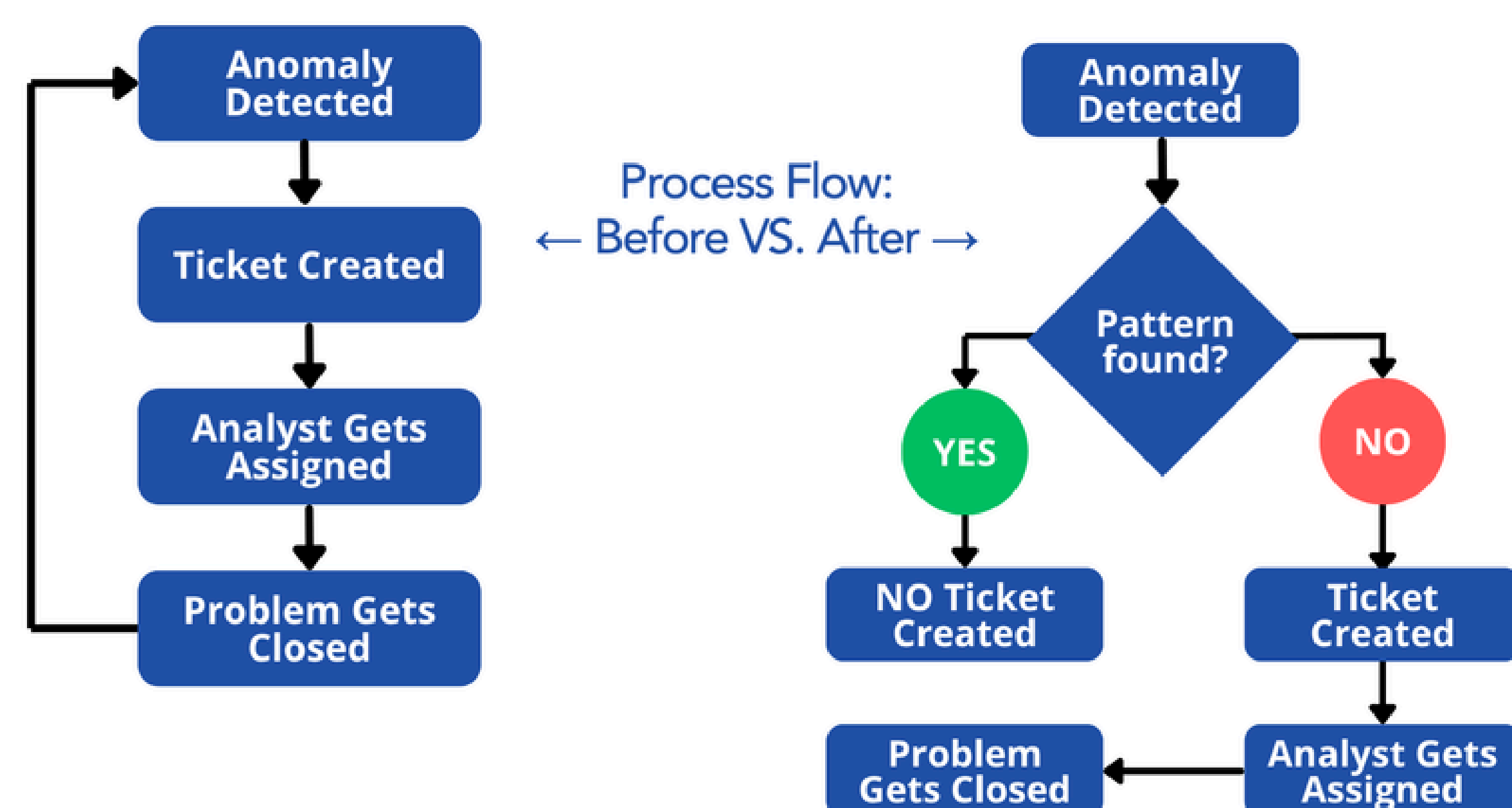




Enhancing Data Quality Monitoring Through Trend Analysis and Automated Anomaly Detection

Overview and Project Scope

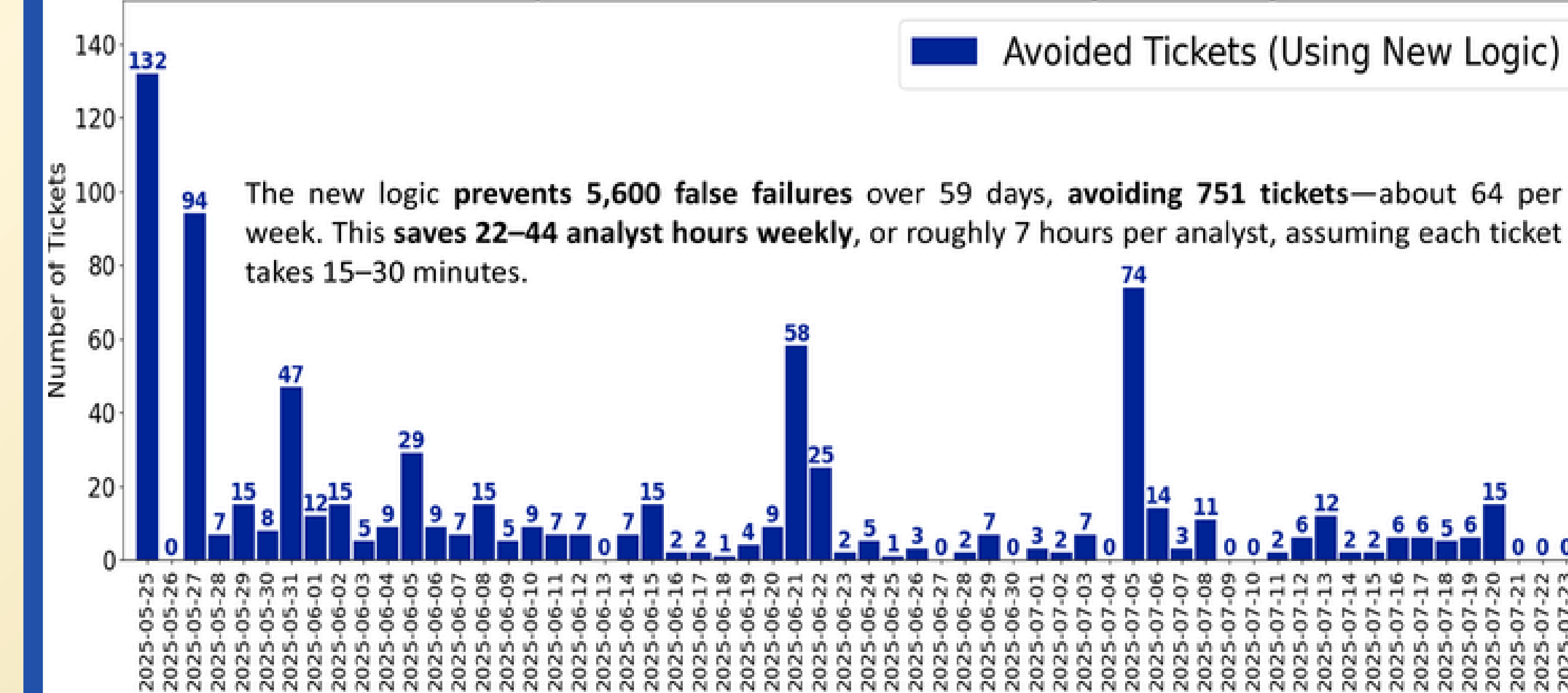
This project focused on improving Penske Truck Leasing's data quality framework to ensure data remains accurate, consistent, and reliable for analysis and decision-making. With millions of records processed daily, maintaining data integrity is essential for efficient operations and trustworthy insights. The project analyzed historical trends, introduced automated anomaly detection, and refined monitoring through validation, pattern recognition, and real-time issue resolution. These improvements enhanced visibility into data behavior, reduced manual investigation time, and minimized false alerts caused by recurring, expected patterns. The before-and-after process flow below shows how automation streamlined anomaly detection and strengthened overall data monitoring.



Roadmap & Process Flow

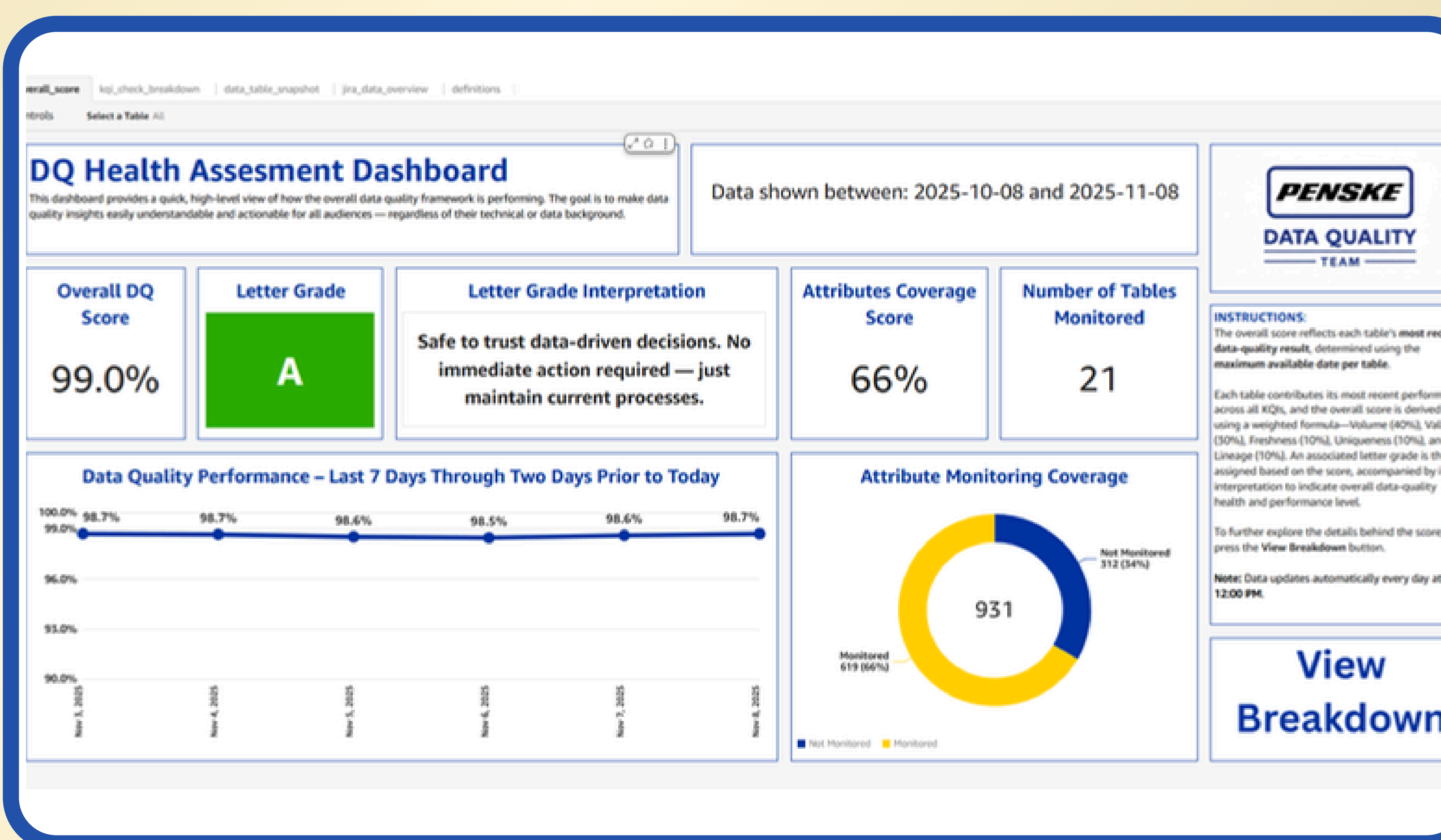
- 01 Data Exploration:** Analyzed 12 months of historical data quality metrics across multiple domains to uncover recurring anomalies and patterns.
- 02 Pattern Recognition:** Identified recurring data failure trends tied to holiday drops, operational fluctuations, provider offboarding, and vehicle inactivity, uncovering consistent behavior patterns that explained repeated quality issues.
- 03 Automation Logic:** Developed anomaly-adjustment rules to distinguish expected variations from real data quality failures.
- 04 Visualization:** Built AWS QuickSight dashboards to validate patterns and display performance across tables and key quality indicators.
- 05 Deployment Preparation:** Validated logic with senior data analysts, documented workflows in Confluence, and prepared for production implementation.

How Many Tickets Would Have Been Avoided Using the New Logic



Quantified Benefits

The proposed anomaly detection logic was tested on historical data to evaluate its effectiveness. Results showed that if this logic had been implemented during the 59-day period analyzed, it could have prevented approximately 5,600 false failures and avoided 751 support tickets, resulting in an estimated ~80% reduction in monthly ticket volume. This improvement could have saved 22–44 analyst hours per week, or about 7 hours per analyst, assuming each ticket takes 15–30 minutes to resolve. By reducing recurring false alerts, the logic allows analysts to focus their expertise on higher-impact analytical work.



Acknowledgements

I would like to sincerely thank the entire Penske team for making this internship possible and for providing an environment of learning and collaboration. I am especially grateful to my manager, Scott Robinson, and mentor, Ken Perna, for their guidance, support, and encouragement throughout this project.

David Sokurovi
dxs1216@case.edu
Master of Business Analytics and Intelligence' 25

