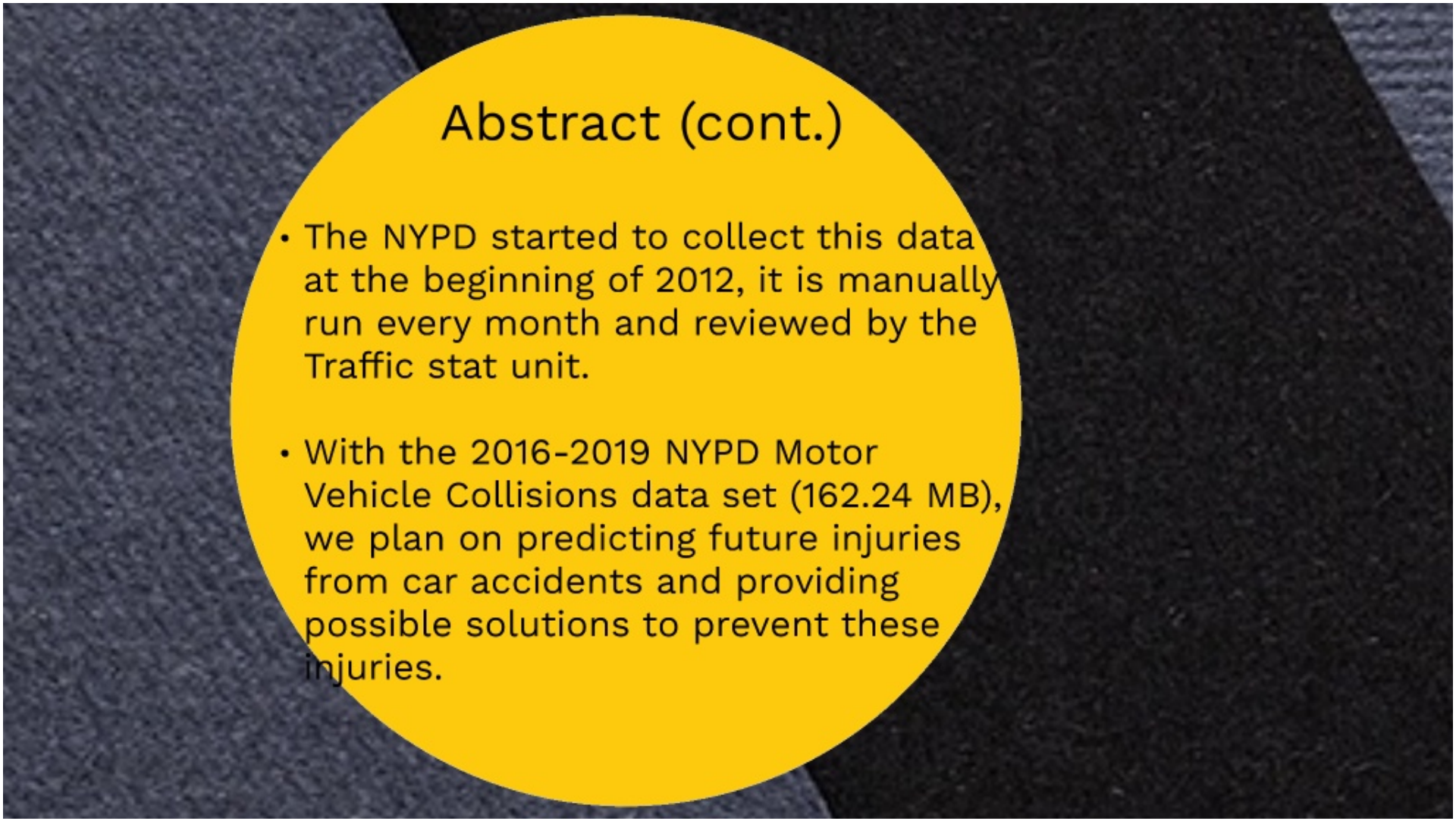
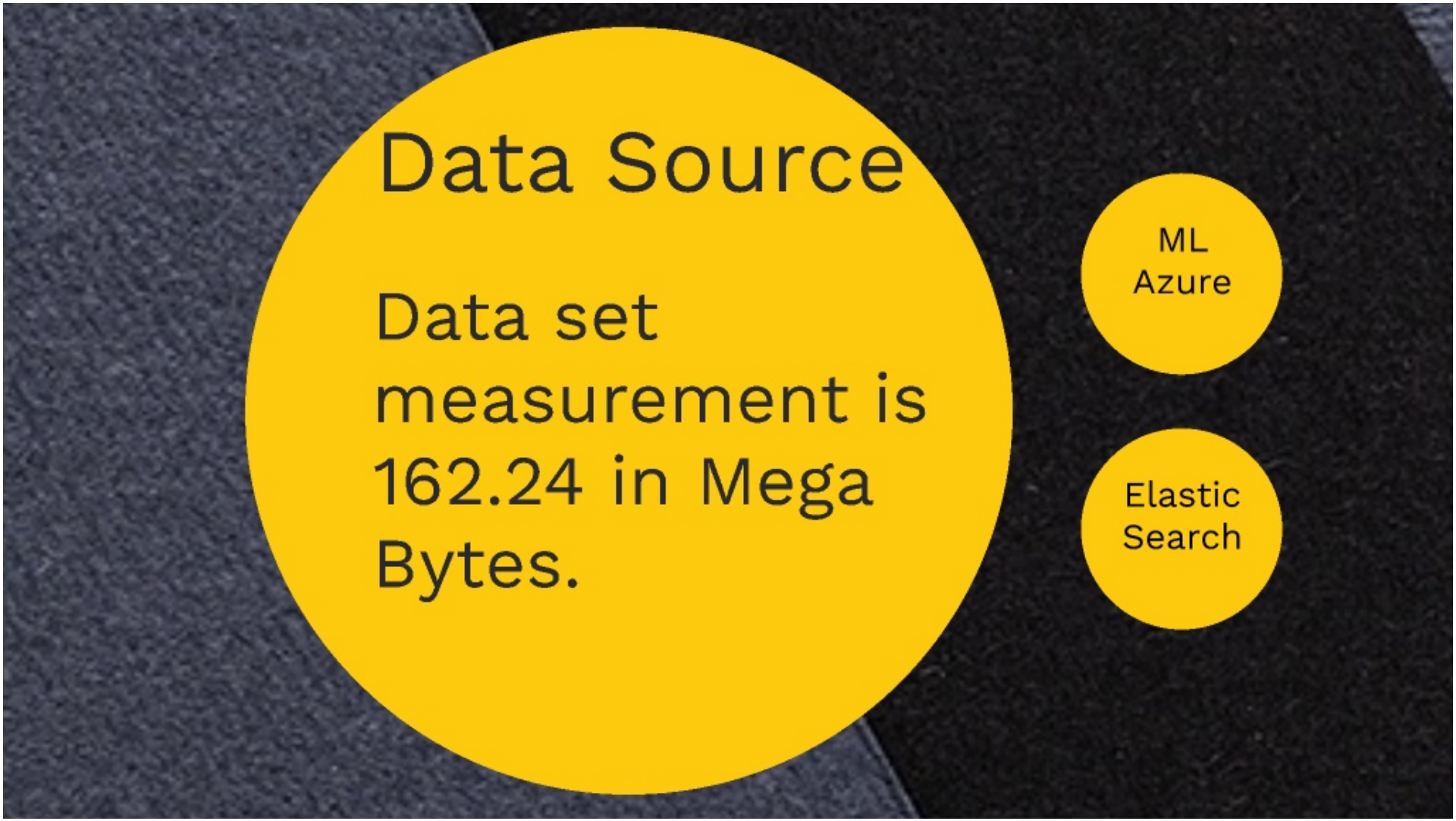
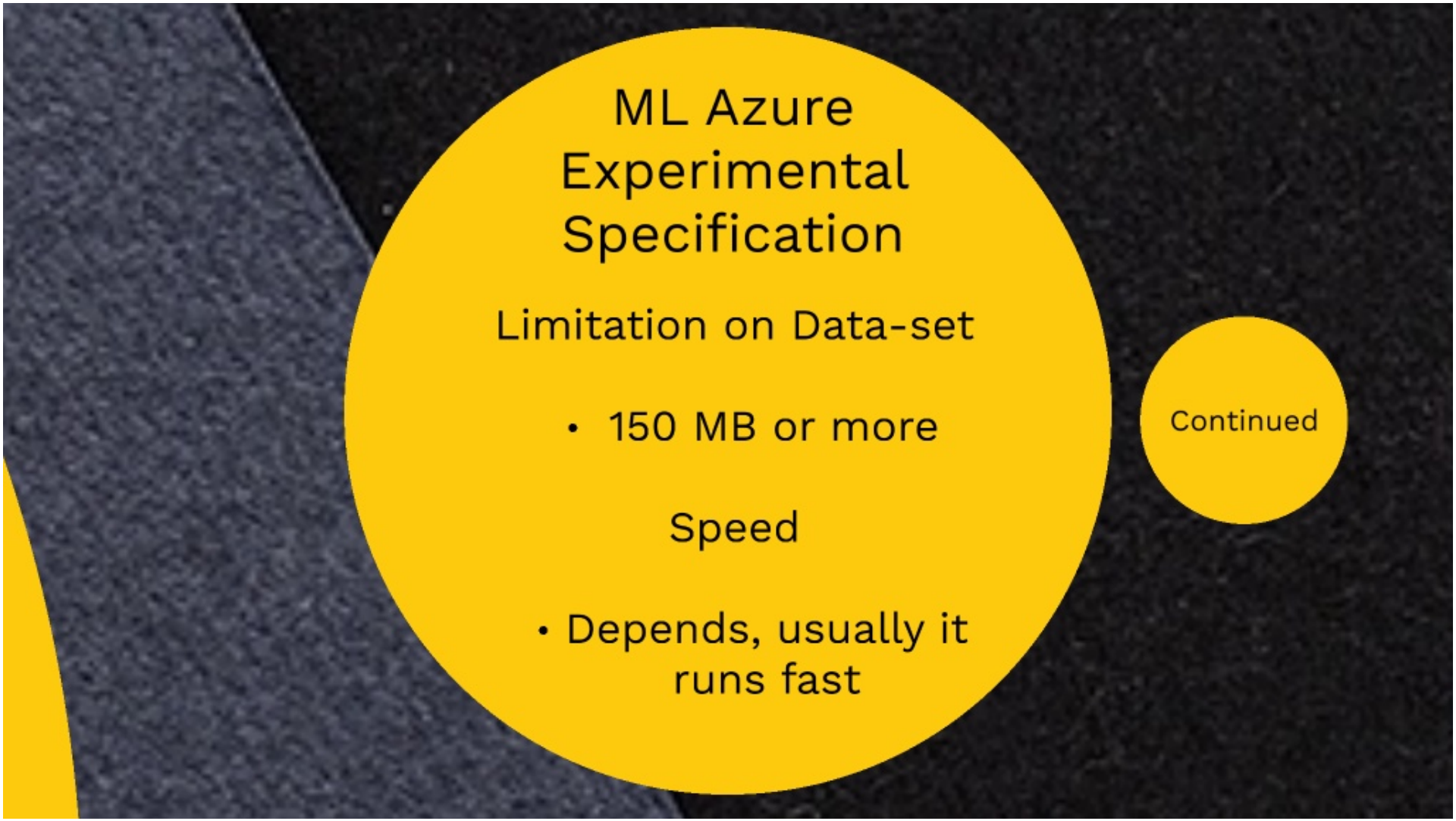
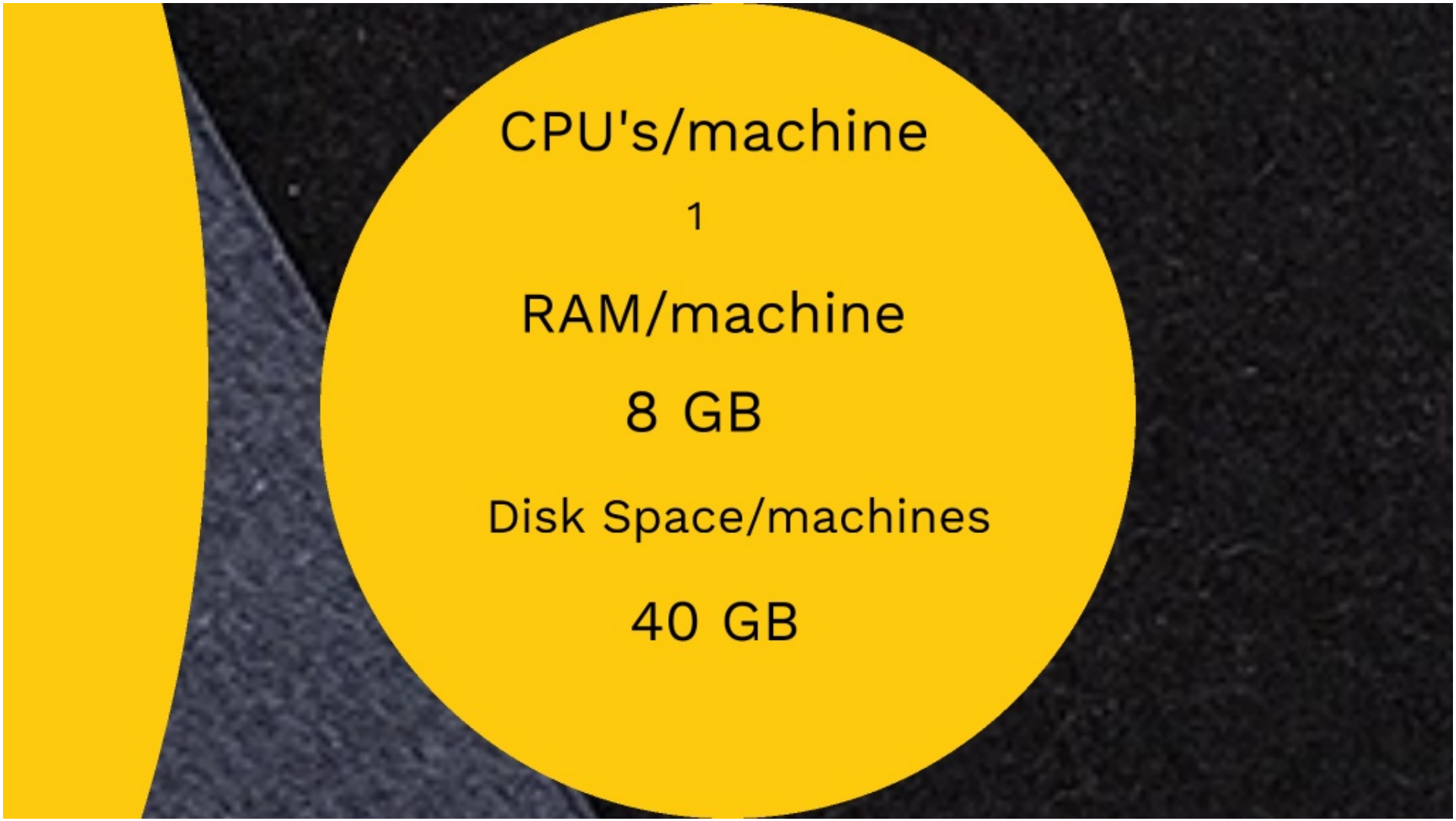

New York Police Department
Injury Report
Presented by
Lucky Trang, Ron Tsang
Stephanie Nguyen, Paul Yi
Intro
Body
CON
REC

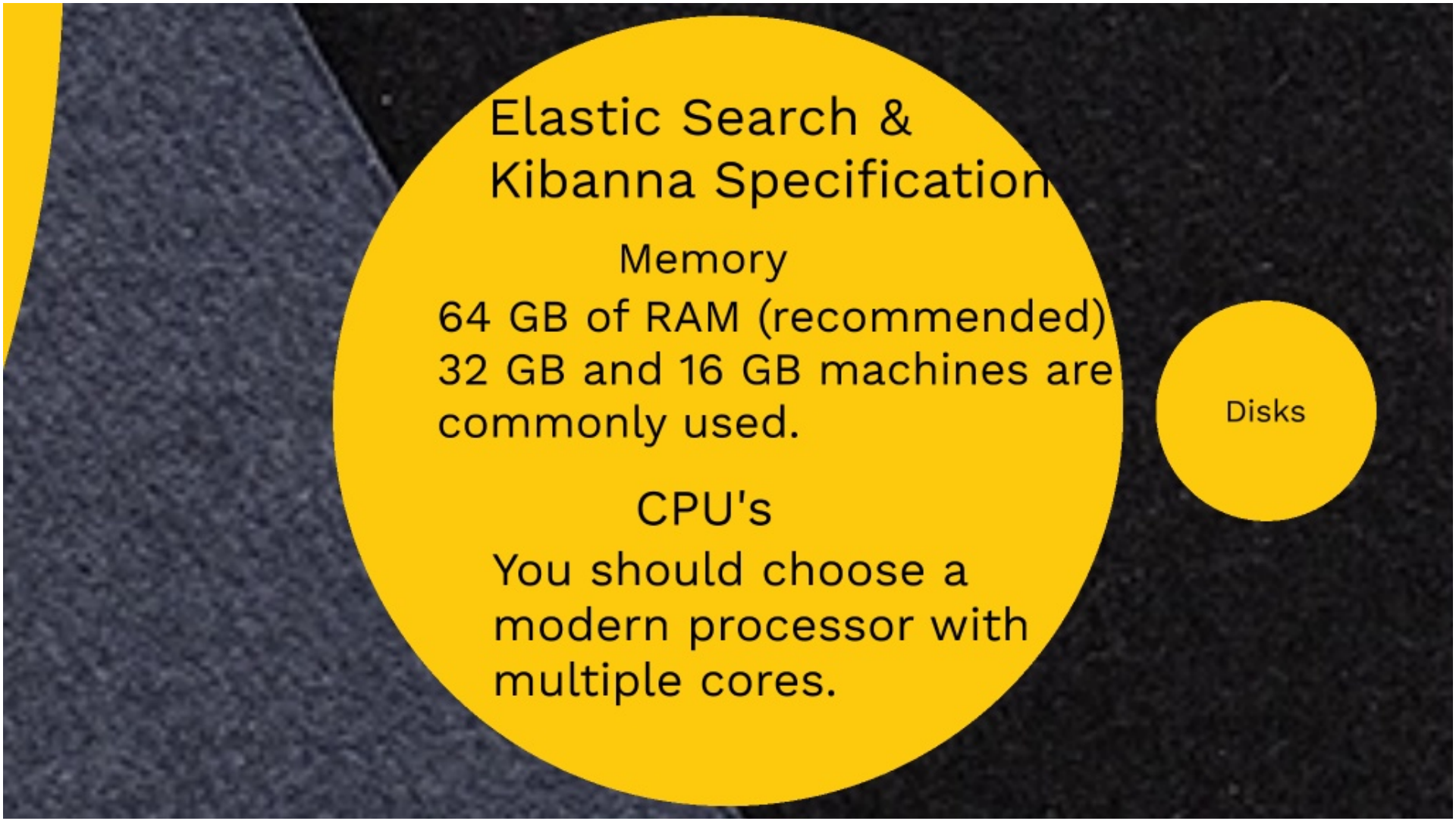

Abstract
According to the World Health Organization, approximately, 1.35 million people die each year from car collisions.
Continue
Data Source
Git Hub

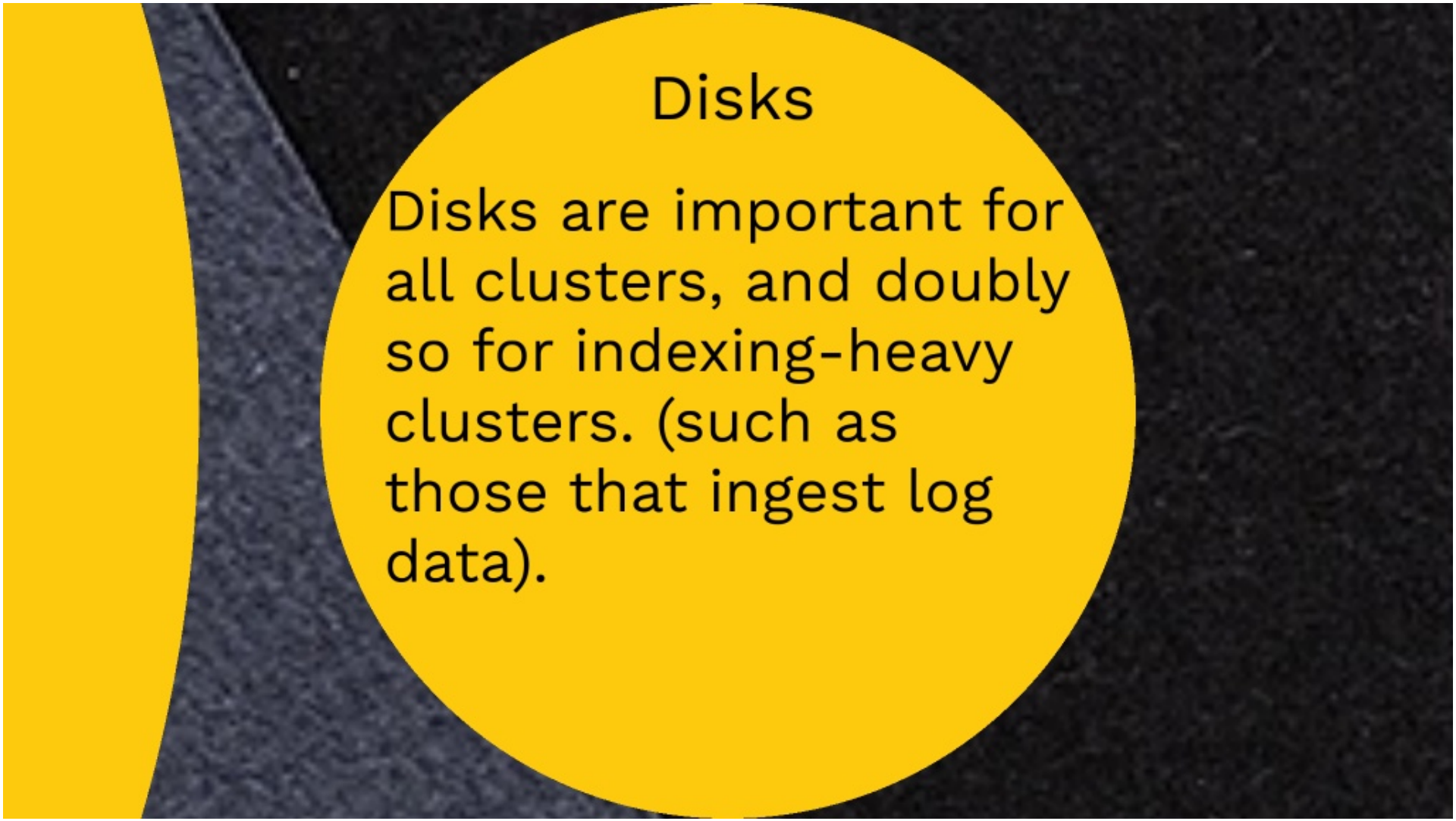

Abstract (cont.)
• The NYPD started to collect this data at the beginning of 2012, it is manually run every month and reviewed by the Traffic stat unit.
• With the 2016-2019 NYPD Motor Vehicle Collisions data set (162.24 MB), we plan on predicting future injuries from car accidents and providing possible solutions to prevent these injuries.


Data Source
Data set measurement is 162.24 in Mega Bytes.
ML Azure
Elastic Search


ML Azure Experimental Specification
Limitation on Data-set
• 150 MB or more
Speed
• Depends, usually it runs fast
Continued

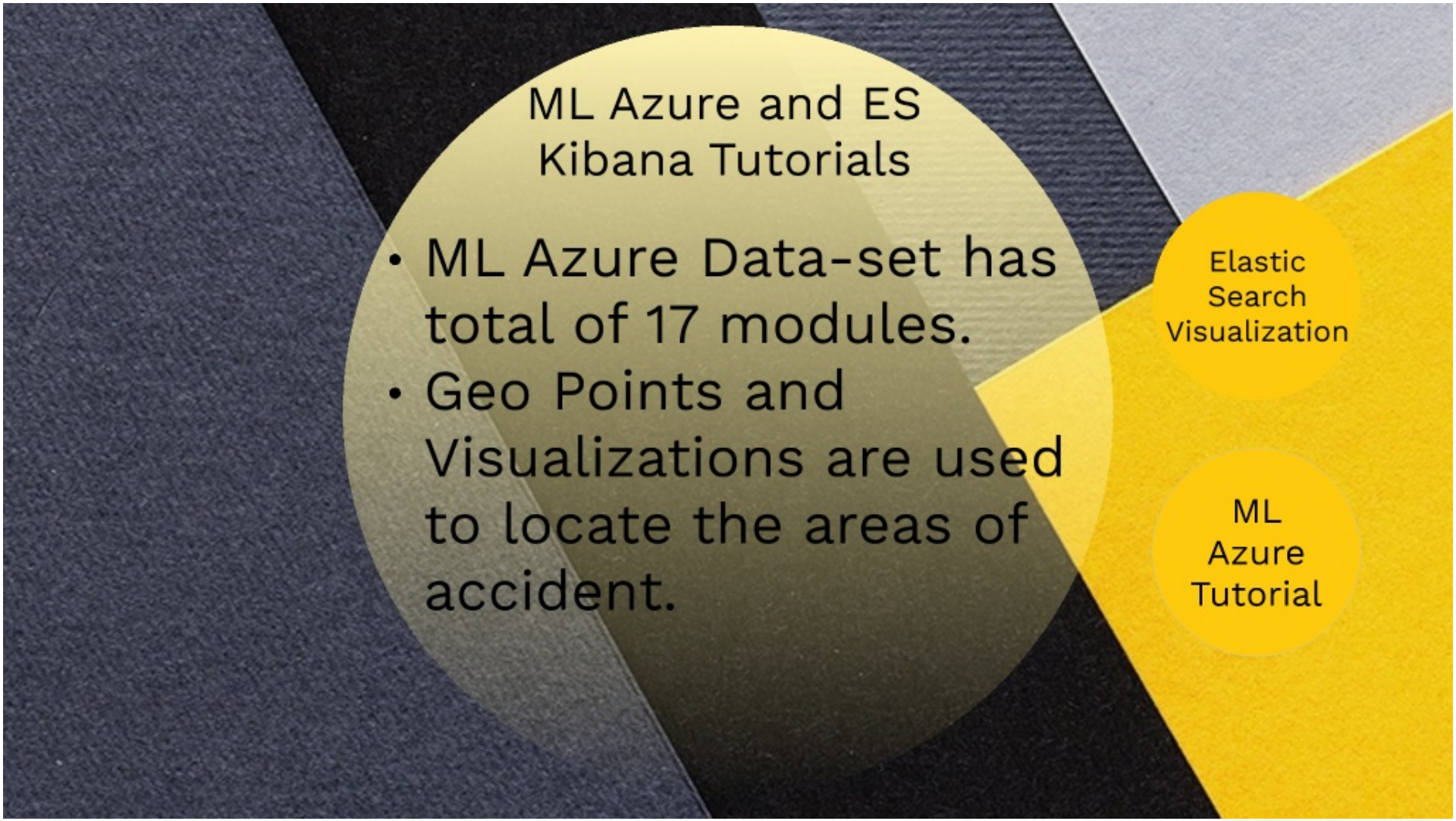

CPU's/machine
1
RAM/machine
8 GB
Disk Space/machines
40 GB


Elastic Search &
Kibanna Specification
Memory
64 GB of RAM (recommended)
32 GB and 16 GB machines are
commonly used.
CPU's
You should choose a
modern processor with
multiple cores.
Disks

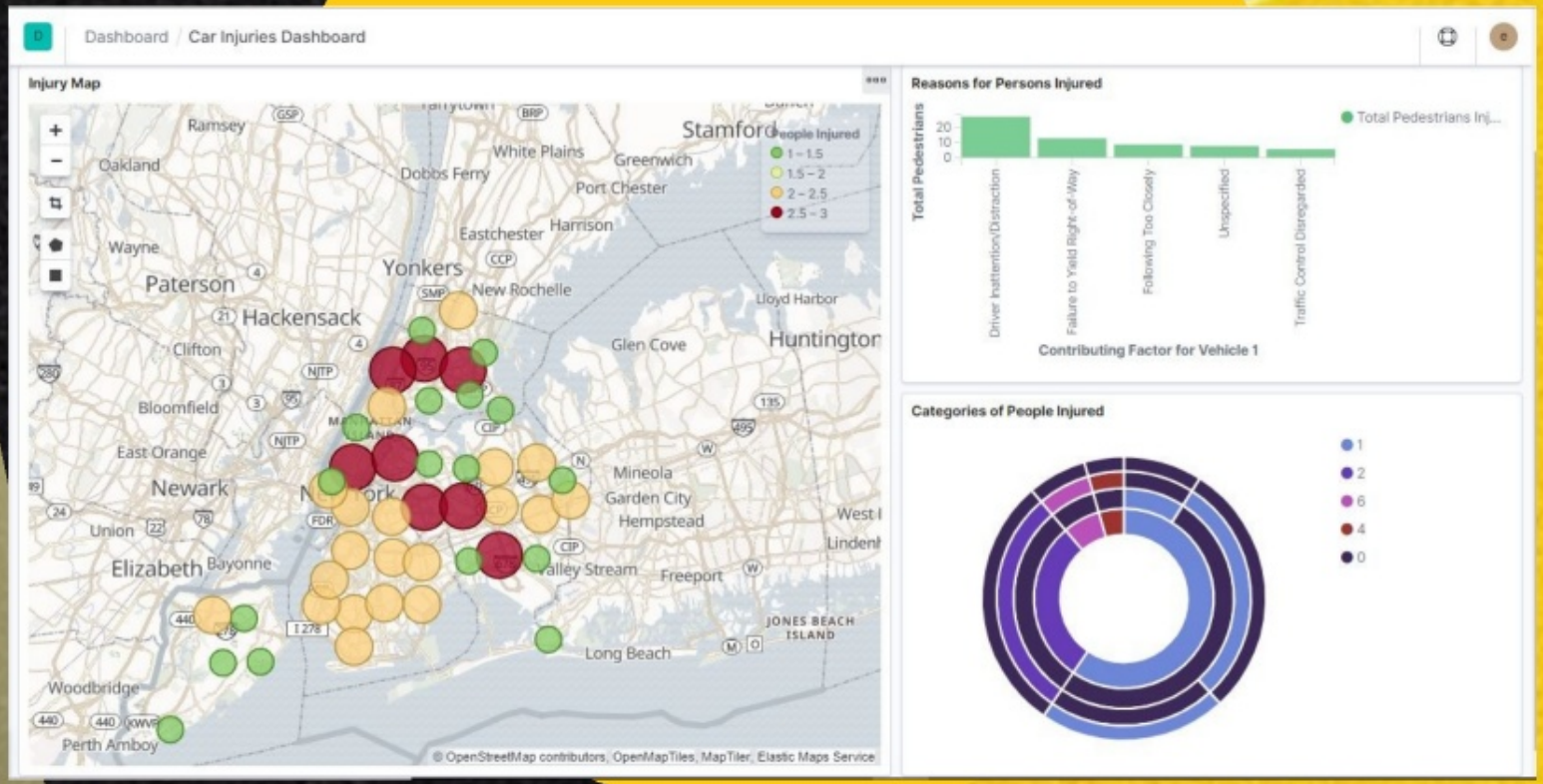

Disks
Disks are important for all clusters, and doubly so for indexing-heavy clusters. (such as those that ingest log data).


Git Hub Link
Link here


New York Police Department
Injury Report
Presented by
Lucky Trang, Ron Tsang
Stephanie Nguyen, Paul Yi
Intro
Body
CON
REC


ML Azure and ES
Kibana Tutorials
• ML Azure Data-set has total of 17 modules.
• Geo Points and Visualizations are used to locate the areas of accident.
Elastic Search Visualization
ML Azure Tutorial

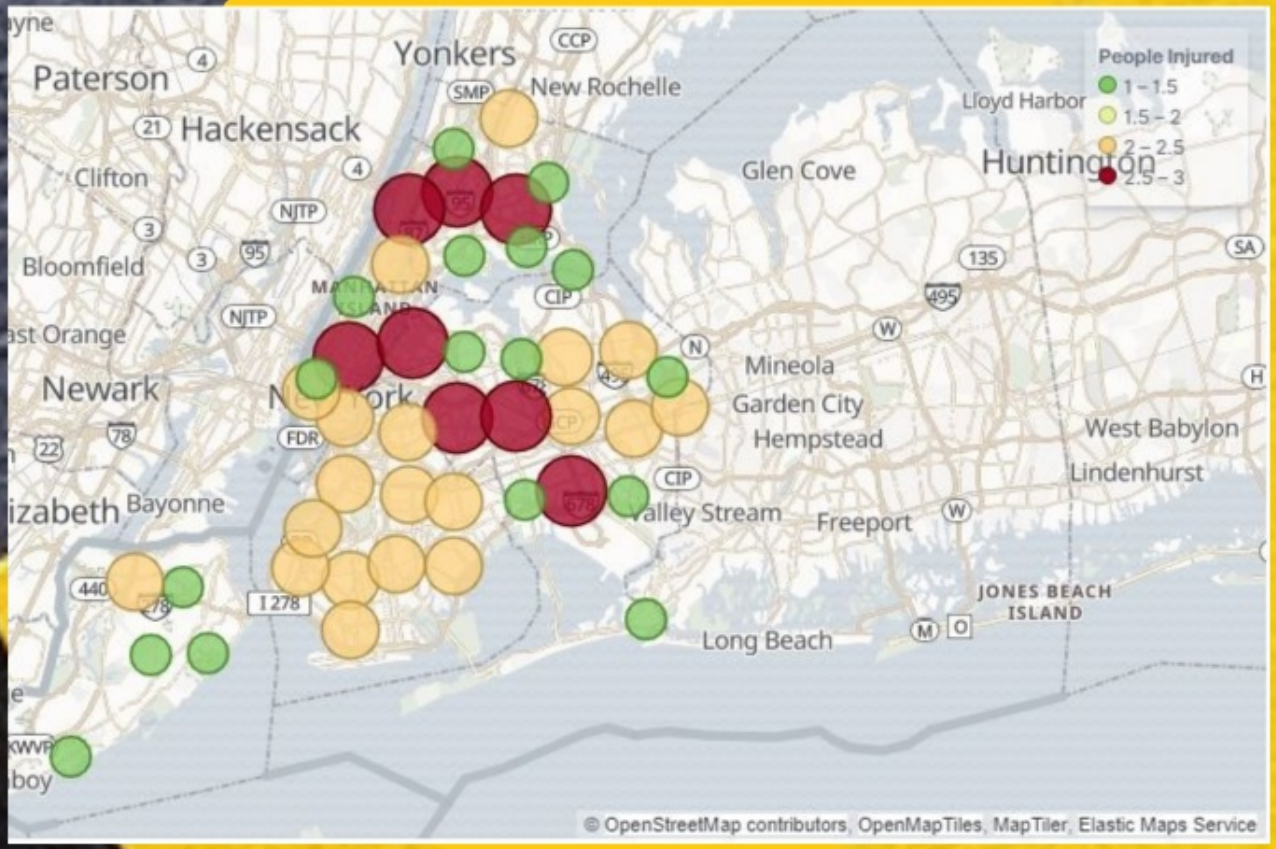
Elastic Search Dashboard



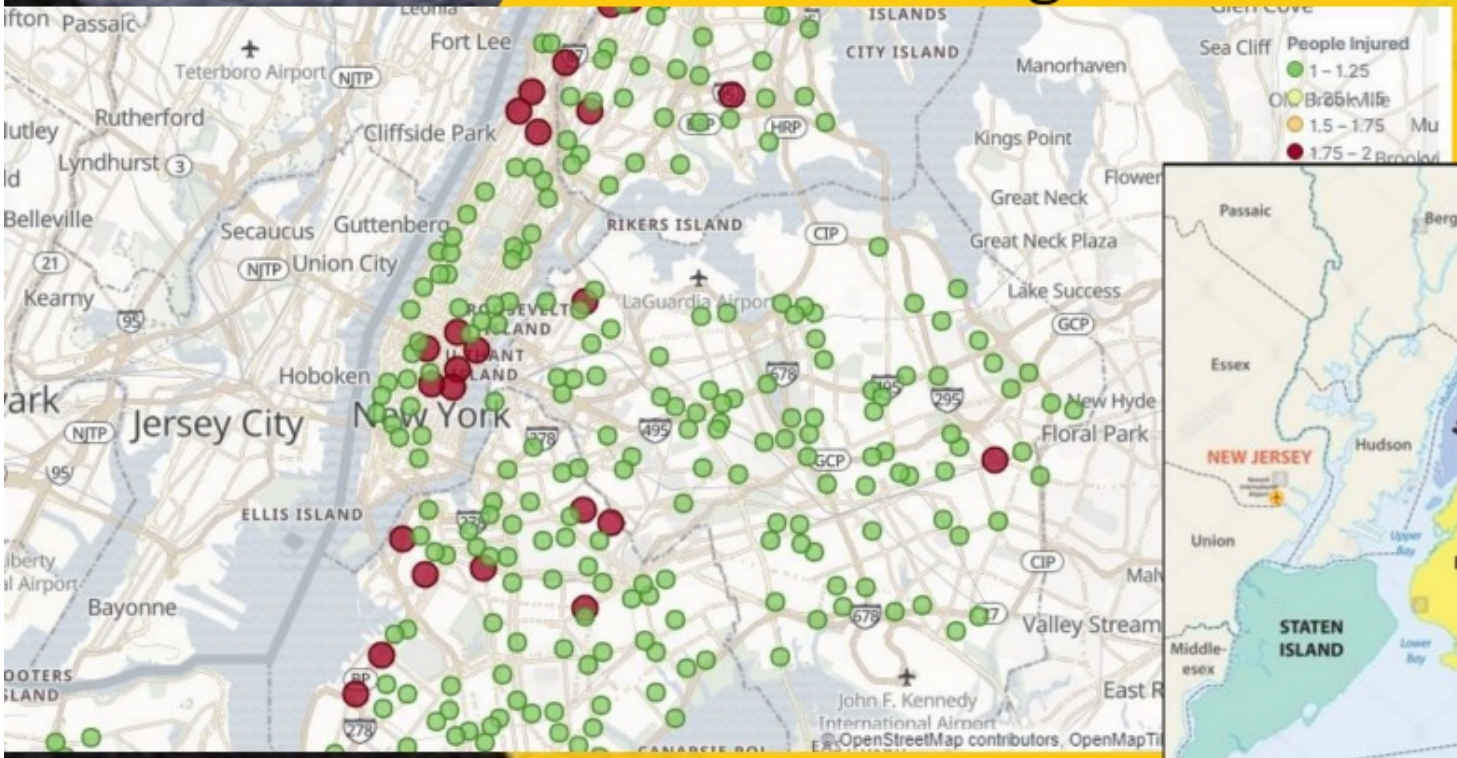
Elast
Searc
Geo Po

Elast
Searc
Map

Elastic Search Map



Elastic Search Map - Boroughs

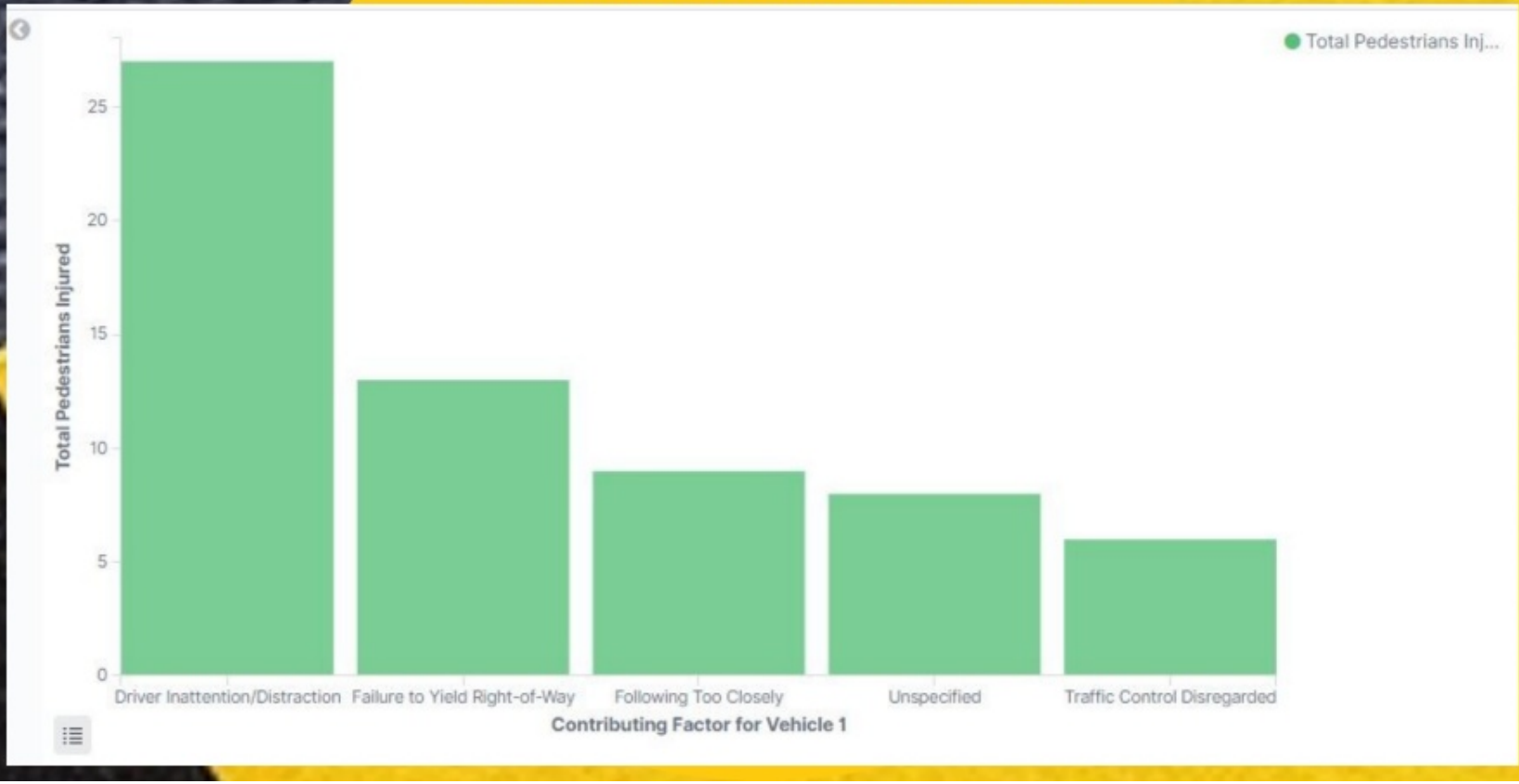


Elastic Search - Categories of Injuries

Total: 137

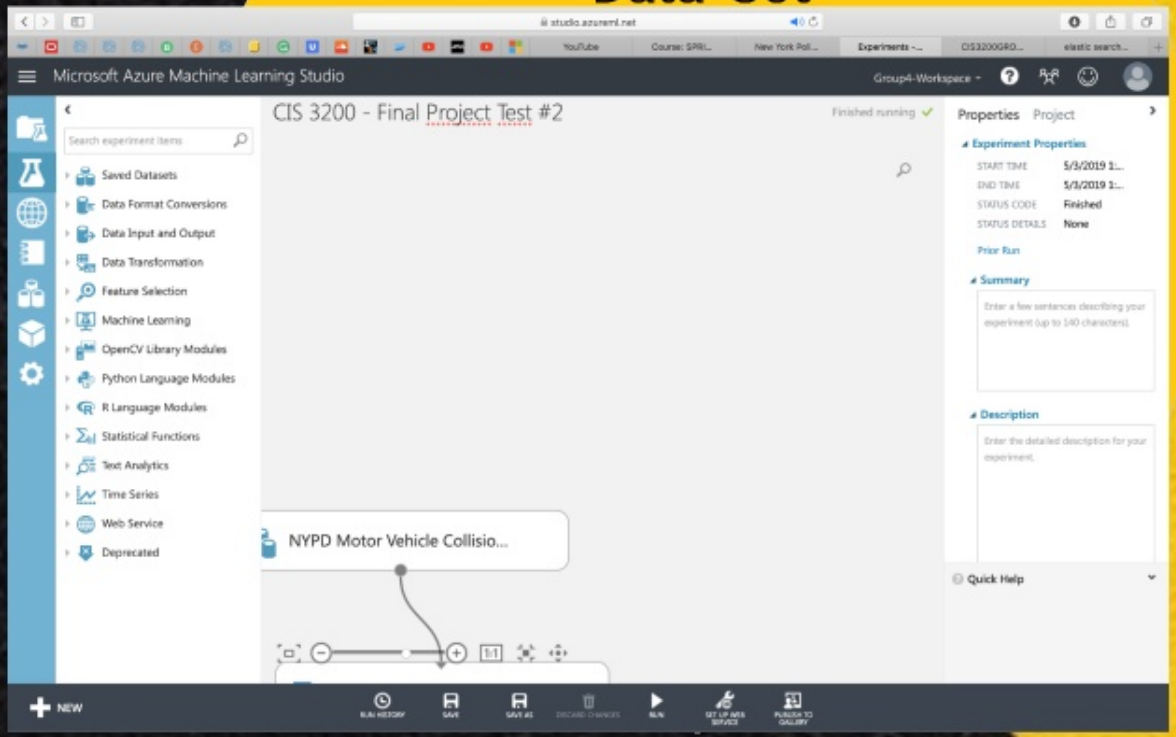
- Cyclists:
  - 1 - 15.09%
- Motorists
  - 1 - 49.06%
  - 2 - 29.21%
  - 4 - 4.49%
  - 6 - 6.74%
- Pedestrians:
  - 1 - 35.85%

Elastic Search - Contribution Factors



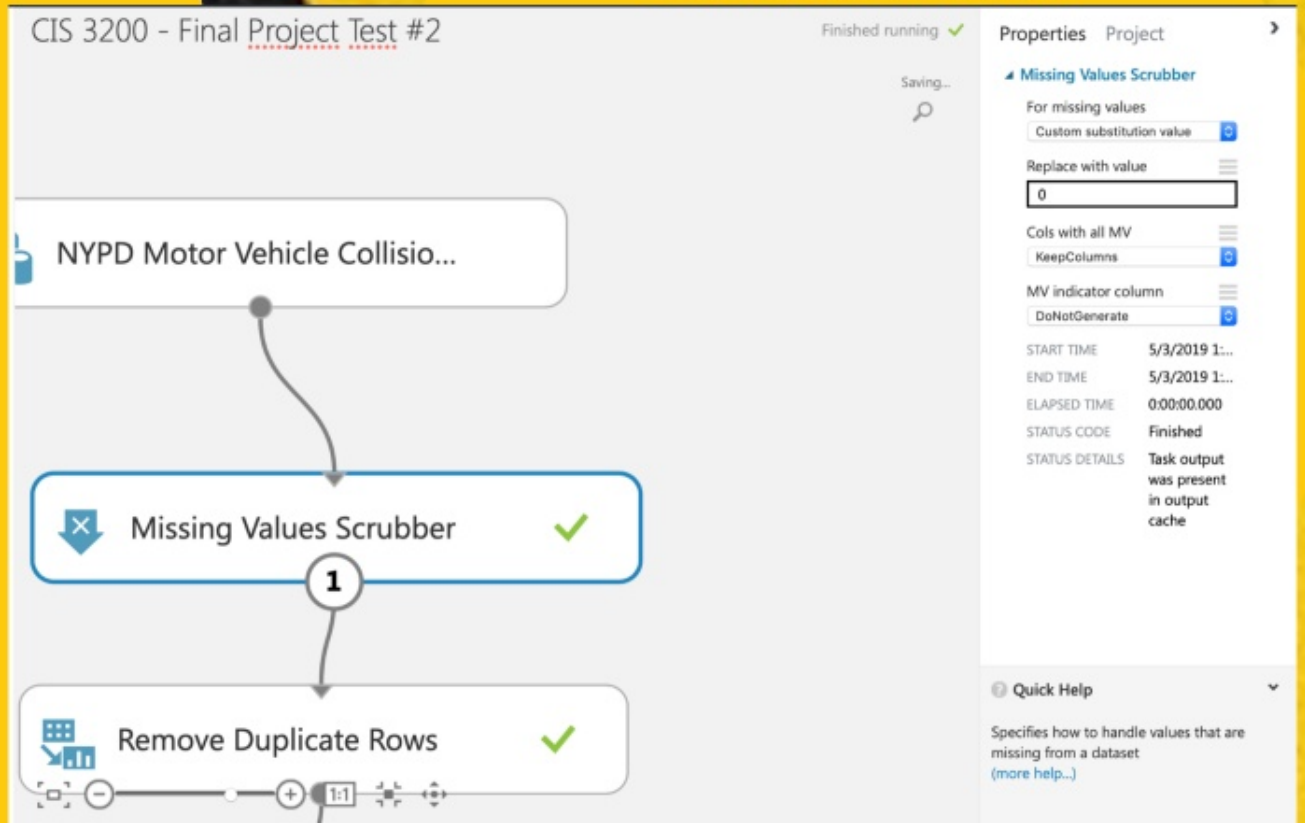
ML Azure Tutorial

Step 1 Upload the Data-Set



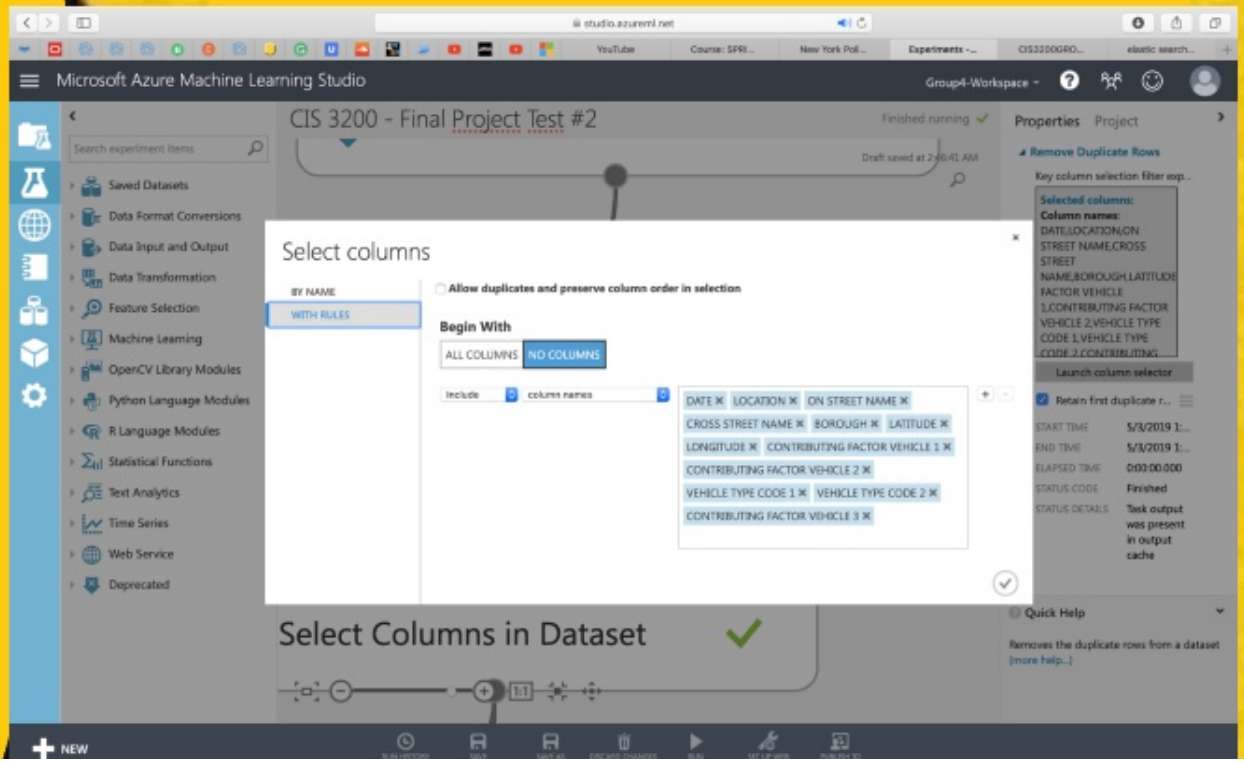
Step 2

Step 2
Missing Values Scrubber module



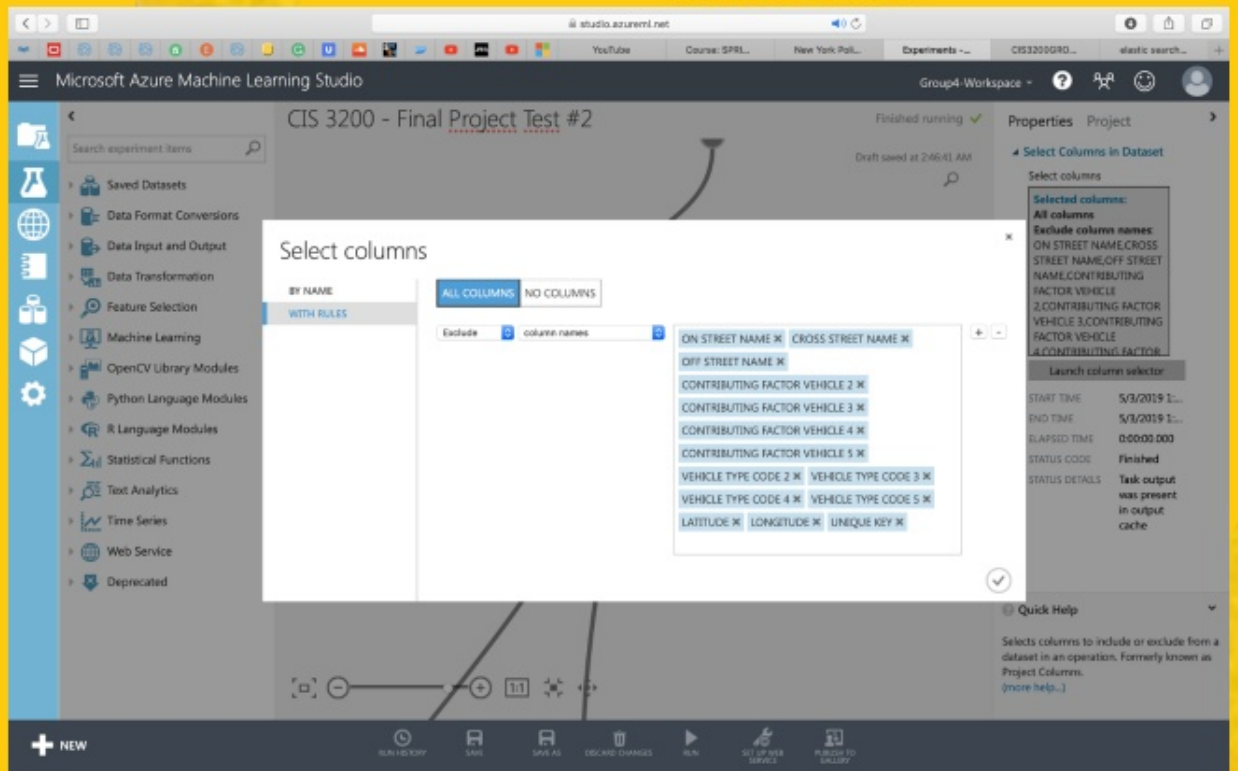
Step 3

Step 3
Remove duplicate rows module



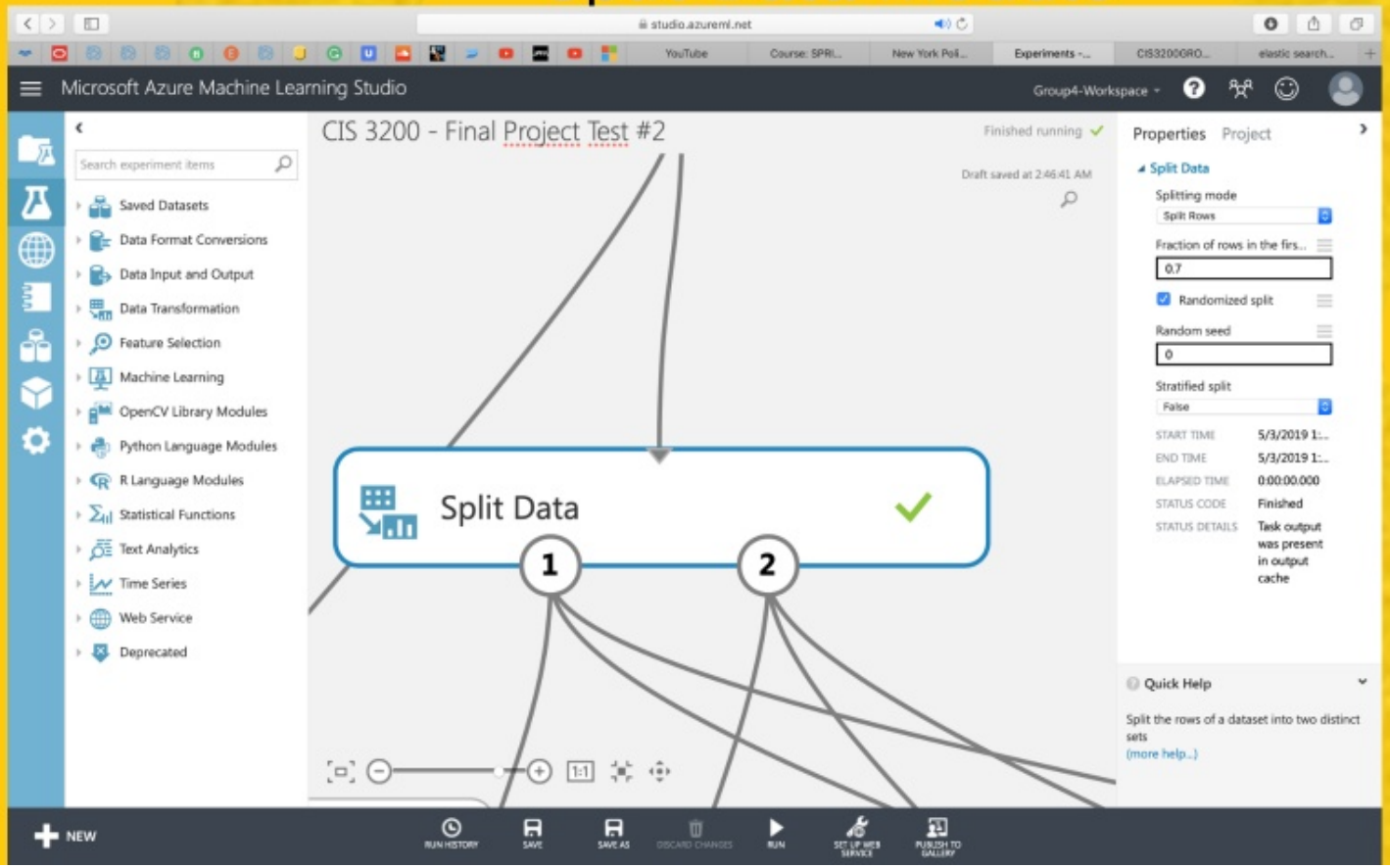
Step 4

Step 4
Select columns in dataset



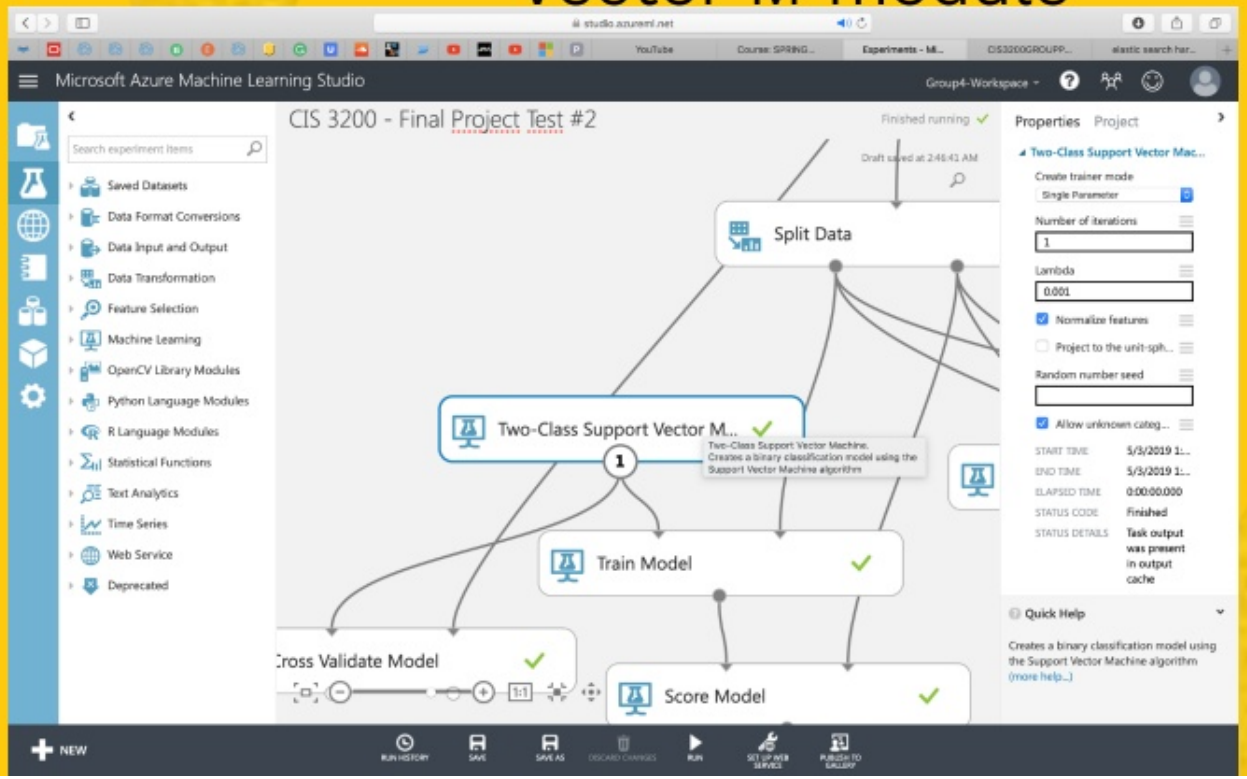
Step 5

Step 5
Split Data module



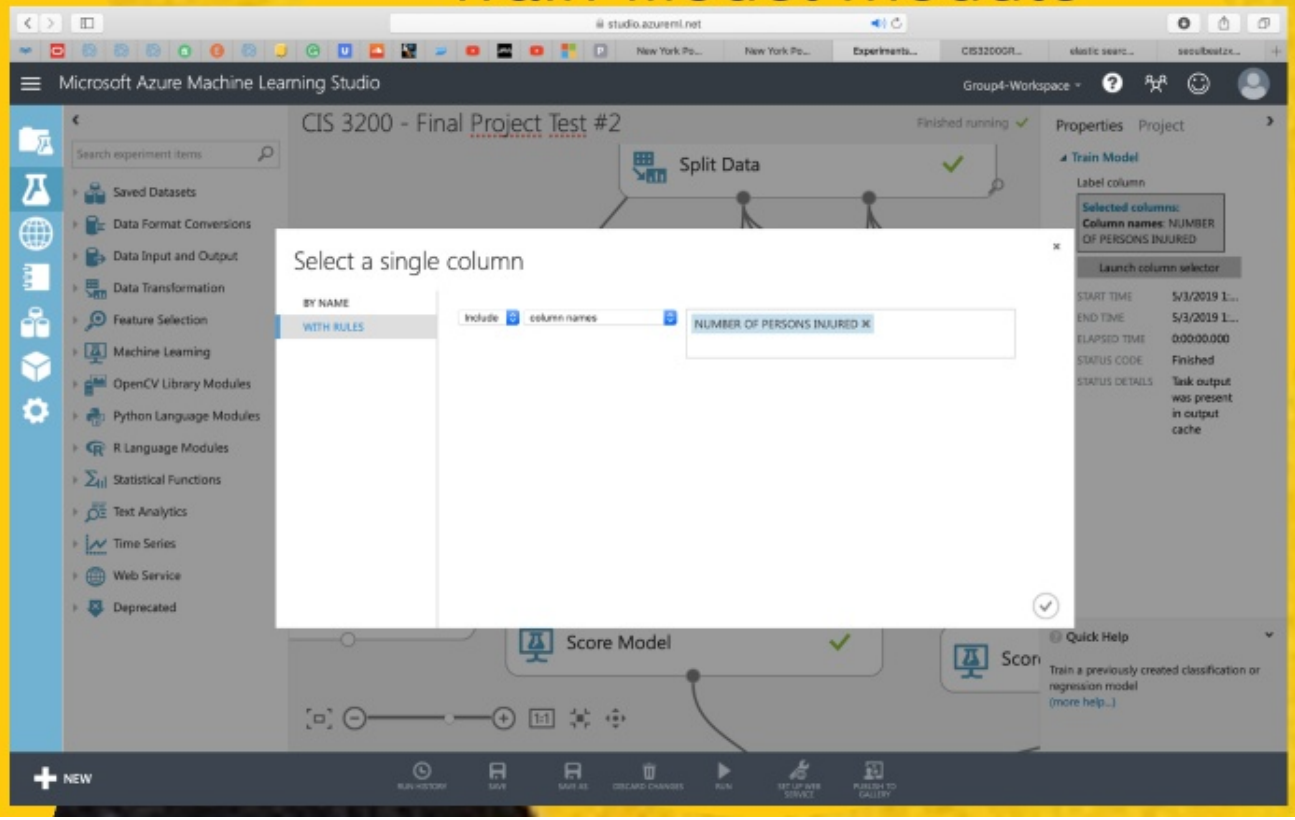
Step 6

Step 6
Two-Class Support Vector M module



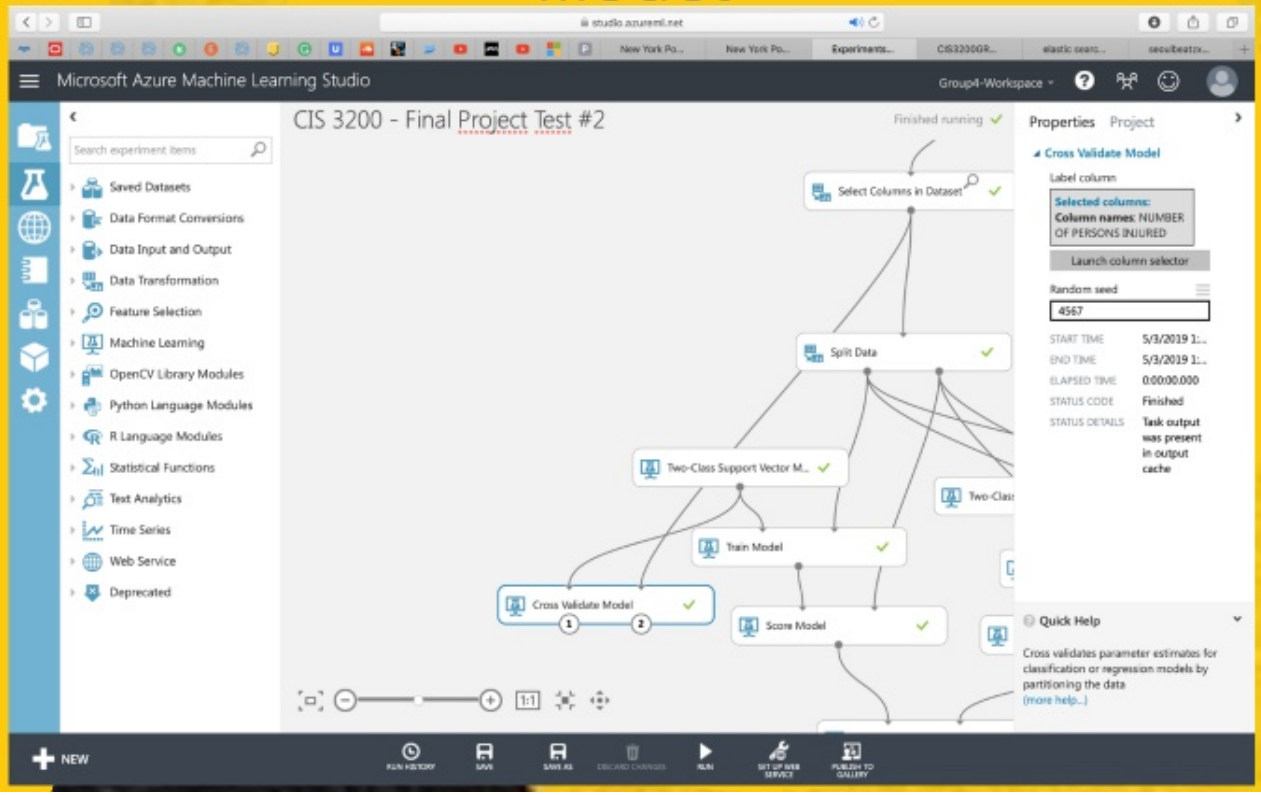
Step 7

Step 7
Train Model module



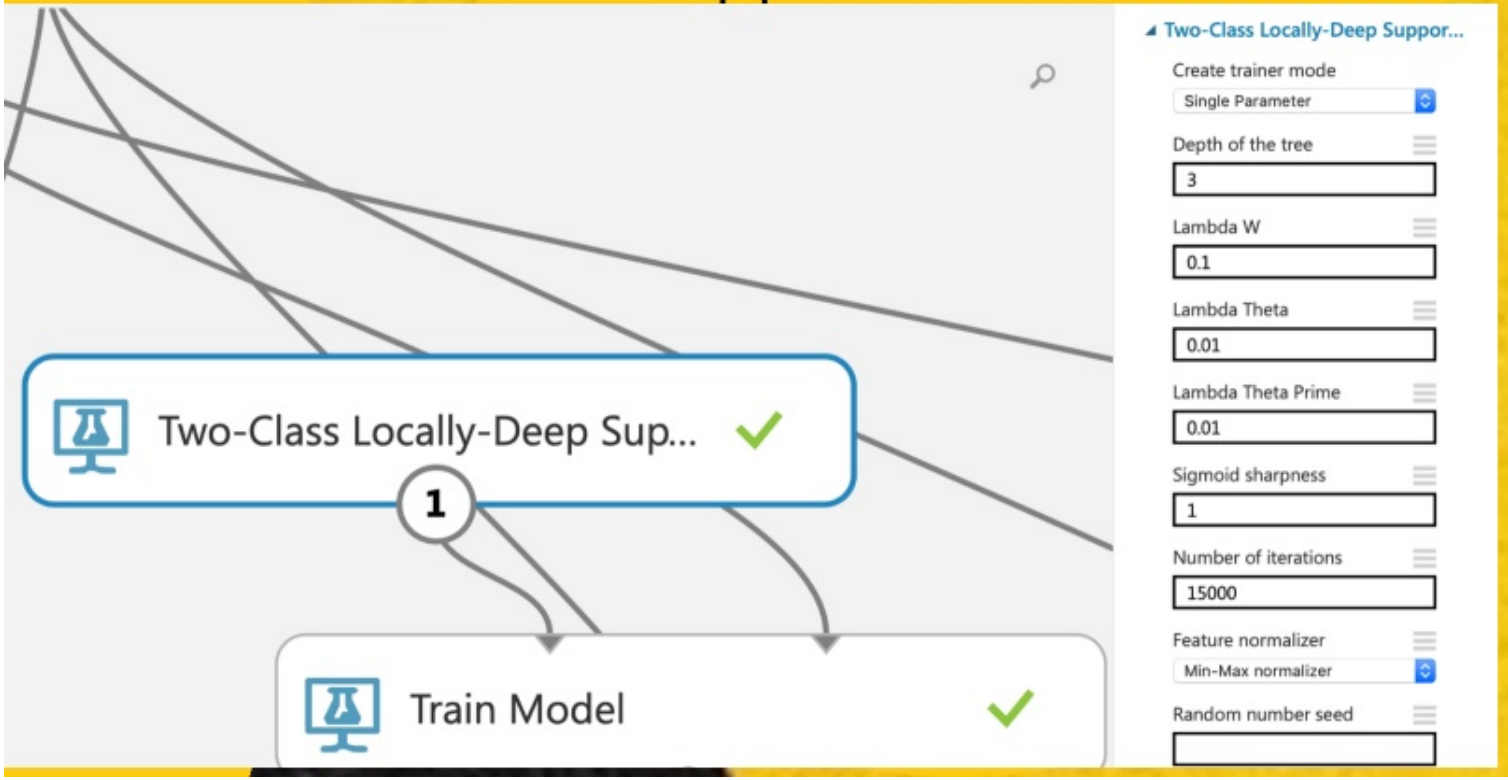
Step 8

Step 8
Cross Validate Model



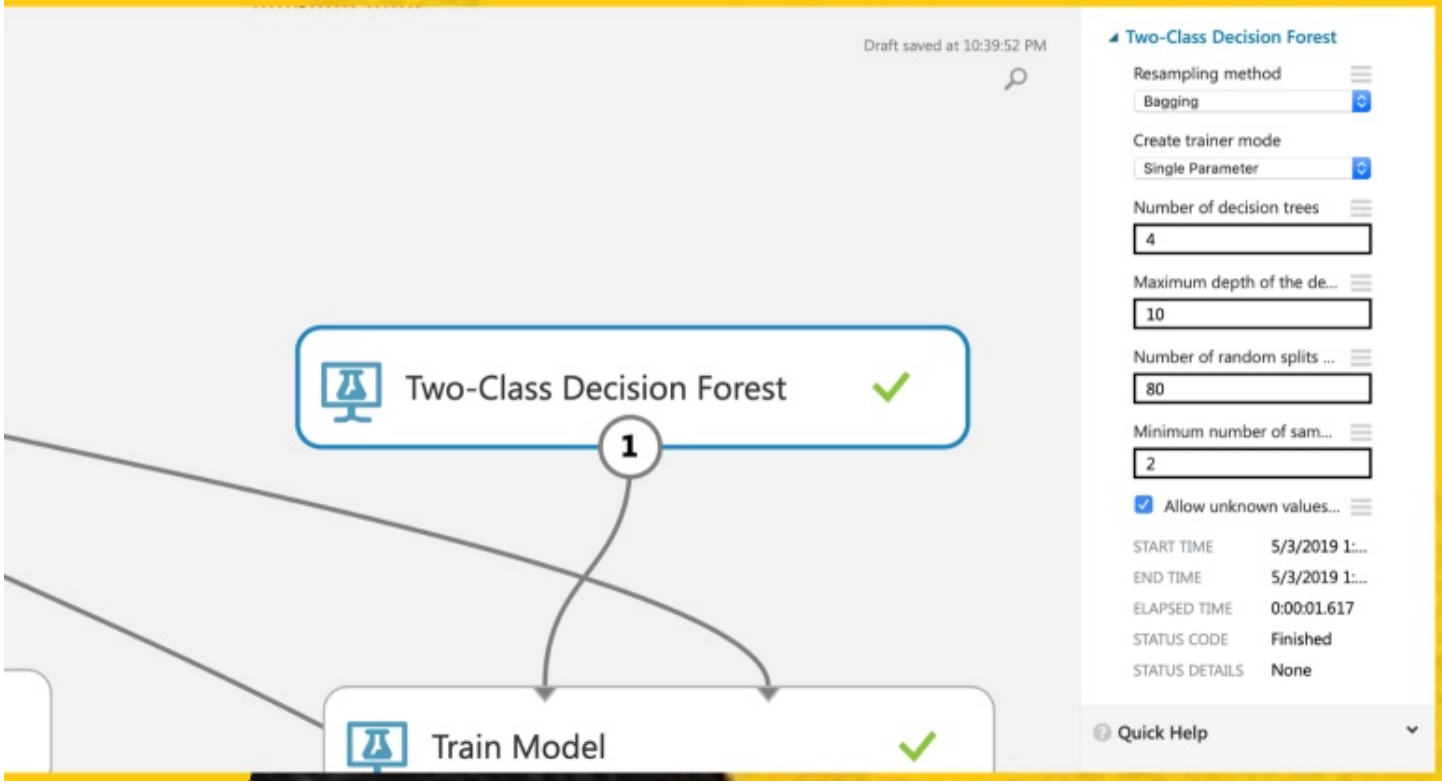
Step 9

Step 9
Two class locally deep support module



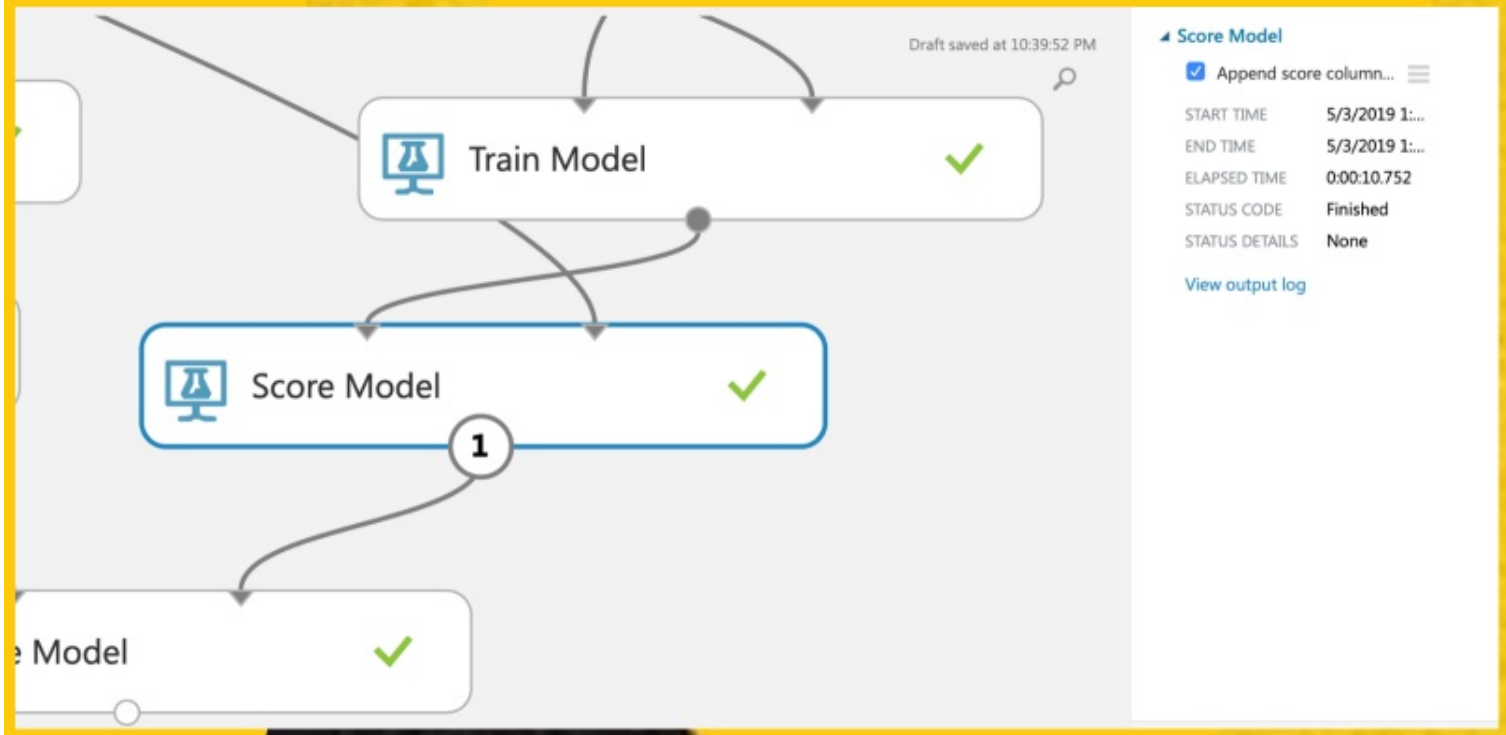
Step 10

Step 10
Two class Decision Forest module



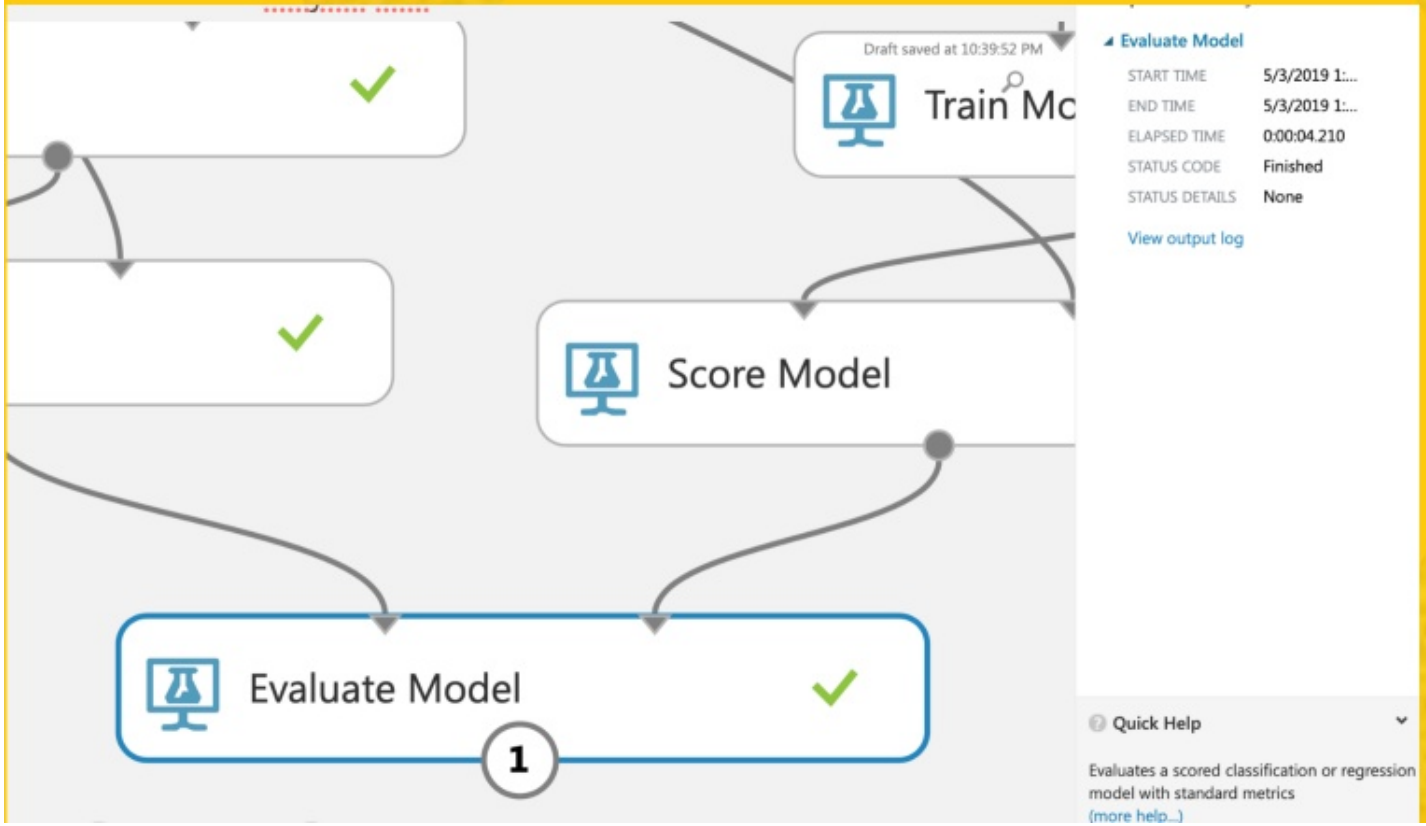
Step 11

Step 11
Score model module



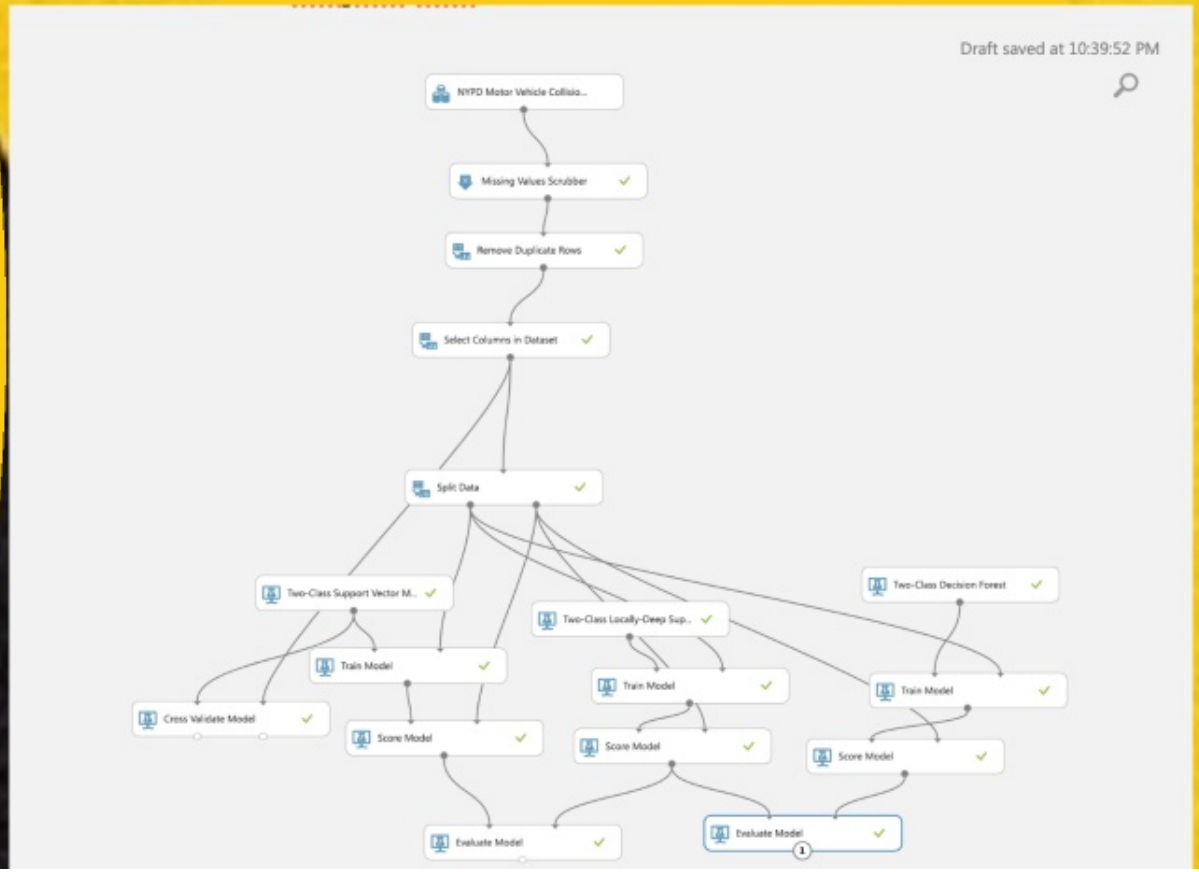
Step 12

Step 12
Evaluate Model module



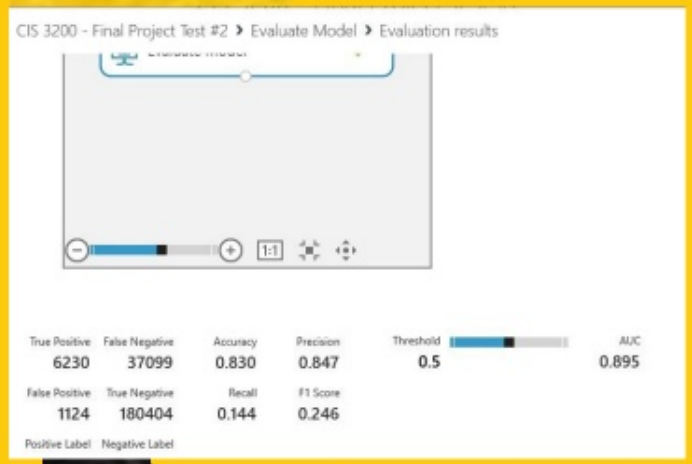
Overall view

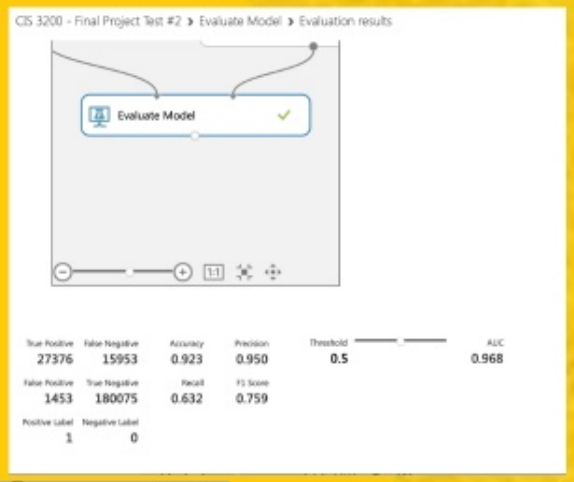
Overview of ML Azure model

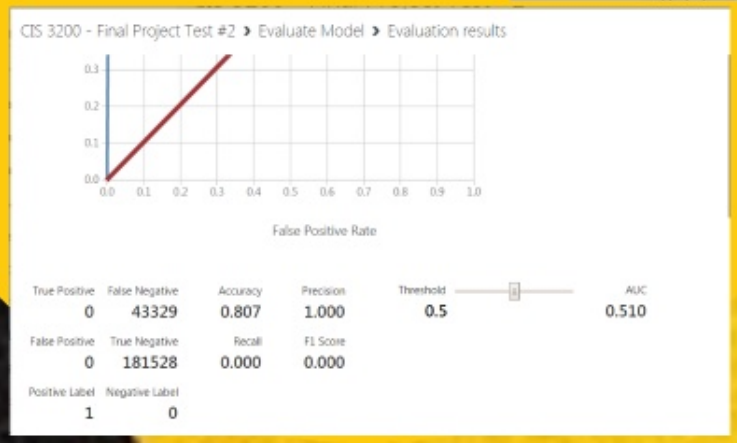


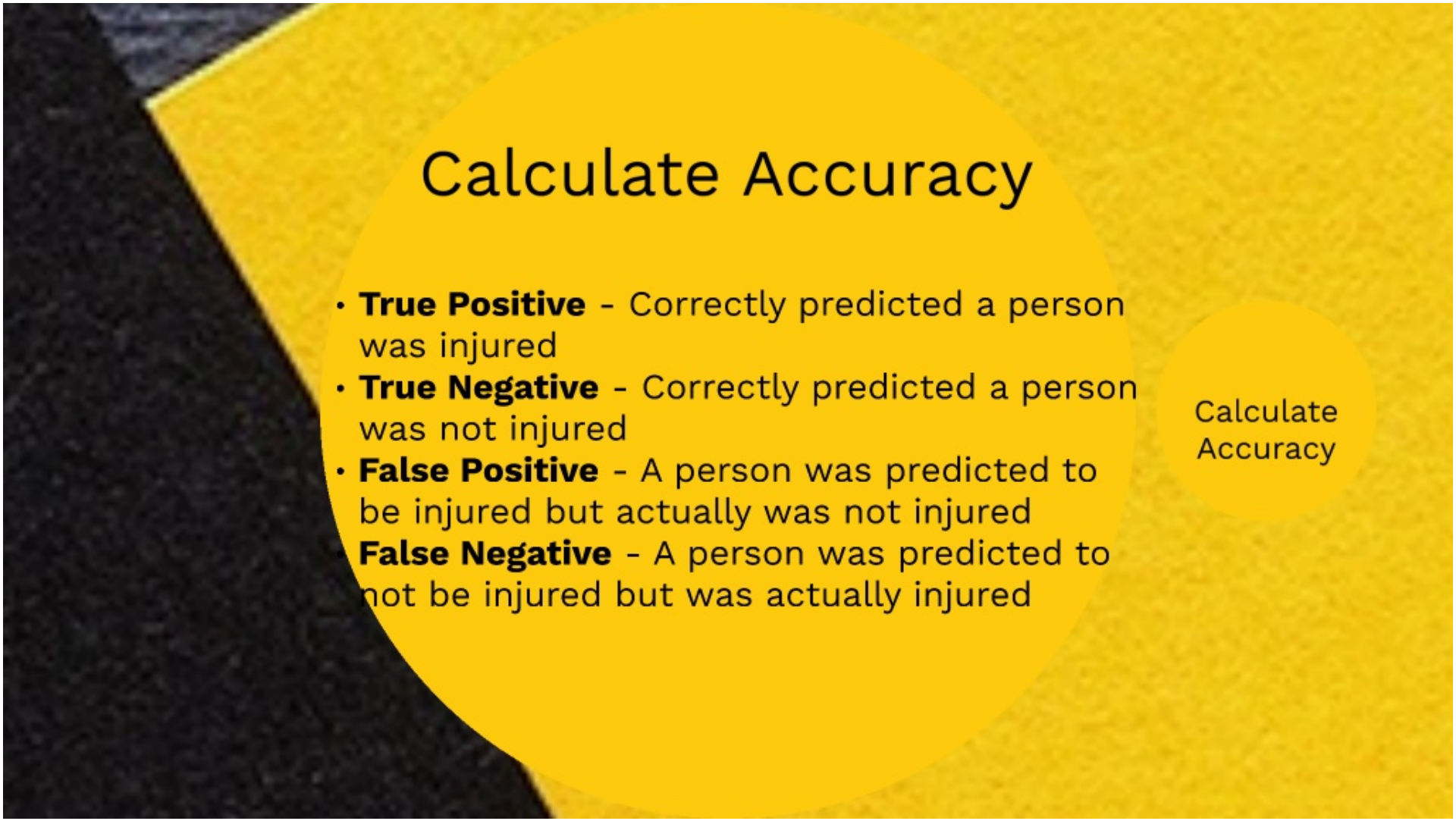
Calculate Accuracy

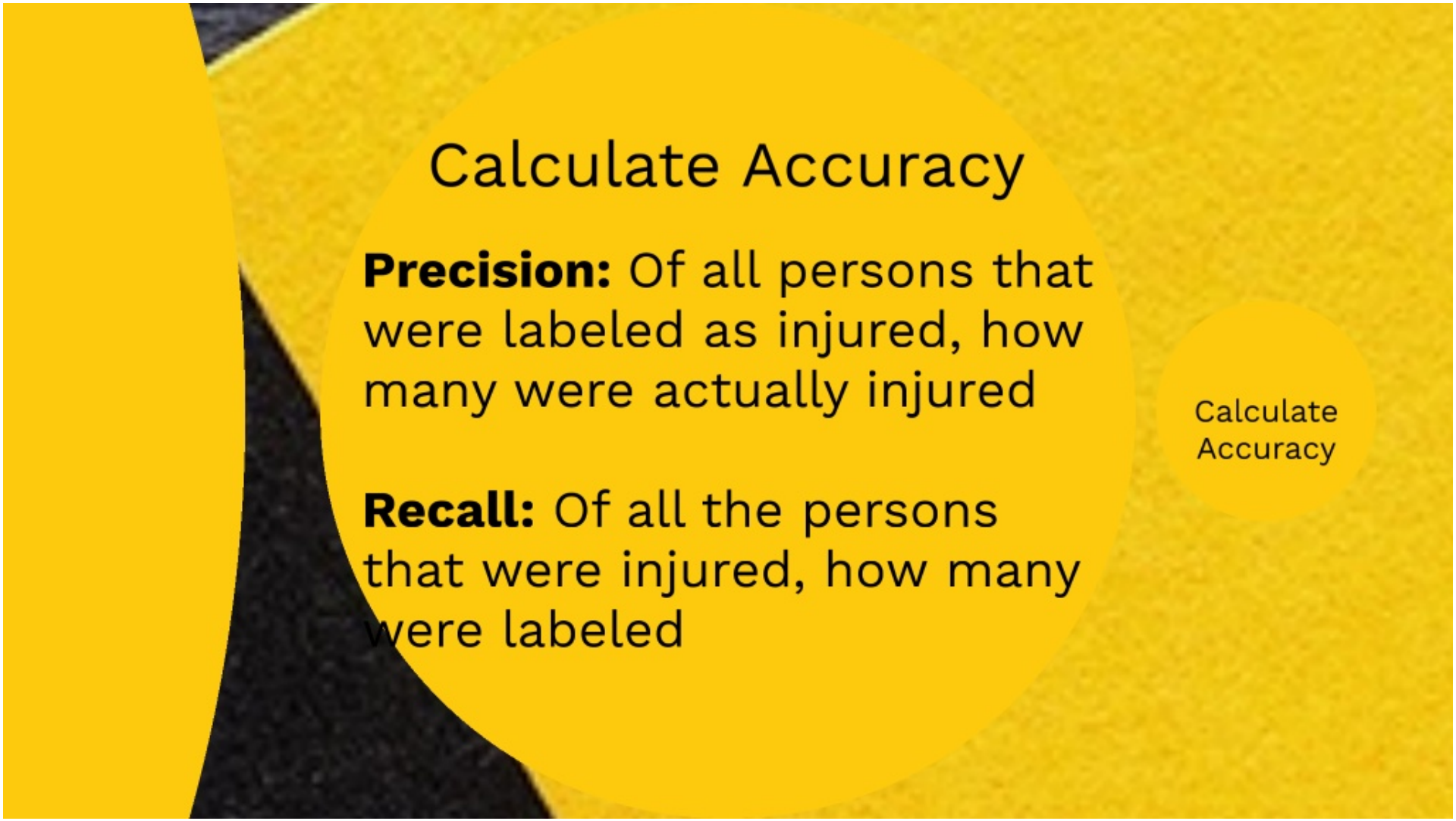
Calculate Accuracy



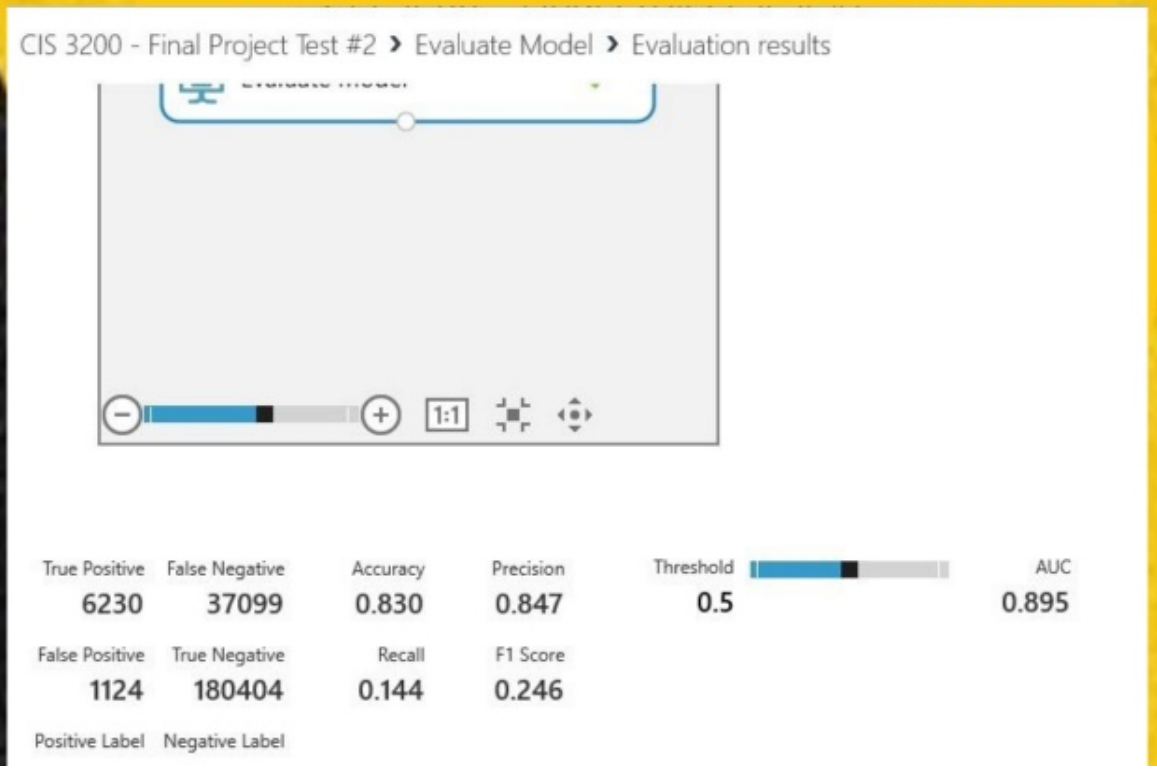





Calculate Accuracy
• True Positive - Correctly predicted a person was injured
• True Negative - Correctly predicted a person was not injured
• False Positive - A person was predicted to be injured but actually was not injured
• False Negative - A person was predicted to not be injured but was actually injured
Calculate Accuracy

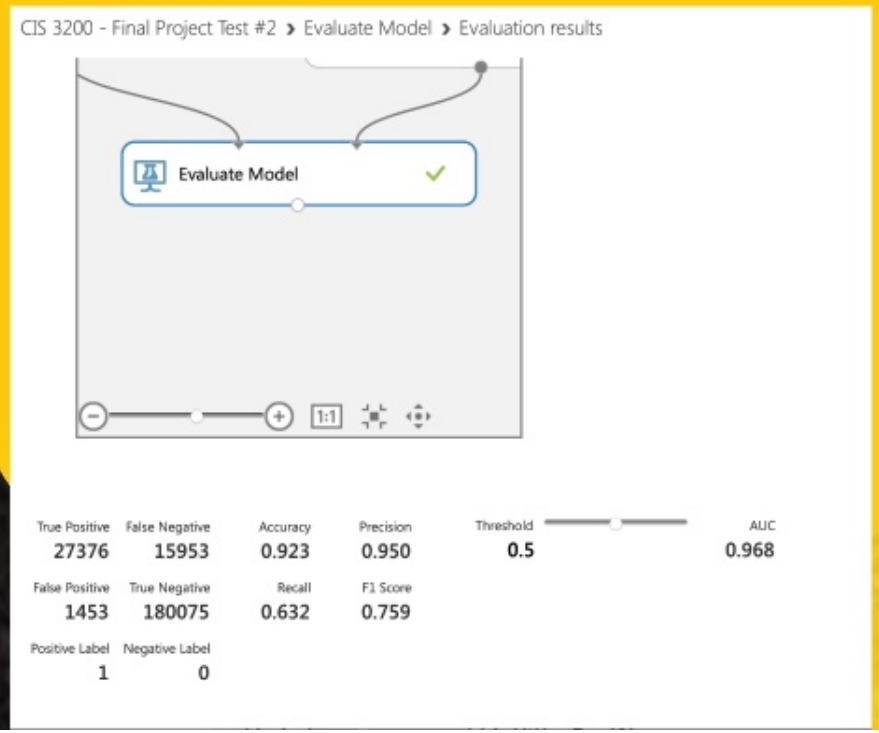

Calculate Accuracy
Precision: Of all persons that were labeled as injured, how many were actually injured
Recall: Of all the persons that were injured, how many were labeled
Calculate Accuracy

Two Class Support Vector Machine Model



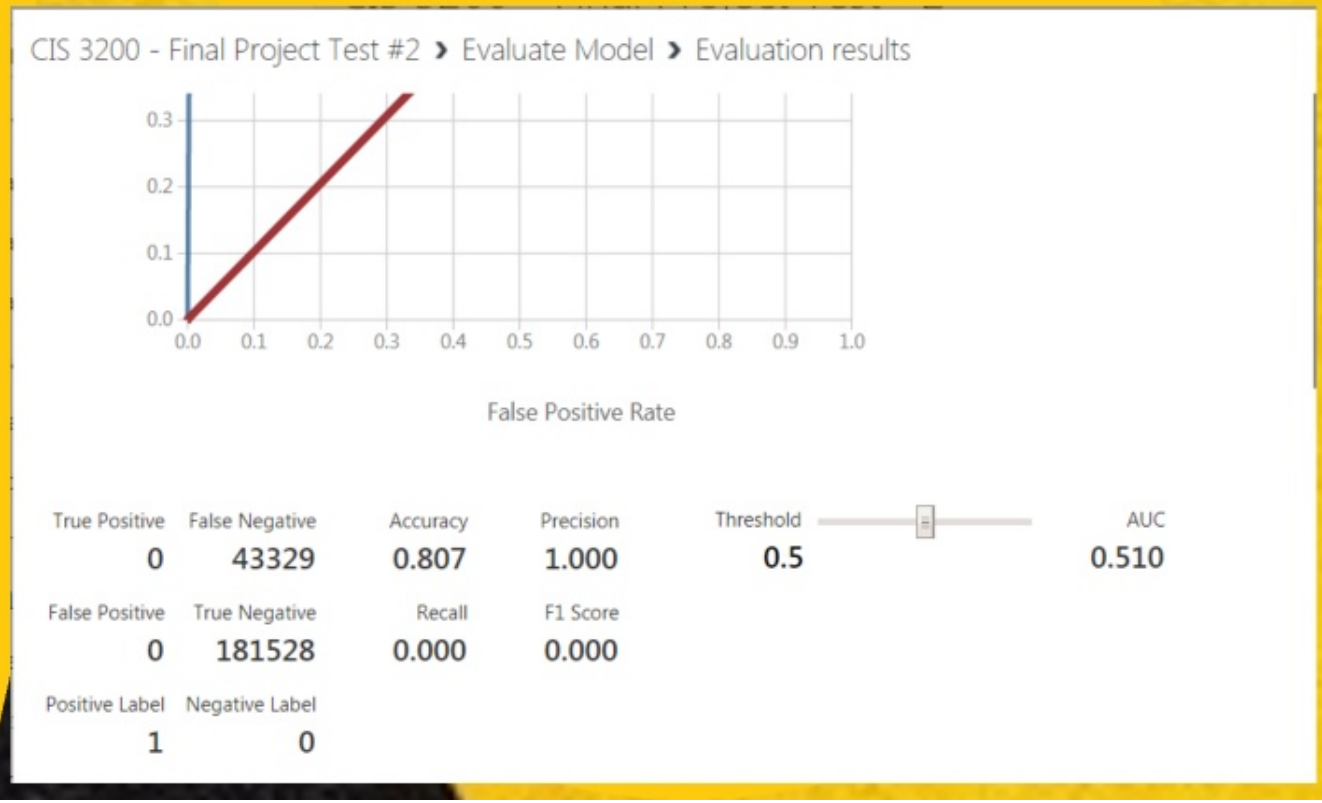
Two Class Locally-Deep Support Vector Machine Model

Two Class Locally-Deep Support Vector Machine Model



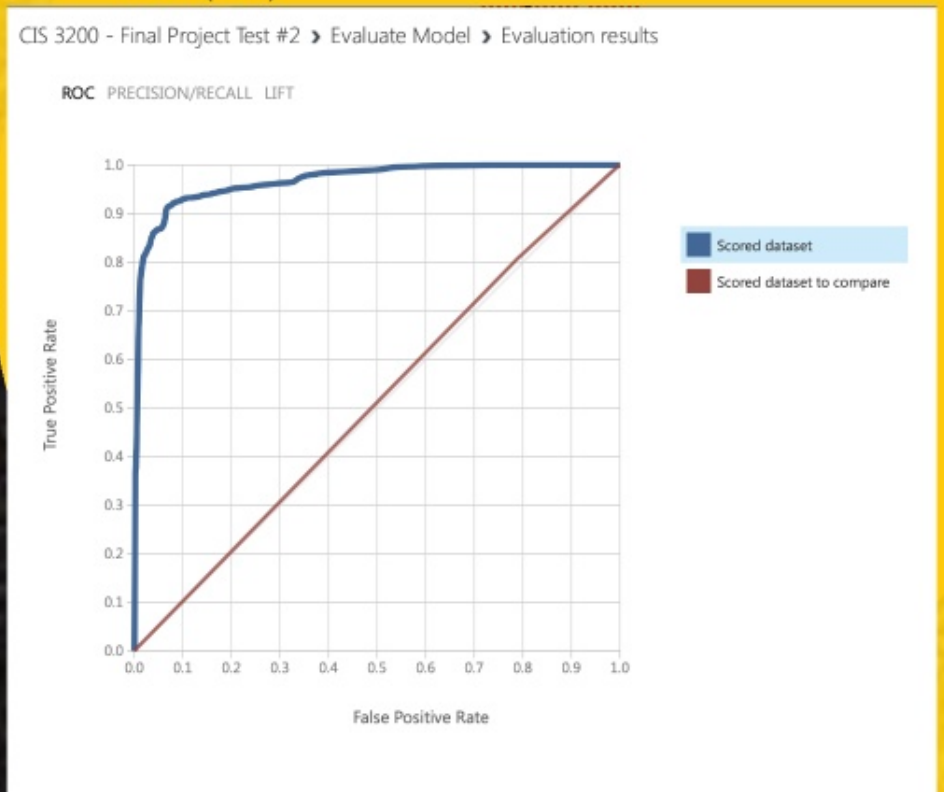
Two Class Decision Forest

Two Class Decision Forest



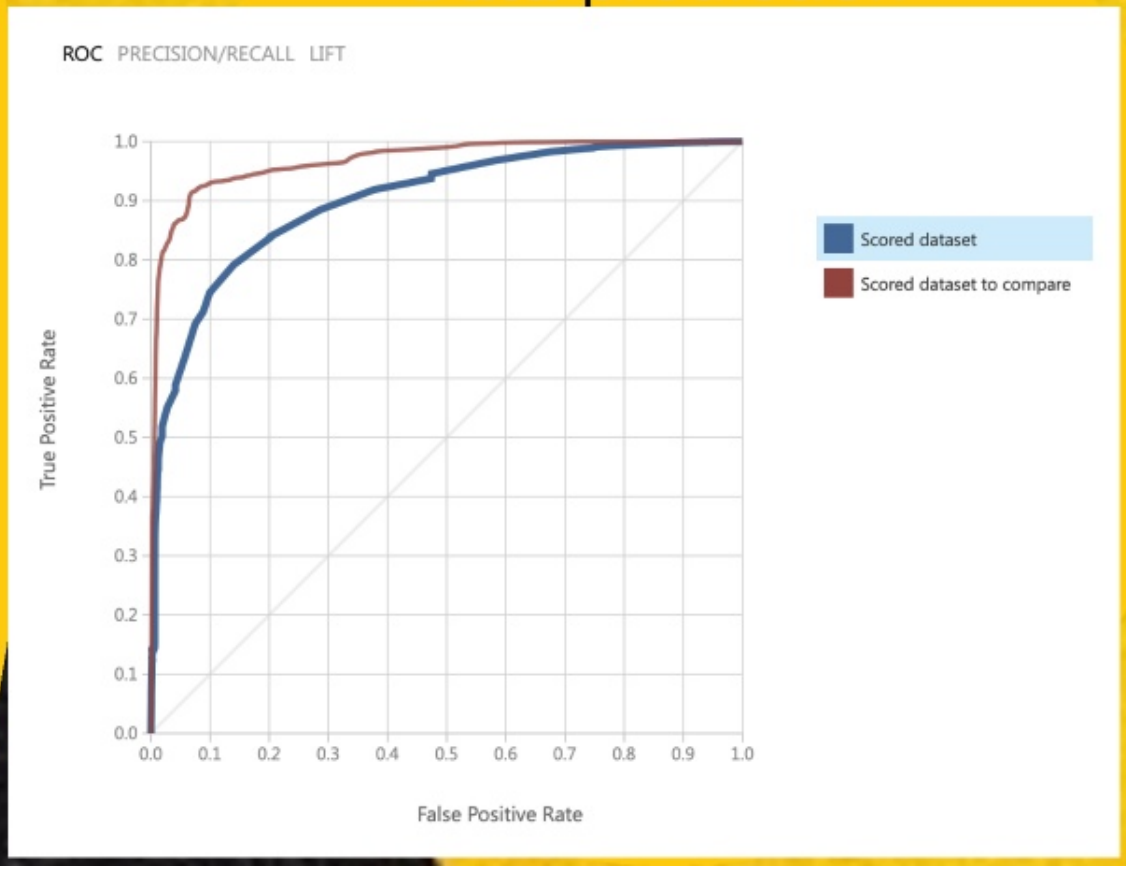
Predictive Result

We used classification as a result for better Area Under the Curve(AUC)

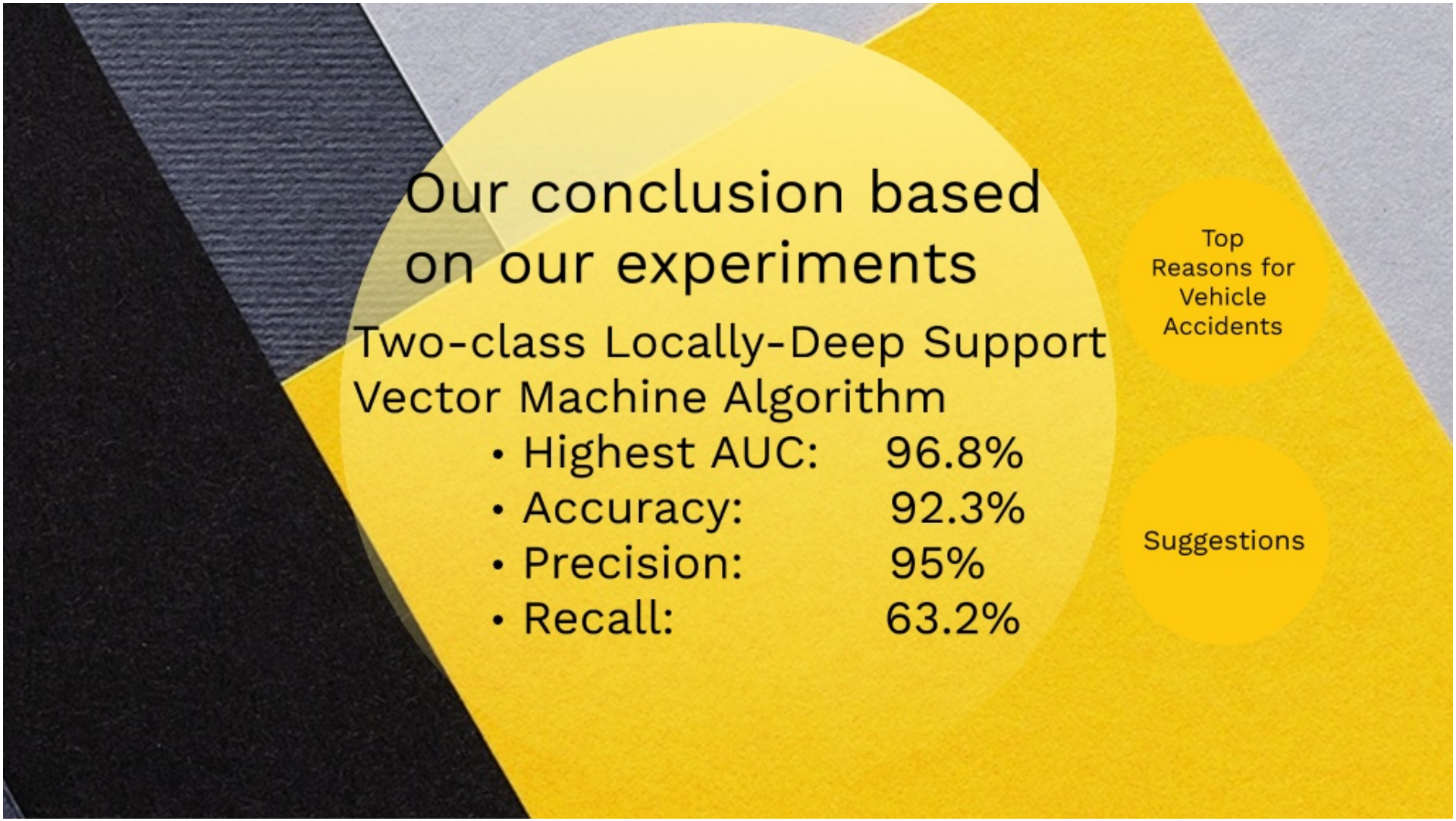


Actual Result

Actual Result of Data prediction




New York Police Department
Injury Report
Presented by
Lucky Trang, Ron Tsang
Stephanie Nguyen, Paul Yi
Intro
Body
CON
REC


Our conclusion based on our experiments
Two-class Locally-Deep Support Vector Machine Algorithm
• Highest AUC: 96.8%
• Accuracy: 92.3%
• Precision: 95%
• Recall: 63.2%
Top Reasons for Vehicle Accidents
Suggestions

Top Reasons for Injuries and accidents

1. Driver inattention / Distraction
2. Failure to yield right of way
3. Following too closely (tailgating)
4. Disregarding Traffic control

Areas to Avoid

1. Manhattan
    - Times Square
        - Tourists
        - Advertisement
            - Colorful billboards
2. Bronx
3. Brooklyn


New York Police Department
Injury Report
Presented by
Lucky Trang, Ron Tsang
Stephanie Nguyen, Paul Yi
Intro
Body
CON
REC

Our Recommendation

Our conclusion to personal injuries report is that we need to emphasize on how much we have to pay attention to our surroundings while driving or walking and possibly avoiding Manhattan, Bronx and Brooklyn red marked areas on a certain predicted time, location and date given by Kibana Geo Point analysis.

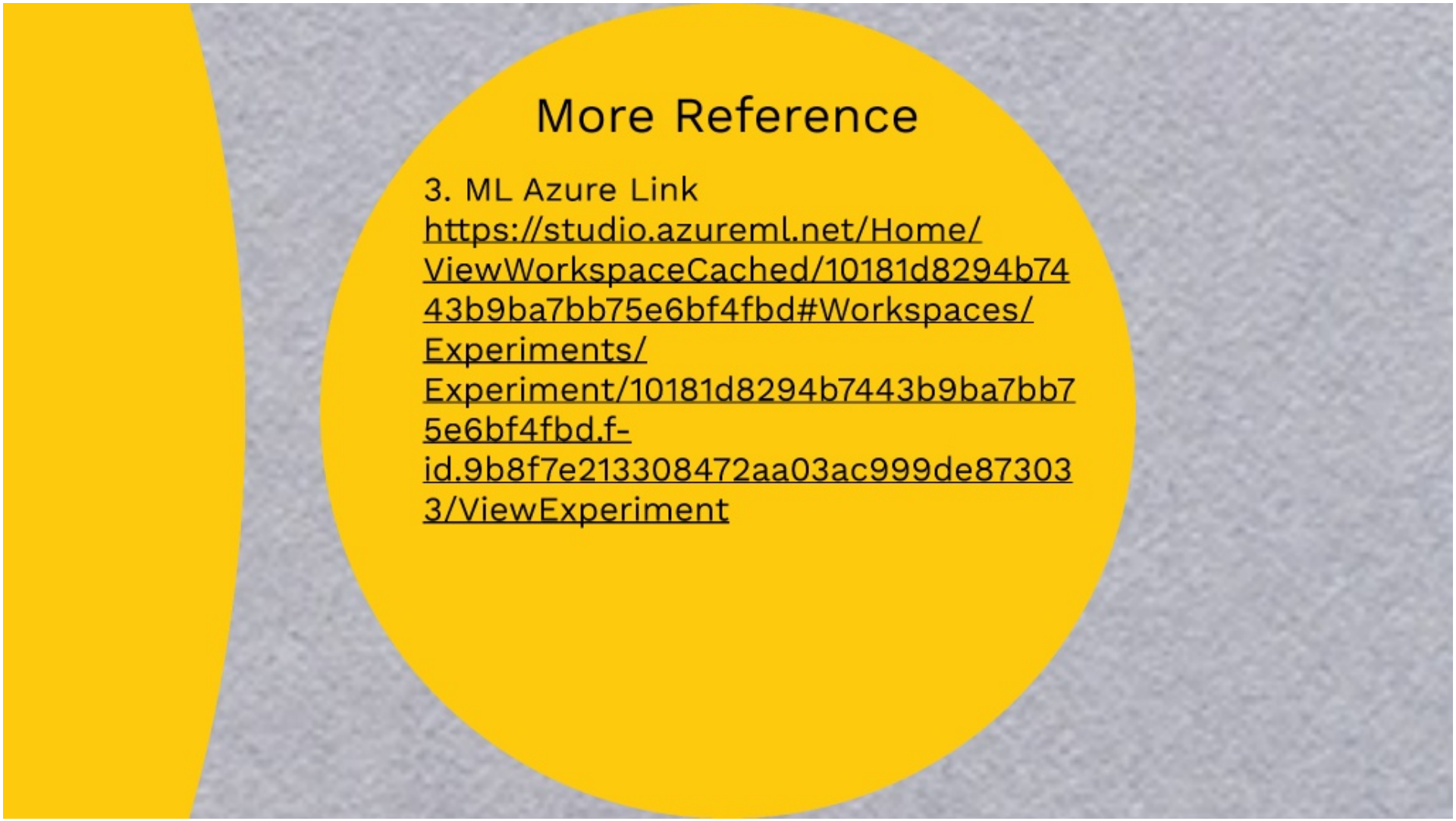
References

References

1. Data Source URL
https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-Collisions/h9gi-nx95

2. Git Hub URL

Continue


More Reference
3. ML Azure Link
https://studio.azureml.net/Home/ViewWorkspaceCached/10181d8294b7443b9ba7bb75e6bf4fbd#Workspaces/Experiments/Experiment/10181d8294b7443b9ba7bb75e6bf4fbd.f-id.9b8f7e213308472aa03ac999de873033/ViewExperiment


New York Police Department
Injury Report
Presented by
Lucky Trang, Ron Tsang
Stephanie Nguyen, Paul Yi
Intro
Body
CON
REC