

## Python for Data Analysis

**By: Atefeh Khazaei**

### Week#10 – Practicals

In this practical sheet, you should analyse the following dataset and present a simple Linear Regression model.

#### Medical Insurance Costs Dataset

This dataset was inspired by the book Machine Learning with R by Brett Lantz. The data contains medical information and costs billed by health insurance companies. It contains 1338 rows of data and the following columns: age, gender, BMI, children, smoker, region and insurance charges.

Link: <https://www.kaggle.com/mirichoi0218/insurance>

Apply what you learned in this class and try to do the following CRISP-DM phases:

- 1- Data Understanding
  - Info
  - Describe
  - Data Visualization
- 2- Data Preparation
  - Handling Missing Values
  - Handling Outliers
  - Normalization
  - Dummy
- 3- Modelling
  - Linear Regression
- 4- Evaluation
  - MAE