



T5 Boot Camp

Data Science

Project Proposal

No.2 (Regression)

No.	G5 - Names
1	Surayyi Alqahtani
2	Nawaf Almutairi

Instructor:

Mejdal Al-Qahtani

- **Question/Problem statement:**

Health insurance covers the costs and offers many other important benefits. Health insurance protects you from unexpected, high medical costs. You pay less for covered in-network health care, even before you meet your deductible. This project is aimed at giving the insurance companies an accurate prediction of costs for every customer.

## **Q\ How the insurance companies can predict the cost for individual patients?**

- **Data Description:**

The dataset is called Medical Insurance Cost form Kaggle

**Columns:** age(int), sex(object), BMI (float), children (int), smoker(object), region(object), charges(float).

**Data size:** (1338 Rows x 7 Columns).

- **Tools:**

**Programs:** Python programming language.

**Libraries:** Pandas, NumPy, Matplotlib, seaborn, Scikit Learn, statsmodels, patsy, BeautifulSoup, requests.

- **MVP Goal:**

The goal of this project is to help insurance companies to predict the price for every individual patient based on several factors. For example, Age, Body mass index and number of children covered by health insurance / Number of dependents.