Data Science and Analytics

With R or Python

Relevance of Digital Skills

With the emergence of the Digital Economy, technological developments and innovation have altered the business landscape and consumer preferences. As a result, companies are having to rewire their businesses, adopt digital technologies and also reskill their workforce. Moreover, the pandemic has fundamentally changed the way we work and live, thereby accentuating the need for reimaging business models and reskilling.

According to McKinsey & Company, **Digital Skills** is one of the four key skill sets needed for an organization's success in the post Covid era. Regardless of the industry sector or function, most employees will need basic digital skills to perform in an environment where interfaces with customers, suppliers and other stakeholders would have changed.

How will Skills in Data Science help you?

Data Science holds the top spot for the most sought after skill and is uppermost in the minds of CEOs and CXOs in their quest to maintain competitive advantage. Leading companies have realized the importance of analytics, AI and ML in improving decision making and enhancing organisational performance. Acquiring skills in data analytics will help you acquire skills in using industry leading tools such as R or Python in applying data analytics, ML and visualization to address business challenges

Learning Objectives

This training program is aimed at facilitating participants to gain an understanding of application of analytics and the tools that can be used to obtain more powerful insights from raw data. Data analytics is effective only when a well-defined model is used. Hence, we have designed a training program which will help participants use the following framework to develop analytics skills:

1 **Key Business Challenges (KBCs)**
Defining KBCs and performance measures (KPIs), framing objectives for analytics and visual reporting.

2 **Data Organization & Preparation**
Identifying the type of data required, sourcing and organizing data.

3 **Data Analysis & Visualization**
Identifying the type of data required, sourcing and organizing data.

4 **Presenting Results**
Tell your story and present the results

Training Program

This training program on *'Data Science and Analytics with R or Python'* will help you master R or Python and use functions and libraries for creating applications and programs for data science in any industry sector. You will acquire skills in uncovering insights by analyzing large volumes of data, learn to use statistical analysis techniques such as hypothesis testing, regression & correlation analysis, classification modeling and machine learning algorithms.

More details in curriculum

Program Delivery

The training will be delivered in an online mode. The format will be a combination of discourse, breakout sessions and knowledge sharing for participants to internalize their learning. The training will be conducted in sessions of 3 hours each spread over 6-8 days.

Who Should Attend

Individual contributors, managers and senior managers from any industry will benefit from this training program.

In the digital era, professionals must reskill and upskill to remain relevant and in demand.

For more information please call 74394 81638 or send a mail to swarup.dutta@kaalp.com

Curriculum

TOPICS	TECHNIQUES IN R	TECHNIQUES IN PYTHON
KEY BUSINESS CHALLENGES		
Evaluating Key Business Challenges	Evaluating Business Background, Framing Business Objectives, Defining the analytics outcomes and identifying the KPIs and constraints.	Evaluating Business Background, Framing Business Objectives, Defining the analytics outcomes and identifying the KPIs and constraints.
DATA ORGANISATION & PRESENTATION		
Data Sourcing	Importing csv files, Tab delimited File, SPSS files, Stata files Connecting to external database. R libraries-Foreign package. ODBC connectivity Data mining -(CRISP DM) methodology	Creating python Variables. Numeric, string and logical operations. Importing csv and Tab delimited files, Connecting to external database. Introduction to Python using Numpy, Pandas library and Matplotlib.
Data Organisation	Merge datasets, Appending datasets. Concept of primary key, Foreign key	Merge datasets, Appending datasets, subsetting datasets using pandas library, Outlier analysis, Shape function
Understanding use of data for decision making	Data Types-Structured vs unstructured data. Variable types-Scale, Nominal and Ordinal. Working with string data, date variable type Summary Statistics-Mean, Median and	Data containers: List, Dictionaries, Tuples and sets. List subsetting, List slicing.

| | Mode, Variance and standard deviation Coefficient of variation, Percentiles, Quantiles, Vingtiles, Deciles. | |

Decision Making Models	Linear Regression, Logistic Regression Model, GLM Models	Simple Linear Regression, Multiple Linear Regression, Logistic Regression Model, Decision Tree using scikit package
Analyzing Data for Decisions	IS NA function, Analyze Missing values, Amelia Package, EM method, Tables, Cross tabulation, Parametric Test and Non Parametric Tests	Is.na function, Analyze Missing values, Amelia Package, Tables, Cross tabulation. Dummy coding, concatenating variables.
Creating an Analysis Plan	Hypotheses Testing- P value , Level of significance, T Test-One sample, , Independent sample T test, Paired samples T Test, F Test.	Hypotheses Testing- P value , Level of significance, T Test-One sample, , Independent sample T test, Paired samples T Test, F Test.
Data Analysis Techniques	Correlation, Heat map, Anova-Single Factor, Two Factor Anova, Difference between Linear Trend Analysis, Kruskall Wallis H Test and Exponential Trend model	Correlation, Heat map, Anova-Single Factor, Two Factor Anova, Difference between Linear Trend Analysis, Kruskall Wallis H Test and Exponential Trend model
Interpretation of Results	Coefficient of Determination-R Squared, Adjusted R Squared, Accuracy, Sensitivity analysis and Specificity Analysis.	Coefficient of Determination-R Squared, Adjusted R Squared, Accuracy, Sensitivity analysis and Specificity Analysis.
Evaluating Alternatives	Comparing accuracy level of multiple models	Comparing accuracy level of multiple models
Validation of results	Receiver Operating Characteristic curve(ROC Curve),Mean Absolute Percentage Error(MAPE) Minimum Absolute Percentage Error(MAPE) to determine the worst case scenario of a business forecast Max Absolute percentage Error (Max APE)	Receiver Operating Characteristic curve(ROC Curve),Mean Absolute Percentage Error(MAPE) Minimum Absolute Percentage Error(MAPE) to determine the worst case scenario of a business forecast Max Absolute percentage Error (Max APE), Silhotte Index, Dendogram
Uncertainty of decisions	Creation of Training, Testing and holdout sample, Cross validation, K fold Validation, Hyper parameter tuning	Creation of Training, Testing and holdout sample, Cross validation, K fold Validation, Hyper parameter tuning, AUC Curve, F1 score, Prevalence score, Precision Score.

Tell Your Story	ROC Curve, Bar chart, Pie chart, Correlogram, Count plot	Dashboard development & modification Interactions between charts, Data Calculations, Aggregate, Logical and User Calculations Create interactive dashboards
Advanced Use of Charts, Dashboards and Visualization	Heat maps, Frequency Distribution plots, Interactive charts.	Explore how to share dashboards and reports and Time sharing tips. Explore comments in reports and dashboards. Format Row Level Security Explore Object Level Security.