



# EECS6414: Data Analytics & Visualization

## Sample Past Project Ideas



Fall 2017



# Group Pattern Discovery of Pedestrian Trajectories

## Project 1



# Pedestrian trajectories





# Network-based Analysis of Public Transportation Systems

## Project 2



# Transit Networks as Graphs

- Graph model result
- Example:
  - San Francisco





# Analysis of the DBLP Collaboration Network

## Project 3





Frank. Timmes  
Astronomy



Robert Rosne  
Astrophysics



Ewing L. Lusk  
parallel computing





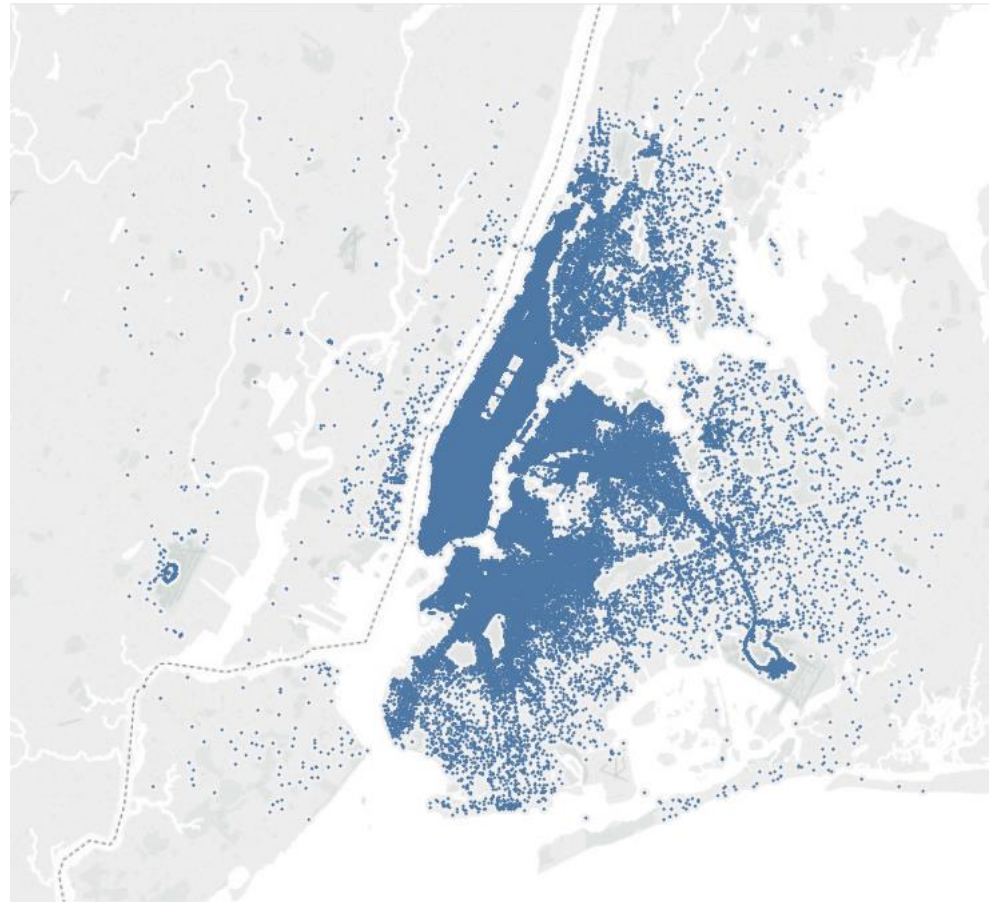
# Core Behavior Analysis of NYC Taxi Trips

## Project 4



# Motivation

- Taxi trip data analysis
  - Study individual taxi
  - Find local patterns
  - Find global patterns



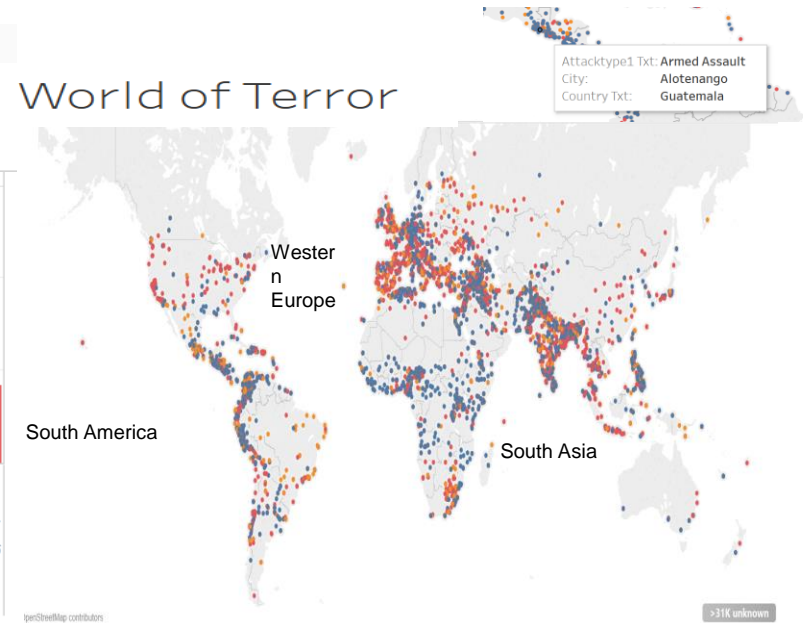
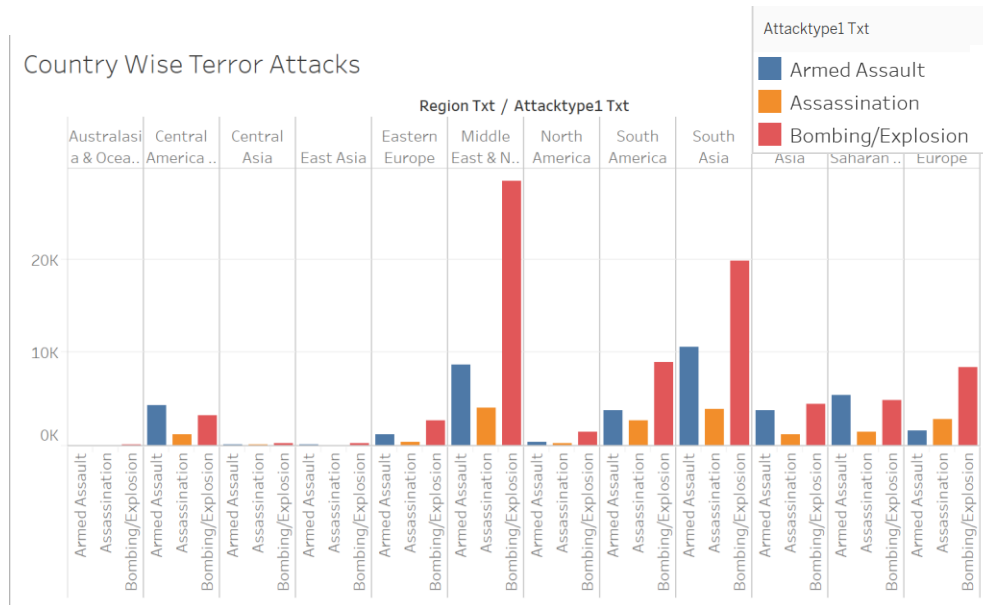


# Global Terrorism Data – An Investigative Visual Perspective

## Project 5



# Visualization trials and Empirical Observations





# Scene Classification

## Project 6



# Motivation



Parking lot  
Highway  
Beach  
Garage  
Street





# Itinerary Planning Using Genetic Algorithm

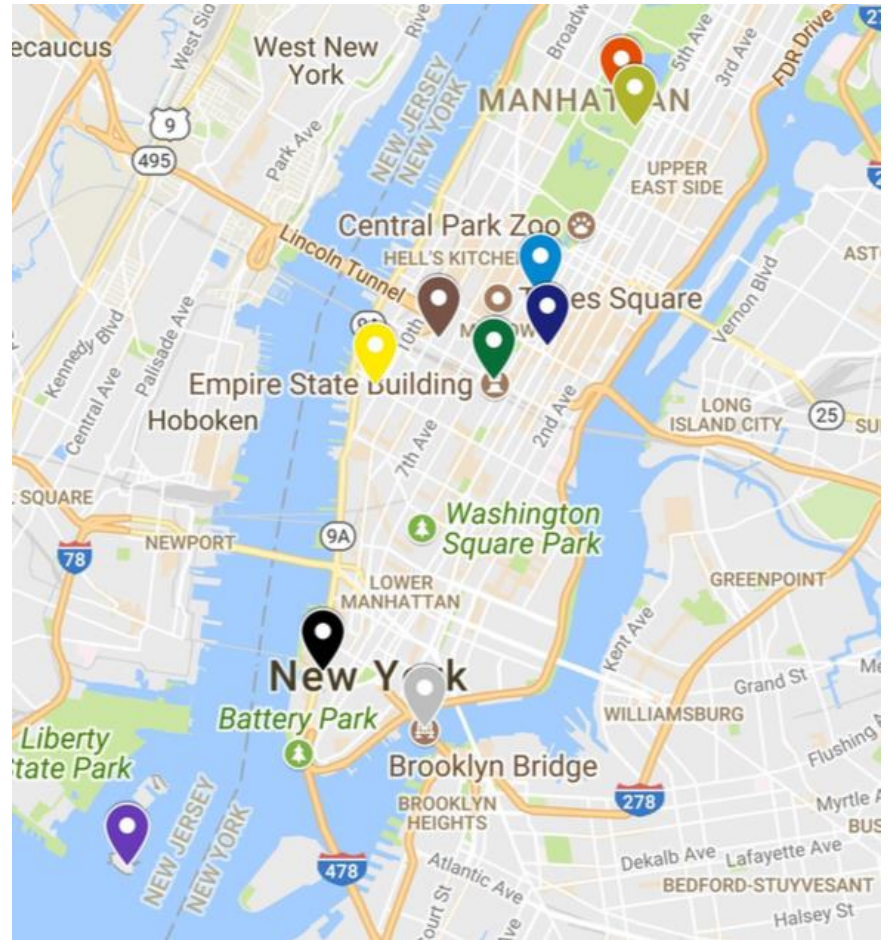
## Project 7



# Itinerary Planning Problem

## Sights

- Central Park
- Statue of Liberty National Monument
- The High Line
- Empire State Building
- Time Square
- Brooklyn Bridge
- Rockefeller Center
- The Metropolitan Museum of Art
- One World Trade Center
- Grand Central Terminal





# User Profiling on News Media

## Project 7



# Motivation

## THE GLOBE AND MAIL

CANADA'S NATIONAL NEWSPAPER • FRIDAY, AUGUST 5, 2016 • globeandmail.com

### REAL ESTATE

## Toronto price jump stirs fear offshore buyers will head east

With B.C.'s tax on foreign buyers taking hold and Toronto house sales smashing records, agents warn overheated market is magnet for speculation

TAMMIE McMAHON  
REAL ESTATE REPORTER

Toronto's housing market cheered out another month of soaring price gains in July, fueling concerns that it could be the next target of a tax on foreign investors.

Almost 10,000 resale homes changed hands in the Greater Toronto Area last month, the highest recorded number for the month of July, the Toronto Real Estate Board reported. Average resale prices jumped 6.6 per cent from the same period last year, to \$524,000.

Real estate agents say it's not too early to tell how much the market's recent rise reflects permanent demand for housing. There are concerns, however, about foreign buyers' home in the C.T.A. as the implementation of the new B.C. tax on foreign buyers in the province's largest city.

### INSIDE

## Quebec waiter arrested over salmon tartare served to man with allergy

Carney's strategy

The Bank of England governor kicks off an aggressive stimulus as confidence plunges

### REPORT ON BUSINESS

### Lakeside manners

David Fiddie explains how to be the kind of cottage guest who gets invited back every year

### GLOBE LIFE

### On the nuclear brink

The Trump threat comes at a critical time for the disarmament movement

### ELIZABETH BENNETT

### IN OPINION

Canada's government has struck a working group to issue recommendations on how to make Vancouver and Toronto's housing more affordable.

### FOOD

## Quebec waiter arrested over salmon tartare served to man with allergy

ANN HUI

In what appears to be the first case of its kind in Canada, Quebec police are recommending a criminal charge of negligent manslaughter against a waiter who served salmon to a man with a severe seafood allergy.

This week, police in Sherbrooke arrested the 22-year-old waiter after a customer went into anaphylactic shock and had to be hospitalized from eating at the restaurant. Experts say it is the first time police in Canada have pushed for criminal charges from such an incident — and with a possible precedent for restaurants' duty of care when it comes to dealing with food allergies.

"The server had almost killed me," Simon Pierre Camuel told The Globe and Mail in an emailed statement. "I know at least an error, but that error had almost taken my life. My boyfriend almost lost his boyfriend."

In late May, Mr. Camuel visited the "Bouquet" restaurant, a tapas bar on the west side of Sherbrooke. According to an online review, the restaurant is "a very

### RIO 2016

## Soccer-mad Brazil recruits world of talent to compete in other sports

STEPHAN LEE

RIO DE JANEIRO

When a country's national football team is competing in the last stages of the country's military dictatorship, they were young intellectuals who saw in future in the country, so they went to the United States where they built a new life and raised an all-American daughter. In 2016, she was graduating from Columbia University and on her way to being a doctor.

And then she had to break some news to her father: He was headed to medical school, but rather to Brazil. To play for the national rugby squad. "So mom said, 'You're doing what?'" his mother said. "They didn't think this was a great idea," he put it mildly.

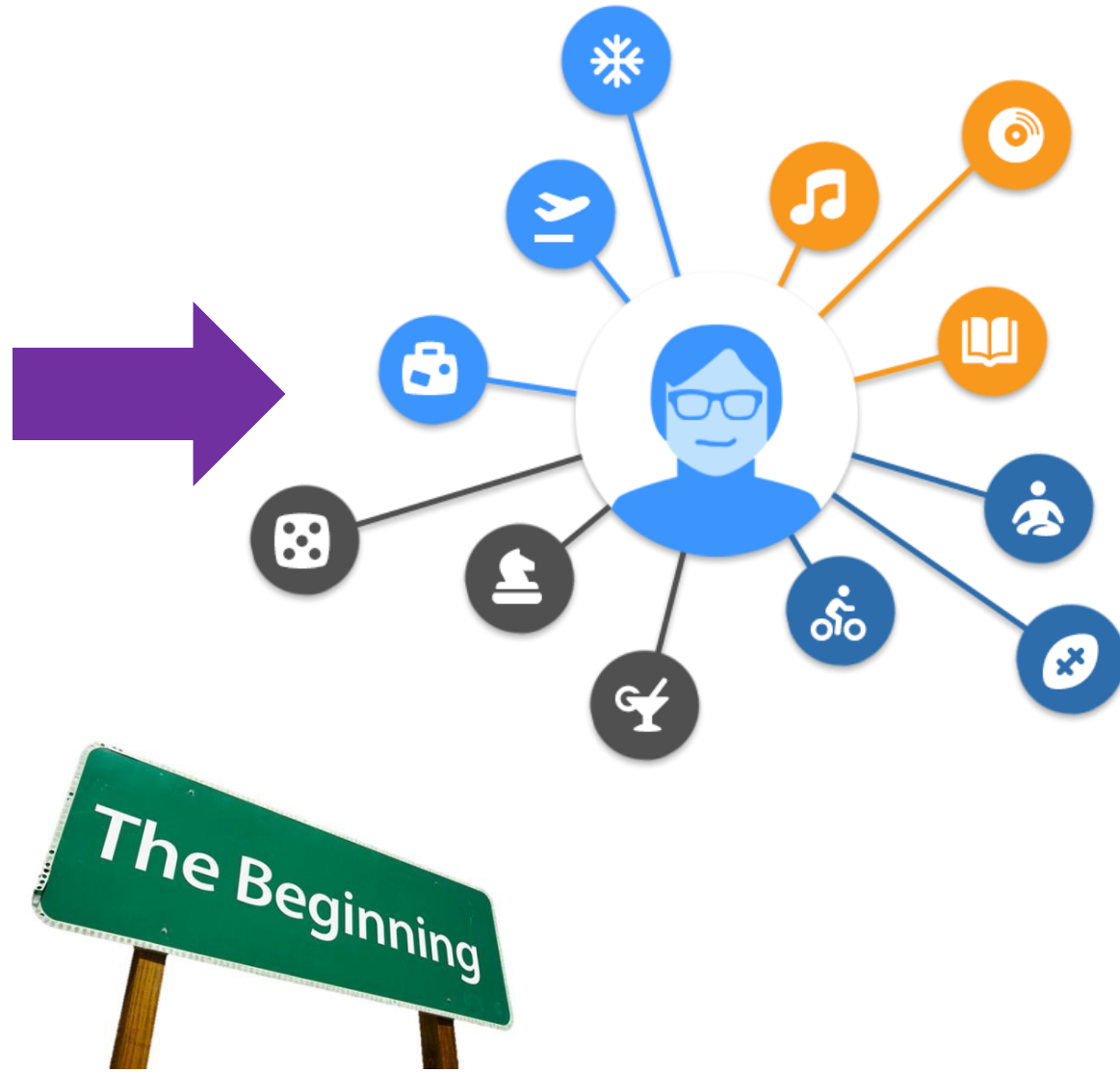
But father might, Mr. Camuel will march into the Maracanã Stadium as an Olympic athlete, part of Brazil's first Olympic rugby squad, and her parents, here to cheer her on, admit that Brazil is not the country they left, and much is possible here today.

As the host country, Brazil qualified automatically to compete in sports for which it otherwise would not have been eligible — and that created a scramble for athletes.

The athletes shortage reflects the fact that Brazil is not, at heart, an Olympic country. There is really only room for one sport, soccer, in this country's heart, with a bit of attention leftover for beach volleyball.

THE GLOBE AND MAIL

7532 0003 6





# Interactive Analysis and Visualization of Large-scale Transit Data

## Project 8

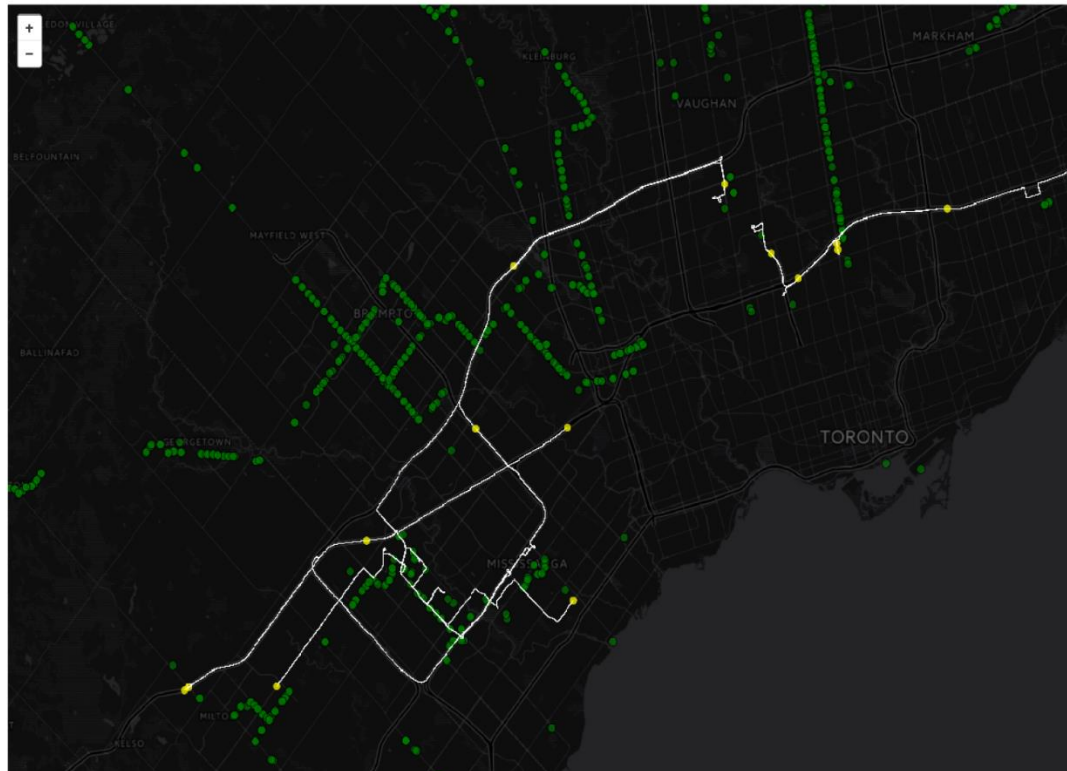


# Motivation

## Transit Data Visualization and Analytics

Animation **Velocity Heatmap** Between-Stops Relations Transit-Based Area Population

Animation



LOAD STOPS

HIDE BUSES LIST

BUS ID	STATUS
8133	<input checked="" type="checkbox"/> Active
2353	<input checked="" type="checkbox"/> Active
2383	<input checked="" type="checkbox"/> Active
8107	<input checked="" type="checkbox"/> Active
2260	<input type="checkbox"/> Active
2428	<input type="checkbox"/> Active
2524	<input type="checkbox"/> Active

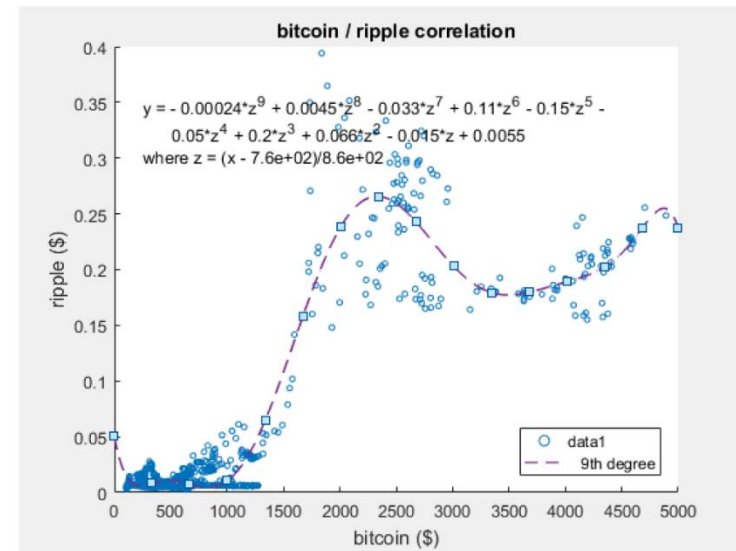
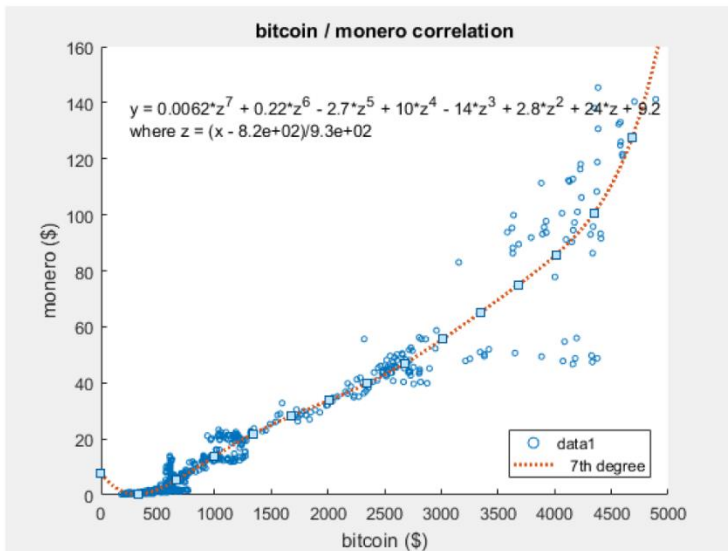
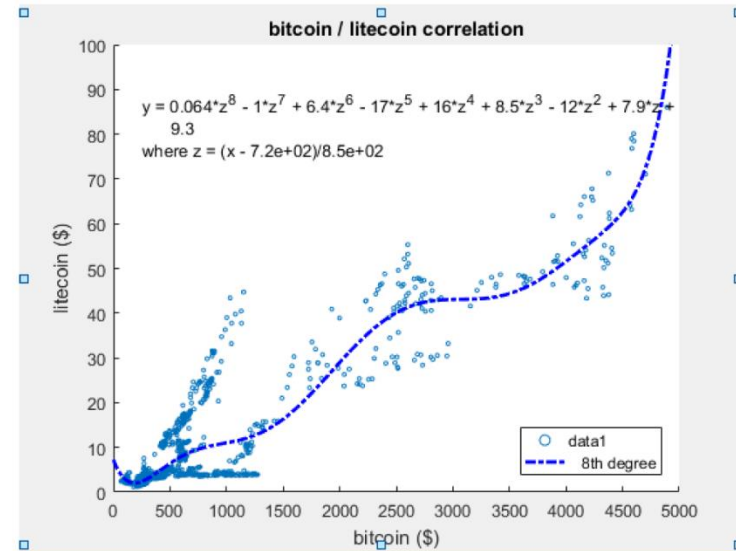
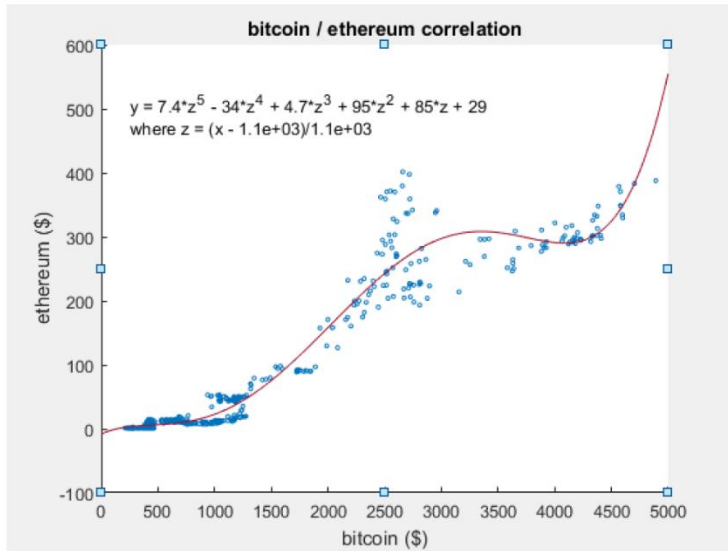


# Modeling and Forecasting Cryptocurrency

## Project 9



# Bitcoin & Other Cryptocurrencies





# CDiscount Image Classification Challenge

Project 10



# Motivation

1<sup>st</sup> - \$ 20,000  
2<sup>nd</sup> - \$ 10,000  
3<sup>rd</sup> - \$ 5,000

OR

Automatically classifying  
incoming products based on  
visual information





Winter 2019



# Kaggle Kernel Mining

## Project 1



# GitHub vs. Kaggle

## GitHub



- For software developers in general
- Create, fork, pull ... on repositories
- Get involved in large OSS projects

## kaggle

- Mostly for data scientists
- Create, fork ... on kernels
- Take part in competitions



# Generating Tactile Graphs in MsExcel

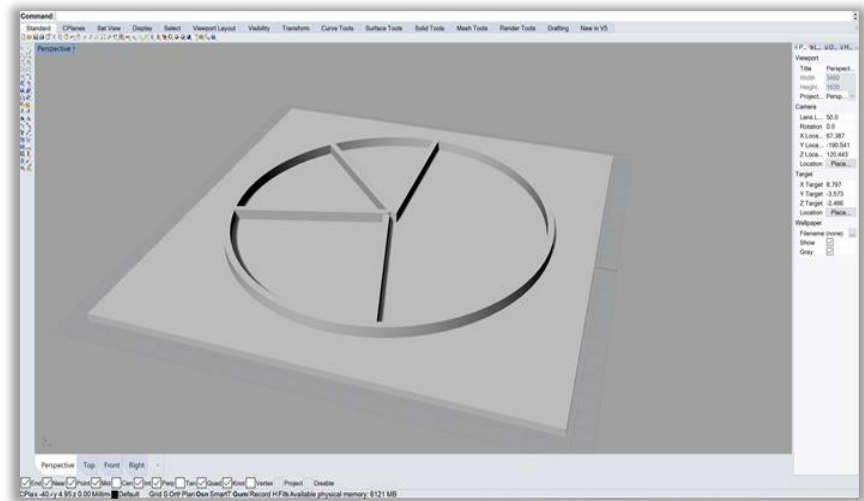
## Project 2



# Methods

A macro for Microsoft Excel to automatically export data to an OpenSCAD program, which would generate and print a tactile visualization of the data

```
1 // Resolution settings
2 $fs = 1; // Don't generate larger facets than 1 mm
3 $fa = 5;
4
5 // Scaling constants
6 mm = 1; // The default scale is millimeters
7 cm = 10; // millimeters per centimeter
8 inches = 25.4; // millimeters per inch
9
10 LineThickness = 3*mm;
11 LineHeight = 3*mm;
12
13 // include(PieChart.dat)
14 // PieChart data file will set the following variables and provide data
  to graph
15 PaperWidth = 7.5*inches;
16 PaperHeight = 11*inches;
17 PaperThickness = 3*mm;
18
19 module smooth_paper () {
20   translate ([0, 0, PaperThickness/2]) {
21     cube ([PaperWidth, PaperHeight, PaperThickness], center = true);
22   }
23 }
24
25 module pie_outline () {
26   translate ([0, 0, {PaperThickness*0.1}]) {
27     difference() {
28       cylinder(d = PaperWidth-(2*inches), h = LineHeight);
29       cylinder(d = PaperWidth-(2*inches)-LineThickness, h = {LineHeight*0.1});
30     }
31   }
32 }
33
34 smooth_paper ();
35 pie_outline ();
```



A code that converts data into a 3D printable file

Picture of the pie chart, ready for 3D-printing.



# Classification of Brain-wave Data for a Virtual Reality Application

## Project 3



# Methods



## PROBLEM DEFINITION

---

VALIDATING THE  
PRECISION RATES OF  
THE MEDITATION LEVELS  
ACHIEVED FROM MUSE  
BY COMPARING THEM  
WITH OUR  
DETERMINISTIC MODEL

THIS IS A  
CLASSIFICATION  
PROBLEM!!



# UK Housing Prices Analysis and Prediction

## Project 4



# Results

## Forecast for top 10 major cities from 2019 to 2020

City	December 2020	Forecast (2019-2020)
London	£882732	12.98%
Bristol	£306110	9.10%
Manchester	£194135	7.61%
Southampton	£228141	6.67%
Birmingham	£200505	6.26%
Leeds	£195442	6.13%
Sheffield	£173098	4.26%
Nottingham	£150079	3.91%
Liverpool	£137795	1.91%
Sunderland	£116784	1.83%



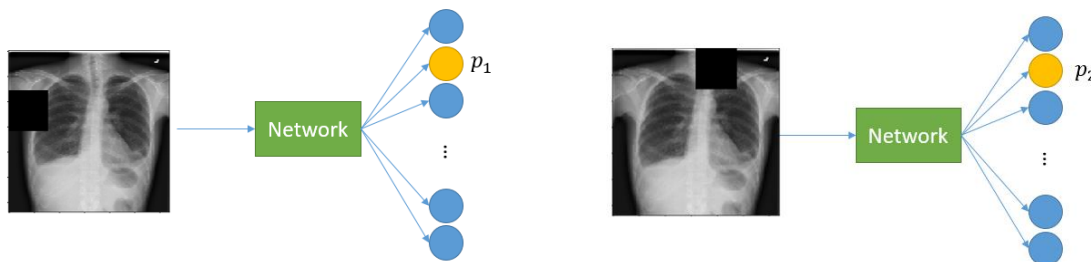
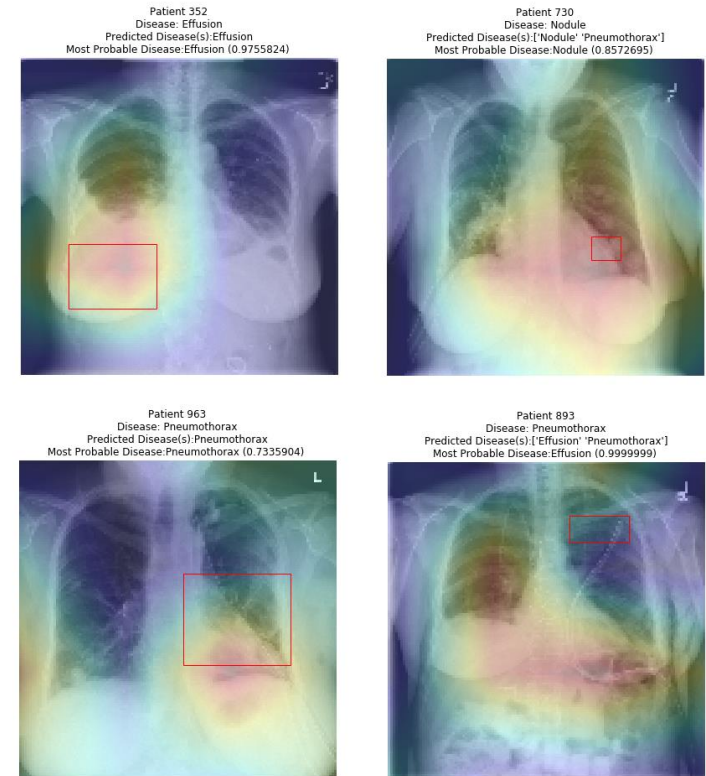
# Visualizing Deep Neural Networks Decision for Medical Imaging

## Project 5



# Results – Prediction Difference Analysis

- In this method, we occlude different parts of image and then track the changes in the output probability of the network
- By doing so, we can generate heatmaps of the parts of the image which are contributing for or against a class





# Canadian Real Estate Rental market - Artificial Hype or Reality? Airbnb vs Traditional Rental Market

## Project 6



# Airbnb Demand Analysis (2015 to 2019)

D3 radial column graph. It shows the percentage of listings that become unavailable in Toronto per month, through 2015 to 2019 with spikes from May to September at Summer times.



AirBnB Monthly Demand and Price Variation



# Go-transit Data Analysis and Visualization

## Project 7



# Motivation



GO Transit serving Greater Golden Horseshoe carrying 70 million passengers over the year

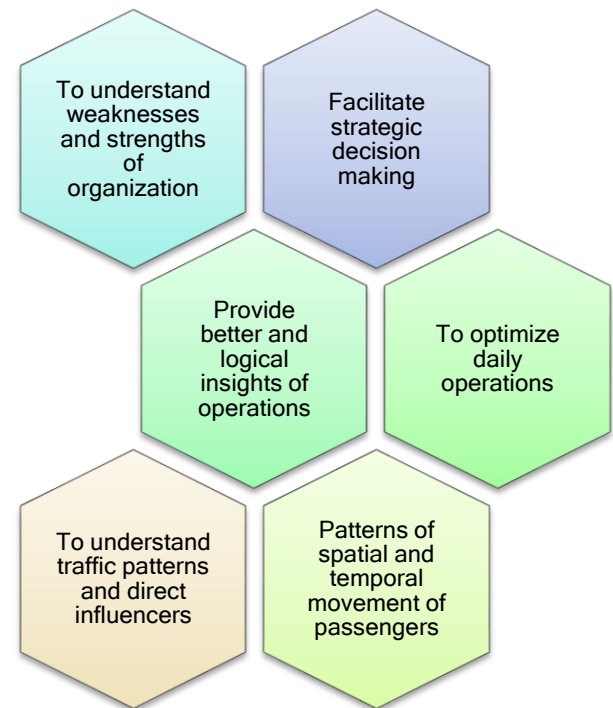
Why GO Transit Data analysis?



Scenario 1: Jam Packed Buses



Scenario 2: Empty Buses





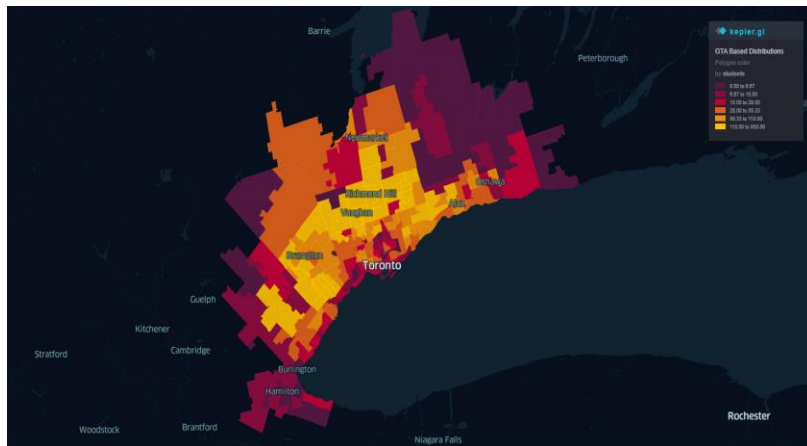
# York Community: Demographics, Analysis and Housing Location Recommendation for New Students

Project 9

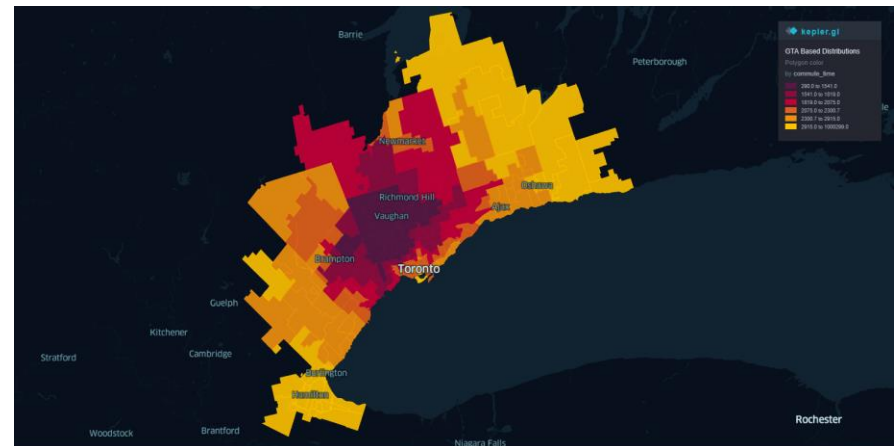


# Results

## Distribution of YorkU Students



## Commuting time needed for each Region



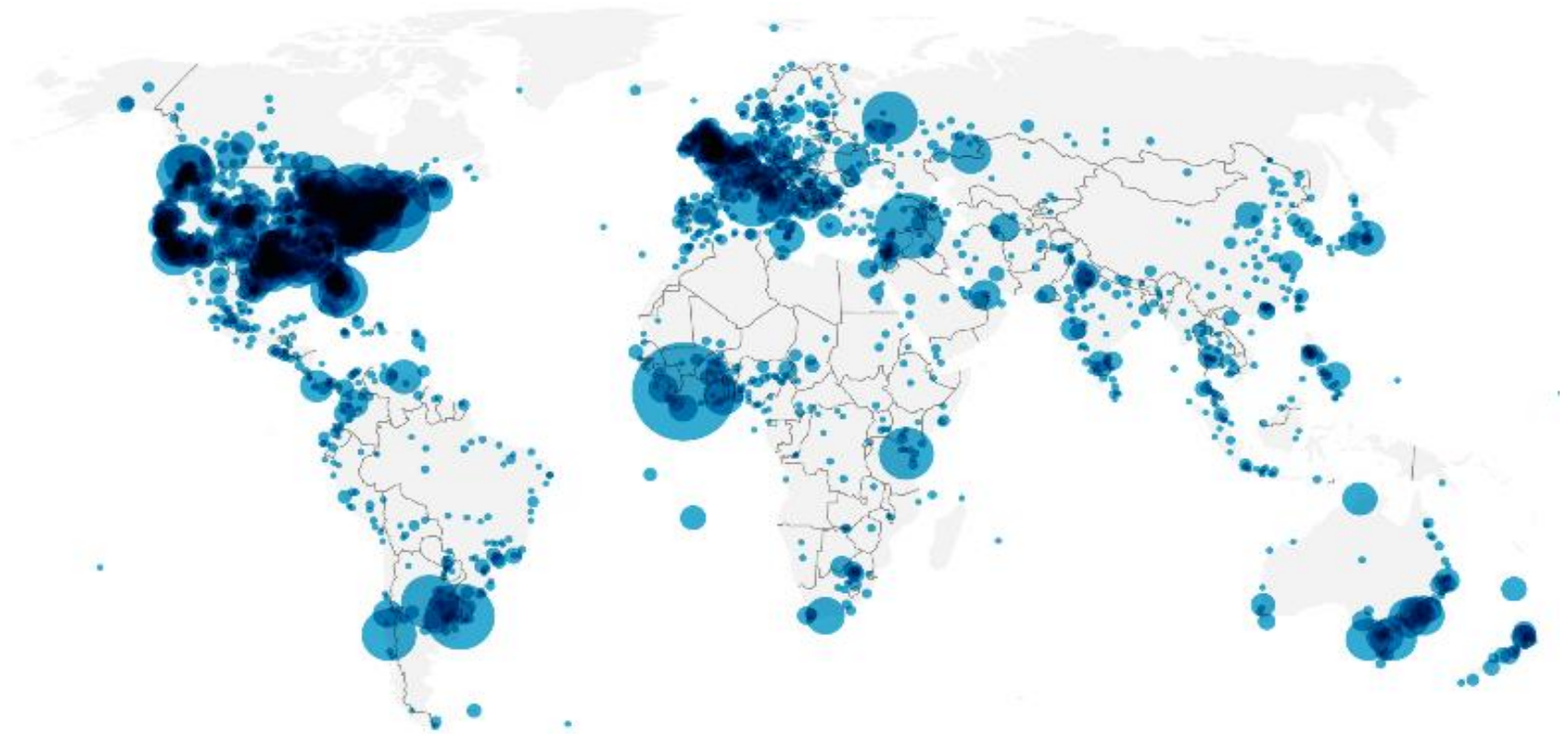


# Modeling and Visualization of User Interest in Reddit

Project 10



# Result & Evaluation (Cont'd)



**Overall distribution of geographical location in all the submissions**



# Mining Opportunities in an Online Shopping Store

Project 11



# Mining Online Shopping Store



Introduction & Research Questions



Data Sources and Cleansing



Speed and Security



Recommendation System



Next Steps





Winter 2020

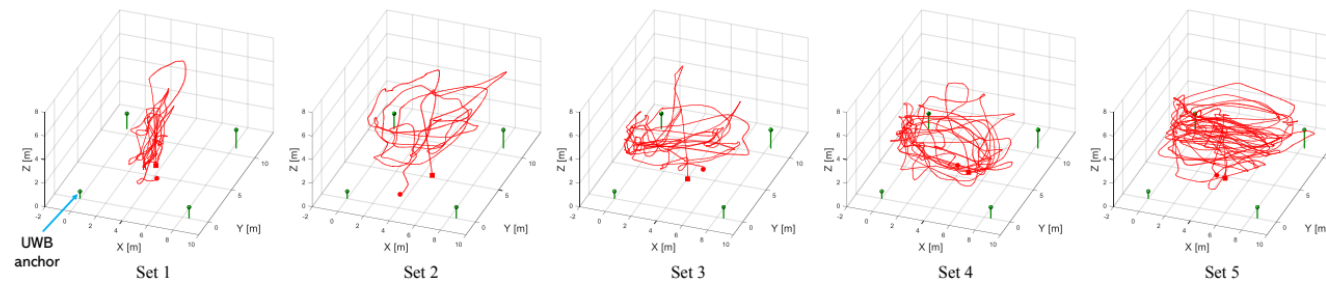
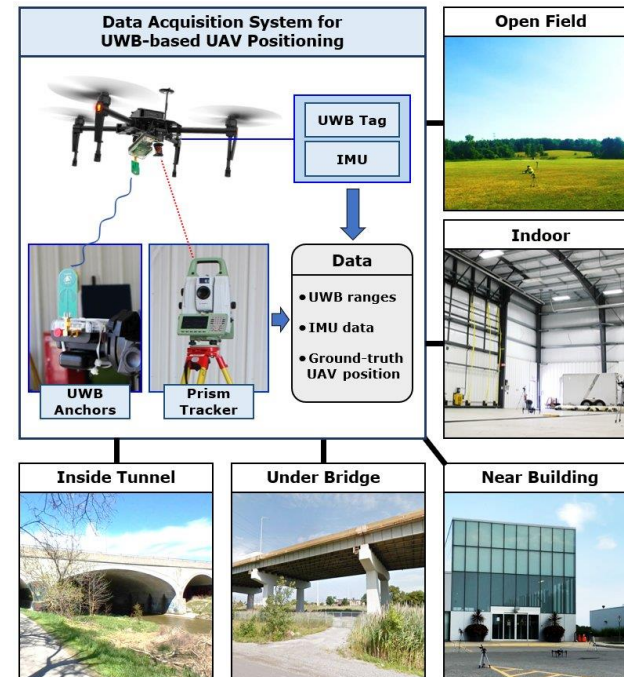
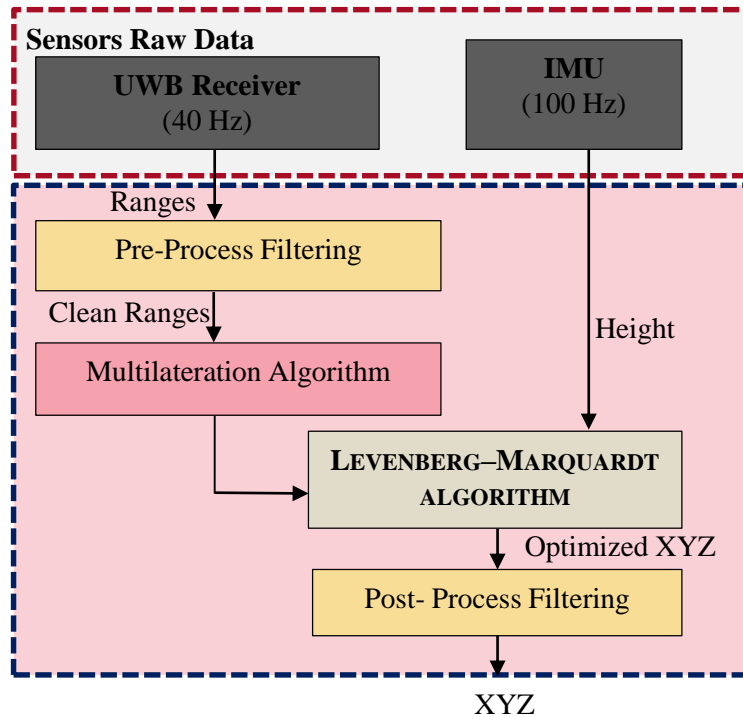


# Drone Trajectory Analysis and Visualization

## Project 1



# Drone Trajectory Analysis and Visualization



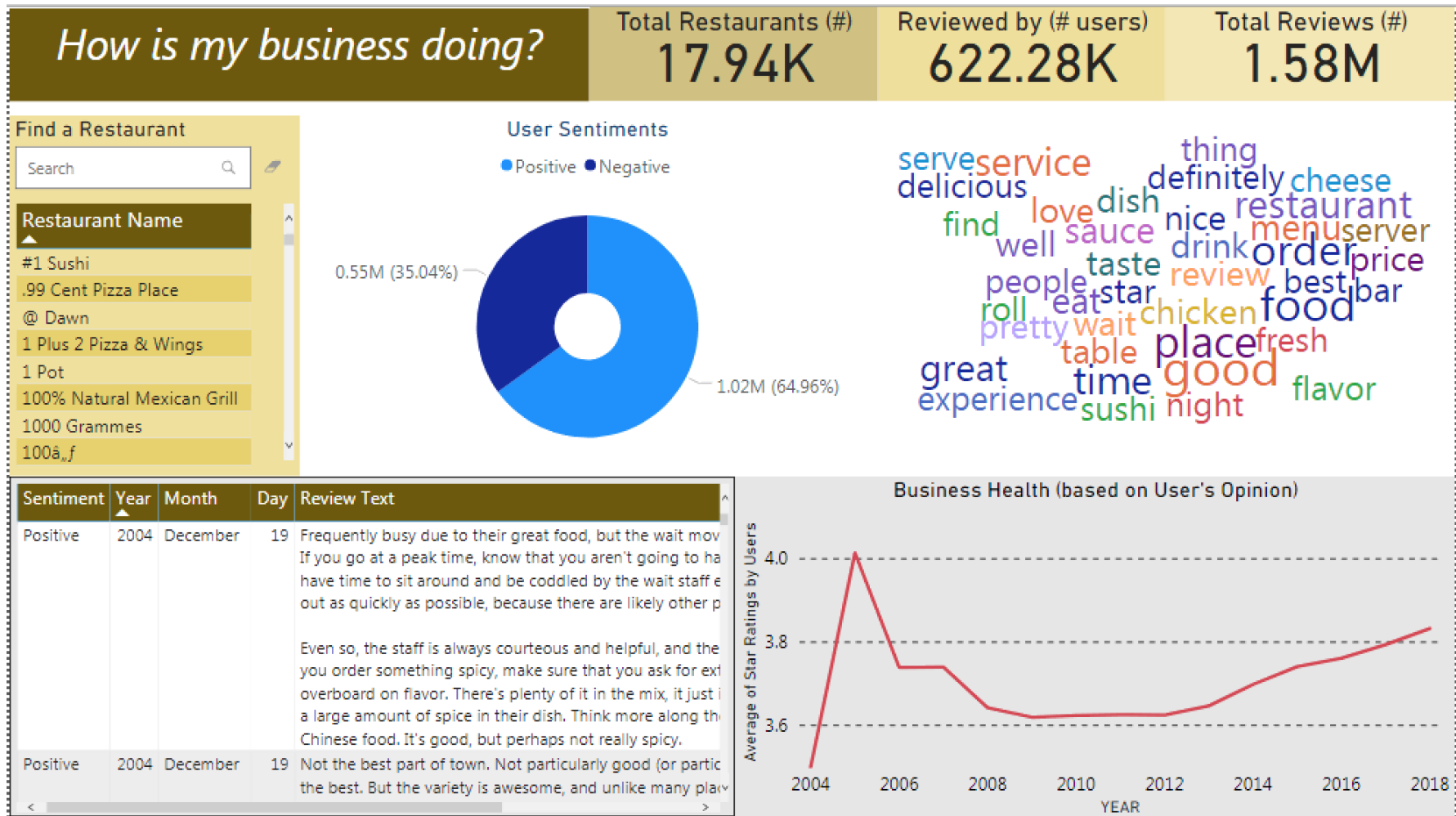


# Yelp Dataset Analysis for Restaurant Improvement Recommendations

## Project 2



# Restaurant Insights



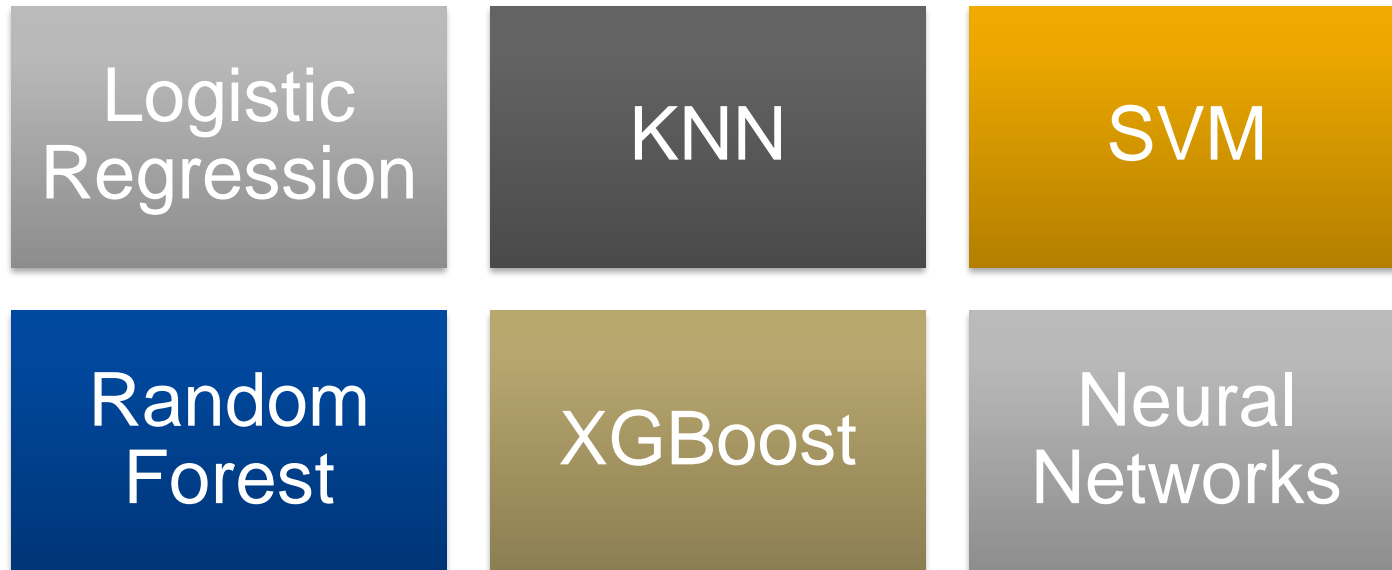


# Enhanced Fraud Detection using IEEE CIS dataset

## Project 3



# Enhanced Fraud Detection



## Classifier Selection



# Toronto BikeShare Data Analysis

## Project 4



# Bike Sharing Analytics

- Who is using Bike Share? Are there distinct differences between members vs. casual users behaviours?
- When are people using Bike Share? How does usage change across the year, the week, and the day?
- How is bike share being used? Commute to work or to explore the city? How does the weather change the way people





# Do Torontonians need the Proposed Pickering Airport

Project 5



# Structural Analysis of the Github Network

## Project 6



# Network Analytics

- Can we detect important developers and structure of the communities in github?
- Can we recommend possible projects that a user would be interested in watching?
- What are the underlying insights in the Github data?

Rank	Username	PageRank $\times 10^4$
1	torvalds	8.82
2	JakeWharton	6.93
3	michaelliao	5.25
4	githubpy	4.44
5	Tj	3.84
6	mojombo	3.30
7	paulirish	3.24
8	defunkt	2.93
9	ruanyf	2.90
10	gaearon	2.76

Rank	Username	Followers
1	torvalds	80184
2	JakeWharton	48120
3	ruanyf	39102
4	Tj	37402
5	addyosmani	32666
6	paulirish	29690
7	yyx990803	29200
8	gaearon	27415
9	sindresorhus	25701
10	mojombo	25112



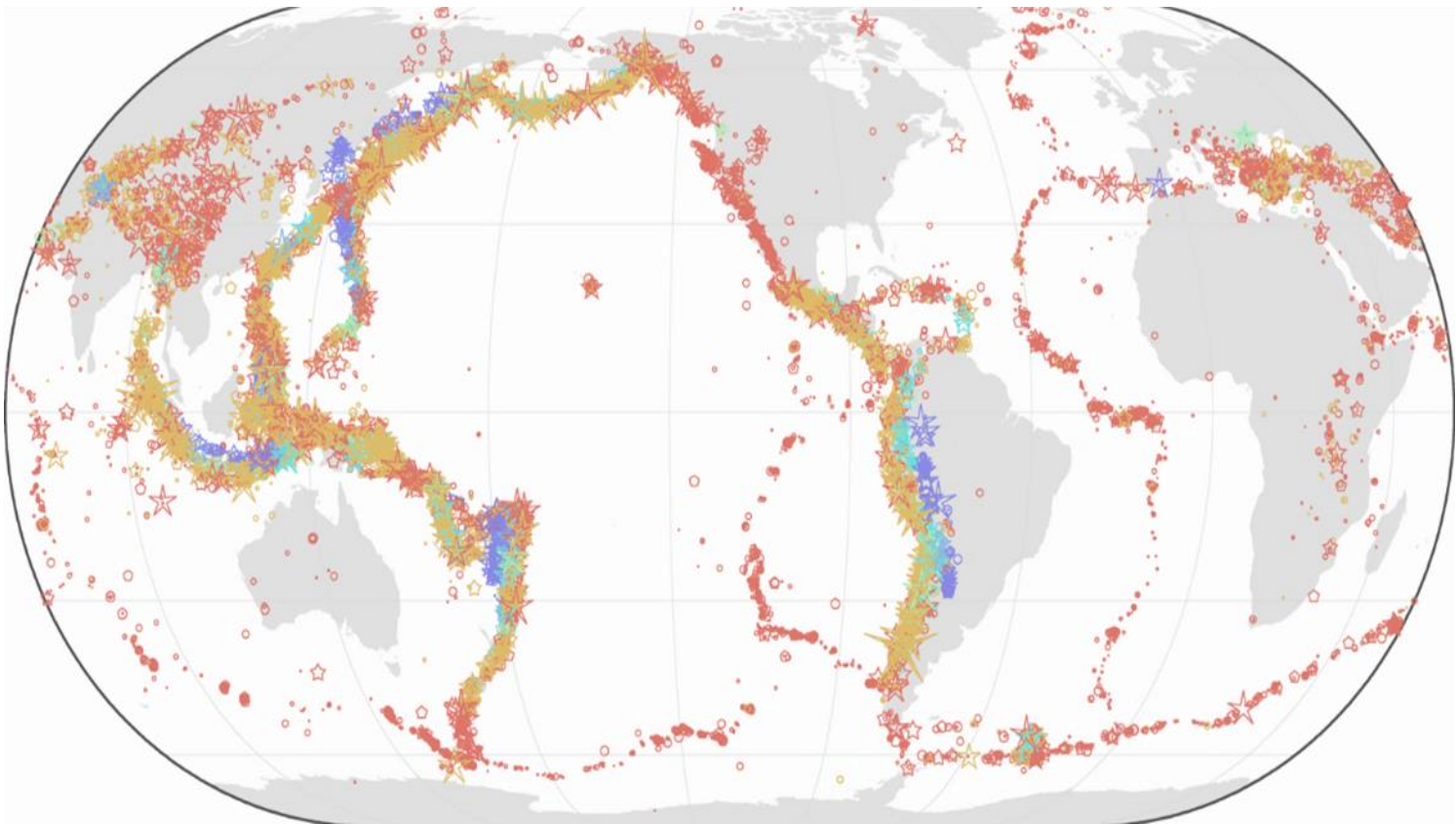
# Earthquake Analysis

## Project 7



# Earthquake Analysis

Analyzing the behaviors and patterns of earthquakes can help us in better preparation and crisis management





# Question Answering using Deep Learning

## Project 8



# Question Answering using Deep Learning

- Multi-hop QA task: for every question, the model needs at least two sentences (supporting facts) in two different documents to answer
- What do we need to do?
  - Find gold documents in the given 10 documents.
  - Find all sentences that help to answer the questions in the gold docs
  - Reason over these supporting facts to predict the answer

## Paragraph A, Return to Olympus:

[1] Return to Olympus is the only album by the alternative rock band Malfunkshun. [2] It was released after the band had broken up and after lead singer Andrew Wood (later of Mother Love Bone) had died of a drug overdose in 1990. [3] Stone Gossard, of Pearl Jam, had compiled the songs and released the album on his label, Loosegroove Records.

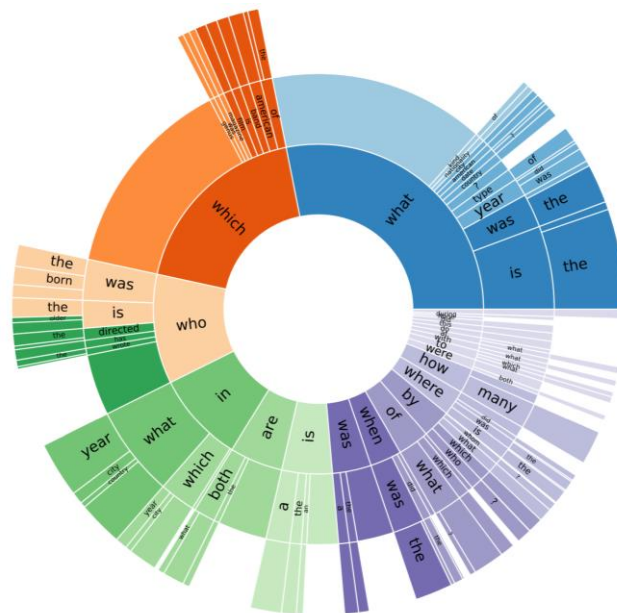
## Paragraph B, Mother Love Bone:

[4] Mother Love Bone was an American rock band that formed in Seattle, Washington in 1987. [5] The band was active from 1987 to 1990. [6] Frontman Andrew Wood's personality and compositions helped to catapult the group to the top of the burgeoning late 1980s/early 1990s Seattle music scene. [7] Wood died only days before the scheduled release of the band's debut album, "Apple", thus ending the group's hopes of success. [8] The album was finally released a few months later.

**Q:** What was the former band of the member of Mother Love Bone who died just before the release of "Apple"?

**A:** Malfunkshun

**Supporting facts:** 1, 2, 4, 6, 7





# Exploring Stack Overflow: Answer Time Prediction and User Profile Analysis

## Project 9



# Problems

## Unasnwered Questions

- Answer time
- prediction [1]

## Tags Accuracy

- Tags
- Prediction [2]

## Answer Quality

- User Profile
- Analysis [3]

[1] D. Wu, S. Johnson, C. Foster, E. Li, H. Elmiligi and M. Rahman, "Improving Response Time Prediction for Stack Overflow Questions," *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, Vancouver, BC, Canada, 2019, pp. 0786-0791.

[2] D. Movshovitz-Attias, Y. Movshovitz-Attias, P. Steenkiste, and C. Faloutsos. 2013. Analysis of the reputation system and user contributions on a question answering website: StackOverflow. In 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013). 886–893. <https://doi.org/10.1145/2492517.250024>

[3] T. Saini and S. Tripathi. 2018. Predicting tags for stack overflow questions using different classifiers. In 2018 4th International Conference on Recent Advances in Information Technology (RAIT). 1–5. <https://doi.org/10.1109/RAIT.2018.8389059>



# Graph-based Identification of Attacks in Anomalous Traffic

Project 10

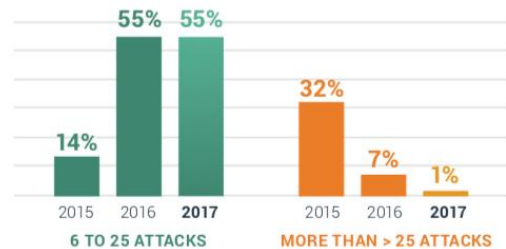


# Motivation

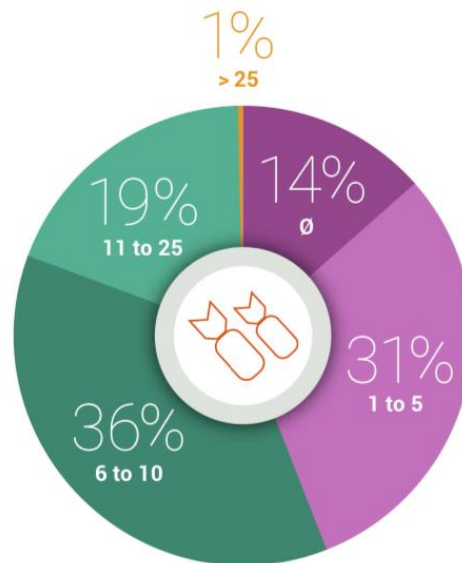
*DDoS attacks have increased in sophistication and intensity, threatening the availability of enterprise services, applications, websites, and networks.*

Most survey respondents (86%) report experiencing at least one DDoS attack in the past 12 months.

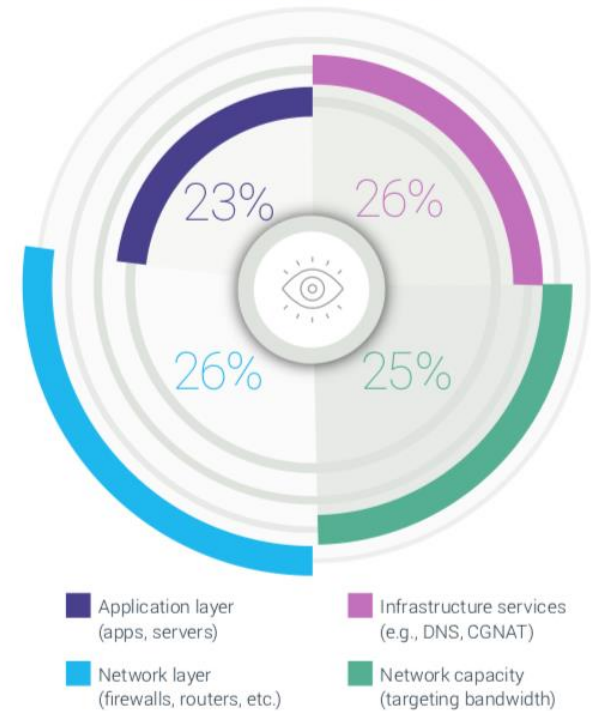
NUMBER OF DDoS ATTACKS – **OVER TIME**



NUMBER OF DDoS ATTACKS – **PAST 12 MONTHS**



PERCENT OF DDoS ATTACKS **TARGETING EACH AREA**





# Data Analytics and Route Navigation in Toronto Based on Accidents Data

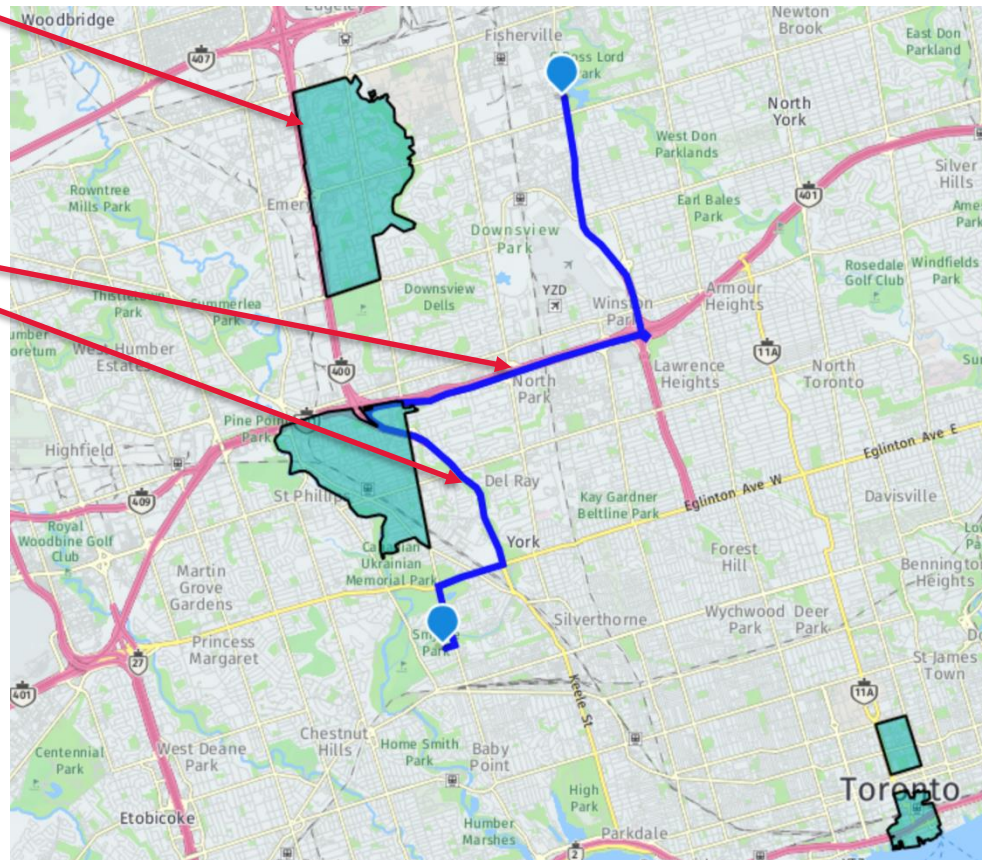
Project 11



# Navigation that Avoids High-risk Areas

Accident areas  
after clustering

Suggested  
Navigation  
Route





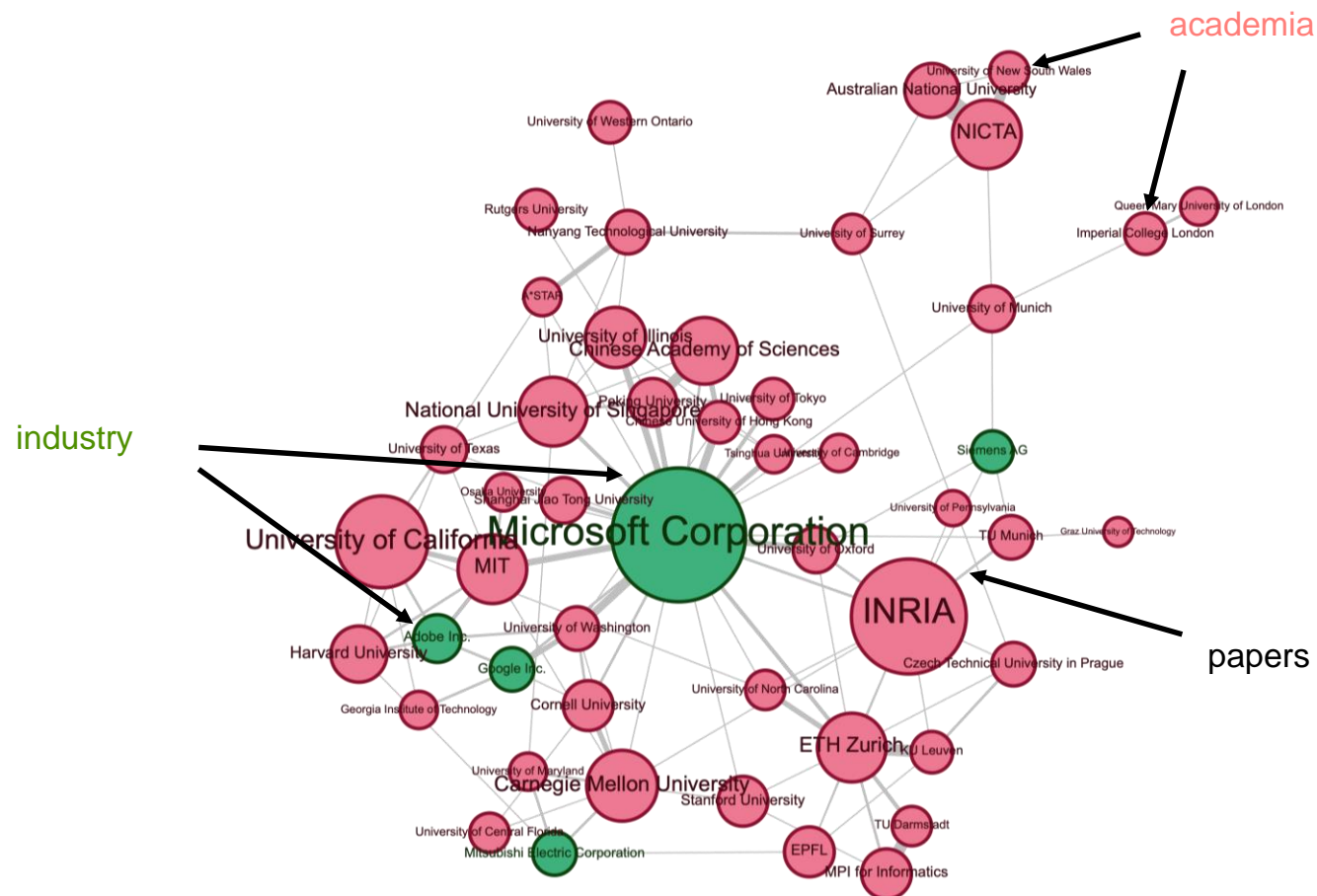
# Impact of Industry on Computer Vision Research

Project 12



# Results

## Affiliation network: 2010





# Bitcoin Network Analytics and Machine Learning

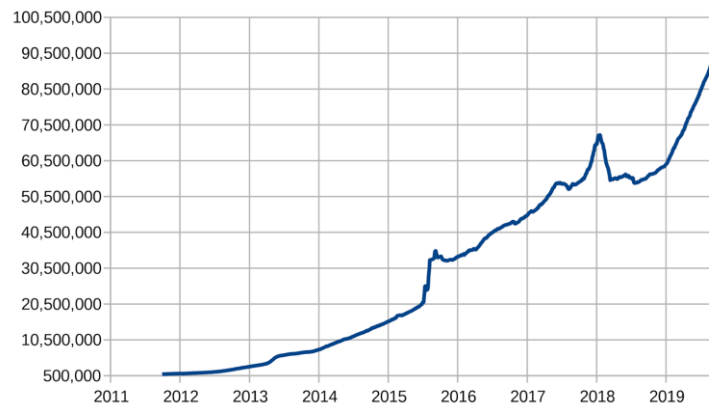
## Project 13



# Motivation

## Bitcoin (□)

- Decentralized digital currency without a central bank that can be sent from user to user on the peer-to-peer bitcoin network without the need for intermediaries
- Exponential rise in use and interest since 2010
- High volatility in stock prices



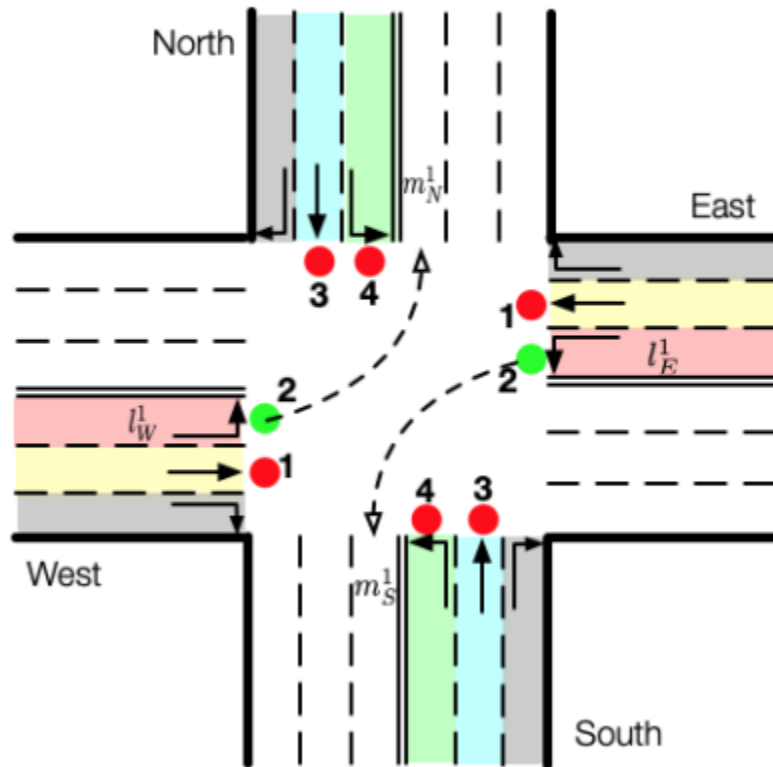


# Adaptive Traffic Signal Control

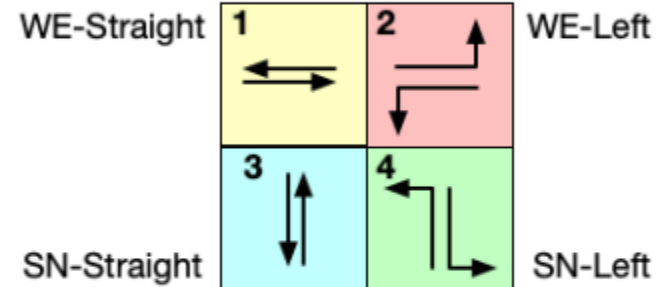
## Project 14



# Motivation



## Four **phases**



Two **traffic movements** (dashed arrows) are allowed in phase #2

$$p_2 = \{(l_W^1, m_N^1), (l_E^1, m_S^1)\}$$



# Prediction of Hit Songs

Project 15



# Motivation



So far, we have a predictive model with **good performance** in **classifying** hit/non-hit songs, based on **high level** and **low level** features.



We need to Train and test our model using **more recent songs**, with a revision on the feature set to select out the **most influential features**.



Github Repository: [https://github.com/shahabty/hit\\_song\\_prediction](https://github.com/shahabty/hit_song_prediction)

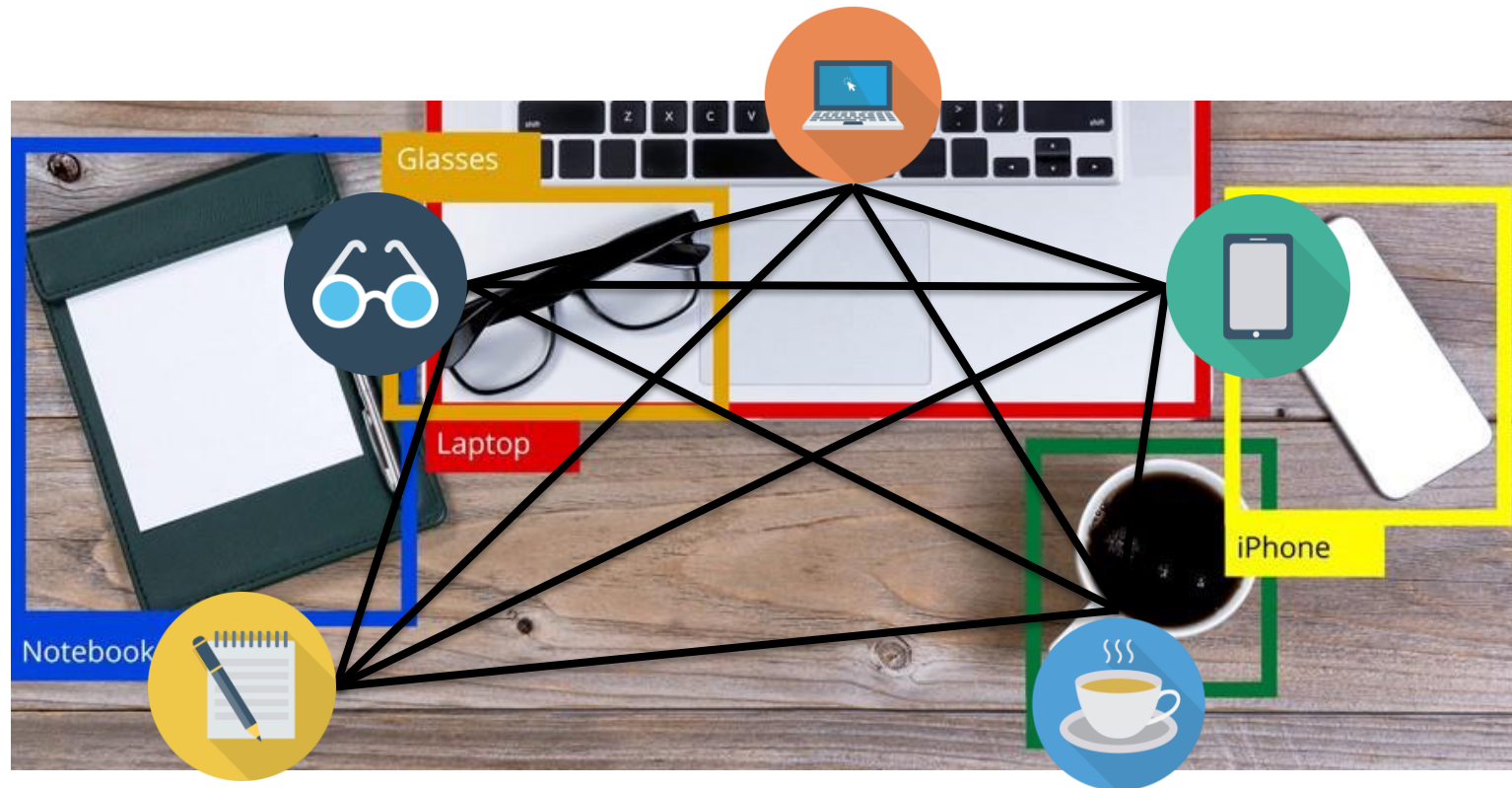


# Knowledge Graph Construction from Images

Project 16



# Main Approach



**Object Detection**

**Generate Graph from detected Objects**



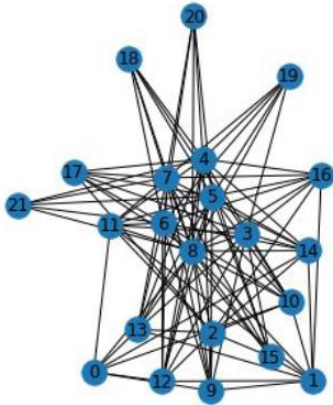
# Predicting Magnetic Interactions of Molecules

Project 17

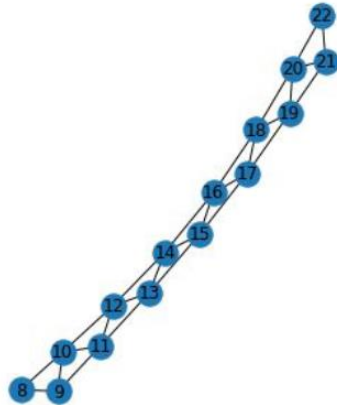


# Scalar Couplings

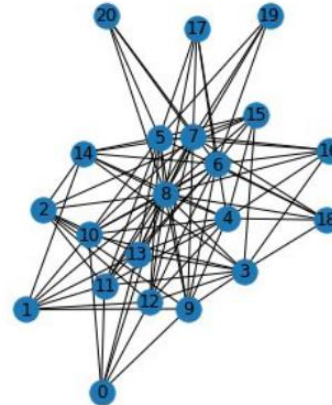
Graph for type 1JHC



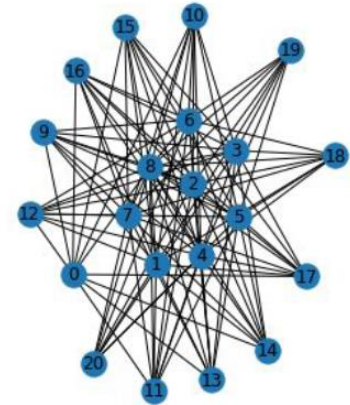
Graph for type 2JHH



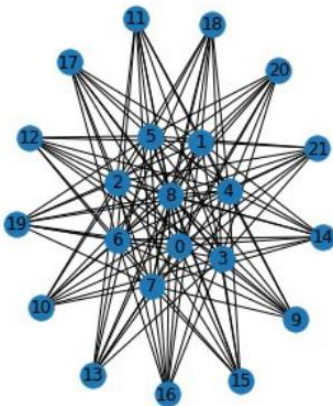
Graph for type 1JHN



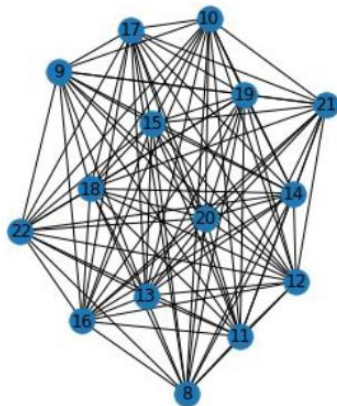
Graph for type 2JHN



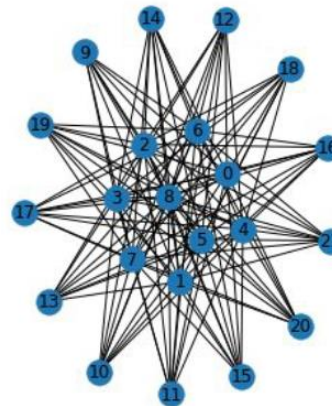
Graph for type 2JHC



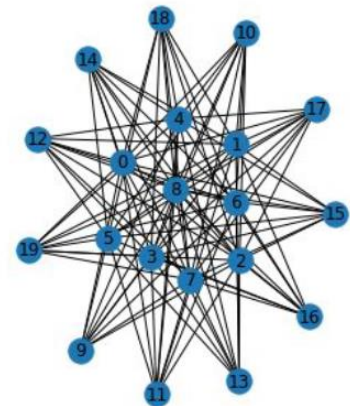
Graph for type 3JHH



Graph for type 3JHC



Graph for type 3JHN



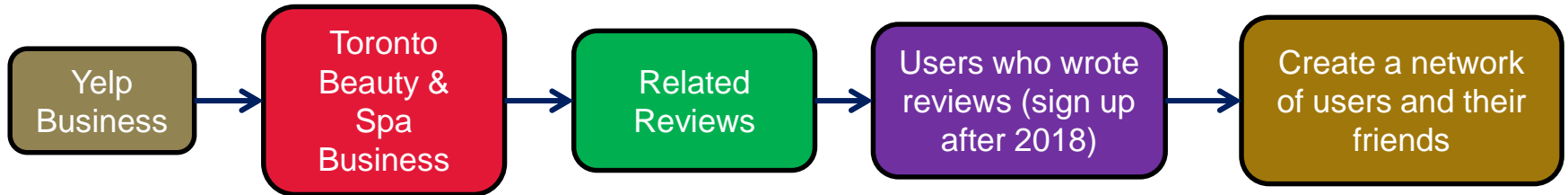


# Yelp Network Analytics and Recommendations

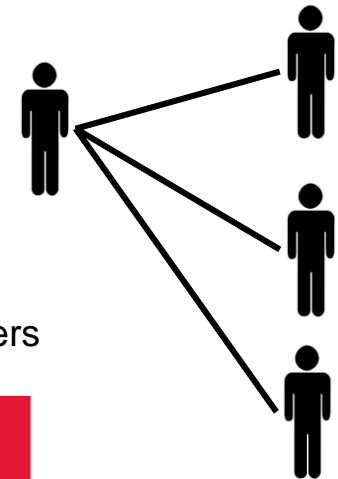
Project 18



# Main Approach



- Data filtering
  - Limited to Toronto City
  - Focus on “Beauty and Spa” business category
  - Concentrate on user signed up after 01-01-2018
- Goal
  - Focus on the study on a city, a category and limit the number of users



Data	Beauty & Spa Business	Review	Users
Count	1618	23620	172



Winter 2021



# Analysis of Google Play Store Reviews

## Project 1



# Our Data Set



App reviews on the  
Google Play Store

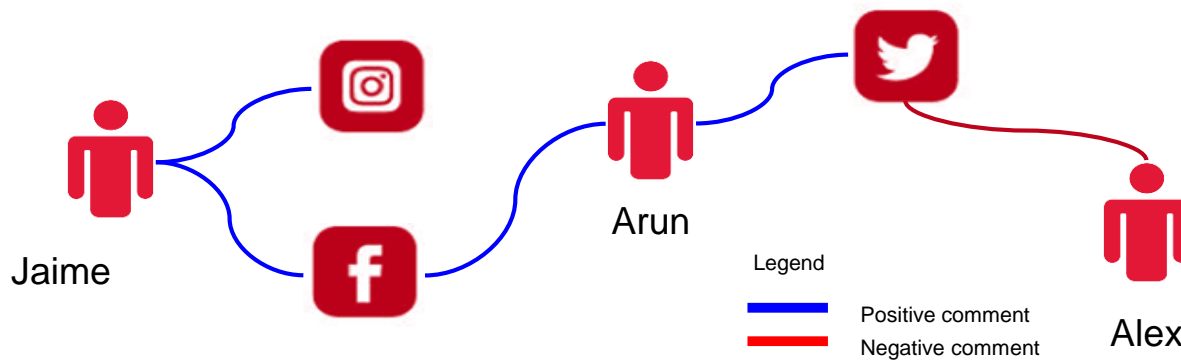


547 apps  
8 million reviews



Several languages

## Network Topology Based on Comment Sentiment





# Human Motion-State Classification from Motion Capture Data

## Project 2



# Animation & Motion Capture Databases

Current motion capture databases rely on manually annotated tags for querying. This approach has problems:

- 1) Manual labeling does not scale well
- 2) Is not always very accurate

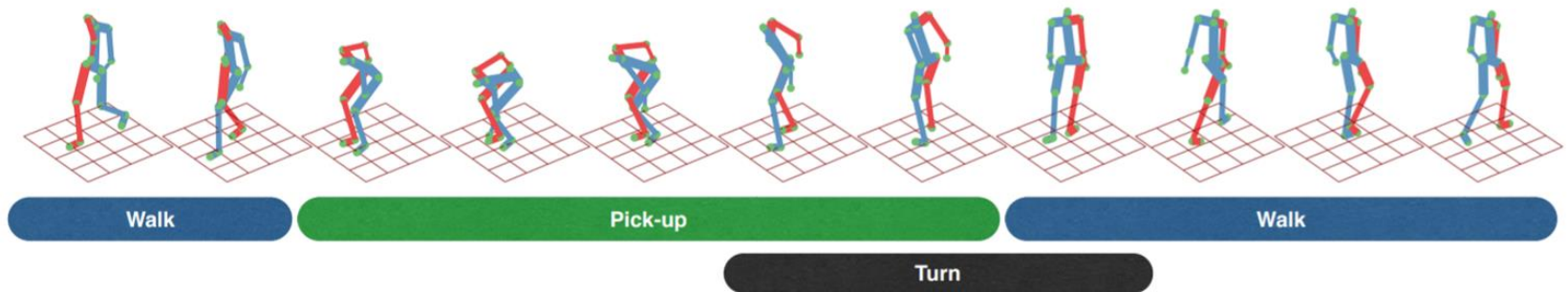


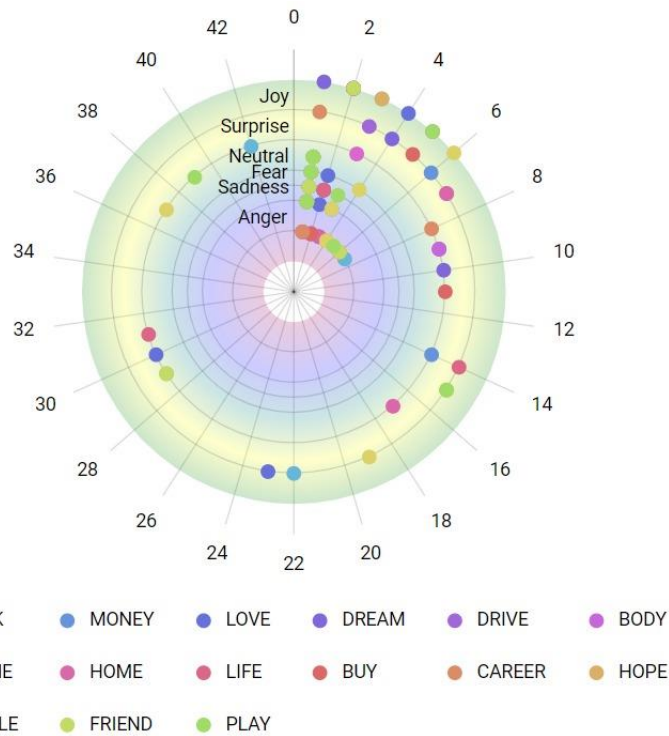
Image Source: Gupta et al. (2016) [1]



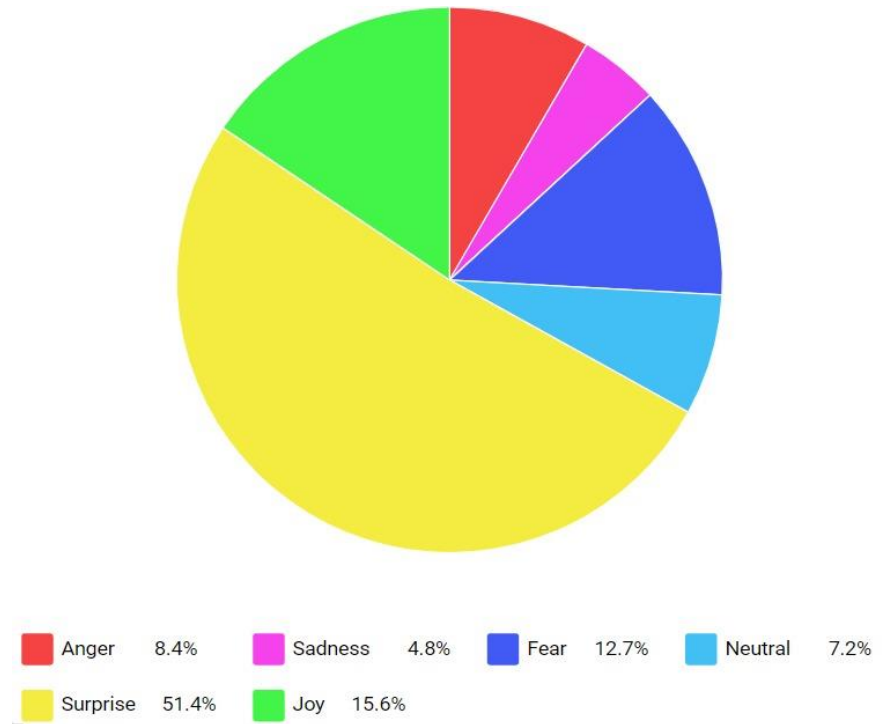
# User Emotional Modeling and Visualization in Social Media Conversations

## Project 3





User emotional association to topics



Overall emotion distribution

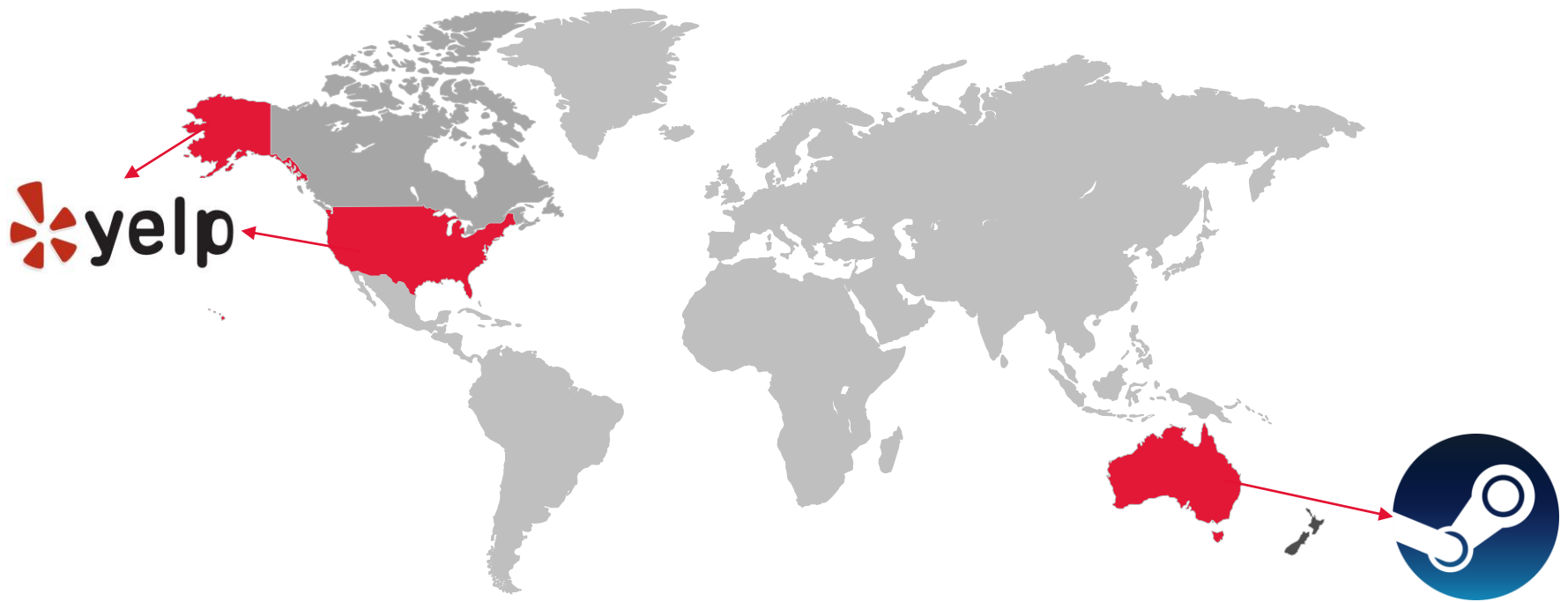


# LLM-based Recommendations on Yelp

## Project 4



# Yelp Dataset





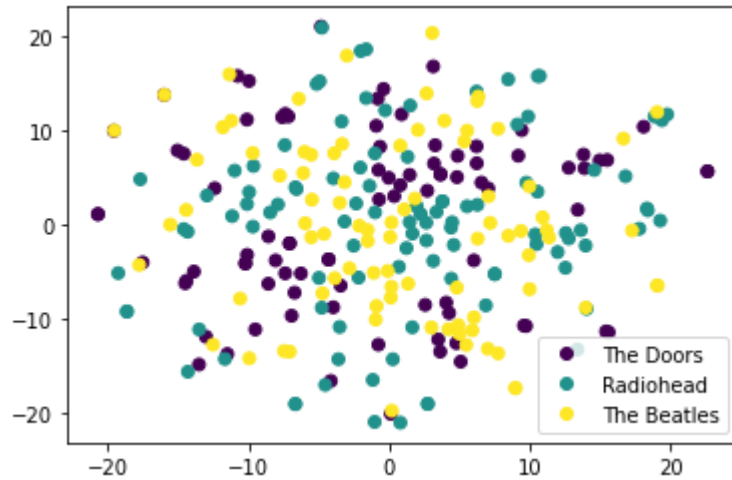
# Deep-learning based Music Recommendations

## Project 5

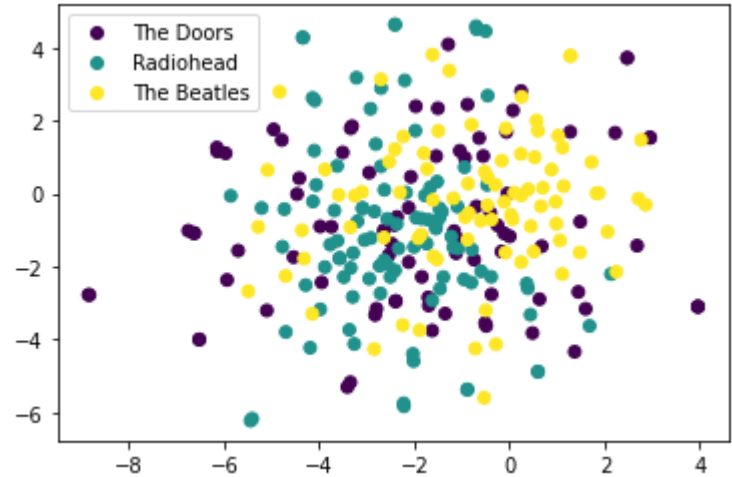


# Lyrics embedding

TF-IDF (only considers frequency of words)  
dim = 500



Doc2Vec (considers the order and meaning of the words)  
dim = 100





# Knowledge Graph for API Method Recommendation

Project 6





What we  
propose

**Self-explainable** and  
**interpretable** API recommendation  
system using a **Knowledge Graph**

Date of birth: April 15, 1452  
Date of death: May 2, 1519  
(age 67 years)



# Movie Metadata Analysis to Guide Recommendations

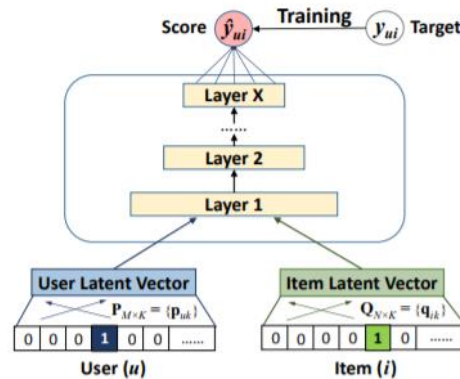
Project 7



# Approach Overview

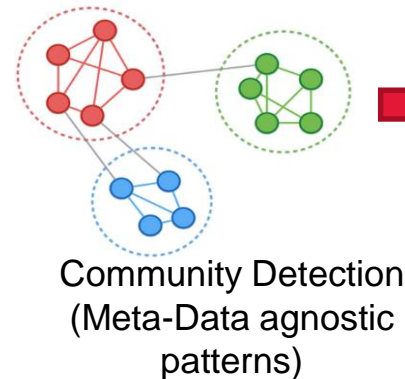
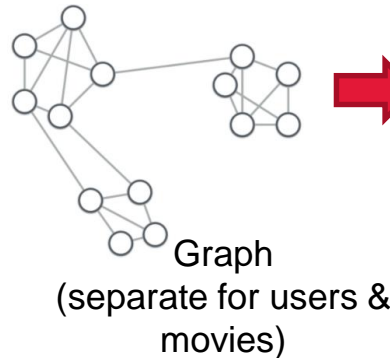
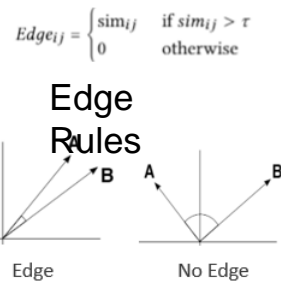
Data

userId	movieId	rating
156	1	5.0
156	2	5.0
156	4	3.0



Movie Embedding Matrix (N x K)

User Embedding Matrix (M x K)





# Visualizing Stories

## Project 8



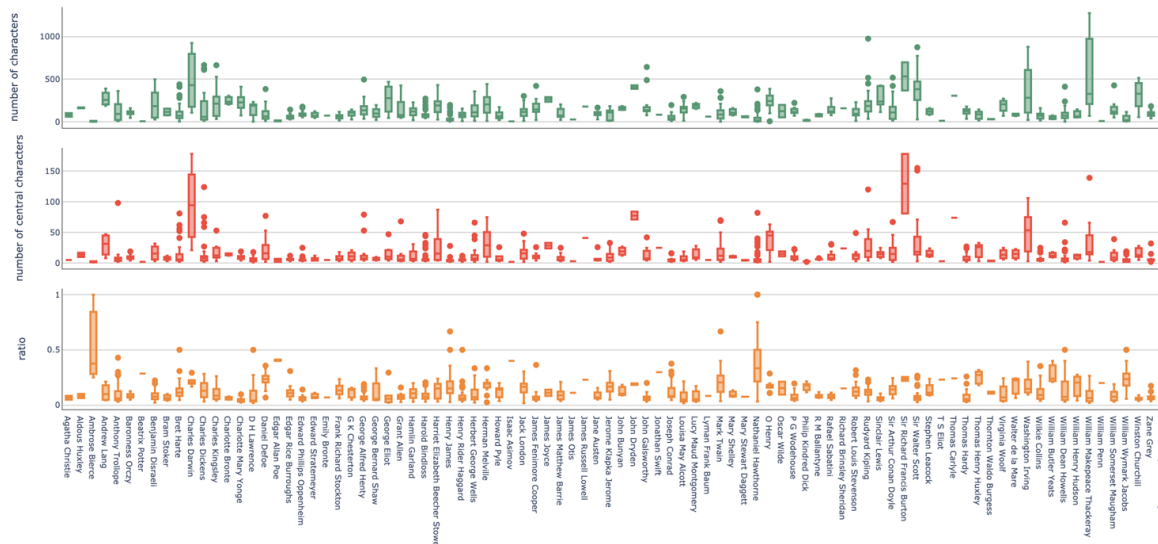
“Do authors use similar patterns?”  
 “Are patterns similar in a genre?”  
 “What pattern appear in real stories like biography  
 and history books?”



Sherlock Holmes and Dr. John Watson By Sidney Paget

Image is retrieved from Wikipedia.

([https://upload.wikimedia.org/wikipedia/commons/7/73/Paget\\_holmes.png](https://upload.wikimedia.org/wikipedia/commons/7/73/Paget_holmes.png))



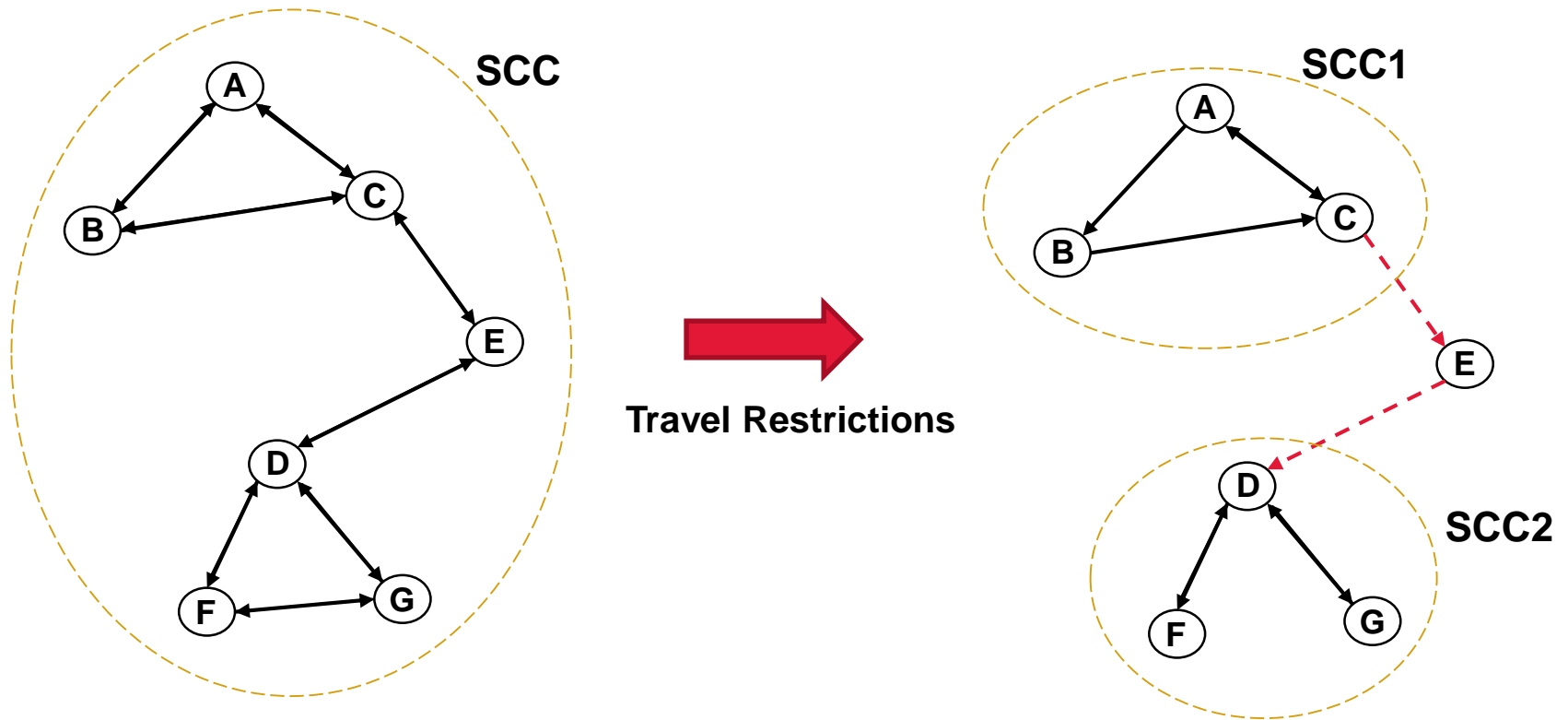


# Flight Network During Covid-19

## Project 9



# Strong Connected Components (SCC)





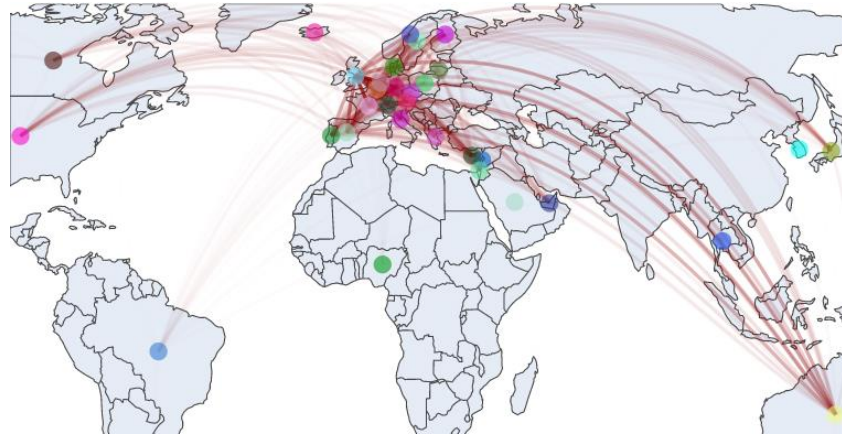
# Product-Customer Networks –Final Presentation

Project 10



# Motivation & Dataset Information

- General Information:
  - 1067371 Transactions
  - 41 Countries
  - 5943 Customers
  - 5699 Products



- Questions:
  - What are the different communities of products/clients?
  - What new products can be recommended to different clients?
  - What products may be of interest to certain regions?



# Movie Casting Problems

## Project 10



# Graph Autoencoder on Actors Network

- Learning a joint representation of both local graph structure and node features
- MSE on link weight prediction task 0.005406 (for edge weights from 0. to 1)

