Candy Power Ranking

Dr. Czuee Morey

29th July 2019

Chocolate candies are the most preferred. Other preferred traits are bar, peanutyalmondy & crispedricewafer.

The customer sentiments are price sensitive. Preference increases with price until 75th percentile and then falls for very high prices.

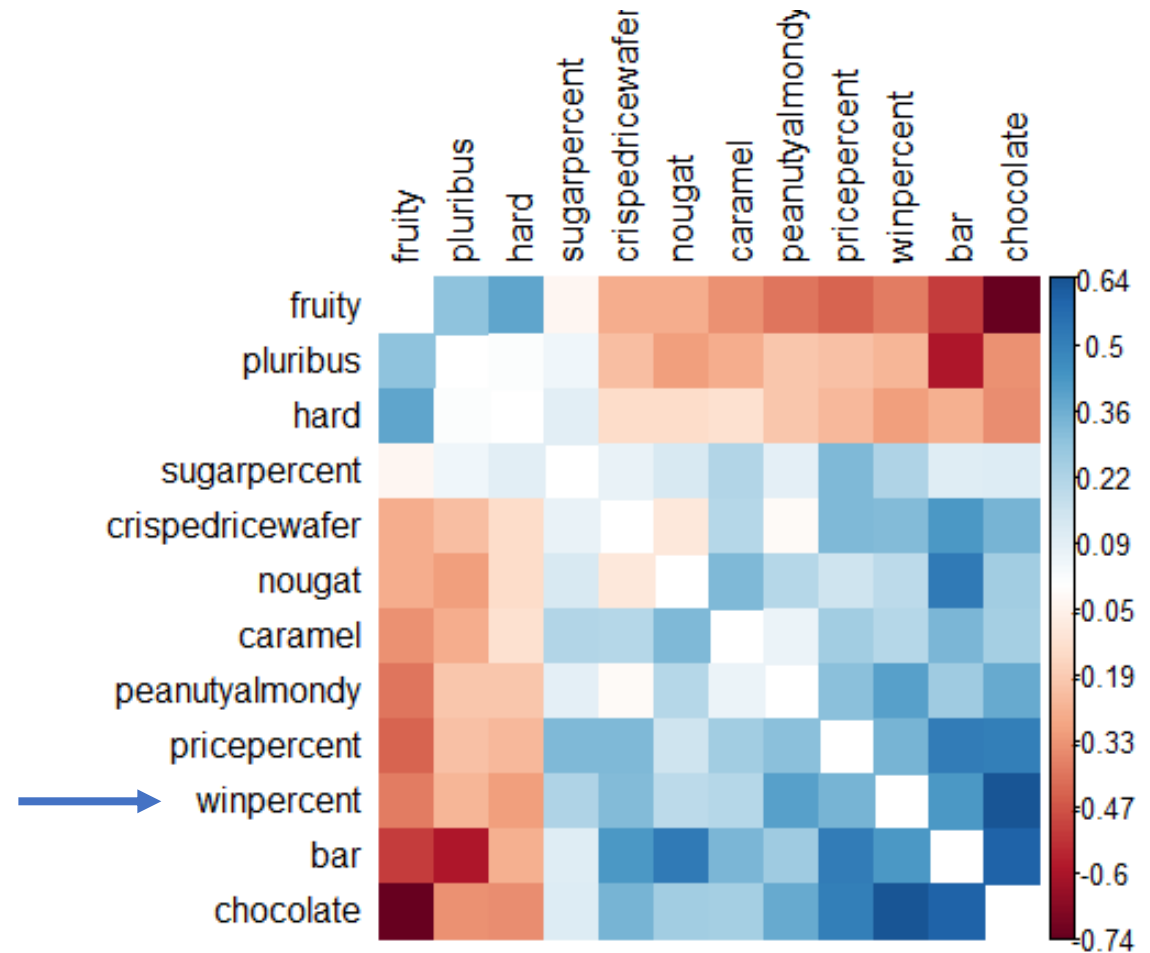The dataset is small with high variability, and additional data can help to make a more robust recommendation.
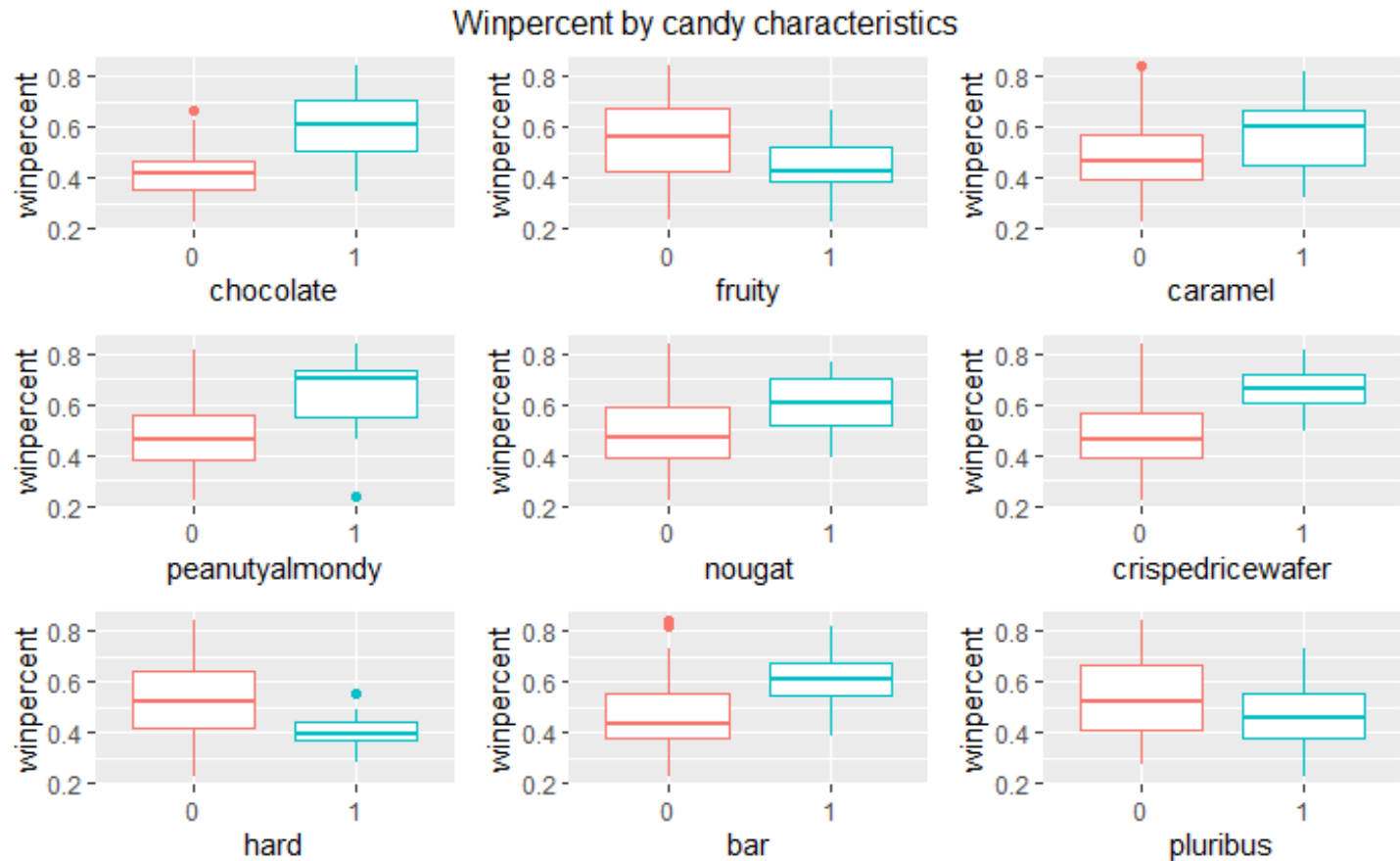
Executive Summary

Data distribution

Which candy traits go together?
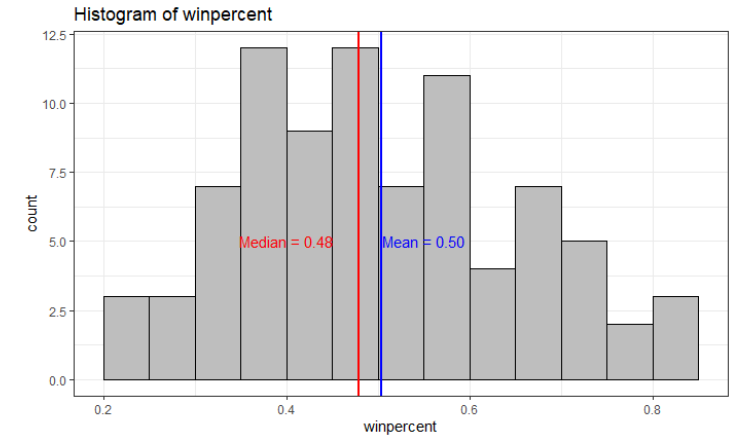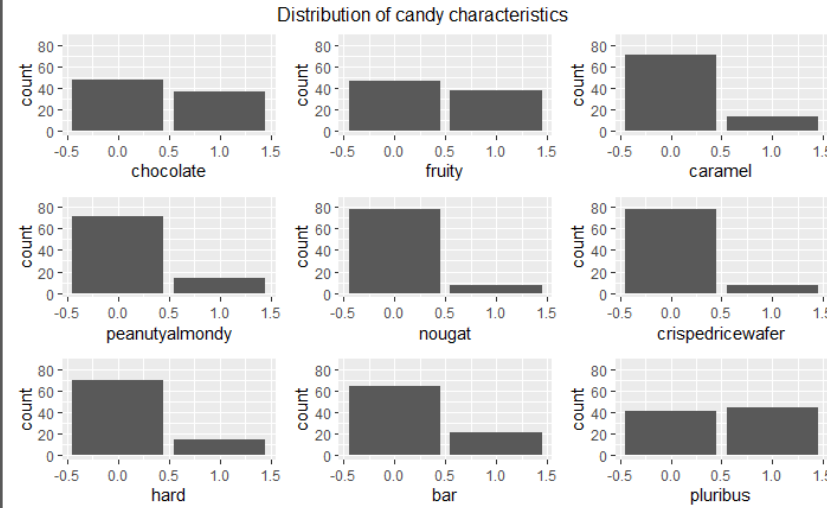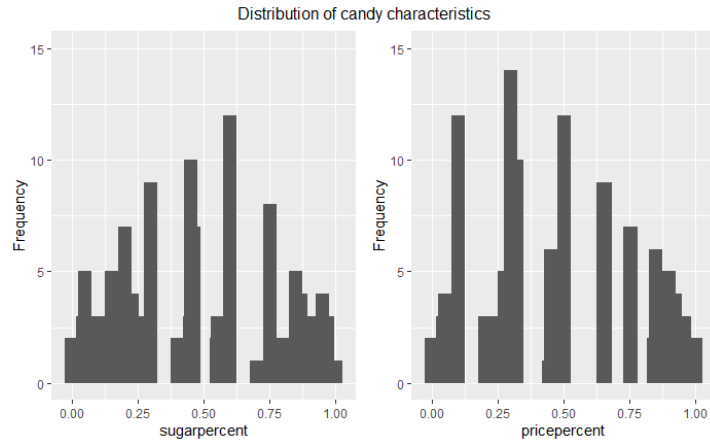
Chocolate is popular! But so are some other traits…

- Peanutyalmondy
- Crispedricewafer
- Bars

Winpercent by candy characteristics

Are customers price and health conscious?

- Sugar and price have weak positive correlations with winning (~0.3).

- Winpercent increases with price upto 75th percentile, but then goes down for more expensive candies

Winpercent by candy characteristics
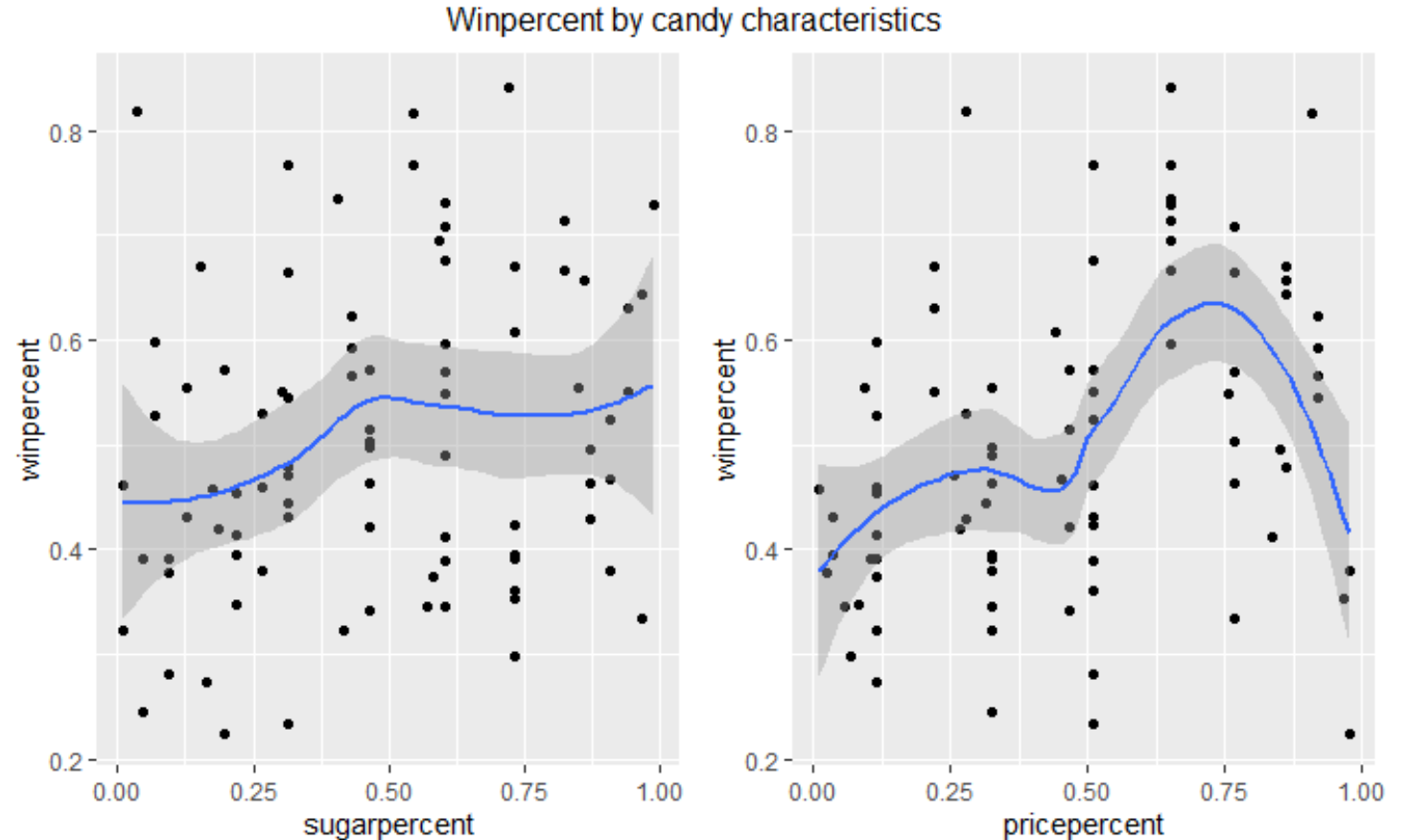
The dataset has high variability

- Only 58% of the variability in the dataset is captured in 3 PCA dimensions.

- Chocolate, bar, pricepercent, fruity, hard, crispedricewafer & nougat are the major contributors in Dim1-3

PCA Biplot clustered by chocolate showing the top contributors
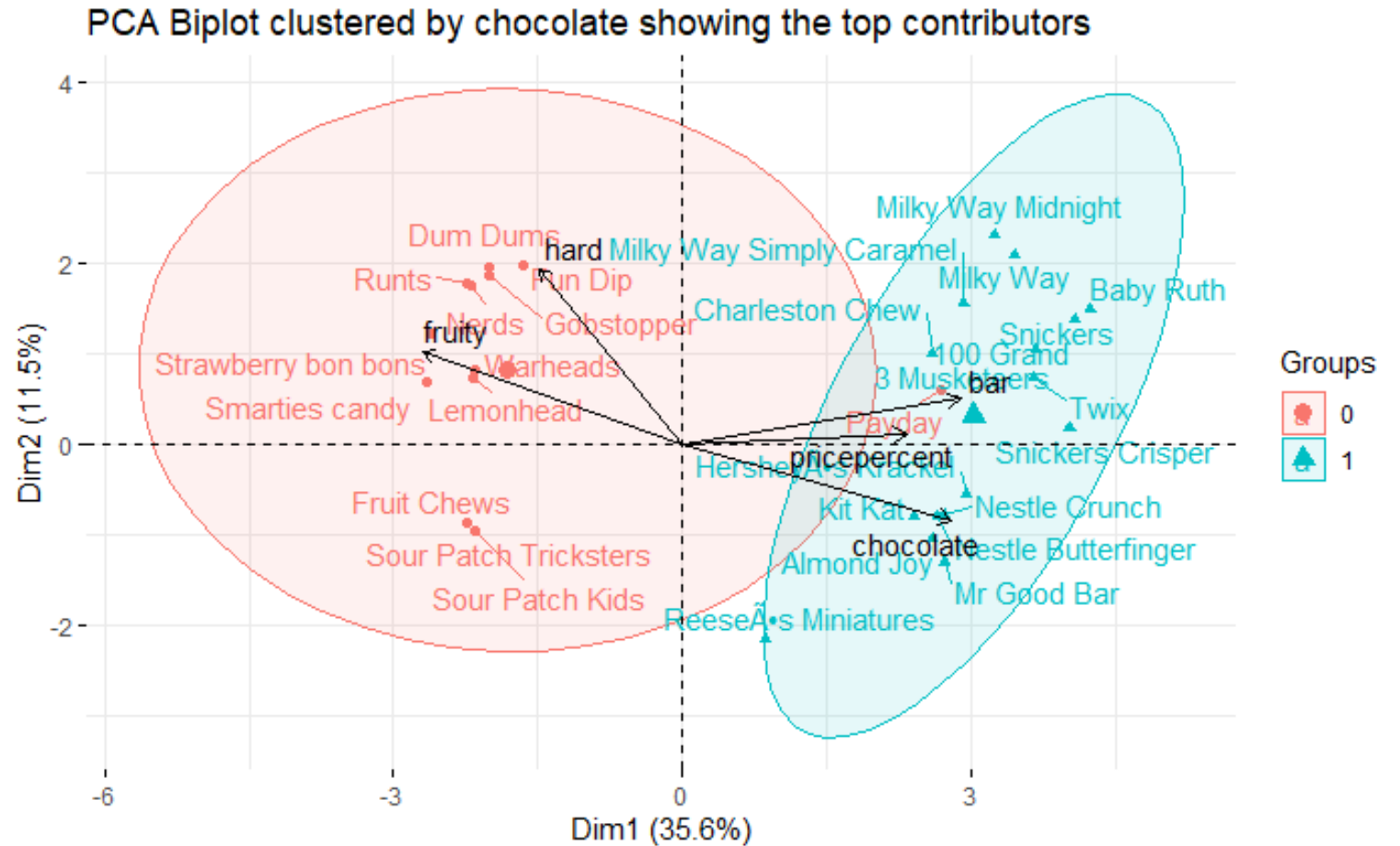
Candy preference predictions

```
Call:
lm(formula = winpercent ~ ., data = ctrain)

Residuals:
     Min       1Q   Median       3Q      Max
-0.17134 -0.06062  0.00484  0.07617  0.19002

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.3183     0.0478    6.65  1.2e-08 ***
chocolate          0.1925     0.0413    4.66  1.9e-05 ***
fruity             0.1139     0.0425    2.68   0.0096 **
caramel            0.0444     0.0447    0.99   0.3255
peanutyalmondy     0.1018     0.0424    2.40   0.0197 *
nougat            -0.0176     0.0581   -0.30   0.7631
crispedricewafer   0.1271     0.0552    2.30   0.0250 *
hard              -0.0620     0.0372   -1.67   0.1007
bar                0.0273     0.0519    0.53   0.6009
pluribus          -0.0168     0.0326   -0.51   0.6089
sugarpercent       0.1432     0.0483    2.97   0.0044 **
pricepercent      -0.0955     0.0581   -1.64   0.1060
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.102 on 57 degrees of freedom
Multiple R-squared:  0.58,     Adjusted R-squared:  0.499
F-statistic: 7.16 on 11 and 57 DF,  p-value: 1.77e-07
```
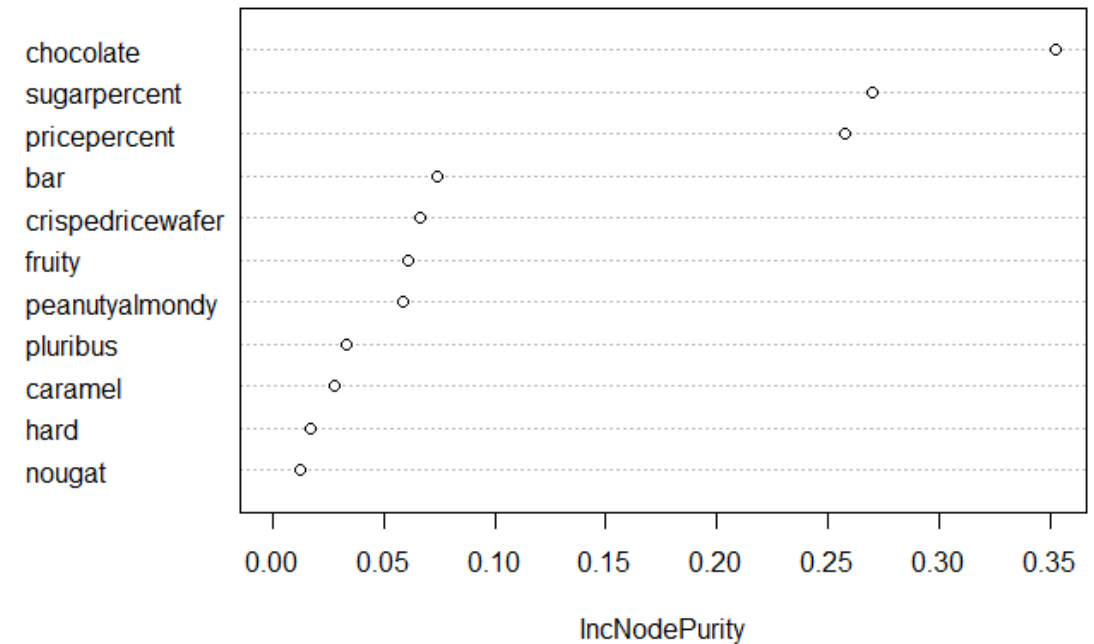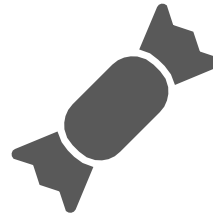

Random Forest variable importance

New Candy Maker

| Predictors	Estimate	Std. Error	Signif	Selection	Prediction
(Intercept)	0.3183	0.0478	***		0.3183
chocolate	0.1925	0.0413	***	1	0.1925
fruity	0.1139	0.0425	**	0	0
caramel	0.0444	0.0447		1	0.0444
peanutyalmondy	0.1018	0.0424	*	1	0.1018
nougat	-0.0176	0.0581		0	0
crispedricewafer	0.1271	0.0552	*	1	0.1271
hard	-0.062	0.0372		0	0
bar	0.0273	0.0519		1	0.0273
pluribus	-0.0168	0.0326		0	0
sugarpercent	0.1432	0.0483	**	0.63	0.090216
pricepercent	-0.0955	0.0581		0.65	-0.062075
Prediction					0.839541

Consider customer segments: age, gender, shop locations

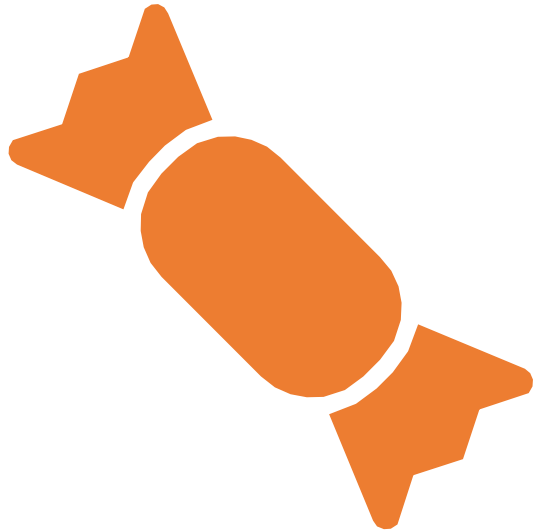Improve experiment design to judge customer satisfaction: Sales instead of matchups

Test findings on a new population of customers

Future directions
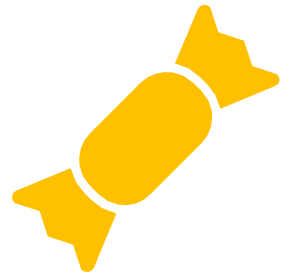
Thank you!

Dr. Czuee Morey

Czuee.Morey@gmail.com