

*Without data you are just another person with an opinion* - W. Edwards Deming

#### About the workshop

Winter workshop in data science is a five-day (intensive) data science program, where a team of professionals from the industry learns fundamentals of data science. The workshop is designed to demystify data science and to empower professionals perform data analytics using the R programming language. We will study and practice a wide range of topics, including R fundamentals, visualization, data manipulation and modeling.

#### Expectations

The workshop **does not** have any pre-requisites, although a little programming experince will be a huge plus. There will be both hands-on sessions and lectures, and the trainees are encouraged (but not required) to bring their own data sets. Towards the end of this workshop the trainee will exhibit in-depth understanding of data science and analysis methods as well as proficiency in R. He/She will produce a portfolio that demonstrates mastery of analysis and visualization methods.

#### Agenda

Each day the session will begin at 9 am and end at 4.30 pm with on-your-own lunch break between 12 noon and 1.30 pm. Lectures will be held between 9-10 am and 1.30-2.30 pm followed by hands-on sessions. The lecture agenda are as follows:

Day 1: Git and R fundamentals

Day 2: R Markdown, Visualize and explore

Day 3: Transform and explore, Exploratory data analysis

Day 4: Wrangle data

Day 5: Data modeling and introduction to machine learning

#### Where to apply and fees

Fill out an application here by November 30, 2017 at 11:59 pm. Due to volume restrictions, late applications will not be considered. Applicants will be notified at the end December with instructions to pay the fees. If you miss the deadline, you can sign up for the general workshop form to be notified of future opportunities. The fee for the entire workshop is SGD 1000/-

#### About the mentor

Dr. Kasthuri Kannan, PhD, is an Assistant Professor of Pathology at New York University (NYU). He has more than 10 years of experience in data science and analytics. He started his data science career in 2006 in a company where he contributed to data mining projects and obtained directed insights in aircraft movement operations at the Tinker Air Force base. He proposed efficient strategies for aircraft logistics through analysis of aircraft movement data using machine learning algorithms. He is trained in mathematics and computer science with postdoctoral training at Memorial Sloan-Kettering Cancer Center (one of the top cancer research institutes in the world), where his data science skills were extensively applied to genomics resulting in top-notch publications. He is actively advising and collaborating with several scientists and physician-scientists, translating data science findings to clinical care, and is passionate about teaching and science communication. Currently, he is directing and teaching a course in data science and biomedical informatics at NYU, a required course for masters students in biomedical informatics. The course content can be viewed at: <https://kannan-kasthuri.github.io/pda.html> and Dr. Kannan's webpage can be accessed here: <https://kannan-kasthuri.github.io>