# **Data Science Candidates - Technical Assessment Test**

## **Assessment Overview**

This technical test will be used to measure and assess your data science skills by providing an unbiased and validated evaluation. There are no right or wrong answers, or you are not expected to understand business terminologies in this assessment, you will be evaluated on your ability to understand problems, analyse data in detail, visualise data meaningfully, find a suitable approach, explain models/algorithms you choose, and interpret results.

The problem statement, data set, and required steps to complete this assessment are given and detailed below. You are expected to use Python software language. In order to deliver your assessment, you will be given access to a GitHub repository by the sonalakedatascience GitHub account. Please check your email to accept the invitation and upload your final Python file to that GitHub repository.

## **Problem Statement**

Mobile operators (eir, Three, Vodafone etc.) face congestion problems in their networks due to increased usage and other factors. Congested sites (cellular base stations) will lose revenue because usage at these sites is curtailed. In an environment where the operators have finite resources to address congestion (be it through building new capacity or upgrading existing sites) they require a solution that will give them future visibility of congestion in order to prioritise their medium to long term investments.

Sonalake’s customer, a successful telecoms operator has engaged us to:

* Get a 14-day forecast on their network usage KPI,
* Find out if they would face congestion, within 14 days in the future,
* Get meaningful insights and recommendations according to your findings.

They have provided a 3-month historical data set with hourly granularity and a threshold level.

## **Information**

* **Data Set:** The data set comprises 3-month (61 days) historical timeseries data with hourly granularity. It includes two fields showing the time stamp (days/hours) and the network value (in Gbps).
* **Threshold Level:** When a value exceeds and breaches the threshold level (16.40 Gbps) given in the data set, it means that the network is congested at that point.

## **Required Steps**

Please feel free to follow the solution, analysis, models and algorithms of your choice. Any data visualisation relies totally on your imagination and preference, however; your work must include all the listed elements below and it will affect your evaluation.

* A brief description of the problem with your own words.
* Descriptive headers to distinguish different sections of your work (e.g.: Section 1: Data Analysis)
* Brief notes under each section to describe the purpose of that section.
* Brief comments in code blocks to explain why you use that code.
* A summary and general discussion of your results and findings.