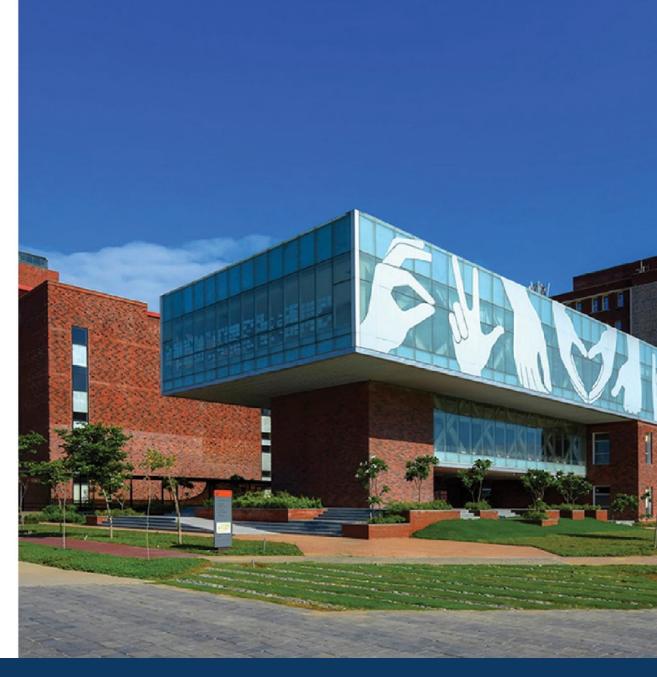# Professional and Executive Development Program in

Data Science for Social Impact

Project 1
Sep. 29 – Nov. 18, 2023

Project 1 – done individually

- ***Task*: analyze datasets using tools that you have learned and submit a report**
  - Choose from one of five themes
    - COVID19, Weather/Air quality, Agriculture, Text analytics, Crime
  - Add datasets of your own if necessary (LCs will help you)
- **Expectations: demonstrate your skills**
  - extract and analyse basic statistics from the dataset (using Excel)
  - generate appropriate charts and graphs for these statistics (using Excel)
  - run clustering and/or classification algorithms and/or any other ML/statistical techniques to answer at least two meaningful questions on the datasets
    - You will formulate these questions on your own as you explore the dataset
  - ability to visualise datasets on a map (this means that the datasets should have a geographical component associated with them)
  - ability to use advanced visualisation tools

Submission

- **At the end of the project, you will submit a report outlining your findings**
  - *Introduction to the datasets*
    - Including those that you found on your own
  - *Preliminary exploration of the datasets*
    - Consisting of basic statistical analysis
  - *Advanced analysis*
    - Consisting of interesting questions that you answered
    - What methods were used to answer these questions
  - *Appropriate visualizations over maps and using advanced visualization techniques*
    - Find datasets that complement the starting datasets
  - *Conclusions*

Next steps

- **LCs will share the spreadsheet of datasets**

- **A form will be floated for you to enter your preferences**
  - We will try our best to assign you your first choice

- Progress update every two weeks

- VMs will be provided on request