

# A Natural Language Query System for Data Visualization

CS Capstone Project

## Competency Demo

Brandon Wilson

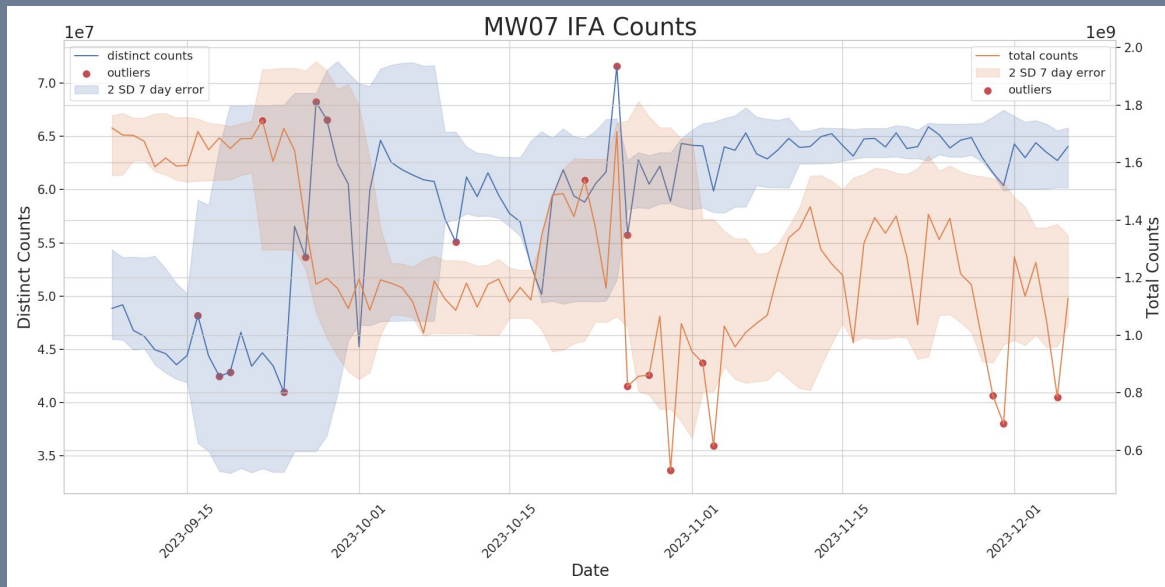
Tufts University  
Fall 2024

# Automated Report Building with Daily Data Flow

I have build an automated reporting script that will daily pull data from S3 and email a summary report of that data in comparison to data from the previous 90 days.

Also calculates a 2 SD error bar to keep an eye on longer term trends (shaded area).

Also a isolation forest outlier algorithm that detects high offending daily outliers.



# Automated Report Building with Daily Data Flow

## Max Date Checks

Athena max date check -- PASS

S3 date check -- PASS

| Source | Location        | Actual Max Date | Expected Max Date | Failure |
|--------|-----------------|-----------------|-------------------|---------|
| Athena | datasource.mw07 | 2023-12-06      | 2023-12-06        | 0       |
| S3     | mw07/           | 2023-12-06      | 2023-12-06        | 0       |

## Distinct IFA counts from the previous week

Android percent valid carrier and model vs. contract target check -- FAILURE. Percent of contract target: 93% with 35921643 valid of 38475000 count contract target

|   | Start Date | End Date   | OS      | Total IFA count | IFA count w valid gps or ip | IFA count w valid gps or ip and model and carrier | IFA count w valid model | IFA count w valid carrier | Percent valid vs. contract target | Percent valid carrier and model vs. contract target | Percent valid total IFAs | Percent model fill total valid IFAs | Percent carrier fill total valid IFAs |
|---|------------|------------|---------|-----------------|-----------------------------|---|-------------------------|---------------------------|-----------------------------------|---|--------------------------|-------------------------------------|---------------------------------------|
| 0 | 2023-11-30 | 2023-12-06 | android | 85,723,633      | 49,770,526                  | 35,921,643  | 49,491,935              | 36,114,639                | 110%                              | 93%   | 58%                      | 99%                                 | 72%                                   |
| 1 | 2023-11-30 | 2023-12-06 | ios     | 116,877,419     | 72,898,235                  | 26,093,065  | 50,529,485              | 34,761,515                | 303%                              | 152%  | 62%                      | 69%                                 | 47%                                   |

Also implemented is a series of calculations of that ensured that the data supplier was still within their contractual limits with us.

There is also a check that we received latest data (green table). The rows would be red if we were missing data.

# Data Loading and Preview with a UI and AWS

[Video Tour](#)