Our team’s role within Walmart was within the Supply Chain Product Management and Analytics team. We were tasked to create a tool which could help supply chain managers diagnose why their items were going out of stock.

This is an important and dire problem for Walmart – when we first released roughly 33% of item-node (fulfillment center) pairs were flagged as being out of stock across the six main Walmart fulfillment centers.

Our team’s final iteration of this tool had 10 different views to show various weighted and unweighted curs of the network’s out of stock situations, as well as over 20 filters.

Every 10-basis point increase in item-node in-stock percentage affects roughly $37,440 of demand per week top-line. By the end of our time with Walmart, the out of stock percentage had dropped by 200 basis points. This is an average increase in stock value of nearly $750,000 in weekly demand.

Utilizing 2,500+ lines of Hive queries and a Hadoop architecture, we engineered an ETL (“extract, transform load”) data pipeline that refreshes automatically each day. The final product of this process were two tables: a warehouse that shows a summary of all relevant stockout metrics as defined by our team at the item-node level as well as a database that breaks each item-node combination down by defect and can be leveraged for more flexibility in data visualization.

We used CART models to predict out of stock situations for every division of items within the Walmart network. Our CART results showed us that primary splits (variables most predictive of out of stock situations aka “root causes”) varied greatly by division. An example for the Everyday Living division is shown at left.

Special thanks to the entire Walmart/Jet team for being incredible sponsors and giving us a great summer, to the entire MIT team for their help and support in the organization of this project, and to all others who made this fun, relevant, and impactful project possible!