Data wrangling with Tableau and Excel

October 11 2016
JRNL 520H

What is data wrangling?

Data wrangling is the process of preparing raw data for use in a data analysis or visualization software.

What are the causes of dirty data?

- Data entry error

What are the causes of dirty data?

- Data entry error
- Incompatible tables

What are the causes of dirty data?

- Data entry error
- Incompatible tables
- Incompatible table format

What should we look out for when cleaning data?

- Table formating

What should we look out for when cleaning data?

- Table formating
- Variable type

What should we look out for when cleaning data?

- Table formating
- Variable type
- Invalid character values

What should we look out for when cleaning data?

- Table formating
- Variable type
- Invalid character values
- Invalid numeric values

What should we look out for when cleaning data?

- Table formating
- Variable type
- Invalid character values
- Invalid numeric values
- Grouping data

What should we look out for when cleaning data?

- Table formating
- Variable type
- Invalid character values
- Invalid numeric values
- Grouping data
- Missing values

Ideal format of data in Tableau

1. Start your data in cell A1. Remove all introductory information and footnotes.
2. Have the first row be the column headers/variable names
3. Have every subsequent row be one observation. No cross-tabulation!

Ideal format of data in Tableau

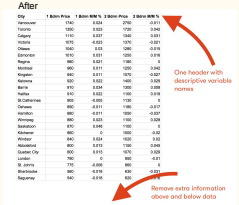
Before

After

Ideal format of data in Tableau

Before

After



Data Interpreter

Tableau's Data Interpreter feature draws out sub-tables and removes some of that extraneous information to help prepare your data source for analysis. Note: the data interpreter only works with Microsoft Excel files, not CSV or other file types.

Data Interpreter

Tableau's Data Interpreter feature draws out sub-tables and removes some of that extraneous information to help prepare your data source for analysis. Note: the data interpreter only works with Microsoft Excel files, not CSV or other file types.

Complete Tableau exercise

Joins

A JOIN is a means for combining columns from one or more tables by using values common to each. There are four main join types: inner, left, right and full outer.

Joins

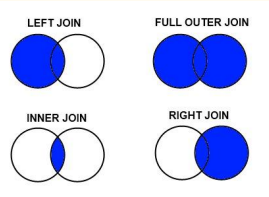
Market rental pricing for July 2016

	1 Bedroom		2 Bedrooms	
City	Price	M/M %	Price	M/M %
Vancouver	1740	0.024	2750	-0.011
Toronto	1350	0.023	1720	0.042
Calgary	1110	0.037	1340	0.031
Victoria	1075	-0.023	1370	-0.021
Ottawa	1040	0.03	1280	-0.015
Edmonton	1010	0.031	1250	0.016
Regina	980	0.021	1180	0
Montreal	960	0.011	1250	0.042
Kingston	940	0.011	1070	-0.027
Kelowna	920	0.022	1400	0.029
Barrie	910	0.034	1300	0.008
Halifax	910	0.022	1100	0.019

Market rental pricing for June 2016

	1 Bedroom		2 Bedrooms	
City	Price	M/M %	Price	M/M %
Vancouver	$1,700	-2.30%	$2,780	-0.70%
Toronto	$1,320	1.50%	$1,650	-1.20%
Calgary	$1,100	1.90%	$1,400	2.90%
Victoria	$1,070	-2.70%	$1,300	-3.70%
Ottawa	$1,010	-1.90%	$1,300	-1.50%
Edmonton	$980	3.20%	$1,230	2.50%
Regina	$960	-4.00%	$1,180	0.00%
Montreal	$950	2.20%	$1,200	-4.00%
Kingston	$930	0.00%	$1,100	4.80%
Kelowna	$910	-4.20%	$1,130	1.80%
Barrie	$900	-2.20%	$1,200	1.70%
Halifax	$900	1.10%	$1,360	1.50%

Joins



Joins

Market rental pricing for July 2016

	1 Bedroom		2 Bedrooms	
City	Price	M/M %	Price	M/M %
Vancouver	1740	0.024	2750	-0.011
Toronto	1350	0.023	1720	0.042
Calgary	1110	0.037	1340	0.031
Victoria	1075	-0.023	1370	-0.021
Ottawa	1040	0.03	1280	-0.015
Edmonton	1010	0.031	1250	0.016
Regina	980	0.021	1180	0
Montreal	960	0.011	1250	0.042
Kingston	940	0.011	1070	-0.027
Kelowna	920	0.022	1400	0.029
Barrie	910	0.034	1300	0.008
Halifax	910	0.022	1100	0.019

Market rental pricing for June 2016

	1 Bedroom		2 Bedrooms	
City	Price	M/M %	Price	M/M %
Vancouver	$1,700	-2.30%	$2,780	-0.70%
Toronto	$1,320	1.50%	$1,650	-1.20%
Calgary	$1,100	1.90%	$1,400	2.90%
Victoria	$1,070	-2.70%	$1,300	-3.70%
Ottawa	$1,010	-1.90%	$1,300	-1.50%
Edmonton	$980	3.20%	$1,230	2.50%
Regina	$960	-4.00%	$1,180	0.00%
Montreal	$950	2.20%	$1,200	-4.00%
Kingston	$930	0.00%	$1,100	4.80%
Kelowna	$910	-4.20%	$1,130	1.80%
Barrie	$900	-2.20%	$1,200	1.70%
Halifax	$900	1.10%	$1,360	1.50%
St Catherines	$890	-3.30%	$1,090	3.80%

Joins

Market rental pricing for July 2016

	1 Bedroom		2 Bedrooms	
City	Price	M/M %	Price	M/M %
Vancouver	1740	0.024	2750	-0.011
Toronto	1350	0.023	1720	0.042
Calgary	1110	0.037	1340	0.031
Victoria	1075	-0.023	1370	-0.021
Ottawa	1040	0.03	1280	-0.015
Edmonton	1010	0.031	1250	0.016
Regina	980	0.021	1180	0
Montreal	960	0.011	1250	0.042
Kingston	940	0.011	1070	-0.027
Kelowna	920	0.022	1400	0.029
Barrie	910	0.034	1300	0.008
Halifax	910	0.022	1100	0.019

Market rental pricing for June 2016

	1 Bedroom		2 Bedrooms	
City	Price	M/M %	Price	M/M %
Vancouver	$1,700	-2.30%	$2,780	-0.70%
Toronto	$1,320	1.50%	$1,650	-1.20%
Calgary	$1,100	1.90%	$1,400	2.90%
Victoria	$1,070	-2.70%	$1,300	-3.70%
Ottawa	$1,010	-1.90%	$1,300	-1.50%
Edmonton	$980	3.20%	$1,230	2.50%
Regina	$960	-4.00%	$1,180	0.00%
Montreal	$950	2.20%	$1,200	-4.00%
Kingston	$930	0.00%	$1,100	4.80%
Kelowna	$910	-4.20%	$1,130	1.80%
Barrie	$900	-2.20%	$1,200	1.70%
Halifax	$900	1.10%	$1,360	1.50%
St Catherines	$890	-3.30%	$1,090	3.80%

Complete Tableau exercise

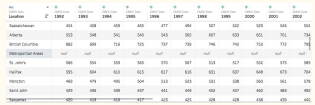
Wrangling in Excel

Sometimes the data interpreter in Tableau isn't able to detect all of the errors in the dataset. In cases like this, you will need to manually clean the data in Tableau.

Complete Tableau exercise

Pivot

Tabular format



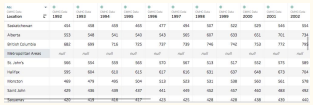
Columnar format

Location	Year	Rent
Canada2	1992	568
Provinces	1992	null
Newfoundland and L...	1992	517
Prince Edward Island	1992	503
Nova Scotia	1992	562
New Brunswick	1992	461
Quebec	1992	473
Ontario	1992	672
Manitoba	1992	545
Saskatchewan	1992	454
Alberta	1992	553
British Columbia	1992	682
Metropolitan Areas	1992	null

Complete Tableau exercise

Pivot

Tabular format



Columnar format

Location	Year	Rent
Canada2	1992	568
Provinces	1992	null
Newfoundland and L...	1992	517
Prince Edward Island	1992	503
Nova Scotia	1992	562
New Brunswick	1992	461
Quebec	1992	473
Ontario	1992	672
Manitoba	1992	545
Saskatchewan	1992	454
Alberta	1992	553
British Columbia	1992	682
Metropolitan Areas	1992	null