

SIT 112 | Data Science Concepts

Lecturer: Dr.Uhtayasanker Thayasivam| rtuthaya@gmail.com

Data Science Project

Due:

Friday 5pm, 28th Feb 2018

Note: This project contributes 25% to your final SIT112 mark. It must be completed individually and submitted before the due date: **5pm, 28/02/2020**.

This Data Science Project aims to apply machine learning techniques to understand and visualize relationships in data. Our task specifications for this project will focus on two supervised learning tasks: linear regression and classification. You must demonstrate through skills acquired in describing the data, exploratory data analysis and prediction.

1. Data and Resources

In the Data Science Project folder, you will find the following files:

Project_instructions.pdf	This is the file contains the instruction to complete your project.
life_expectancy.csv	this contains data related to life expectancy
Data_Description.pdf	Description of life_expectancy columns
project_notebook.ipynb	This is the Jupyter notebook which has been prepared and pre-filled for you to complete the programming task.

2. Task Description

A python note book file **project_notebook.ipynb** has been prepared for you to complete this task. Download this notebook, load it up and follow instructions inside the notebook to complete the task. *You are required to submit your solution in a Jupyter Notebook format as well as its exported version in html.*

3. Summary for submission

This project is to be completed individually and submitted online. By the due date, you are required to submit the following files to the corresponding *Assessments/DSProject* in *SIT112* site:

1. **[YourID]_project_solution.ipynp** : your IPython notebook solution source file.
2. **[YourID]_project_output.html**: the output of your IPython notebook solution in html