

NAMES: Foong Min Wong

PROGRESS REPORT

Tasks Attempted

- Reformatted data sets (with year columns) so that they can be read into BigML for creating time series model

Conjectures/Observations Made

- According to the research papers related to dengue hemorrhagic fever (DHF) data analysis, the highest dengue fever outbreak location was Selangor based on 2010-2015 data and the age group of people who are 10 - 35 years old have the highest number of DHF cases among all age groups.
- Based on other dengue fever related research results, I predict that Selangor will have the highest dengue hemorrhagic fever cases from 2010 - 2017.

Pictures/Illustrations that would be relevant

The screenshot shows a Jupyter Notebook titled 'dengue-test' with a last checkpoint from Wednesday at 4:18 PM. The code in cell [58] renames the 'KUMPULAN UMUR' column to 'Age Group' for datasets dhf0 through dhf7. Cell [59] displays the 'dhf6' dataset as a table.

	Age Group	JOHOR	KEDAH	KELANTAN	MELAKA	NEGERI SEMBILAN	PAHANG	PERAK	PERLIS	PULAU PINANG	SABAH	SARAWAK	SELANGOR	TERENGGANU	LL
2	0-4	1	1	1	0	0	0	0	0	0	1	1	12	0	
3	5-9	0	1	1	0	2	0	0	0	1	1	0	13	0	
4	10-14	2	1	0	1	0	2	0	0	2	1	0	18	5	
5	15-19	5	1	1	2	5	1	3	0	1	2	0	38	3	
6	20-24	3	0	4	2	4	1	5	0	2	1	1	52	2	
7	25-29	7	0	3	0	2	0	2	0	0	0	2	50	1	

New formatted Dengue Hemorrhagic Fever Datasets on Jupyter Notebook

Tasks Yet To Do

- Forecast which age group and locations has more DHF cases
- Create slides and explain more on the research progress and codes for generating the new formatted data
- Generate time series model from the reformatted data on age groups and locations
- Analyze dengue hemorrhagic fever data downloaded from data.gov.my (2010-2017)

Additional Comments/Concerns