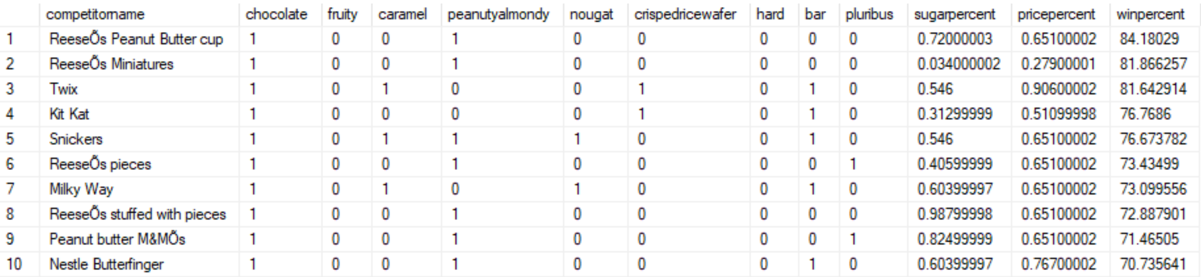
Candy case study

1. At first , I input the Data into MS SQL , and build a Table, to have a quick look.
   I assumed that,chocolate have the most positive correlation with the winpercent, and peanutyalmondy have the second.



	competitorname	chocolate	fruity	caramel	peanutyalmondy	nougat	crispedricewafer	hard	bar	pluribus	sugarpercent	pricepercent	winpercent
1	ReeseÕs Peanut Butter cup	1	0	0	1	0	0	0	0	0	0.72000003	0.65100002	84.18029
2	ReeseÕs Miniatures	1	0	0	1	0	0	0	0	0	0.034000002	0.27900001	81.866257
3	Twix	1	0	1	0	0	1	0	1	0	0.546	0.90600002	81.642914
4	Kit Kat	1	0	0	0	0	1	0	1	0	0.31299999	0.51099998	76.7686
5	Snickers	1	0	1	1	1	0	0	1	0	0.546	0.65100002	76.673782
6	ReeseÕs pieces	1	0	0	1	0	0	0	0	1	0.40599999	0.65100002	73.43499
7	Milky Way	1	0	1	0	1	0	0	1	0	0.60399997	0.65100002	73.099556
8	ReeseÕs stuffed with pieces	1	0	0	1	0	0	0	0	0	0.98799998	0.65100002	72.887901
9	Peanut butter M&MÕs	1	0	0	1	0	0	0	0	1	0.82499999	0.65100002	71.46505
10	Nestle Butterfinger	1	0	0	1	0	0	0	1	0	0.60399997	0.76700002	70.735641

2. Then I used LinearRegression to analyse the Daten, and used AIC model to validation the features. I found the best combination of the features:
   'chocolate','fruity','peanutyalmondy','crispedricewafer','hard','sugarpercent'

3. I found out the chocolate and peanutyalmondy have the biggest coef, and the P value of this to features are smaller than 0.01,just like I assumed.