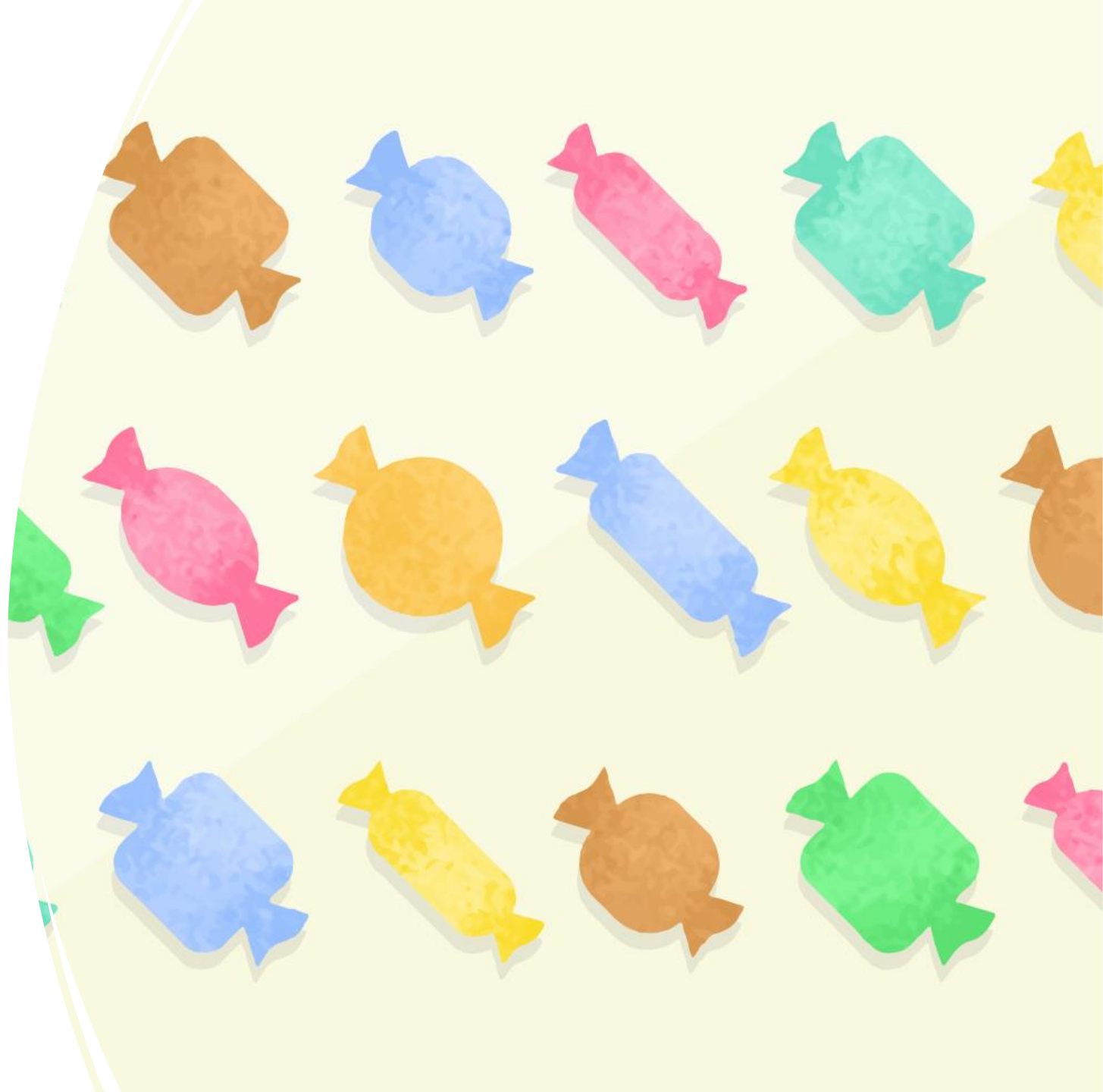# CANDY RECOMMENDATION

Anežka Lhotáková

MANAGEMENT SUMMARY

PROBLEM

The Lidl purchasing group wants to expand our candy offering. The idea is to create a brand-new product.

The market research data are available, and a data-driven approach is required.

The task is to **present a recommendation of new candy** product, which is driven from the data as a **statistical solution.**

SOLUTION

Data explanation via descriptive statistics

Designed possible approaches and strategies

Applied multiple linear regression

Proposed a solution

REVIEW

The data lack further context such as quality of the ingredients, the earnings of the product, the price ad sugar content were relativized etc.

The results (including comparison) are simplified due to already mentioned lack of context and imperfect translation between real and predicted value.

PROCEDURE & ASSUMPTIONS

MULTIPLE LINEAR REGRESSION

is a statistical technique that uses **several explanatory variables** to predict the outcome of a response variable.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_n x_{in} + \epsilon$$

The model performance is evaluated based on **AIC and R^2 statistics**. Also **mean squared error** is observed.

ASSUMPTIONS

- It is more crucial to correctly describe the behaviour for the **more popular candies** than for the less liked.

- Recommendation should consider **trends** published in the candy market [research](#).

    > rising demand for sugar-free candies
    > decline in gums and mint-candy

- Let's assume the **quality of each ingredient in the characteristics is equal**. In other words, the difference in product's prices does not come from different qualities of chocolate, nougat, etc.

DATA HIGHLIGHTS

The **DATA** consist of 85 observations, which include **attributes for each candy along with its ranking**.

Data are available [here](#).

| variable	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.
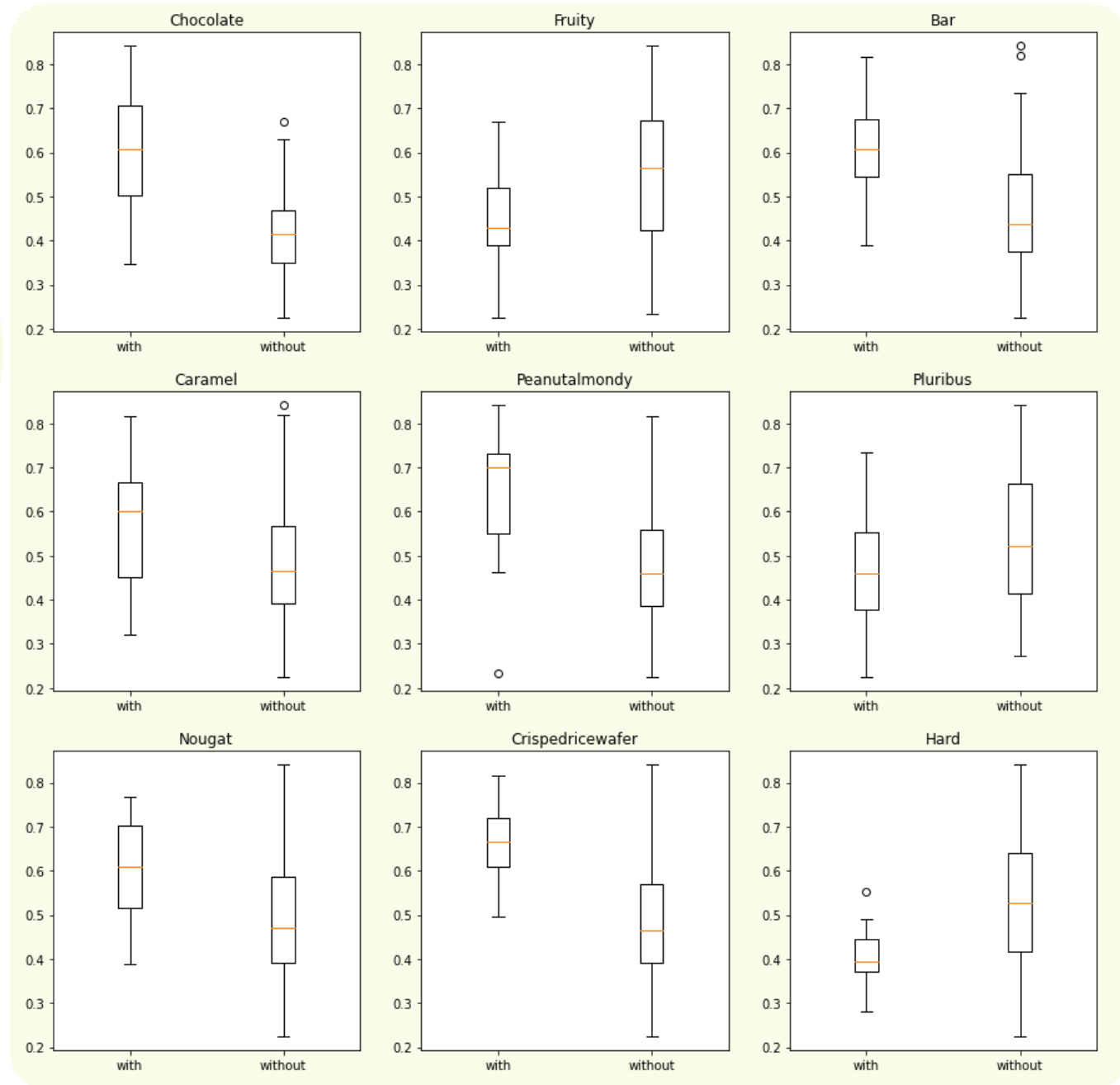
Fig 1.: Boxplot of each ingredient. Axis y represents target variable „winpercent'.

MODEL

While building a multiple linear regression model, the dataset was modified in order to put **more weight to the best ranked candies**. By doing so, we can **bring more attention to the well-ranked observations** and the linear fit will be **well adjusted to the prediction of good candy,** at the expense of the accuracy in the less liked candies.
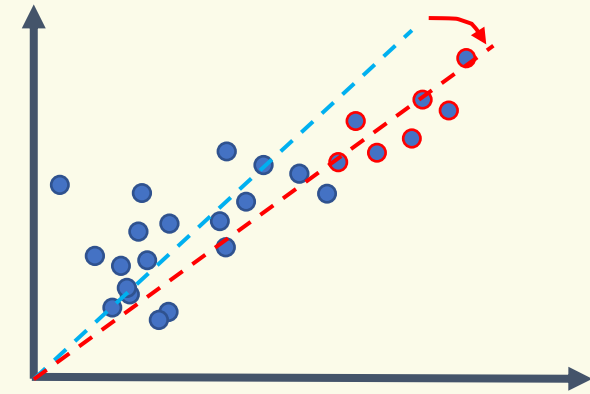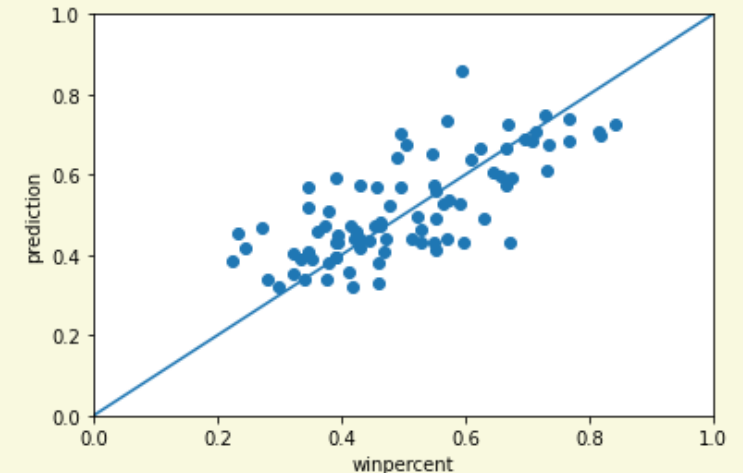

FINAL MODEL

$$winpercent = 0.3618 + 0.2220 \cdot chocolate + 0.1003 \cdot fruity + 0.0306 \cdot caramel + 0.1312 \cdot peanutyalmondy + 0.0215 \cdot nougat + 0.1379 \cdot crispedricewafer - 0.0656 \cdot hard - 0.0264 \cdot bar - 0.0275 \cdot pluribus + 0.0774 \cdot sugarpercent - 0.0676 \cdot pricepercent$$

$R^2 = 0.641$

$R^2_{adj} = 0.596$

MSE = 0.015

CANDY RECOMMENDATION

Pure data-driven candy recommendation

Choose the positive effect as 1 and negative effect as 0.

chocolate	fruity	caramel	peanutyal.	nougat	crispedrice.	hard	bar	pluribus	sugarpercent	pricepercent	winpercent
1	1	1	1	1	1	0	0	0	1	0	1.00

Low-sugar candy recommendation

Choose the positive effect as 1 and negative effect as 0. Let's set the sugarpercent as the value of 25th percentile in the original data and pricepercent as the mean value.

chocolate	fruity	caramel	peanutyal.	nougat	crispedrice.	hard	bar	pluribus	sugarpercent	pricepercent	winpercent
1	1	1	1	1	1	0	0	0	0.22	0.46	0.91

Affordable candy recommendation

Choose the positive effect as 1 and negative effect as 0. Let's set the pricepercent as the value of 25th percentile in the original data and sugarpercent as the mean value.

chocolate	fruity	caramel	peanutyal.	nougat	crispedrice.	hard	bar	pluribus	sugarpercent	pricepercent	winpercent
1	1	1	1	1	1	0	0	0	0.46	0.25	0.94

THANK YOU FOR YOUR ATTENTION