Maven Halloween Challenge

~ Monil Shah

Maven Halloween Challenge

Goal: Use data to find the 3 types of Halloween candy that will make you the most popular house on the block.

Details: Using online votes ranking 85 types of candy, your task is to find the 3 treats you'll give out on Halloween to guarantee that trick-or-treaters of all tastes find something they'll love and present the data to back up your decision.

Approach

Python Pandas: Load and View the data.

Machine Learning: PCA and jitter to prepare the data for the scatter plot.

Tableau: Explore the scatter plot and make recommendations

Python Pandas

[1]:
```
import pandas as pd
```

[2]:
```
candies = pd.read_csv('candy-data.csv')
candies.head()
```

[2]:

	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

[3]:
```
flavours_subset = candies.iloc[:,1:-3]
flavours_subset.head()
```

[3]:

	chocolate	fruity	caramel	peanutyalmondy	nougat	crispedricewafer	hard	bar	pluribus
0	1	0	1	0	0	1	0	1	0
1	1	0	0	0	1	0	0	1	0
2	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0
4	0	1	0	0	0	0	0	0	0

Machine Learning

Principal Component Analysis (PCA)

[4]:
```
from sklearn.decomposition import PCA
```

[5]:
```
pca = PCA(n_components=2)
```

[6]:
```
pca.fit(flavours_subset)
```

[6]:
```
PCA
PCA(n_components=2)
```

[7]:
```
pca.explained_variance_ratio_
```

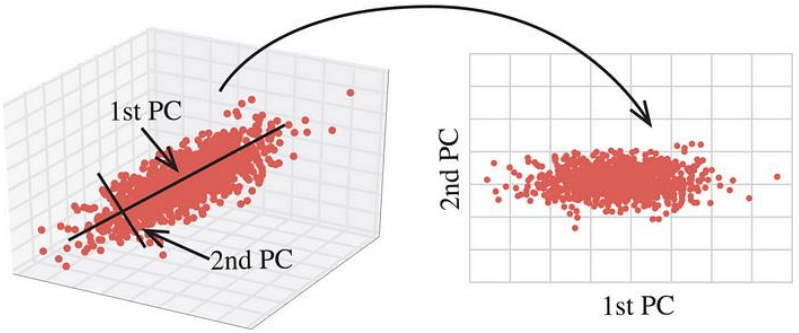
[7]:
```
array([0.45736631, 0.15454393])
```

These 2 newly created colums out of the 9 original columns are able to capture around 60% of the information in the original 9 colums. which is a decent representation.

[8]:
```
candies_2d = pd.DataFrame(pca.transform(flavours_subset))
candies_2d.head()
```

[8]:

	0	1
0	1.305021	-0.335978
1	1.123628	-0.294698
2	0.034733	-0.204088
3	0.034733	-0.204088
4	-0.469043	-0.565369



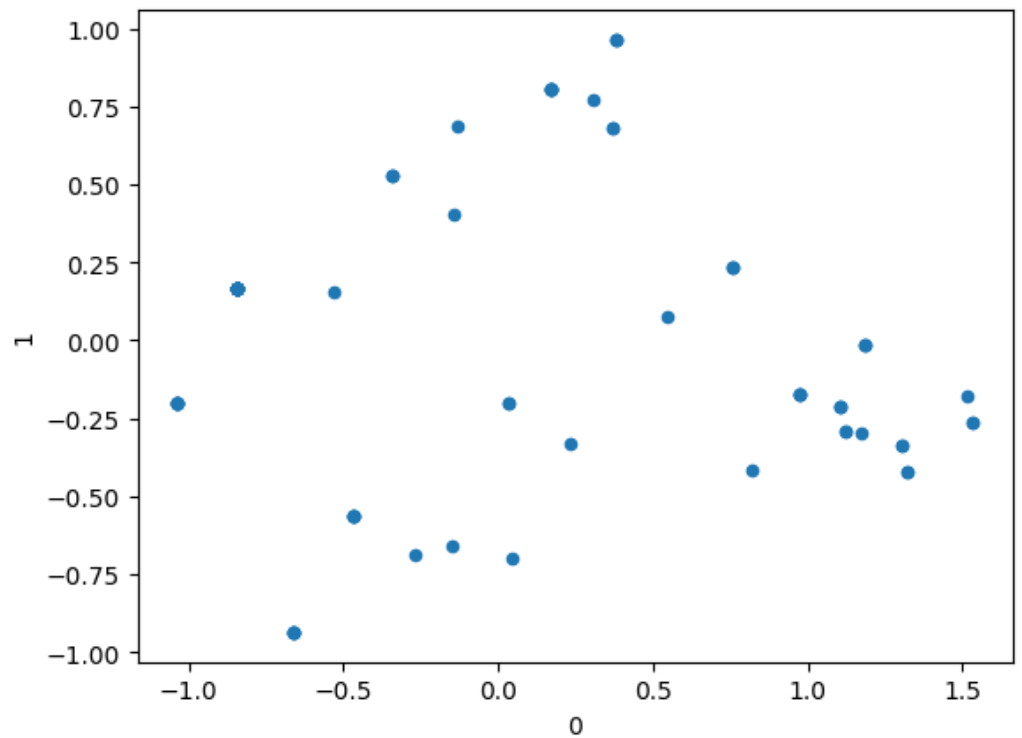
[8]:
```
candies_2d = pd.DataFrame(pca.transform(flavours_subset))
candies_2d.head()
```

[8]:

	0	1
0	1.305021	-0.335978
1	1.123628	-0.294698
2	0.034733	-0.204088
3	0.034733	-0.204088
4	-0.469043	-0.565369

[19]:
```
candies_2d.plot(kind='scatter',x=0,y=1)
```

[19]:
```
<Axes: xlabel='0', ylabel='1'>
```



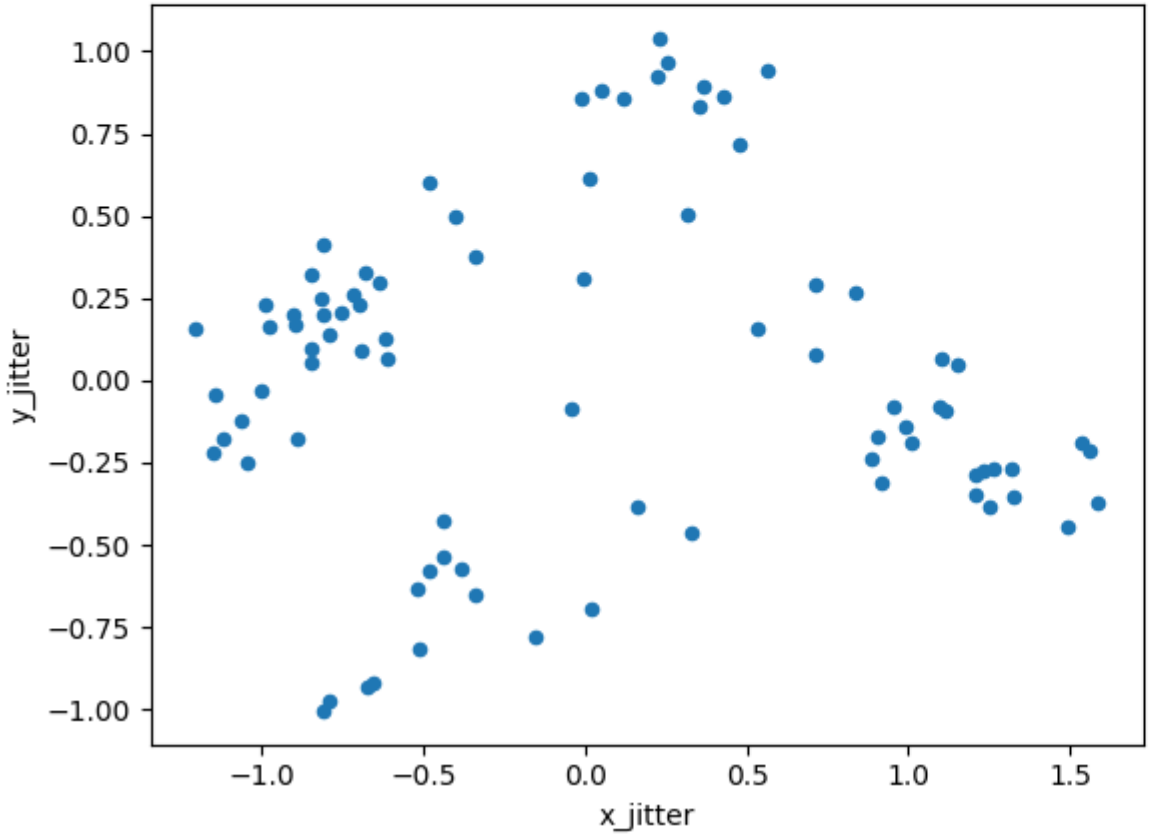
since some of the candies have similar values of these 9 features they are overlapping so we should add a little bit of random noise to have less overlap

Jitter (Random Noise)

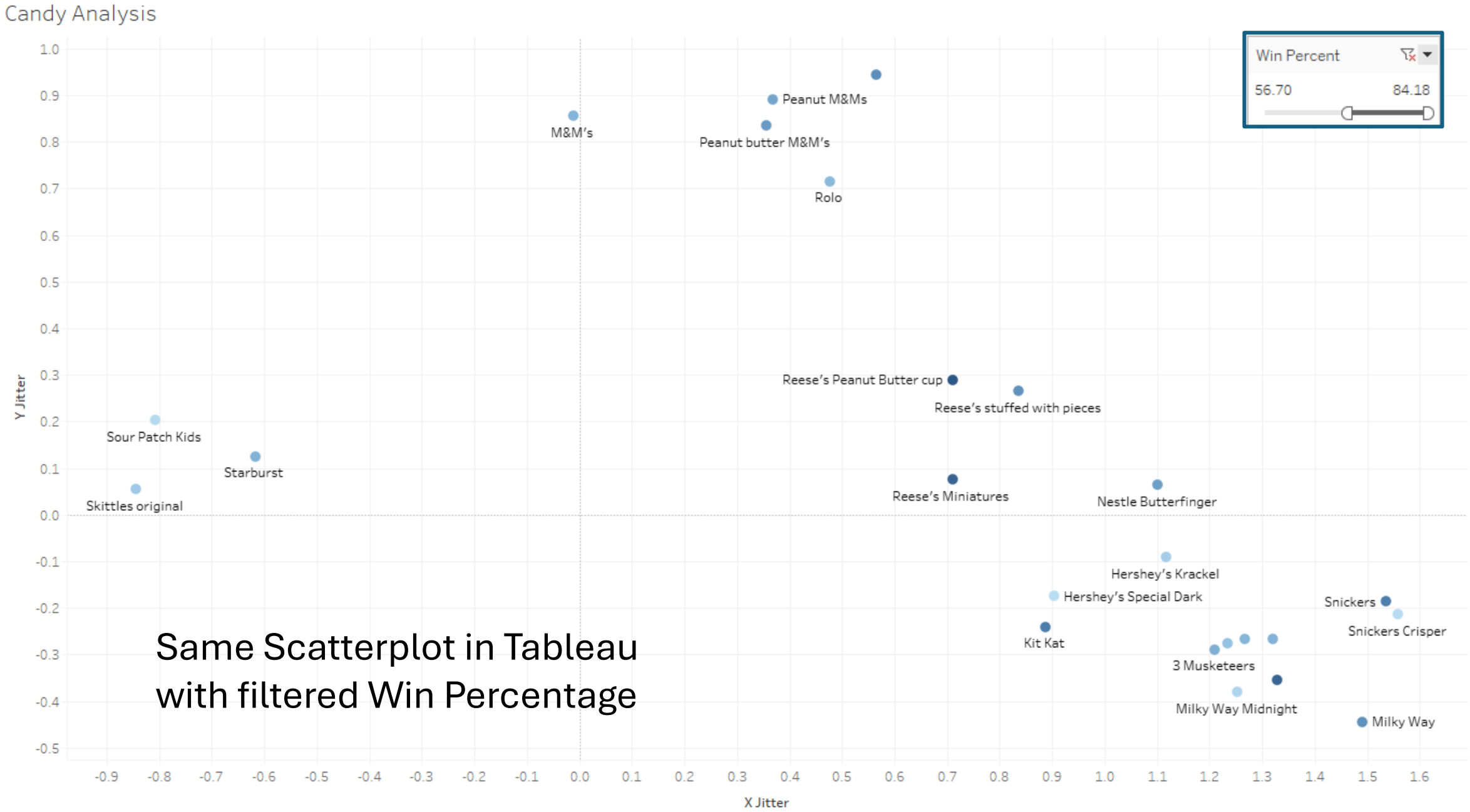
since some of the candies have similar values of these 9 features they are overlapping so we should add a little bit of random noise to have less overlap

[38]:
```
import numpy as np

candies['x_jitter'] = candies_2d[0] + np.random.randn(85)*0.1
candies['y_jitter'] = candies_2d[1] + np.random.randn(85)*0.1

candies.plot(kind='scatter',x='x_jitter',y='y_jitter')
```

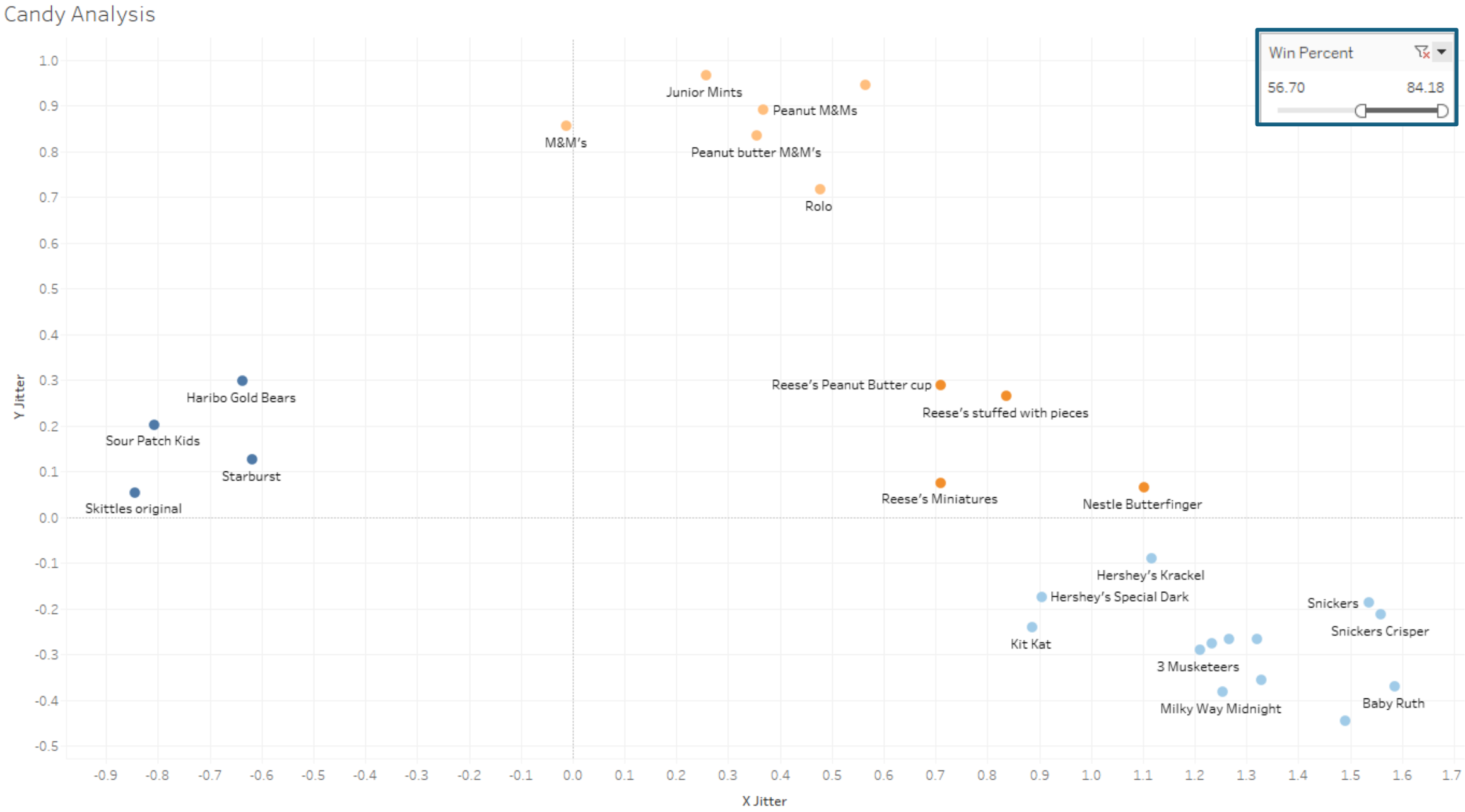
[38]:
```
<Axes: xlabel='x_jitter', ylabel='y_jitter'>
```



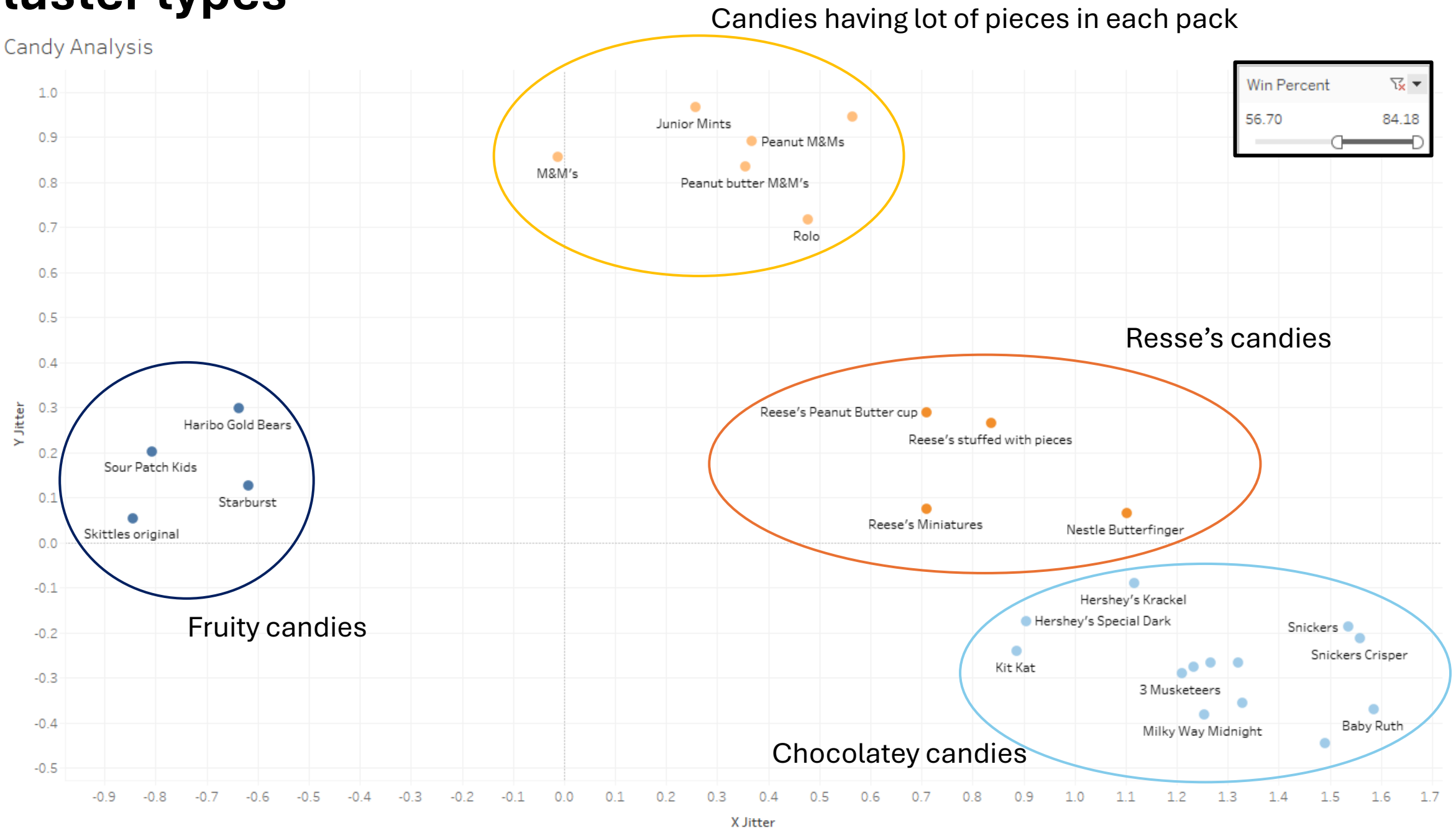
Tableau


Candy Analysis
Win Percent
56.70
84.18
Peanut M&Ms
M&M's
Peanut butter M&M's
Rolo
Reese's Peanut Butter cup
Reese's stuffed with pieces
Sour Patch Kids
Starburst
Skittles original
Reese's Miniatures
Nestle Butterfinger
Hershey's Krackel
Hershey's Special Dark
Snickers
Snickers Crisper
Kit Kat
3 Musketeers
Milky Way Midnight
Milky Way
Y Jitter
X Jitter
Same Scatterplot in Tableau
with filtered Win Percentage

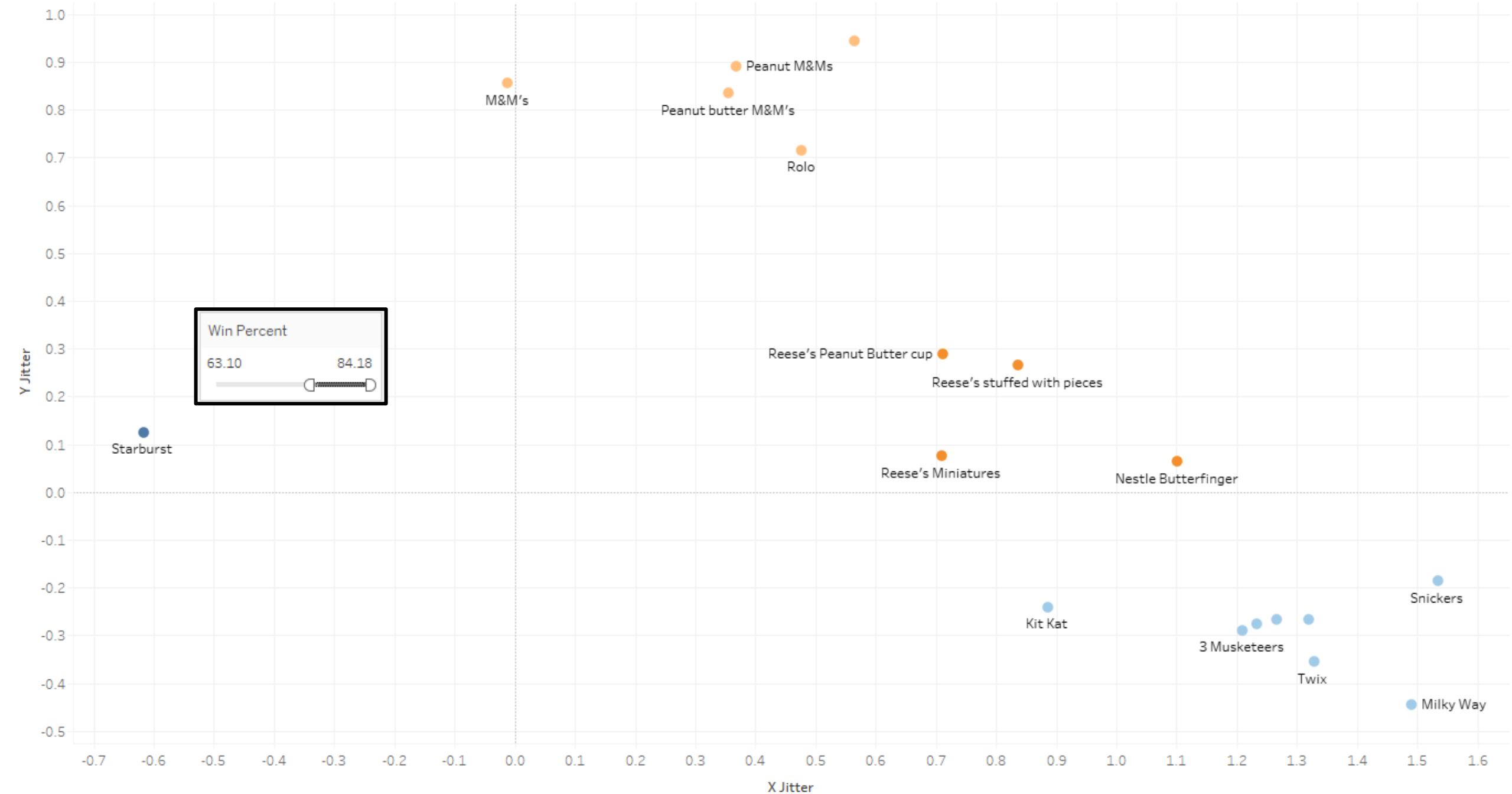
4 Clusters


Candy Analysis
Win Percent
56.70
84.18
Junior Mints
Peanut M&Ms
M&M's
Peanut butter M&M's
Rolo
Reese's Peanut Butter cup
Reese's stuffed with pieces
Haribo Gold Bears
Sour Patch Kids
Starburst
Skittles original
Reese's Miniatures
Nestle Butterfinger
Hershey's Krackel
Hershey's Special Dark
Snickers
Snickers Crisper
Kit Kat
3 Musketeers
Milky Way Midnight
Baby Ruth
Y Jitter
X Jitter

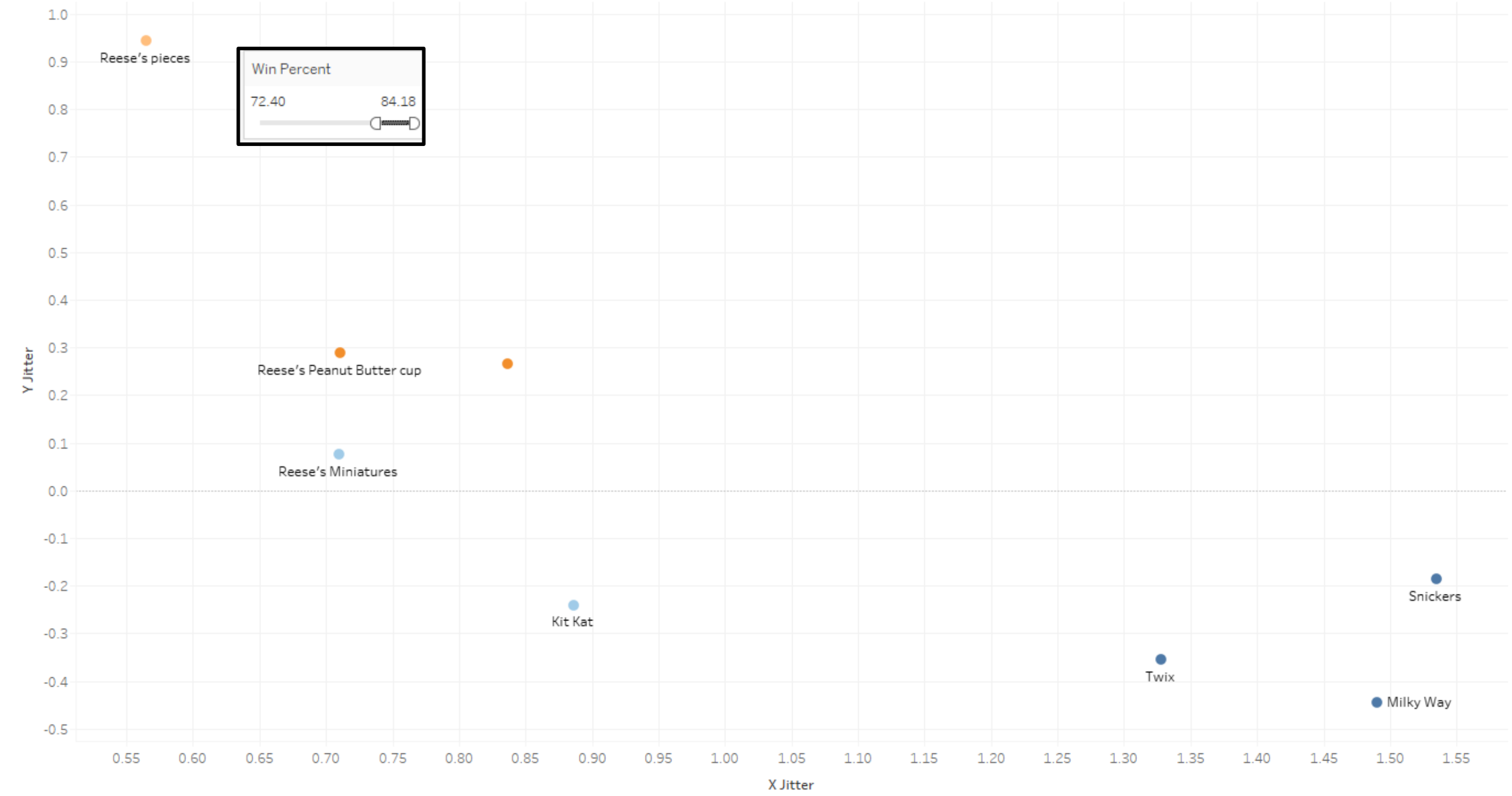
Cluster types



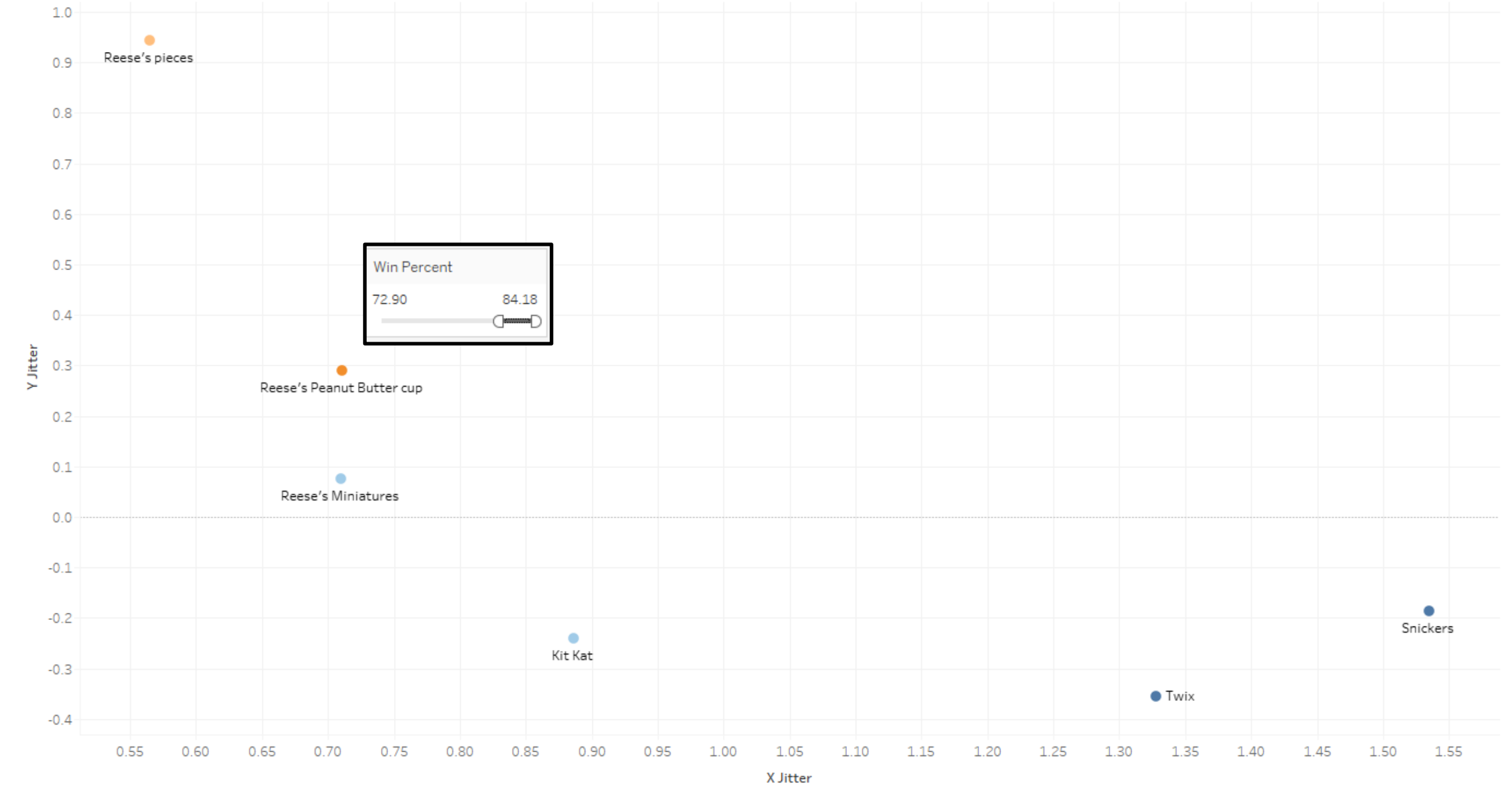
Candy Analysis



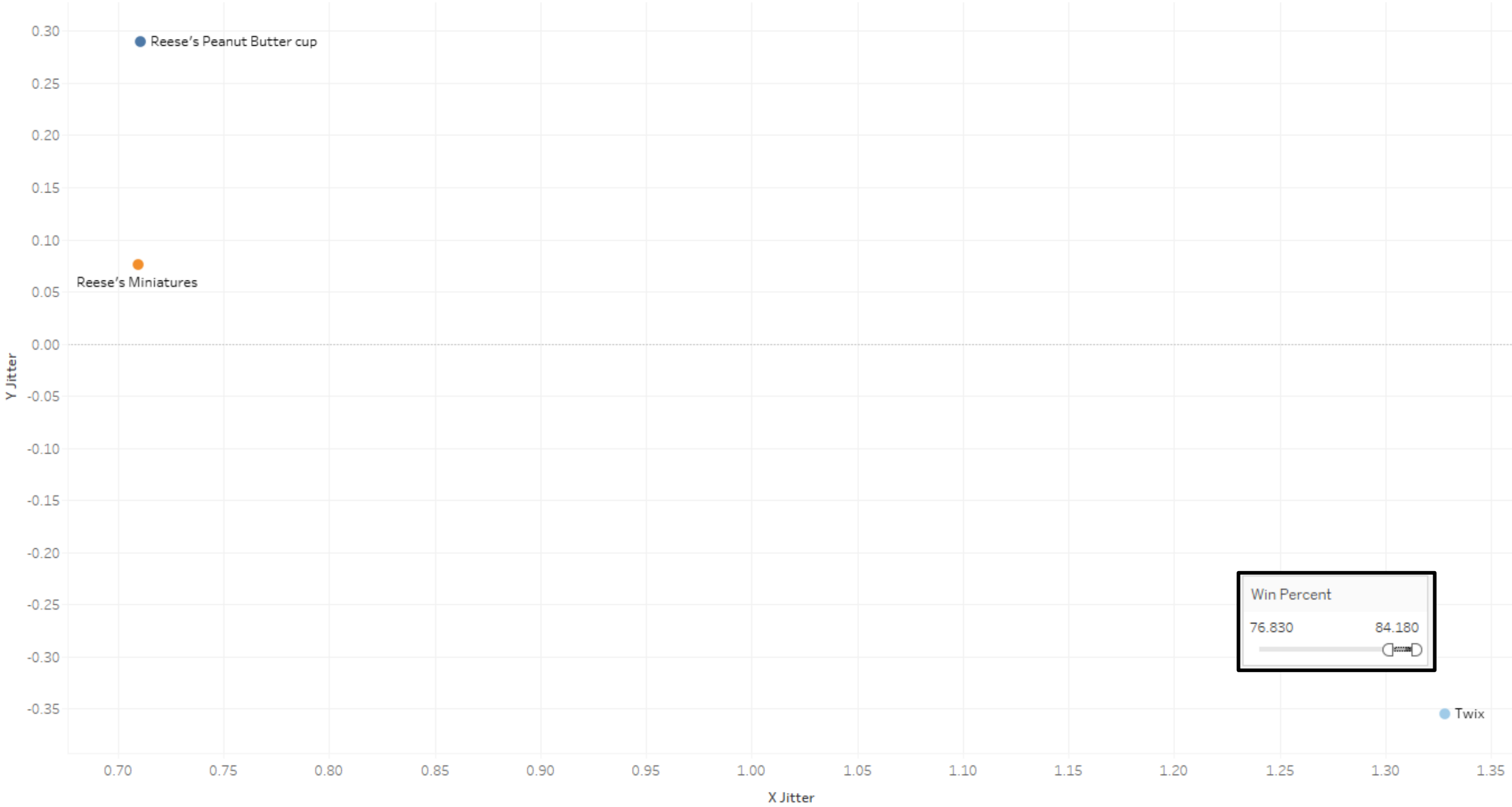
Candy Analysis



Candy Analysis

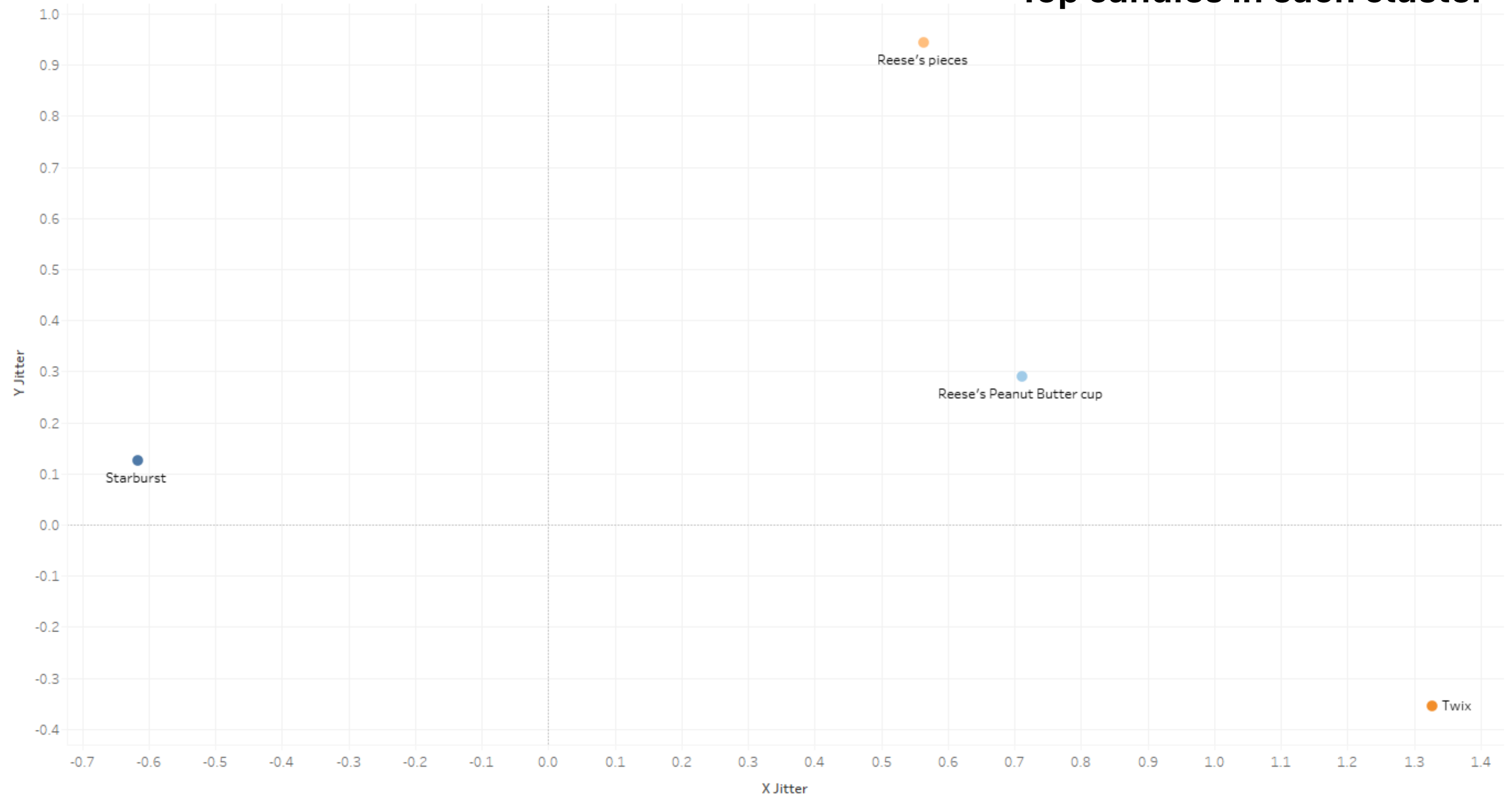


Candy Analysis



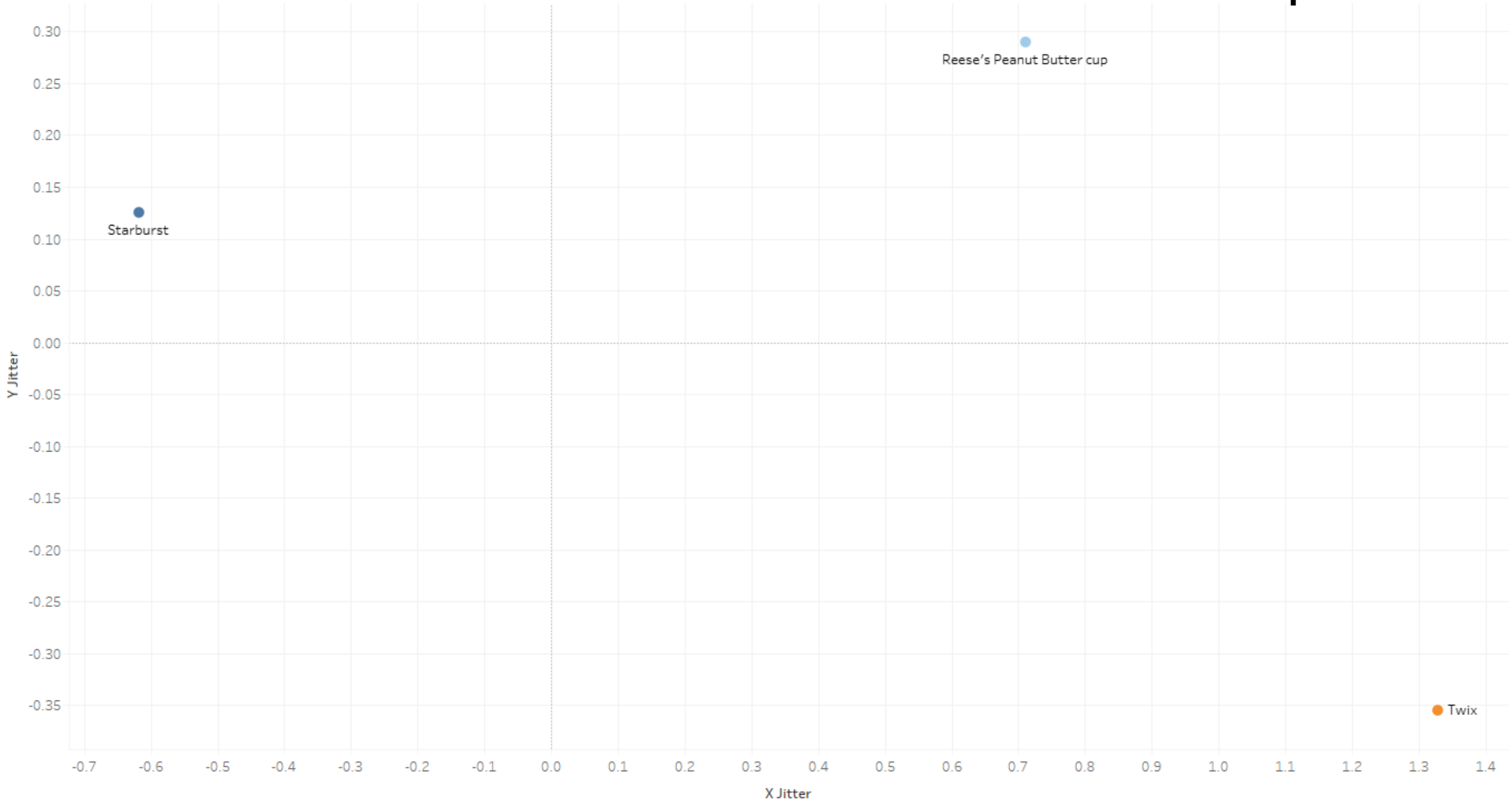
Candy Analysis

Top candies in each cluster



Candy Analysis

Top 3 candies



Top 3 recommendations

1. Reese’s Peanut Butter Cup
2. Twix
3. Starburst

Thank You

www.linkedin.com/in/monil09 https://github.com/MJshah001