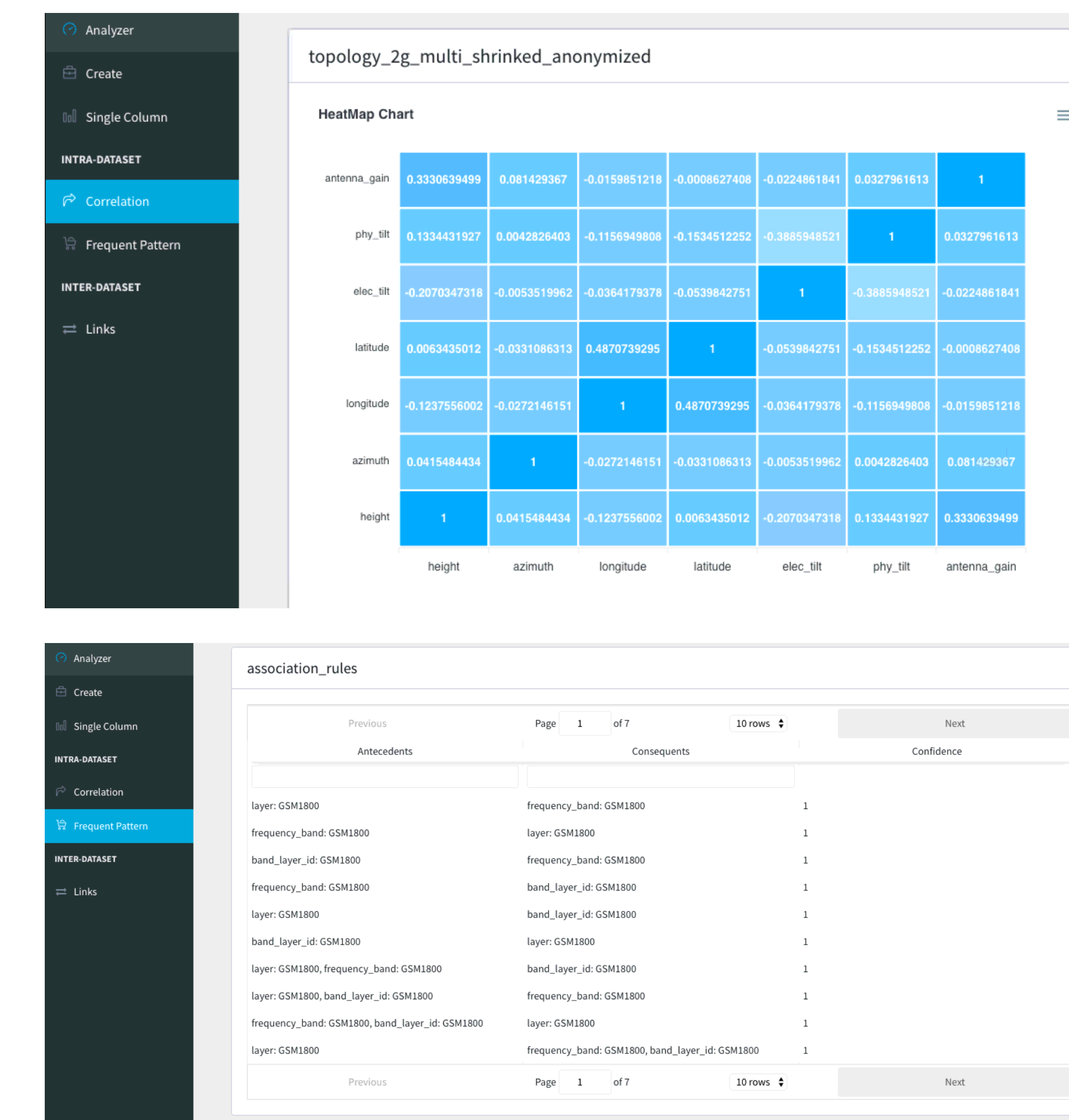
- User can create a new project / selects an existing project
- Inputs the parameters required for analyzing the data and triggers the analysis job

Single Column Analysis



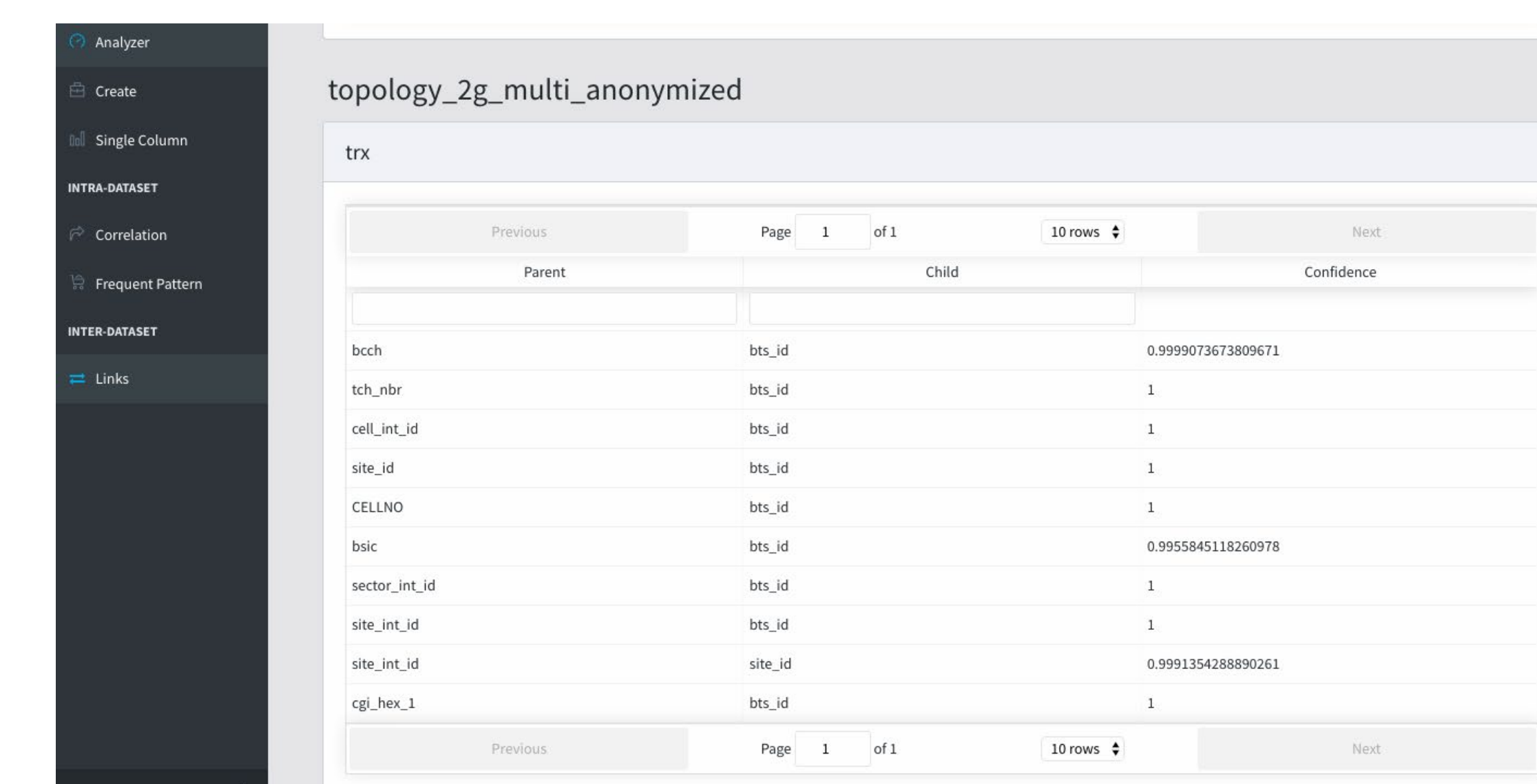
- This section displays the results for the analysis of each individual column according to the labels generated by the backend
- Pie charts are displayed for Categorical, ID and Semi-ID data, Distribution charts are displayed for Numerical Data

Intra-Dataset Analysis



- This section displays correlation between numerical columns & frequent patterns discovered
- Heat Map is displayed for the correlation and the mined patterns are shown in a tabular format

Inter-Dataset Analysis



- This section displays links generated between columns of two or more datasets
- These linked columns across datasets are calculated based on their Jaccard similarity.

Future Work

- More features like user registration / login, choosing a specific analysis result or all the results generated on any day will be added
- Automatically generate a report of the analysis in Natural language
- Create graph of semi-ID fields and apply graph algorithms (e.g., page-rank) to extract useful information