**UDACITY – DATA SCIENTIST NANODEGREE**

**TERM II**

**PROJECT I: WRITING A DATA SCIENCE BLOG**

# CONTENTS TABLE:

1. [Pre-requirements.](#_PRE-REQUIREMENTS:)
2. [Project Motivation.](#_PROJECT_MOTIVATION:)
3. [Description of files.](#_DESCRIPTION_OF_FILES:)
4. [Findings](#_FINDINGS:)
5. [License.](#_LICENSE:)
6. [Future Work.](#_FUTURE_WORK:)

# PRE-REQUIREMENTS:

For this project, there is no exceptional requirements, just the user needs to install python and run the tool.

The following python documentation will provide more information.

<https://docs.python.org/3/>

Also, as this script is to compare between two years data of stack overflow surveys. You need to have access files for the stack overflow survey results. It can be accessed through the below link.

<https://insights.stackoverflow.com/survey>

For this project, we used the 2017 and 2019 data, you can unzip the downloaded survey results through WinRAR application.

<https://www.win-rar.com/>

# PROJECT MOTIVATION:

Stack overflow is one of the main community platforms for developers or business peoples worked in the software development industry. It’s also, useful for the end-user to analyze the data and compare it with. In this project, I selected the most three themes that developers monitor which is programming languages, database, and platforms.

This tool, I will answer for the below six questions.

1. What's the most worked programming Language in 2017 & 2019?
2. What's the most wanted language was needed in 2017 by the developers to be used in 2018 compared with the 2019 desired language?
3. What's the most worked Database Engines in 2017 & 2019?
4. What's the most wanted Database Engines was needed in 2017 by the developers to be used in 2018 compared with the 2019 desired Database Engines?
5. What's the most worked Platform in 2017 & 2019?
6. What's the most wanted platform was needed in 2017 by the developers to be used in 2018 compared with the 2019 desired platform?

# DESCRIPTION OF FILES:

The below will help you to know about each of the available files:

* 1. Raw Data for 2017 & 2019: The current version, has two data files for stack overflow survey and the data can be downloaded from the following link.

Each year data will be downloaded as a zip file, the zip files include the data and schema of the data. For the sake of this too, the schema file is not needed, but the data file which named as **survey\_results\_public** available in 2017 and 2019 folders.

<https://insights.stackoverflow.com/survey>

3.2 ipynb file [**term\_2\_project\_one**] which includes all the analysis of the data based on the CRISP-DM methodology.

3.3 **term\_2\_project\_one\_V3.html**, is the outcome of the tools based on static user inputs.

# FINDINGS:

Based on the user selection, the tool can compare between two years in the same country and responders' gender for the survey.

All the details can be available through my blog.

<https://development4geeks.wordpress.com/2019/11/03/understanding-stack-overflow-developers-survey/>

# LICENSE:

The raw data for the developers' survey is licensed for stack overflow under the [stack overflow privacy policy.](https://stackoverflow.com/legal/privacy-policy?_ga=2.188457640.1669360686.1572983339-326182324.1566123578)

# FUTURE WORK:

This script can be updated to be analysis for different years of the stack overflow developers survey and adding more features to compare rather than the language, database engine, and platforms.