# TM351

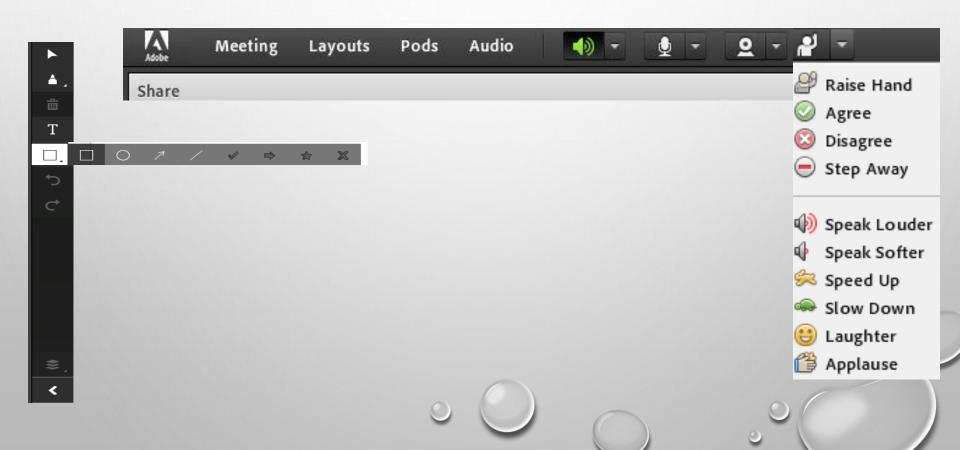
DATA MANAGEMENT AND ANALYSIS

Jo Buxton, 2017

# GET INVOLVED...

- EVERY SO OFTEN I'LL ASK YOU TO VOTE OR CONFIRM, USE THE ICONS TO DO SO
- CLICK TALK IF YOU WANT TO SPEAK (MAX 4 SPEAKING AT ANY TIME)
- YOU CAN ALSO TEXT QUESTIONS AS WE GO ALONG AND I'LL TRY AND RESPOND
- WE'LL EXPERIMENT WITH FEATURES OF ADOBE CONNECT AS WE GO!
- WE WILL HAVE A GENERAL DISCUSSION AT THE END OF EACH SECTION

# ADOBE CONNECT – A QUICK OVERVIEW!



# WHAT WE'LL COVER TODAY...

- MODULE WEBSITE
- SOFTWARE AND THE VM
- NOTEBOOKS
- OVERVIEW OF THE COURSE MODULE
- ASSESSMENT: ICMAS, TMAS AND EMA.
- TUTOR CONTACT AND TUTORIALS



- JO BUXTON
- SHARON DAWES
- MARION EDWARDS
- ROGER SHAW

- KEITH HASTELEY
- AVON HUXOR
- STEVEN SELF

- YOU CAN ATTEND ANY OF THE TUTORIALS IN YOUR CLUSTER.
- MIXTURE OF FACE TO FACE AND ONLINE

# WHICH TUTOR GROUP ARE YOU IN?

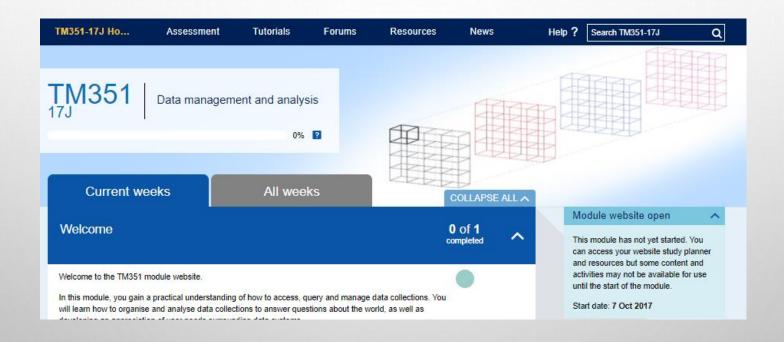
Jo Buxton	Sharon Dawes	Keith Hasteley	Steven Self
Marion Edwards	Roger Shaw	Avon Huxor	

# MODULE WEBSITE

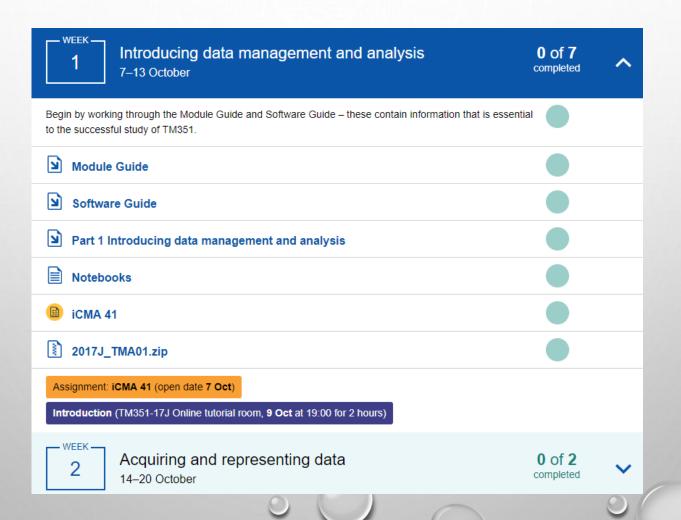
# MODULE MATERIAL

- MATERIALS (PARTS 1 − 26)
- NOTEBOOKS
- ICMAS
- EACH PART CAN BE VIEWED AS A SINGLE PAGE AND SAVED AS PDF FOR THOSE THAT LIKE TO HAVE AN OFF-LINE VERSION OR PRINT OFF SECTIONS.

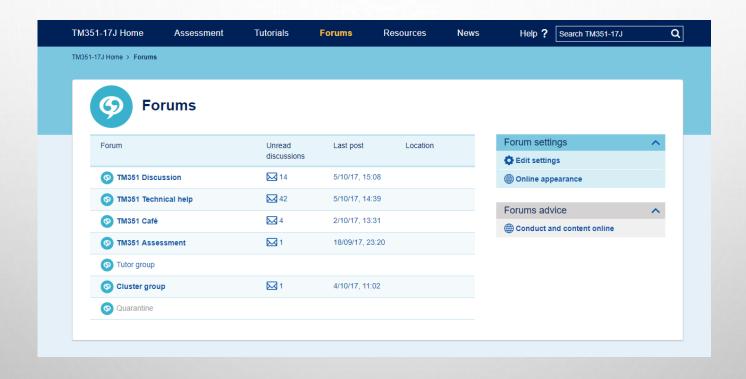
# MODULE WEBSITE



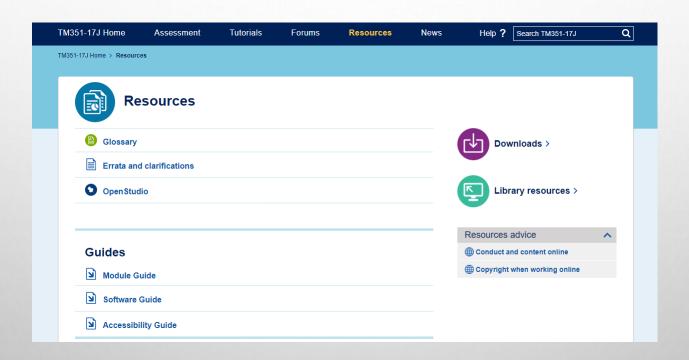
# **CONTENTS**



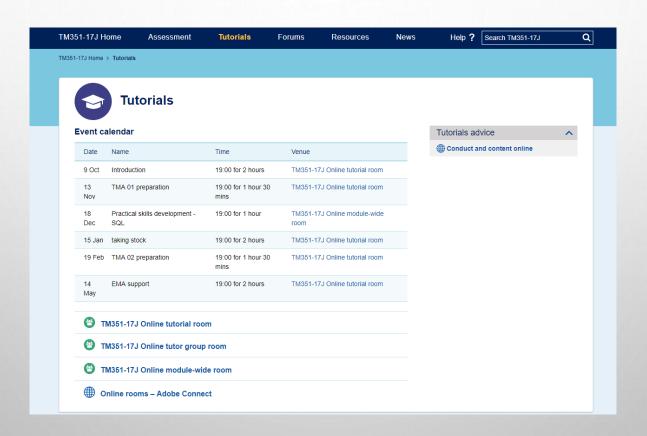
# **FORUMS**



# **RESOURCES**



# **TUTORIALS**



# SOFTWARE AND THE VM

# **SOFTWARE**

•	SOFTWARE	VERSION DOCUMENTATION
		TENDIOI 1 DOCUMENTATION

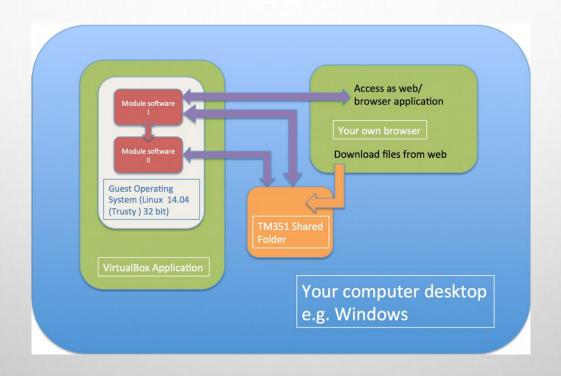
• P	PANDAS	0.20.2	HTTP://PANDAS.PYDATA.ORG/
-----	--------	--------	---------------------------

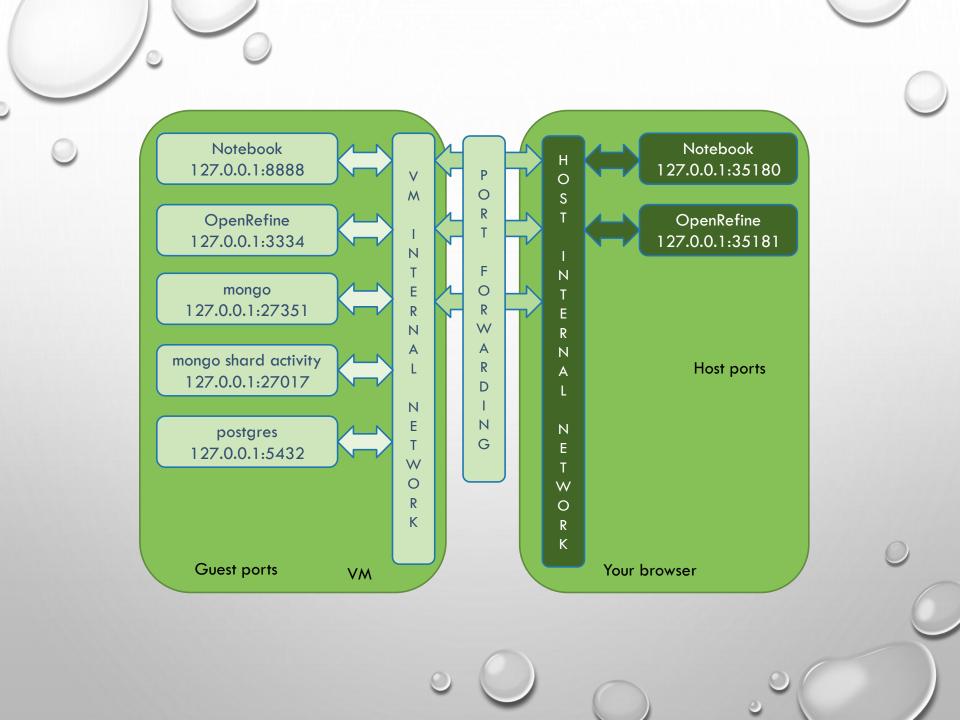
- PYTHON 3.5.2 HTTPS://DOCS.PYTHON.ORG/
- POSTGRESQL
   9.5.7
   HTTP://WWW.POSTGRESQL.ORG/DOCS/
- MONGODB 3.4.4 HTTPS://DOCS.MONGODB.ORG/MANUAL/
- OPENREFINE 2.7 HTTP://OPENREFINE.ORG/DOCUMENTATION.HTML

# INTRODUCTION

- INSTALL THE VM AND NOTEBOOKS
- OVERVIEW OF THE COURSE
- ASSESSMENT

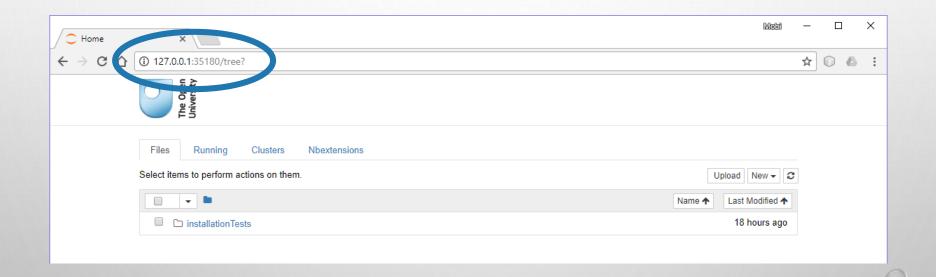
# VM AND SOFTWARE





# NOTEBOOKS

# **NOTEBOOKS**





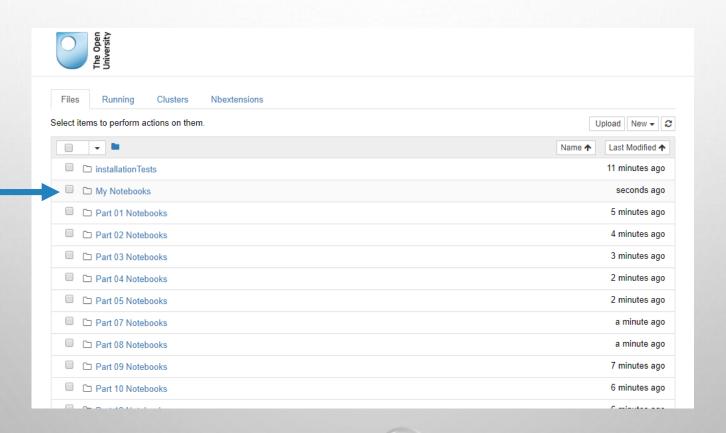
NAVIGATE TO
INSTALLATION TEST
AND FOLLOW
INSTRUCTIONS TO
MAKE SURE
EVERYTHING IS
RUNNING OK

## TM351 VM Installation Test This notebook provides a series of tests to ensure that the virtual machine is running correctly. Run each cell in turn by clicking the play button or keyboard shortcut shift-return. (A full list of keyboard shortcuts can be found from the Help menu or via the keyboard shortcut ESC-h.) The cells should run without error. Versions Display the VM build version and build time, as well as database service versions and pandas version. In [1]: !cat /opt/version.txt Version: TM351\_17J\_0.01 - build time: 2017-06-07/09:31.03. In [2]: psql --version psql (PostgreSQL) 9.5.7 In [3]: mongod --version db version v3.4.4 git version: 888390515874a9debd1b6c5d36559ca86b44babd OpenSSL version: OpenSSL 1.0.2g 1 Mar 2016 allocator: tcmalloc modules: none build environment: distmod: ubuntu1604 distarch: x86\_64 target\_arch: x86\_64 mport pandas as pd d.\_\_version\_\_ Out[4]: '0.20.2' ▼ Test Core Packages

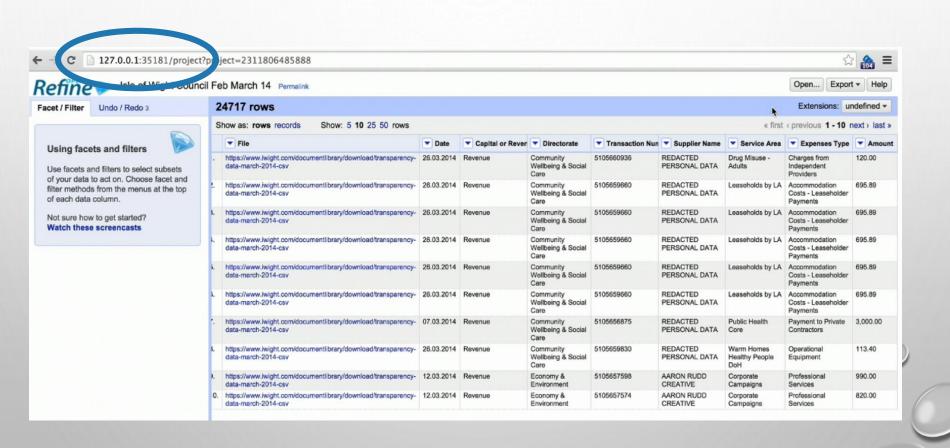
# DOWNLOAD AND INSTALL NOTEBOOKS

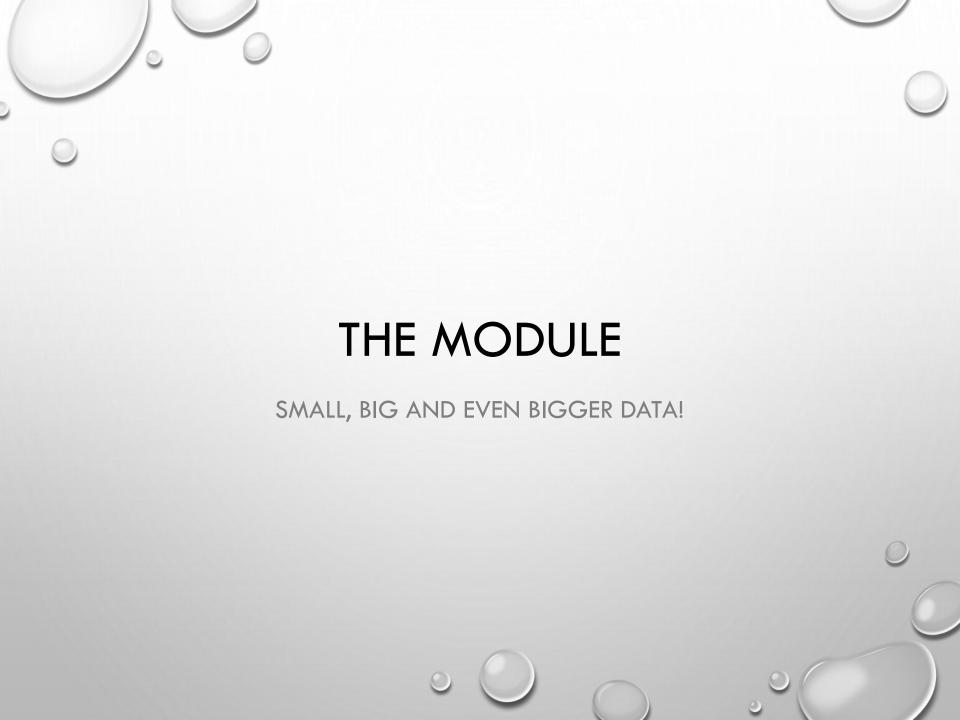
- LINK TO NOTEBOOKS ZIPS FROM WEEK 1
- MUST INSTALL TO THE TM351VM\NOTEBOOKS FOLDER (SHARED WITH VM)
- UNZIP /CREATE FOLDERS IF NEEDED

# ADD PERSONAL FOLDER



# **OPENREFINE**





# CONTENT

Subject	Weeks	Parts
Introduction	1	1
Small data	2-7	2-7
Bigger data	8-25	8-23
Even bigger data	26-31	24-26





- INTRODUCING DATA MANAGEMENT & ANALYSIS
- ACQUIRING AND REPRESENTING DATA
- DATA PREPARATION
- DATA ANALYSIS
- PRESENTATION: TELLING THE STORY
- WITH DATA COME RESPONSIBILITY
- WHEN SPREADSHEETS FAIL

# **BIGGER DATA**

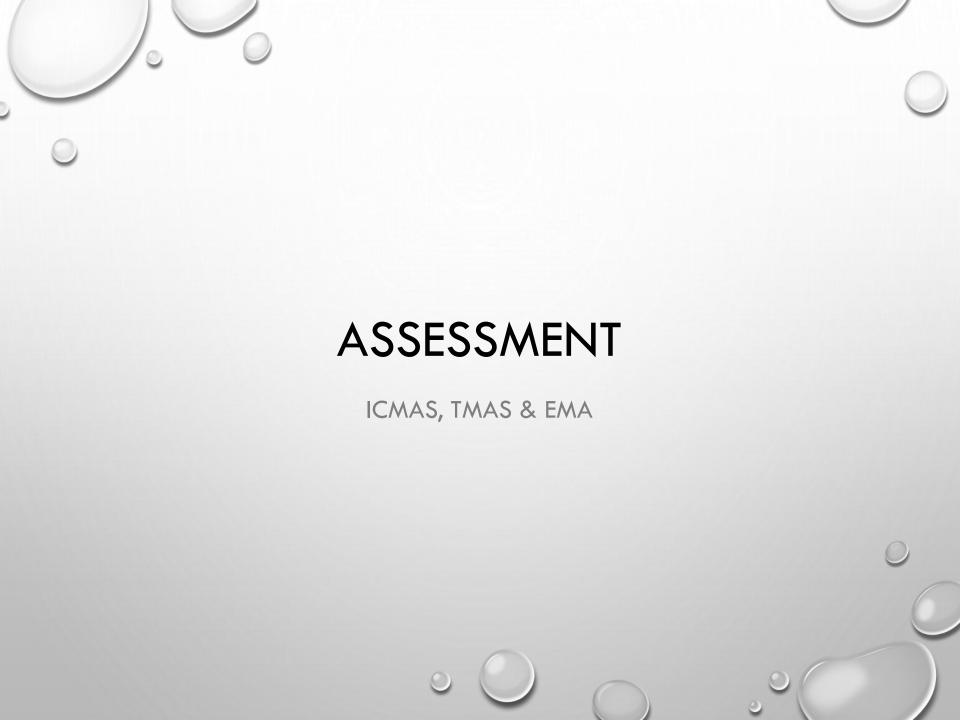
- DATABASES
- RELATIONAL DATABASES
- NORMALISATION
- DATA INTEGRITY
- ALTERNATIVES TO RELATIONAL DATABASES
- DOCUMENT DATABASES
- COMPLEX QUERIES AND ANALYSIS

# ...BIGGER DATA

- SCALING OUT: REPLICATION AND SHARDING
- DISTRIBUTED TRANSACTIONS
- DATA WAREHOUSING
- DATA MINING: CLASSIFICATION
- DATA MINING: CLUSTERING

# EVEN BIGGER DATA

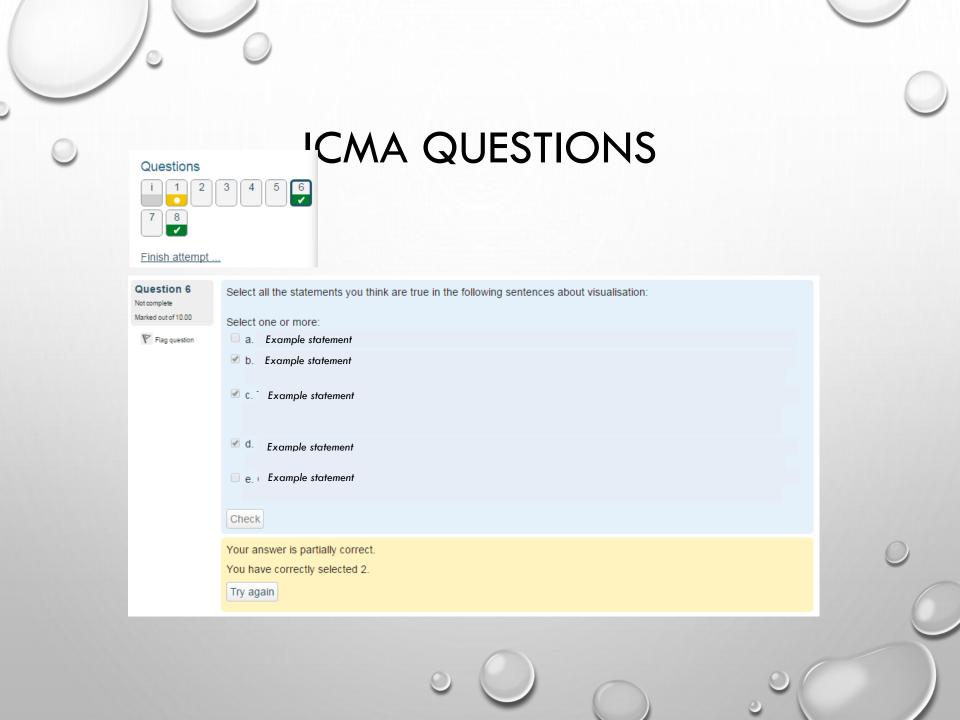
- LINKED DATA ON THE SEMANTIC WEB
- QUERYING AND MANIPULATING LINKED DATA
- APPLICATION OF SEMANTIC WEB DATA



# **ICMAS**

iCMA	Cut off date
iCMA41	26 Oct 2017
iCMA42	23 Nov 2017
iCMA43	11 Dec 2017
iCMA44	1 Feb 2018
iCMA45	1 Mar 2018
iCMA46	20 Apr 2018
iCMA47	10 May 2018

- DO NOT COUNT TO FINAL COURSE MARK BUT
- MUST PASS 5/7 TO PASS THE MODULE
- 30% PASS MARK
- CAN RETAKE AS MANY TIMES AS YOU LIKE
- QUESTIONS WILL CHANGE
- NO EXTENSIONS, MUST SUBMIT BY THE DEADLINE





Question 2	Drag and drop terms	s form the list below	v to complete the foll	owing sentences.				
Tries remaining: 2	text text text text text text text text							
Marked out of 10.00	lext text to	text text	text text text text text	text text text text text				
Flag question text text text text text text text tex					text text text	text text text text text tex		
		is	text text text text t	ext text text tex				
	text text text text te	combine	text text text text t	t	text text text text text text text text			
		text text text text tex	t text text text text text	text text text text text	text text text text text	text text text text te		
	text text te	text text	text text text text t	ext ttext text text text	text text text text	text text text text te		
	theta-join	output	joins	aggregation	inner	combine		
	cartesian product	inequality	equi-join	false equality	equality	full-outer		
	select	left-outer	project	true	horizonal	condition		
	Ohaak							
	Check							
	Next							

# TMAS/EMA

TMA/EMA	Cut off date
TMA 01	07 Dec 2017
TMA 02	08 Mar 2018
EMA	04 Jun 2018

- NEED 40% OVER THE TWO TMAS
- 40% ON THE EMA

# FULL ASSESSMENT DATE LIST

Assessment	Cut off date
iCMA41	26 Oct 2017
iCMA42	23 Nov 2017
TMA 01	07 Dec 2017
iCMA43	11 Dec 2017
iCMA44	01 Feb 2018
iCMA45	01 Mar 2018
TMA 02	08 Mar 2018
iCMA46	20 Apr 2018
iCMA47	10 May 2018
EMA	04 Jun 2018



# LEARNING EVENTS (TUTORIALS)

- INTRODUCTION
- PRACTICAL SKILLS DEVELOPMENT PYTHON
- PRACTICAL SKILLS DEVELOPMENT PANDAS
- TMA01 PREPARATION
- PRACTICAL SKILLS DEVELOPMENT SQL
- TAKING STOCK
- TMA02 PREPARATION
- EMA SUPPORT

# **TUTORIAL TYPES**

- ONLINE AND FACE TO FACE
- CLUSTER AND MODULE WIDE
- THREE ROOMS:
  - ONLINE TUTORIAL ROOM FOR CLUSTER TUTORIALS
  - ONLINE TUTOR GROUP FORUM FOR ONLINE MEETINGS WITH YOUR TUTOR
  - ONLINE MODULE-WIDE ROOM FOR MODULE TUTORIALS

# MY TUTORIALS

Date	Name	Time	Venue
9 Oct	Introduction	19:00 for 2 hours	TM351-17J Online tutorial room
13 Nov	TMA 01 preparation	19:00 for 1 hour 30 mins	TM351-17J Online tutorial room
18 Dec	Practical skills development - SQL	19:00 for 1 hour	TM351-17J Online module-wide room
15 Jan	Taking stock	19:00 for 2 hours	TM351-17J Online tutorial room
19 Feb	TMA 02 preparation	19:00 for 1 hour 30 mins	TM351-17J Online tutorial room
14 May	EMA support	19:00 for 2 hours	TM351-17J Online tutorial room

# FACE TO FACE TUTORIALS (CLUSTER)

Description	Tutor	Date	Time	Location	Site
Taking stock	Keith Hasteley	Sat 16 Dec 17	11:00	Reading	Reading College
Taking stock	Marion Edwards	Sat 13 Jan 18	10:30	London	Kings College Waterloo Campus
Taking stock	Sharon Dawes	Sat 20 Jan 18	10:30	Brighton	University of Brighton
Taking stock	Roger Shaw	Sat 03 Feb 18	10:30	London South East	Kings College Waterloo Campus
Taking stock	Avon Huxor	Sat 03 Feb 18	11:00	Bristol	St Mary Redcliffe School
Taking stock	Steven Self	Sat 24 Feb 18	11:00	Southampton	Richard Taunton Sixth Form College

Event Type	Description	Tutor	Date	Time	Location
Cluster	Introduction	Jo Buxton	Mon 09 Oct 17 (Mon)	19:00	Online
Cluster	Introduction	Roger Shaw	Thu 12 Oct 17 (Thu)	19:30	Online
Cluster	Introduction	Sharon Dawes	Sat 14 Oct 17 (Sat)	16:00	Online
Module	Practical skills development – Python	Roger Shaw	Mon 16 Oct 17 (Mon)	19:30	Online
Cluster	Introduction	Keith Hasteley	Tue 17 Oct 17 (Tue)	19:30	Online
Module	Practical skills development – Pandas	Chris Thomson	Thu 19 Oct 17 (Thu)	19:30	Online
Module	Practical skills development – Python	Steve Abbott	Sat 21 Oct 17 (Sat)	10:00	Online
Cluster	Introduction	Avon Huxor	Sun 22 Oct 17 (Sun)	16:00	Online
Cluster	Introduction	Marion Edwards	Wed 25 Oct 17 (Wed)	19:30	Online
Module	Practical skills development – Pandas	Joanne Dunster	Thu 26 Oct 17 (Thu)	19:00	Online
Module	Practical skills development – Python	Sharon Dawes	Sun 29 Oct 17 (Sun)	16:00	Online
Cluster	Introduction	Steven Self	Wed 01 Nov 17 (Wed)	19:00	Online
Module	Practical skills development – Pandas	Marion Edwards	Wed 01 Nov 17 (Wed)	19:30	Online
Module	Practical skills development – Python	Avon Huxor	Sat 04 Nov 17 (Sat)	10:30	Online
Module	Practical skills development – Pandas	Andrew Gargett	Sun 05 Nov 17 (Sun)	14:00	Online
Cluster	TMA 01 preparation	Jo Buxton	Mon 13 Nov 17 (Mon)	19:00	Online
Cluster	TMA 01 preparation	Keith Hasteley	Tue 14 Nov 17 (Tue)	19:30	Online
Cluster	TMA 01 preparation	Roger Shaw	Thu 16 Nov 17 (Thu)	19:00	Online
Cluster	TMA 01 preparation	Sharon Dawes	Sat 18 Nov 17 (Sat)	16:00	Online
Cluster	TMA 01 preparation	Marion Edwards	Wed 22 Nov 17 (Wed)	19:30	Online
Cluster	TMA 01 preparation	Avon Huxor	Sun 26 Nov 17 (Sun)	16:00	Online
Module	Practical skills development – SQL	Mary Garvey/Charly Lowndes	Tue 28 Nov 17 (Tue)	19:00	Online
Cluster	TMA 01 preparation	Steven Self	Wed 29 Nov 17 (Wed)	19:00	Online
Module	Practical skills development - SQL	Colin Gallagher	Mon 04 Dec 17 (Mon)	19:30	Online



0.		mil and the management of the latter than the second and the secon			
Event Type	Description	Tutor	Date	Time	Location
Module	Practical skills development - SQL	Colin Gallagher	Mon 04 Dec 17 (Mon)	19:30	Online
Module	Practical skills development - SQL	Keith Hasteley	Sat 09 Dec 17 (Sat)	19:00	Online
Cluster	Taking stock	Keith Hasteley	Sat 16 Dec 17 (Sat)	11:00	Reading
Module	Practical skills development - SQL	Jo Buxton	Mon 18 Dec 17 (Mon)	19:00	Online
Module	Practical skills development - SQL	Steven Self	Wed 20 Dec 17 (Wed)	19:00	Online
Cluster	Taking stock	Marion Edwards	Sat 13 Jan 18 (Sat)	10:30	London South East
Cluster	Taking stock	Jo Buxton	Mon 15 Jan 18 (Mon)	19:00	Online
Cluster	Taking stock	Sharon Dawes	Sat 20 Jan 18 (Sat)	10:30	Brighton
Cluster	Taking stock	Roger Shaw	Sat 03 Feb 18 (Sat)	10:30	London South East
Cluster	Taking stock	Avon Huxor	Sat 03 Feb 18 (Sat)	11:00	Bristol
Cluster	TMA 02 preparation	Sharon Dawes	Sat 10 Feb 18 (Sat)	15:00	Online
Cluster	TMA 02 preparation	Keith Hasteley	Tue 13 Feb 18 (Tue)	19:30	Online
Cluster	TMA 02 preparation	Jo Buxton	Mon 19 Feb 18 (Mon)	19:00	Online
Cluster	TMA 02 preparation	Marion Edwards	Wed 21 Feb 18 (Wed)	19:30	Online
Cluster	TMA 02 preparation	Roger Shaw	Thu 22 Feb 18 (Thu)	19:30	Online
Cluster	Taking stock	Steven Self	Sat 24 Feb 18 (Sat)	11:00	Southampton
Cluster	TMA 02 preparation	Avon Huxor	Sun 25 Feb 18 (Sun)	16:00	Online
Cluster	TMA 02 preparation	Steven Self	Wed 28 Feb 18 (Wed)	19:00	Online
Cluster	EMA support	Sharon Dawes	Sat 07 Apr 18 (Sat)	15:00	Online
Cluster	EMA support	Keith Hasteley	Tue 08 May 18 (Tue)	19:00	Online
Cluster	EMA support	Avon Huxor	Sun 13 May 18 (Sun)	16:00	Online
Cluster	EMA support	Jo Buxton	Mon 14 May 18 (Mon)	19:00	Online
Cluster	EMA support	Marion Edwards	Wed 16 May 18 (Wed)	19:30	Online
Cluster	EMA support	Roger Shaw	Sat 19 May 18 (Sat)	11:00	Online
Cluster	EMA support	Steven Self	Wed 23 May 18 (Wed)	19:00	Online
100000				11.50	

# PRACTICAL DEMONSTRATION

SCREEN SHARE — PYTHON NOTEBOOKS AND OPENREFINE





- EMAIL ME J.BUXTON@OPEN.AC.UK
- TUTOR GROUP FORUM
- TEXT ON: 07900 783316
- MONDAY BETWEEN 19:00 AND 21:00