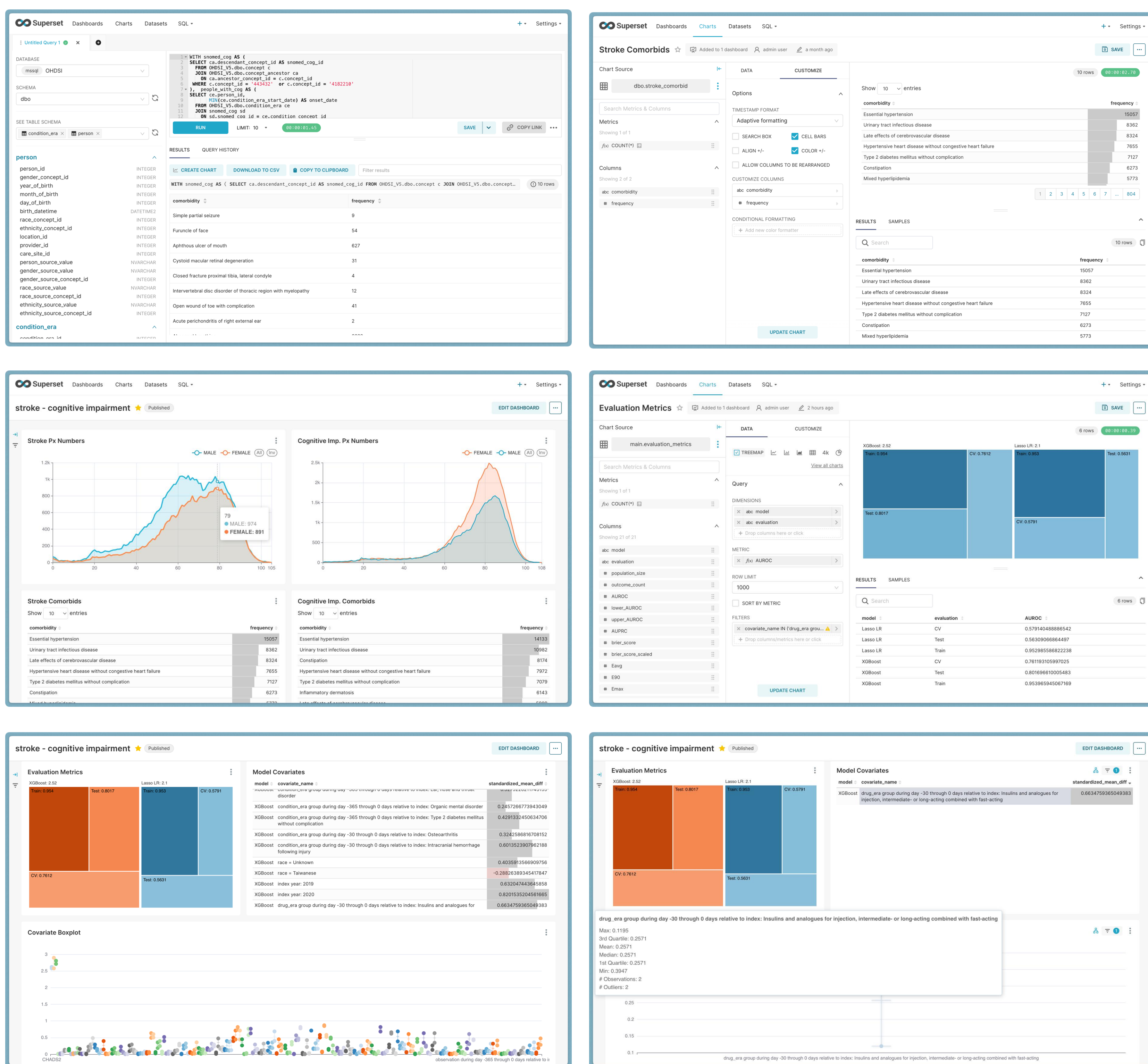


Apache Superset for rapid exploration tools for OHDSI's OMOP CDM

Exploring Stroke and Cognitive Impairment in the TMU-CRD OMOP CDM Dataset Using Apache Superset

Background: This study utilizes Apache Superset, an open-source BI tool as an alternative to Shiny App, to explore stroke and cognitive impairment data within the OMOP OHDSI framework, specifically leveraging the TMU Clinical Research Database (TMU-CRD), and to explore the OHDSI's PLP study result as well.

Methods: Exploratory data analysis conducted with Superset's intuitive drag-and-drop interface and SQL capabilities for dashboard creation. Data queries were built using OHDSI's QueryLibrary, facilitating analysis of comorbidities and demographic trends, extracted from TMU-CRD CDM database. Visualization for PLP Results were derived from OHDSI's PLP package result, including XGBoost and LASSO regression, revealing predictive insights from patient features.



Results: Some tabular and charts were presented interactively, and some adjustments could be easily facilitated through internal SQL Lab across multiple databases.

Conclusion: This approach showcases Apache Superset's flexibility and accessibility for exploring large-scale health datasets. Its BI capabilities empower researchers to visualize and interpret complex patterns in stroke and cognitive impairment, paving the way for further clinical insights.

Limitation: While Apache Superset excels in flexibility and ease of use, it faces challenges in replicating standard statistical and evaluation metrics curves (e.g., ROC or calibration curves) commonly produced in dedicated statistical software. These limitations may require external tools or programming to supplement Superset's functionality for advanced model evaluations.

