

Proud Cockroaches



Second phase report

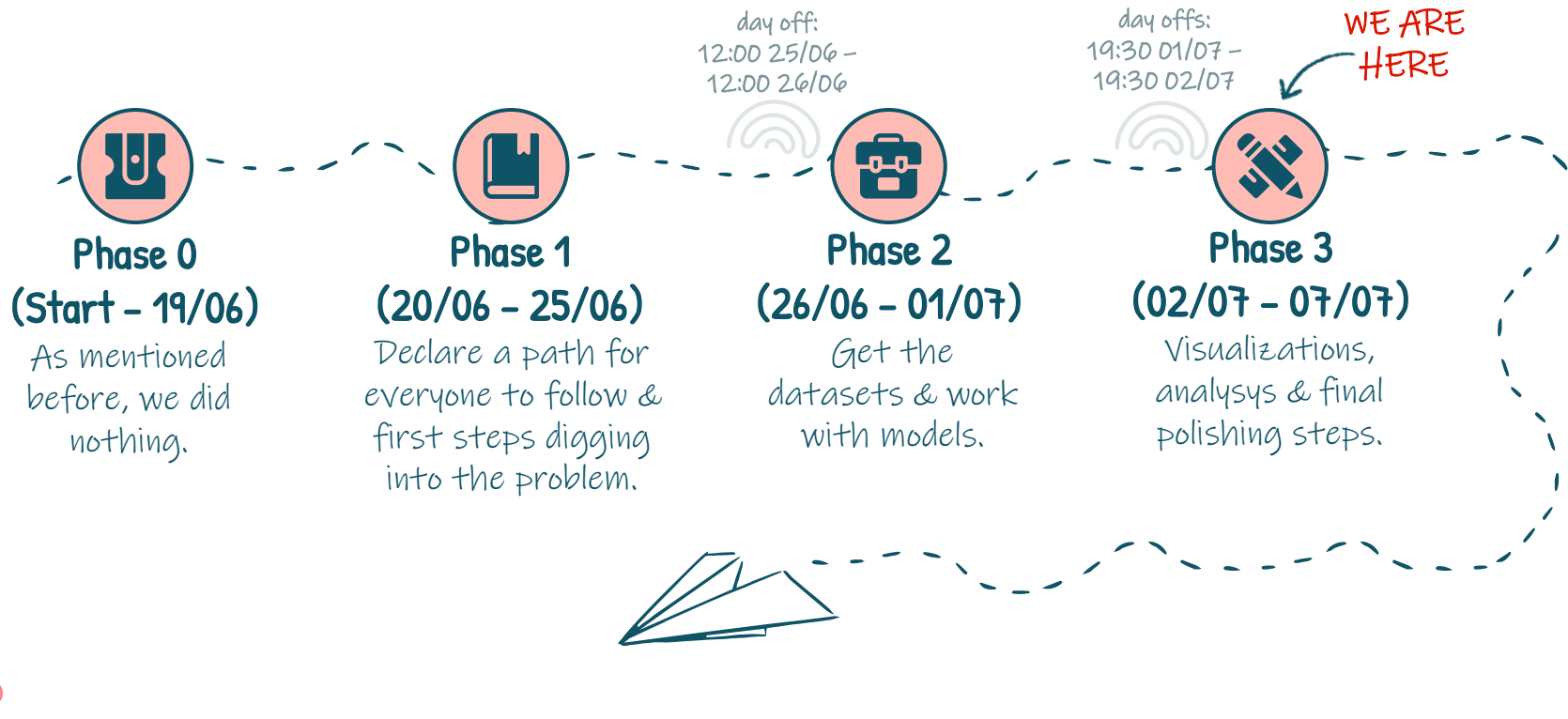
03/07/2021



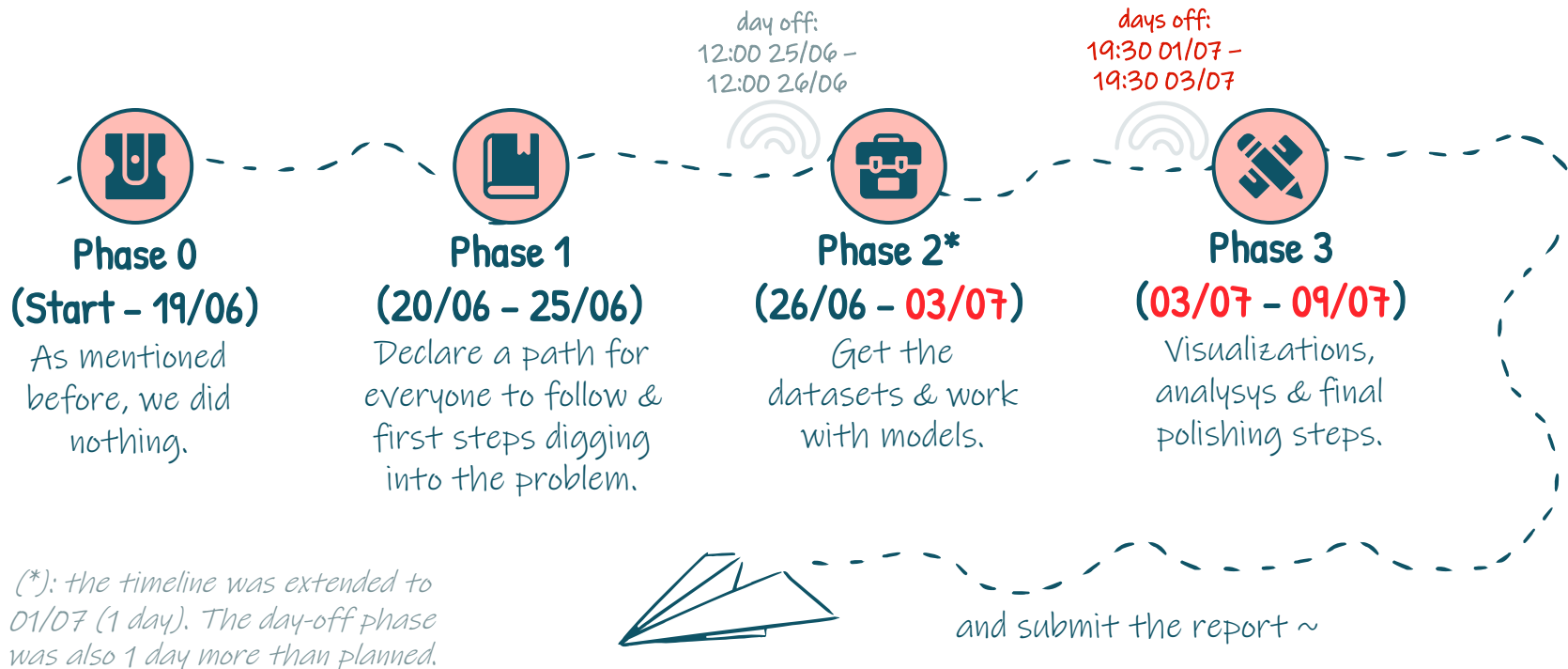
Abstract

In this report, we recall the progress timeline with some modifications, review the main tasks of second phase and consider remaining problems. From that, we come up with a plan for the last phase.

The progress timeline (previous)



The progress timeline (with updates)



How much have we done (approximately)?

100%

- Task 1: Data obtaining

And that took a whole week. Not good. Not good.

60%

- Task 2: Data analysis

No missing value, separated datasets for separated countries.
That's fine.



60%

- Task 3: Model building

Oh we have A WHOLE lot of tasks to get done from now.

10%

- Task 4: Model analysis

Along with model building, some parts of model analysis has been achieved. That's... a good thing actually (?)

(Temporary) Evaluation scores based on contribution

	First phase (40%)	Second phase (30%)	Third phase (30%)	Total
Hằng (Team cap)	✓	✓	💡	65%
Vinh	✓	✓	💡	70%
An	✓	✓	💡	70%
Đại	✓	✗	💡	40%
Nghiệp	✓	✓	💡	70%

✓ Participated

✗ Absent

💡 Working on it

What were the tasks in phase 2?



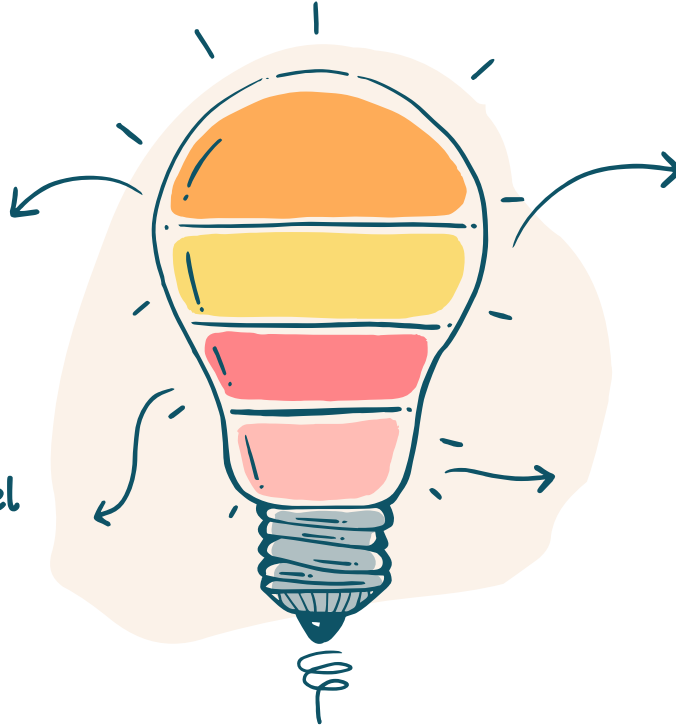
Study SIR model

Understands how it works
and get the code to run on
our dataset



Study Random Forest model

Understands how it works
and get the code to run on
our dataset (again)



Study ARIMA model

Understands how it works
and get the code to run on
our dataset (again)



Deal with problems from last phase (and extras)

...I'll talk about it right
next.



01

Tackle past problems & doing random stuffs (?)


Assigned to team captain




Past problems

Problems of last phase:

- ☒ Not yet having the target of this project (*Solved: Our target is now **the number of Infected***)
 - ☐ Not yet having a question set
 - ☐ Not yet calculating other attributes
 - ☒ Not yet understanding Python model syntaxes (*Solved: lol I'm sure y'all do now*)
- (2 remaining)



I was not capable to solve all of them while following the model tasks. It could have been done better. (sorry guys :<)







Past problems (cont.)



Why do we need question set? We need to develop one to visualize the data and models based on it. All figures has to be the answers of something.

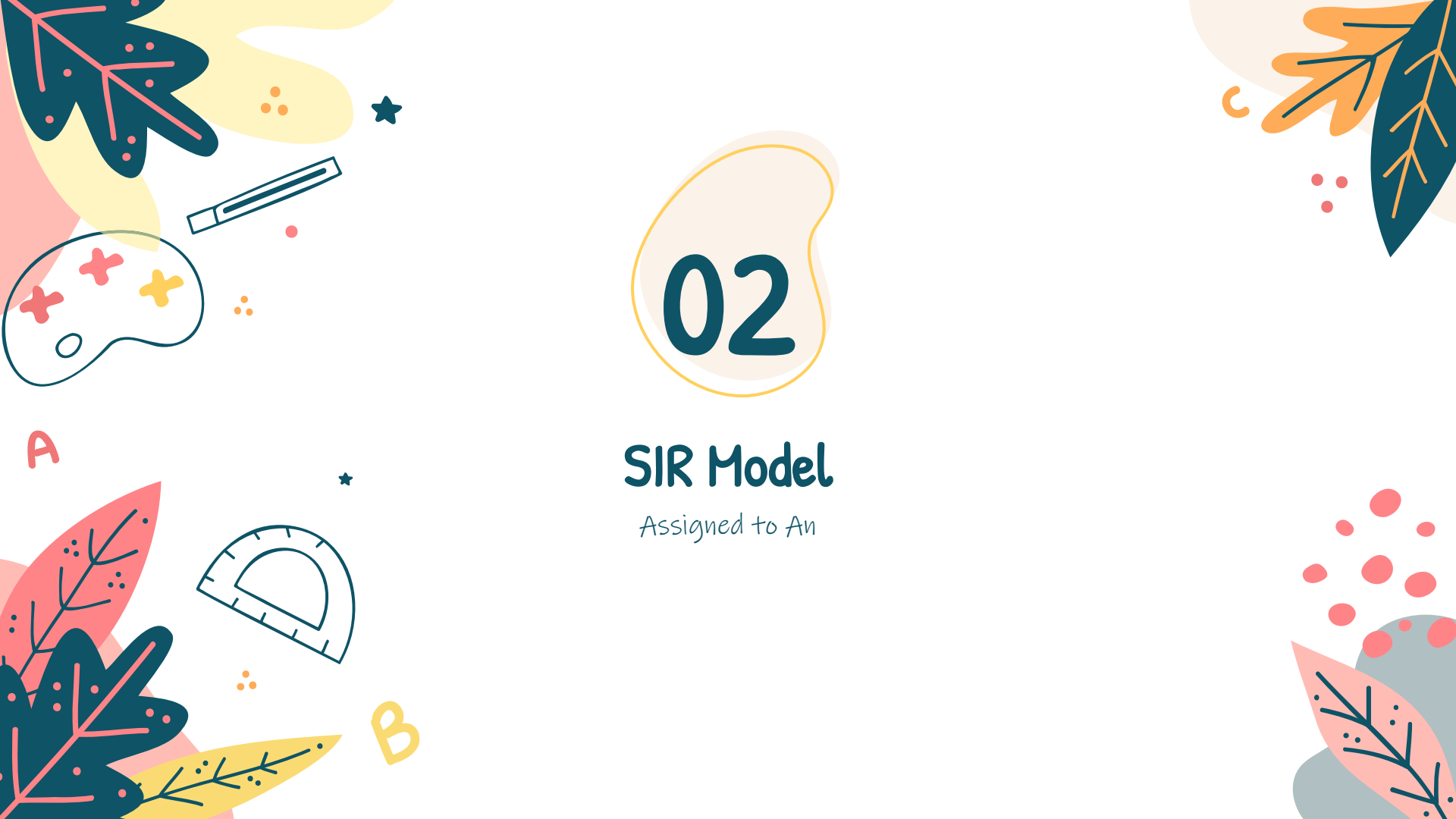
What about other attributes? Besides Confirmed, Deaths and Recovered, we might consider the following:

- _ $\text{Active/Infected} = \text{Confirmed} - \text{Deaths} - \text{Recovered}$
 - _ $\text{Susceptible (as of basic SIR model)} = \text{Population} - \text{Confirmed} - \text{Deaths} - \text{Recovered}$
 - _ Mortality Rate
 - _ Vaccination data
- and some more...
- 
- 



Uncategorized works I have done in phase 2(?)

1. Dig in a data source that contains extra attributes/datasets we need to prepare for visualization step. <https://github.com/owid/>
 2. Made some of the basic visualizations in RapidMiner tool.
 3. Made a list of figures we may need to make.
 4. Follow and support all Modeling tasks.
- 
- 



02

SIR Model



Assigned to An



03

ARIMA Model

Assigned to Nghiep



04

Random Forest Model

Assigned to Vinh



BRUH

