

Candy Task

By: Kumar Awanish

Dataset

	competitorname	chocolate	fruity	caramel	peanutyalmondy	nougat	crispedricewafer	hard	bar	pluribus	sugarpercent	pricepercent	winpercent
0	100 Grand	1	0	1	0	0	1	0	1	0	0.732	0.860	66.971725
1	3 Musketeers	1	0	0	0	1	0	0	1	0	0.604	0.511	67.602936
2	One dime	0	0	0	0	0	0	0	0	0	0.011	0.116	32.261086
3	One quarter	0	0	0	0	0	0	0	0	0	0.011	0.511	46.116505
4	Air Heads	0	1	0	0	0	0	0	0	0	0.906	0.511	52.341465

12 variables that describe a candy.

9 of them are categorical and the rest, i.e. 3 are numerical variables.

Data cleaning

```
Out[4]: competitorname    0  
         chocolate        0  
         fruity           0  
         caramel          0  
         peanutyalmondy   0  
         nougat           0  
         crispedricewafer 0  
         hard             0  
         bar              0  
         pluribus         0  
         sugarpercent     0  
         pricepercent     0  
         winpercent       0  
         dtype: int64
```

Data is not messy and its clean to use.

No missing values found

Feature Correlation :Heat map



1. Chocolate, Peanutyalmondy, Crispedricwafer, and Bar has the highest correlation among our attributes with winning percentage.
2. Candy with Chocolate and Fruit seems rare.
3. Chocolate has very high influence on winpercent.

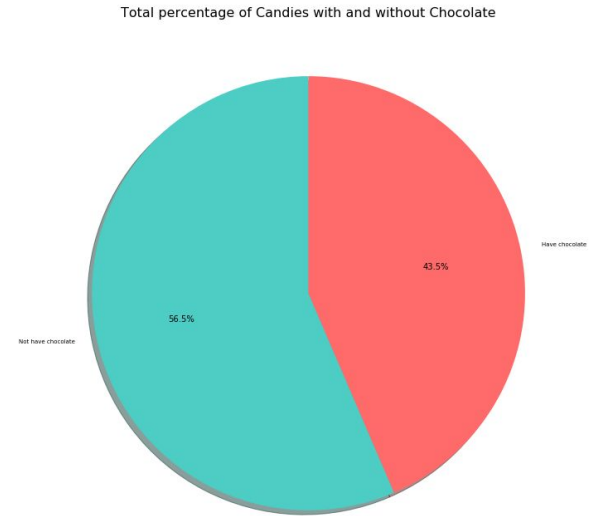
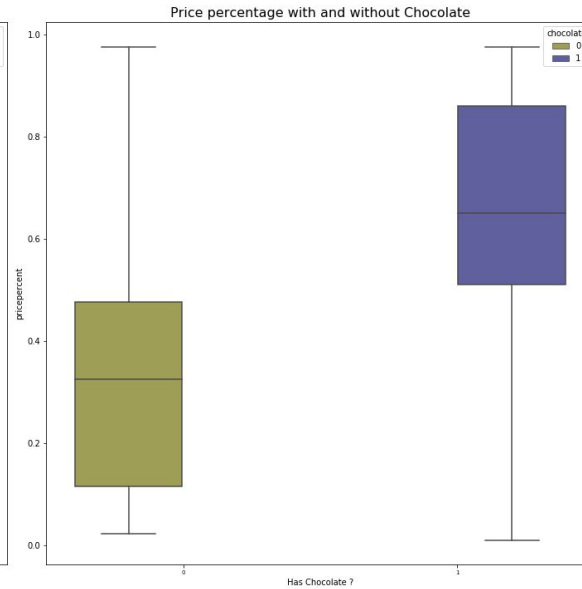
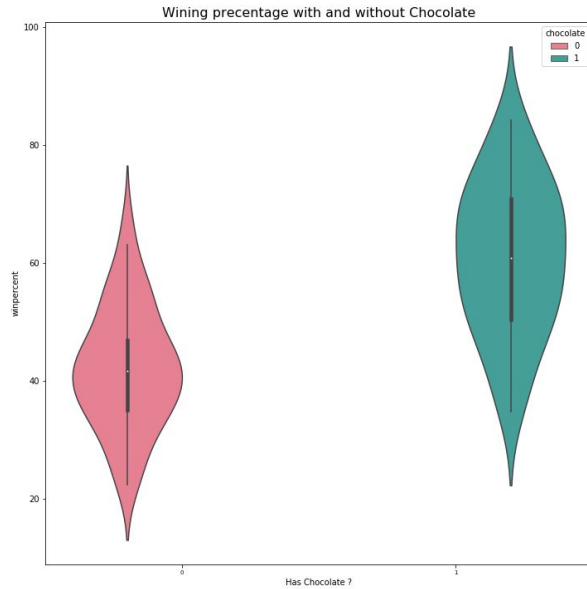
Quick Data insight: Top 10 candies by winpercent

Out[27]:

	competitorname	chocolate	fruity	caramel	peanutyalmondy	nougat	crispedricewafer	hard	bar	pluribus	sugarpercent	pricepercent	winpercent
52	Reese's Peanut Butter cup	1	0	0	1	0	0	0	0	0	0.720	0.651	84.180290
51	Reese's Miniatures	1	0	0	1	0	0	0	0	0	0.034	0.279	81.866257
79	Twix	1	0	1	0	0	1	0	1	0	0.546	0.906	81.642914
28	Kit Kat	1	0	0	0	0	1	0	1	0	0.313	0.511	76.768600
64	Snickers	1	0	1	1	1	0	0	1	0	0.546	0.651	76.673782
53	Reese's pieces	1	0	0	1	0	0	0	0	1	0.406	0.651	73.434990
36	Milky Way	1	0	1	0	1	0	0	1	0	0.604	0.651	73.099556
54	Reese's stuffed with pieces	1	0	0	1	0	0	0	0	0	0.988	0.651	72.887901
32	Peanut butter M&M's	1	0	0	1	0	0	0	0	1	0.825	0.651	71.465050
42	Nestle Butterfinger	1	0	0	1	0	0	0	1	0	0.604	0.767	70.735641

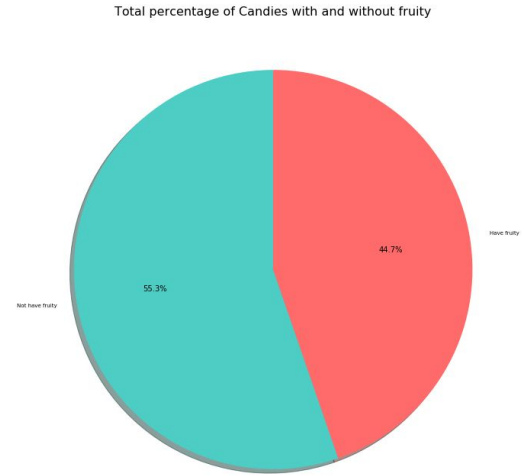
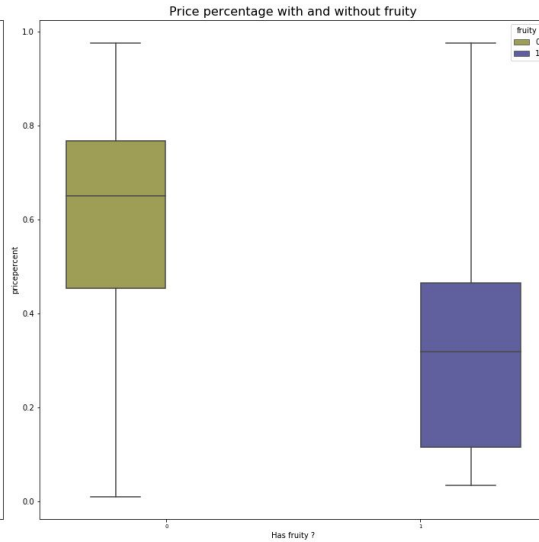
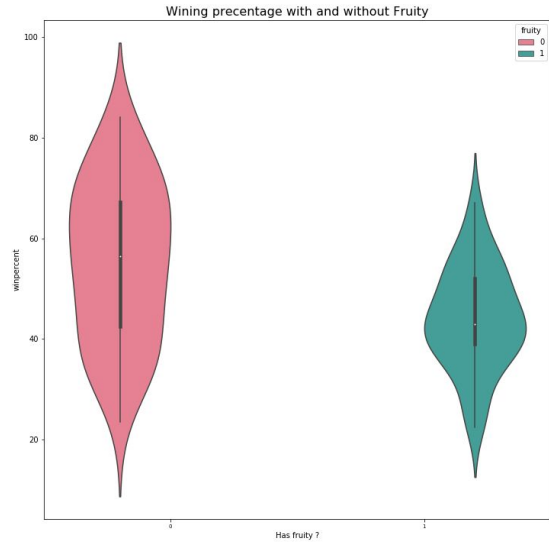
1. Reese's seem to be a favourite.
2. All the top competitors are chocolaty as well.
3. Also, Reese's Miniatures is very cheap when compared to top competitors and overall as well.

Data Visualisation & Descriptive analytics : Chocolate plot insight



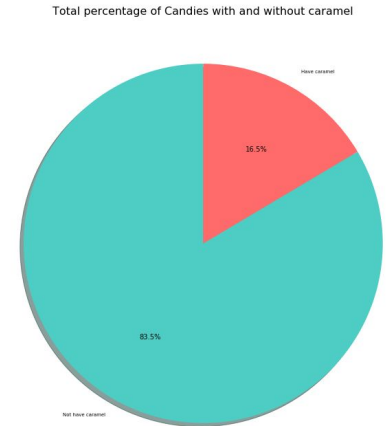
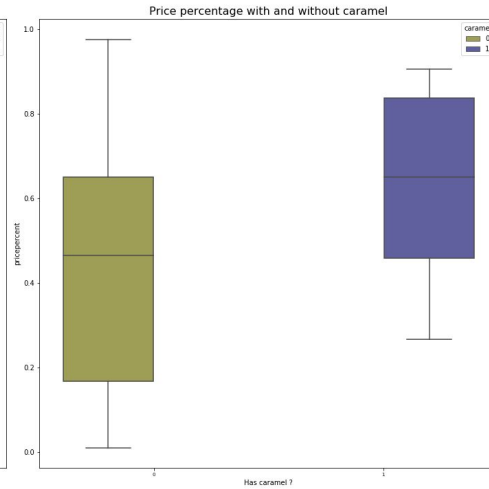
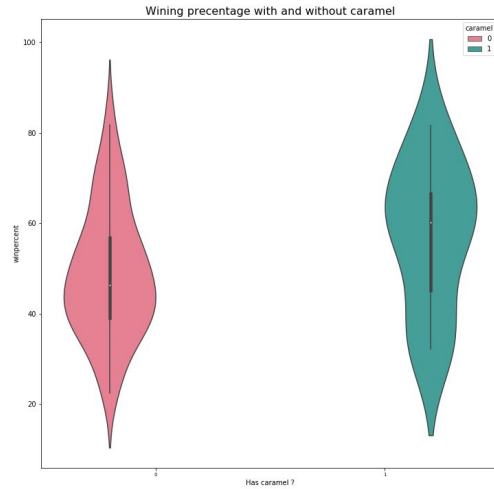
1. Chocolate is very much related to winpercent, it is also in heatmap.
2. Candy with Chocolate content is bit expensive as we can see from the Boxplot.
3. Cany with Chocolate content is in total 43.5%, which tells us manufacturers like to use chocolate for making candies.

Data Visualisation & Descriptive analytics : Fruity plot insight



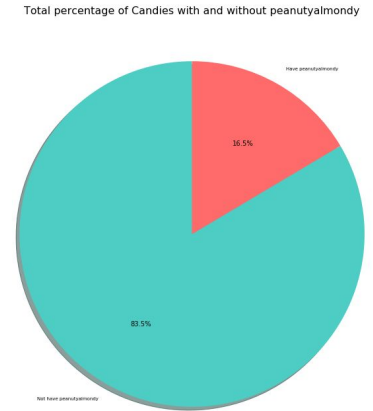
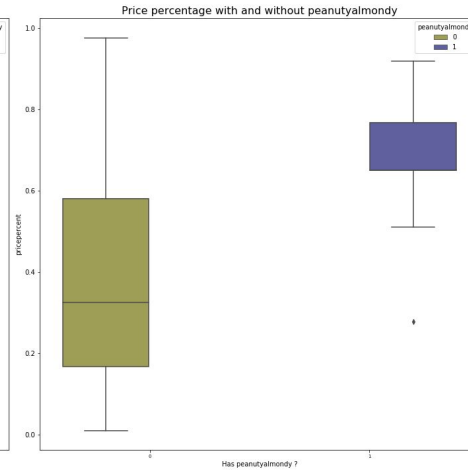
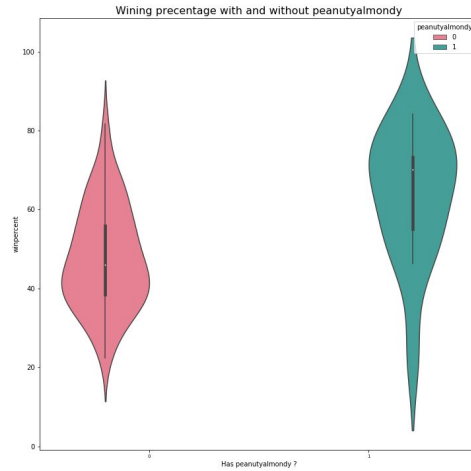
1. Winpercent is not influenced by fruit content.
2. Candy containing fruits are cheaper.
3. Most of the Candy uses fruit(around 44%).

Data Visualisation & Descriptive analytics : Caramel plot insight



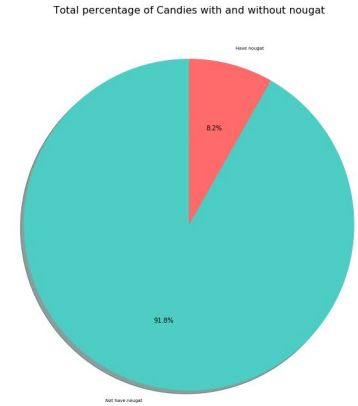
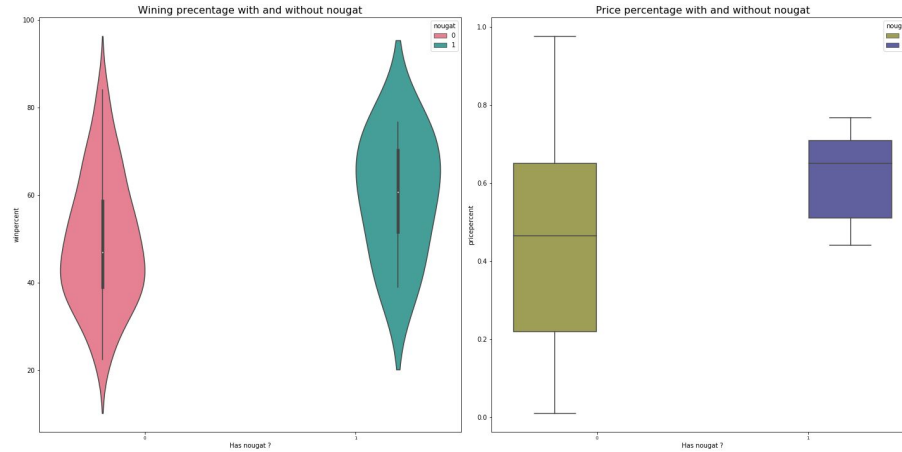
1. Caramel content doesn't have clear influence on winpercent. Its hard to observe any inference from the violin plot.
2. Candy containing Caramel are in the mid-range price bar.
3. Candy makers doesn't favour Caramel, i.e only 16.5% candy contains Caramel.

Data Visualisation & Descriptive analytics : Peanut plot insight



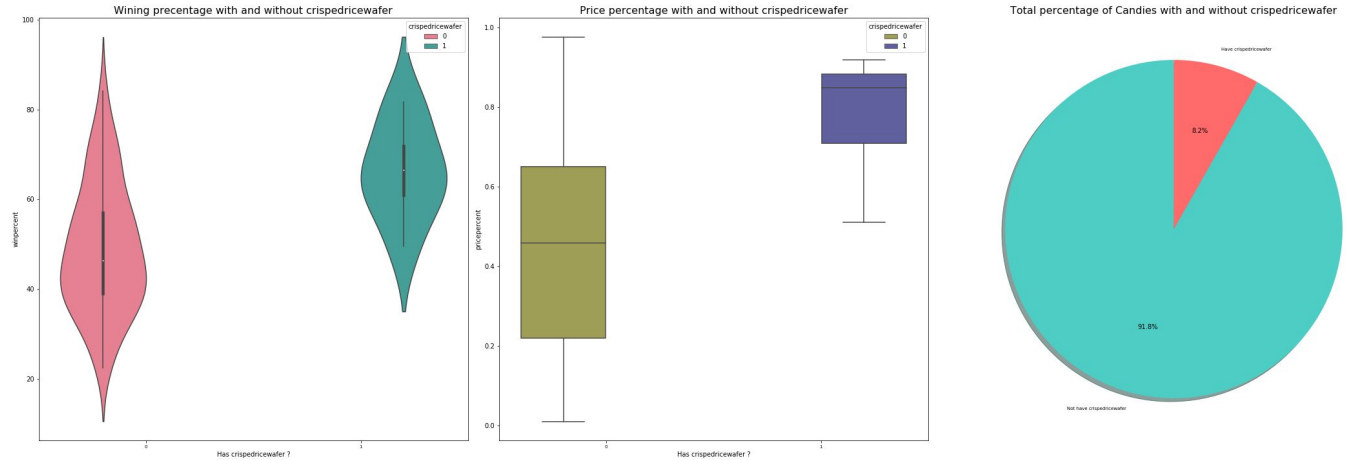
1. Winpercent is influenced by peanut & almond content like Choocolate and it can also be verified in heatMap plot.
2. Peanut & almond content candies are expensive compare to others.
3. From pie chart its evident that makers doesn't have peanut made candy.

Data Visualisation & Descriptive analytics : nougat plot insight



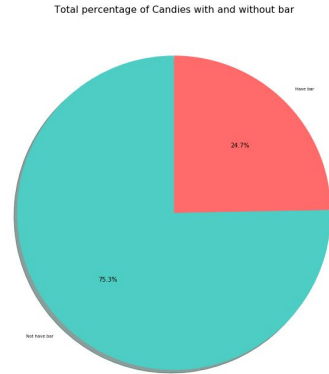
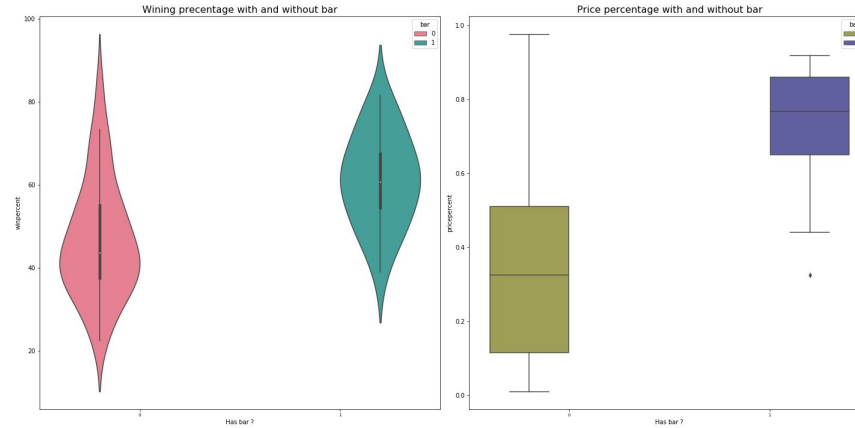
1. nougat has very slight influence on winpercent.
2. Its more expensive than others.
3. Not very popular among candy makers.

Data Visualisation & Descriptive analytics : crispedricewafer plot insight



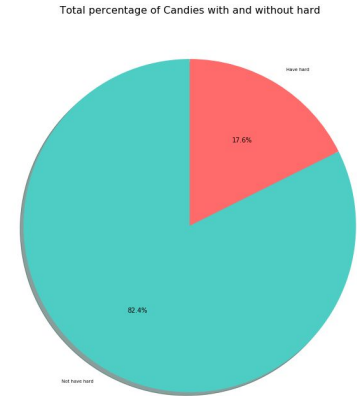
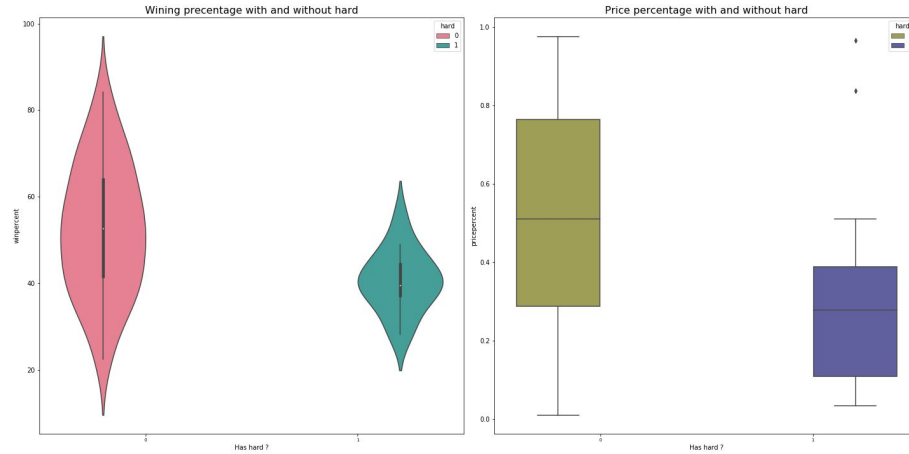
1. It has considerably good effect on winning. Minimum score is 40, which is a good sign.
2. Its bit more expensive.
3. Not very popular among candy makers.

Data Visualisation & Descriptive analytics : Bar plot insight



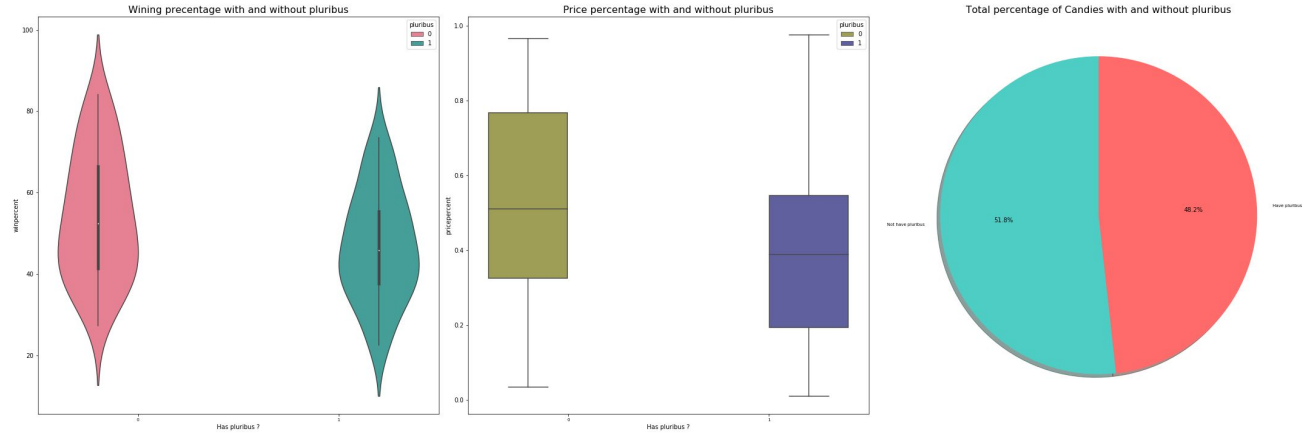
1. It has got influence on winpercent.
2. Its on bit expensive side.
3. More candy makers favours bar compare to hard candy.

Data Visualisation & Descriptive analytics : Hard plot insight



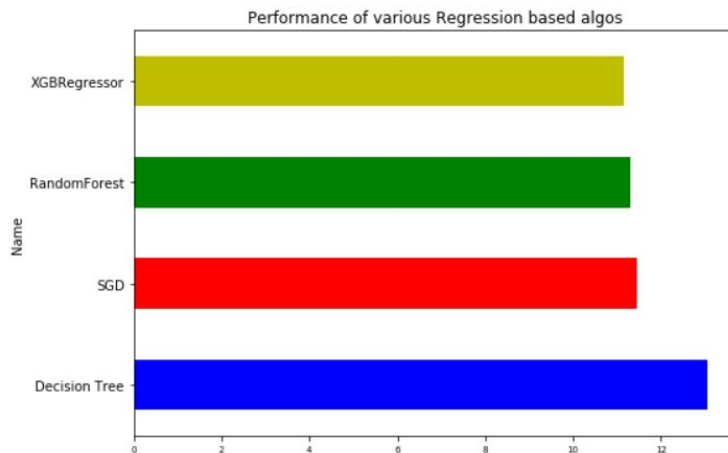
1. Hard candies are not favourite for winpercent.
2. Its not very expensive.
3. Not popular among candy makers.

Data Visualisation & Descriptive analytics : pluribus plot insight



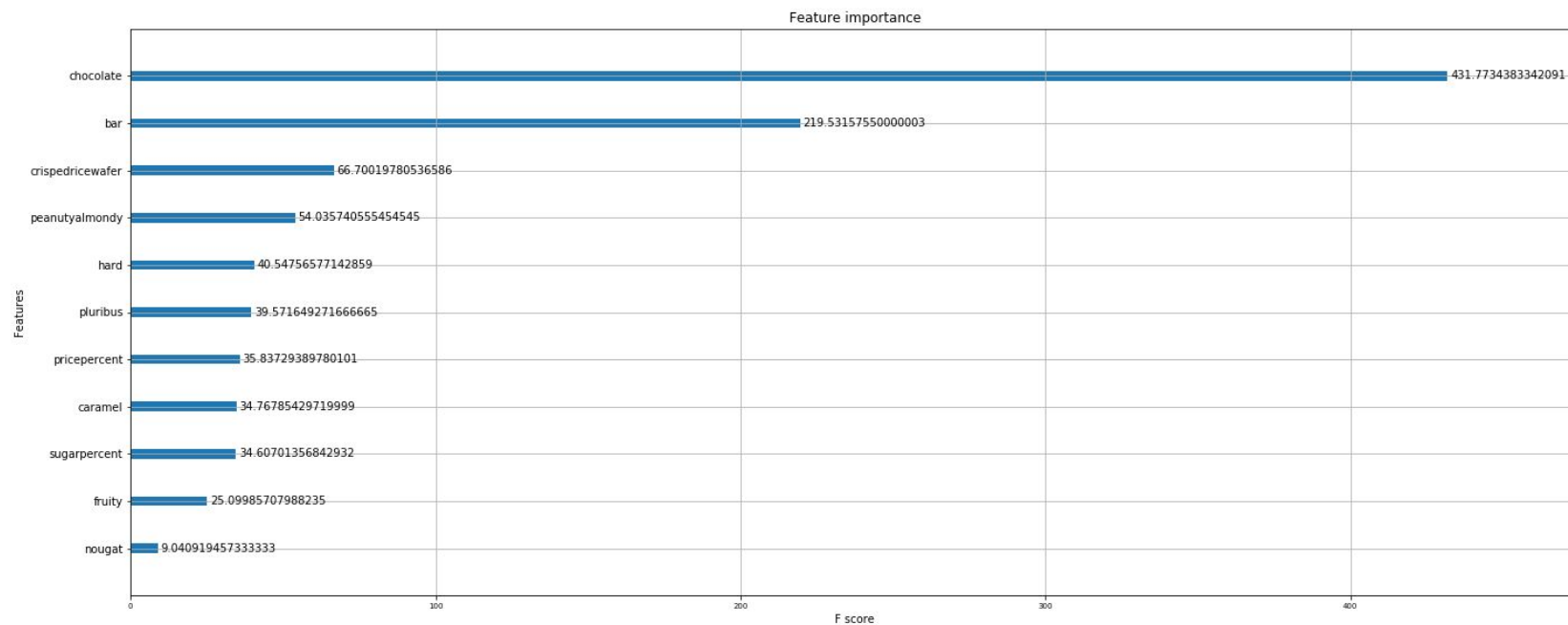
1. Candy without pluribus has high chance of winpercent.
2. Its moderately priced.
3. Almost half of candy makers favour this varient.

Regressor Performance : Model evaluation



1. XGB Regressor works best for modelling this data set.
2. On training the data on 75% of the record and tested on 25% of records, best r^2 Score i.e (1 - residual sum of square / total sum of squares) is 0.36 and lowest $Rmse_error$ is 11.16 from XGB Regressor.
3. Above plot shows performance of Regrssors.

Feature Importance



Summary

1. As per the analysis, candy with chocolate is more favourable.
2. Reese's seem to be a favourite brand.
3. Candy with Chocolate and Fruit seems rare.
4. Candy in bar form seems to be more famous among candy makers.

Thank You