MICHIGAN STATE UNIVERSITY

# Beta Presentation

## Security Analytics Suite: Dataset Merger Tool

## The Capstone Experience


Team Avata

Jonny Dowdall
Paige Henderson
Matt Scheffler
Aasir Walajahi
Zac Wellmer

Department of Computer Science and Engineering
Michigan State University

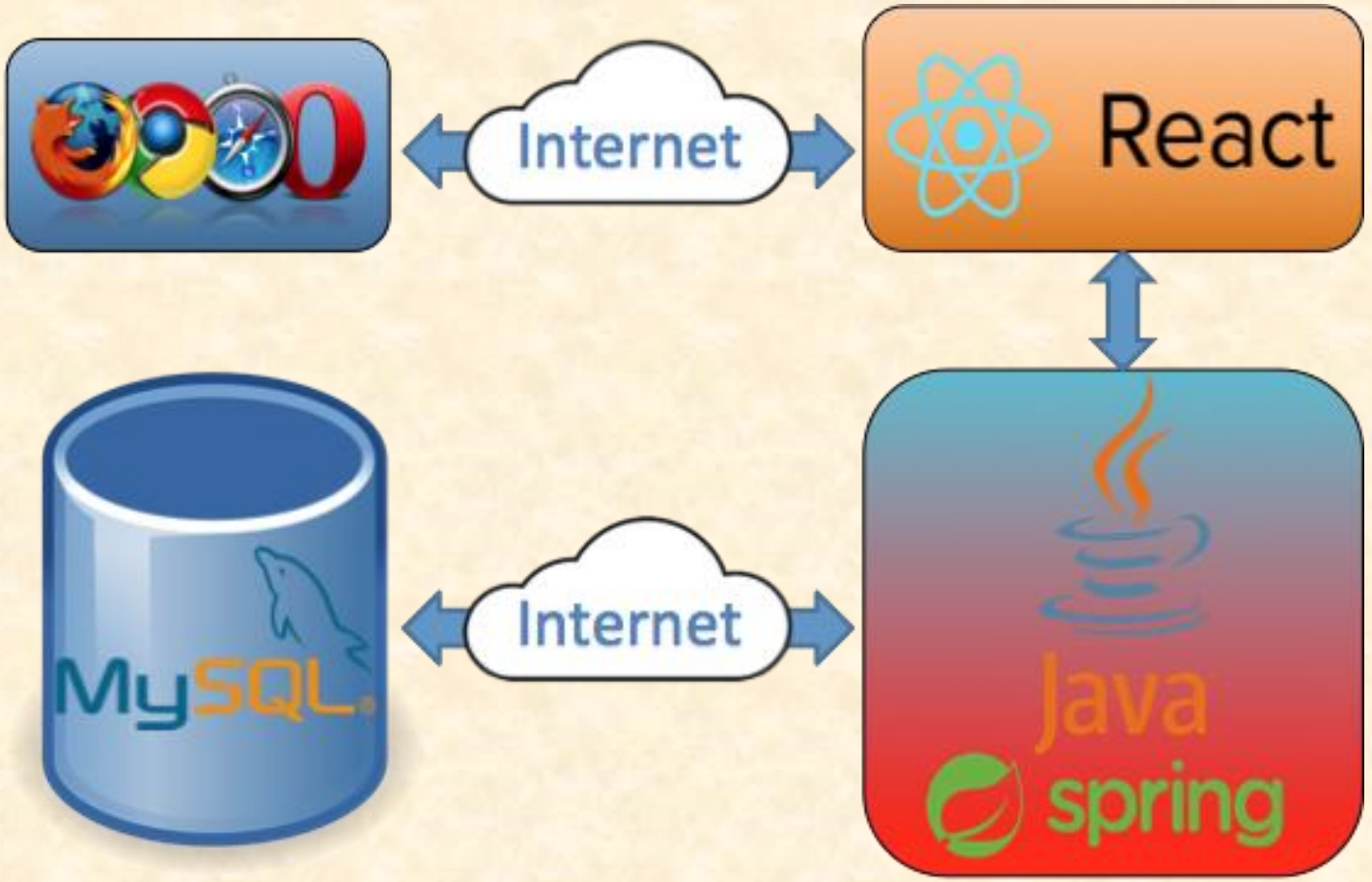

Fall 2016

*From Students…*
*…to Professionals*

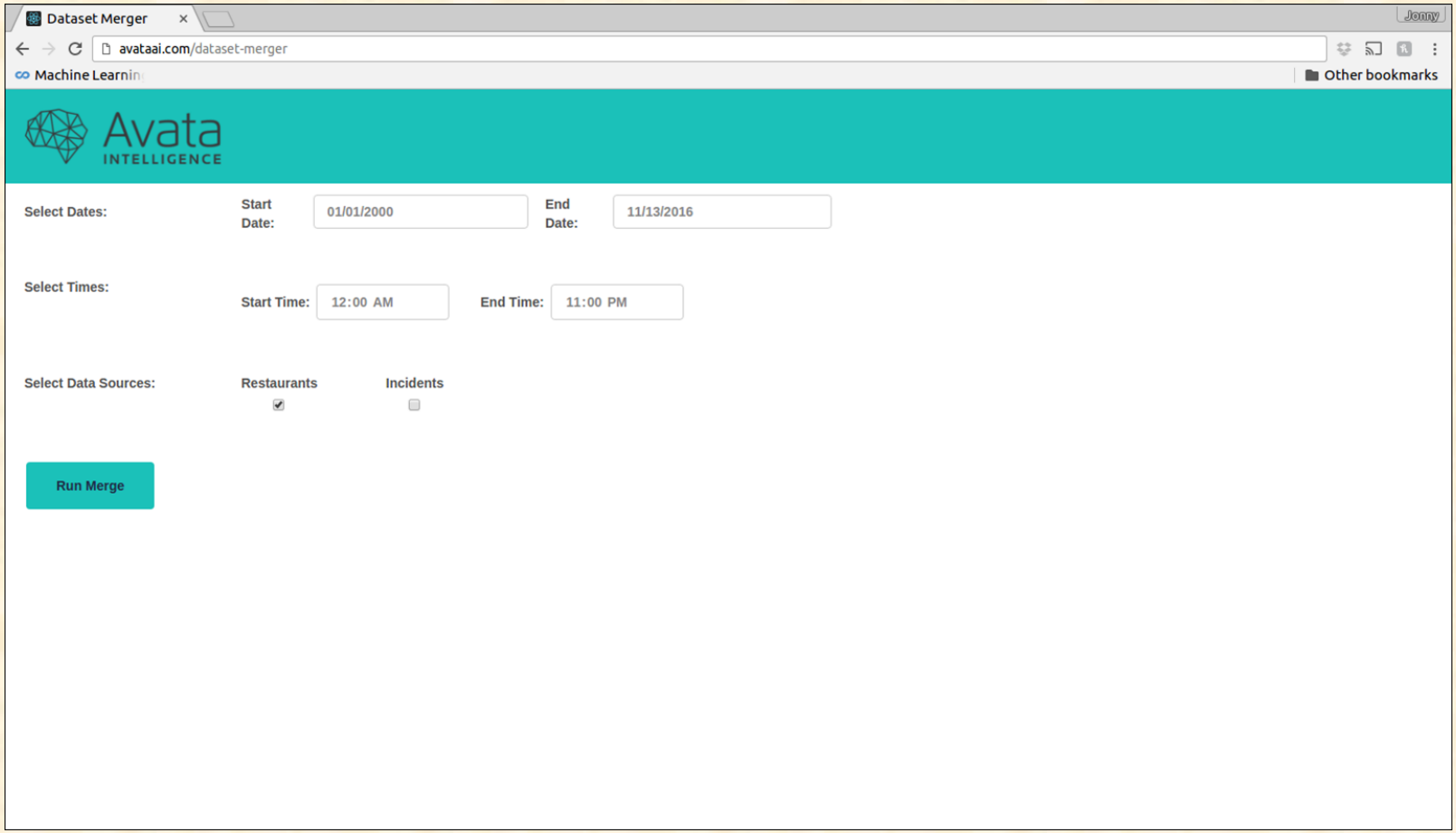
# Project Overview

- Automatically identify and merge duplicate records within and across datasets.
- Allow the user to to approve or disapprove the merging of similar records.
- Present the results as a report containing tables and charts.

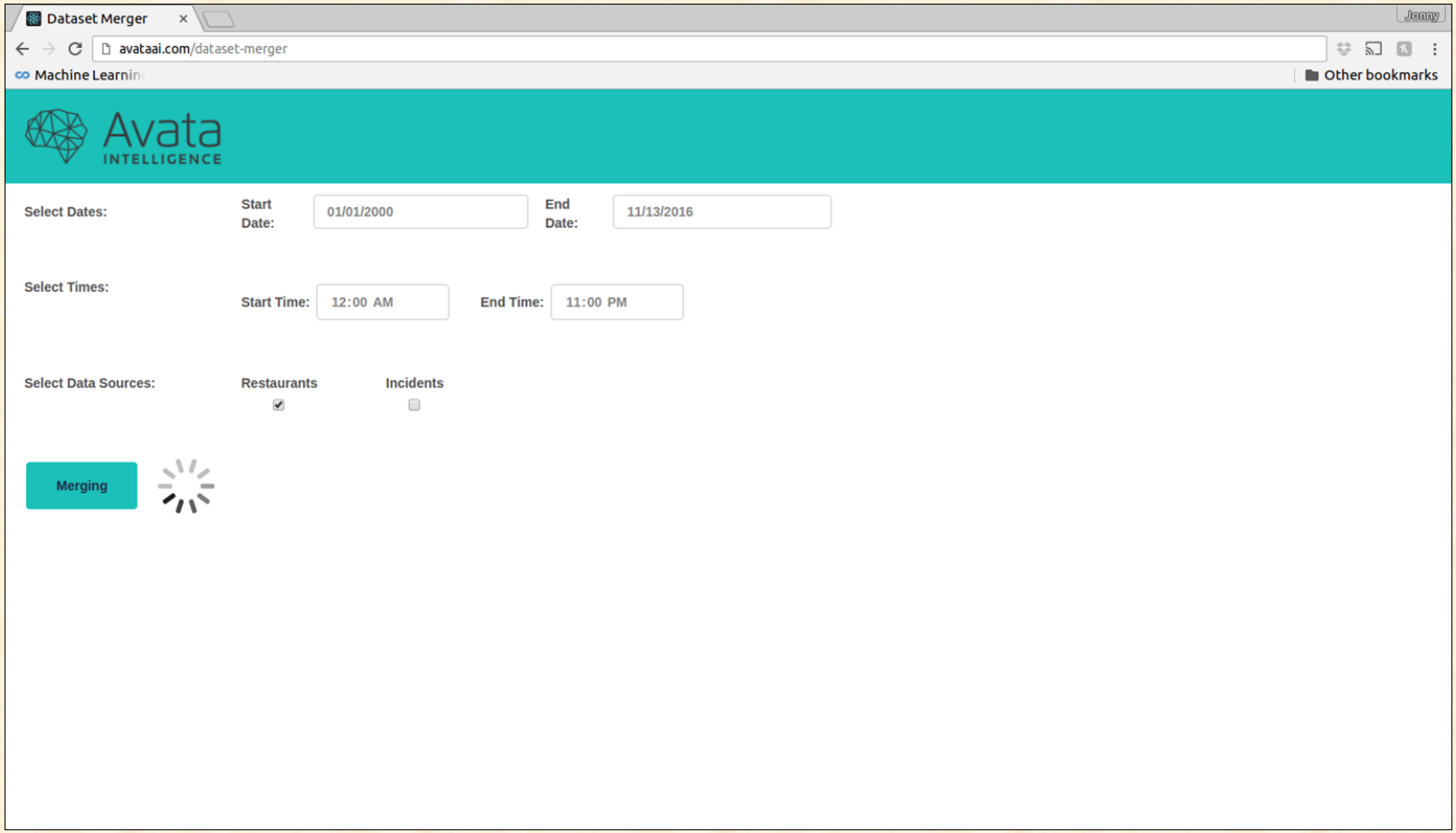
# System Architecture



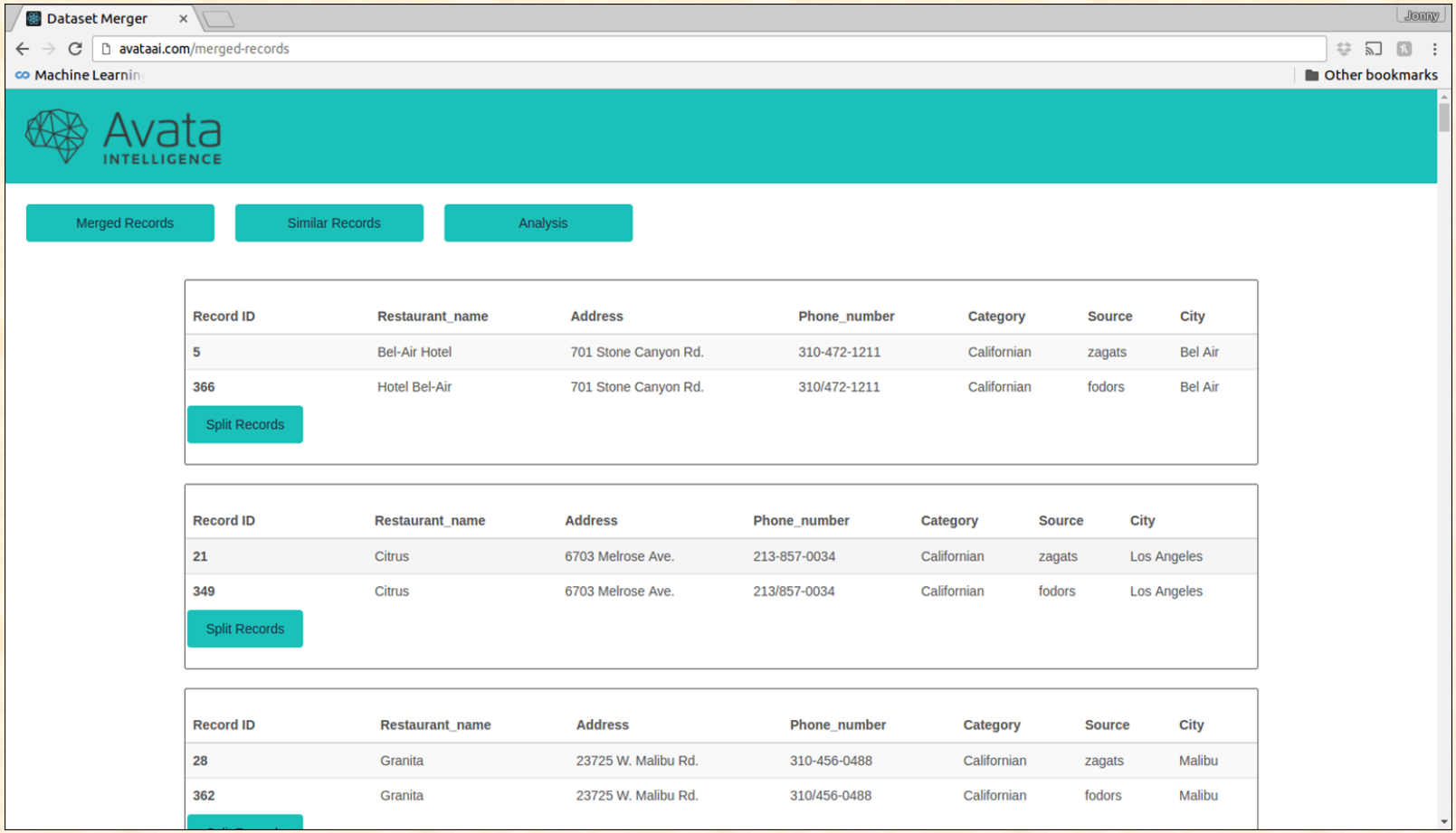
# Run Page


Dataset Merger
Jonny
avataai.com/dataset-merger
Machine Learnin
Other bookmarks
Avata
INTELLIGENCE
Select Dates:
Start Date:
01/01/2000
End Date:
11/13/2016
Select Times:
Start Time:
12:00 AM
End Time:
11:00 PM
Select Data Sources:
Restaurants
Incidents
Run Merge

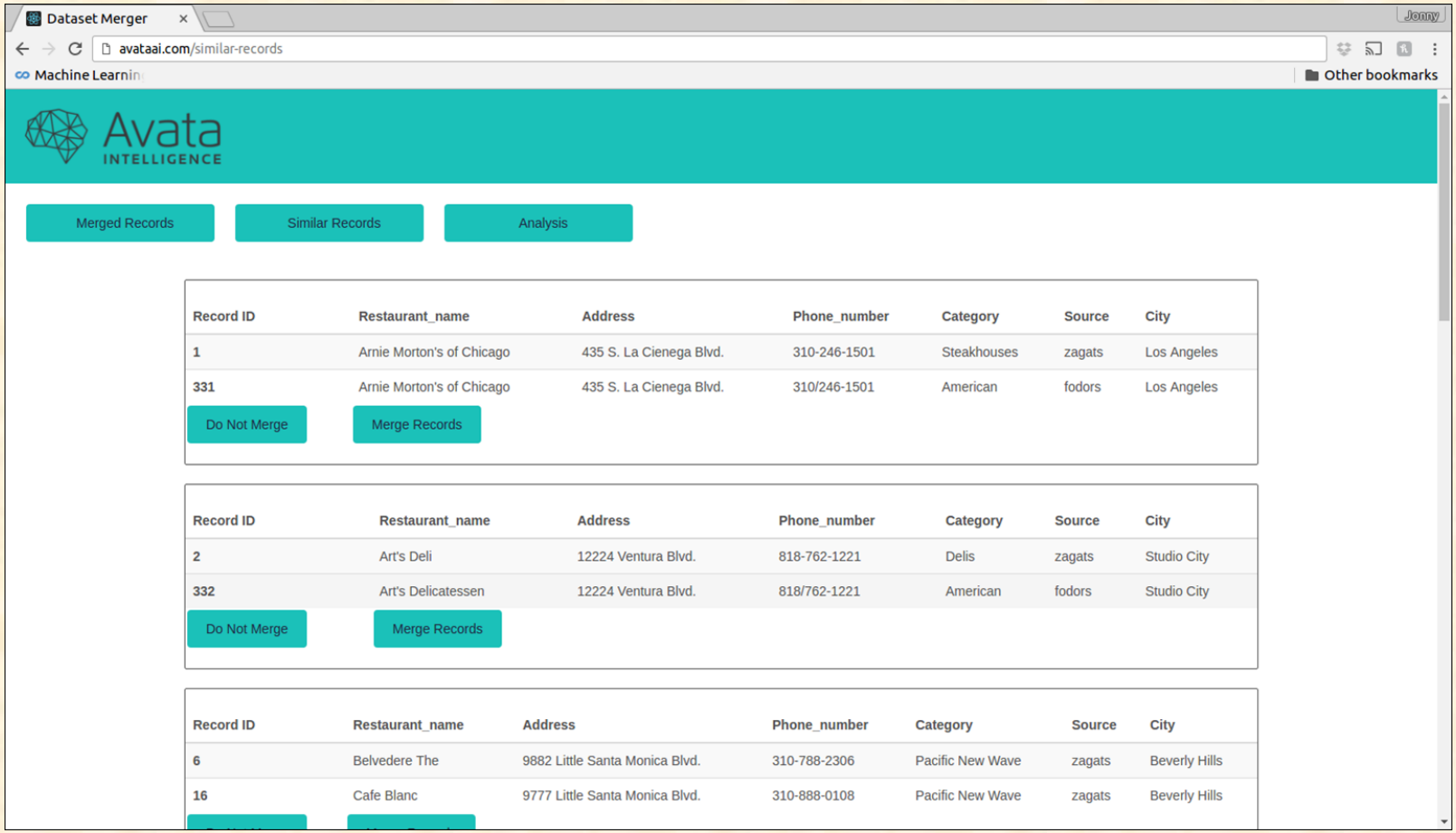
# Run Page


Dataset Merger
Jonny
avataai.com/dataset-merger
Machine Learnin
Other bookmarks
Avata
INTELLIGENCE
Select Dates:
Start Date:
01/01/2000
End Date:
11/13/2016
Select Times:
Start Time:
12:00 AM
End Time:
11:00 PM
Select Data Sources:
Restaurants
Incidents
Merging

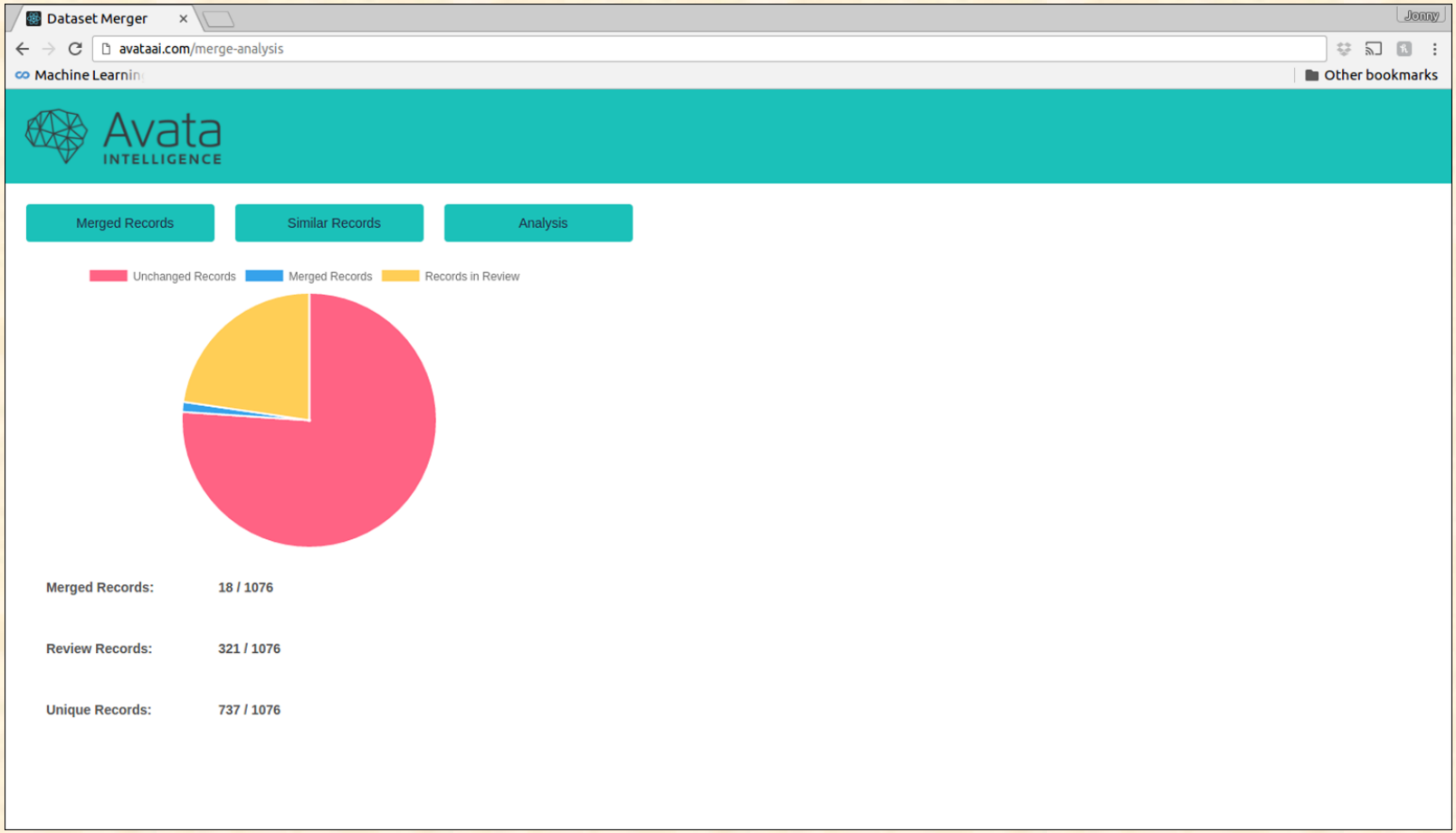
# Merged Records Page



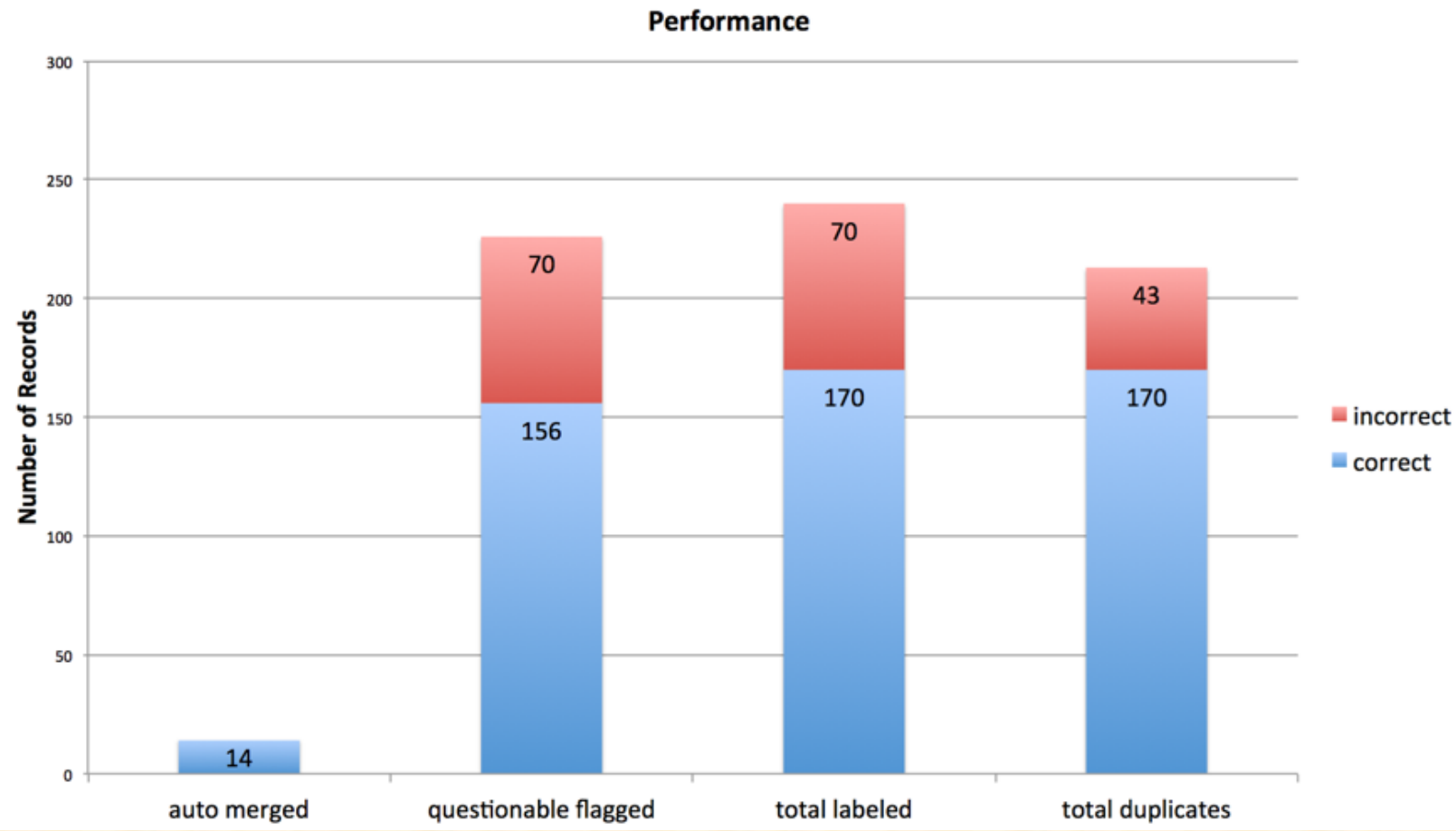
# Similar Records Page



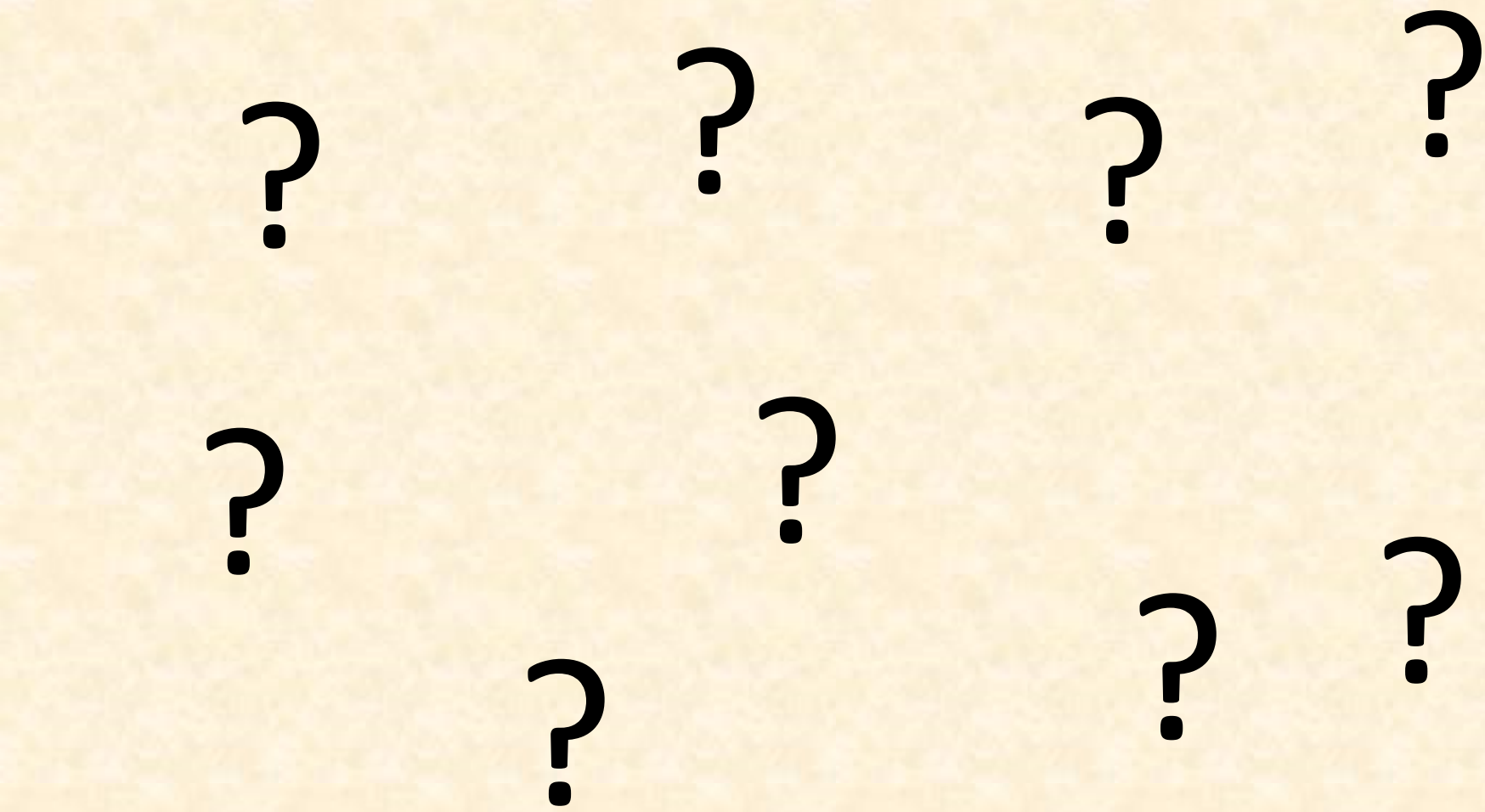
# Analysis Page



# Results


Performance
Number of Records
300
250
200
150
100
50
0
14
70
156
70
170
43
170
auto merged
questionable flagged
total labeled
total duplicates
incorrect
correct

# What's left to do?

- Improve SQL performance and security
- Improve parameterized query UX
- Styling fixes
- Add more statistics for analysis
- Finish testing
- Minor bug fixes

# Questions?