The Impact Of COVID-19 On Public Transport In The Sydney Metropolitan Area

Thomas Hall - 46495193
Sam Huynh - 45431507
Kavindu Nilhan Rankothge - 46172106
S S Herath Mudiyanselage Nehan Irushika Senevirathna - 45848009

Aim of the project

This project explores the impact that COVID-19 has had on the Sydney transport system through the following:

- Analysing whether lockdown regulations had a negative impact on the use of public transport
- Analysis into which mode of public transport had a greater decrease in usage over the pandemic period
- Understanding trends between different regions and looking for relationships between public transport usage and socio-economic status
- Exploring different machine learning models for predicting the future of public transport usage in the metropolitan area

Data Sources

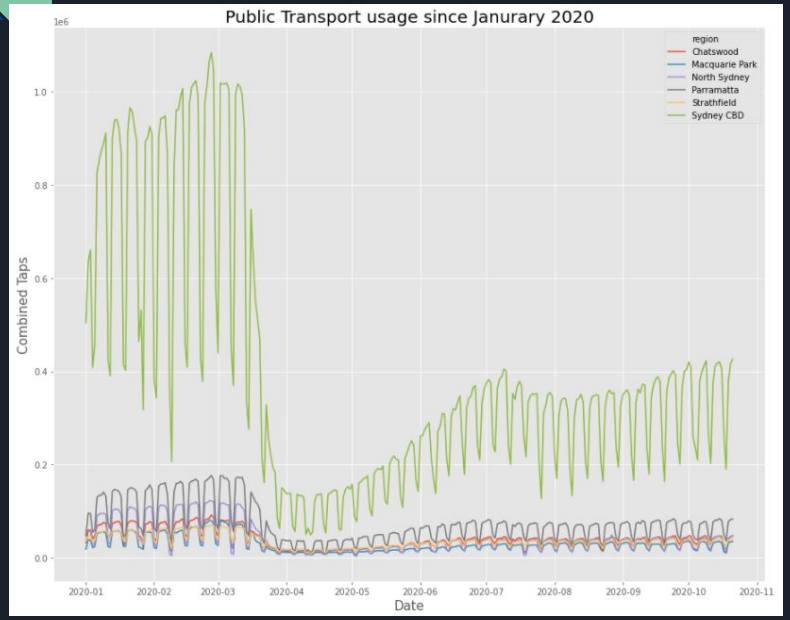
- Data Transport NSW, Opal-Patronage
  - The Opal Patronage dataset gave information into the number of tap ons and tap offs for each area
- Data NSW, Covid-19 cases throughout NSW
  - Data NSW dataset provides information on the number of covid cases throughout the pandemic
- Australian Bureau of Statistics, Socio-Economic status
  - The ABS dataset provided information into the SES of the regions we focussed on

Methods Used

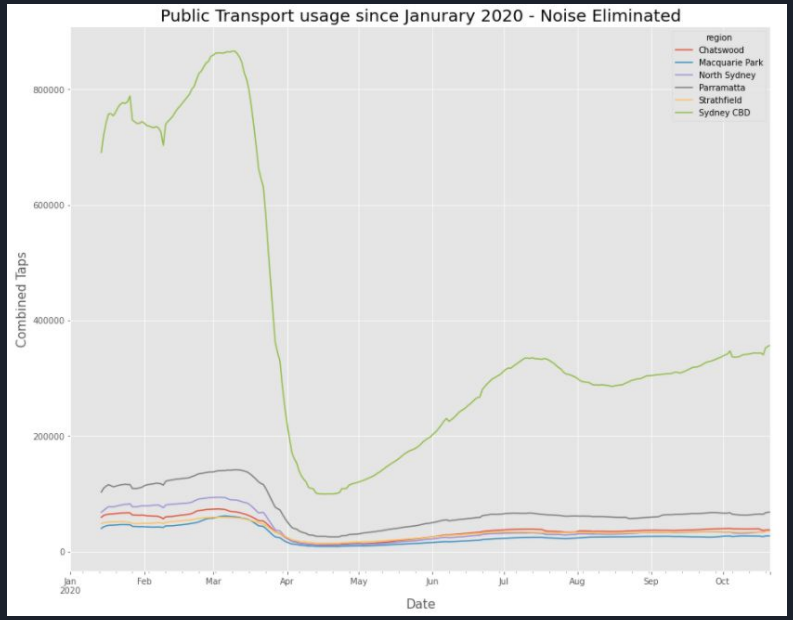
- Data preparation and descriptive statistics
  - Data was prepped and explored to help find relationships as well as provided a visual analysis of the datasets

- Predator-Prey Model
  a. Neural network model
     - Used to attempt a prediction of the next day cases and total transport usage
  b. Regression model for next day predictions
     - Used for next day prediction of cases and total transport usage
     - Used to forecast when public transport may return to normal

Investigating Opal Data

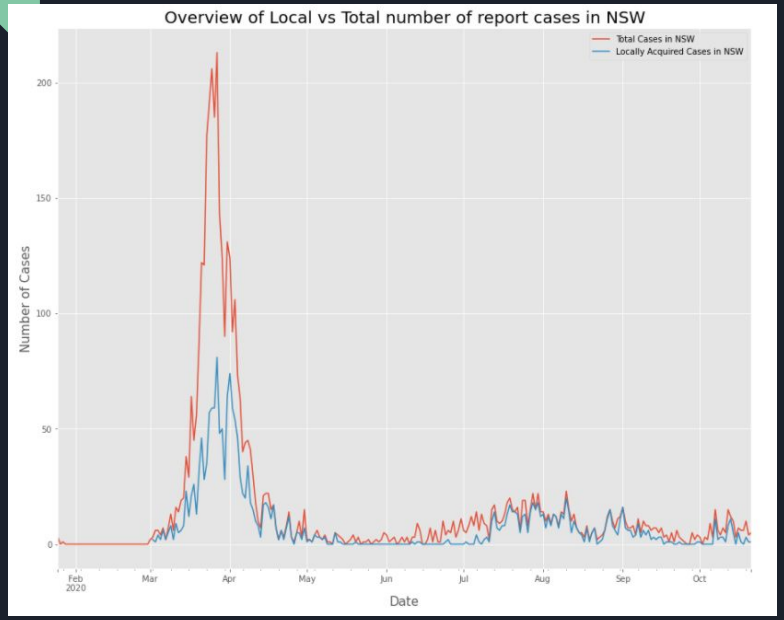


Without Rolling Average

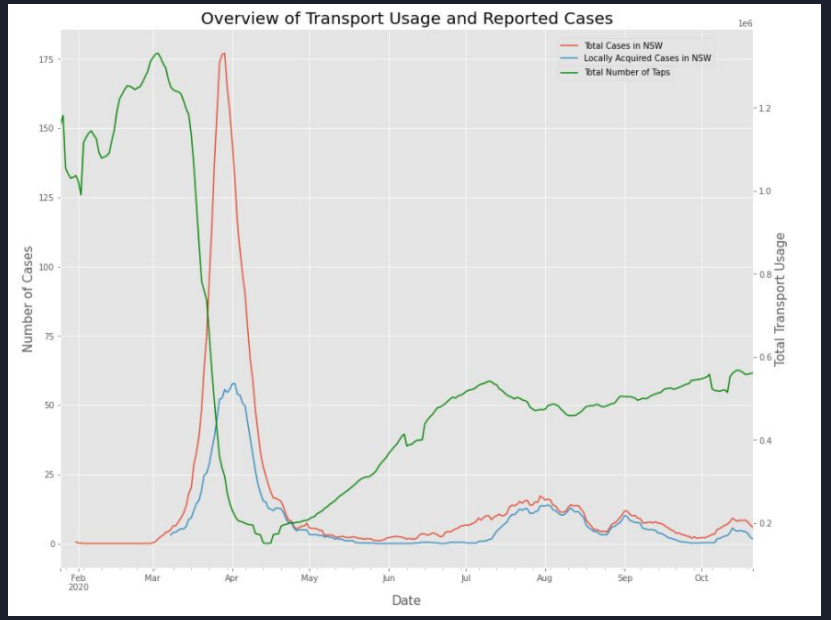


With Rolling Average

Covid Cases Graph and Overlay Graph

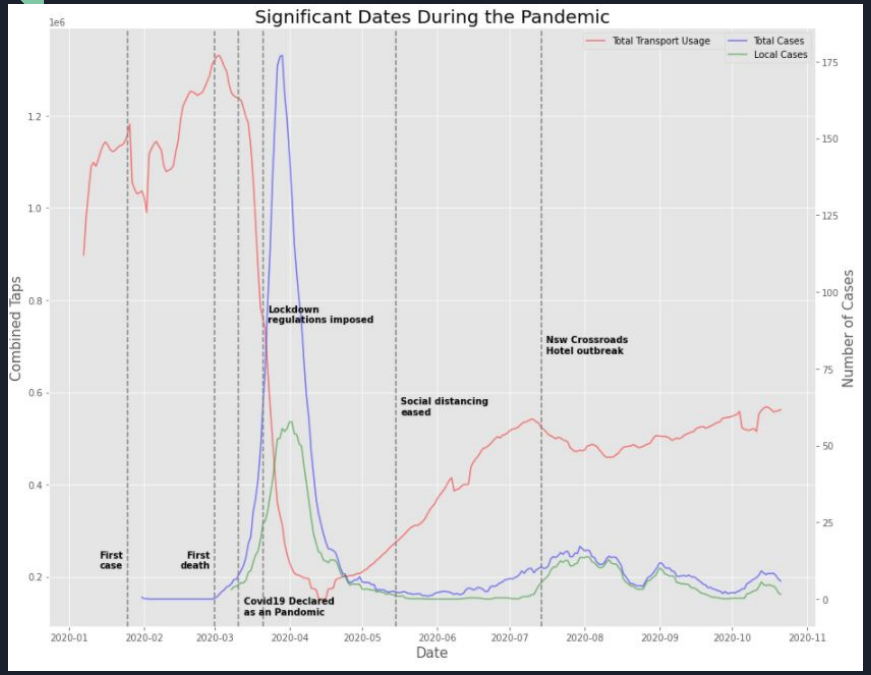


Covid data without Rolling Average



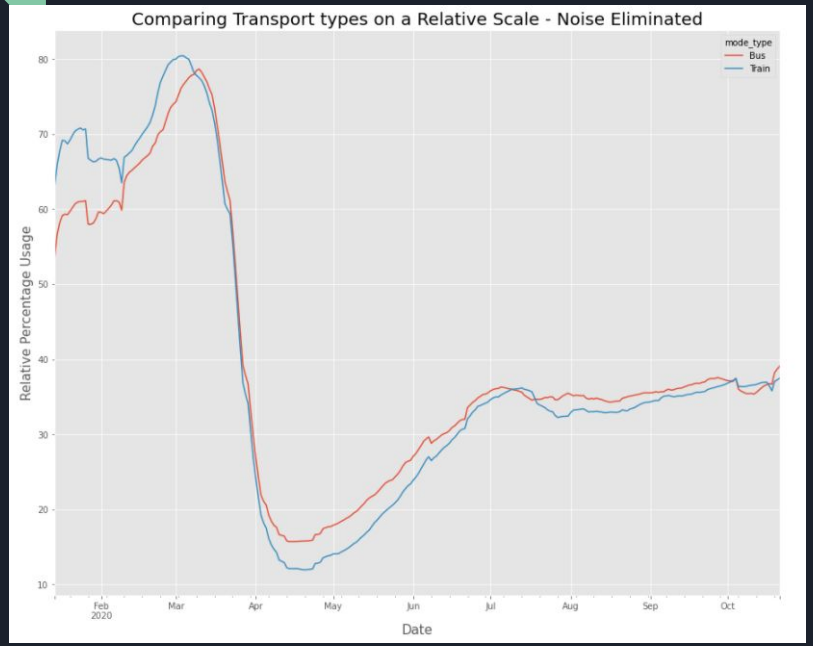
Overlay of both graphs with rolling average

Analysing Significant Events with the data

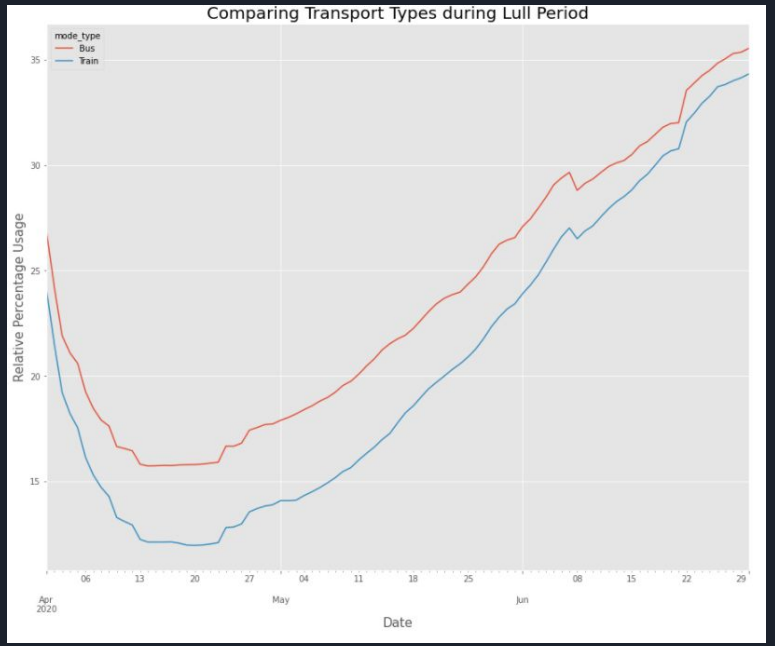


Key Events	Date
First Case in Australia	25/01/2020
First Death in Australia	01/03/2020
Covid19 Declared as a Pandemic	11/03/2020
Lockdown regulations imposed	21/03/2020
Social distancing eased	15/05/2020
NSW Crossroads Hotel outbreak	14/07/2020

Comparing Public Transport Types



Bus and Train Usage with rolling average



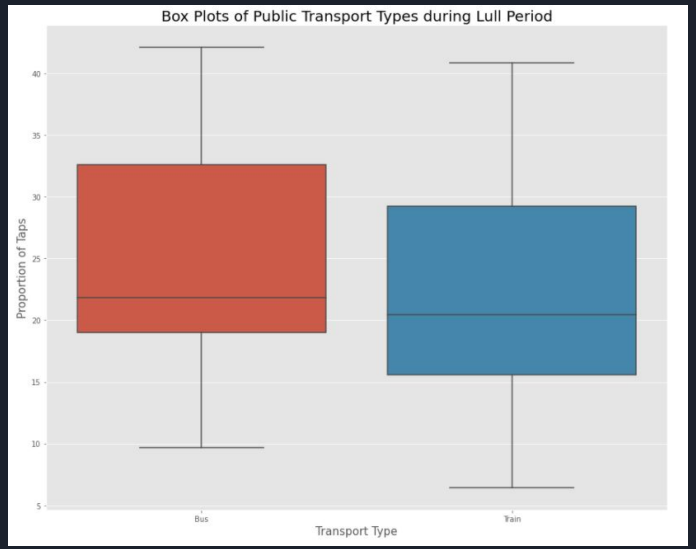
Bus and Train Usage during the Lull Period
01-04-2020 to 01-07-2020

Further Investigating Bus and Train Transport Usage

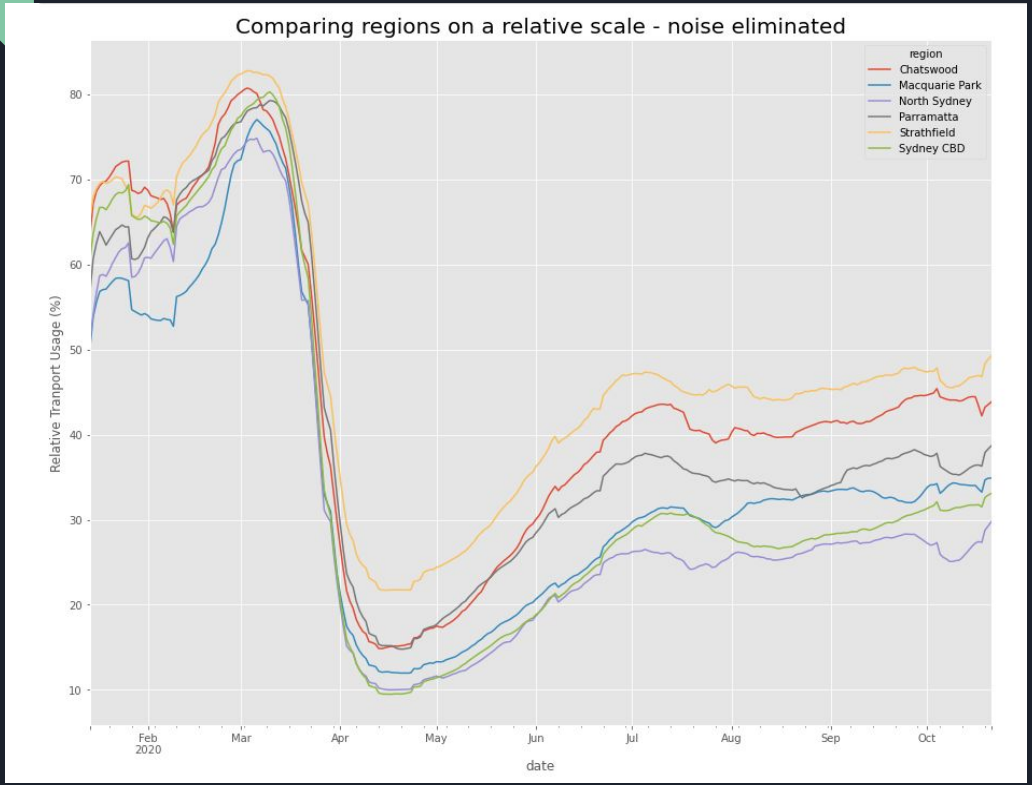
Average of Bus and Train Proportion taps

	Proportion_taps
mode_type	
Bus	24.788626
Train	21.825513

Box Plots of Bus and Train



Analysis of Public Transport Usage

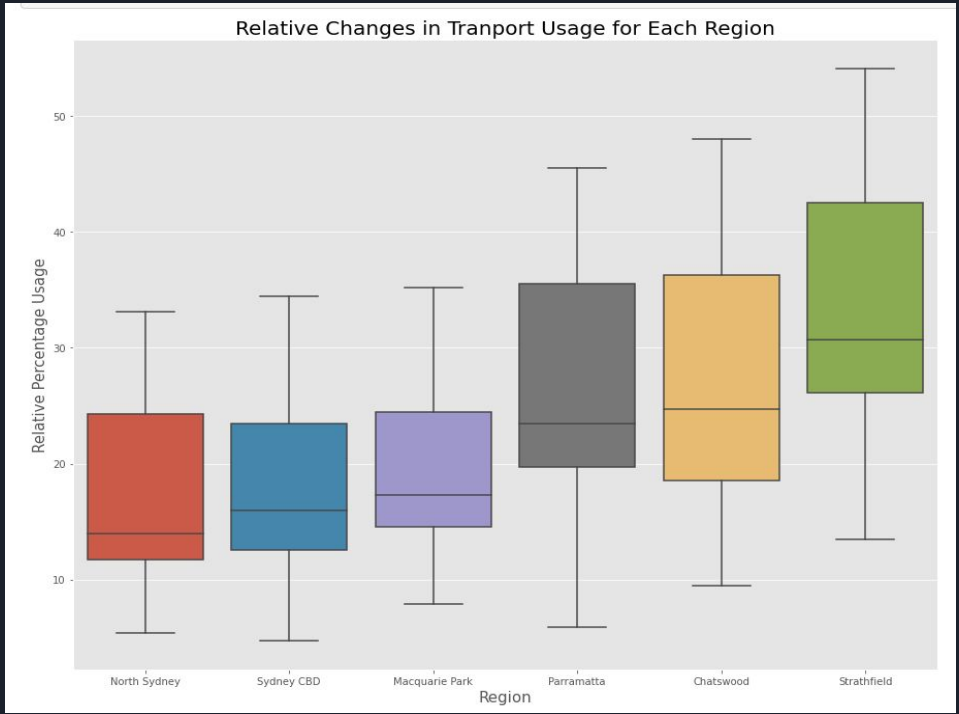


With Rolling Average

Transport Usage in Lull Period

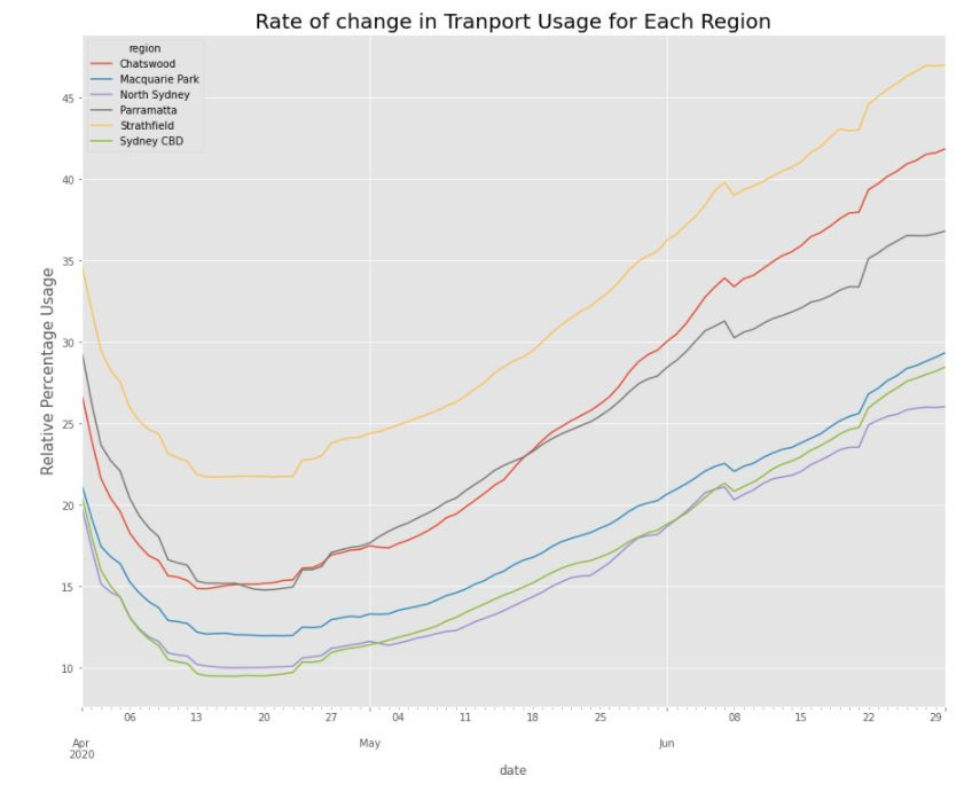
	Proportion_taps
region	
North Sydney	17.078462
Sydney CBD	17.694835
Macquarie Park	19.261154
Parramatta	25.532308
Chatswood	27.070385
Strathfield	33.205275

Table of Tap ons within the period 01-04-2020 TO 01-07-2020



Bar Plot comparing relative change in Transport

Rate in Change of Transport Usage



Line Plot comparing relative change in Transport

Identifying the Socio Economic status of an Area

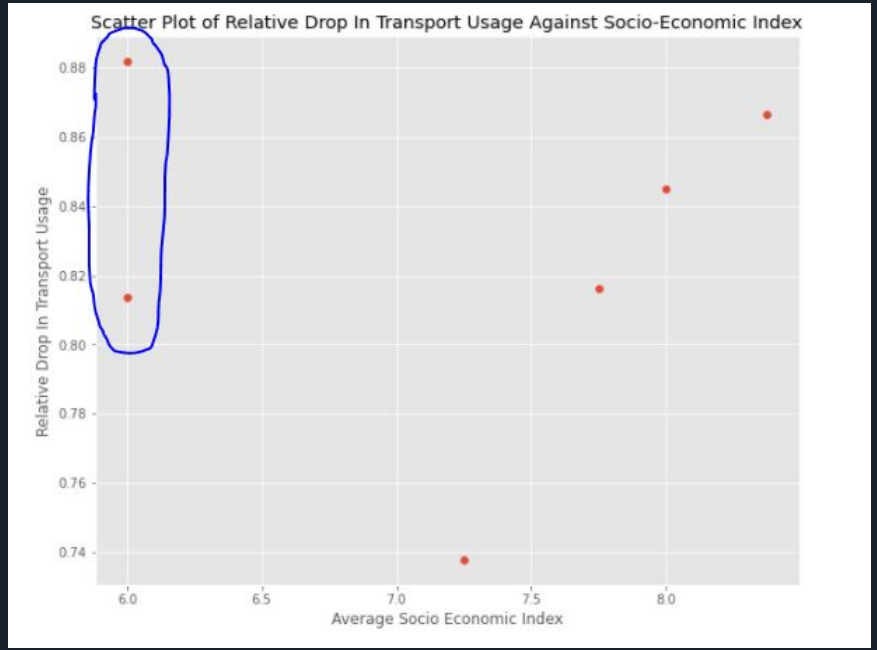
Area	Relative SE Dis	Relative SE Adv & Dis	Index of Economic Resources.1	Index of Education and Occupation.1
CBD	4.0	9.0	1.0	10.0
Chatswood	8.0	10.0	3.0	10.0
Macquarie_Park	8.0	10.0	4.0	10.0
North_Sydney	10.0	10.0	3.5	10.0
Parramatta	6.0	8.0	1.0	9.0
Strathfield	7.0	9.0	4.0	9.0

Table uses a Decile Scoring method - Each Region is put into 1 of 10 groups

Comparing Socio Economic Index with Drop in Transport Usage

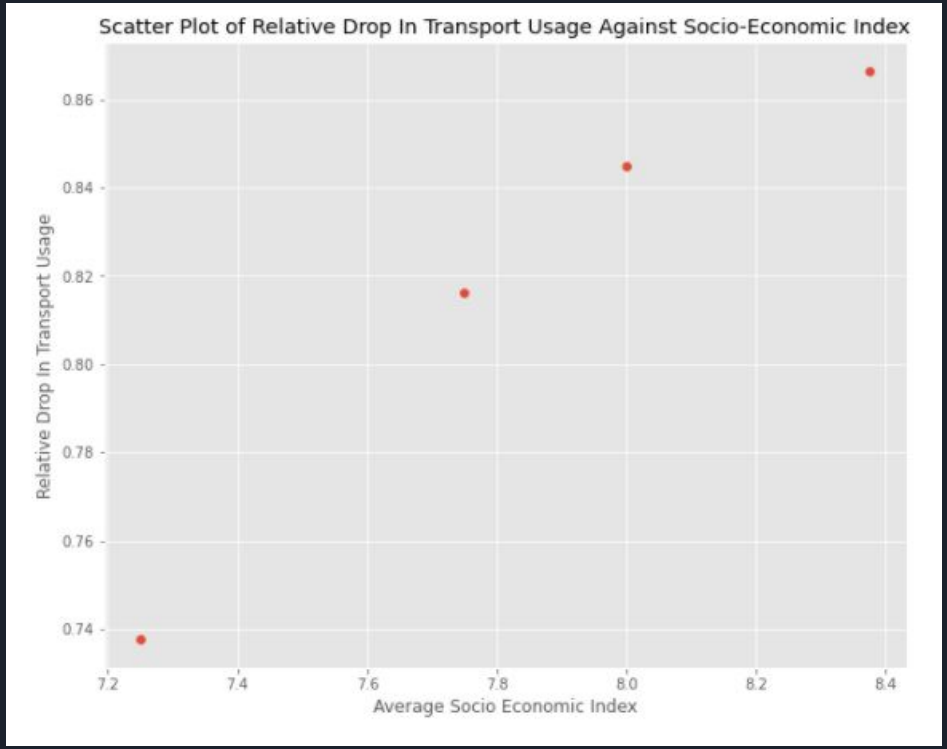
Area	average_index	transport_drop
CBD	6.000	0.881981
Chatswood	7.750	0.816140
Macquarie_Park	8.000	0.844832
North_Sydney	8.375	0.866469
Parramatta	6.000	0.813627
Strathfield	7.250	0.737721

Table Comparing the two variables



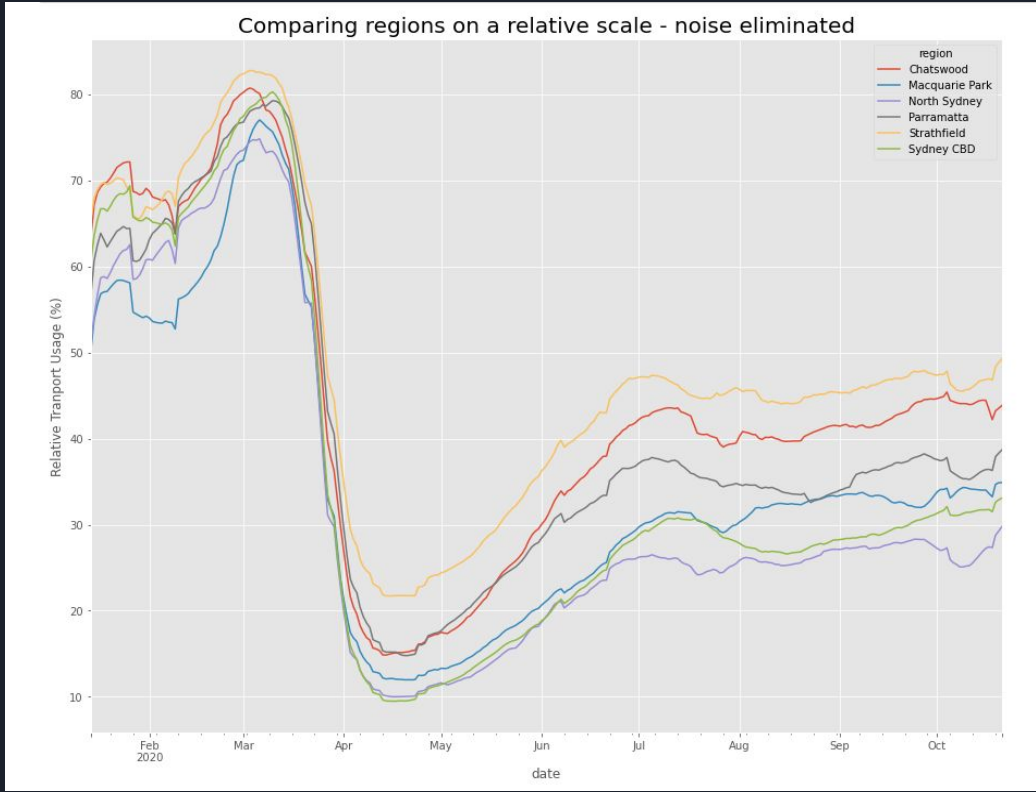
Graph identifying possible anomalies

Comparing Socio Economic Index with Drop in Transport Usage (Without anomalies)



Graph referencing changes

Percentage of Tap Ons per Region



Predator-Prey Model

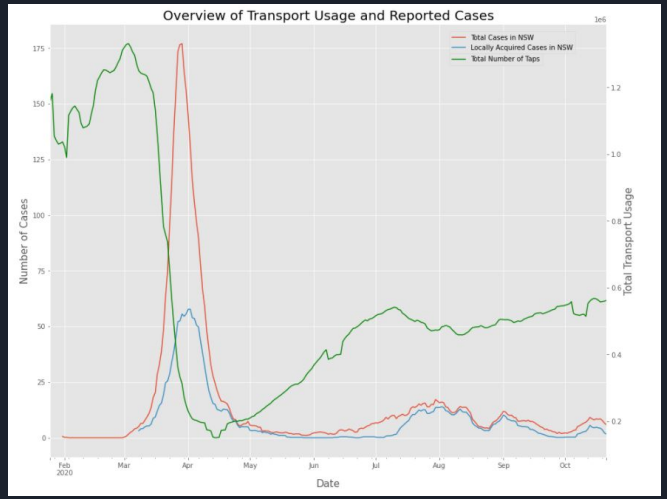
- Two populations
- Each population changes based on its own population size and the other population size.

$$\frac{dx}{dt} = x \cdot f(x, y)$$

$$\frac{dy}{dt} = y \cdot g(x, y)$$

Why use a model like this?

- It is a building block for advanced models in biology
- Data has similar characteristics between other predator-prey models



Feature Engineering

These equations take into account the changes in population size. We need to do that for our models too!

$$\frac{dx}{dt} = x \cdot f(x, y)$$

$$\frac{dy}{dt} = y \cdot g(x, y)$$

Our data:

- 14 day averages for taps and COVID cases

New features

- Changes in averages from 5, 7, 14, 21, 28 days ago
- The next 14 day averages

Prediction Models

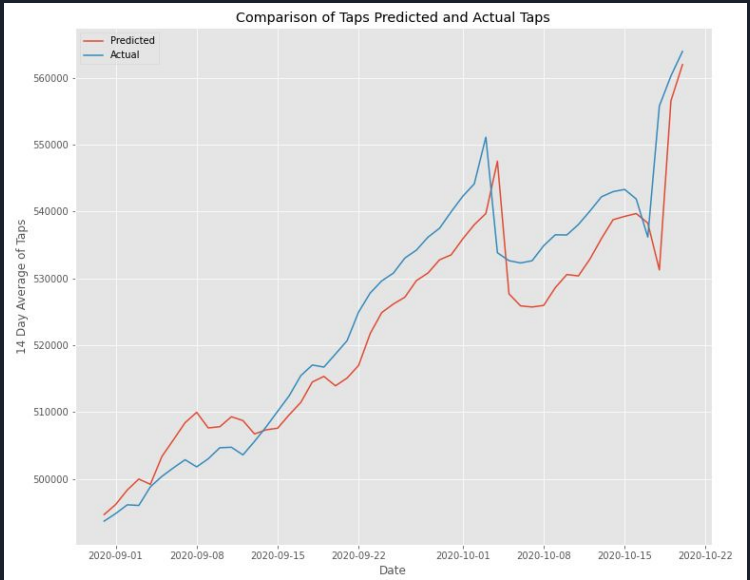
Linear Regressor

MLP Regressor

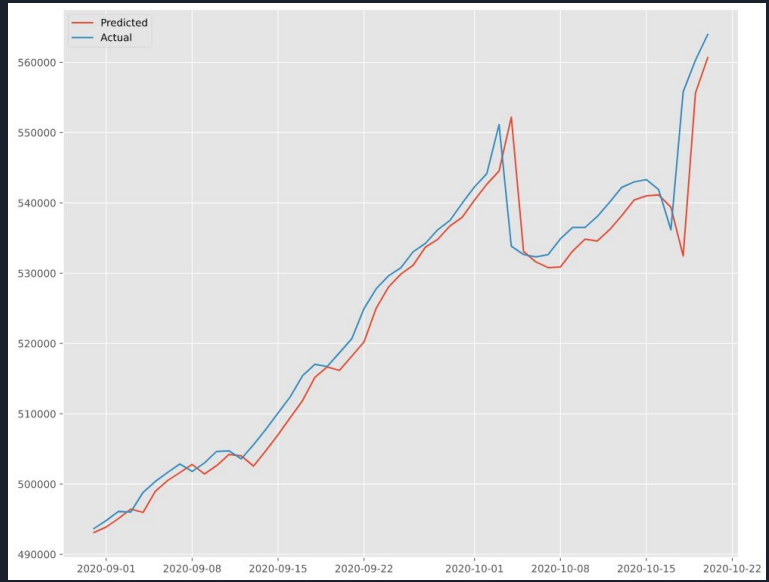
Have to create models for the next 14 day average of taps AND the next 14 day average of COVID cases

Comparing taps predicted for both models

Linear Regressor

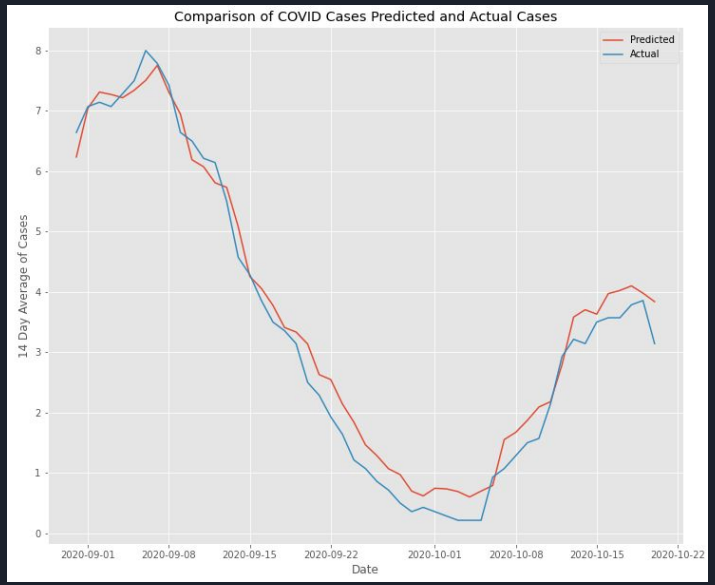


MLP Regressor

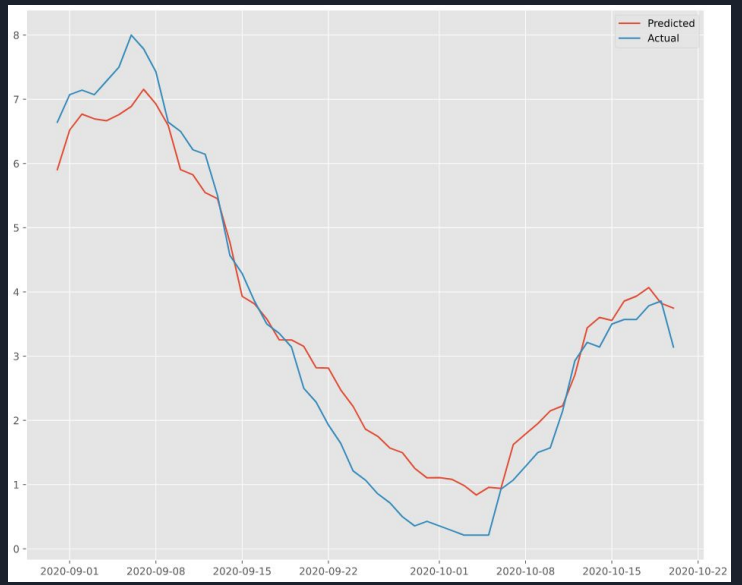


Comparing cases predicted for both models

Linear Regressor

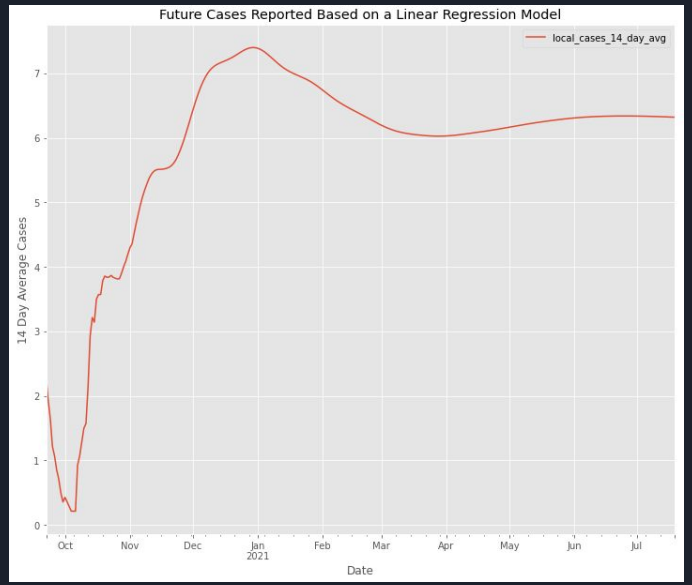


MLP Regressor

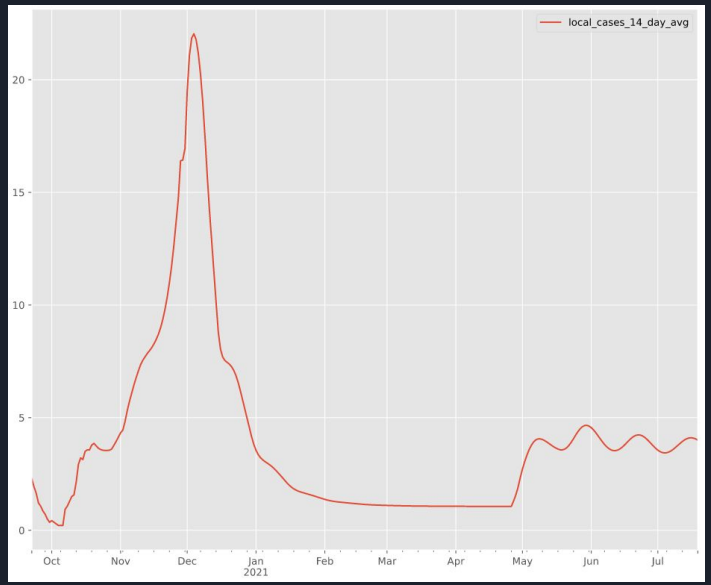


Predicting cases further into the future

Linear Regressor

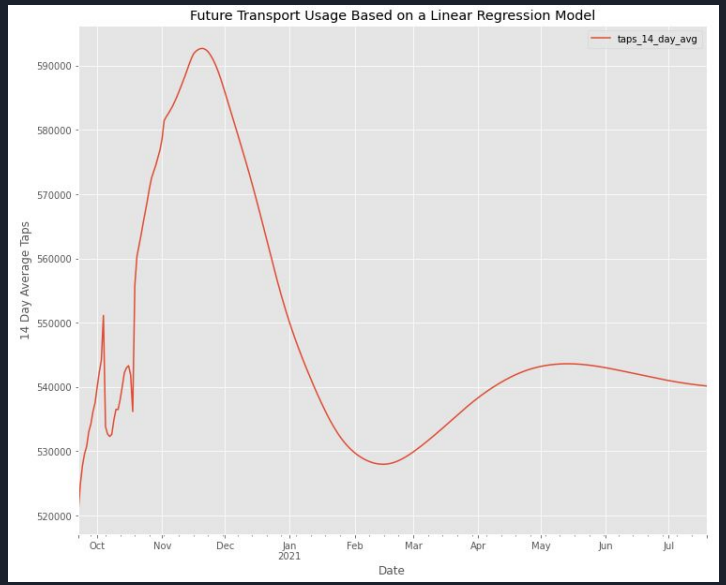


MLP Regressor

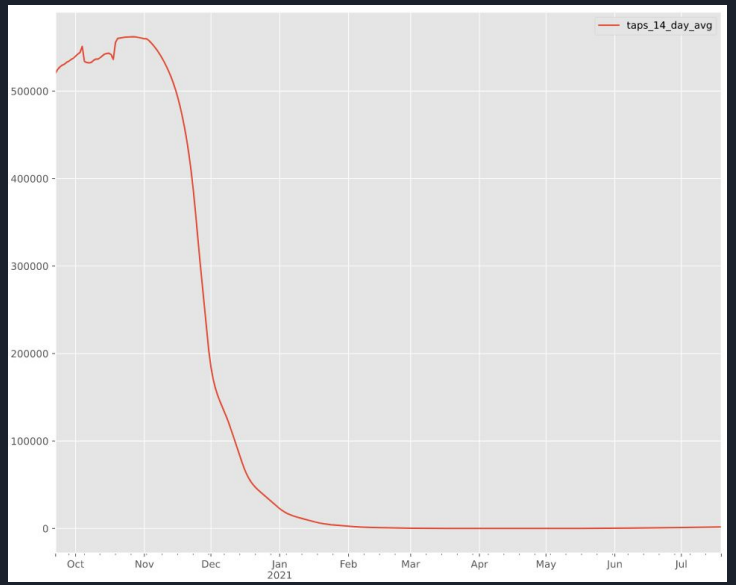


Predicting taps further into the future

Linear Regressor



MLP Regressor



Limitation of the model

- Assumption of populations
  - Model assumes the populations are dependent on each other
  - No external influences
- Features used
  - Only features based on transport usage and COVID cases reported are used in the model
  - No feature about lockdown regulations, telecommuting for work, etc.

The amount of people using public transport and COVID cases reported are two populations that are related but not entirely dependent on each other.

Better models would take into account geographical features, socio-economic features and much more.