

## Data Analysis Training Course for Professionals in Career Transition

Chief Lecturer: *Michael Peng, PhD*  
Managing Director, *Witzcredit.com*  
Tel 2153009631  
[Michael.Peng@Witzcredit.com](mailto:Michael.Peng@Witzcredit.com)  
[mkpeng2007@gmail.com](mailto:mkpeng2007@gmail.com)

Data Analysis has become a critical and integral part of business decision making. To be successful in today's business landscape, all companies need to enhance their data analysis capability. The job prospects in data analysis are getting brighter and brighter. This overview and introductory course is designed to prepare future data scientist/analyzer and (entry level) quantitative modeler (who were not working in the area) with necessary knowledge, experience and confidence to be 'self-propelled' into to higher altitude in this fast growing area. The goal is to help students 'over the hump' (i.e. taking the most challenging first few steps) in their career transition. For those who are interested in advanced trainings in quantitative modeling, we will provide customized follow-up courses. While data analysis and modeling have been widely used in any field of business, we will focused on applications in three major sectors: 1) Finance, 2) Pharmacy and 3) Marketing/Market research.

*Who should take this course?*

- Business/Banking Analysts who want to shift their career path towards more technical and data related work (within or outside her current company);
- Professionals with hard science majors (Pharmaceutical/Medical, Physics/chemistry and Engineering) who wants to enhance their data mining and statistical skills.
- DBA and system support professional who want to gain knowledge beyond relational database and want to step into data driven business analytics
- Accountant/finance professionals who want to 'upgrade' their number crunching skill to a more analytical level.
- Anyone who wants to get a practical warm-up training before embarking on a new journey of becoming a certified data scientist, with confidence boosted ;
- Any non-quantitative professionals who wants to learn about *fundamental* skills and knowledge of *data-driven decision process*
- Anyone who would like to 'upgrade her/his current role and apply to an higher level internal job position (within the firm) that requires data analysis and modeling skills
- Anyone who wants to increase her/his work performance/efficiency by automating –via basic programming--data processing previously done manually

### *What Can I expect from the course?*

By the end of the course, you can expect to:

- be able to conduct common “data munging or data wrangling “ (i.e. converting or mapping data from one "raw" form into another format that allows for further analysis) ;
- be able to acquire, parse and query data using: Excel/VBA, plus sophisticated SQL.
- be able to conduct common types of data analysis and basic statistical modeling
- become tuned into basic computer programming and acquainted with at least one useful tool (SAS, Python and R), eliminating the fear as a “programmer” in data processing.
- become confident and ready to step deeper into data modeling and quantitative modeling area (either to get specialized certification or to take a formal advanced degree program)
- be ready to answer interview questions for a junior position in data science area;
- get a significant boost in analytical & data-presenting skill in your current job

### *Important Note:*

0) We will provide a sealed assessment letter to each individual student at the end of the class regarding his/her performance and honest ranking among his/her peers. Opinions about his/her potential /prospect in this field will also be offered if requested. Suggestions for actions will also be included in this assessment letter.

1) For top 25% students, we will i) offer free consultation service on career paths in finance and quantitative area, including resume editing, interview coaching (e.g. how to answer the annoying HR questions such as “What do you; your weaknesses are?”) ii) provide references and help students find a related job via our networks and connections; iii) iv) consider partial refunding (up to 20% of tuition) if a proven financial stress (e.g. unemployment, medical) is presented.

2) Student will be given one-time opportunity to retake the course within 12 months (free of charge) under the following condition: **a)** the student failed the class measured by the end test and project performance; **b)** the student missed more than 1/3 of the class due to factors beyond his/her control (e.g. medical reason) ; **c)** the student expressed his/her desire to take a second run of the class in order to solidify or reinforce what he learned.

3) For all students we will provide advice and sources on further learning in this area and offer our help as to how to use the learned knowledge in their current and future work.

### *The characteristic style of the course:*

### *Coverage extent*

Since each item could be a huge topic, we don't intend to provide A-Z exposition on each one. Rather, we use case study to expose student directly to real problems and explain in exquisite detail about each aspect of the problem solving issues (including the programming part). If any prior knowledge is needed, we will assign it as home work before or after the related session.

### *Complementary learning*

Students are required to read related text (to be assigned at the end of each session) in the designated reference books (provided by the lecturer); and bring questions to class for discussion. It is estimated that class time is only half (or less) of the total time needed to be devoted to this class.

### *Tool & Programming language*

Considering the fact that different job positions have different requirements for tool used in data analysis, we are not restricted in one specific language (such as popular SAS) to students. Instead, we exposure student to several major languages (SAS, Python, VBA/Excel and R) and convince them that the logic are quite similar and not hard to master any of these tools. Once they got their feet into the door, they will have the confidence and be able to step much deeper to learn much more to meet the requirement of their job. It is emphasized here that *programming does not need advanced math or computer science knowledge, it is more of logic rules being artfully and creatively assembled together to solve a practical problem.*

It is our belief that learning by comparison is an effective way to understand the concepts— notwithstanding the risk of confusing the syntax of one language with another (say SAS and Python) --- but it is OK for this class because students will be able to know what exact syntax should be for a particular language to used in a real project.

### *Teaching method*

Course content will be delivered via class lecture, web seminar, topical discussions, computer demonstrations, and regular homework assignments involving data analysis exercises and programming applications. Students are expected to ask questions, share experiences, and to actively participate either in class or web discussions.

*Prerequisite:* College level math and statistics

*Hardware requirement:* a Laptop (64bit OS preferable but not necessary) with at least 2.5G memory, with Wifi capability.

*Software preparation:* 1) Microsoft Office and Adobe PDF) 2) SAS (a free version from SAS University will be used and the installment guide will be sent separately); 3) Python/Panda package (free, open source package, installment is easy buy a step-by-step guidance will be provided to student) ; 4) R package (self installment) .

*Cost:* \$2500

(Note: 1) Student has the choice of taking only some of the session (for some specific topic) instead of the whole course and the tuition will be prorated. For example, some remote students could have option to take only web sessions (4 out of 8 sessions will be in web sessions, with full access to full textbook and reference and all assignments, and up to 1 hours coaching/tutoring time/per session, and the total cost will be \$1600)

2) The price also includes: i) proprietary textbook, densely and meticulously written (rather than in cursory PPT style as in many other training courses); ii) 10 valuable, full length reference books that provide detailed supplementary material for each related topics in the class

*Refund policy:*

Full refund will be granted should the student expresses dissatisfaction about either the course itself or the instructor before the second session. 6/8, 5/8 refund will be granted respectively if student decided to withdraw after second session and third session. Future course credit will be given in lieu of refunding after forth session.

*Job placement:*

As specified in “Important Note “above, we will use our network and connections to help students find a data related positions and provide references via associated companies. We will also work with reputed recruiters in the tri-state (PA, NJ, NY) area to explore opportunities for our students.

Class Locations:

**New York: 32 Broadway Suite 1702** (near Wall Street bull), for the coming class,

Princeton: NJ, Princeton Public Library (the following class)

New Town, PA, Holy Family College

*Expected Starting Date: October 31, 2015*

### Class Arrangement and Schedule

Class will be run on each Saturday but will meet in person only every other Saturday and for the rest of Saturday

Time: 10:00-11:00 Session 1

11:00-11:20 Break

11:20-12:15 Session 2

12:15-12:50 Lunch (complementary)

12:50- 1:45 Session 3

1:45--2:00 Q and A

The following is a tentative class schedule, subject to adjustment and modification

### ***Session 1 (3H)***

1.1 An overview of Data Analysis: past, present and future

1.2 Data vs. Big Data: Is big data just a hype?

Examples of actual big data in business and finance

1.3 The scope and content of data analysis

Data→Knowledge pyramid

Data Analysis vs Quantitative Modeling

Data structure

- Basic concepts:
  - population, sample, observation, variable/field, dataset/record,
  - table and list, key, index
- The concept of relational database (in MS Access or MS SQL)
- What 's the problem with flat files;

Data Query

Raw data

Derived data field

Common tasks in data analysis

1.4. An overview of important statistical concepts (must-know knowledge)

i) review about key probability concept and statistical measure on data sample

ii) put statistics in perspective (can statistics lie?) : how to avoid spurious conclusion

([http://cseweb.ucsd.edu/~ricko/CSE3/Lie\\_with\\_Statistics.pdf](http://cseweb.ucsd.edu/~ricko/CSE3/Lie_with_Statistics.pdf))

Assignment from reference #2

1.5.Examples of data Analysis (in three area)

1.5.1 Banking (retail loan portfolio data analysis)

(Model parameter inputs/data preparation for bond credit risk modeling)

1.5.2. Pharmaceutical (Data preparation for quantitative modeling)

1.5.3 Marketing (Survey data processing)

1.5.2 (An application of basic statist concepts): Risk management/finance: Using SAS to Model Portfolio Drawdown, Recovery, and Measuring Risk (“Value at Risk”)

(reference : <http://www2.sas.com/proceedings/sugi29/199-29.pdf>)

1.6. Computer programming demystified: Your first taste of computer programming

(Thinking like a programmer without knowing too math!)

A super simple code example

*Reading* Assignment (before class):

Ref #6, page 3-5, Ref#3 2-10

*Practice* Assignment (*review*): <https://www.youtube.com/watch?v=AIhKNNXzZLM>

## **Session 2** Tools and Technique

2.1.. Introduction to several popular computer programming language in data analysis

a) A comparison of major programming language used in data analysis: SAS, R, Python, VBA---What you need to learn to get started in the area.

i) *VBA/Excel*: What is a VBA Macro?

ii) *SAS*: Is SAS going to be phased out?

iii) *Python*: Why is Python so popular in data analytics world?

iii) *R*: Do you need to know a little R?

iv) *SQL*: Why SQL can be a powerful yet intuitive tool in parsing data?

What if the data set is so large?

b) A simple but practical programming example:

(Will provide an overview and demonstration of programming example)

c) Illustrative Example: Risk management/finance(Model Parameter inputs/data preparation for bond credit risk modeling) with excel VBA, R, Python.

d) Project: *Read in data file from extern sources*

*(Class will demonstrate how this is done via SAS, R, and leave Python implementation as homework)*

2.2. More on programming concepts and their use in data manipulation

2.2.1. How to make computer do repetitive data work under different condition (Loop and If.. command)

A practical example data processing automation: reconciliation of two transaction files

2.2.2 Concept of ‘Procedure’ (as compared to Object –oriented programming)

SAS Procedure in detail

An example of Object-oriented programming (in R or Python)

**Project 2:** Create descriptive statistics summary for a sample of large data set

2.2.3 . SQL in 360 degree review: how to combine

An anatomy to a complicated query statement .

**Session 3** Case Study (I)

Solving a data real data query problem with SAS

( A Step-by-Step explanation on an actual SAS procedure with sophisticated SQL,

1.4.1 Banking (retail loan portfolio data analysis)

**Session 4 :** Important tasks in pre-processing data for further analysis

(Python as compared with SAS)

4.1 Data Wrangling: Cleaning, Transform, Merge, Reshape

Chapter 7, page 175 Ref# 5:

4.2 Interacting with database

Page 172-173, Ref#5

**Session 5 Case Study (II)**

*Solving a real problem in R as compared to Python 1.4.4*

*Risk management/finance: Using SAS to Model Portfolio Drawdown, Recovery, and Measuring Risk (“Value at Risk”)*

**Session 6 Advanced topic (I) :** *From data processing to business intelligence*

Linea regression analysis and logistic modeling

(SAS and R)

**Session 6 Advanced Topic (II) Financial modeling**

The whole process in financial data modeling : An Example in rating scoring model

( SAS)

**Session 7**

An example in Marketing : Handling survey data and discrete data



**Session 8** Project Discussion and comprehensive review of the course.

**Exam/Test** (40 minutes, closed book)

**Reference:**

1. Introduction To Data Science (With R)  
[https://ischool.syr.edu/media/documents/2012/3/DataScienceBook1\\_1.pdf](https://ischool.syr.edu/media/documents/2012/3/DataScienceBook1_1.pdf)
2. Statistics in a nut shell
3. Thinking like a programmer
4. Practical Data Science with R
5. Python for Data Analysis
6. Automate the boring Stuff with Python