Big Data Analytics & Machine Learning for Smart Cities

ASOR Thursday Afternoon Seminar
9 July 2020

Peter Ryan[1] and Richard Watson[2]

1 – Defence Science & Technology Group
2 – Research Scientist

Peter Ryan

- Education: BSc (Hons), PhD (Physics) Melbourne University
- Academia: Postdoc (US)
- Defence Science & Technology:
  - Research Scientist, Scientific Advisor, Scientific Manager, Visiting Scientist (UK)
- **Now:**
  - **Honorary Research Fellow, Defence Science**
  - **Chair of Standards Australia Committee in M&S**
  - **Australian delegate for several ISO Committees including Smart Cities**
  - **Private researcher**

Richard Watson

- Education: MSc (Melb), PhD (ANU), Grad Dip OR (Canberra)
- Research Scientist, Central Studies Establishment, Canberra
- Senior Research Scientist, DSTO Melbourne
- Lecturer in IT, Swinburne University of Technology
- Various senior roles in the IT industry in Melbourne & Auckland NZ
- **Now:**
  - **Private researcher & consultant**

Overview of Presentation

- Smart Cities
- Big Data Analytics
- Council Datasets
- Student Project – Pedestrian Counts
- Pilot Project – Victorian Road Accidents
- Student Project – Web APIs
- Emerging Projects – Waste Management
- Emerging Projects – Social Indicators

Smart Cities

- A Smart City
  - exploits modern ICT technologies to provide greater efficiencies for urban areas
  - integrates ICT and IoT networked physical devices to optimize the efficiency of city operations and services
  - allows city officials to interact directly with both community and city infrastructure to monitor city activities and demonstrate how the city is evolving.

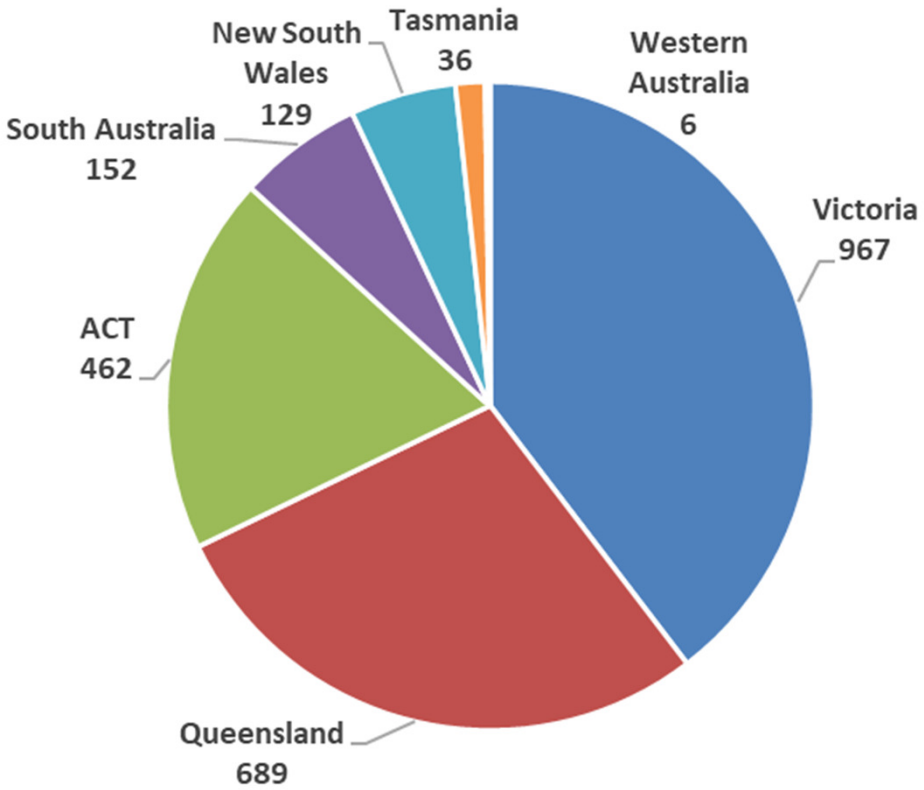
Big Data in Smart Cities

- Smart Cities generate massive amounts of data through their networked sensors
- Referred to as Big Data
  - datasets that are too large and/or complex to be effectively and efficiently handled by traditional data-related theories, technologies, and tools
- Such datasets are collected at local, state, and federal levels in Australia

Australian Government Open Datasets

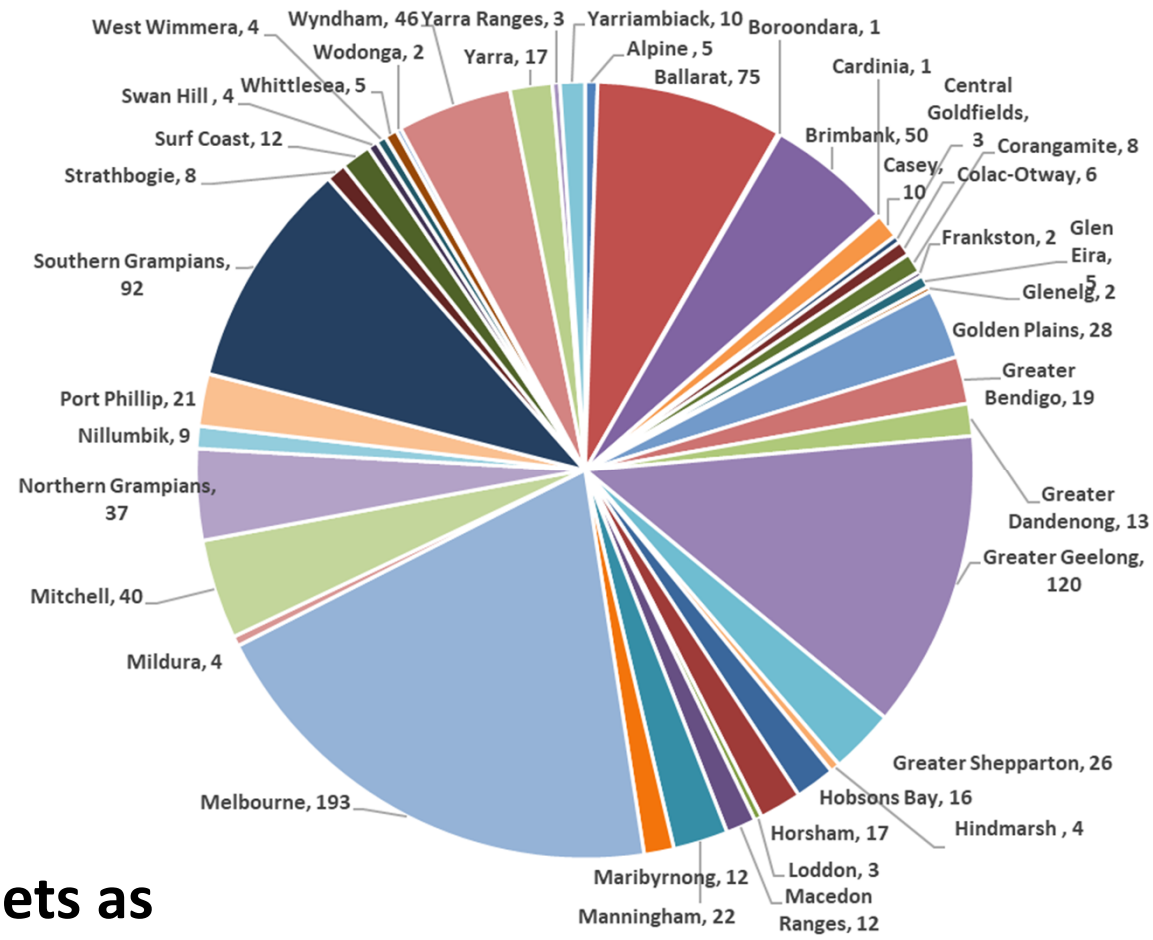
- City of Melbourne:
  - https://data.melbourne.vic.gov.au/
- DataVic:
  - https://www.data.vic.gov.au/
- Australian datasets:
  - https://data.gov.au/
- Council datasets

Open Datasets from Local Councils in Each State



2441 Datasets
as of April 2020

Victorian Council Open Datasets

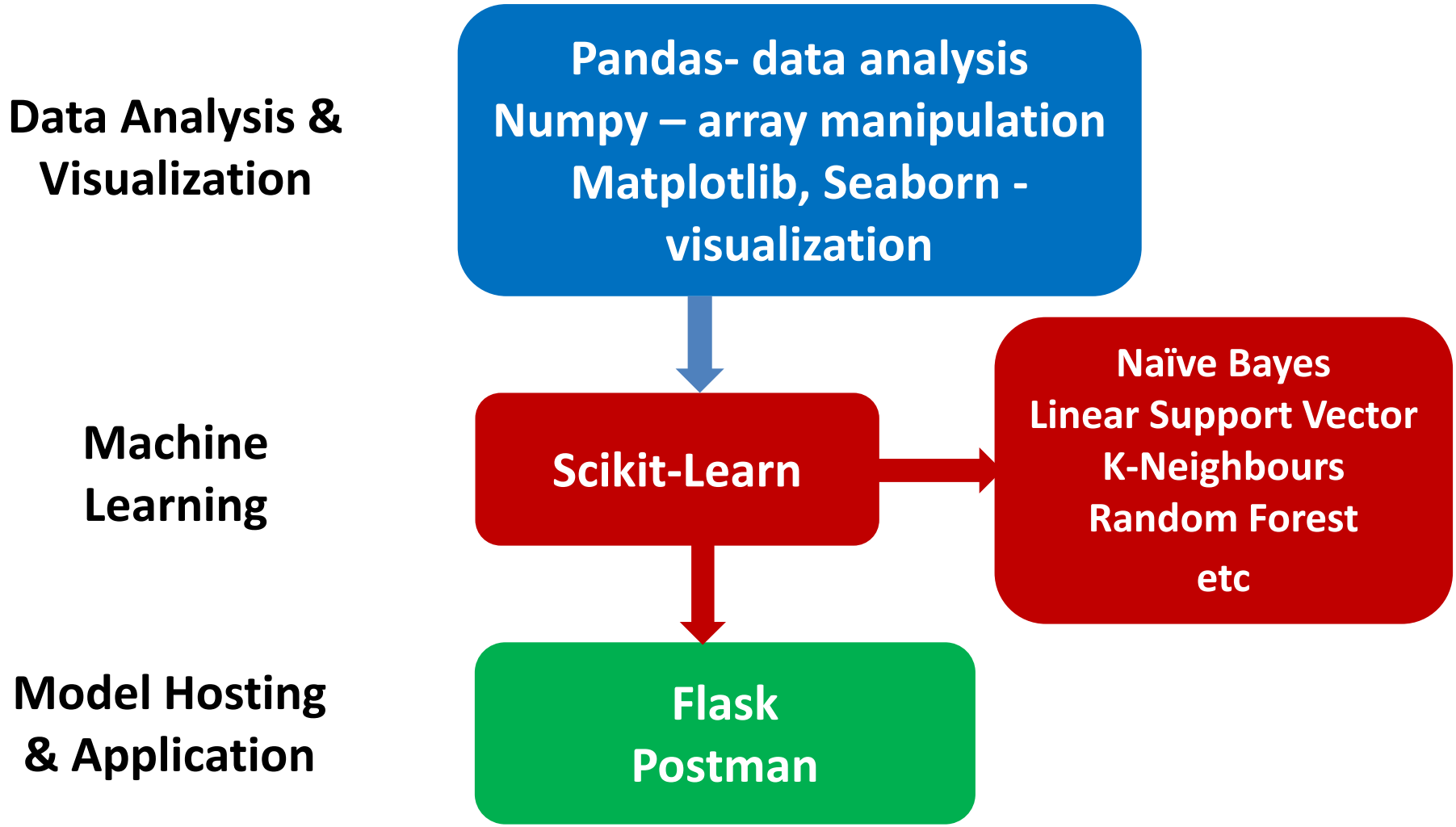


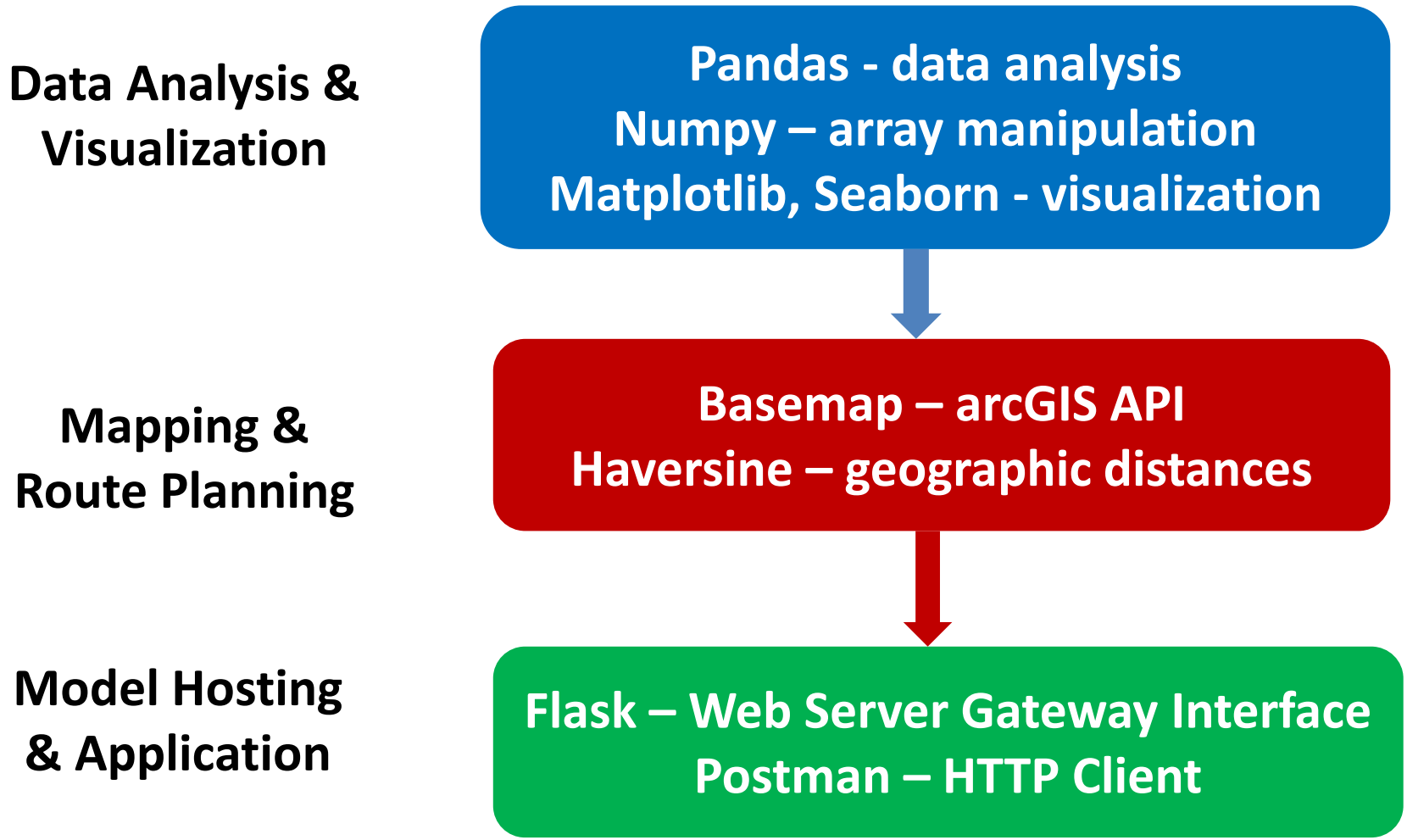
967 Datasets as of April 2020

Analysis of Big Data with Machine Learning

- ML - answer questions and make predictions
- Stages: dataset preparation and preprocessing, dataset splitting, modelling, model deployment
- Machine learning as a service:
  - Major: Amazon, Microsoft, Google and IBM
  - Other: DataRobot, RStudio, BigML
- Python Libraries:
  - *Pandas, numpy, matplotlib, seaborn, scikit-learn*

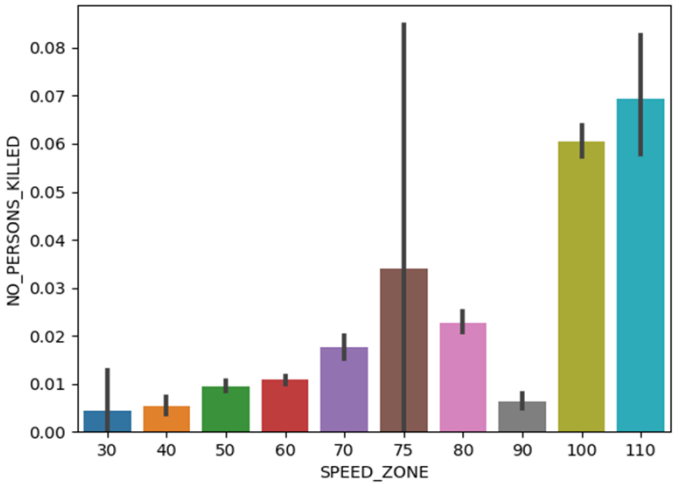
Software Modules

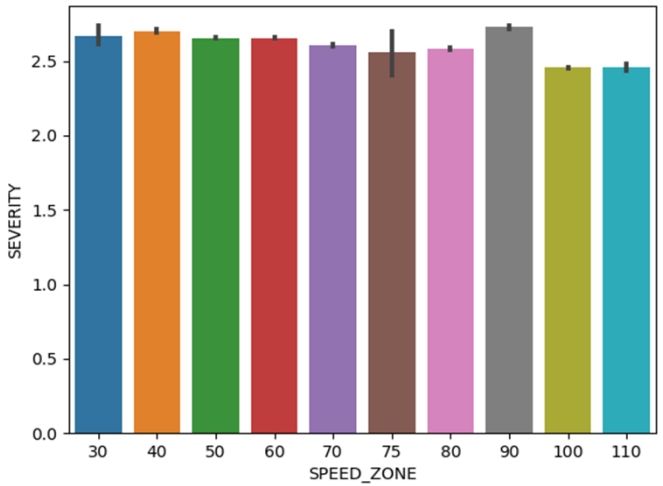



Data Analysis & Visualization
Pandas - data analysis
Numpy – array manipulation
Matplotlib, Seaborn - visualization
Mapping & Route Planning
Basemap – arcGIS API
Haversine – geographic distances
Model Hosting & Application
Flask – Web Server Gateway Interface
Postman – HTTP Client

Pilot Project – Victorian Crash Data

- 40 MB dataset 186546 rows x 28 columns
- Analysis using *pandas*
- Visualization using *matplotlib* and *seaborn*
- Explore dependencies of target outcomes on features such as speed zone





Predictive Machine Learning

- Predict outcomes using *Naïve Bayes, Linear Support Vector Classification, K Neighbors Classifier, Random Forest* or other models
- *scikit-learn* machine learning tools
- Compare model accuracy to inform ML model
- Save model using *joblib* library
- Create Application Programming Interface (API) using python *flask* library
- Test model over web using *Postman* client

Pilot Project Results

Scores for ML Models

Model	Score
Random Forest	**63.4%**
Naïve Bayes	62.9%
K-Neighbors Classifier	58.7%
Linear Support Vector Classification	45.1%

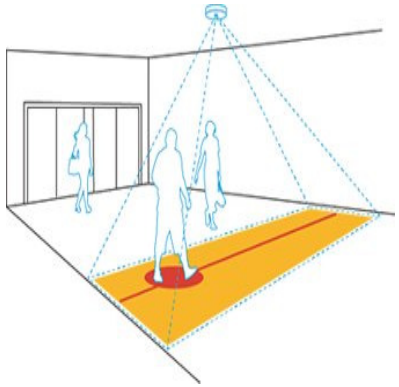
Feature Importance for Random Forest

Feature	Importance
LIGHT_CONDITION	**0.344**
ROAD_GEOMETRY	**0.175**
SPEED_ZONE	**0.480**

Model deployed and tested on Postman

Swinburne 2019 (1)

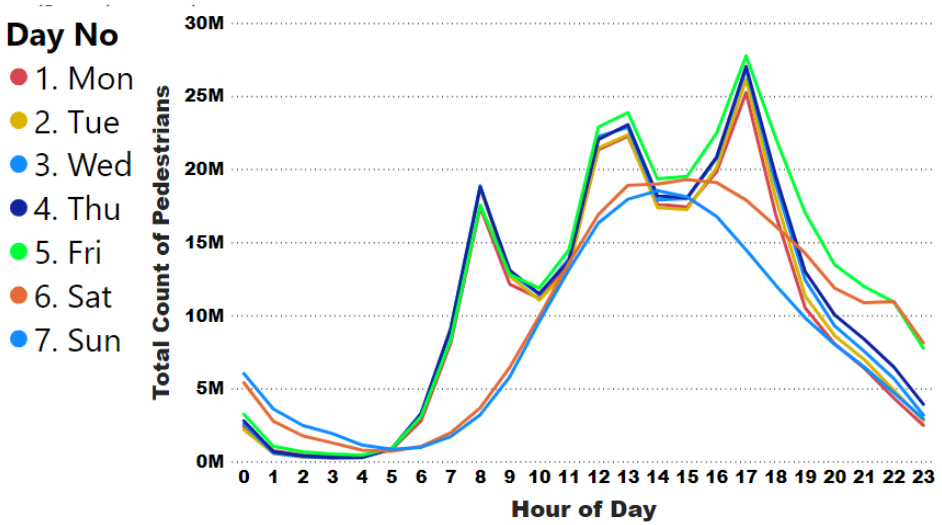
- Enhancing Pedestrian Mobility within Smart Cities
- City of Melbourne (COM) pedestrian counting system
- http://www.pedestrian.melbourne.vic.gov.au/



SWINBURNE UNIVERSITY OF TECHNOLOGY

Swinburne 2019 (2)

- MS PowerBI analysis
  - 3 datasets: (1) Pedestrian Counting System (Past Hr), (2) Pedestrian Counting System All Time, (3) Pedestrian Sensor Location
  - Datasets downloaded from COM web portal



SWINBURNE UNIVERSITY OF TECHNOLOGY

Swinburne 2019 (3)

- Future Directions
  - Apply ML to optimize pedestrian traffic flow
  - Link with social media data from mobile phones and fitness or health data from wearable devices

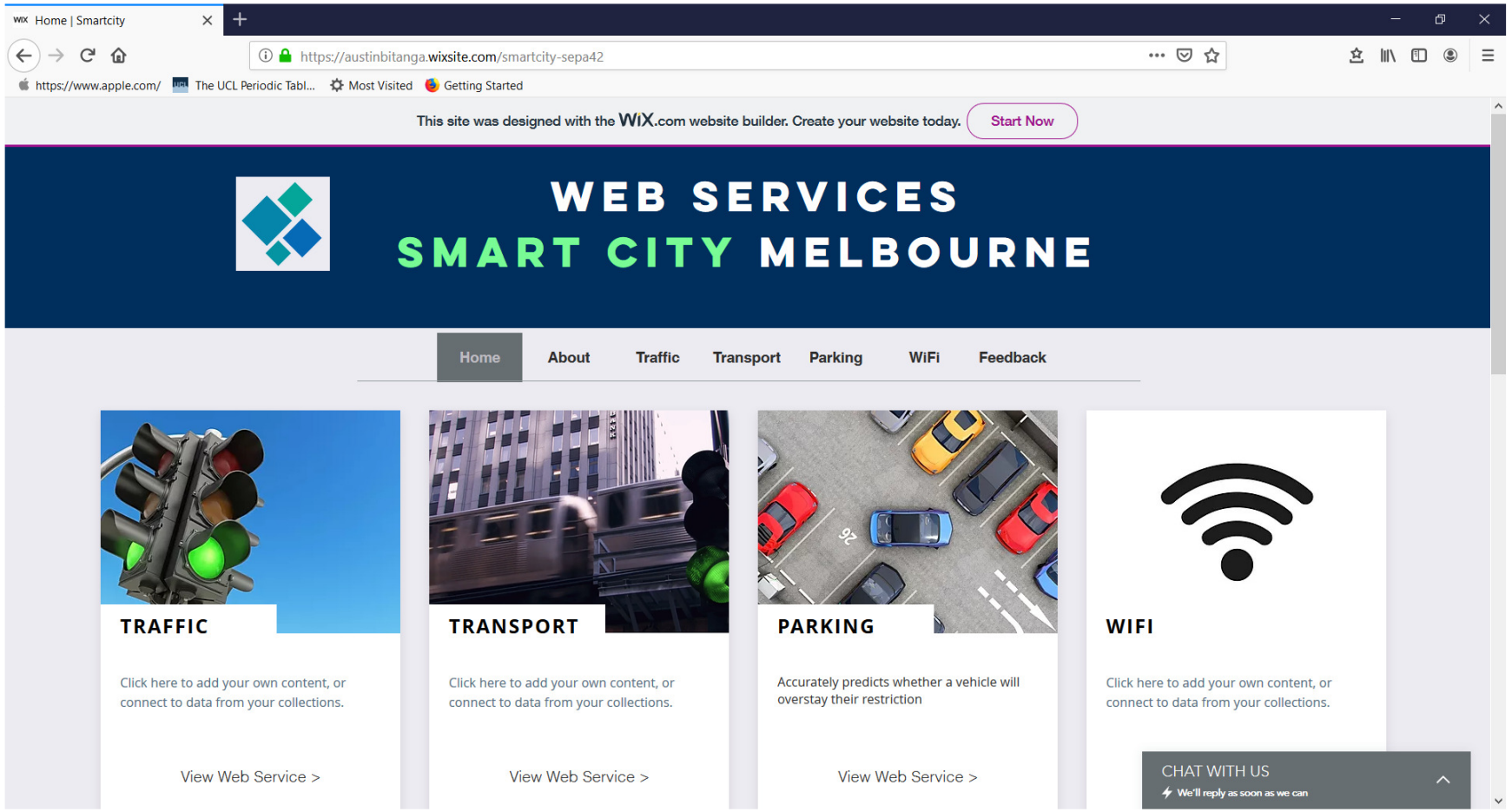
- Published as:
  - Carter, E.; Adam, P.; Tsakis, D.; Shaw, S.; Watson, R.; Ryan, P., Enhancing pedestrian mobility in Smart Cities using Big Data. *Journal of Management Analytics* **2020,** *7* (*2*), 173-188.

Swinburne 2020

- Explore local council datasets – Melbourne, Casey, Wyndham, Geelong, Yarra, Adelaide, etc
- Transport theme
- Big Data Analytics and Machine Learning
- Approaches
  - Machine Learning as a Service (MLaaS)
  - Python Libraries
  - API Development and Cloud Deployment
  - REST Client Development

SWINBURNE UNIVERSITY OF TECHNOLOGY

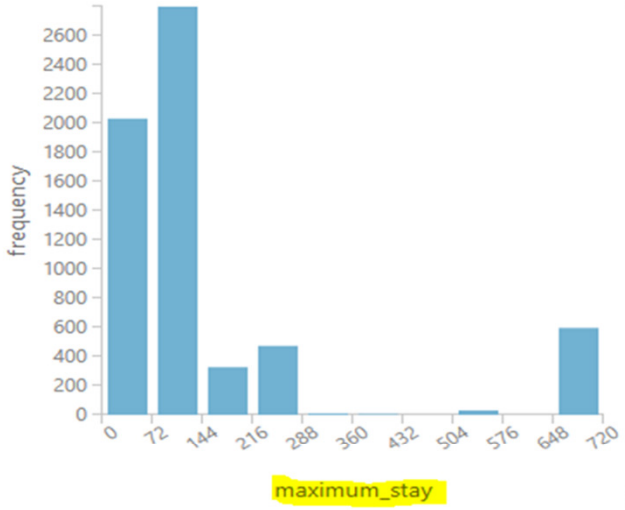
Swinburne Prototype Web Site

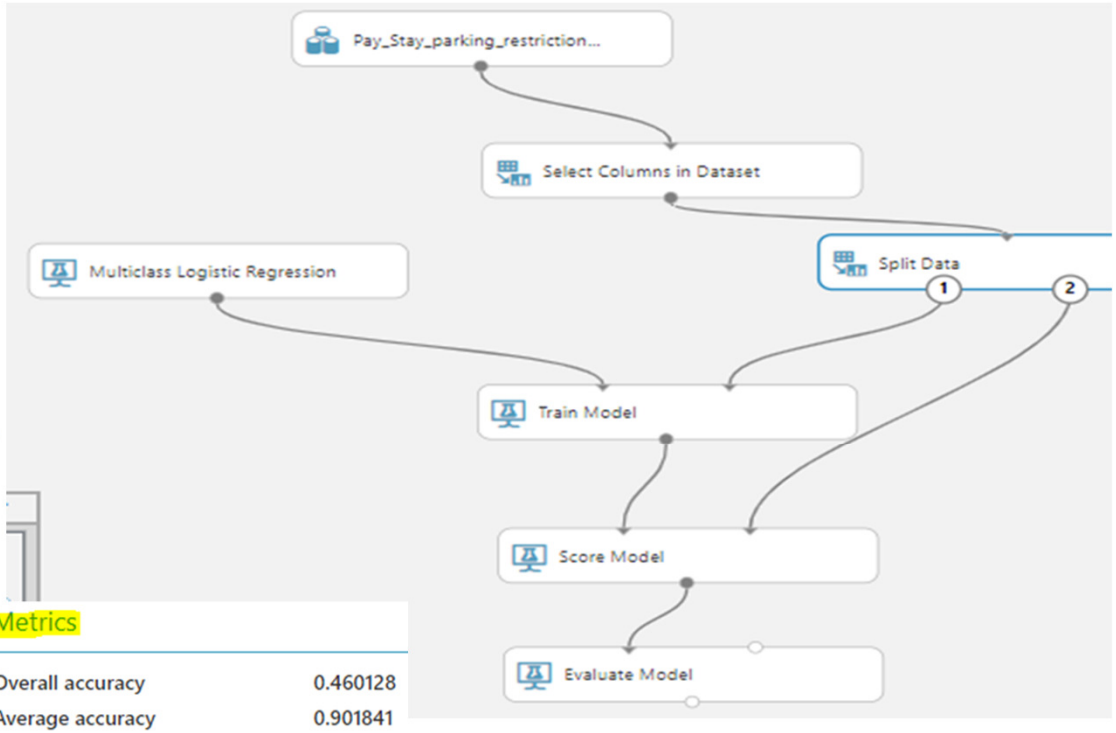


https://austinbitanga.wixsite.com/smartcity-sepa42

MS Azure Machine Learning

- Transport dataset analysed: COM parking





Metrics

Overall accuracy	0.460128
Average accuracy	0.901841
Micro-averaged precision	0.460128
Macro-averaged precision	NaN
Micro-averaged recall	0.460128
Macro-averaged recall	NaN

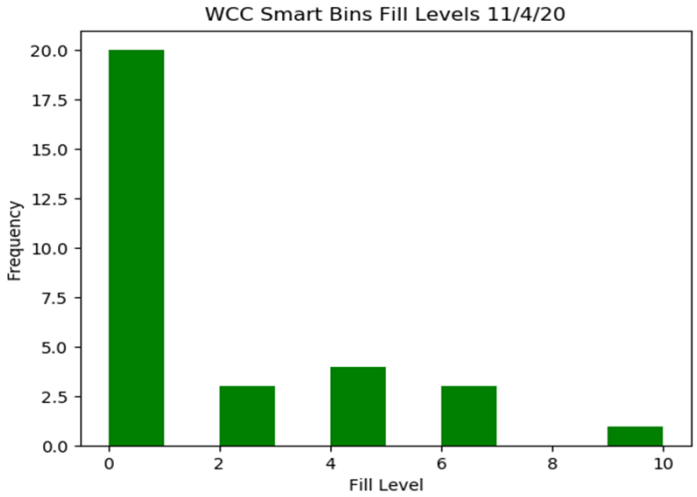
APIs for Web Services

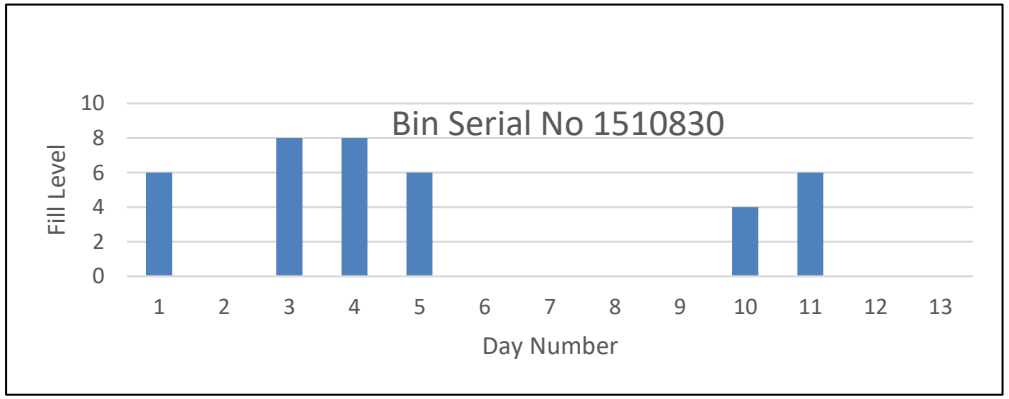
- API Calls, Protocols
  (REST, SOAP, GraphQL, etc.)
  https://ffeathers.wordpress.com/2014/02/16/api-types/
- URIs (URLs)
  - Uniform Resource Locator
  - Uniform Resource Identifier
- API Verbs:
  (http://www.restapitutorial.com/lessons/httpmethods.html)
  - GET, POST, PUT, PATCH, DELETE, Etc.

(from Roman Smolkin, https://www.youtube.com/watch?v=16PAM5a8PpY)

Emerging Project – Waste Management (1)

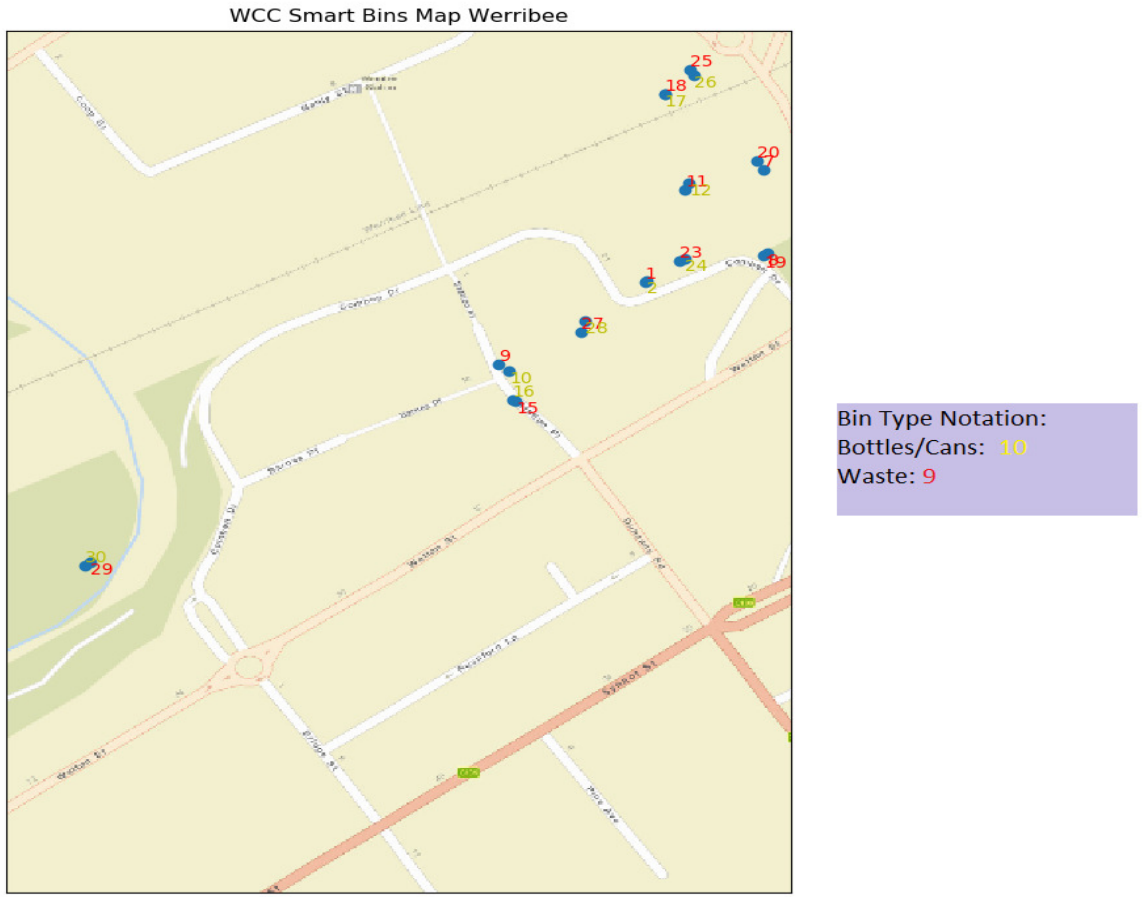
- Waste management for local councils
- Smart bin system deployed in Australia, eg City of Wyndham, Victoria
- Preliminary analysis





Emerging Project – Waste Management (2)

- Python *basemap/matplotlib* libraries used to draw map of smart bin sites and types in City of Wyndham
- Optimal bin emptying routes can be determined by algorithm

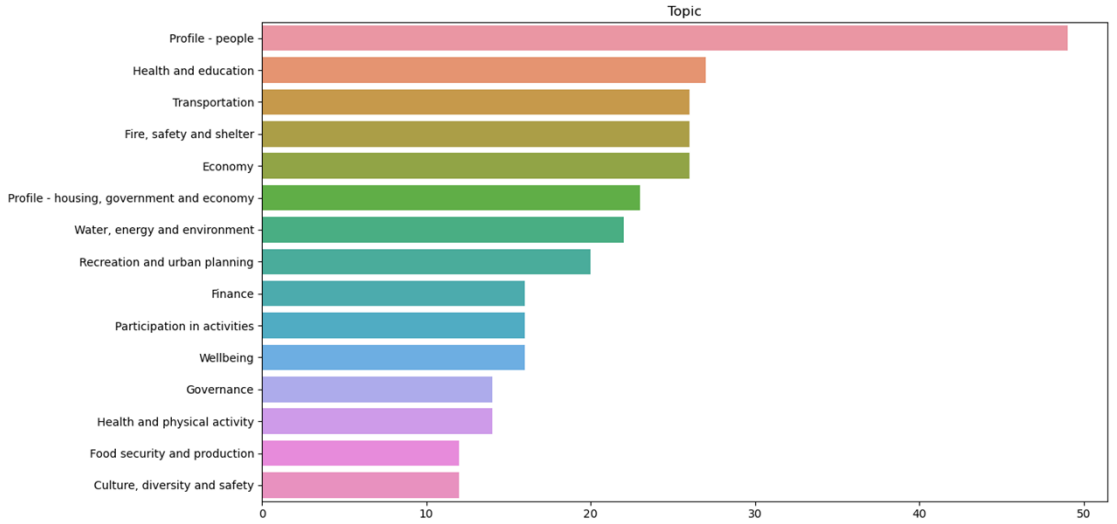


Emerging Project – Social Indicators (1)

- The City of Melbourne Social Indicators Survey conducted in 2018
- Collected data about the state of health, well-being, participation and connection of communities

CITY OF MELBOURNE

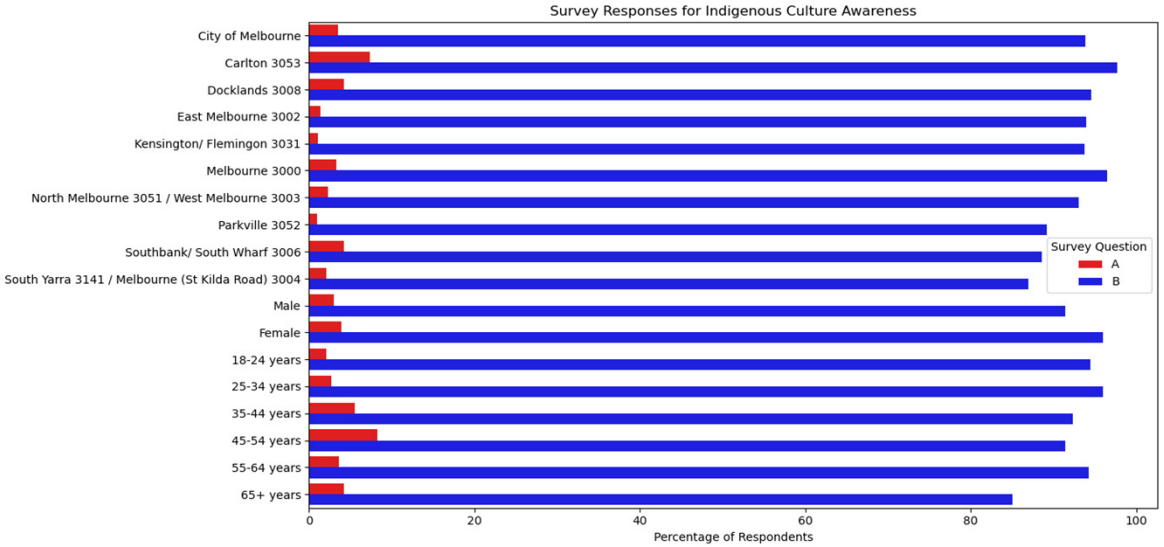
- **Topics for liveability survey**
- **Number of questions of each type**



Emerging Project – Social Indicators (2)

- Question on names of two local indigenous tribes

  - Fraction who could identify tribes (A)
  - Fraction who rated relationships with indigenous peoples as significant (B)



Conclusions

- Australian governments at local, state, and federal levels are exploiting open datasets to enable communities to benefit from smart city initiatives
- These datasets are amenable to big data analytics and machine learning techniques to assist governments to improve the efficiency of operations and planning
- A pilot study using python data analytics on a dataset of Victorian road accidents was described
- Student projects sponsored in two Melbourne Universities were described with some indicative results