

# Project Proposal

This repo is one of the T5 Bootcamp requirements.

## Goals

The project aim is to get answers for the following questions through obtaining, cleaning and analyzing a certain dataset using Exploratory Data Analysis (EDA) approach.

- What is the most gender answering the surveys in Saudi Arabia and worldwide?
- What's the most popular programming language in 2020 and 2021?
- What's the most commonly used cloud platforms in 2021?
- What's the most common framework in 2021?
- What's the most popular Integrated Development Environment (IDE) in 2021?
- What's the most popular ML algorithm used for Data Science 2021?
- What's the most common industry rely on Machine Learning products in 2021?

## Dataset

The data comes from the survey that Kaggle performed this year. The survey questions range from demographic questions, such as gender and level of higher education, to questions about programming languages, tools, and machine learning algorithms used. The dataset is available as a comma-separated values (.csv) file.

## Tools

There are several tools that can be used to achieve the goal of this study, such as: Numpy and Pandas for data manipulation and Matplotlib and Seaborn for plotting, The work will be done through Jupyter notebook.

## TO DO

- Explore the data and come up with EDA phases.
- **NOTE:** the used features may be increased or changed.