

NAMES: Foong Min Wong

PROGRESS REPORT

Tasks Attempted

- Researched on different statistical techniques for dengue data analysis
- Started a draft of slides showing the research methods used on the Dengue Hemorrhagic Fever (DHF) data

Conjectures/Observations Made

- According to the research papers related to dengue fever (which is slightly different from dengue hemorrhagic fever) data analysis, the highest dengue fever outbreak location was Selangor based on 2010-2015 data.
- 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 seven different data frames (dhf0 to dhf7). Cell [59] displays the 'dhf6' data frame, which is a table of dengue hemorrhagic fever cases across various Malaysian states and age groups.

	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	

Reading and formatting Dengue Hemorrhagic Fever data on Jupyter Notebook

Tasks Yet To Do

- Explore more on K Nearest Neighbor Algorithm and try to apply it on the dengue dataset
- Forecast which age group and locations has more DHF
- Create slides and explain more on the research progress and codes when doing data formatting
- Analyze dengue hemorrhagic fever data downloaded from data.gov.my (2010-2017)

Additional Comments/Concerns