









# Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

## Data Collected from the Internet and Entered into Excel





	A	B	C	D	E	F	G
1							
2	U302 SAT - Fast Food Outlets - Population + Outlets						
3							
4	<b>Regional Cities</b>	<b>Population (2011 Census)</b>		<b>Mc Donalds</b>	<b>KFC</b>	<b>Hungry J's</b>	<b>Subway</b>
5	Ballarat	86600		4	3	1	8
6	Bendigo	79600		4	4	1	6
7	Geelong	157700		7	6	1	8
8	Shepparton	45400		3	3	1	5
9	Mildura	47500		2	2	1	2
10	Wodonga	31100		2	2	1	2
11							
12	<b>Suburbs / Towns</b>	<b>Population (2011 Census)</b>					
13	Berwick(Nth and Sth)	45237		3	1	1	3
14	Narre Warren (Nth and Sth)	54476		2	2	0	3
15	Cranbourne (NSEW)	56862		3	1	0	5
16	Warragul	14074		1	1	0	1
17	Dandenong	25923		3	2	1	6
18	Frankston	21583		3	3	1	5
19	Ringwood	16009		1	2	1	4
20	Werribee	40218		3	2	1	3
21	Hoppers Crossing	37598		1	1	0	2
22	Footscray	13203		1	1	1	3
23	Brighton	21257		0	1	0	3
24	Hampton	16876		0	0	0	2
25	Hawthorn	20704		1	0	0	2
26	Toorak	12871		0	0	0	1
27	South Yarra	20661		1	0	0	1
28	Torradin	1359		0	0	0	0
29	Cheltenham	20365		1	1	0	2
30	Mordialloc	18722		0	0	0	1
31	Rosebud	19622		1	1	0	1
32	Mornington	22421		1	1	1	2
33	Lalor	19873		0	0	0	0
34	Epping	26855		2	2	1	4
35	Northcote	22920		0	0	0	0
36	Brunswick	22764		2	0	0	2
37	Carlton	22404		0	1	0	2
38	Fitzroy (and Nth)	20903		0	0	0	0
39	Elsternwick	10741		1	0	0	1
40	St Kilda	22550		1	0	0	3

# Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

## 1) Creating Some Column Totals

First added some Column Totals (Do Row Totals Later after Sorting)

We used =SUM() formulas like this one: =SUM(D5:D10) and so on

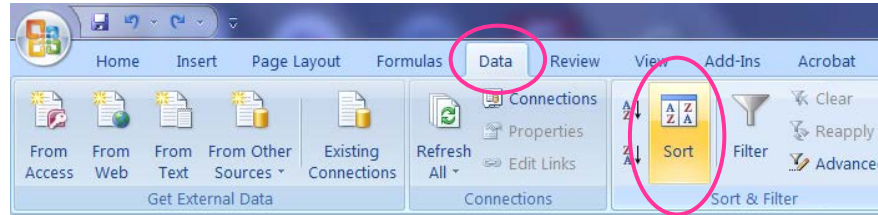
	A	B	C	D	E	F	G
1							
2	U302 SAT - Fast Food Outlets - Population + Outlets						
3							
4	<b>Regional Cities</b>	<b>Population (2011 Census)</b>		<b>Mc Donalds</b>	<b>KFC</b>	<b>Hungry J's</b>	<b>Subway</b>
5	Geelong	157700		7	6	1	8
6	Ballarat	86600		4	3	1	8
7	Bendigo	79600		4	4	1	6
8	Mildura	47500		2	2	1	2
9	Shepparton	45400		3	3	1	5
10	Wodonga	31100		2	2	1	2
11		<b>CITY TOTALS</b>		<b>22</b>	<b>20</b>	<b>6</b>	<b>31</b>

Then we did the same thing for the Towns/Suburbs Group:

	Suburbs / Towns	Population (2011 Census)				
14	Cranbourne (NSEW)	56862	3	1	0	5
15	Narre Warren (Nth and Sth)	54476	2	2	0	3
16	Berwick(Nth and Sth)	45237	3	1	1	3
17	Werribee	40218	3	2	1	3
18	Hoppers Crossing	37598	1	1	0	2
19	Epping	26855	2	2	1	4
20	Dandenong	25923	3	2	1	9
21	Northcote	22920	0	0	0	0
22	Brunswick	22764	2	0	0	2
23	St Kilda	22550	1	0	0	3
24	Mornington	22421	1	1	1	2
25	Carlton	22404	0	1	0	2
26	Frankston	21583	3	3	1	5
27	Brighton	21257	0	1	0	3
28	Fitzroy (and Nth)	20903	0	0	0	0
29	Hawthorn	20704	1	0	0	2
30	South Yarra	20661	1	0	0	1
31	Cheltenham	20365	1	1	0	2
32	Lalor	19873	0	0	0	0
33	Rosebud	19622	1	1	0	1
34	Mordialloc	18722	0	0	0	1
35	Hampton	16876	0	0	0	2
36	Ringwood	16009	1	2	1	4
37	Warragul	14074	1	1	0	1
38	Footscray	13203	1	1	1	3
39	Toorak	12871	0	0	0	1
40	Elsternwick	10741	1	0	0	1
41	Torradin	1359	0	0	0	0
42		<b>SUBURB TOTALS</b>	<b>32</b>	<b>23</b>	<b>8</b>	<b>65</b>

# Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

## 2) Sorting By Population Values







Produces our Cities Sorted on Column B “Population” from biggest to smallest as shown on the next page:

The screenshot shows an Excel spreadsheet with a table of fast food outlets. The table is sorted by population in descending order. A 'Sort' dialog box is open over the spreadsheet, showing the sorting criteria: 'Sort by Column B', 'Sort On Values', and 'Order Largest to Smallest'. The spreadsheet data is as follows:

Regional Cities	Population (2011 Census)	Mc Donalds	KFC	Hungry J's	Subway
Ballarat	86600	4	3	1	8
Bendigo	79600	4	4	1	6
Geelong	157700	7	6	1	8
Shepparton	45400	3	3	1	5
Mildura	47500	2	2	1	2
Wodonga	31100	2	2	1	2
<b>CITY TOTALS</b>		<b>22</b>	<b>20</b>	<b>6</b>	<b>31</b>
Suburbs / Towns	Population (2011 Census)				
Brunswick	22764	2	0	0	2
Carlton	22404	0	1	0	2
Fitzroy (and Nth)	20903	0	0	0	0
Elsternwick	10741	1	0	0	1
St Kilda	22550	1	0	0	3

## Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

**OUTPUT: Regional Cities Sorted from Biggest to Smallest:**

	A	B	C	D	E	F	G
1							
2	U302 SAT - Fast Food Outlets - Population + Outlets						
3							
4	<b>Regional Cities</b>	<b>Population (2011 Census)</b>		<b>Mc Donalds</b>	<b>KFC</b>	<b>Hungry J's</b>	<b>Subway</b>
5	Geelong	157700		7	6	1	8
6	Ballarat	86600		4	3	1	8
7	Bendigo	79600		4	4	1	6
8	Mildura	47500		2	2	1	2
9	Shepparton	45400		3	3	1	5
10	Wodonga	31100		2	2	1	2
11	<b>CITY TOTALS</b>			<b>22</b>	<b>20</b>	<b>6</b>	<b>31</b>





Then we did the same thing for the Towns/Suburbs Data:

13	Suburbs / Towns	Population (2011 Census)				
14	Cranbourne (NSEW)	56862	3	1	0	5
15	Narre Warren (Nth and Sth)	54476	2	2	0	3
16	Berwick(Nth and Sth)	45237	3	1	1	3
17	Werribee	40218	3	2	1	3
18	Hoppers Crossing	37598	1	1	0	2
19	Epping	26855	2	2	1	4
20	Dandenong	25923	3	2	1	9
21	Northcote	22920	0	0	0	0
22	Brunswick	22764	2	0	0	2
23	St Kilda	22550	1	0	0	3
24	Mornington	22421	1	1	1	2
25	Carlton	22404	0	1	0	2
26	Frankston	21583	3	3	1	5
27	Brighton	21257	0	1	0	3
28	Fitzroy (and Nth)	20903	0	0	0	0
29	Hawthorn	20704	1	0	0	2
30	South Yarra	20661	1	0	0	1
31	Cheltenham	20365	1	1	0	2
32	Lalor	19873	0	0	0	0
33	Rosebud	19622	1	1	0	1
34	Mordialloc	18722	0	0	0	1
35	Hampton	16876	0	0	0	2
36	Ringwood	16009	1	2	1	4
37	Warragul	14074	1	1	0	1
38	Footscray	13203	1	1	1	3
39	Toorak	12871	0	0	0	1
40	Elsternwick	10741	1	0	0	1
41	Torradin	1359	0	0	0	0
42	<b>SUBURB TOTALS</b>		<b>32</b>	<b>23</b>	<b>8</b>	<b>65</b>

# Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

## 3) Adding Horizontal Totals

We use =SUM() formula to work out the Total number of Fast Food outlets for each City, Suburb, and Town. Eg. =SUM(D5:G5) etc

	A	B	C	D	E	F	G	H
1								
2	U302 SAT - Fast Food Outlets - Population + Outlets							
3								
4	<b>Regional Cities</b>	<b>Population (2011 Census)</b>		<b>Mc Donalds</b>	<b>KFC</b>	<b>Hungry J's</b>	<b>Subway</b>	<b>TOTALS</b>
5	Geelong	157700		7	6	1	8	22
6	Ballarat	86600		4	3	1	8	16
7	Bendigo	79600		4	4	1	6	15
8	Mildura	47500		2	2	1	2	7
9	Shepparton	45400		3	3	1	5	12
10	Wodonga	31100		2	2	1	2	7
11		<b>CITY TOTALS</b>		<b>22</b>	<b>20</b>	<b>6</b>	<b>31</b>	<b>79</b>
12								
13	<b>Suburbs / Towns</b>	<b>Population (2011 Census)</b>						
14	Cranbourne (NSEW)	56862		3	1	0	5	9
15	Narre Warren (Nth and Sth)	54476		2	2	0	3	7
16	Berwick(Nth and Sth)	45237		3	1	1	3	8
17	Werribee	40218		3	2	1	3	9
18	Hoppers Crossing	37598		1	1	0	2	4
19	Epping	26855		2	2	1	4	9
20	Dandenong	25923		3	2	1	6	12
21	Northcote	22920		0	0	0	0	0
22	Brunswick	22764		2	0	0	2	4
23	St Kilda	22550		1	0	0	3	4
24	Mornington	22421		1	1	1	2	5
25	Carlton	22404		0	1	0	2	3
26	Frankston	21583		3	3	1	5	12
27	Brighton	21257		0	1	0	3	4
28	Fitzroy (and Nth)	20903		0	0	0	0	0
29	Hawthorn	20704		1	0	0	2	3
30	South Yarra	20661		1	0	0	1	2
31	Cheltenham	20365		1	1	0	2	4
32	Lalor	19873		0	0	0	0	0
33	Rosebud	19622		1	1	0	1	3
34	Mordialloc	18722		0	0	0	1	1
35	Hampton	16876		0	0	0	2	2
36	Ringwood	16009		1	2	1	4	8
37	Warragul	14074		1	1	0	1	3
38	Footscray	13203		1	1	1	3	6
39	Toorak	12871		0	0	0	1	1
40	Elsternwick	10741		1	0	0	1	2
41	Torradin	1359		0	0	0	0	0
42		<b>SUBURB TOTALS</b>		<b>32</b>	<b>23</b>	<b>8</b>	<b>62</b>	<b>125</b>

# Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

## 4) Applying Filtering

We want to filter so that we can get a list just of the Locations with Zero Fast Food outlets.

First we highlight all of the Data, (first make sure all of our columns have headings on them), then on the ribbon we do Data and click the Filter Funnel. (Note that clicking the Funnel is a toggle and turns filtering on and off).

Suburbs / Towns	Population (2011 Census)	Mc Donalds	KFC	Hungry J's	Subway	TOTALS
CITY TOTALS		22	20	6	31	79
Suburbs / Towns	Population (2011 Census)	Mc Donalds	KFC	Hungry J's	Subway	TOTALS
Cranbourne (NSEW)	56862	3	1	0	5	9
Narre Warren (Nth and Sth)	54476	2	2	0	3	7
Berwick(Nth and Sth)	45237	3	1	1	3	8
Werribee	40218	3	2	1	3	9
Hoppers Crossing	37598	1	1	0	2	4
Epping	26855	2	2	1	4	9
Dandenong	25923	3	2	1	6	12
Northcote	22920	0	0	0	0	0
Brunswick	22764	2	0	0	2	4
St Kilda	22550	1	0	0	3	4
Mornington	22421	1	1	1	2	5
Carlton	22404	0	1	0	2	3
Frankston	21583	3	3	1	5	12
Brighton	21257	0	1	0	3	4
Fitzroy (and Nth)	20903	0	0	0	0	0
Hawthorn	20704	1	0	0	2	3
South Yarra	20661	1	0	0	1	2
Cheltenham	20365	1	1	0	2	4
Lalor	19873	0	0	0	0	0
Rosebud	19622	1	1	0	1	3
Mordialloc	18722	0	0	0	1	1
Hampton	16876	0	0	0	2	2
Ringwood	16009	1	2	1	4	8
Warragul	14074	1	1	0	1	3
Footscray	13203	1	1	1	3	6
Toorak	12871	0	0	0	1	1
Elsternwick	10741	1	0	0	1	2
Tooradin	1359	0	0	0	0	0
SUBURB TOTALS		32	23	8	62	125

When we click filter we get a series of down arrows added to our sheet:

# Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

We need to click the Totals far right hand column, and then set it up to choose only values that have a zero in them:

The screenshot shows the Microsoft Excel interface with the 'Data' tab selected in the ribbon. The 'Filter' icon is highlighted. A data table is displayed with columns for 'Suburbs / Towns', 'Population (2011 Census)', 'Mc Donald', 'KFC', 'Hungry J', 'Subway', and 'TOTAL'. The 'TOTAL' column is highlighted, and a filter menu is open over it. The filter menu is set to 'Filter by Value' and '0' is selected in the list. The 'Data' tab in the ribbon is also highlighted.

Suburbs / Towns	Population (2011 Census)	Mc Donald	KFC	Hungry J	Subway	TOTAL
Cranbourne (NSEW)	56862	3	1			
Narre Warren (Nth and Sth)	54476	2	2			
Berwick(Nth and Sth)	45237	3	1			
Werribee	40218	3	2			
Hoppers Crossing	37598	1	1			
Epping	26855	2	2			
Dandenong	25923	3	2			
Northcote	22920	0	0			
Brunswick	22764	2	0			
St Kilda	22550	1	0			
Mornington	22421	1	1			
Carlton	22404	0	1			
Frankston	21583	3	3			
Brighton	21257	0	1			
Fitzroy (and Nth)	20903	0	0			
Hawthorn	20704	1	0			
South Yarra	20661	1	0			
Cheltenham	20365	1	1			
Lalor	19873	0	0			
Rosebud	19622	1	1			
Mordialloc	18722	0	0	0	1	1
Hampton	16876	0	0	0	2	2
Ringwood	16009	1	2	1	4	8
Warragul	14074	1	1	0	1	3
Footscray	13203	1	1	1	3	6
Toorak	12871	0	0	0	1	1
Elsternwick	10741	1	0	0	1	2
Tooradin	1359	0	0	0	0	0
<b>SUBURB TOTALS</b>		<b>32</b>	<b>23</b>	<b>8</b>	<b>62</b>	<b>125</b>

## Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

When we click okay we get our “Filtered” set of Data like this:

### 4) Applying Filtering - OUTPUT

	A	B	C	D	E	F	G	H
12								
13	Suburbs / Towns	Population (2011 Census)		Mc Donald	KFC	Hungry J	Subway	TOTAL
21	Northcote	22920		0	0	0	0	0
28	Fitzroy (and Nth)	20903		0	0	0	0	0
32	Lalor	19873		0	0	0	0	0
41	Tooradin	1359		0	0	0	0	0

*We need to do some investigation into why these particular Suburbs, especially the ones with large populations have no American fast Food outlets at all! Because they are “outliers” or “anomalies” or “abnormal” we may need to explain them away and remove them from our data at some stage.*

**If we now click the big filter symbol in the ribbon, we can turn all the filtering off and return our spreadsheet to normal.**

### 5) Rounding Off Population Values

Our original Hypothesis says the number of Fast Food outlets enables us to “ESTIMATE” the population, not work it out to the nearest exact person figure:

**“The population of a city, town, or suburb can be **estimated** from the number of American Fast Food outlets it has, because bigger towns have enough people to sustain a larger number of outlets.”**





This means that it is okay to round our Population Figures off the the nearest 1000, 5000, or 10 000.

We want to be able to do some Grouping of our data based on rounded off Populations, so we have decided to round off all of our populations to the NEAREST TEN THOUSAND.

Our new Spreadsheet of rounded off values is shown below:



Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

	A	B	C	D	E	F	G	H
1	<b>U302 SAT - Fast Food Outlets - Population + Outlets</b>							
2	<b>(Populations Rounded off to nearest 10 000)</b>							
3								
4	<b>Regional Cities</b>	<b>Population (2011 Census)</b>		<b>Mc Donalds</b>	<b>KFC</b>	<b>Hungry J's</b>	<b>Subway</b>	<b>TOTALS</b>
5	Geelong	160000		7	6	1	8	22
6	Ballarat	90000		4	3	1	8	16
7	Bendigo	80000		4	4	1	6	15
8	Mildura	50000		2	2	1	2	7
9	Shepparton	50000		3	3	1	5	12
10	Wodonga	30000		2	2	1	2	7
11		<b>CITY TOTALS</b>		<b>22</b>	<b>20</b>	<b>6</b>	<b>31</b>	<b>79</b>
12								
13	<b>Suburbs / Towns</b>	<b>Population (2011 Census)</b>		<b>Mc Donalds</b>	<b>KFC</b>	<b>Hungry J's</b>	<b>Subway</b>	<b>TOTALS</b>
14	Cranbourne (NSEW)	60000		3	1	0	5	9
15	Narre Warren (Nth and Sth)	60000		2	2	0	3	7
16	Berwick(Nth and Sth)	40000		3	1	1	3	8
17	Werribee	40000		3	2	1	3	9
18	Hoppers Crossing	40000		1	1	0	2	4
19	Epping	30000		2	2	1	4	9
20	Dandenong	30000		3	2	1	6	12
21	Northcote	20000		0	0	0	0	0
22	Brunswick	20000		2	0	0	2	4
23	St Kilda	20000		1	0	0	3	4
24	Mornington	20000		1	1	1	2	5
25	Carlton	20000		0	1	0	2	3
26	Frankston	20000		3	3	1	5	12
27	Brighton	20000		0	1	0	3	4
28	Fitzroy (and Nth)	20000		0	0	0	0	0
29	Hawthorn	20000		1	0	0	2	3
30	South Yarra	20000		1	0	0	1	2
31	Cheltenham	20000		1	1	0	2	4
32	Lalor	20000		0	0	0	0	0
33	Rosebud	20000		1	1	0	1	3

## Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

### 6) Group Rounded Off Data and Investigate Zero Outlets Locations

We want to Group the Data and then calculate: Mean, Median, Mode, and Standard Deviation for each of the Groups.

The Zero outlets locations need to be removed, as they will drag averages down.

Here again (from when we filtered) are the Zero Fast Food Outlets Locations:

	A	B	C	D	E	F	G	H
12								
13	Suburbs / Towns	Population (2011 Census)		Mc Donald	KFC	Hungry J	Subway	TOTAL
21	Northcote	22920		0	0	0	0	0
28	Fitzroy (and Nth)	20903		0	0	0	0	0
32	Lalor	19873		0	0	0	0	0
41	Tooradin	1359		0	0	0	0	0

Tooradin is a very small country town that does NOT have any of the following factors that would attract Fast Food Outlets:

- 1) Is not on a major state highway
- 2) Does not attract a large number of visitors
- 3) Does not have a large regional high school
- 4) Is not on a major train line
- 5) Does not have a large High School
- 6) Does not have housing estates with lots of young families
- 7) Does not have lots of Tradies working or living there
- 8) Does not have major sporting complexes
- 9) Has an extremely small population: 1300 people.

Northcote and Fitzroy are both inner city suburbs heavily populated by university students and young urban professionals, eg. Very few young Families, High School students, or Tradies.

They also have all of the same 1) to 8) characteristics as Tooradin.

## Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

Lalor has around 20,000 residents and it appears to be a normal suburb, eg. <http://lalor.localstats.com.au/demographics/vic/melbourne/north/lalor> HOWEVER Only 31% of the people living there speak English as a first language, and it does have one of Melbourne’s highest unemployment rates.





Lalor has a train station, a high school, has a shopping centre with a Coles and 27 specialty shops. It does not have any major sports venues.

In the Yellow pages, Lalor has several Pizza Shops, Fish and Chip Shops, Kebab shops, and a Red Rooster, in all about 14 shops. <http://www.yellowpages.com.au/find/fast-food/lalor-vic>

There are McDonalds, KFC, etc in the neighbouring suburb of Epping.

There are not any real reasons for it not to have major take away shops, perhaps there are not any good locations to place a shop, because Lalor is an old suburb with old narrow roads, and not open areas or room to create parking lots and drive throughs.

**To further investigate the abnormality for Lalor, we created a Spreadsheet of the Top 10 Victorian Poor Towns and Suburbs.**

	A	B	C	D	E	F	G	H	I	J	K
1	<b>U302 SAT - Fast Food Outlets - Poorest Suburbs</b>										
2											
3										ZERO OUTLETS	
4	<b>Top 10 Poverty Locations</b>	<b>Population (2011 Census)</b>		<b>Mc Donalds</b>	<b>KFC</b>	<b>Hungry J's</b>	<b>Subway</b>	<b>TOTALS</b>		<b>Comments</b>	
5	Broadmeadows	10578		1	2		2	5			
6	Corio	14919		2	1		1	4			
7	Doveton	8404		1	0		1	2			
8	Frankston North	5626		1	0		0	1			
9	Mayborough	7174		1	1		1	3			
10	Morwell	13691		1	1		2	4			
11	Ardeer	2823		0	0		0	0		Too Small for Outlets	
12	Braybrook	8184		1	0		1	2			
13	Coolaroo	3261		0	0		0	0		Too Small for Outlets	
14	Campbellfield	5467		1	0		1	2			
15	Rosebud West	4579		0	0		0	0		Too Small for Outlets	
16	Wendouree	10752		1	1		1	3			
17	Heathcote	2776		0	0		0	0		Too Small for Outlets	
18	Moe	15292		1	1		1	3			
19	Eaglehawk	5546		0	0		0	0		Too Small for Outlets	
20	Lalor	19873		0	0	0	0	0		Why No Outlets?	
21	St Albans	35091		0	0	0	0	0		Why No Outlets?	
22	St Arnaud	2619		0	0		0	0		Too Small for Outlets	
23											
24											
25	Reference for List of Poorest Areas:										
26	<a href="http://www.heraldsun.com.au/news/victoria/victorias-worst-struggletowns-revealed-in-dropping-off-the-edge-report/news-story/b9fc9b474177285e73c1690144b860bf">http://www.heraldsun.com.au/news/victoria/victorias-worst-struggletowns-revealed-in-dropping-off-the-edge-report/news-story/b9fc9b474177285e73c1690144b860bf</a>										
27											
28	Lalor - Only 31% of people living there speak English as their main language.										
29	St Albans - Only 24% of people speak English as their main language										
30											
31	This seems to be the only difference between these "poor" locations and the others.										

## Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT





We found amongst the “poor” suburbs, that St Albans, with 1.5 as many people as Lalor also has no outlets, but it also has a Train Station, Shops, High School, etc.

Two things which Lalor and St Albans do have in common is a low percentage of people who speak English as their main language.

**For our SAT, Lalor is an uncommon exception, and so we are going to remove it from our data. We will need to discuss this in the SAT conclusion.**

However, we will do this later on in our manipulation, after we have done Mean, Median, Mode, and Standard Deviation calculations on our Groups.

### Rounded Off Grouped Data

	A	B	C	D	E	F	G	H
1	<b>U302 SAT - Fast Food Outlets - Population + Outlets</b>							
2	<i>(Populations Rounded off to nearest 10 000)</i>							
3								
4	<b>Regional Cities</b>	<b>Population (2011 Census)</b>		<b>Mc Donalds</b>	<b>KFC</b>	<b>Hungry J's</b>	<b>Subway</b>	<b>TOTALS</b>
5	Geelong	160000		7	6	1	8	22
6	Ballarat	90000		4	3	1	8	16
7	Bendigo	80000		4	4	1	6	15
8	Mildura	50000		2	2	1	2	7
9	Shepparton	50000		3	3	1	5	12
10	Wodonga	30000		2	2	1	2	7
11		<b>CITY TOTALS</b>		<b>22</b>	<b>20</b>	<b>6</b>	<b>31</b>	<b>79</b>
12								
13	<b>Suburbs / Towns</b>	<b>Population (2011 Census)</b>		<b>Mc Donalds</b>	<b>KFC</b>	<b>Hungry J's</b>	<b>Subway</b>	<b>TOTALS</b>
14	Cranbourne (NSEW)	60000		3	1	0	5	9
15	Narre Warren (Nth and Sth)	60000		2	2	0	3	7
16								
17								
18	Berwick(Nth and Sth)	40000		3	1	1	3	8
19	Werribee	40000		3	2	1	3	9
20	Hoppers Crossing	40000		1	1	0	2	4
21								
22	Epping	30000		2	2	1	4	9
23	Dandenong	30000		3	2	1	6	12
24								
25	Northcote	20000		0	0	0	0	0
26	Brunswick	20000		2	0	0	2	4
27	St Kilda	20000		1	0	0	3	4
28	Mornington	20000		1	1	1	2	5
29	Carlton	20000		0	1	0	2	3
30	Frankston	20000		3	3	1	5	12
31	Brighton	20000		0	1	0	3	4
32	Fitzroy (and Nth)	20000		0	0	0	0	0
33	Hawthorn	20000		1	0	0	2	3
34	South Yarra	20000		1	0	0	1	2
35	Cheltenham	20000		1	1	0	2	4
36	Lalor	20000		0	0	0	0	0
37	Rosebud	20000		1	1	0	1	3
38	Mordialloc	20000		0	0	0	1	1
39	Hampton	20000		0	0	0	2	2
40	Ringwood	20000		1	2	1	4	8
41								
42	Warragul	10000		1	1	0	1	3
43	Footscray	10000		1	1	1	3	6
44	Toorak	10000		0	0	0	1	1
45	Elsternwick	10000		1	0	0	1	2
46								
47	Tooradin	1000		0	0	0	0	0
48		<b>SUBURB TOTALS</b>		<b>32</b>	<b>23</b>	<b>8</b>	<b>62</b>	<b>125</b>





# Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

## 7) Sorting on Number of Outlets Totals

So far our Sorting has all been based on Sorting and Grouping “Population” values.

We will now do some Sorting on our other Hypothesis variable: “Total Outlets”.

To do this we made another copy of our data spreadsheet that has Totals in it: (and saved it as TotalOutlets\_SORTED.XLS)

	A	B	C	D	E	F	G	H
1								
2	U302 SAT - Fast Food Outlets - Population + Outlets							
3								
4	<b>Regional Cities</b>	<b>Population (2011 Census)</b>		<b>Mc Donalds</b>	<b>KFC</b>	<b>Hungry J's</b>	<b>Subway</b>	<b>TOTALS</b>
5	Geelong	157700		7	6	1	8	22
6	Ballarat	86600		4	3	1	8	16
7	Bendigo	79600		4	4	1	6	15
8	Mildura	47500		2	2	1	2	7
9	Shepparton	45400		3	3	1	5	12
10	Wodonga	31100		2	2	1	2	7
11		<b>CITY TOTALS</b>		<b>22</b>	<b>20</b>	<b>6</b>	<b>31</b>	<b>79</b>
12								
13	<b>Suburbs / Towns</b>	<b>Population (2011 Census)</b>		<b>Mc Donalds</b>	<b>KFC</b>	<b>Hungry J's</b>	<b>Subway</b>	<b>TOTALS</b>
14	Cranbourne (NSEW)	56862		3	1	0	5	9
15	Narre Warren (Nth and Sth)	54476		2	2	0	3	7
16	Berwick(Nth and Sth)	45237		3	1	1	3	8
17	Werribee	40218		3	2	1	3	9
18	Hoppers Crossing	37598		1	1	0	2	4
19	Epping	26855		2	2	1	4	9
20	Dandenong	25923		3	2	1	6	12
21	Northcote	22920		0	0	0	0	0
22	Brunswick	22764		2	0	0	2	4
23	St Kilda	22550		1	0	0	3	4
24	Mornington	22421		1	1	1	2	5
25	Carlton	22404		0	1	0	2	3
26	Frankston	21583		3	3	1	5	12
27	Brighton	21257		0	1	0	3	4
28	Fitzroy (and Nth)	20903		0	0	0	0	0
29	Hawthorn	20704		1	0	0	2	3
30	South Yarra	20661		1	0	0	1	2
31	Cheltenham	20365		1	1	0	2	4
32	Lalor	19873		0	0	0	0	0
33	Rosebud	19622		1	1	0	1	3
34	Mordialloc	18722		0	0	0	1	1
35	Hampton	16876		0	0	0	2	2
36	Ringwood	16009		1	2	1	4	8
37	Warragul	14074		1	1	0	1	3
38	Footscray	13203		1	1	1	3	6
39	Toorak	12871		0	0	0	1	1
40	Elsternwick	10741		1	0	0	1	2
41	Tooradin	1359		0	0	0	0	0
42		<b>SUBURB TOTALS</b>		<b>32</b>	<b>23</b>	<b>8</b>	<b>62</b>	<b>125</b>

## Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT





We then had to remove the two yellow horizontal Totals Rows, and then set up a Criteria Sort on the Cities, to sort on Column H, like this:

The screenshot shows the Microsoft Excel interface with the 'Data' tab selected. The 'Sort' button in the 'Sort & Filter' group is circled in pink. Below the ribbon, a table titled 'Regional Cities' is displayed. The table has columns for City, Population (2011 Census), and counts for four fast-food chains: McDonald's, KFC, Hungry Jack's, and Subway. A 'TOTALS' column is also present. The 'TOTALS' column is circled in pink. A 'Sort' dialog box is open, showing 'Sort by Column H' and 'Order Largest to Smallest'. The 'OK' button in the dialog box is also circled in pink.

Regional Cities	Population (2011 Census)	Mc Donalds	KFC	Hungry J's	Subway	TOTALS
Geelong	157700	7	6	1	8	22
Ballarat	86600	4	3	1	8	16
Bendigo	79600	4	4	1	6	15
Mildura	47500	2	2	1	2	7
Shepparton	45400	3	3	1	5	12
Wodonga	31100	2	2	1	2	7

## Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

The Sorted (Highest to Lowest) Number of Outlets for Cities ended up as shown below:

	A	B	C	D	E	F	G	H
1								
2	U302 SAT - Fast Food Outlets - Population + Outlets							
3								
4	<b>Regional Cities</b>	<b>Population (2011 Census)</b>		<b>Mc Donalds</b>	<b>KFC</b>	<b>Hungry J's</b>	<b>Subway</b>	<b>TOTALS</b>
5	Geelong	157700		7	6	1	8	22
6	Ballarat	86600		4	3	1	8	16
7	Bendigo	79600		4	4	1	6	15
8	Shepparton	45400		3	3	1	5	12
9	Mildura	47500		2	2	1	2	7
10	Wodonga	31100		2	2	1	2	7

The “Cities” group sorted results are not that surprising, the biggest population cities tend to have the largest number of outlets. (except for Shepparton which is on a major through Hwy and so maybe it has extra outlets because lots of people travel through this central Victorian Town and do a food and fuel stop along the way.)

We then did the same type of Sort on the Suburbs, and this revealed the towns highlighted in Orange each had an abnormally large number of Fast Food Outlets for their population size, when they are compared to other towns with the same or bigger populations.

12	Suburbs / Towns	Population (2011 Census)	Mc Donalds	KFC	Hungry J's	Subway	TOTALS
13	Dandenong	25923	3	2	1	6	12
14	Frankston	21583	3	3	1	5	12
15	Cranbourne (NSEW)	56862	3	1	0	5	9
16	Werribee	40218	3	2	1	3	9
17	Epping	26855	2	2	1	4	9
18	Berwick(Nth and Sth)	45237	3	1	1	3	8
19	Ringwood	16009	1	2	1	4	8
20	Narre Warren (Nth and Sth)	54476	2	2	0	3	7
21	Footscray	13203	1	1	1	3	6
22	Mornington	22421	1	1	1	2	5
23	Hoppers Crossing	37598	1	1	0	2	4
24	Brunswick	22764	2	0	0	2	4
25	St Kilda	22550	1	0	0	3	4
26	Brighton	21257	0	1	0	3	4
27	Cheltenham	20365	1	1	0	2	4
28	Carlton	22404	0	1	0	2	3
29	Hawthorn	20704	1	0	0	2	3





## Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

The “Median” is the usual value used in Real Estate Prices and other Suburban statistics, and so from the above results we can say that for a Town of around 65000 people we would expect there to be 4 McDonalds, 3 KFC’s, 1 Hungry Jacks, 6 Subways.

### For the other Suburban and Country Locations Grouped by Population:

	A	B	C	D	E	F	G	H
19	<b>Suburbs / Towns</b>	<b>Population (2011 Census)</b>		<b>Mc Donalds</b>	<b>KFC</b>	<b>Hungry J's</b>	<b>Subway</b>	<b>TOTALS</b>
20	Cranbourne (NSEW)	60000		3	1	0	5	9
21	Narre Warren (Nth and Sth)	60000		2	2	0	3	7
22	<b>MEAN</b> =AVERAGE(____)	<b>60000</b>		<b>2.5</b>	<b>1.5</b>	<b>0</b>	<b>4</b>	<b>8</b>
23	<b>STD DEVIATION</b> =STDDEV(____)	<b>0</b>		<b>0.7</b>	<b>0.7</b>	<b>0.0</b>	<b>1.4</b>	<b>1.4</b>
24	<b>MEDIAN</b> =MEDIAN(____)	<b>60000</b>		<b>2.5</b>	<b>1.5</b>	<b>0</b>	<b>4</b>	<b>8</b>
25	<b>MODE</b> =MODE(____)	<b>60000</b>		<b>#N/A</b>	<b>#N/A</b>	<b>0</b>	<b>#N/A</b>	<b>#N/A</b>
26								
27								
28	Berwick(Nth and Sth)	40000		3	1	1	3	8
29	Werribee	40000		3	2	1	3	9
30	Hoppers Crossing	40000		1	1	0	2	4
31	<b>MEAN</b> =AVERAGE(____)	<b>40000</b>		<b>2.3</b>	<b>1.3</b>	<b>0.7</b>	<b>2.7</b>	<b>7.0</b>
32	<b>STD DEVIATION</b> =STDDEV(____)	<b>0</b>		<b>1.2</b>	<b>0.6</b>	<b>0.6</b>	<b>0.6</b>	<b>2.6</b>
33	<b>MEDIAN</b> =MEDIAN(____)	<b>40000</b>		<b>3</b>	<b>1</b>	<b>1</b>	<b>3</b>	<b>8</b>
34	<b>MODE</b> =MODE(____)	<b>40000</b>		<b>3</b>	<b>1</b>	<b>1</b>	<b>3</b>	<b>#N/A</b>
35								
36	Epping	30000		2	2	1	4	9
37	Dandenong	30000		3	2	1	6	12
38	<b>MEAN</b> =AVERAGE(____)	<b>30000</b>		<b>2.5</b>	<b>2</b>	<b>1</b>	<b>5</b>	<b>10.5</b>
39	<b>STD DEVIATION</b> =STDDEV(____)	<b>0</b>		<b>0.7</b>	<b>0.0</b>	<b>0.0</b>	<b>1.4</b>	<b>2.1</b>
40	<b>MEDIAN</b> =MEDIAN(____)	<b>30000</b>		<b>2.5</b>	<b>2</b>	<b>1</b>	<b>5</b>	<b>10.5</b>
41	<b>MODE</b> =MODE(____)	<b>30000</b>		<b>#N/A</b>	<b>2</b>	<b>1</b>	<b>#N/A</b>	<b>#N/A</b>
42								
43	Northcote	20000		0	0	0	0	0
44	Brunswick	20000		2	0	0	2	4

Note that Mean, Median, Mode, and Standard Deviation have been formatted as 1 decimal place using right click then Format Cells, then Number, and 1 decimal place.

## Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

### 8) Using Standard Deviation to Remove Outliers

In our data we will often have some “outlier” values that are either way bigger than most of the other data values, or a lot smaller.

Sorting and Filtering the Fast Foods Data has identified the following outliers:

Locations which have no outlets at all:

	A	B	C	D	E	F	G	H
12								
13	Suburbs / Towns	Population (2011 Census)		Mc Donalds	KFC	Hungry J	Subway	TOTAL
21	Northcote	22920		0	0	0	0	0
28	Fitzroy (and Nth)	20903		0	0	0	0	0
32	Lalor	19873		0	0	0	0	0
41	Tooradin	1359		0	0	0	0	0

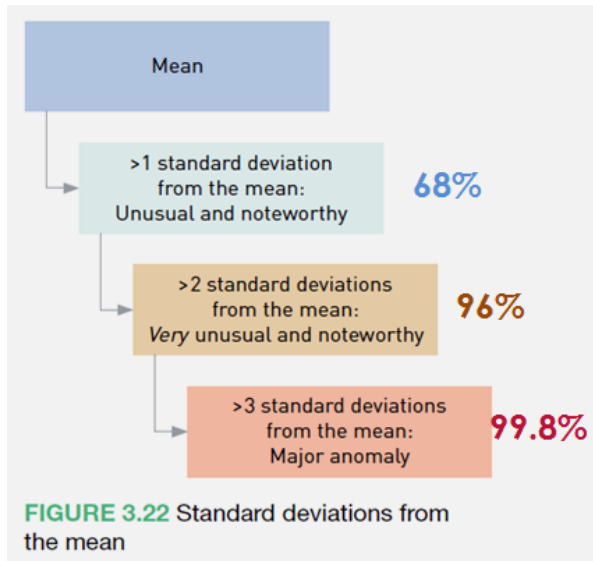
And Locations with abnormally high numbers of outlets for their Population Size:

	Suburbs / Towns	Population (2011 Census)	Mc Donalds	KFC	Hungry J's	Subway	TOTALS
13	Dandenong	25923	3	2	1	6	12
14	Frankston	21583	3	3	1	5	12
15	Cranbourne (NSEW)	56862	3	1	0	5	9
16	Werribee	40218	3	2	1	3	9
17	Epping	26855	2	2	1	4	9
18	Berwick(Nth and Sth)	45237	3	1	1	3	8
19	Ringwood	16009	1	2	1	4	8
20	Narre Warren (Nth and Sth)	54476	2	2	0	3	7
21	Footscray	13203	1	1	1	3	6
22	Mornington	22421	1	1	1	2	5
23	Hoppers Crossing	37598	1	1	0	2	4

Although we can usually Sort and Filter our data and then visually identify these “outliers”, we also use the mathematics of the Standard Deviation to confirm that the values are in fact statistical “outliers”.

The Standard Deviation combined with the Mean tells us the normally expected spread of values. (See Textbook pages 154 to 156).





## Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT



This means that there is a 96% certainty that the number of Fast Food Outlets for a particular sized suburb should lie between 2 standard deviations of the Mean.

We can use Standard Deviation to check the Outlet values we have determined, and see if statistically any of these are abnormally High or abnormally low.

The Standard Deviation check can also reveal that we have sample sizes which are too small.

		   					
28	Suburbs / Towns	Population (2011 Census)	Mc Donalds	KFC	Hungry J's	Subway	TOTALS
29	Werribee	40000	3	2	1	3	9
30	Berwick(Nth and Sth)	40000	3	1	1	3	8
31	Hoppers Crossing	40000	1	1	0	2	4
32	<b>MEAN</b> =AVERAGE( : )	40000	2.3	1.3	0.7	2.7	7.0
33	<b>STD DEVIATION</b> =STDDEV( : )	0	1.2	0.6	0.6	0.6	2.6

If we look at the Fast Food Suburbs with a population of 40000 people then the Mean of 7.0 and SD of 2.6 tells us the expected 2SD 96% certainty range is:

## Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

Total 96% Expected Outlets =  $(7.0 - 2 \times 2.6 \text{ to } 7.0 + (2 \times 2.6)) = 2 \text{ outlets to } 12 \text{ outlets}$

**This indicates that the 96% statistical certainty normal range for the number of outlets in a 40000 Town is 2 to 12 outlets.**

This is a huge spread in outlets, due to our small sample size of only three suburbs with 4 to 9 outlets in them.

We cannot make reliable predictions based on only data for these three widely varying suburbs.

We also used Standard Deviation checking to work out if the abnormally high and low suburbs which we identified by sorting, are in fact statistically abnormal:

45	Suburbs / Towns	Population (2011 Census)	Mc Donalds	KFC	Hungry J's	Subway	TOTALS
46	Frankston	20000	3	3	1	5	12
47	Ringwood	20000	1	2	1	4	8
48	Mornington	20000	1	1	1	2	5
49	Brunswick	20000	2	0	0	2	4
50	St Kilda	20000	1	0	0	3	4
51	Brighton	20000	0	1	0	3	4
52	Cheltenham	20000	1	1	0	2	4
53	Carlton	20000	0	1	0	2	3
54	Hawthorn	20000	1	0	0	2	3
55	Rosebud	20000	1	1	0	1	3
56	South Yarra	20000	1	0	0	1	2
57	Hampton	20000	0	0	0	2	2
58	Mordialloc	20000	0	0	0	1	1
59	Northcote	20000	0	0	0	0	0
60	Fitzroy (and Nth)	20000	0	0	0	0	0
61	Lalor	20000	0	0	0	0	0
62	MEAN =AVERAGE(____)	20000	0.8	0.6	0.2	1.9	3.4
63	STD DEVIATION =STDDEV(____)	0	0.9	0.9	0.4	1.4	3.1

For the 20000 suburbs the 2SD 96% expected range is  $(3.4 - 2 \times 3.1) \text{ to } 3.4 + 2 \times 3.1$  or **-2.8 to 10 outlets.**

This means the 12 for Frankston is abnormally high, but Ringwood, Northcote, Lalor, and Fitzroy are all within the 96% expected 2 SD possibility range.

**Frankston can be removed from our data as an “outlier” but the other suburbs highlighted in orange cannot.**

## Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

**Frankston has a statistically abnormally high number of fast food outlets for the following reasons:**

**It has a lot of extra temporary visitors to the Suburb who are likely to eat fast food because of:**

- **Tourist destination with beaches and a long pier as well as amusement arcades**
- **Has a very busy public boat ramp for fishing and jet ski launching**
- **Has a busy train station that is the main station for the Mornington Peninsula**
- **has the only large shopping complex with Movie theatres on the Mornington Peninsula**
- **Has many more large entertainment venues (Pubs and Nightclubs) than any nearby suburbs**

As mentioned previously, Northcote and Fitzroy are both inner city suburbs heavily populated by university students and young urban professionals, eg. Very few young Families, High School students, or Tradies. There is also not large enough land to build outlets like McDonald's with parking Lots and Drive Throughs. This is why they have zero Fast Food outlets in them.

Ringwood has a high number of outlets compared to other suburbs of 20000 people.

This happens because of similar reasons to Frankston:

- Ringwood has the large Eastland regional shopping complex with movie theatres where people from out of town come to shop and get entertainment
- Ringwood has a large and busy train station

# Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

## 9) Scatter Plots and Correlations

We copied the Rounded off Populations, (with the Zero Outlets Suburbs Removed), and only looking at Total Outlets per Location, as per our Hypothesis.

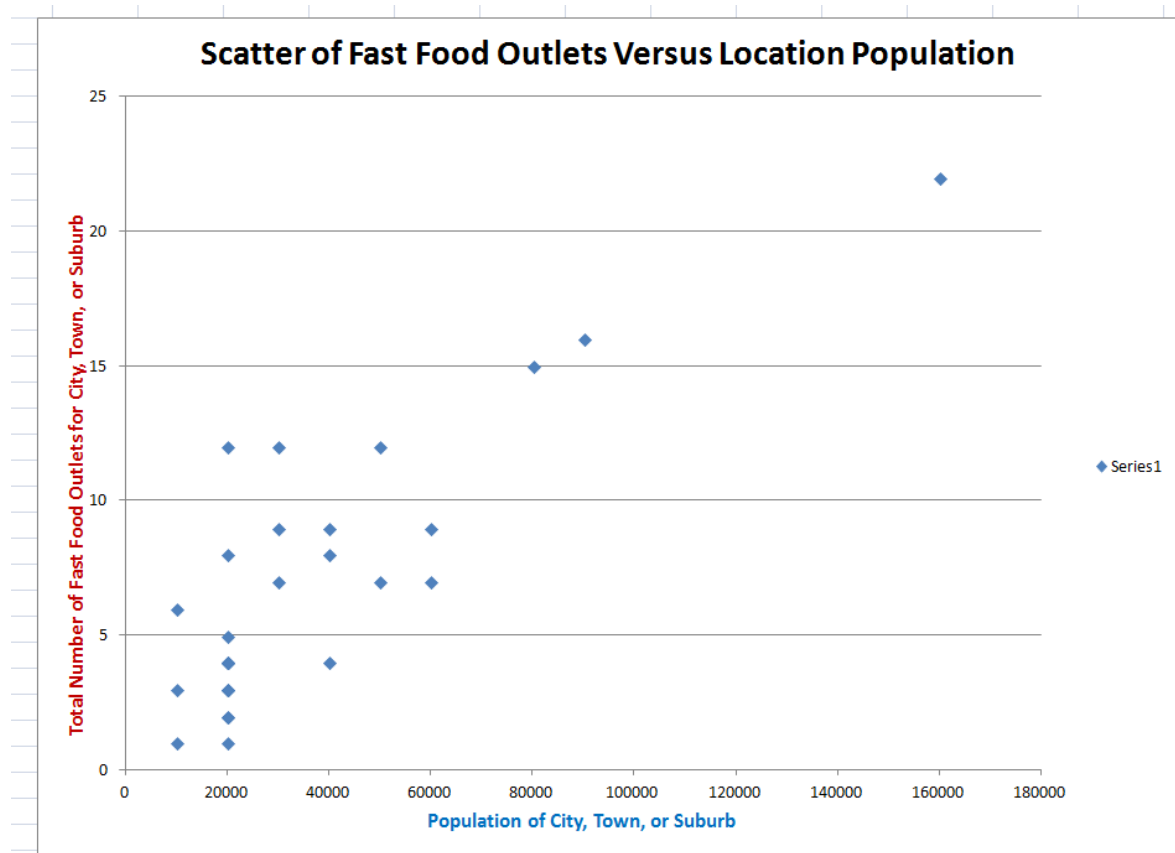
We had to Manually type in our Totals as numbers into the Scatter Plot xls sheet, by having a snipping tool image from our other sheet temporarily copied and pasted onto our Scatter Plot sheet. Here is the set of input data for the Total Outlets Correlation.

	A	B	C	D	E
1	<b>U302 SAT - Fast Food Outlets - Scatter Plot Diagram</b>				
2	<b>(Populations Rounded off to nearest 10 000)</b>				
3					<b>Total</b>
4	<b>Regional Cities</b>	<b>Population (2011 Census)</b>			<b>Outlets</b>
5	Geelong	160000			22
6	Ballarat	90000			16
7	Bendigo	80000			15
8	Mildura	50000			7
9	Shepparton	50000			12
10	Wodonga	30000			7
11	Cranbourne (NSEW)	60000			9
12	Narre Warren (Nth and Sth)	60000			7
13	Berwick(Nth and Sth)	40000			8
14	Werribee	40000			9
15	Hoppers Crossing	40000			4
16	Epping	30000			9
17	Dandenong	30000			12
18	Brunswick	20000			4
19	St Kilda	20000			4
20	Mornington	20000			5
21	Carlton	20000			3
22	Frankston	20000			12
23	Brighton	20000			4
24	Hawthorn	20000			3
25	South Yarra	20000			2
26	Cheltenham	20000			4
27	Rosebud	20000			3
28	Mordialloc	20000			1
29	Hampton	20000			2
30	Ringwood	20000			8
31	Warragul	10000			3
32	Footscray	10000			6
33	Toorak	10000			1
34	Elsternwick	10000			2

## Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

We then had to hold down Ctrl key, and only mouse highlight the number values (not Headings) in our XLS like this:

	A	B	C	D	E
1	U302 SAT - Fast Food Outlets - Scatter Plot Diagram				
2	(Populations Rounded off to nearest 10 000)				
3					
4	Regional Cities	Population (2011 Census)			Total Outlets
5	Geelong	160000			22
6	Ballarat	90000			16
7	Bendigo	80000			15
8	Mildura	50000			7
9	Shepparton	50000			12
10	Wodonga	30000			7
11	Cranbourne (NSEW)	60000			9
12	Narre Warren (Nth and Sth)	60000			7
13	Berwick(Nth and Sth)	40000			8
14	Werribee	40000			9
15	Hoppers Crossing	40000			4
16	Epping	30000			9
17	Dandenong	30000			12
18	Brunswick	20000			4
19	St Kilda	20000			4
20	Mornington	20000			5
21	Carlton	20000			3
22	Frankston	20000			12
23	Brighton	20000			4
24	Hawthorn	20000			3
25	South Yarra	20000			2
26	Cheltenham	20000			4
27	Rosebud	20000			3
28	Mordialloc	20000			1
29	Hampton	20000			2
30	Ringwood	20000			8
31	Warragul	10000			3
32	Footscray	10000			6
33	Toorak	10000			1
34	Elsternwick	10000			2



And then do Insert Tab > Charts > Scatter Plot and choose the first top option of just dots, to make the initial Scatter Chart. Then we had to click on the chart and go to the Layout Tab and Do Chart Title, Axes, etc to get our chart fully labelled.

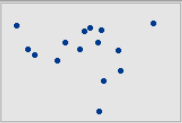
The following Video shows how to do Scatter Plot, as well as Correlation Coefficient: <https://www.youtube.com/watch?v=s2TVkYmmCAs>

## Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

Calculating the Correlation Coefficient Value is done from the 1 minute mark onwards of the previously mentioned video: eg. Find an empty cell on your spreadsheet, click the just below the ribbon fx button, search for the CORREL function, click Go, then mouse select and enter each of the dataset ranges separately then click OK.

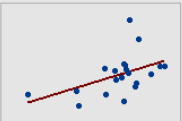
**We did this for our previous Suburbs and Total Outlets Spreadsheet, and Correlation = 0.83**  
which indicates only a moderately strong positive relationship.

### Interpreting Correlation Coefficient Value



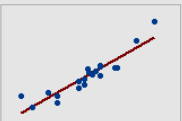
No relationship: Pearson  $r = 0$

The points fall randomly on the plot, which indicates that there is no linear relationship between the variables.



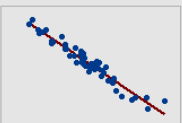
Moderate positive relationship: Pearson  $r = 0.476$

Some points are close to the line but other points are far from it, which indicates only a moderate linear relationship between the variables.



Large positive relationship: Pearson  $r = 0.93$

The points fall close to the line, which indicates that there is a strong linear relationship between the variables. The relationship is positive because as one variable increases, the other variable also increases.



Large negative relationship: Pearson  $r = -0.968$

The points fall close to the line, which indicates that there is a strong negative relationship between the variables. The relationship is negative because, as one variable increases, the other variable decreases.

If your Correlation Coefficient comes out lower than 0.4 then there is not really an identifiable relationship between the two sets of data.

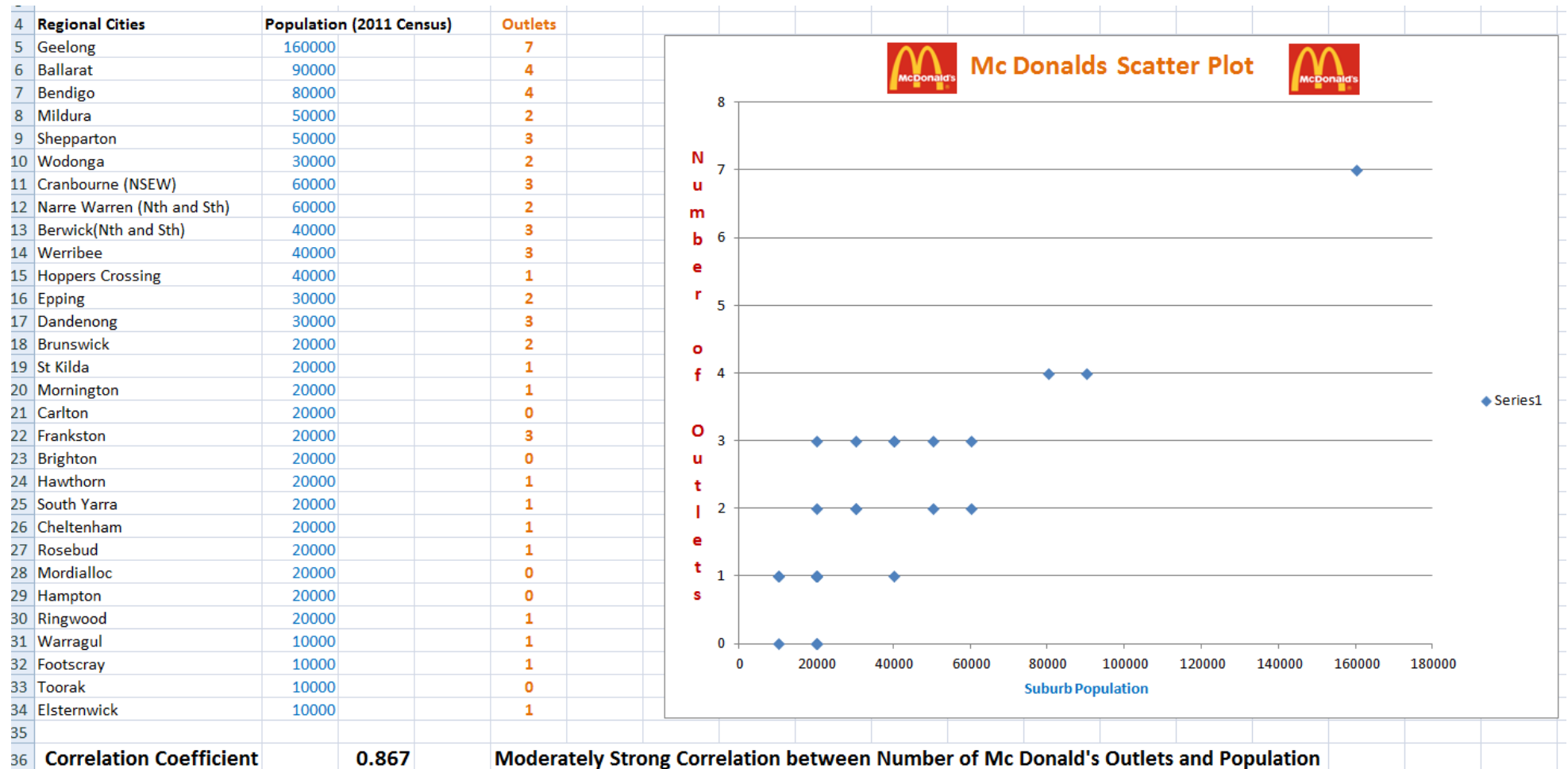
If your Correlation Coefficient comes out higher than 0.9 then there is a very strong mathematical relationship between the two sets of data.



## Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

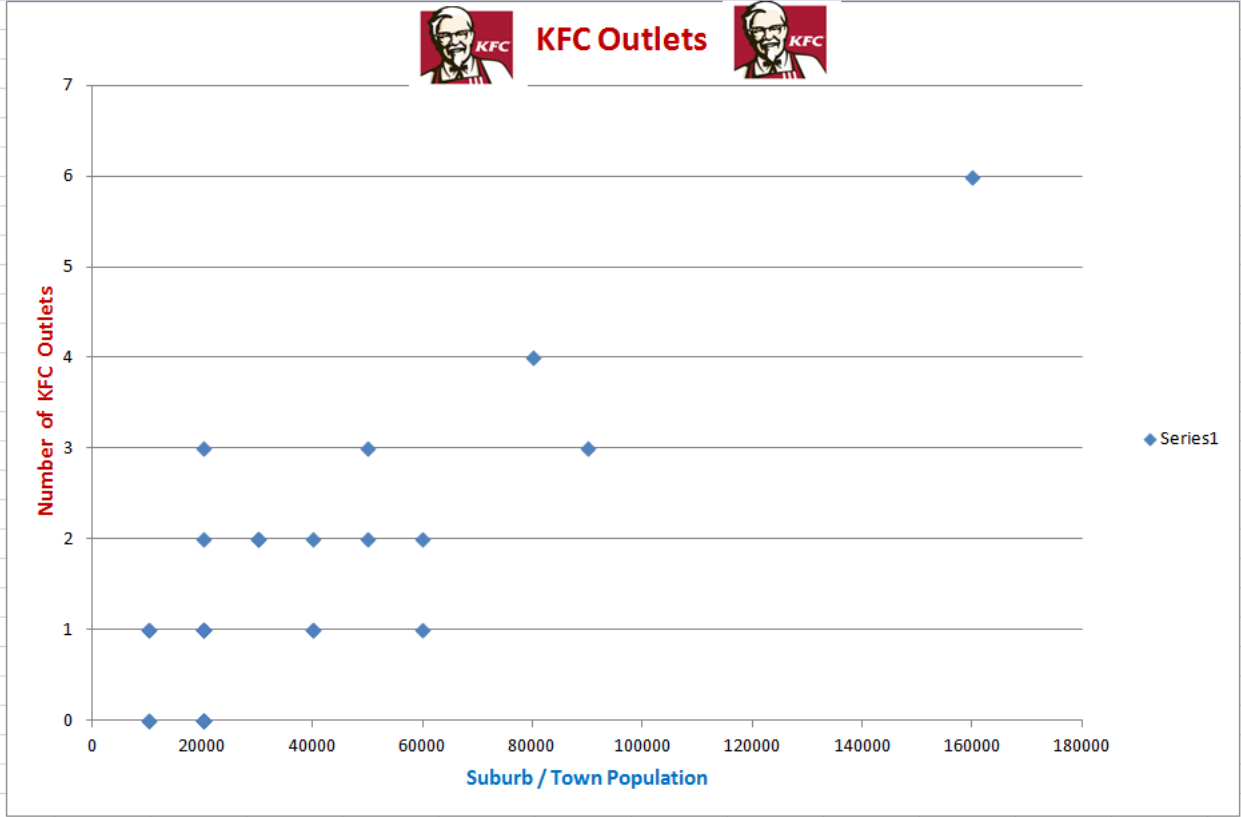
On our Spreadsheet, we then made Tabs and did separate Scatter Plots and Correlation Coefficients for each of McDonalds, KFC, Hungry Jacks, Subway.

MAKE SURE YOUR DATA IS FORMATTED AS NUMBER, OR EXCEL WILL JUST DO A STACK OF DOTS VERTICALLY AT 1.



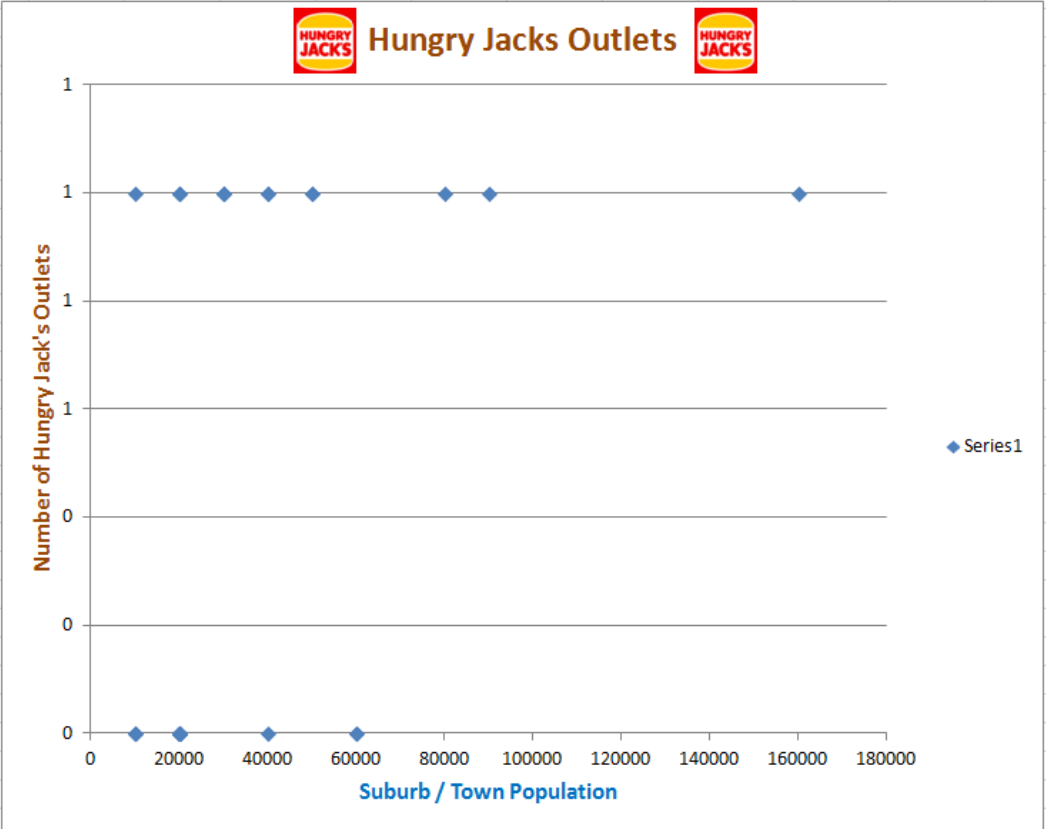
# Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
4	<b>Regional Cities</b>	<b>Population (2011 Census)</b>			<b>Outlets</b>														
5	Geelong	160000			6														
6	Ballarat	90000			3														
7	Bendigo	80000			4														
8	Mildura	50000			3														
9	Shepparton	50000			2														
10	Wodonga	30000			2														
11	Cranbourne (NSEW)	60000			1														
12	Narre Warren (Nth and Sth)	60000			2														
13	Berwick(Nth and Sth)	40000			1														
14	Werribee	40000			2														
15	Hoppers Crossing	40000			1														
16	Epping	30000			2														
17	Dandenong	30000			2														
18	Brunswick	20000			0														
19	St Kilda	20000			0														
20	Mornington	20000			1														
21	Carlton	20000			1														
22	Frankston	20000			3														
23	Brighton	20000			1														
24	Hawthorn	20000			0														
25	South Yarra	20000			0														
26	Cheltenham	20000			1														
27	Rosebud	20000			1														
28	Mordialloc	20000			0														
29	Hampton	20000			0														
30	Ringwood	20000			2														
31	Warragul	10000			1														
32	Footscray	10000			1														
33	Toorak	10000			0														
34	Elsternwick	10000			0														
35																			
36	<b>Correlation Coefficient</b>			<b>0.816</b>		<b>Moderately Strong Correlation between Number of KFC Outlets and Population</b>													



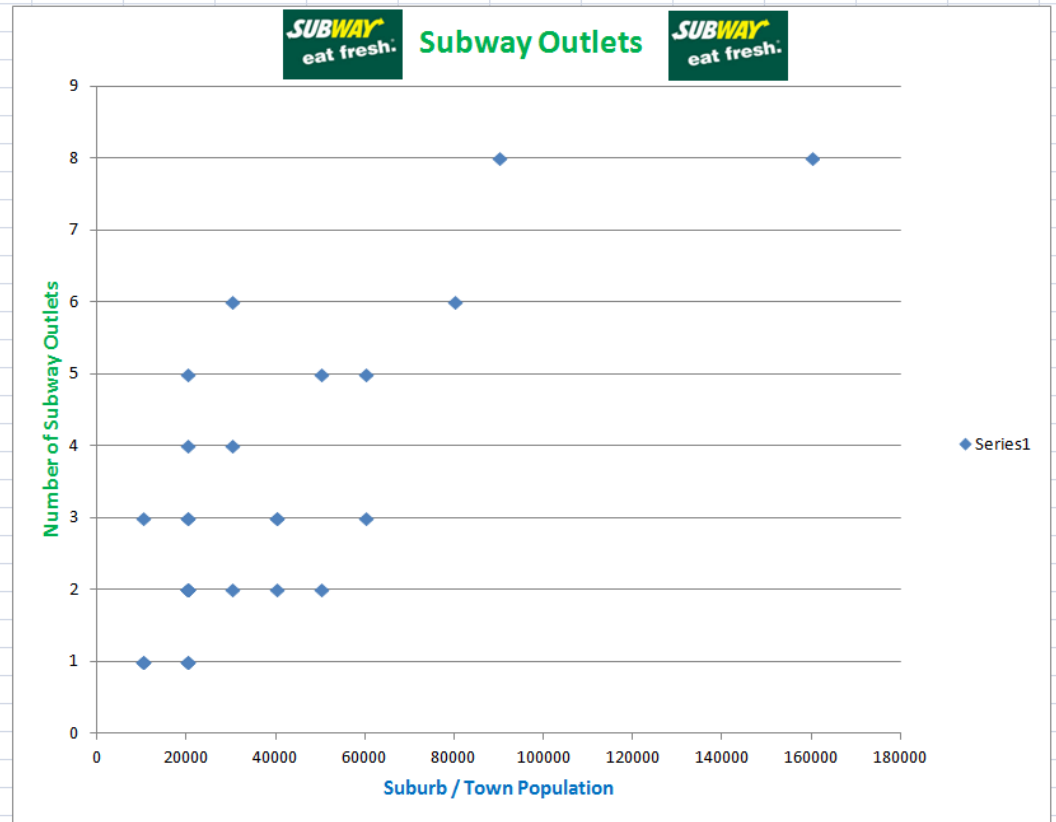
# Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
1	<b>U302 SAT - Fast Food Outlets - Scatter Plot Diagram - Hungry Jacks</b>																			
2	<b>(Populations Rounded off to nearest 10 000 - Zero Outlets Suburbs have been removed)</b>																			
3																				
4	<b>Regional Cities</b>	<b>Population (2011 Census)</b>																		<b>Outlets</b>
5	Geelong	160000																		1
6	Ballarat	90000																		1
7	Bendigo	80000																		1
8	Mildura	50000																		1
9	Shepparton	50000																		1
10	Wodonga	30000																		1
11	Cranbourne (NSEW)	60000																		0
12	Narre Warren (Nth and Sth)	60000																		0
13	Berwick(Nth and Sth)	40000																		1
14	Werribee	40000																		1
15	Hoppers Crossing	40000																		0
16	Epping	30000																		1
17	Dandenong	30000																		1
18	Brunswick	20000																		0
19	St Kilda	20000																		0
20	Mornington	20000																		1
21	Carlton	20000																		0
22	Frankston	20000																		1
23	Brighton	20000																		0
24	Hawthorn	20000																		0
25	South Yarra	20000																		0
26	Cheltenham	20000																		0
27	Rosebud	20000																		0
28	Mordialloc	20000																		0
29	Hampton	20000																		0
30	Ringwood	20000																		1
31	Warragul	10000																		0
32	Footscray	10000																		1
33	Toorak	10000																		0
34	Elsternwick	10000																		0
35																				
36	<b>Correlation Coefficient</b>	<b>0.3838</b>																		<b>Very Limited Correlation between Number of Hungry Jack's Outlets and Population</b>



## Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
4	<b>Regional Cities</b>	<b>Population (2011 Census)</b>			<b>Outlets</b>														
5	Geelong	160000			8														
6	Ballarat	90000			8														
7	Bendigo	80000			6														
8	Mildura	50000			2														
9	Shepparton	50000			5														
10	Wodonga	30000			2														
11	Cranbourne (NSEW)	60000			5														
12	Narre Warren (Nth and Sth)	60000			3														
13	Berwick(Nth and Sth)	40000			3														
14	Werribee	40000			3														
15	Hoppers Crossing	40000			2														
16	Epping	30000			4														
17	Dandenong	30000			6														
18	Brunswick	20000			2														
19	St Kilda	20000			3														
20	Mornington	20000			2														
21	Carlton	20000			2														
22	Frankston	20000			5														
23	Brighton	20000			3														
24	Hawthorn	20000			2														
25	South Yarra	20000			1														
26	Cheltenham	20000			2														
27	Rosebud	20000			1														
28	Mordialloc	20000			1														
29	Hampton	20000			2														
30	Ringwood	20000			4														
31	Warragul	10000			1														
32	Footscray	10000			3														
33	Toorak	10000			1														
34	Elsternwick	10000			1														
35																			
36	<b>Correlation Coefficient</b>				<b>0.762</b>			<b>Moderate Correlation between Number of Subway Outlets and Population</b>											



**None of the Correlations show a very strong relationship between the Number of Outlets and the Population of a Town or Suburb.**

**Counting either the Mc Donald's or KFC outlets gives a Moderately strong correlation, counting the Subways does not work as well, and counting the Hungry Jack's outlets is basically a complete waste of time.**

# Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT





## 10) Pivot Tables

These are a bit like doing Database type Queries on an Excel Spreadsheet, which allow instant Filtering and Summarising of Data. To do this we use a pick and click drag and drop type Interface that Excel builds for us within the sheet.

We can extract data for certain categories or groups and examine them individually in detail using a Pivot Table.

The following video explains Pivot Tables really well: <https://www.youtube.com/watch?v=peNTp5fuKfg>

For our Fast Food Outlets, we took our data that was already grouped on Population, and added an Additional Category of “Demographic” which is a fancy word for “people type” and added this as a new column on our Spreadsheet. The Category values that we set up were: Wealthy, Poor, Young Families, Regional Hub, Inner City, and Town.

	A	B	C	D	E	F	G
1	<b>U302 SAT - Fast Food Outlets - Population + Outlets</b>						
2	(Populations Rounded off to nearest 10 000)						
3							
4	<b>Regional Cities</b>	<b>Population (2011 Census)</b>	<b>Category</b>	<b>Mc Donalds</b>	<b>KFC</b>	<b>Hungry J's</b>	<b>Subway</b>
5	Geelong	160000	Regional Hub	7	6	1	8
6	Ballarat	90000	Regional Hub	4	3	1	8
7	Bendigo	80000	Regional Hub	4	4	1	6
8	Mildura	50000	Town	2	2	1	2
9	Shepparton	50000	Regional Hub	3	3	1	5
10	Wodonga	30000	Town	2	2	1	2
11	Cranbourne (NSEW)	60000	Families	3	1	0	5
12	Narre Warren (Nth and Sth)	60000	Families	2	2	0	3
13	Berwick(Nth and Sth)	40000	Families	3	1	1	3
14	Werribee	40000	Families	3	2	1	3
15	Hoppers Crossing	40000	Families	1	1	0	2
16	Epping	30000	Families	2	2	1	4
17	Dandenong	30000	Regional Hub	3	2	1	6
18	Northcote	20000	Inner City	0	0	0	0
19	Brunswick	20000	Inner City	2	0	0	2
20	St Kilda	20000	Inner City	1	0	0	3
21	Mornington	20000	Families	1	1	1	2
22	Carlton	20000	Inner City	0	1	0	2
23	Frankston	20000	Regional Hub	3	3	1	5
24	Brighton	20000	Wealthy	0	1	0	3
25	Fitzroy (and Nth)	20000	Inner City	0	0	0	0
26	Hawthorn	20000	Wealthy	1	0	0	2
27	South Yarra	20000	Inner City	1	0	0	1
28	Cheltenham	20000	Families	1	1	0	2
29	Lalor	20000	Poor	0	0	0	0
30	Rosebud	20000	Families	1	1	0	1
31	Mordialloc	20000	Families	0	0	0	1
32	Hampton	20000	Wealthy	0	0	0	2
33	Ringwood	20000	Regional Hub	1	2	1	4
34	Warragul	10000	Town	1	1	0	1
35	Footscray	10000	Inner City	1	1	1	3
36	Toorak	10000	Wealthy	0	0	0	1
37	Elsternwick	10000	Wealthy	1	0	0	1
38	Tooradin	1000	Town	0	0	0	0

# Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

Inserted Pivot Table in a new Tab like this: (Filter = Category)

Category	(All)		
<b>Values</b>			
<b>Row Labels</b>	<b>Sum of Population (2011 Census)</b>	<b>Sum of Mc Donalds</b>	
Ballarat	90000	4	
Bendigo	80000	4	
Berwick(Nth and Sth)	40000	3	
Brighton	20000	0	
Brunswick	20000	2	
Carlton	20000	0	
Cheltenham	20000	1	
Cranbourne (NSEW)	60000	3	
Dandenong	30000	3	
Elsternwick	10000	1	
Epping	30000	2	
Fitzroy (and Nth)	20000	0	
Footscray	10000	1	
Frankston	20000	3	
Geelong	160000	7	
Hampton	20000	0	
Hawthorn	20000	1	
Hoppers Crossing	40000	1	
Lalor	20000	0	
Mildura	50000	2	
Mordialloc	20000	0	
Mornington	20000	1	
Narre Warren (Nth and Sth)	60000	2	
Northcote	20000	0	
Ringwood	20000	1	
Rosebud	20000	1	
Shepparton	50000	3	

**PivotTable Field List**

Choose fields to add to report:

- Regional Cities
- Population (2011 Census)
- Category
- Mc Donalds
- KFC
- Hungry J's
- Subway

Drag fields between areas below:

<p><b>Report Filter</b></p> <p>Category</p>	<p><b>Column Labels</b></p> <p>Σ Values</p>
<p><b>Row Labels</b></p> <p>Regional Cities</p>	<p><b>Values</b></p> <p>Sum of Population (...)</p> <p>Sum of Mc Donalds</p>

Defer Layout Update Update

We chose McDonald's because they had one of the better Correlation Coefficients. We then looked at Filtering on Families, Wealthy, and Inner Suburban with results shown below:

## Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

We set up the filter on “Families” and then did a Sort on the resulting Pivot Table on Population:

	A	B	C
1	Category	Families	
2			
3		<b>Values</b>	
4	<b>Row Labels</b>	<b>Sum of Population (2011 Census)</b>	<b>Sum of Mc Donalds</b>
5	Narre Warren (Nth and Sth)	60000	2
6	Cranbourne (NSEW)	60000	3
7	Werribee	40000	3
8	Berwick(Nth and Sth)	40000	3
9	Hoppers Crossing	40000	1
10	Epping	30000	2
11	Cheltenham	20000	1
12	Rosebud	20000	1
13	Mornington	20000	1
14	Mordialloc	20000	0
15	<b>Grand Total</b>	<b>350000</b>	<b>17</b>

In Family type suburbs, the Number of McDonald’s outlets is not a great indicator of Population, because we get so much variation ☹️  
 But all Family type suburbs we investigated have at least one McDonalds, (except for Mordialloc which is on the Bay and big on Fish and Chip Shops).

	A	B	C
1	Category	Wealthy	
2			
3		<b>Values</b>	
4	<b>Row Labels</b>	<b>Sum of Population (2011 Census)</b>	<b>Sum of Mc Donalds</b>
5	Brighton	20000	0
6	Hawthorn	20000	1
7	Hampton	20000	0
8	Toorak	10000	0
9	Elsternwick	10000	1
10	<b>Grand Total</b>	<b>80000</b>	<b>2</b>

Wealthy Suburbs do not have many McDonald’s outlets, (perhaps influential rich people use lawyers to stop the council from approving them?)

## Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

	A	B	C
1	Category	Inner City	
2			
3	<b>Values</b>		
4	<b>Row Labels</b>	<b>Sum of Population (2011 Census)</b>	<b>Sum of Mc Donalds</b>
5	Brunswick	20000	2
6	South Yarra	20000	1
7	Northcote	20000	0
8	Carlton	20000	0
9	St Kilda	20000	1
10	Fitzroy (and Nth)	20000	0
11	Footscray	10000	1
12	<b>Grand Total</b>	<b>130000</b>	<b>5</b>

Inner City areas often do not have a McDonald's, perhaps because it is hard to get enough of the very expensive land to build outlets on.

	A	B	C
1	Category	Regional Hub	
2			
3	<b>Values</b>		
4	<b>Row Labels</b>	<b>Sum of Population (2011 Census)</b>	<b>Sum of Mc Donalds</b>
5	Geelong	160000	7
6	Ballarat	90000	4
7	Bendigo	80000	4
8	Shepparton	50000	3
9	Dandenong	30000	3
10	Ringwood	20000	1
11	Frankston	20000	3
12	<b>Grand Total</b>	<b>450000</b>	<b>25</b>

Regional Hubs with lots of out of town visitors passing through tend to have the most McDonald's outlets.

But once again the number of outlets is not directly related to the permanent population for the town or suburb ☹



## Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

If your Pivot Table Build Menu disappears at any time, then just click on the Pivot Table report, and it should re-appear:

The screenshot shows an Excel spreadsheet with a PivotTable and the PivotTable Field List task pane. The PivotTable is located in the range A3:C12. The task pane is on the right side of the spreadsheet. A blue arrow points from the PivotTable to the task pane, and another blue arrow points from a text box to the task pane.

Row Labels	Sum of Population (2011 Census)	Sum of Mc Donalds
Geelong	160000	7
Ballarat	90000	4
Bendigo	80000	4
Shepparton	50000	3
Dandenong	30000	3
Ringwood	20000	1
Frankston	20000	3
<b>Grand Total</b>	<b>450000</b>	<b>25</b>

Click anywhere in the generated report to make The field list menu re-appear.

PivotTable Field List

Choose fields to add to report:

- Regional Cities
- Population (2011 Census)
- Category
- Mc Donalds
- KFC
- Hungry J's
- Subway

Drag fields between areas below:

Report Filter: Category

Column Labels: Σ Values

Row Labels: Regional Cities

Values: Σ Sum of Population (...), Σ Sum of Mc Donalds

Defer Layout Update:  Update

## Data Manipulation Report for Internet Data – Fast Food Outlets Example SAT

This completes all of our Data Manipulations in Excel for the Fast Food Outlets SAT. 😊

To get maximum marks for the Data Manipulation part of the SAT, you need to demonstrate use of all of the Excel processes we have used for the Fast Foods SAT including:

Grouping  
Totals  
Sorting  
Filtering  
Mean Median Mode  
Standard Deviation  
Pivot Tables

Doing the Data Manipulations is very time consuming, but keep in mind that it is a very important part of the SAT.