COVID19 And Travel (CAT)

Managing Urban Mobility and COVID Spread using Data Science

Bell-The- CAT (COVID19 And Travel)

To bell the CAT means to attempt or agree to attempt an impossibly difficult task of balancing mobility and infection spread, that if achieved, will benefit the entire community. This is my attempt.

Archit Narain – a 7th grader from Bengaluru, India

Overview

Problem

- The COVID crisis has forced public transport to completely shut down
- And has prevented people from moving around for their livelihood

Objective

- Leverage data science to enable effective operation of Bengaluru Metropolitan Transport Corporation (BMTC) as well as contain spread of infection in Bengaluru City

Around 35,00,000 commuters daily

41% of the population

Archit Narain

I saw the news about **COVID** and noticed that it was **getting serious every day** and saw that people were **struggling** to make ends meet and thought if there was a way of **moving around effectively and containing** the Covid spread at the same time

The Tool

Why

Able to analyze COVID impact in **every** Bruhat Bengaluru Mahanagara Palike (BBMP) **Ward** due to **operation** of BMTC buses

How

The tool uses **SIR model** to predict the spread of the coronavirus and helps in analyzing the extent of spread due to urban mobility

What: The **visual** analysis tool **identifies** BMTC **routes to run**, based on factors like **transmission rate, recovery period, infection rate and peak period**. It generates the graphs & reports for Bengaluru and its Wards.

Archit Narain

Data Science

Data science is the study of data. It involves developing methods of recording, storing, and analyzing data to effectively extract useful information. The goal of data science is to gain insights and knowledge from any type of data — both structured and unstructured

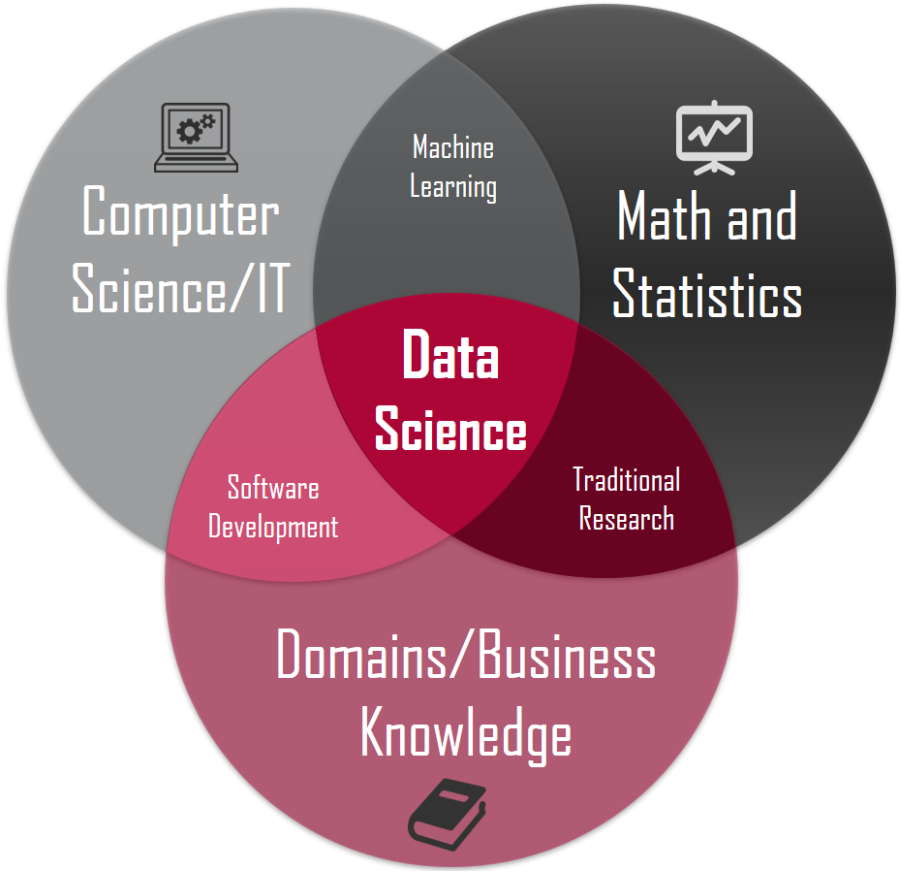


Image Source: https://www.inf.elte.hu/en/content/data-science-in-computer-science-msc.t.1732?m=361

Archit Narain

Building the Tool

Important Considerations

Belling-The-CAT

- Choose the correct epidemiological model [SIR vs SEIR]
- Origin-Destination (OD) for Urban Mobility - BMTC
- BBMP Ward data: Population, Infections, etc

Build

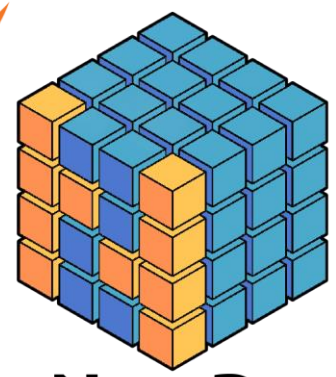
Steps …

- Get BMTC data (Routes, Bus Stops, …)
- Get BBMP ward details
  - Zone, Number, Population, Geo Boundaries, …
- Get ward wise infection data
- Get ward wise commercial activity (tax collection)
- Generate Origin-Destination (OD) data for urban mobility
- Program SIR model in Python
- Run simulations, and validate. Deploy
- Develop dashboard for visual analysis

Tools

Organize and Execute

- Chose **Python**
- Setup **Jupyter** Notebook
  - Data Processing
  - Programming and Testing
  - Run Simulations
- **Data** from **Internet**
- Python **Data Science** modules (NumPy, Pandas, …)
- **Dashboard** modules (Dash, Plotly, …)



Public Transport Data

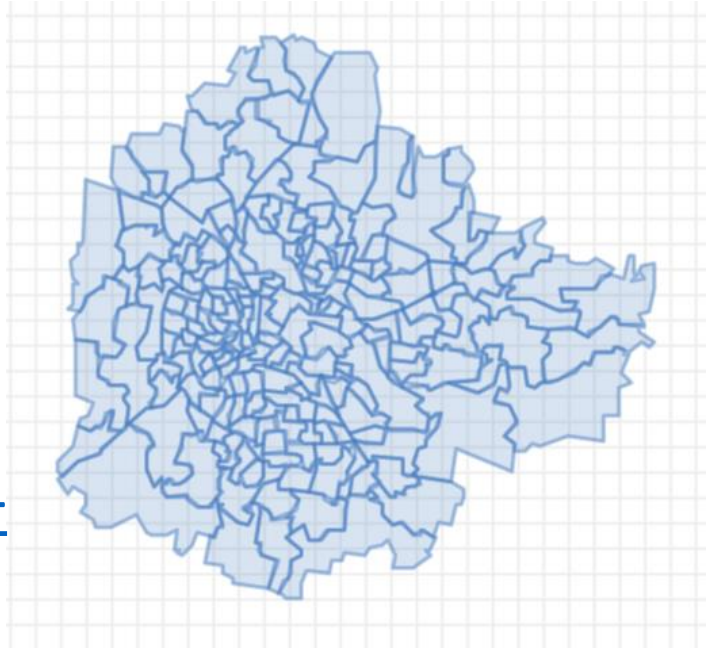
Get BMTC Routes, Bus Stops & Geo Locations

- Found https://github.com/geohacker/bmtc
- Data available for 2045 routes
- Generated all possible route combinations
- Found 1053486 Origin-Destination - ODs

Archit Narain

City Data

Get BBMP data



- Found https://github.com/openbangalore/bangalore
- Got the ward wise population data from https://indikosh.com/city/708740/bruhat-bengaluru-mahanagara-palike
- For the **1053486** ODs, Identified Origin & Destination wards
  - By getting the ward geo boundaries
  - Checking if the O or D bus stop geo location is in a ward
- Removed ODs, where either O or D ward number was missing
- Finally for the BBMP's 198 wards, found **656913** values
- Got **1313826** ODs for both directions

Infection Data

Get BBMP ward wise data

- Challenge to find updated data ward wise
- And found this one to be the closest https://indianexpress.com/article/cities/bangalore/covid-19-101-cases-in-bengaluru-so-far-heres-the-list-of-wards-affected-6368503/

Number of people traveling in BMTC?

Understand Ward Wise Activity to determine Mobility

- **Actual OD mobility data is not available**
- So, assumed that BBMP property tax collection can give an idea of commercial activity, and hence mobility
- Found http://bbmp.gov.in/documents/10180/13438554/2018-2019+PT+Collections.pdf/d61e8b49-e90e-4c03-9604-a47ff15a452e
- And used this to determine proportion of people moving to a ward

	ward_no	revenue	proportion
0	1	594.90	0.308801
1	2	446.67	0.231858
2	3	904.79	0.469659
3	4	1092.02	0.566847
4	5	1675.88	0.869918
..	...	...	...
193	194	691.50	0.358945
194	195	1044.82	0.542346
195	196	411.90	0.213809
196	197	1021.87	0.530433
197	198	1892.88	0.982558

Final OD Data Set

Preparation

- OD data for every ward with route numbers and OD counts
- Assumed at least 1 person travels in every OD combination in a day
- So, 1313826 can travel
- Group them ward wise
- And add population data

	ward_no	origin_ward	route_num	route_counts
0	1.0	1.0	281	6
1	1.0	1.0	281G	12
2	1.0	1.0	281K	12
3	1.0	1.0	282	12
4	1.0	1.0	282A	12
...	...	...	...	...
251441	198.0	198.0	502HC	30
251442	198.0	198.0	502V	2
251443	198.0	198.0	FDR-7	12
251444	198.0	198.0	G-6	12
251445	198.0	198.0	MP-501A	6

	ward_no	population	origin_ward	origin_population	route_num
0	1	34783	1	34783	281
1	1	34783	1	34783	281G
2	1	34783	1	34783	281K
3	1	34783	1	34783	282
4	1	34783	1	34783	282A
...	...	...	...	...	...
251441	198	50440	198	50440	502HC
251442	198	50440	198	50440	502V
251443	198	50440	198	50440	FDR-7
251444	198	50440	198	50440	G-6
251445	198	50440	198	50440	MP-501A

Archit Narain

Final OD Data Set

Clean up and validation

- Assuming 1 person travels in OD per day, can be misleading, as it does not show where people travel most
- So, computed possible number of people traveling based on commercial activity

	ward_no	proportion_factor	origin_ward	total	movement
0	1	0.683457	1	608	415
1	1	0.683457	2	105	71
2	1	0.683457	3	227	155
3	1	0.683457	4	636	434
4	1	0.683457	5	331	226
...	...	...	...	...	...
18740	198	0.771239	192	80	61
18741	198	0.771239	194	60	46
18742	198	0.771239	196	50	38
18743	198	0.771239	197	1578	1217
18744	198	0.771239	198	3238	2497

Modelling

SIR model

- The Susceptible-Infectious-Recovered (SIR) model is a mathematical model of epidemics
- Programmed the model
- Ran simulations
- Validated outputs
- Determined β, γ , α ranges

$$S_{j,t+1} = S_{j,t} - \frac{\beta_{j,t} S_{j,t} I_{j,t}}{N_j} - \frac{\alpha S_{j,t} \sum_k m^t_{j,k} x_{k,t} \beta_{k,t}}{N_j + \sum_k m^t_{j,k}}$$

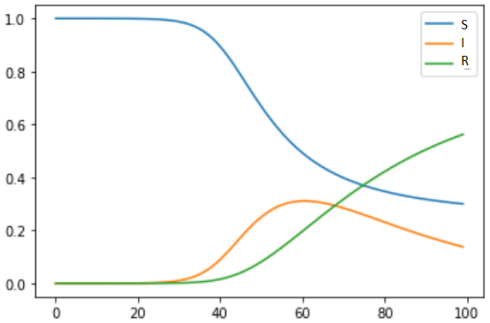
$$I_{j,t+1} = I_{j,t} + \frac{\beta_{j,t} S_{j,t} I_{j,t}}{N_j} + \frac{\alpha S_{j,t} \sum_k m^t_{j,k} x_{k,t} \beta_{k,t}}{N_j + \sum_k m^t_{j,k}} - \gamma I_{j,t}$$

$$R_{j,t+1} = R_{j,t} + \gamma I_{j,t},$$

This formula calculates the S, I and R of the model, where

S - Likely or liable susceptible population to be infected
I - Affected with a virus.
R - Return to a normal state of health or removed

β (transmission rate) - The rate at which a virus spreads
γ (recovery rate) – The rate at which a person recovers
m - Number of people moving
K - origin
J – Destination
α - Public transport factor
t - time
N – Population at K and J
X – fraction of infected people at K and at time t



Archit Narain

Dashboard

Heatmap showing mobility intensity (OD counts)

Use this slider to change the percentage of routes operated

Graphs, change accordingly to the inputs

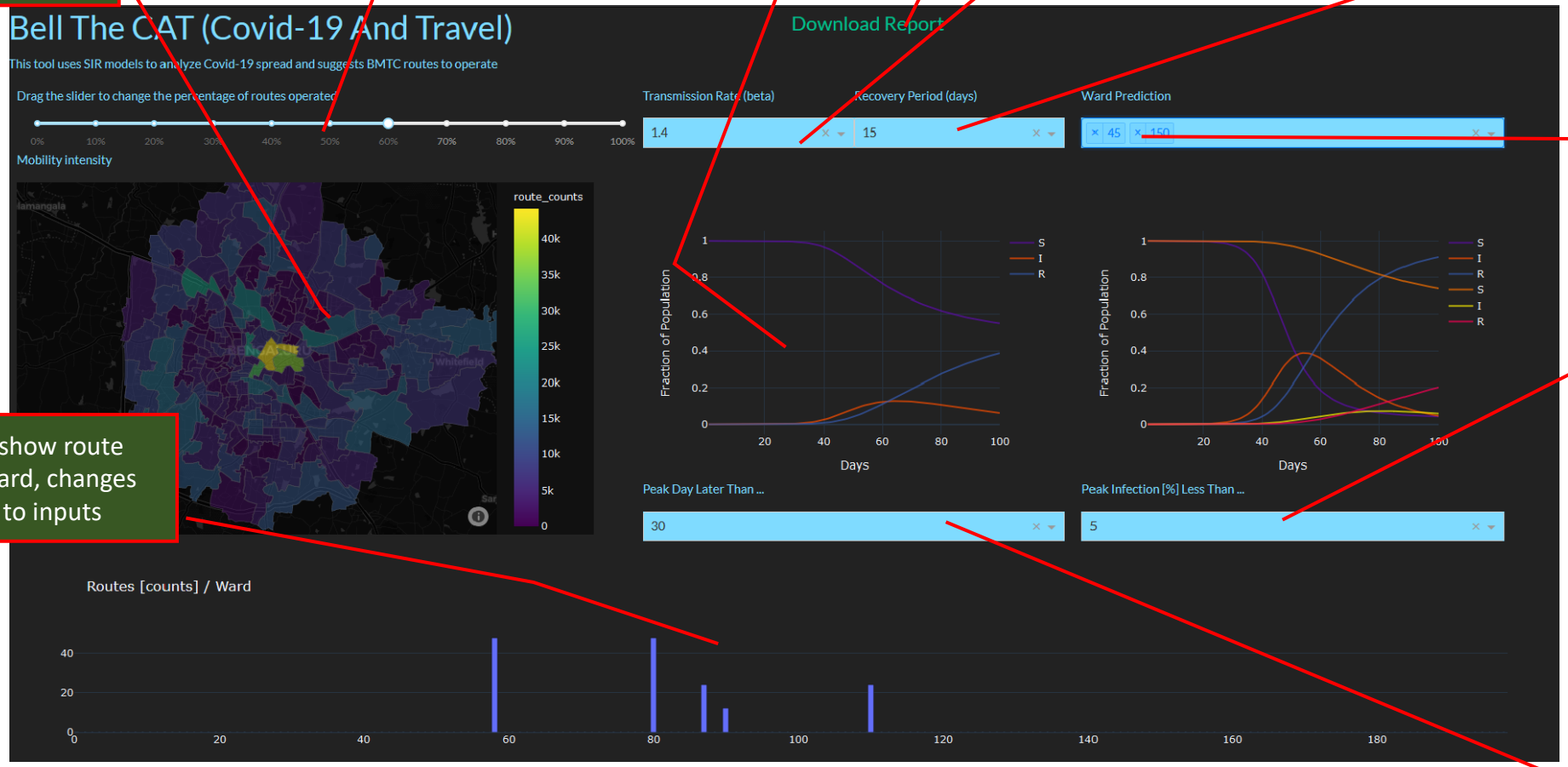
Shows fraction of population (SIR) over 100 days

Download Report

Use this drop down to change the Transmission Rate

Use this drop down to change the Recovery Period

Use this drop down to specify the ward you want to view



Use this drop down to change the Peak Infection

Bar graphs, show route counts per ward, changes according to inputs

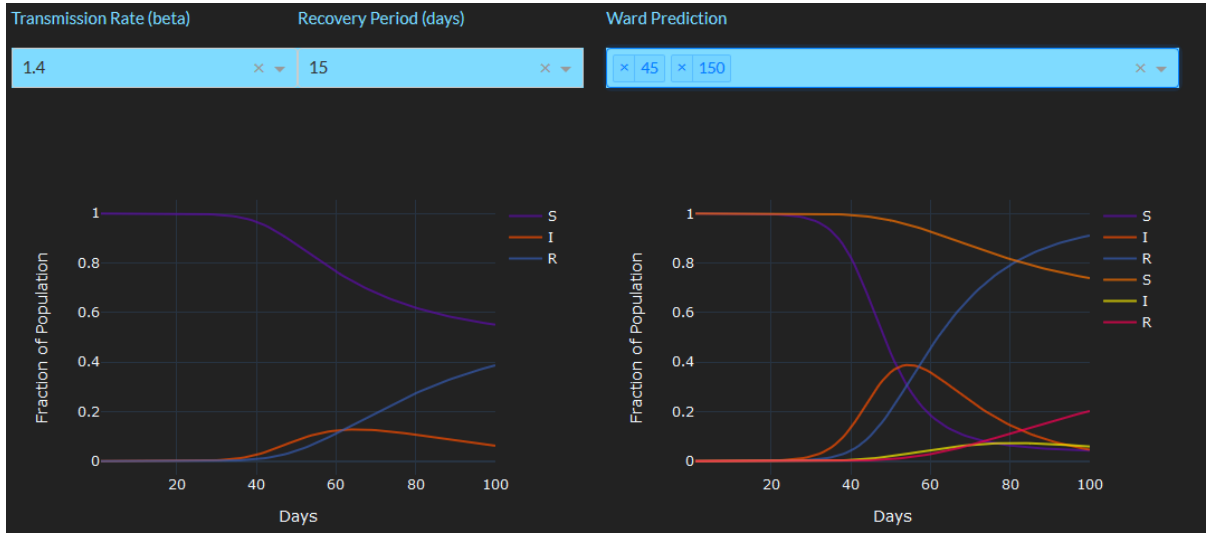
Use this drop down to change the Peak Day

Archit Narain

Highly Interactive Dashboard

Analyze Impact

- Change Transmission rates on a daily basis based on actuals
- Define thresholds and change peak projections to decide on conditions for running buses
- See impact across wards and overall for the input params
- Ex: The peak of ‘I’ with value ‘0.24’ indicates 24% infected



- S starts with 1 and reduces over time, and R, recovery increases. I increases, peaks and decreases

Archit Narain

Operation Report

Suggested routes to operate

- A CSV operation report is generated by the data science tool
- An example of the report is shown here
- This ward number of Originating wards (with bus stops) and their Destination wards. And, the route numbers are listed

The existing simple excel sheet that can be downloaded after analysis

WARD NO	ORIGIN WARD	ROUTE	OD COUNT
1	1	409A	12
1	2	401BR	1
1	2	401J	4
1	2	401JA	8
1	2	401JC	8
1	2	401T	2
1	2	401Y	8
1	2	404C	2
1	2	407C	10
1	2	500DA	2
1	3	401	2
1	3	401A	2
1	3	401AC	6
1	3	401AE	2
1	3	401BR	4
1	3	401J	16
1	3	401JA	24

Archit Narain

How the tool can be improved?

Actual Data Needed

- Ward wise infection numbers
- The actual movement of people between bus stops
- Expert help to refine models (SEIR, …)
- More conditions for decision making (to run buses)
- Reports needed? Enhance reports, on screen data display

The tool is published on the **Internet**, and can be accessed using the **link** below

https://bell-the-cat.herokuapp.com/

Thank You

References

- https://towardsdatascience.com/modelling-the-coronavirus-epidemic-spreading-in-a-city-with-python-babd14d82fa2
- https://towardsdatascience.com/infectious-disease-modelling-beyond-the-basic-sir-model-216369c584c4

About Me

I am a 7^{th} grader and like data science. And have been trying to explore how Python and data science tools can be used to solve problems

Archit Narain