

AICPA & CIMA Data Analytics Summer School

10 - 14 August at Gonville & Caius College, Cambridge, UK

CPE/CPD: 40+ hours

Day 1: Data Analysis Fundamentals

9:00 Opening remarks

9:05 **Disruption**

- Identify hurdles to becoming a data-driven organization
- Analyze data practices in the organization
- Identify how data can benefit the organization
- Distinguish how to be a proactive data practitioner

9:25 **Business Analysis Principles**

- Identify the categories of analytical people
- Distinguish and define roles and responsibilities of professionals in data analysis

9:45 **Data Driven Decision Making**

- Identify cultural barriers
- Distinguish solutions to cultural and cross-functional barriers
- Identify six steps of the data-driven decision-making model

10:05 Morning coffee break

10:30 **Ecosystem - Relational Databases**

- Identify the nature of relational databases
- Recall the purpose of the SQL language
- Distinguish between the key aspects of ACID
- Identify the meaning of ETL

11:30 **Ecosystem – Not Only SQL**

- Distinguish between big data and other data storage tools
- Identify ways of interacting with MongoDB
- Distinguish between document stores and graph stores

12:30 Lunch

<https://informaconnect.com/aicpa-cima-data-analytics-summer-school/>

13:30 **Ecosystem – Big Data**

- Identify key functions of big data technologies
- Determine the utility of Hadoop
- Identify the purpose of MapReduce

14:15 **Ecosystem – Statistical Tool, Machine Learning, and Data Visualization**

- Identify tools for statistical analysis
- Distinguish between Python and R for statistical analysis purposes
- Identify purpose of machine learning
- Distinguish between visualization tools

15:15 Afternoon break

15:40 **Data Lifecycle Management**

- Identify the stages in the data life cycle
- Distinguish between ways that data enters the organization
- Identify the forms data takes as it is stored and used within the organization
- Identify the stages in the data life cycle

16:10 **Requirements Gathering**

- Analyze why requirement gathering process is critical to proper analysis
- Distinguish between the ways data is consumed (the three V's of data)
- Understand how requirement gathering fits with the development of a customer journey map
- Distinguish between the stages of the customer journey map

17:10 End of day

Day 2: Application of Data Analysis Essentials

9:00 Opening remarks

9:05 **Types of Variables**

- Determine the nature of variables in data analysis
- Differentiate between numerical and categorical variables
- Distinguish between nominal and ordinal variables
- Differentiate between interval and ratio
- Distinguish between continuous and discrete

9:50 **Central Tendency of Data**

- Identify the components of central tendency
- Calculate mean/median/mode
- Identify the steps in calculating weighted/geometric/harmonic means

10:20 Morning coffee break

10:40 **Measurement and Variability**

- Determine core aspects of measurement and variability
- Calculate range
- Calculate quartiles
- Calculate interquartile range
- Calculate variance
- Calculate standard deviation
- Analyze permutation with repetition
- Analyze combinations without repetition

11:40 **Basic Probability**

- Analyze the uses of probability
- Differentiate between sample space, event, independent and dependent
- Calculate probability

12:10 **Probability and Ven Diagramming**

- Analyze “this” OR “that” diagram
- Analyze “this” AND “that” diagram
- Analyze exclusive diagram
- Calculate join probability
- Calculate conditional probability

12:55 Lunch

13:45 **Calculating Probability**

- Calculate P using a contingency table
- Calculate P from trees
- Calculate Bayes’ theorem
- Calculate the mean in terms of probabilities
- Calculate the variance and standard deviation in terms of probabilities

14:45 Distributions

- Analyze distributions
- Identify discrete distributions
- Calculate binomial distributions
- Calculate Poisson distributions

15:30 Afternoon break

15:45 Continuous Distributions

- Identify continuous distributions
- Calculate continuous distributions
- Identify cumulative distributions
- Identify normal distributions
- Calculate normal distributions
- Compare quartiles and normal distributions
- Identify skew

16:30 Statistics in R Case Study

- Apply Vectors in R
- Use Data Frames in R
- Use data from an external file in R
- Apply mean/median/standard deviation in R

16:50 Distributions in R Case Study

- Functions for working with:
- Use Normal distribution function in R
- Use Poisson distribution function in R
- Apply Scatter plot in R
- Apply Histogram in R
- Apply Box Plot in R

17:10 Fraud Detection Case Study

- Apply scripts in R
- Create reusable, user defined function in R
- Use Bayes' Theorem in R
- Choose a function flexible to allow for different input parameters

17:30 End of day

Day 3: Data Analytics Modeling

9:00 Opening remarks

9:05 Understanding Your Business

- Identify business value
- Determine how business and corporate drivers impact the strategic direction of the business
- Analyze different project processes used in working with data
- Compare different types of data

9:35 Data Profiling

- Identify core data profiling tasks
- Identify outliers
- Use tools for data profiling

10:20 Morning coffee break

10:40 Data Cleansing

- Identify core data cleansing tasks
- Use tools for cleansing

11:30 Data Quality and MDM

- Identify role of data quality in organization
- Identify role of MDM in organization
- Use tools for data quality and MDM

12:30 Lunch

13:30 ETL Tools

- Distinguish between ETL processes
- Use Talend Data Integration
- Use MSSQL SSIS

14:30 Data Warehousing

- Identify purpose of data warehousing
- Identify between key components of a data warehouse
- Distinguish between data warehouses and data lakes
- Determine the role of different warehousing techniques

15:00 Afternoon break

15:15 **Data Warehousing Tools**

- Differentiate between utility of relational DW, cubes, and in-memory scenarios
- Compare techniques for data integration with regards to warehousing
- Use warehousing tools
- Use integration tools for warehousing

15:45 **Understanding Data – Overview**

- Identify data roles in the organization
- Determine how data moves through the data lifecycle

16:15 **Understanding Data – Data Modeling**

- Identify the role of data modeling in the organization
- Analyze data modeling techniques
- Use tools for data modeling

17:15 End of day

Required post-work:

Understanding Data – Structured Data Tools (90 mins)

Identify core tools for RDBMS's (structured storage)
Use SQL to perform CRUD tasks against a database

Understanding Data – Unstructured Data Tools (90 mins)

Identify tools in unstructured stack
Use tools for unstructured data management

Day 4: Forecasting and Predictive Analytics

9:00 Opening remarks

9:05 **Bivariate regression**

- Identify linear regression models and their application in data analysis

- Calculate linear bivariate regressions in order to model commonly encountered problems
- Determine the quality of fit of a linear model, establishing its usefulness for prediction

10:20 Morning coffee break

10:45 **Multivariate regression**

- Identify multivariate linear regression models in data analytics
- Calculate linear multivariate regressions in order to model commonly encountered problems
- Determine the quality of fit of a multivariate linear model, establishing its usefulness for prediction

12:00 **Other regressions**

- Identify logistic regression models and apply those models to common business problems
- Determine the quality of fit of a multivariate linear model

12:45 **Forecasting in time**

- Identify the components of a time forecast in order to predict future values from a model
- Differentiate seasonal variations from trends in order to improve prediction of future values from a model
- Calculate seasonal indices so that seasonal variations can be qualified in the model

13:15 Lunch

14:00 **Classification K-Nearest Neighbor (KNN)**

- Determine core aspects of classification in order to understand when it's an appropriate technique
- Calculate kNN algorithm with fixed or variable number of k and assess the quality of the results

14:45 **Classification Naïve Bayes**

- Identify Naïve Bayes classification and when it is applicable
- Apply a Naïve Bayes classification to a common business problem

15:15 Afternoon break

15:30 **Clustering**

- Determine core aspects and types of clustering in order to properly apply the algorithms to business problems
- Apply various clustering algorithms to data sets in order to solve common, applicable business problems

16:15 **Optimization**

- Identify the goals and constraints of a linear optimization
- Calculate a linear optimization in order to solve a business problem

17:00 **Simulation**

- Use data analysis performed on historical data and any applicable theory to construct a model
- Use a Monte Carlo analysis with the model in order to generate and assess the likelihood of predictions from the model

17:45 End of day

Required post-work:

Measuring Distance (45 mins)

Identify the meaning of dimensionality and its implications in analytics
Calculate different types of distances and identify scenarios when each type is applicable

Classification Support Vector Machine (1 hour)

Identify the basics of the support vector machine (SVM) classification algorithm
Identify what a support vector is and its use in the classification algorithm
Calculate a SVM classification to solve common business problems

Classification Decision Trees (1 hour)

Identify the steps to build a decision tree classifier
Apply the steps to create a basic decision tree
Use a decision tree algorithm and appropriate metrics to solve a business problem and assess the quality of the solution

Day 5: Data Visualisation

9:00 Opening remarks

9:05 Chart and Graph Selection

- Differentiate between the following graph and chart types and when to use them:
 - Categorical
 - Continuous
 - Time series
 - Bivariate
 - Distribution

9:50 Introduction to Business Intelligence (BI)

- Identify business intelligence (BI) concepts and their common applications
- Differentiate between the major enterprise BI platforms and when each is appropriate
- Apply BI techniques (online analytical processing [OLAP], drill through, drill down, drill up)

10:35 Morning coffee break

11:00 Report Building and Online Analytical Processing (OLAP)

- Apply techniques for building reports and identify the best uses of the learned techniques
- Distinguish the elements of OLAP (online analytical processing)

12:00 Business Intelligence in Action

- Identify the steps and benefits of scorecard and strategy management and apply the steps in appropriate situations
- Distinguish between effective dashboard techniques

12:45 Lunch

13:45 Mashups

- Identify event streaming techniques and their uses
- Differentiate between types of mashups in order to determine the best mashup for a dashboard presentation
- Analyze trusted data in order to derive information

14:30 **Data Visualization Best Practices**

- Identify core data visualization rules in order to ensure quality report development
- Identify the goals of data visualization and how data visualization can benefit an entity
- Differentiate between the needs of different report users and identify appropriate report styles for each

15:15 Afternoon break

15:30 **Reporting Options**

- Differentiate between alerts and trends and the appropriate uses for each
- Distinguish the appropriateness of Self-Serve reports and their application in common scenarios
- Differentiate between best uses for BI techniques in order to ensure effective reporting

16:00 **What Makes a Good Report**

- Compare layout options and identify appropriate presentation choices for maximum effectiveness
- Differentiate between good and bad report implementations and the potential impact of poor implementation

16:30 **Constructing Reports**

- Distinguish between when to use reporting tools versus analysis tools
- Use core tools to build report and dashboard layouts to achieve maximum effectiveness

17:30 End of day