

## 學習對象：

想了解巨量資料概念、進而做巨量資料分析的人員。本課程將提供由基礎概念至應用實務等各種層面的內容，讓學員一步步踏入巨量資料科學家的領域。

## 課程描述：

此一基礎套裝課程，精選 SAS 高效能商業分析軟體做實機體驗。學員全程參加課程後，將獲得一張 SAS 提供之資料科學家基礎班之證書。

## 必備條件：

本課程由基礎課程開始，學員無需具備相關知識與技能。

## 課程天數：

共有 5 堂不同課程，總計 10 天的訓練時間。

## 課程內容：

### 【Course1】

#### 揭開巨量資料&資料科學家的面紗

##### 巨量資料與資料科學家

- 巨量資料之趨勢
- 資料科學家的角色及需具備技能

##### 巨量資料分析之企業範例

- 整合性客戶服務 – 全通路服務案例
- 洞察需求創造企業利潤 - 客戶案例

##### 巨量資料之資料集建置

- 巨量資料之衝擊與挑戰
- 建置分析性資料集
- 分散式資料庫 Hadoop 架構
- 巨量資料與 Hadoop 平台之結合

##### 巨量資料之管理機制

- 巨量資料之人員及角色管理
- 巨量資料之資料集管理

##### 巨量資料之高性能分析架構

- 平行運算(Grid Computing)架構
- 內植資料庫(In-Database)架構
- 內植記憶體(In-Memory)架構

##### 巨量資料之分析

- 視覺化分析
- 資料採礦及文字採礦分析
- 預測分析

### 【Course2】

#### 快速整理&分析巨量資料

##### 快速資料讀入及合併

- 交談式介面讀入各式外部檔案
- 交談式介面合併資料檔
- SAS® Enterprise Guide 實機操作

##### 快速巨量分析資料集建置

- 函數運用、欄位新增、資料篩選
- 依照不同分組彙總成巨量分析資料集
- SAS® Enterprise Guide 實機操作

##### 運用 SAS 視覺化進行資料勘查

- 視覺化勘查視窗之互動設計
- 資料勘查之相關與決策樹分析

##### SAS® Visual Analytics 實機操作

- 多視窗下鑽及互動報表製作
- 指標與儀表板及連結報表製作
- SAS® Visual Analytics 實機操作

##### 行動裝置(iPad)呈現巨量資料報表

課程內容：

## 【Course3】

### 輕鬆踏入 5 大統計分析殿堂

#### 基本統計介紹

- 統計常用名詞介紹
- 單樣本檢定

#### 變異數分析(ANOVA)

- 兩組樣本一維變異數分析
- 多組樣本一維變異數分析

#### 迴歸分析

- 簡單、多元與逐步迴歸分析
- 檢視殘差、找具影響力之觀察值與共線性檢定

#### 邏輯斯迴歸模型

- 建模資料的整備與結果的評估
- 利用邏輯斯迴歸模型進行預測

#### 集群分析方法

- 階層式與分割式集群分析
- 集群個數選取方法及特徵的描述

#### 時間序列模型

- 穩定式時間序列模型與預測
- 時間序列模型適合度的評估

## 【Course4】

### 精準建立採礦預測模型

#### Enterprise Miner 導論

- 資料採礦分析流程方法論(SEMMA)
- 分析資料欄位屬性定義

#### 決策樹

- 使用二元和多元裂分法建構決策樹
- 決策樹模型結果解讀、修剪和調整

#### 邏輯斯迴歸分析

- 演算方法與相關參數設定
- 模型資料處理方法及結果解讀

#### 類神經網路

- 演算方法及相關參數設定
- 模型變數選擇方法

#### 模型評估診斷和評分

- 模型比較、樣本比率調整
- 評分程式碼的產生和使用

#### 關聯性分析

- 集群分析及購物籃分析

#### 文字採礦

- 文件歸類
- 文件規則產生器

## 【Course5】

### 建置進階資料集(Data Mart)之語法

#### 外部資料讀入

- 讀入 Excel 工作表
- 讀入標準區隔、非標準區隔之資料
- 讀入資料庫資料

#### 合併資料檔

- 垂直增補合併與上下完整資料合併
- 1 對 1、1 對多及多對多之左右合併資料檔

#### 資料處理

- 函數運用、欄位新增、資料篩選及資料輸出
- 依照不同分組彙總資料
- 條件式執行迴圈(Do Loop)及陣列(Array)之運用

#### SAS 巨集語言

- 巨集變數的宣告與呼叫
- 巨集定義的宣告與呼叫

#### 運用程式語法建置進階資料集(Data Mart)實例

- 零售業之銷售資料集範例
- 醫療業之損益資料集範例

