







Australian Clothing Retail Expansion: Data Cleaning Steps


Step 1: Raw data load

- Raw data loaded into Tableau with a row filter, so we only see data in date range 2010-2015

 Data1 F1	 Data1 Turnover : New South Wales : ...	 Data1 Turnover : New South W...	 Data1 Turnover : New South W...	 Data1 Turnover : New South W...	 Data1 Turnover : New South W...	 Data1 Turnover : New South W...
2010-01-01	2,053.500	233.700	255.400	2,542.60	312.000	491.900
2010-02-...	1,817.400	199.000	215.500	2,231.90	273.000	392.700
2010-03-...	2,017.700	222.600	229.600	2,469.80	288.800	428.700
2010-04-...	1,951.100	222.200	214.400	2,387.70	263.500	432.600
2010-05-...	1,988.600	213.400	224.500	2,426.50	289.600	465.400
2010-06-...	1,888.200	206.300	205.600	2,300.10	309.200	516.300
2010-07-01	2,035.800	206.700	198.700	2,441.20	302.500	476.500
2010-08-...	2,012.800	212.900	203.300	2,429.00	300.000	472.500
2010-09-...	1,984.200	224.300	203.200	2,411.60	284.800	437.800
2010-10-01	2,075.900	238.700	214.400	2,529.00	309.300	461.800
2010-11-01	2,159.700	251.200	201.400	2,612.40	315.500	490.400
2010-12-01	2,360.100	380.000	251.800	2,991.90	357.600	720.900
2011-01-01	2,157.300	250.000	205.800	2,613.10	318.000	439.100
2011-02-01	1,953.800	216.800	194.600	2,365.20	269.000	384.100
2011-03-01	2,108.100	240.500	200.400	2,549.10	282.900	418.300
2011-04-01	2,062.300	244.400	196.200	2,502.80	283.900	394.400

Step 2: Wide to long format. Pivot, and split.

- Data is currently in wide format, but it will be turned into long format using tableau pivot to make it easier for machine analysis.

 Data1	Abc Pivot	# Pivot
F1	Pivot Field Names	Pivot Field Values
2010-01-01	Turnover ; Australian Capital Territory ; Cafes, restaurants and catering services ;	27.80
2010-01-01	Turnover ; Australian Capital Territory ; Cafes, restaurants and takeaway food services ;	42.80
2010-01-01	Turnover ; Australian Capital Territory ; Clothing retailing ;	17.10
2010-01-01	Turnover ; Australian Capital Territory ; Clothing, footwear and personal accessory retailing ;	27.30
2010-01-01	Turnover ; Australian Capital Territory ; Department stores ;	29.80
2010-01-01	Turnover ; Australian Capital Territory ; Electrical and electronic goods retailing ;	35.80
2010-01-01	Turnover ; Australian Capital Territory ; Food retailing ;	148.70
2010-01-01	Turnover ; Australian Capital Territory ; Footwear and other personal accessory retailing ;	10.20
2010-01-01	Turnover ; Australian Capital Territory ; Furniture, floor coverings, houseware and textile goods retailing ;	20.00
2010-01-01	Turnover ; Australian Capital Territory ; Hardware, building and garden supplies retailing ;	23.50
2010-01-01	Turnover ; Australian Capital Territory ; Household goods retailing ;	79.30
2010-01-01	Turnover ; Australian Capital Territory ; Liquor retailing ;	11.10
2010-01-01	Turnover ; Australian Capital Territory ; Newspaper and book retailing ;	8.10
2010-01-01	Turnover ; Australian Capital Territory ; Other recreational goods retailing ;	10.80
2010-01-01	Turnover ; Australian Capital Territory ; Other retailing ;	37.80
2010-01-01	Turnover ; Australian Capital Territory ; Other retailing n.e.c. ;	6.90

- Further cleaning needed to split pivot field name column into state, and industry columns
- Similar data cleaning done to population data

Abc Data1 F1	# Pivot Population	≠Abc Calculation Gender	≠Abc Calculation State
Jun-1981	7,474,993	Female	Australia
Jun-1981	113,976	Female	Australian Capital Territory
Jun-1981	2,626,538	Female	New South Wales
Jun-1981	57,223	Female	Northern Territory
Jun-1981	1,166,761	Female	Queensland
Jun-1981	664,829	Female	South Australia
Jun-1981	214,659	Female	Tasmania
Jun-1981	1,988,200	Female	Victoria
Jun-1981	642,807	Female	Western Australia
Jun-1981	7,448,267	Male	Australia
Jun-1981	113,605	Male	Australian Capital Territory
Jun-1981	2,608,351	Male	New South Wales
Jun-1981	65,393	Male	Northern Territory

<div> <div></div> <div>Data1</div> </div> <div>F1</div>	<div> <div></div> <div>Pivot</div> </div> <div>Pivot Field Values</div>	<div> <div></div> <div>Calculation</div> </div> <div>State</div>	<div> <div></div> <div>Calculation</div> </div> <div>Industry</div>
2010-01-01	27.80	Australian Capital Territory	Cafes, restaurants and cateri...
2010-01-01	42.80	Australian Capital Territory	Cafes, restaurants and takea...
2010-01-01	17.10	Australian Capital Territory	Clothing retailing
2010-01-01	27.30	Australian Capital Territory	Clothing, footwear and perso...
2010-01-01	29.80	Australian Capital Territory	Department stores
2010-01-01	35.80	Australian Capital Territory	Electrical and electronic goo...
2010-01-01	148.70	Australian Capital Territory	Food retailing
2010-01-01	10.20	Australian Capital Territory	Footwear and other personal ...
2010-01-01	20.00	Australian Capital Territory	Furniture, floor coverings, ho...
2010-01-01	23.50	Australian Capital Territory	Hardware, building and garde...
2010-01-01	79.30	Australian Capital Territory	Household goods retailing
2010-01-01	11.10	Australian Capital Territory	Liquor retailing
2010-01-01	8.10	Australian Capital Territory	Newspaper and book retailing
2010-01-01	10.80	Australian Capital Territory	Other recreational goods ret...
2010-01-01	37.80	Australian Capital Territory	Other retailing
2010-01-01	6.90	Australian Capital Territory	Other retailing n.e.c.