Candy Brand

Lidl Analytics – Data Science Case Study

Sagarnil Dasgupta

dasgupta.sagarnil@gmail.com

Brief Result of the case study

When this experience is conducted using 25000 user choices not the ingredient but the packaging and the colour and smell of the candy had a part to playing in deciding the win percentage from an user perspective.

These parameters like sugar percentage, chocolate and caramel is not the only parameters influencing a user likeness of the candy. Boolean values like 0 and 1 for chocolate and caramel says if that ingredient is present or not but how much grams of chocolate in 100 grams of the candy could help to better analyse.

Like for example

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | competitorname | chocolate | fruity | caramel | peanutyalmondy | nougat | crispedricewafer | hard | bar | pluribus | sugarpercent | pricepercent | winpercent |
| 46 | Nik L Nip | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.20 | 0.98 | 22.45 |
| 70 | Starburst | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.15 | 0.22 | 67.04 |

Nik L Nip and Starburst have almost all the intergradient present with almost same sugar with a difference in price but there is a significant difference in win percentage which some other attributes influence the win percentage of a candy.

Few examples with similar feature but significant difference in win percentage.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | competitorname | chocolate | fruity | caramel | peanutyalmondy | nougat | crispedricewafer | hard | bar | pluribus | sugarpercent | pricepercent | winpercent |
| 33 | Lifesavers big ring gummies | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.27 | 0.28 | 52.91 |
| 74 | Super Bubble | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.16 | 0.12 | 27.30 |
| 35 | M&M’s | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.82 | 0.65 | 66.57 |
| 61 | Sixlets | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.22 | 0.08 | 34.72 |

On further analysis it is found that chocolate and fruity are highly corelated to each other as they are mutually exclusive except one- Hershey’s Special Dark where chocolate and fruity both are present. It somehow seems that any candy manufacture finds it tough to add both chocolate and fruity in a single candy.

When each ingredient is taken into account following trend is observed :

|  |  |
| --- | --- |
| CANDY TYPE | AVG. WIN SHARE |
| Nougat | 66% |
| Crispy | 64% |
| Chocolate | 61% |
| Candy bar | 61% |
| Caramel | 60% |
| Peanuts & nuts | 57% |
| Hard candy | 47% |
| Fruit | 44% |
| Multi-piece | 41% |

The above chart shows the average win percentage of each ingredient in candy.

Whereas when predicting the win percentage based on the 85 candies it is found Nougat, Bar, Caramel and pluribus or multi-piece have no such significant influence in determining the win percentage where Chocolate, Fruit, Peanut & Nuts, Crispy and a high sugar percentage is liked by most of the user. Ideal candy should be soft not a jaw breaker and price should be reasonable range.

|  |  |
| --- | --- |
| **CANDY TYPE** | **Value added to win percentage** |
| Chocolate | 19.9873 |
| Fruit | 8.6228 |
| Peanuts & nuts | 10.0435 |
| Crispy | 9.4243 |
| Hard Candy | -6.0456 |
| sugar percentage | 9.5396 |
| price percentage | -5.4628 |

From the above table we can say that Chocolate and Peanut are a must in the ideal candy as its adds a value of (+20) and (+10) to the overall win percentage. Adding Fruits will increase the percentage but in the sample Chocolate and Fruit combo are very rare and one taste could be killing the other. Crispy normally add on the win percentage but chocolate + Peanuts + Crispy like (Snickers Crisper) doesn’t do as well as Reeves’s Peanut Butter cup and its spinoffs.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | competitorname | chocolate | fruity | caramel | peanutyalmondy | nougat | crispedricewafer | hard | bar | pluribus | sugarpercent | pricepercent | winpercent |
| 53 | Reeves’s Miniatures | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0.03 | 0.28 | 81.87 |
| 54 | Reeves’s Peanut Butter cup | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0.72 | 0.65 | 84.18 |
| 56 | Reeves’s stuffed with pieces | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0.99 | 0.65 | 72.89 |

There are only 3 candies all from same brand Reese’s which have same composition of only chocolate and Peanuts and have an average win percentage of 80% which says either this combination works out for most of the customers or the buyers are more connected with the Reese’s brand. As we also see 4 out of 10 most win percentage is also from Reese’s brand.

Following conclusion are made based on the observation obtained by applying analytics technique to candy dataset described below Halloween Candy Power Ranking.

List of figures

[Figure1: Test Cycle Timeline 7](#_Toc3919937)

[Figure2: Overall test overview 8](#_Toc3919938)

## List of Observation

[Observation 1: Reese’s Peanut Butter Cups and their spinoffs come out huge here, taking four of the top 10 spots and appearing pretty synonymous with the platonic ideal of Halloween candy. 6](#_Toc21093406)

[Observation 2: All the top 10 candies has chocolate attribute present. 6](#_Toc21093407)

[Observation 3: Reese's Miniatures is very cheap when compared to top competitors and overall as well. 6](#_Toc21093408)

[Observation 3: When non-chocolate candies are taken into consideration Starburst tops the chart. 6](#_Toc21093409)

[Observation ReeseÕs stuffed with pieces is the top sugary candy 7](#_Toc21093410)

[Observation 4: Tootsie Roll Midgies seems to perform better when price and win percentages are taken. 8](#_Toc21093411)

[Observation 5: Dum Dums seems to perform better when price and sugar percentages are considered. 8](#_Toc21093412)

[Observation 11:When checked in detail regarding the chocolate and fruity attribute it is seen except 1(Tootsie Pop) which is chocolate as well as fruit. Other than this candy there is no candy which is have a chocolatey and fruity combo. Either it is chocolatey or fruity or no of these attribute. 9](#_Toc21093413)

[Observation 10: Chocolate, fruity, peanut-almond, crispy and sugar percentage has positive impact on win percentage but hard could be avoided and expected with any retail product less price is more linked by the customer. 11](#_Toc21093414)

[Table 11: Glossary table 14](#_Toc21093415)

Table of contents

[List of Observation 4](#_Toc21268732)

[Table of contents 5](#_Toc21268733)

[1 Introduction 6](#_Toc21268734)

[1.1 Scenario 6](#_Toc21268735)

[1.2 Data 6](#_Toc21268736)

[1.3 Data Description 6](#_Toc21268737)

[1.4 Data Manipulation 7](#_Toc21268738)

[1.4.1 Missing value check 7](#_Toc21268739)

[1.4.1.1 Sagatnil 7](#_Toc21268740)

[1.4.2 Rectifying scale 7](#_Toc21268741)

[1.5 Generic 8](#_Toc21268742)

[1.6 Top 10 candy by win percentage 8](#_Toc21268743)

[1.7 Top 5 and bottom 5 candies by win percentage 8](#_Toc21268744)

[1.8 Top candy without chocolate 9](#_Toc21268745)

[1.9 Top 10 sugary candy 9](#_Toc21268746)

[1.10 Top 5 and bottom 5 candies by sugar percentage 10](#_Toc21268747)

[1.11 Impact of price on win and sugar 10](#_Toc21268748)

[1.12 Correlation 12](#_Toc21268749)

[Test Summary 13](#_Toc21268750)

[1.13 Predicting Winpercentage 13](#_Toc21268751)

[2 Summary & Conclusions 16](#_Toc21268752)

[2.1 Making the perfect Candy 16](#_Toc21268753)

[2.2 Ingredient of perfect candy 16](#_Toc21268754)

[2.3 Best Candy with ingredient 17](#_Toc21268755)

[2.4 Recommendations 17](#_Toc21268756)

[3 Glossary 18](#_Toc21268757)

[**Day 2: Plot a Histogram** 18](#_Toc21268758)

[**Day 3: Perform a T-test** 18](#_Toc21268759)

[**Day 4: Draw a Bar Chart** 18](#_Toc21268760)

[**Day 5: Perform a Chi-Square Test** 18](#_Toc21268761)

# Introduction

This document details about the approach taken to analyse the best candy afer retrivng

## Scenario

The Lidl purchasing group wants to expand their candy offering. These are store brand candies that are sold along the brand offerings. The idea is to create a brand-new product. The team is discussing various options at the moment.

Some prefer cookie-based sweets while others think that it should be gummies. The Divisional Director responsible for purchasing has decided to use a more data-driven approach. He contracted with a market research group to collect data on products in the market and their characteristics and customer sentiment.

The market research data is now available and it is your job to find out which product characteristics drive customer sentiment and subsequently make a recommendation on a new product.

## Data

The data set is located (incl. a short description) here:

<https://github.com/fivethirtyeight/data/tree/master/candy-power-ranking>

The data set is provided by FiveThirtyEight under the Creative Commons Attribution 4.0 International license (<https://creativecommons.org/licenses/by/4.0/> )

## Data Description

| **Header** | **Description** |
| --- | --- |
| chocolate | Does it contain chocolate? |
| fruity | Is it fruit flavored? |
| caramel | Is there caramel in the candy? |
| peanutalmondy | Does it contain peanuts, peanut butter or almonds? |
| nougat | Does it contain nougat? |
| crispedricewafer | Does it contain crisped rice, wafers, or a cookie component? |
| hard | Is it a hard candy? |
| bar | Is it a candy bar? |
| pluribus | Is it one of many candies in a bag or box? |
| sugarpercent | The percentile of sugar it falls under within the data set. |
| pricepercent | The unit price percentile compared to the rest of the set. |
| winpercent | The overall win percentage according to 269,000 matchups. |

# Data Manipulation

## Missing value check

There are no missing values in the dataset.

## Sagatnil

## Rectifying scale

Sugar percentage and price percentage are in the scale from 0 to 1 but winpercentage are in the scale of 0 to 100. As this is a regression problem with Winpercentage as target variable its ok to keep in different scale.

# Data Analysis

candy-data.csv file consist of 85 rows which are 85 candy types and each one of them is defined based on attributes like chocolate, fruity etc… and the target variable while determines of the user liked it is winpercentage.

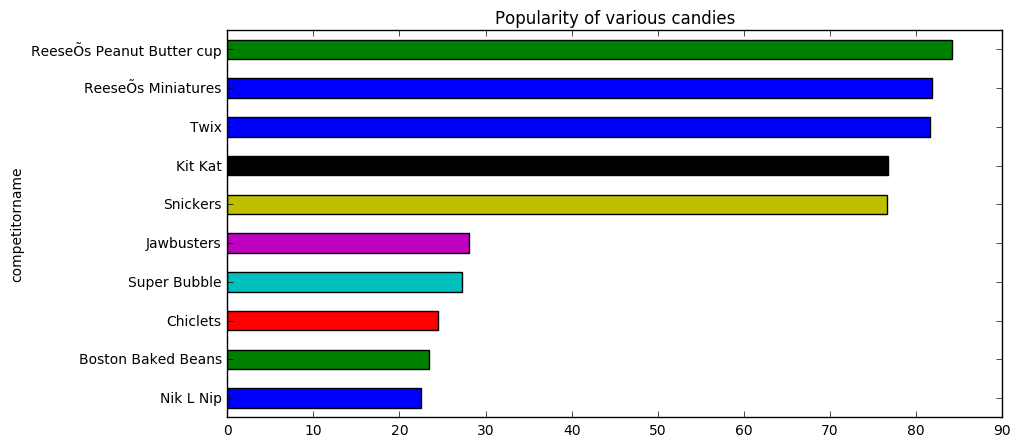
# Top 10 candy by win percentage

As Winpercentage is the attribute which defines whether user likes it compared to other candy brand. It is assumed a higher win percentage means that candy is licked better than other candies.

Top 10 candies when terms of winpercentage are as follows:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | competitorname | chocolate | fruity | caramel | peanutyalmondy | nougat | crispedricewafer | hard | bar | pluribus | sugarpercent | pricepercent | winpercent |
| 52 | ReeseÕs Peanut Butter cup | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0.72 | 0.65 | 84.18 |
| 51 | ReeseÕs Miniatures | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0.03 | 0.28 | 81.87 |
| 79 | Twix | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0.55 | 0.91 | 81.64 |
| 28 | Kit Kat | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0.31 | 0.51 | 76.77 |
| 64 | Snickers | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0.55 | 0.65 | 76.67 |
| 53 | ReeseÕs pieces | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0.41 | 0.65 | 73.43 |
| 36 | Milky Way | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0.60 | 0.65 | 73.10 |
| 54 | ReeseÕs stuffed with pieces | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0.99 | 0.65 | 72.89 |
| 32 | Peanut butter M&MÕs | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0.83 | 0.65 | 71.47 |
| 42 | Nestle Butterfinger | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0.60 | 0.77 | 70.74 |

## Top 5 and bottom 5 candies by win percentage



Observation 1: Reese’s Peanut Butter Cups and their spinoffs come out huge here, taking four of the top 10 spots and appearing pretty synonymous with the platonic ideal of Halloween candy.

Observation 2: All the top 10 candies has chocolate attribute present.

Observation 3: Reese's Miniatures is very cheap when compared to top competitors and overall as well.

Code details can be found at Jupyter notebook section 5

## Top candy without chocolate

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | competitorname | chocolate | fruity | caramel | peanutyalmondy | nougat | crispedricewafer | hard | bar | pluribus | sugarpercent | pricepercent | winpercent |
| 68 | Starburst | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.15 | 0.22 | 67.04 |
| 60 | Skittles original | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.94 | 0.22 | 63.09 |
| 66 | Sour Patch Kids | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.07 | 0.12 | 59.86 |

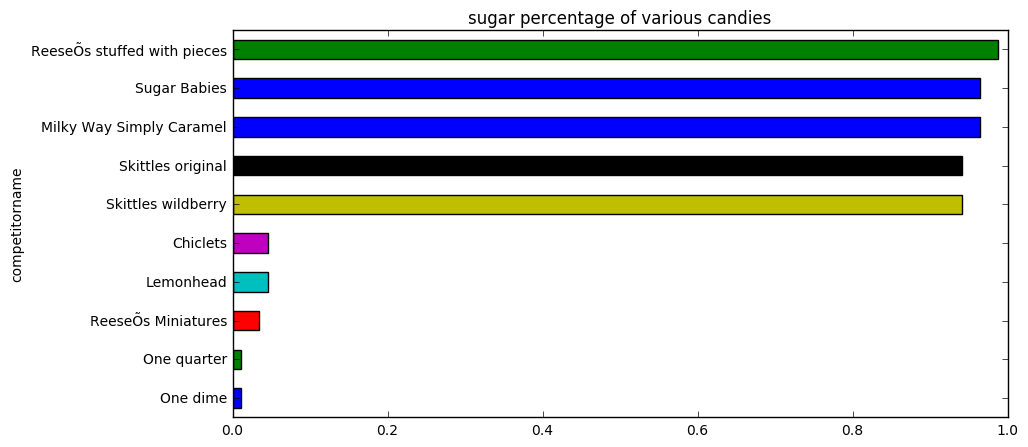
Observation 3: When non-chocolate candies are taken into consideration Starburst tops the chart.

Code details can be found at Jupyter notebook section 5

## Top 10 sugary candy

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | competitorname | chocolate | fruity | caramel | peanutyalmondy | nougat | crispedricewafer | hard | bar | pluribus | sugarpercent | pricepercent | winpercent |
| 54 | ReeseÕs stuffed with pieces | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0.988 | 0.651 | 72.887901 |
| 70 | Sugar Babies | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0.965 | 0.767 | 33.437550 |
| 38 | Milky Way Simply Caramel | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0.965 | 0.860 | 64.353340 |
| 61 | Skittles wildberry | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.941 | 0.220 | 55.103695 |
| 60 | Skittles original | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.941 | 0.220 | 63.085140 |
| 17 | Gobstopper | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0.906 | 0.453 | 46.783348 |
| 4 | Air Heads | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.906 | 0.511 | 52.341465 |
| 8 | Candy Corn | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.906 | 0.325 | 38.010963 |
| 34 | Mike & Ike | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.872 | 0.325 | 46.411716 |
| 84 | Whoppers | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0.872 | 0.848 | 49.524113 |

## Top 5 and bottom 5 candies by sugar percentage



Observation: ReeseÕs stuffed with pieces is the top sugary candy.

## Impact of price on win and sugar

Two new derived attributes are used sugarbyprice and winbyprice.

Higher sugarbyprice value means the candy is sweet as well as cheap.

Higher winbyprice value means the candy is more liked as well as cheap.

Top 10 candies are identified by sorting on the basis of winbyprice.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | competitorname | chocolate | fruity | caramel | peanutyalmondy | nougat | crispedricewafer | hard | bar | pluribus | sugarpercent | pricepercent | winpercent | sugarbyprice | winbyprice |
| 76 | Tootsie Roll Midgies | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.17 | 0.01 | 45.74 | 15.82 | 4157.89 |
| 48 | Pixie Sticks | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.09 | 0.02 | 37.72 | 4.04 | 1640.10 |
| 15 | Fruit Chews | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.13 | 0.03 | 43.09 | 3.74 | 1267.32 |
| 14 | Dum Dums | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0.73 | 0.03 | 39.46 | 21.53 | 1160.60 |
| 69 | Strawberry bon bons | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0.57 | 0.06 | 34.58 | 9.81 | 596.19 |
| 22 | HersheyÕs Kisses | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.13 | 0.09 | 55.38 | 1.37 | 595.43 |
| 66 | Sour Patch Kids | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.07 | 0.12 | 59.86 | 0.59 | 516.07 |
| 67 | Sour Patch Tricksters | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.07 | 0.12 | 52.83 | 0.59 | 455.40 |
| 57 | Root Beer Barrels | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0.73 | 0.07 | 29.70 | 10.61 | 430.49 |
| 59 | Sixlets | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.22 | 0.08 | 34.72 | 2.72 | 428.67 |

Observation 4: Tootsie Roll Midgies seems to perform better when price and win percentages are taken.

Top 10 candies are identified by sorting on the basis of sugarbyprice.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | competitorname | chocolate | fruity | caramel | peanutyalmondy | nougat | crispedricewafer | hard | bar | pluribus | sugarpercent | pricepercent | winpercent | sugarbyprice | winbyprice |
| 14 | Dum Dums | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0.73 | 0.03 | 39.46 | 21.53 | 1160.60 |
| 76 | Tootsie Roll Midgies | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.17 | 0.01 | 45.74 | 15.82 | 4157.89 |
| 57 | Root Beer Barrels | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0.73 | 0.07 | 29.70 | 10.61 | 430.49 |
| 69 | Strawberry bon bons | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0.57 | 0.06 | 34.58 | 9.81 | 596.19 |
| 50 | Red vines | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.58 | 0.12 | 37.35 | 5.01 | 321.97 |
| 61 | Skittles wildberry | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.94 | 0.22 | 55.10 | 4.28 | 250.47 |
| 60 | Skittles original | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.94 | 0.22 | 63.09 | 4.28 | 286.75 |
| 48 | Pixie Sticks | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.09 | 0.02 | 37.72 | 4.04 | 1640.10 |
| 15 | Fruit Chews | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.13 | 0.03 | 43.09 | 3.74 | 1267.32 |
| 58 | Runts | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0.87 | 0.28 | 42.85 | 3.13 | 153.58 |

Observation 5: Dum Dums seems to perform better when price and sugar percentages are considered.

## Correlation

Top 10 correlation relationship:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Feature1 | Feature2 | corr |
| 0 | chocolate | fruity | 0.741721 |
| 1 | sugarbyprice | winbyprice | 0.675094 |
| 2 | chocolate | win percent | 0.636517 |
| 3 | chocolate | bar | 0.597421 |
| 4 | bar | pluribus | 0.593409 |
| 5 | nougat | bar | 0.522976 |
| 6 | bar | pricepercent | 0.518407 |
| 7 | fruity | bar | 0.515066 |
| 8 | chocolate | pricepercent | 0.504675 |
| 9 | pricepercent | winbyprice | 0.471809 |

Observation 11: When checked in detail regarding the chocolate and fruity attribute it is seen except 1(Tootsie Pop) which is chocolate as well as fruity. Other than this candy there is no candy which is have a chocolatey and fruity combo. Either it is chocolatey or fruity or none of these attribute.

Identifying the component that make a good candy association.plot(kind='barh',x='itemsets',y='support',title=f'Most Frequently Used Composition',sort\_columns=True,figsize = (10,5),legend=False)

# Test Summary

## Predicting Winpercentage

Winpercentage can be dependent on 11 IDV variable

| **Header** | **Description** |
| --- | --- |
| chocolate | Does it contain chocolate? |
| fruity | Is it fruit flavored? |
| caramel | Is there caramel in the candy? |
| peanutyalmondy | Does it contain peanuts, peanut butter or almonds? |
| nougat | Does it contain nougat? |
| crispedricewafer | Does it contain crisped rice, wafers, or a cookie component? |
| hard | Is it a hard candy? |
| bar | Is it a candy bar? |
| pluribus | Is it one of many candies in a bag or box? |
| sugarpercent | The percentile of sugar it falls under within the data set. |
| pricepercent | The unit price percentile compared to the rest of the set. |

Linear Regression can used to predict the winpercentage and the influence of each variable

Step 1:

Split dataset into 80% train and 20% test.

Step 2: Below table shows the implementation of OLS Regression when 11 IDV are used.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **attribute** | **coef** | **std err** | **t** | **P>|t|** | **[95.0% Conf. Int.]** |
| **const** |  | 34.534 | 4.32 | 7.994 | 0 | 25.924 43.144 |
| **x1** | chocolate | 19.7481 | 3.899 | 5.065 | 0 | 11.978 27.518 |
| **x2** | fruity | 9.4223 | 3.763 | 2.504 | 0.015 | 1.923 16.922 |
| **x3** | caramel | 2.2245 | 3.657 | 0.608 | 0.545 | -5.065 9.514 |
| **x4** | peanutyalmondy | 10.0707 | 3.616 | 2.785 | 0.007 | 2.864 17.277 |
| **x5** | nougat | 0.8043 | 5.716 | 0.141 | 0.888 | -10.588 12.197 |
| **x6** | crispedricewafer | 8.919 | 5.268 | 1.693 | 0.095 | -1.580 19.418 |
| **x7** | hard | -6.1653 | 3.455 | -1.784 | 0.079 | -13.051 0.721 |
| **x8** | bar | 0.4415 | 5.061 | 0.087 | 0.931 | -9.645 10.528 |
| **x9** | pluribus | -0.8545 | 3.04 | -0.281 | 0.779 | -6.913 5.204 |
| **x10** | sugarpercent | 9.0868 | 4.659 | 1.95 | 0.055 | -0.200 18.373 |
| **x11** | pricepercent | -5.9284 | 5.513 | -1.075 | 0.286 | -16.916 5.060 |

Clearly Carmel, nougat, bar , pluribus and price percentage seems to have very little impact on win percentage as there p-value is more than 0.05 when industry standard of 5% significant level is considered.

Back propagation is used to get rid of non-significant variable one by one after considering the change in R-sqaure and adjusted R-square the below model best suits the purpose.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **attribute** | **coef** | **std err** | **t** | **P>|t|** | **[95.0% Conf. Int.]** |
| **const** |  | 34.3934 | 3.769 | 9.125 | 0 | 26.888 41.898 |
| x1 | chocolate | 19.9873 | 3.669 | 5.447 | 0 | 12.681 27.294 |
| x2 | fruity | 8.6228 | 3.566 | 2.418 | 0.018 | 1.523 15.723 |
| x3 | peanutyalmondy | 10.0435 | 3.483 | 2.883 | 0.005 | 3.108 16.979 |
| x4 | crispedricewafer | 9.4243 | 4.585 | 2.055 | 0.043 | 0.294 18.554 |
| x5 | hard | -6.0456 | 3.305 | -1.829 | 0.071 | -12.626 0.535 |
| x6 | sugarpercentage | 9.5396 | 4.377 | 2.18 | 0.032 | 0.824 18.255 |
| x7 | pricepercentage | -5.4628 | 5.12 | -1.067 | 0.289 | -15.658 4.733 |

Better display of win percentage:

|  |  |
| --- | --- |
| **CANDY TYPE** | **Value added to win percentage** |
| chocolate | 19.9873 |
| fruity | 8.6228 |
| peanutyalmondy | 10.0435 |
| crispedricewafer | 9.4243 |
| hard | -6.0456 |
| sugar percentage | 9.5396 |
| price percentage | -5.4628 |



Observation 10: Chocolate, fruity, peanut-almond, crispy and sugar percentage has positive impact on win percentage but hard could be avoided and expected with any retail product less price is more liked by the customer.

# Summary & Conclusions

## Making the perfect Candy

|  |  |
| --- | --- |
| **CANDY TYPE** | **Value added to win percentage** |
| chocolate | 19.9873 |
| fruity | 8.6228 |
| peanutyalmondy | 10.0435 |
| crispedricewafer | 9.4243 |
| hard | -6.0456 |
| sugar percentage | 9.5396 |
| price percentage | -5.4628 |

The table simplifies that “coef.” column. If a hypothetical candy had chocolate in it, we’d expect its win percentage to rise by about 20 points. If it’s fruity, we’d expect it to rise by 9. If it had nuts, we’d also expect its win percentage to rise by 10, with wafers or crisped rice rising by 9. And nougat and caramel don’t bring a ton to the table. A candy being hard — like a lollipop or jawbreaker — actually knocks about 6 points off its win percentage. Whether it’s in bar form or a bunch of little candies makes no major difference.

A candy can be one or two or all of these things: A 100 Grand which is chocolate (+20), crispedricewafer (+9), not hard and a sugar percent of 0.73 that we’d expect to have in the ballpark of 66 percent win rate, and it does in fact have a 67 percent win rate.

--- pie chart or graph of coeficents

## Ingredient of perfect candy

We’ve got to have chocolate — the win percentage of contenders containing chocolate was about 11 points higher than the average contender and 19 points higher than contenders that did not have chocolate. The same goes for both crispiness and nuttiness: Entrants with peanuts or almonds had win percentages about 13 points higher than the average contender, and ones with crispy wafers or puffed rice were nearly 16 points higher than average. Nougat and caramel are net positives for sure, so throw them in.

Now we get to some major qualitative components. Candies in bar form generally had a higher overall win rate than those in pieces, so we’ll want a bar. Yes, fruitiness can be fine, but things that had a fruity taste had a win percentage 11 points lower than those that did not. This is due to the near mutual exclusivity of fruity flavors and chocolate, with Tootsie Pops being the sole exception.[3](https://fivethirtyeight.com/features/the-ultimate-halloween-candy-power-ranking/#fn-3)

So, in the end, the best Frankencandy has the chocolate of a Hershey bar, the nougat of a Baby Ruth, the caramel of a Milky Way, the peanut butter of a Reese’s Cup and the wafer of a Twix, and it’s assembled in a castle looming over an Eastern European village. Or if you’re trying to make this at home: Maybe take a Twix bar and smush it on a PayDay, or roll a Snickers around in rice crispies.

## Best Candy with ingredient

Best candy with chocolate- Reese’s Peanut Butter cup

Best candy with fruit- Starburst

Best candy with nuts- Reese’s Peanut Butter cup

Best Candy with crispedricewafer- Twix

Best Candy which is not hard - Reese’s Peanut Butter cup

Candy with more sugar- Reese’s stuffed with pieces

Cheapest Candy- Tootsie Roll Midgies

In a perfect world we can make a candy with chocolate, nuts and soft as like Reese’s Peanut Butter cup, as wafery as Twix, as fruity as Starburst and price should be like Roll Midges.

Tabular column in picture

## Recommendations

# Glossary

|  |  |  |
| --- | --- | --- |
| Ref | Acronym / Abbreviation | Definition |
| **1** | CAM | Conditional Access Module |
| **2** | MP | Mass Production |
| **3** | DE | Germany |
| **4** | UM | UnityMedia |
| **5** | ITC | International Test Center |
| **6** | CSC | Chipeset Secure Channel |
| **7** | DRM | Digital Rights Management |
| **8** | ECM | Entitled Control Message |
| **9** | STU-B | Stuttgart B signal-lab signal |
| **10** | BW | Baden-Württemberg |
| **11** | CA | Conditional Access |
| **12** | CI+ | Common Interface |
| **13** | MPEG | Moving Picture Experts Group |
| **14** | TS | Transport Stream |
| **15** | OTA | Over The Air |
| **16** | FSK | Freiwillige Selbstkontrolle der Filmwirtschaft |

Table 11: Glossary table

Things to do:

## **Day 2: Plot a Histogram**

Now, it's time to plot a histogram of the different candies by their sugar percentiles.

## **Day 3: Perform a T-test**

I'll be performing a t-test to see if there's a difference in sugar percentiles in candies that contain chocolate vs. those that don't.

## **Day 4: Draw a Bar Chart**

I'm going to be drawing a bar chart of the number of candies with chocolate vs. no chocolate.

## **Day 5: Perform a Chi-Square Test**

I'm going to run a chi-squared test to see if there is a relationship between a candy having chocolate and a candy having caramel.

<https://www.kaggle.com/nadintamer/introductory-data-challenge-r>

*# 1) Can you predict if a candy is chocolate or not based on its other features?*

*# 2) Fit a GLM model of the appropriate family.*

*# 3) Plot diagnostic plots for your model. Does it seem like your model is a good fit for your data? Are the residuals normally distributed (no patterns in the first plot and the points in the second plot are all in a line)? Are there any influential outliers?*

*# 4) Plot your two variables & use "geom\_smooth" and the appropriate family to fit and plot a model*

Tools used section in the top- software, algorithms, programming language, which os