Course Description  
In this course you WILL experience firsthand all of the PAIN a Data Scientist goes through on a daily basis. Corrupt data, anomalies, irregularities - you name it!

This course will give you a full overview of the Data Science journey. Upon completing this course you will know:

* How to clean and prepare your data for analysis
* How to perform basic visualisation of your data
* How to model your data
* How to curve-fit your data
* And finally, how to present your findings and wow the audience

This course will give you so much practical exercises that real world will seem like a piece of cake when you graduate this class. This course has homework exercises that are so thought provoking and challenging that you will want to cry... But you won't give up! You will crush it. In this course you will develop a good understanding of the following tools:

* SQL
* SSIS
* Tableau
* Gretl

**What you’ll learn**

* Successfully perform all steps in a complex Data Science project
* Create Basic Tableau Visualisations
* Perform Data Mining in Tableau
* Understand how to apply the Chi-Squared statistical test
* Apply Ordinary Least Squares method to Create Linear Regressions
* Assess R-Squared for all types of models
* Assess the Adjusted R-Squared for all types of models
* Create a Simple Linear Regression (SLR)
* Create a Multiple Linear Regression (MLR)
* Create Dummy Variables
* Interpret coefficients of an MLR
* Read statistical software output for created models
* Use Backward Elimination, Forward Selection, and Bidirectional Elimination methods to create statistical models
* Create a Logistic Regression
* Intuitively understand a Logistic Regression
* Operate with False Positives and False Negatives and know the difference
* Read a Confusion Matrix
* Create a Robust Geodemographic Segmentation Model
* Transform independent variables for modelling purposes
* Derive new independent variables for modelling purposes
* Check for multicollinearity using VIF and the correlation matrix
* Understand the intuition of multicollinearity
* Apply the Cumulative Accuracy Profile (CAP) to assess models
* Build the CAP curve in Excel
* Use Training and Test data to build robust models
* Derive insights from the CAP curve
* Understand the Odds Ratio
* Derive business insights from the coefficients of a logistic regression
* Understand what model deterioration actually looks like
* Apply three levels of model maintenance to prevent model deterioration
* Install and navigate SQL Server
* Install and navigate Microsoft Visual Studio Shell
* Clean data and look for anomalies
* Use SQL Server Integration Services (SSIS) to upload data into a database
* Create Conditional Splits in SSIS
* Deal with Text Qualifier errors in RAW data
* Create Scripts in SQL
* Apply SQL to Data Science projects
* Create stored procedures in SQL
* Present Data Science projects to stakeholders

Course Pathways

Section 2: Data Science

Section 3: Part 1: Visualisation

Section 4: Introduction to Tableau

Installing Tableau Desktop: [www.tableau.com](http://www.tableau.com)

Sample Data Source: [www.superdatascience.com/pages/training](http://www.superdatascience.com/pages/training)

Connecting Tableau to Data File: Open Tableau, on left click Connect 🡪 Text File, Select officeSupplies.csv

After opening the text file: Sorted fields are in the arranged in order of String, Date, Number for all data’s.

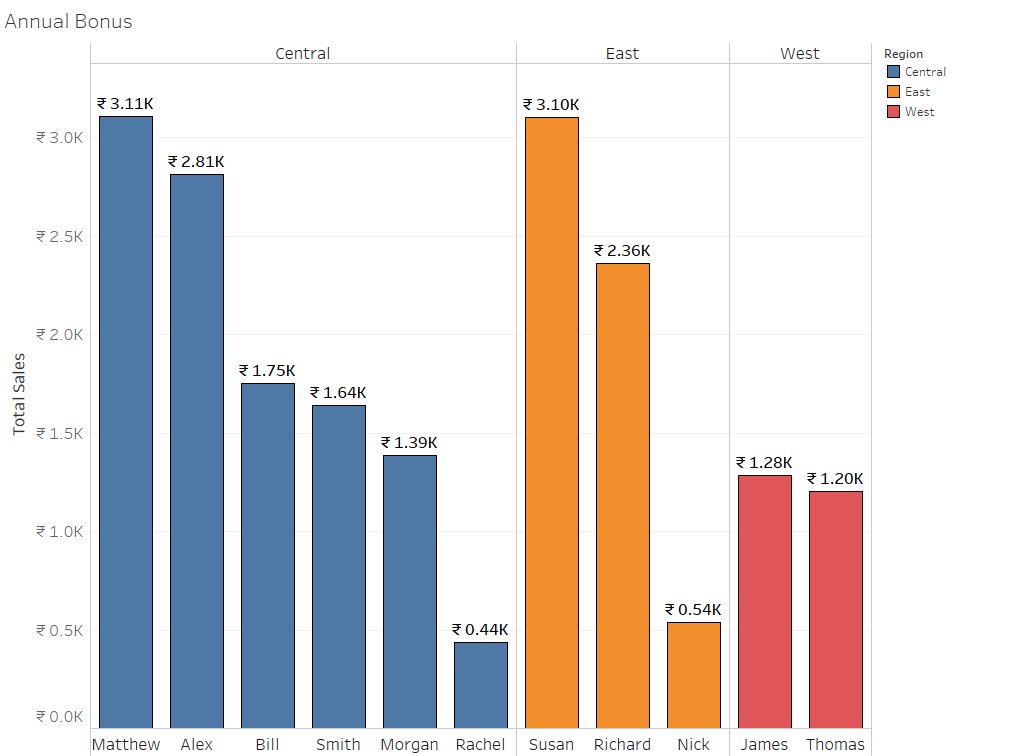
14. Creating a calculated field: Right Click on Measures, Create Calculated Field, Name the Field: Use [] to get the values of Measures. Example: [Units]\*[UnitPrice] 🡪 Click OK to get the TotalSales

15. Colors: i). Drag and Drop Dimension field on Color, this will give each value fo field a different color OR ii). Hold the Click, and pull Field, bring it on Top of the datas and press control, holding the click, a plus sign will appear and move and drop it on top of Color. (if we have multiple columns and rows, easy way)

We can edit color and use a palette

16. Label: Drag and Drop, press control and drop on to Label. If Label is not Shown, Right click on Label 🡪 Mark Label 🡪 Always Show

Right Click on Label or Axis, Click Format to change to currency, numbers, change decimal places

Sample: 

Section 5: How to use Tableau for Data Mining

Section 6: Advance Data Mining with Tableau

Section 7: Part 2: Modelling

Section 8: Stats Refresher

Section 9: Simple Linear Regression

Section 10: Multiple Linear Regression

Section 11: Logistic Regression

Section 12: Building a robust geodemographic segmentation model

Section 13: Assessing your model

Section 14: Drawing insights from the model

Section 15: Model Maintenance

Section 16: Part 3: Data Preparation

Section 17: Business Intelligence (BI) Tools

Section 18: ETL Phase 1: Data Wrangling before the Load

Section 19: ETL Phase 2: Step-by-step guide to uploading data using SSIS

Section 20: Handling errors during ETL (Phase 1 & 2)

Section 21: SQL Programming for Data Science

Section 22: ETL Phase 3: Data Wrangling after the load

Section 23: Handling errors during ETL (Phase 3)

Section 24: Part 4: Communication

Section 25: Working with people

Section 26: Presenting for Data Scientists

Section 27: Homework Solutions

Section 28: Bonus Lectures

