DATA SCIENCE
11 WEEK PART TIME COURSE

Week 7 – Hands On SQL
Monday 1st February 2016

1. Guest Speaker - Thom Mackey
2. SQL
3. Lab
4. Review

SQL

What is SQL?

- Structured Query Language (SQL)
- Data manipulation language from ~1970s
- Prior to this there was very little standardisation of database programming languages
- S.Q.L. or Sequel?

Why use SQL?

‣ Handles large data better than Excel

‣ Traceable

‣ Extendable

‣ Expressive

‣ It's a necessity for handling data
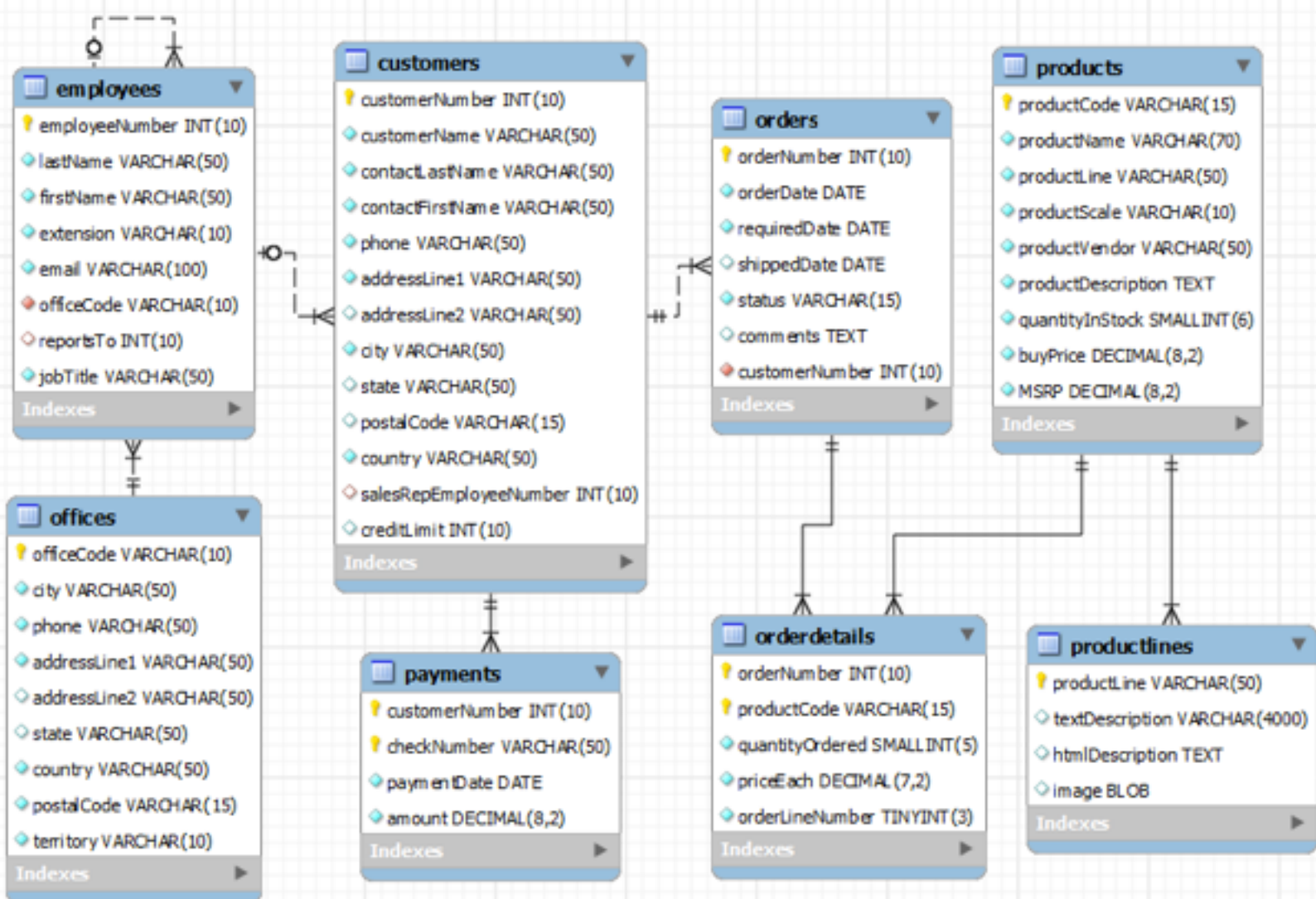
Why use SQL? (Non-technical reasons)

‣ Most organisations have data in some form of SQL data base somewhere (even if it's underlying another program like finance system)

‣ Getting at this data is a powerful skill to have, means less time asking someone to extract it for you and more time analysing it and making decisions

Looking at a SQL Database for a retail store

‣ Employees

‣ Offices

‣ Customers

‣ Products

‣ Payments

‣ Orders

E14 f_x lbondur@classicmodelcars.com

	A	B	C	D	E	F	G	H
1	employeeNumber	lastName	firstName	extension	email	officeCode	reportsTo	jobTitle
2	1002	Murphy	Diane	x5800	dmurphy@classicmodelcars.com	1	NULL	President
3	1056	Patterson	Mary	x4611	mpatterso@classicmodelcars.com	1	1002	VP Sales
4	1076	Firrelli	Jeff	x9273	jfirrelli@classicmodelcars.com	1	1002	VP Marketing
5	1088	Patterson	William	x4871	wpatterson@classicmodelcars.com	6	1056	Sales Manager (APAC)
6	1102	Bondur	Gerard	x5408	gbondur@classicmodelcars.com	4	1056	Sale Manager (EMEA)
7	1143	Bow	Anthony	x5428	abow@classicmodelcars.com	1	1056	Sales Manager (NA)
8	1165	Jennings	Leslie	x3291	ljennings@classicmodelcars.com	1	1143	Sales Rep
9	1166	Thompson	Leslie	x4065	lthompson@classicmodelcars.com	1	1143	Sales Rep
10	1188	Firrelli	Julie	x2173	jfirrelli@classicmodelcars.com	2	1143	Sales Rep
11	1216	Patterson	Steve	x4334	spatterson@classicmodelcars.com	2	1143	Sales Rep
12	1286	Tseng	Foon Yue	x2248	ftseng@classicmodelcars.com	3	1143	Sales Rep
13	1323	Vanauf	George	x4102	gvanauf@classicmodelcars.com	3	1143	Sales Rep
14	1337	Bondur	Loui	x6493	lbondur@classicmodelcars.com	4	1102	Sales Rep
15	1370	Hernandez	Gerard	x2028	ghernande@classicmodelcars.com	4	1102	Sales Rep
16	1401	Castillo	Pamela	x2759	pcastillo@classicmodelcars.com	4	1102	Sales Rep
17	1501	Bott	Larry	x2311	lbott@classicmodelcars.com	7	1102	Sales Rep
18	1504	Jones	Barry	x102	bjones@classicmodelcars.com	7	1102	Sales Rep
19	1611	Fixter	Andy	x101	afixter@classicmodelcars.com	6	1088	Sales Rep
20	1612	Marsh	Peter	x102	pmarsh@classicmodelcars.com	6	1088	Sales Rep
21	1619	King	Tom	x103	tking@classicmodelcars.com	6	1088	Sales Rep
22	1621	Nishi	Mami	x101	mnishi@classicmodelcars.com	5	1056	Sales Rep
23	1625	Kato	Yoshimi	x102	ykato@classicmodelcars.com	5	1621	Sales Rep
24	1702	Gerard	Martin	x2312	mgerard@classicmodelcars.com	4	1102	Sales Rep

What is the SELECT statement?

‣ SELECT is how we specify which columns to include in the results, like when we click on a column in Excel

‣ We can either name the columns we want, or if we want all the columns we can just use '*' (the asterix)

‣ If we are selecting columns we need to separate the column names by a comma

What is the FROM clause?

‣ FROM is the table we get the data from

‣ We can think of this like a tab in an excel spreadsheet

What is the WHERE clause?

‣ WHERE is used to filter the data you have to meet some criteria, e.g. select all the employees that are Sales Representatives

‣ We can think of it like filtering in Excel

Other WHERE clauses?

‣ BETWEEN selects values within a range of values

‣ LIKE matches value based on pattern matching

‣ IN specifies if the value matches any value in a list

‣ IS NULL checks if the value is NULL (like a missing or empty cell)

What is the ORDER BY clause?

‣ ORDER BY will make sure the results are returned in an order you specify by selecting a column

‣ We can think of this like the sort button in Excel

‣ By default it sorts in ascending order, i.e. A, then B, then C, and similarly 1, 2, 3, 4

‣ If we want descending order we can specify it at the end of the column we specify with DESC

What is the GROUP BY clause?

‣ GROUP BY is a way to group rows into subgroups, so we can summarise the data

‣ This is similar to the Pivot table functionality in Excel

‣ We can use functions such as averages and sums to summarise the data in a table

Aggregate Functions

‣ SUM() - add all the values in the column for the group

‣ AVG() - average the values in the column for the group

‣ MIN() - take the minimum value in the column for the group

‣ MAX() - take the maximum value in the column for the group

‣ COUNT(*) - count the number of rows in a given group

What is a JOIN?

‣ JOIN is a way to combine 2 or more different tables with matching values in specified column

‣ You can think of it like a vlookup in Excel, however there are many important distinctions
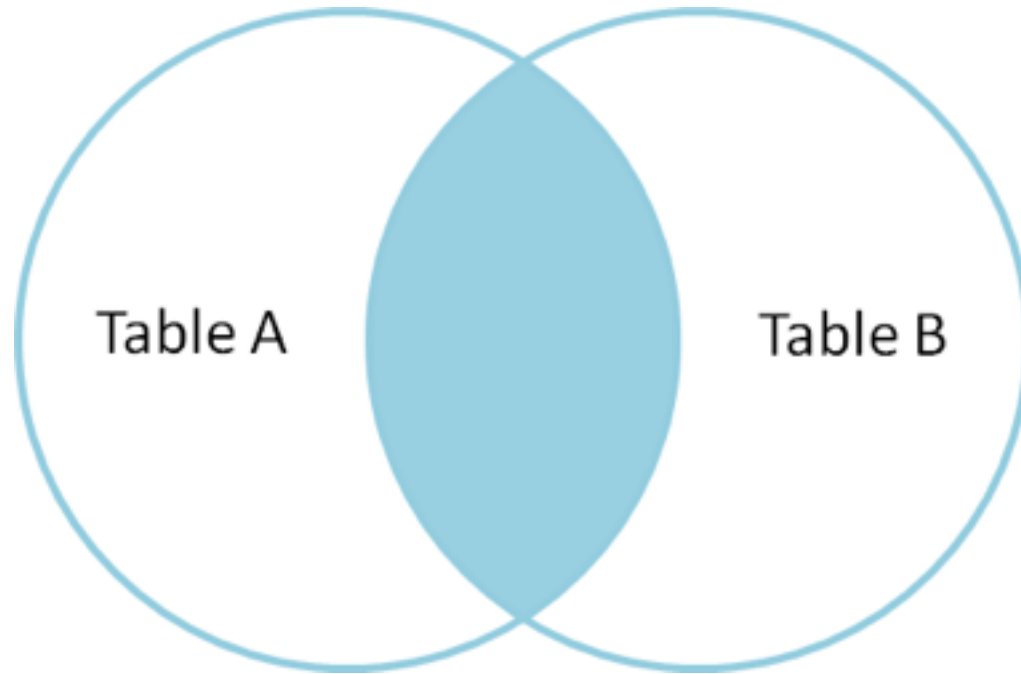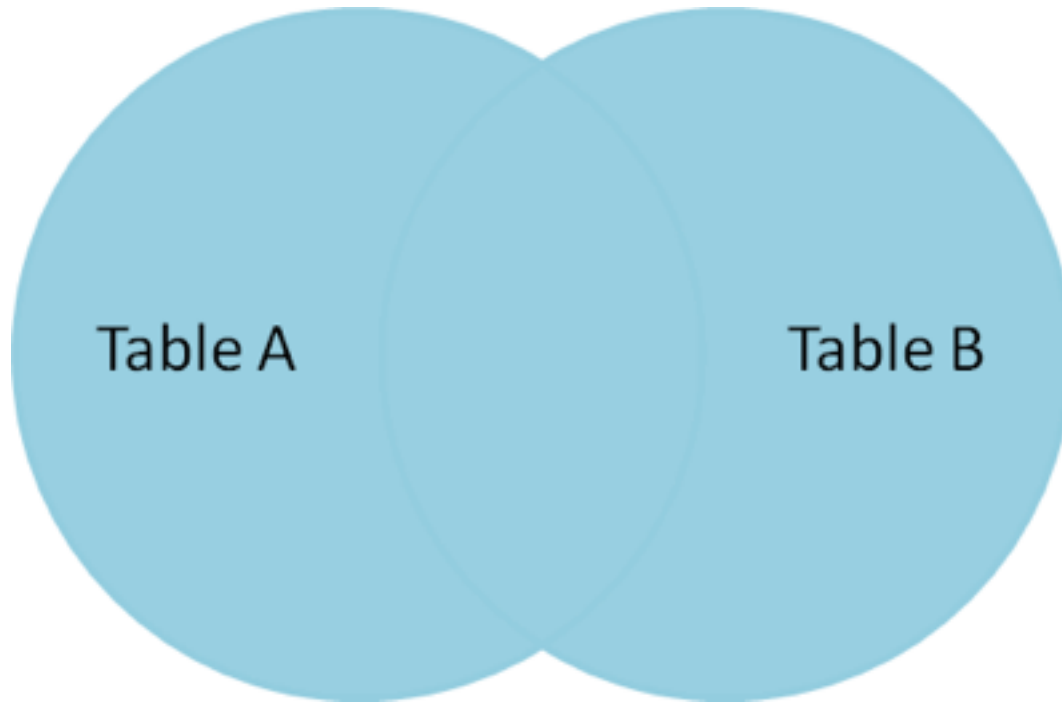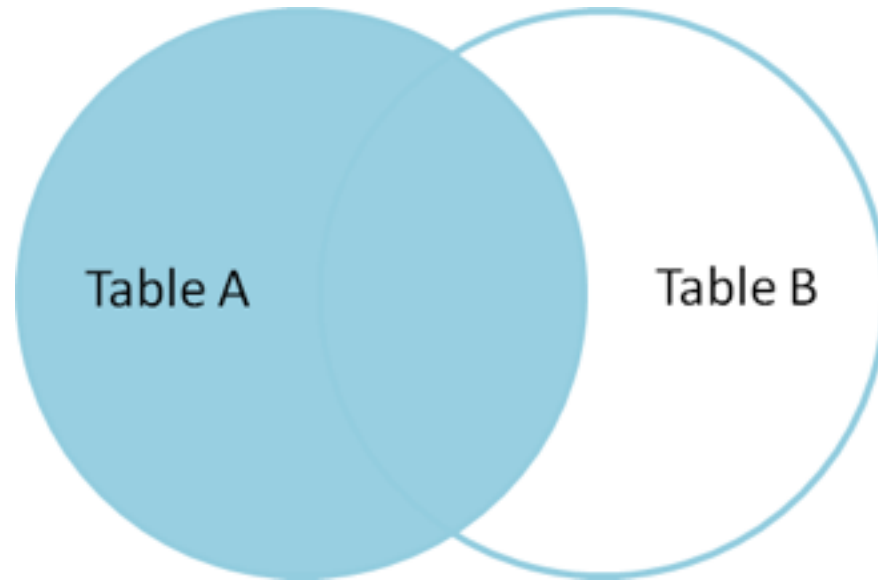
DATA SCIENCE PART TIME COURSE

LAB

http://ec2-54-183-194-31.us-west-1.compute.amazonaws.com:3000/

‣ Username: example@ga.com
‣ Password: 'ok123'

DISCUSSION TIME

- ‣ Review of last week
- ‣ Extra Learning + Tasks

HACKDAY

- ‣ **Saturday 30th January 12:30 – 3:30**
- ‣ **Review topics completed done so far**
- ‣ **AWS**
- ‣ **Projects**

Task List

- ☐ Email me anything you'd like to review (previously been covered) by Thursday for Saturday
- ☐ Email me anything you'd like to cover (not covered) by Thursday for Saturday
- ☐ Work on your project
- ☐ Look up information on the company giving the guest talk on Monday and come up with at least one question you'd like to ask someone from that company

Task List

- [ ] Download and install Putty and Cygwin
- [ ] Download and install Postgres SQL and PGadmin on your local machines
- [ ] Lookup and write down the describe why you would use; ec2, rds, mongoDB, Redshift, Spark