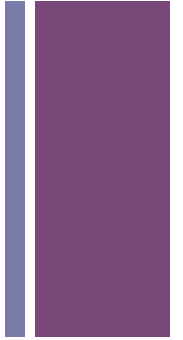




GA DS 30 Final Project

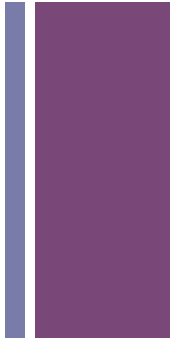
Ravi Kiran Tamiri

+ Problem



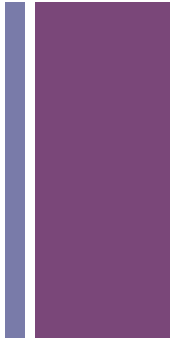
- Data Received from 150 different locations with different formats .
- A suite of scripts capture these files and run validation and Normalization
- A specific set of thresholds were set up on a legacy basis.
- Manual overrides happen to capture any changes in the thresholds.
- The file failures are ongoing and keep increasing on a dialy basis.

+ Solution



- Variables – Source data and type of database solution used.
- X = Company id , Historical failure metrics
- Y = Predicted New Threshold per metric
- Low training Error and good fit.

+ Challenges



- Specific model for customer or a an Overarching model
- Choosing the sample data
- Trusting the model to override the data (Explaining why)
- Time crunch due to personal milestones.