**EXECUTIVE SUMMARY**

In this project, a **Python** based **Exploratory Data Analytics (EDA)** notebook is prepared to analyze the **Google Analytics Customer Revenue Prediction Dataset** obtained from **Kaggle**. The **Data Science and Analytics** is one of the leading areas now-a-days. EDA is the foremost component of Data Science & Analytics, as it plays a vital role to general the initial insights from the data after performing descriptive statistics, inferential statistics, or both, and visualization techniques to generate actionable insights from data.

In this project, I have analyzed the data obtained from a Kaggle challenge worth US$45000, which is known as Google Analytics Customer Revenue Prediction. This dataset is analyzed for the basic analytics to obtain the information about the most popular browser, search type, region, visitors and other research questions. This project explores the data using data aggregation & grouping techniques to understand the patterns affecting the customer revenue, after the application of data cleaning and preparation phase.

The regression models are fitted to the dataset to predict the revenue based upon several quantitative and qualitative parameters such as visitor count, month, year, day of month, weekday, continent, browser, operating system, etc. Several regression models are applied to the dataset, which includes ordinary least squares (OLS), ridge regression, random forest regression, gradient boosting regression and adaboost regression. Random Forest (R2 – 0.97) and Gradient Boosting (R2 – 0.95) are found to be the best regression models, and correctly describe 95% of data.

For the selection of analytics stack, the Python programming with Jupyter Notebooks IDE has been utilized, which are coupled with popular Python Libraries like Numpy, Scipy, Pandas, Matplotlib, Seaborn and many others.