**IBM Final Capstone Project**

To keep pace with changing technologies and remain competitive, an organization (a global IT and business consulting services firm that is known for their expertise in IT solutions and their team of highly experienced IT consultants) regularly analyzes data to help identify future skill requirements.

As a Data Analyst, I will be assisting with this initiative and have been tasked with collecting data from various sources and identifying trends for this year's report on emerging skills.

My first task is to collect the top programming skills that are most in demand from various sources including:

* Job postings
* Training portals
* Surveys

Once I have collected enough data, I will begin analyzing the data and identify insights and trends that may include the following:

* What are the top programming languages in demand?
* What are the top database skills in demand?
* What are the popular IDEs?

I will begin by scraping internet web sites and accessing APIs to collect data in various formats like .csv files, excel sheets, and databases.

Once this is completed, I will make that data ready for analysis using data wrangling techniques.

When the data is ready, I will then apply statistical techniques to analyze the data. Then bring all your information together by using IBM Cognos Analytics to create the dashboard. And finally, show off my storytelling skills by sharing my findings in a presentation/report.

Throughout this guided project, IBM Skills Network Labs (SN Labs), which is a virtual lab environment, was used in this course.

Tasks:

* Collecting Data Using APIs/ WebScraping
* Exploring Data
* Data Wrangling or Munging:
  + Identify duplicate rows in the data frame.
  + Remove duplicate rows from the dataframe.
  + Find the number of missing values for all columns.
  + Find the value counts for the column "Employment".
  + Normalize the data using two existing columns
* Analysis:

1. Plot a distribution curve, and histogram.
2. Find the median, and outliers of particular columns.
3. Compute the Inter Quartile Range.
4. Find out the upper and lower bounds, and find correlations between numerical columns.
5. Create a new dataframe.

* Visualization skills by working with the Stack Overflow Developer Survey 2019 dataset:

1. Create a histogram to show the distribution of data.
2. Create different plots such as a scatter, bubble or boxplot.
3. Create a pie chart, bar chart, and stacked chart to show medians and counts.
4. Final dashboard using IBM Cognos Analytics with the Kaggle survey 2023 data to create a dashboard. In this dashboard, I will create the following Dashboard Components

A. Current Technology Usage Tab

1. Top 10 Languages
   * Visualize the top 10 programming languages respondents have worked with (`LanguageHaveWorkedWith`).
2. Top 10 Databases
   * Visualize the top 10 databases respondents have worked with (`DatabaseHaveWorkedWith`).
3. Platforms
   * Visualize the different platforms respondents have worked with (`PlatformHaveWorkedWith`).
4. Top 10 Web Frameworks
   * Visualize the top 10 web frameworks respondents have worked with (`WebframeHaveWorkedWith`).

B. Future Technology Trends Tab

1. Top 10 Languages Desired for the Next Year
   * Visualize the top 10 programming languages respondents desire to learn next year (`LanguageWantToWorkWith`).
2. Top 10 Databases Desired for the Next Year
   * Visualize the top 10 databases respondents desire to learn next year (`DatabaseWantToWorkWith`).
3. Desired Platforms for the Next Year
   * Visualize the platforms respondents desire to work with next year (`PlatformWantToWorkWith`).
4. Top 10 Web Frameworks Desired for the Next Year
   * Visualize the top 10 web frameworks respondents desire to learn next year (`WebframeWantToWorkWith`).

C. Demographics Tab

1. Respondent Classified by Work Experience
   * Visualize the number of respondents classified by work experience (` WorkExp `).
2. Respondent Count for Countries
   * Visualize the number of respondents from different countries (`Country`).
3. Respondent Count by Age
   * Visualize the number of respondents in different age groups (`Age`).
4. Respondent Count by Gender and Classified by Education Level
   * Visualize the respondents classified by gender (Gender) and further categorized by their education level (`EdLevel`).