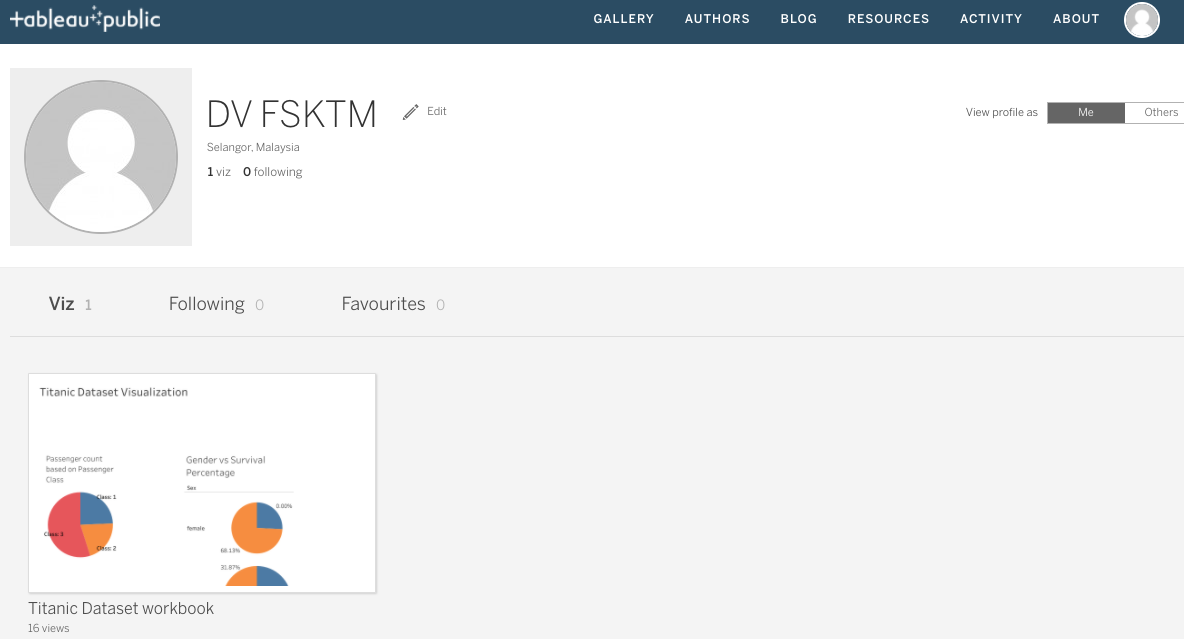
**SSK4409 BIG DATA ANALYTICS PROJECT**

1. This project consists of 3 parts. Part 1 is on the Big Data Platform, Part 2 is on the Data Analytics and Part 3 is on the Data Visualization. You need to answer all 3 parts to complete this project. Part 2 and Part 3 requires you to use the same dataset.
2. **Part 1 (Big Data Platform) (20 marks) (CPS 9)** - In this part, you need to install Hadoop in your machine, and you need to show the followings:
   1. Show evidences in your report that you have successfully installed Apache Hadoop in your machine. (10 marks)
   2. Show evidences in your report that you managed to run MapReduce program on word counting on a text file. (10 marks)
3. **Part 2 (Data Analytics) (50 marks) (CPS 8)** – In this part, you need to perform data analytics tasks. First, you need to choose a dataset from the datasets available in Kaggle. Then you need to install the relevant tool to perform your data analytics task (e.g Jupyter Notebook for Python or R Studio for R). Then you need to show the followings:
   1. Show evidences in your report that you have perform data analytics task on the selected dataset
      1. Pre-processing, data cleaning – 10 marks
      2. Modelling – 10 marks
      3. Results (findings, numerical results and visualizations) – 20 marks
      4. Summary (recommendations, insights) – 10 marks
   2. Publish your R code to pdf formats (https://rmarkdown.rstudio.com/lesson-9.html) or for Python you can publish it to HTML format and then save it as pdf. (You may publish it to PDF format but it needs a lot of steps, if you are okay with than you can proceed)
4. **Part 3 (Data Visualization) (30 marks) (CPS 9)**– In this part, you need to visualize your selected dataset and show the followings:
   1. Install public Tableau in your machine. Create Tableau account in Public Tableau site.



* 1. Use Tableau to analyze and reveal various relationships within your selected dataset.
  2. All Tableau worksheets must be combined within a single Story in which it may be a combination of multiple Dashboards. (Each dashboard may contain multiple worksheets). Publish your Tableau story in the Public Tableau page. At least 3 dashboards must be created to complete this part. Each dashboard may contain one or more worksheets. (30 marks)

**Submit your report to Putrablast by 6th January 2024.**