

6CS030 Worksheet One – 10%

Hand-out: Week 3. Due: Week 4

Student Number:	1828421
Student Name:	Prakash

My Remainder Value:	2
----------------------------	---

A. Cleaning data:

As per my remainder value, 2, my dataset is No Qualification.

Before Cleaning data, it is good to make copy of it.

289	Redditch	17.6	19.4	12.9	2
290	Rugby	11.2	8.4	6.8	2
291	Sandwell	23	25.4	24.3	2
292	Shropshire	16	14.9	12.7	2
293	Solihull	13.1	10.1	11.5	2
No Qualification Employment Rate Degree-Level Qualifications no_qualification					

Default data representation before cleaning are:

Upper Header Section:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1															
2	annual population survey														
3	ONS Crown Copyright Reserved [from Nomis on 29 November 2018]														
4	Analysis :	variable													
5	Confidence :	95% confidence interval of percent figure (+/-)													
6	Variable :	% with no qualifications (NVQ) - aged 16-64													
7															
8	local authority:	Jan 2004-Dec 2004	Jan 2005-	Jan 2006-	Jan 2007-	Jan 2008-	Jan 2009-	Jan 2010-	Jan 2011-	Jan 2012-	Jan 2013-	Jan 2014-	Jan 2015-	Jan 2016-	Jan 2017-Dec 2017
9															
10	Babergh	18.5	12.6	14	13.6	11.8	9.8	12	4.6	5.1	9.1	7	9.2	4.3	4.9
11	Basildon	18.4	14.6	16.1	15	20	17	12.3	13	9.8	8.9	5.5	7.6	11.4	10.1
12	Bedford	13.1	15.1	15.5	11.8	11.7	11.6	8	8.7	6.4	5.3	7.8	7.9	5.4	8.3
13	Braintree	16.9	16.5	19.4	12.9	15.6	16.4	12.4	7.2	8.2	9.6	8.7	7.4	7.5	7
14	Breckland	17.1	18.7	17.5	16.1	16.3	13.5	13.5	13.9	9.8	5.6	11	10.8	6.1	9.3
15	Brentwood	13.2	13.2	12.5	13	14.7	8.6	10.8	5.9	3.9	9.7	5.9	7.2	3.6	!
16	Broadland	12.2	11.7	11	13.3	9.4	10.1	10.7	10.5	5.9	8.1	6.1	3.6	7.7	4.3
17	Broxbourne	10	13.7	12.2	6.6	10	8.2	15.1	9.5	6.4	15.7	6.7	5.6	6.3	6.6
18	Cambridge	8.9	7.5	5.2	9.1	5.5	5.2	3.1	7.1	5.5	5.3	9.1	4.5	4.3	4.3
19	Castle Point	14.6	16.8	16.7	16.8	13.6	18.3	15.2	12.1	8.7	10.1	14	8.3	7.4	11.8
20	Central Bedford	11.1	9.1	14.3	8.8	8.3	8.2	8.7	7.3	9.6	10.2	6.1	6.8	7.2	4.9
21	Chelmsford	11.9	12.9	12.9	9.1	13.2	10.5	7.6	6.3	5.6	6.5	6.9	7.1	8.4	4.7
22	Colchester	17.9	20.9	17.7	9.4	12.6	9.3	7.5	4.8	6.6	9.4	8.6	7.9	6.9	4.8
23	Dacorum	11.7	5.6	5.7	10.8	11	6.8	7.5	10.6	10.4	8.3	6.5	7.9	3.4	2
24	East Cambridge	11.5	8.7	7.5	12.3	8.8	5.2	7.4	5.5	7.8	4	8.9	10.1	6.9	5.1
25	East Hertfordsh	10.8	7.8	12.2	10	5.7	6.4	5.3	4.8	4	4.7	4.5	4.4	5.3	3.4
26	Epping Forest	11.2	13.9	13.8	15.2	17	17.3	11.1	8.5	9	3.9	6.2	7.2	8.2	4.7
27	Fenland	11.8	10.7	14.4	10.3	13.8	16.3	15.3	13.3	14	18.8	12.4	10.1	7.4	9.3
28	Forest Heath	28.4	20	15.2	21.4	15.3	12.6	12.9	10.9	8.9	12.3	8.2	18.1	10.3	10.4
29	Great Yarmouth	15.5	18.2	16.8	11.3	24.6	22.4	19.1	14.2	14	15.1	11.4	7.6	13.8	21.7
Degree-Level Qualifications No Qualifications															

Lower Footer Section:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
324	Kirklees	16.5	17.1	14.3	12.6	15.3	13.8	13.3	15.8	12.6	12.8	11	10.5	11.2	13.1
325	Leeds	14.1	13.1	12.3	13.3	11.8	11.6	10.7	9.1	10.3	10.3	9.6	8.8	9.5	11.2
326	North East Lincolnshire	14.4	12.1	11.6	11.2	11.8	12.8	11.6	10.7	11.2	10.3	8.3	11.1	10.5	9.7
327	North Lincolnshire	12.3	10.3	14.1	10.6	12.8	10.3	10.2	9.4	8.8	8.6	8.4	8.1	6.8	7.8
328	Richmondshire	8.4	15.2	9.5	8.6	11.5	6.9	4.8	4.4	5.5	4.9	3.8	7.6	13.3	8.2
329	Rotherham	17.2	17.2	16.5	18.3	17.8	15.6	13.2	12.3	10.8	12.4	9.9	10	11.7	8.7
330	Ryedale	9.3	11.5	10.7	18.6	9.1	5.5	16	13.6	7.9	3.3	7	12.4	5.4	!
331	Scarborough	16.5	17.7	19.2	17.9	12.9	11.8	10.5	14	20.6	12.7	13.6	7.8	7.4	10.3
332	Selby	9.4	16.3	19.5	11.5	8.3	14.2	10.1	8.1	6.5	8.2	7.9	8.2	8.9	7.6
333	Sheffield	14.3	15.3	13.4	14.8	16.2	11	13.5	10	10.7	10.7	10.7	10	8.4	8.5
334	Wakefield	18	16.6	15.8	16.2	15.2	12.9	14.6	14	12.7	10.7	10.8	11.3	10.7	10.4
335	York	11.8	11.6	12.8	9.6	8.1	8.1	7.2	7	6.5	6.9	4.9	4.6	6.1	4.4
336	Column Total	14.9	14.2	13.7	13.1	13.3	12.1	11.1	10.5	9.5	9.2	8.6	8.4	7.8	7.6
337															
338															
339	#														
340															
341	!														
342															
343															
344	*														
345															
346															
347	~														
348															
349	-														
350															
351															
352															

It is necessary to clean the data. Several cleaning activities will be carried out.

While importing data in oracle database, the data should be in tabular format. Renaming the table name into no_qualification. The given data has some header section which can create problem. So, removing header section of the data up to 7th row.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2												
3												
4												
5												
6												
7												
8	local authority:	Jan 2004-Dec 2004	Jan 2005-[Jan 2006-[Jan 2007-[Jan 2008-[Jan 2009-[Jan 2010-[Jan 2011-[Jan 2012-[Jan 2013-[Jan 2014-[
9												
10	Babergh	18.5	12.6	14	13.6	11.8	9.8	12	4.6	5.1	9.1	7
11	Basildon	18.4	14.6	16.1	15	20	17	12.3	13	9.8	8.9	5.5
12	Bedford	13.1	15.1	15.5	11.8	11.7	11.6	8	8.7	6.4	5.3	7.8
13	Braintree	16.9	16.5	19.4	12.9	15.6	16.4	12.4	7.2	8.2	9.6	8.7
14	Breckland	17.1	18.7	17.5	16.1	16.3	13.5	13.5	13.9	9.8	5.6	11
15	Brentwood	13.2	13.2	12.5	13	14.7	8.6	10.8	5.9	3.9	9.7	5.9
16	Broadland	12.2	11.7	11	13.3	9.4	10.1	10.7	10.5	5.9	8.1	6.1
17	Broxbourne	10	13.7	12.2	6.6	10	8.2	15.1	9.5	6.4	15.7	6.7
18	Cambridge	8.9	7.5	5.2	9.1	5.5	5.2	3.1	7.1	5.5	5.3	9.1
19	Castle Point	14.6	16.8	16.7	16.8	13.6	18.3	15.2	12.1	8.7	10.1	14
20	Central Bedford	11.1	9.1	14.3	8.8	8.3	8.2	8.7	7.3	9.6	10.2	6.1
21	Chelmsford	11.9	12.9	12.9	9.1	13.2	10.5	7.6	6.3	5.6	6.5	6.9
22	Colchester	17.9	20.9	17.7	9.4	12.6	9.3	7.5	4.8	6.6	9.4	8.6
23	Dacorum	11.7	5.6	5.7	10.8	11	6.8	7.5	10.6	10.4	8.3	6.5
24	East Cambridge	11.5	8.7	7.5	12.3	8.8	5.2	7.4	5.5	7.8	4	8.9
25	East Hertfordsh	10.8	7.8	12.2	10	5.7	6.4	5.3	4.8	4	4.7	4.5
26	Epping Forest	11.2	13.9	13.8	15.2	17	17.3	11.1	8.5	9	3.9	6.2
27	Fenland	11.8	10.7	14.4	10.3	13.8	16.3	15.3	13.3	14	18.8	12.4
28	Forest Heath	28.4	20	15.2	21.4	15.3	12.6	12.9	10.9	8.9	12.3	8.2
29	Great Yarmouth	15.5	18.2	16.8	11.3	24.6	22.4	19.1	14.2	14	15.1	11.4
	...	Degree-Level Qualifications	no_qualification									

Similarly, erasing data from lower section from 336th row to 349th row.

331	Scarborough	16.5	17.7	19.2	17.9	12.9
332	Selby	9.4	16.3	19.5	11.5	8.3
333	Sheffield	14.3	15.3	13.4	14.8	16.2
334	Wakefield	18	16.6	15.8	16.2	15.2
335	York	11.8	11.6	12.8	9.6	8.1
336						
337						
338						
339						
340						
341						
342						
343						
344						
345						
346						
347						
348						
349						
350						

Upper 7 rows are empty so to maintain that, moving the table to A1.

	A	B	C	D	E	F	G	H
1	local authority:	Jan 2004-Dec 2004	Jan 2005-[Jan 2006-[Jan 2007-[Jan 2008-[Jan 2009-[Jan 2010-[
2								
3	Babergh	18.5	12.6	14	13.6	11.8	9.8	12
4	Basildon	18.4	14.6	16.1	15	20	17	12.3
5	Bedford	13.1	15.1	15.5	11.8	11.7	11.6	8
6	Braintree	16.9	16.5	19.4	12.9	15.6	16.4	12.4
7	Breckland	17.1	18.7	17.5	16.1	16.3	13.5	13.5
8	Brentwood	13.2	13.2	12.5	13	14.7	8.6	10.8
9	Broadland	12.2	11.7	11	13.3	9.4	10.1	10.7
10	Broxbourne	10	13.7	12.2	6.6	10	8.2	15.1
11	Cambridge	8.9	7.5	5.2	9.1	5.5	5.2	3.1
12	Castle Point	14.6	16.8	16.7	16.8	13.6	18.3	15.2
13	Central Bedford	11.1	9.1	14.3	8.8	8.3	8.2	8.7
14	Chelmsford	11.9	12.9	12.9	9.1	13.2	10.5	7.6
15	Colchester	17.9	20.9	17.7	9.4	12.6	9.3	7.5
16	Dacorum	11.7	5.6	5.7	10.8	11	6.8	7.5
17	East Cambridge	11.5	8.7	7.5	12.3	8.8	5.2	7.4
18	East Hertfordsh	10.8	7.8	12.2	10	5.7	6.4	5.3
19	Enning Forest	11.2	12.0	12.8	15.2	17	17.2	11.1

Now, the table is in structured format. But still there are some missing values which can create problem in database.

	A	B	C	D	E	F
1	local authority: district / unitary (prior to April 2015)	Jan 2004-Dec 2004	Jan 2005-[Jan 2006-[Jan 2007-[Jan 2008-[
2						
3	Babergh	18.5	12.6	14	13.6	11.8
4	Basildon	18.4	14.6	16.1	15	20
5	Bedford	13.1	15.1	15.5	11.8	11.7
6	Braintree	16.9	16.5	19.4	12.9	15.6
7	Breckland	17.1	18.7	17.5	16.1	16.3
8	Brentwood	13.2	13.2	12.5	13	14.7
9	Broadland	12.2	11.7	11	13.3	9.4
10	Broxbourne	10	13.7	12.2	6.6	10
11	Cambridge	8.9	7.5	5.2	9.1	5.5
12	Castle Point	14.6	15.8	15.7	15.8	12.6

93	Brent	18.4	13.3	12.5	8.9	8.4	9.7	5.1	1
94	Bromley	9.8	9.4	9.3	7.7	9.5	5.5	6.8	
95	Camden	8.7	7.6	14.2	10.4	7.7	8.9	8	
96	City of London	!	!	!	!	!	!	!	
Degree-Level Qualifications no_qualification (+)									
Ready									

250	Exeter	8.1	8.4	8.1	8	11	8	5.8
251	Forest of Dean	11.1	11.8	11.9	6.9	9.7	11.6	7.7
252	Gloucester	9.6	11.7	12.1	11.9	9.9	9.4	7.2
253	Isles of Scilly	-	-	-	-	-	-	-
254	Mendip	11.3	12.4	9.5	6.1	9.1	12.4	6.5
255	Mid Devon	11.2	10.3	6	9.1	10.1	11.3	11.4
256	North Devon	15.5	14.8	12.1	8.5	7.3	6.2	6.1
257	North Dorset	11.9	8.8	11.8	8.6	9.9	7	8.1
258	North Somerset	9.8	7.7	7.6	8	6.2	7.3	6.5
259	Plymouth	13.4	11	10.2	10	12	8.5	10.1
Degree-Level Qualifications no_qualification (+)								
Ready								

The column names, empty rows or #, ! -, values can be problematic.

So, removing empty row, renaming the columns name and representing in graph format for analysis.

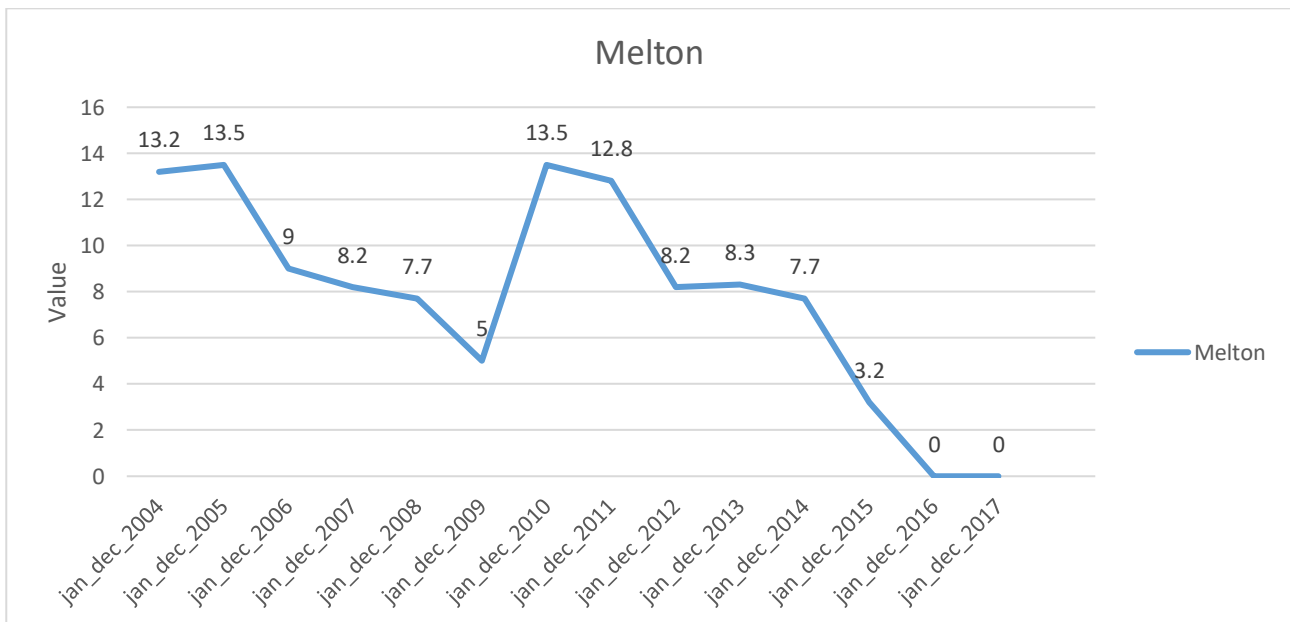
	A	B	C	D	E	F	G	H
1	local_authority	jan_dec_2004	jan_dec_2005	jan_dec_2006	jan_dec_2	jan_dec_2	jan_dec_2	jan_dec_2
2	Babergh	18.5	12.6	14	13.6	11.8	9.8	12
3	Basildon	18.4	14.6	16.1	15	20	17	12.3
4	Bedford	13.1	15.1	15.5	11.8	11.7	11.6	8
5	Braintree	16.9	16.5	19.4	12.9	15.6	16.4	12.4
6	Breckland	17.1	18.7	17.5	16.1	16.3	13.5	13.5
7	Brentwood	13.2	13.2	12.5	13	14.7	8.6	10.8
8	Broadland	12.2	11.7	11	13.3	9.4	10.1	10.7
9	Broxbourne	10	13.7	12.2	6.6	10	8.2	15.1

Removing row of data of Isles of Scilly, South Bucks and city of London which has no values at all.

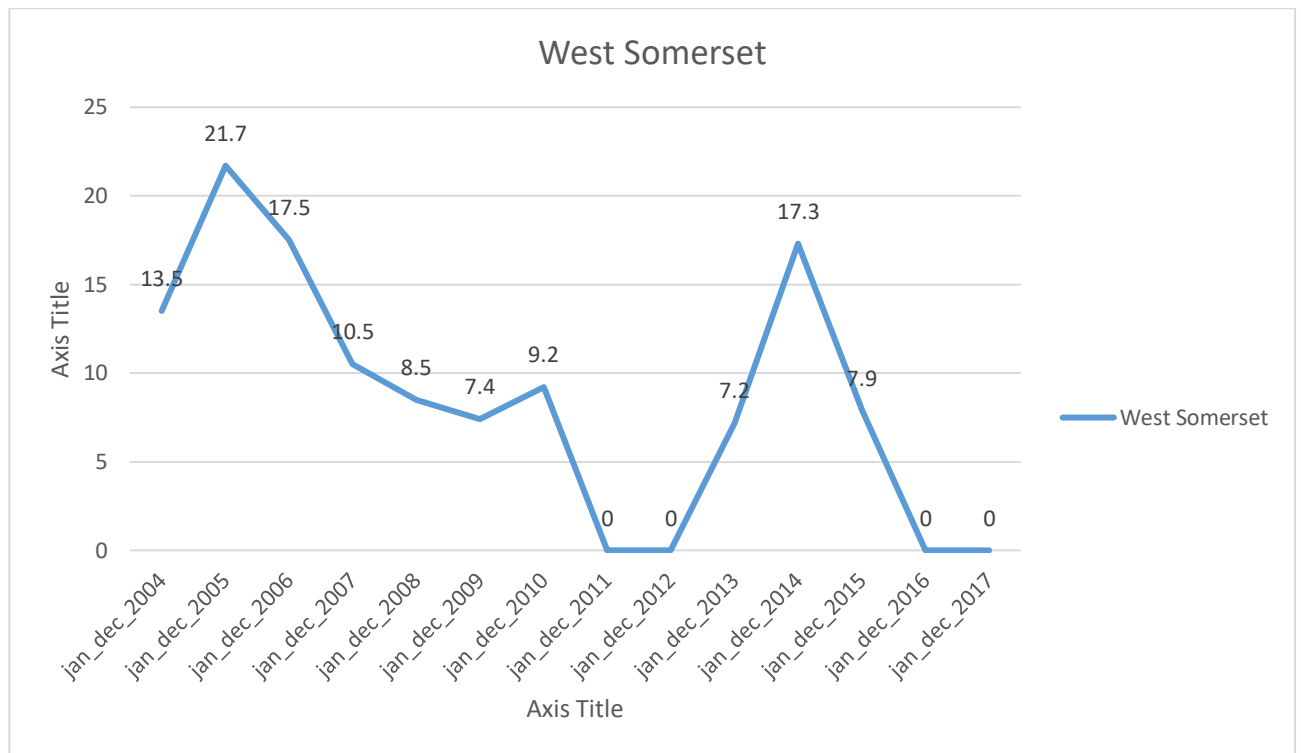
251	Gloucester	9.6	11.7	12.1	11.9	9
252	Isles of Scilly	-	-	-	-	-
253	Mendip	11.3	12.4	9.5	6.1	9
		No Qualification	Employment Rate	Degree-Level Qualifications	no_qualification	+
Ready						

93	Bromley	9.8	9.4	9.3	7.7	9.5	5.5	6
94	Camden	8.7	7.6	14.2	10.4	7.7	8.9	
95	City of London	!	!	!	!	!	!	!
96	Croydon	10.6	11.2	13.7	16.1	12.9	11.8	6
97	Ealing	15.3	9.6	11.6	12.3	12.1	13.7	12
		No Qualification	Employment Rate	Degree-Level Qualifications	no_qualification	+	:	
Ready								

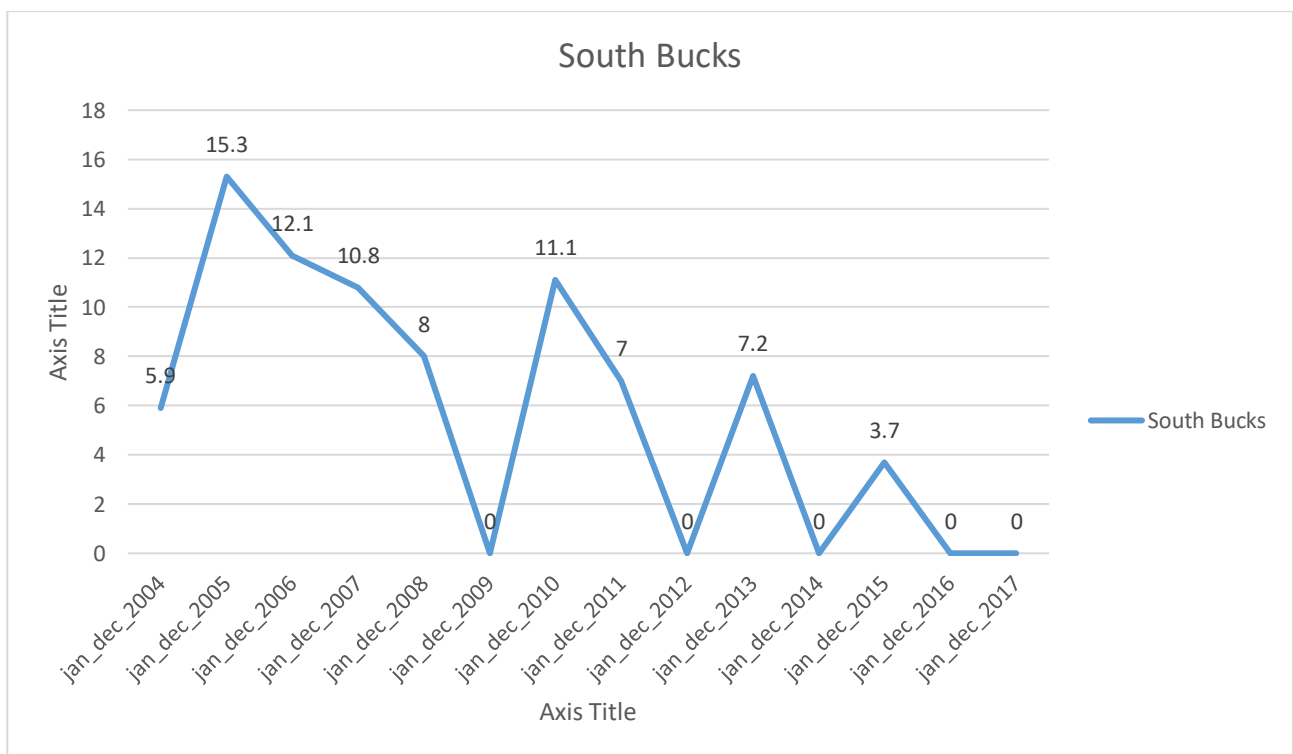
Some values are missing in some cities.



Last two years values of Melton are missing



Year 2011, 2012, 2016 and 2017 values are missing in city West Somerset.



Here are some values of South Bucks are missing which is represented as 0.

There are other cities whose values are missing. So, these things create problem, therefore it must be maintained. There are two different technique to add missing value.

- a. Average: Taking average of the whole column and assigning the value in missing cell.

Average is useful if more than one values are missing in the column.

- b. Median: Taking median of preceding and succeeding shell of the missing value and assigning the median value in missing cell.

Median is useful if only few values are missing in the column.

In some cities, values are null. So, if many values are null average is suitable but if only one value is null then median is suitable.

In South Bucks, many values are missing so taking average and assigning value in it.

212	Rother	12.4	11.1	9.7	12.2	13.3	12.3	7.9	11	7	8.9	5.3	9	5.1	7.3
213	Runnymede	12.6	13.6	11.6	8.4	8.4	8.2	14.5	10.1	9.6	7.7	9.7	7.5	2.7	4.3
214	Rushmoor	9.3	10.2	10.4	15.1	12.4	5.8	8.4	12.3	9.4	5.7	5.8	7.5	7.5	7.4
215	Sevenoaks	7.6	13.6	14.8	15.9	14	11.2	13.6	12	6.1	6.1	6.6	7.7	8.2	9.3
216	Folkestone and Hythe	14.9	14.4	7.8	9.9	9.3	7	7.9	9.7	7.3	9.5	6.3	7.5	7.3	8.6
217	Slough	13.8	13.3	14.2	13.4	11.8	11.2	12.1	13.3	11.5	9.7	10	8	6.5	6.2
218	South Bucks	5.9	15.3	12.1	10.8	8	11.1	7	7.2	3.7	!	!	!	!	=AVERAGE(B218:O218)
219	South Oxfordshire	9.7	6	9	10.6	8.9	4.2	6.9	4.5	5.3	4.1	4.8	3.3	4.2	3.4
220	Southampton	12.9	12.3	10.9	12.5	11.1	9.9	8.8	11.3	8.6	8.6	6.5	6.4	7.2	AVERAGE(number1, [number2], ...)
221	Spelthorne	11.2	5.8	11.7	11.4	10.7	6.8	10	9.4	7.1	7	5.4	7.3	4.8	6.4
222	Surrey Heath	6.1	9.9	4.6	5.7	8.8	9.5	4.3	10.7	8.6	2.6	2.1	5.8	3.6	3.3
223	Swale	14.7	13.9	14.5	19.6	14.3	14.4	18	12	11.2	8.3	8.5	10.2	8.9	8.3
224	Tandridge	7.3	6.9	8	8.6	6.9	7.2	6.1	4.8	4.4	5.2	5.2	10	!	5.4
225	Test Valley	11.7	8.7	8	8.5	6.9	8.4	5	4.4	4.2	3.2	3.7	2.3	4	3
226	Thanet	15.6	17.2	16.4	16	16.4	13.6	14.2	7.9	8.3	8.1	9.8	8	5.5	7.6

Obtained average value is assigned in ! values of South Bucks.

If few data are null then Median is taken and assigned.

97	Hastings	14.5	12.8	11.7	14.6	10.8	7.4	8.1	13.6	1
98	Havant	15.7	13.6	17.7	10.7	10.7	11.2	13.5	10.3	1
99	Horsham	4.4	5.3	5.9	5.1	5.8	4.6	5.7	5.1	
00	Isle of Wight	14.3	11.8	11	11.4	11	10.7	11.9	10.6	
01	Lewes	14.4	10.9	8.7	12.3	12.6	5.1	7.1	11.5	
02	Maidstone	10.1	12	9.1	13.5	12.4	11.8	12.1	5.5	
03	Medway	13	12.3	12.2	11.8	10.8	13.4	13.6	13.1	
04	Mid Sussex	8.5	7.6	4.4	3.6	7.4	4.2	3.1	3.8	
05	Milton Keynes	12.7	11.1	11.9	13.2	10.4	9.2	8.9	9	
06	Mole Valley	7.2	9	7.2	4.9	4.6	4.6	!	=MEDIAN(G206,I206)	
07	New Forest	9.5	12.4	7.5	10.4	11	10.7	!	MEDIAN(number1, [number2], [number3], ...)	
08	Oxford	12.2	11.2	7.1	8.2	5.8	8.5	9.5	7.7	
09	Portsmouth	12.5	14.1	13.6	13.9	12.5	12.7	11.4	11.7	
10	Reading	10.1	10.9	9.4	9.7	11.3	10	9	9	
11	Reigate and Banstead	8.7	9.1	12.1	7.9	8.2	9.4	6.5	4.1	
12	Rother	12.4	11.1	9.7	12.2	13.3	12.3	7.9	11	
13	Runnymede	12.6	13.6	11.6	8.4	8.4	8.2	14.5	10.1	
14	Rushmoor	9.3	10.2	10.4	15.1	12.4	5.8	8.4	12.3	

Similarly adding vales in another empty cell.

B. Formatting and Importing

At first renaming Data Name “No Qualification” to “no_qualification”.

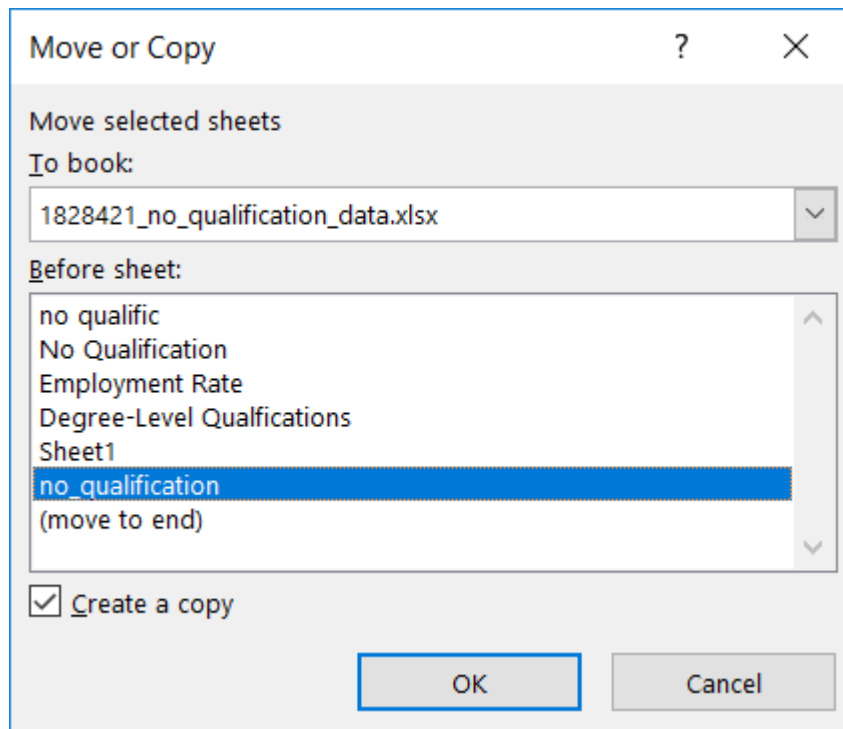
24	East Cambridge	11.5	8.7	7.5	12.3
25	East Hertfordsh	10.8	7.8	12.2	10
26	Epping Forest	11.2	13.9	13.8	15.2
27	Fenland	11.8	10.7	14.4	10.3
28	Forest Heath	28.4	20	15.2	21.4
29	Great Yarmouth	15.5	18.2	16.8	11.3

Columns name should be formatted well before importing it in the data base.

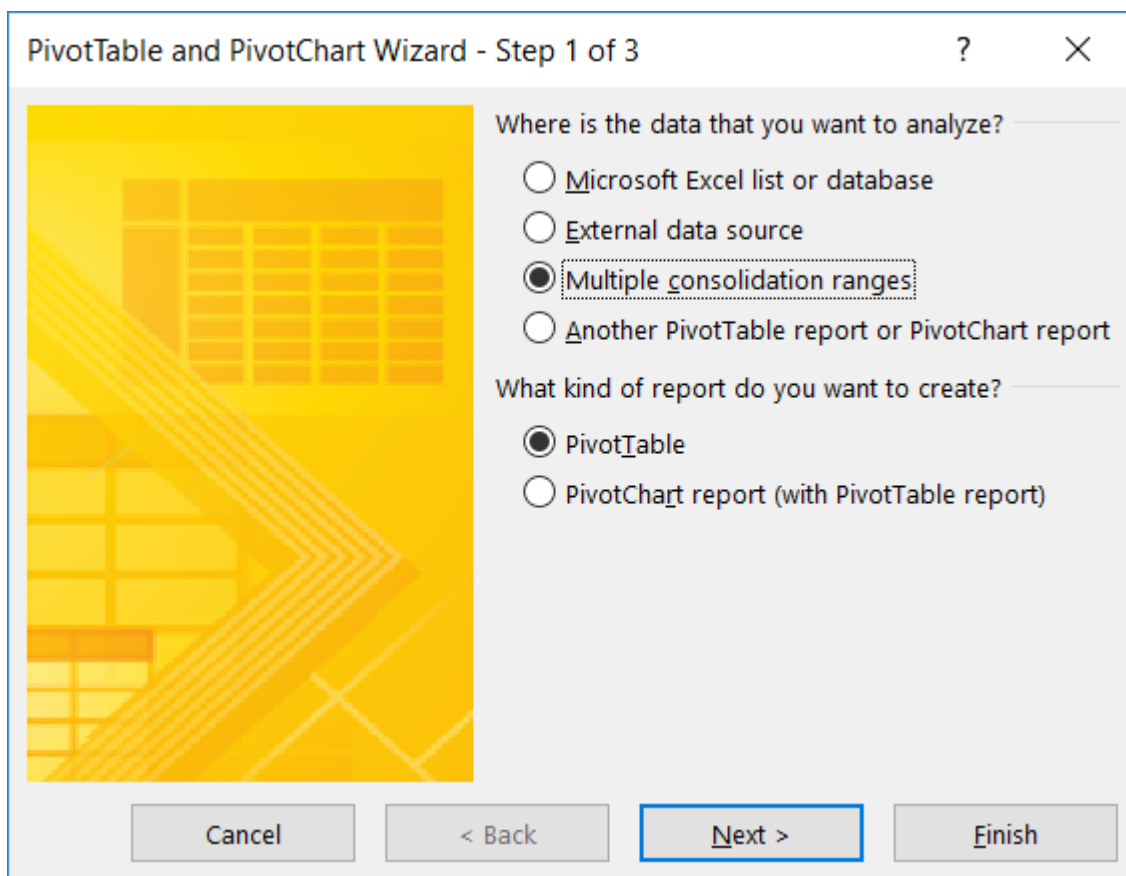
	A	B	C	D	E
1	local_authority	jan_dec_2004	jan_dec_2005	jan_dec_2006	jan_dec_2jan_2007
2	Babergh	18.5	12.6	14	13.6
3	Basildon	18.4	14.6	16.1	15
4	Bedford	13.1	15.1	15.5	11.8

Now the data format is appropriate to import in oracle. But still the size of column is large and confusing which can be maintained by converting the table into pivot table.

For this at first creating a copy of the page



Now renaming the file name as pivot_data. Pressing Ctrl + D and then P which gives the dialog box.



Selecting multiple consolidation ranges and Pivot table then hitting next.

PivotTable and PivotChart Wizard - Step 2a of 3

You can create a PivotTable report that uses ranges from one or more worksheets, and that has no page fields or up to four page fields.

How many page fields do you want?

☐ Create a single page field for me
☒ I will create the page fields

Cancel < Back Next > Finish

Select I will create the page field.

PivotTable and PivotChart Wizard - Step 2b of 3

Where are the worksheet ranges that you want to consolidate?

Range:

pivot_data!\$A\$1:\$O\$325

Add Delete Browse...

All ranges:

How many page fields do you want?

☒ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4

What item labels do you want each page field to use to identify the selected data range?

Field one: Field two:

Field three: Field four:

Cancel < Back Next > Finish

Selecting All ranges and hitting next.

PivotTable and PivotChart Wizard - Step 3 of 3

Where do you want to put the PivotTable report?

☒ New worksheet

☐ Existing worksheet

Click Finish to create your PivotTable report.

Layout... Options... Cancel < Back Next > Finish

Leaving the default New worksheet and finish.

New sheet is displayed

	A	B	C	D	E	F	G	H	I
1									
2									
3	Sum of Value	Column Labels							
4	Row Labels	jan_dec_2004	jan_dec_2005	jan_dec_2006	jan_dec_2007	jan_dec_2008	jan_dec_2009	jan_dec_2010	jan_dec_2011
5	Adur	15.7	13.5	13.1	8.7	9.9	4.8	8.4	
6	Allerdale	14.1	9.3	5.4	13.1	12.6	10.6	9.7	
7	Amber Valley	23.6	22.1	11.4	12.5	15	12.4	14.8	
8	Arun	13.7	10.8	9.7	8.6	14.4	14.6	13.6	
9	Ashfield	20.3	16.5	15.4	13.6	18	10.8	15.4	
10	Ashford	11.6	14.6	12.3	7.6	10.1	12.4	9.8	
11	Aylesbury Vale	11.3	8.1	10.4	11.9	10.8	6.6	3.1	
12	Babergh	18.5	12.6	14	13.6	11.8	9.8	12	
13	Barking and Dagenham	22.7	23.8	20.3	23.8	23.9	18.7	16.9	
14	Barnet	9.8	10.1	7.5	11.4	8.8	9.9	6.4	
15	Barnsley	18.7	19.5	18.3	17.7	19.9	15.3	13.5	
16	Barrow-in-Furness	11.1	12.3	13.8	13.2	12.5	8.5	4.4	
17	Basildon	18.4	14.6	16.1	15	20	17	12.3	
18	Basingstoke and Deane	8.4	6.4	7.9	9.6	10.1	10.3	7.9	
19	Bassetlaw	17.9	17.5	16.5	14.5	21.2	14.4	11.7	
20	Bath and North East Somerset	9.3	9.1	7.5	7.5	7.5	7.1	6.5	
21	Bedford	13.1	15.1	15.5	11.8	11.7	11.6	8	
22	Bexley	11.2	12.3	9.8	13.6	12.5	10.9	9.5	
23	Birmingham	19.4	21.4	20	19.1	21.4	20.6	19.4	
24	Blaby	11.3	15.2	14.7	9.2	16.9	15.1	18.7	
25	Blackburn with Darwen	18.5	20.2	20	19.1	20	18.2	18.1	
26	Blackpool	21.1	20.9	23.6	22.1	20.5	18.8	16.5	
27	Bolsover	31.1	23.2	15.4	15.5	18.3	11.2	15.8	
28	Bolton	15.1	19.3	17	13.9	15.8	15.7	12.6	
29	Boston	15.5	18.8	17.5	15.2	15.8	12.2	21.5	

Unchecking the row and column which gives the total sum.

PivotTable Active Field Group Filter

A3 Sum of Value

	A	B	C
1			
2			
3	Sum of Value		
4	47144.53369		
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			

PivotTable Fields

Choose fields to add to report:

Search

☐ Row
☐ Column
☒ **Value**

More Tables...

Drag fields between areas below:

Filters

Double click on the value (cell A4).

Properties

Tools

A1

✕

✓

fx

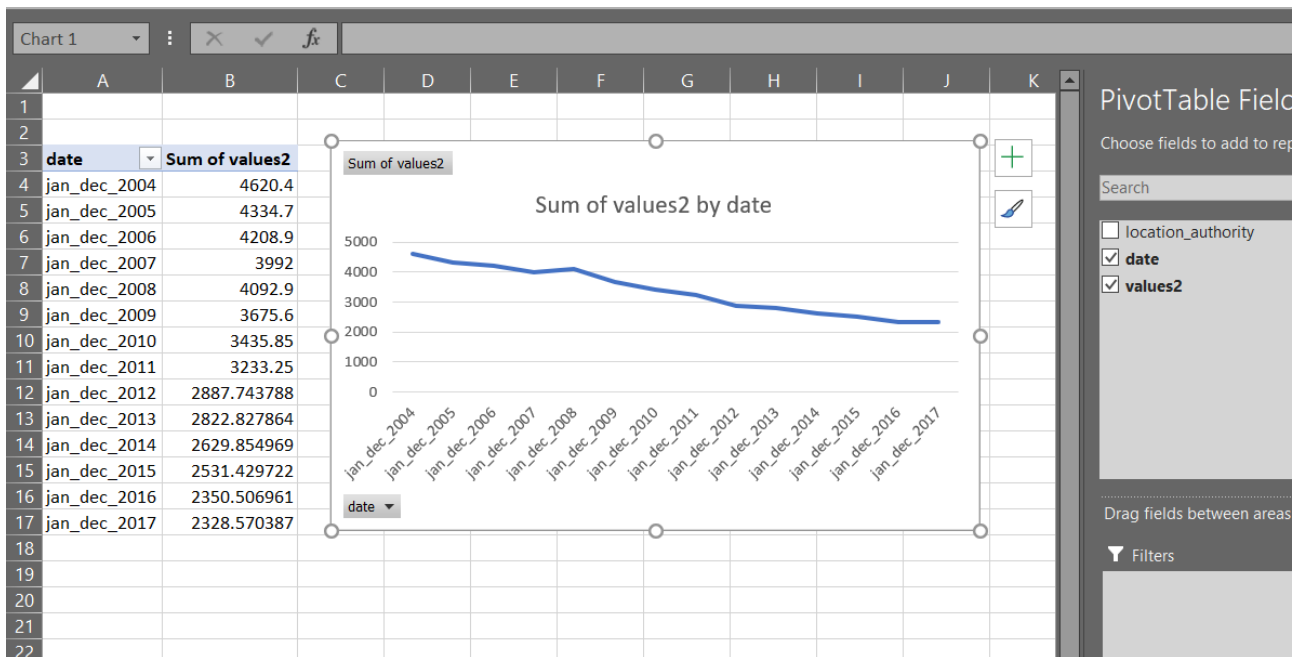
	A	B	C	D
1	Row	Column	Value	
2	Babergh	jan_dec_20	18.5	
3	Babergh	jan_dec_20	12.6	
4	Babergh	jan_dec_20	14	
5	Babergh	jan_dec_20	13.6	
6	Babergh	jan_dec_20	11.8	
7	Babergh	jan_dec_20	9.8	
8	Babergh	jan_dec_20	12	
9	Babergh	jan_dec_20	4.6	
10	Babergh	jan_dec_20	5.1	
11	Babergh	jan_dec_20	9.1	
12	Babergh	jan_dec_20	7	
13	Babergh	jan_dec_20	9.2	

The new pivot table is ready. Renaming the row column and value as location_authority, date and values.

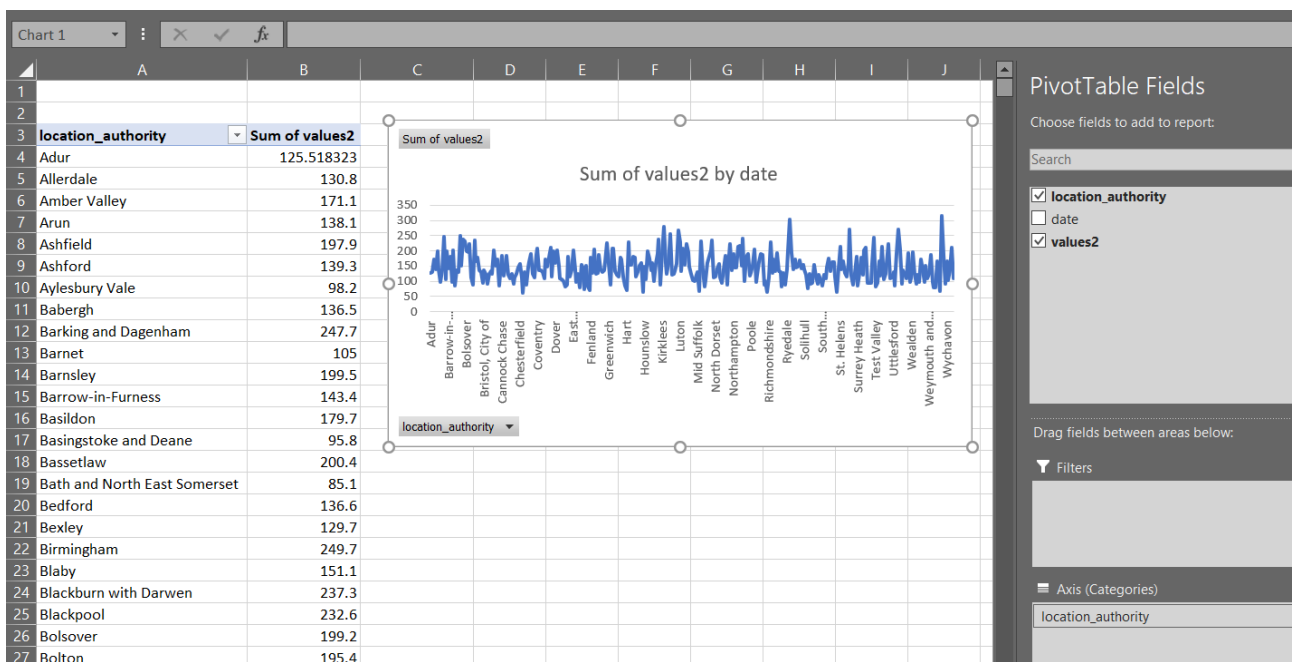
D4				fx	
	A	B	C		
1	location_authority	date	values		
2	Babergh	jan_dec_2004	18.5		
3	Babergh	jan_dec_2005	12.6		
4	Babergh	jan_dec_2006	14		
5	Babergh	jan_dec_2007	13.6		
6	Babergh	jan_dec_2008	11.8		
7	Babergh	jan_dec_2009	9.8		
8	Babergh	jan_dec_2010	12		
9	Babergh	jan_dec_2011	4.6		
10	Babergh	jan_dec_2012	5.1		
11	Babergh	jan_dec_2013	9.1		
12	Babergh	jan_dec_2014	7		
13	Babergh	jan_dec_2015	9.2		

Renaming the sheet name as pivot_table

The obtained data representation:

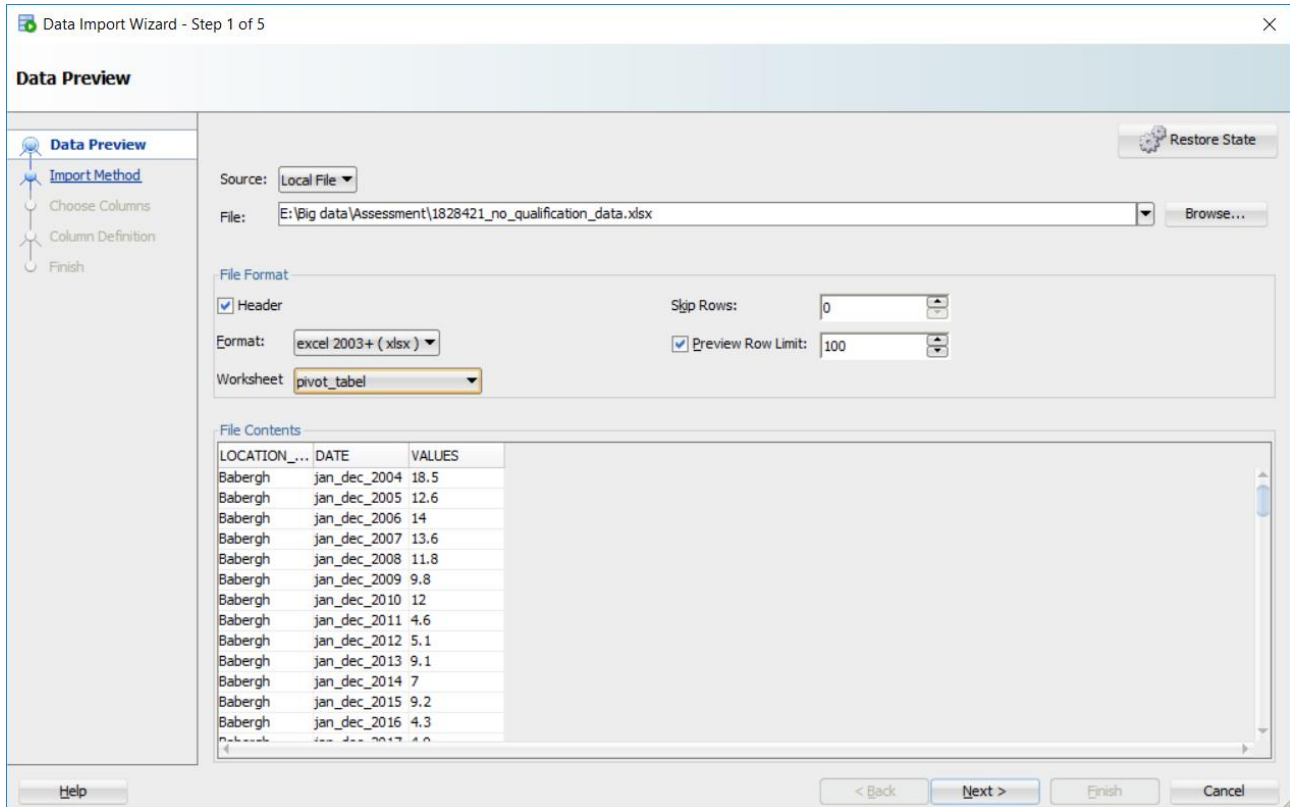


Without location_authority



Without date field, since there are huge locations, so the line graph is huge.

Now the data is ready to import in oracle.



Data Import Wizard - Step 1 of 5

Data Preview

Source: Local File

File: E:\Big data\Assessment\1828421_no_qualification_data.xlsx

File Format

☒ Header

Format: excel 2003+ (xlsx)

Worksheet: pivot_table

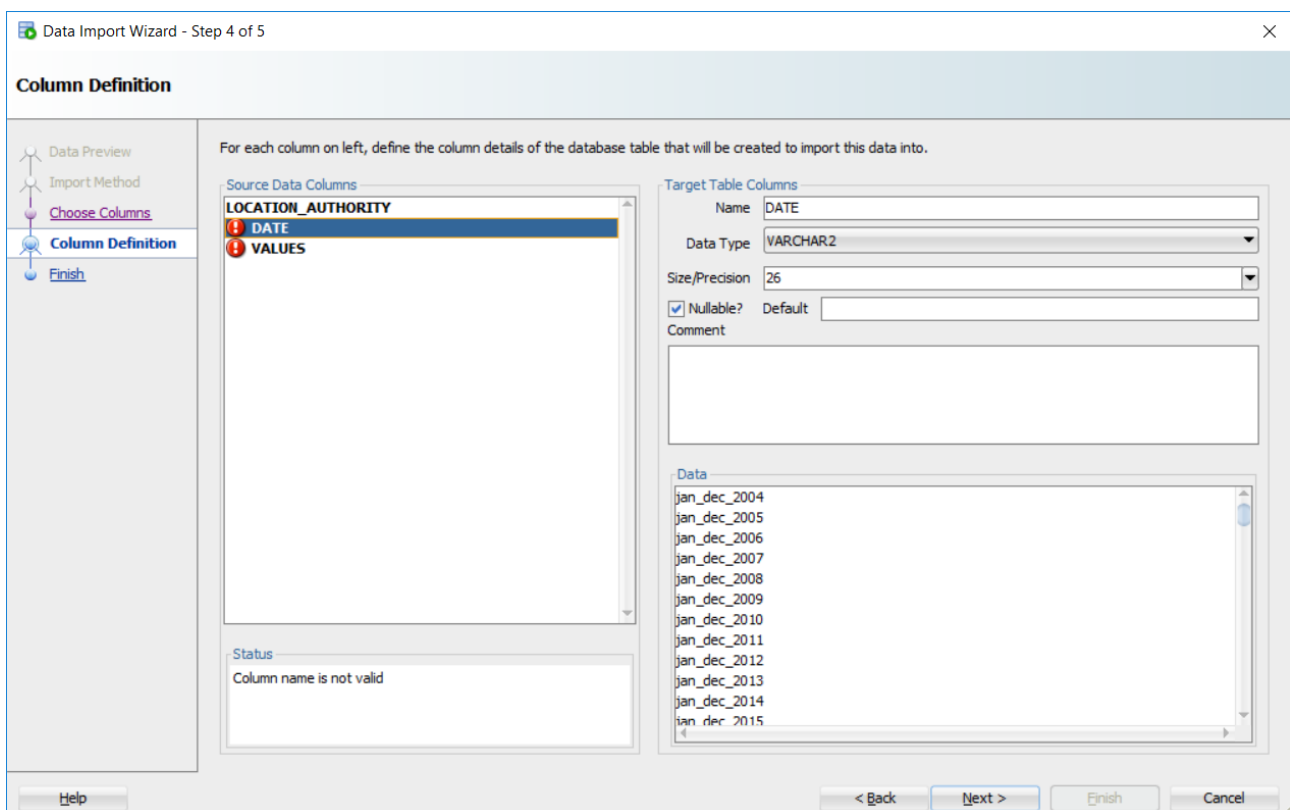
Skip Rows: 0

☒ Preview Row Limit: 100

File Contents

LOCATION_...	DATE	VALUES
Babergh	jan_dec_2004	18.5
Babergh	jan_dec_2005	12.6
Babergh	jan_dec_2006	14
Babergh	jan_dec_2007	13.6
Babergh	jan_dec_2008	11.8
Babergh	jan_dec_2009	9.8
Babergh	jan_dec_2010	12
Babergh	jan_dec_2011	4.6
Babergh	jan_dec_2012	5.1
Babergh	jan_dec_2013	9.1
Babergh	jan_dec_2014	7
Babergh	jan_dec_2015	9.2
Babergh	jan_dec_2016	4.3

Help < Back Next > Finish Cancel



Data Import Wizard - Step 4 of 5

Column Definition

For each column on left, define the column details of the database table that will be created to import this data into.

Source Data Columns

- LOCATION_AUTHORITY
- DATE
- VALUES

Target Table Columns

Name: DATE

Data Type: VARCHAR2

Size/Precision: 26

☒ Nullable? Default

Comment

Data

jan_dec_2004
jan_dec_2005
jan_dec_2006
jan_dec_2007
jan_dec_2008
jan_dec_2009
jan_dec_2010
jan_dec_2011
jan_dec_2012
jan_dec_2013
jan_dec_2014
jan_dec_2015

Status

Column name is not valid

Help < Back Next > Finish Cancel

local_authority is column name is checked not null and increasing the size/precision from 38 to 70 since location name are larger than 38. The date and values column

is not valid because it is sql keywords of oracle which can create problem while extracting data. So, changing date as year and values as result.

Data Import Wizard - Step 4 of 5

Column Definition

For each column on left, define the column details of the database table that will be created to import this data into.

Source Data Columns

Source Data Columns
LOCATION_AUTHORITY
DATE
VALUES

Target Table Columns

Target Table Columns
Name: RESULT
Data Type: NUMBER
Size/Precision: 38
Scale: 6
Nullable? <input checked="" type="checkbox"/>
Default:
Comment:

Data

Data
18.5
12.6
13.6
11.8
9.8
4.6
5.1
9.1
12

Status

Data is not compatible with column definition or is not available for a not nullable column

Help < Back Next > Finish Cancel

Values gives warning because of its datatype. Some values in result are integer and some are in decimal. Therefore, warning is given. If datatype is changed to varchar2 it won't give any warning, but we must need its values in number format, neglecting the warning and leaving it as NUMBER datatype.

The default Data Type is in number format which gives warning because all values do not have decimal numbers. Since arithmetic operation is required leaving the data type as it is.

Import Data

Import Data into table pivot_table from file E:\Big data\Assessment\1828421_no_qualification_data.xlsx . Task successful and import committed.

OK

The structured table is imported successfully into database.

Connections

- MyConnection
 - prakash
 - Tables
 - PIVOT_TABLE
 - Views
 - Indexes
 - Packages
 - Procedures
 - Functions
 - Operators
 - Queues
 - Queues Tables
 - Triggers
 - Types
 - Sequences
 - Materialized Views
 - Materialized View Logs

Reports

- All Reports
- Analytic View Reports
- Data Dictionary Reports
- Data Modeler Reports

...sql | prakash | Import-1828421_no_qualification_data-xlsx-bad_2019

Columns | Data | Model | Constraints | Grants | Statistics | Triggers | Flashback | Depend

Sort.. | Filter:

	LOCATION_AUTHORITY	YEAR	RESULT
1	Norwich	jan_dec_2008	9.8
2	Norwich	jan_dec_2009	8.6
3	Norwich	jan_dec_2010	10.9
4	Norwich	jan_dec_2011	12.3
5	Norwich	jan_dec_2012	9.2
6	Norwich	jan_dec_2013	10.3
7	Norwich	jan_dec_2014	8.5
8	Norwich	jan_dec_2015	6.5
9	Norwich	jan_dec_2016	6.1
10	Norwich	jan_dec_2017	7
11	Peterborough	jan_dec_2004	16.5
12	Peterborough	jan_dec_2005	15.3
13	Peterborough	jan_dec_2006	16.7
14	Peterborough	jan_dec_2007	15.3
15	Peterborough	jan_dec_2008	16
16	Peterborough	jan_dec_2009	14.2
17	Peterborough	jan_dec_2010	12.9
18	Peterborough	jan_dec_2011	12.3

C. SQL

ROLLUP query presents data in hierarchical format where CUBE presents query in combination form. Therefore, more null values will be obtained from CUBE query.

Two sql queries which demonstrates an Online Analytical Processing (OLAP) query is presented below:

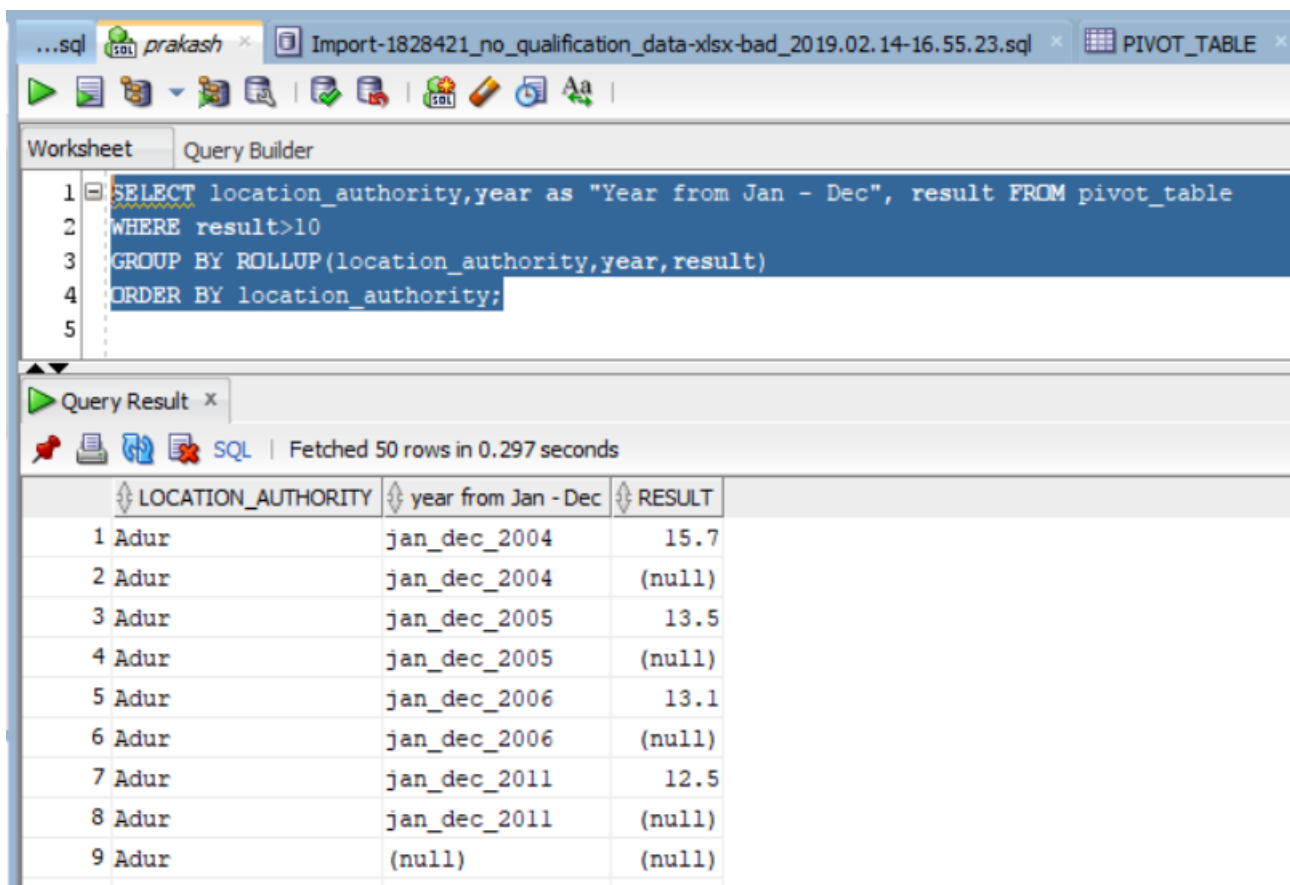
a. ROLLUP Query:

```
SELECT location_authority,year as "Year from Jan - Dec", result FROM  
pivot_table
```

```
WHERE result>10
```

```
GROUP BY ROLLUP(location_authority,year,result)
```

```
ORDER BY location_authority;
```



	LOCATION_AUTHORITY	year from Jan - Dec	RESULT
1	Adur	jan_dec_2004	15.7
2	Adur	jan_dec_2004	(null)
3	Adur	jan_dec_2005	13.5
4	Adur	jan_dec_2005	(null)
5	Adur	jan_dec_2006	13.1
6	Adur	jan_dec_2006	(null)
7	Adur	jan_dec_2011	12.5
8	Adur	jan_dec_2011	(null)
9	Adur	(null)	(null)

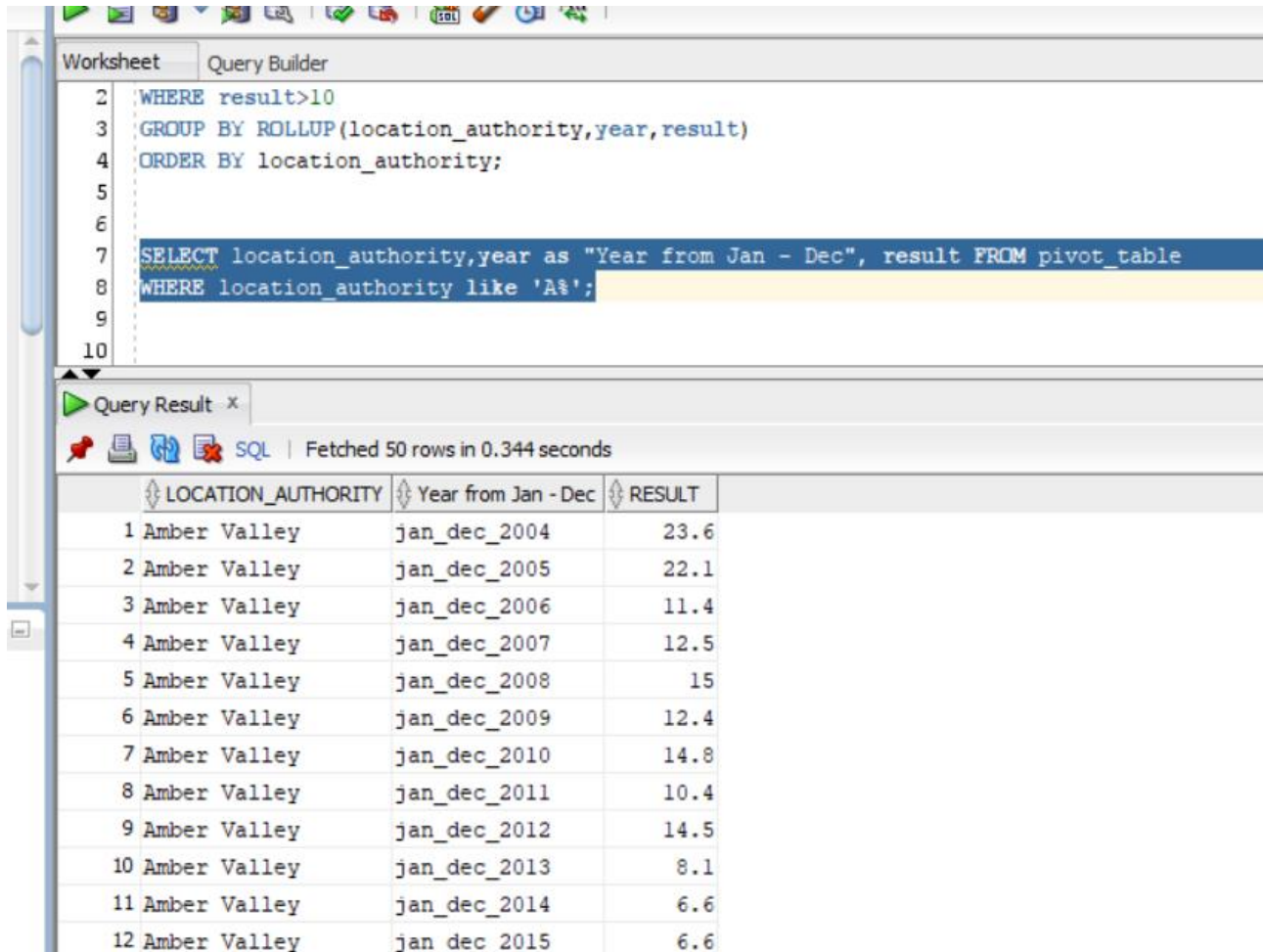
4566	Wyre Forest	jan_dec_2013	11.9
4567	Wyre Forest	jan_dec_2013	(null)
4568	Wyre Forest	jan_dec_2014	15.1
4569	Wyre Forest	jan_dec_2014	(null)
4570	Wyre Forest	jan_dec_2015	14.7
4571	Wyre Forest	jan_dec_2015	(null)
4572	Wyre Forest	jan_dec_2017	10.4
4573	Wyre Forest	jan_dec_2017	(null)
4574	Wyre Forest	(null)	(null)
4575	York	jan_dec_2004	11.8
4576	York	jan_dec_2004	(null)
4577	York	jan_dec_2005	11.6
4578	York	jan_dec_2005	(null)
4579	York	jan_dec_2006	12.8
4580	York	jan_dec_2006	(null)
4581	York	(null)	(null)
4582	(null)	(null)	(null)

This query selects all three columns, year as given allies, from the table pivot_table. It only displays the output if the result values are greater than 10. It is groped using ROLLUP function which gives result in hierarchy, which means result column can be null but if its upper column is null then it must be null similarly if location_authority is null then year is also null, so the result column is also null. Finally, when it gives output, it is sorted in an ascending order (i.e. A-Z) by location_authority column.

b. Group By Query:

SELECT location_authority, year as "Year from Jan - Dec", result FROM pivot_table

WHERE location_authority like 'A%';



The screenshot shows a database query builder interface. The top section is labeled 'Query Builder' and contains a SQL query. The query is as follows:

```
2 WHERE result > 10
3 GROUP BY ROLLUP(location_authority, year, result)
4 ORDER BY location_authority;
5
6
7 SELECT location_authority, year as "Year from Jan - Dec", result FROM pivot_table
8 WHERE location_authority like 'A%';
9
10
```

The bottom section is labeled 'Query Result' and shows the results of the query. It indicates that 50 rows were fetched in 0.344 seconds. The results are displayed in a table with three columns: LOCATION_AUTHORITY, Year from Jan - Dec, and RESULT.

	LOCATION_AUTHORITY	Year from Jan - Dec	RESULT
1	Amber Valley	jan_dec_2004	23.6
2	Amber Valley	jan_dec_2005	22.1
3	Amber Valley	jan_dec_2006	11.4
4	Amber Valley	jan_dec_2007	12.5
5	Amber Valley	jan_dec_2008	15
6	Amber Valley	jan_dec_2009	12.4
7	Amber Valley	jan_dec_2010	14.8
8	Amber Valley	jan_dec_2011	10.4
9	Amber Valley	jan_dec_2012	14.5
10	Amber Valley	jan_dec_2013	8.1
11	Amber Valley	jan_dec_2014	6.6
12	Amber Valley	jan_dec_2015	6.6

This query selects all three columns from pivot_table of only the column location_authority which starts with A.

D. Advantages

The advantage of using this approach in big data are explained below:

- i. Data cleaning:

The data which we may not be in proper format but in this approach since the data should be in an appropriate format, it must be cleaned so Null, missing values can be assigned taking its average or median. Because of this data analysis can be effective.

E. Disadvantages

Disadvantage of this approach for using big data are given below:

- i. Deletion of inappropriate data:

If the data is not in a structured format, then it must be cleaned. If there are null data, then it should be filled. But if many values are missing then it is difficult to assign values in it. So, the whole row or column must be removed which may result in the deletion of important values as well.