Data Science for Power Users

Course Synopsis:

Predictive analytics generates future insights with a significant degree of precision based on historical data. Through this course, learners will demonstrate their competencies in developing a predictive analytics project in teams by applying data privacy and ethical principles in the collection, use and disposal of data. With the data collected, learners will demonstrate their competencies to impute data format, transform and reshape the business data before applying relevant predictive modelling techniques to predict the desired business outcomes. Learners will then perform data exploratory analysis to discover patterns and trends and develop data stories for an effective narrative and visual representation of their predictive analytics project.

Learning Objectives:

At the end of this course, learners will be able to:

1. Collect data from multiple sources using appropriate collection tools and techniques that comply with data and privacy ethics.
2. Perform data preprocessing techniques to impute data format, transform, reshape and protect the data in accordance with the business requirements and data protection principles.
3. Develop interactive and effective visualisations with a global perspective to address international and cultural differences, diverse needs and expectations of the key stakeholders using visualisation tools.
4. Perform data exploratory analysis to identify underlying data patterns, trends and analytical insights using visualisation tools.
5. Apply relevant predictive modelling techniques to predict the desired business outcomes to meet the service expectation of the key stakeholders.
6. Work collaboratively in a team to develop dashboards using data storytelling approach for an effective narrative and visual representation of their predictive analytics project.

Course Duration:

14 days

Analytical Software:

- Microsoft Excel
- Power BI / Tableau / QlikSense
- Python

Course Fee:

This is a course under the SkillsFuture Series. The course fee and subsidy table as follows (exclude GST):

Course Name	Total Programme Fee	Singapore Citizen < 40 yrs old and Permanent Residents	Singapore Citizens ≥ 40 yrs old	Singapore Citizens and Permanent Residents under PR
Data Science for Power User	$2,400.00	$720.00	$240.00	$240.00

The course fee is inclusive of venue and course materials. Certificate of completion will be provided for participants who have attended 75% of the course. Two refreshment breaks will be provided.

Course Programme:

Program	
Day 1	**Introduction to Visualisation and Tool** **Basic Visualisation Techniques** **Preparing data for analysis** <u>Lab 1:</u> - Importing data from external data sources - Combine multiple tables into a single table - Profile data

	Visualising Data
	 <u>Lab 2:</u>
	 - Add visualisation items to reports
	 - Choose an effective visualisation
	     * table and matrix
	     * bar and columns chart
	     * line an area charts
	     * pie chart, donut chart, and treemaps
	     * combo charts
	     * card visualisation
	     * funnel visualisation
	     * gauge chart
	     * waterfall chart
	     * scatter chart
	     * slicer visualisation
	     * Q&A visualisation *(depend on selected analytical software)*
	 - Format and configure visualisations

Day 2	**Visualising Data** <u>Lab 3:</u> - Import custom visual - Work with key performance indicators

Program		
	Visualising Spatial Data Lab 4: - Putting data on Singapore Map	
	Create data-driven story - Best practices for dashboard/report design - MICE Framework for Data Visualisation Lab 5: - Design report layout - Design report navigation - Use basic interactions	
Day 3	**Advanced Interactive Report** Lab 6: - Use advanced interactions and drill through - Configure conditional formatting - Apply slicing, filtering and sorting - Publish and export reports	
	Introduction to Python Lab 7a: - Python Environment Setup - Python Variables - Data Types - Operators	
Day 4	**Introduction to Python** Lab 7b: - Python Collections (List, Tuple, Set, Dictionary) - Control Flow (If else, loops) - Functions - Modules - Handling Files	
	Descriptive Statistics - Descriptive vs Inferential statistics - Types of data (quantitative vs qualitative) and level of measurement - Central tendency (mean, median, mode) - Shape of distributions, deviation, variance and standard deviation Lab 8: - Descriptive statistics using Python	
Day 5	**Data Analysis using Python** Lab 9: - Numpy - Pandas - Matplotlib	

Program	
Day 6	**Data Preparation Techniques** - What leads to poor data and its impact - Understand the data extraction, transformation and loading (ETL) processes - Data sources, data meta data and dictionary - Handle common data issues: missing data, duplicate data, etc <u>Lab 10:</u> - Data Preparation using Python
Day 7	**Introduction to Machine Learning Techniques** - Data Analytics Lifecycle (CRSIP-DM) - Machine Learning use cases - Machine Learning vs AI vs Deep Learning
	Machine Learning I - Predictive Modelling: Supervised Learning - Supervised Learning: Regression <u>Lab 11:</u> - Linear Regression
Day 8	**Machine Learning II** - Supervised Learning: Classification <u>Lab 12:</u> - Logistic Regression <u>Lab 13:</u> - K-Nearest-Neighbors
Day 9	<u>Lab 14:</u> - Decision Trees and Random Forests <u>Lab 15:</u> - Support Vector Machines
Day 10	**Machine Learning III** - Predictive Modelling: Unsupervised Learning <u>Lab 16:</u> - K-Means Clustering <u>Lab 17:</u> - Association Rule Mining
Day 11	**Feature Engineering** <u>Lab 18:</u> - Forward and Backward Feature selection - Dimension Reduction
	Model Comparison and Evaluation - Comparison of models: strength & weakness - Compare black box and white box approaches - Evaluate and improve on the models - Case Study using CRISP-DM

Program	
Day 12	**Introduction to Natural Language Processing (NLP)** - Text Preprocessing - Text Classification <u>Lab 19:</u> - Text Classification using Python
Day 13-14	Predictive Modelling Capstone Project