



Lidl Analytics Consulting Fallstudie

Case Study Präsentation – Jonathan Ulrich

Berlin, 01.09.2020

Die beliebteste Süßigkeit sind Schokoladenriegel mit Karamell kombiniert mit zwei weiteren Zutaten



Erkenntnisse

Beliebteste Eigenschaften



Schokolade



Erdnuss /
Mandel



Reiswaffel

Beliebteste Gruppe

- Ø Zuckeranteil: 62%
- Ø Preis: 0,31 €
- Ø Zutaten: 2
- Enthalten alle Schokolade
- Sind alle Riegel und einzeln verpackt
- Kombination mit anderen Erdnuss/Mandel, Reiswaffel, Karamell



Handlungs- empfehlungen

Durchführung weiterer Marktforschung:

- Fokus auf beliebteste Eigenschaften und Gruppen
- **Ermittlung von spezifischen Produktmerkmalen:** Qualität, Gewicht, Menge an Einzelnen Zutaten, Verpackungseigenschaften
- **Ermittlung von spezifischen Kundenmerkmalen:** Alter, Vorlieben, Einkaufsmenge, Zahlungsbereitschaft
- **Ermittlung von Marketingmerkmalen:** Werbeintensität, Werbemittel, Werbeausgaben
- **Ermittlung von Preis- und Kostenmerkmalen:** Kosten der Zutaten, Kosten der Herstellung, Zusammenhänge zwischen Preis und Verkaufsmenge, Gewinnmargen



Zielstellung d. Analyse

- Erweiterung des Süßigkeiten Sortiments
- Ableitung von Eigenschaften einer neuen Süßigkeit auf Basis von Marktstudie



Marktstudie

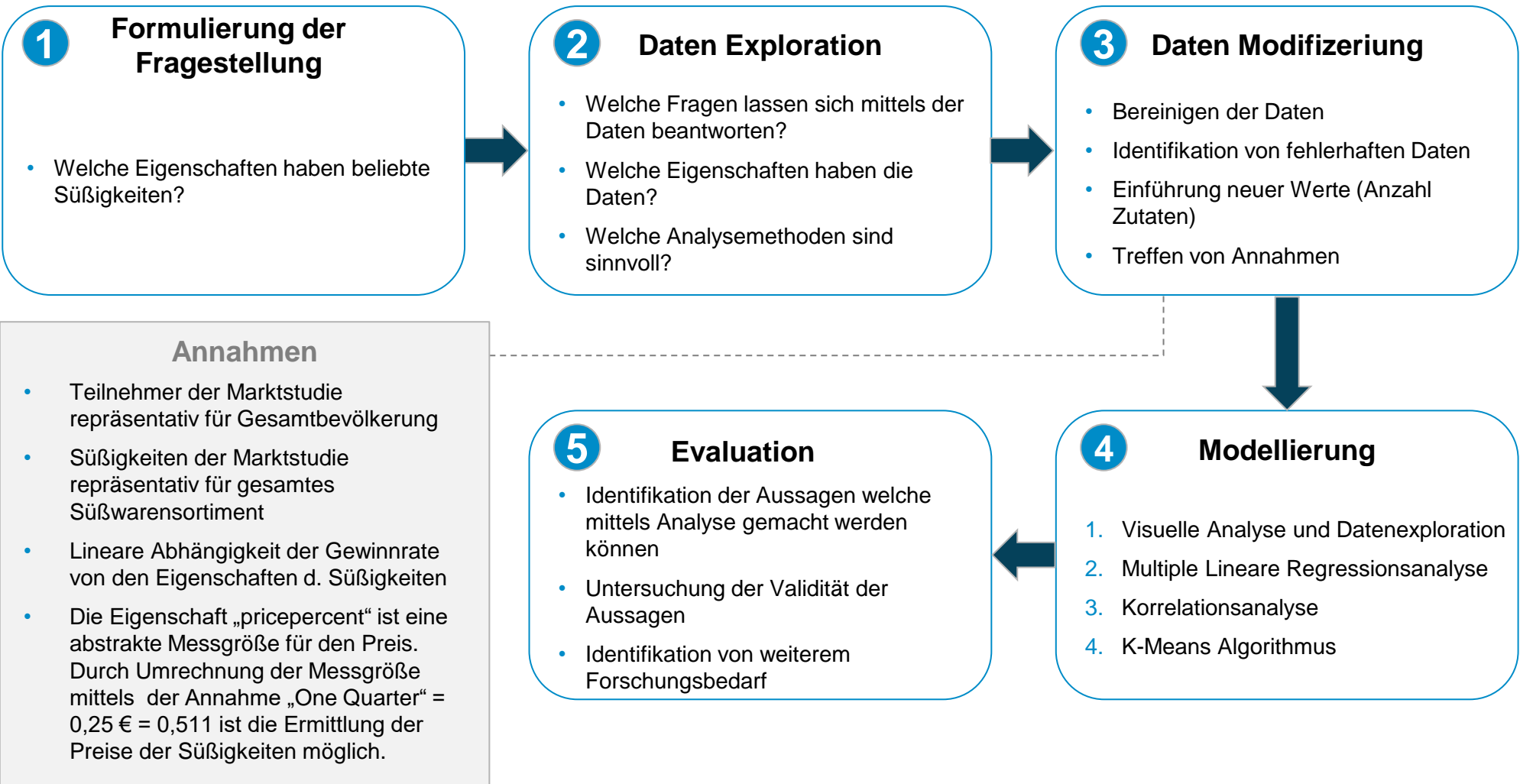
- Probanden entscheiden sich für eine von zwei Süßigkeiten bei zufälligem Matchup
- 83 Süßigkeiten in 269.000 Matchups
- 8361 IP-Adressen als Probanden



Analyse Methoden

- Visuelle Datenanalyse
- Lineare Regressionsanalyse
- Korrelationsanalyse
- K-Means Algorithmus

Mittels statistischer Modelle wurden die beliebtesten Eigenschaften und die beliebteste Gruppe von Süßigkeiten ermittelt



Visuelle Analyse und Datenexploration ermöglichen ein tieferes Verständnis der Daten und liefert Erkenntnisse



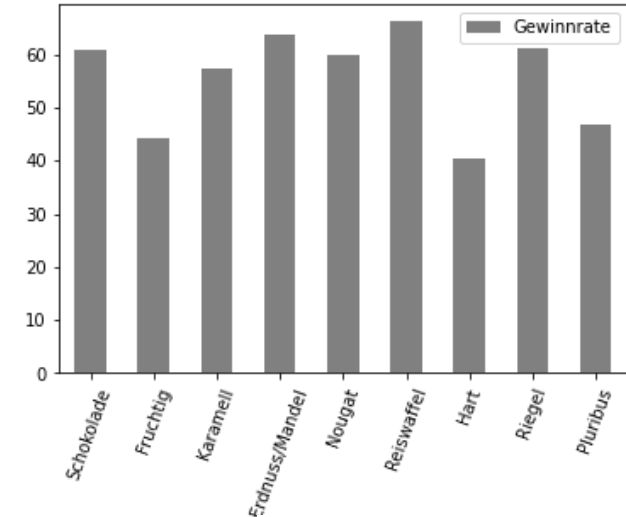
METHODIK

(1) **Visuelle Analyse und Datenexploration:** Explorieren des Datensatzes, aufgreifen Nennenswerter Informationen und **Visualisierung der Daten.**

ERKENNTNISSE

- Höchste **durchschnittliche Gewinnraten*** haben die Eigenschaften **Reiswaffel (66%)**, **Erdnuss/Mandel (63%)** und **Riegel (61%)**
- Die am **meisten vertretene Eigenschaft** ist **Pluribus (44%)**, **Fruchtig (38%)** und **Schokolade (37%)**
- Die **3 Süßigkeiten mit der höchsten Gewinnrate** haben sehr ähnliche Eigenschaften und eine **Preisspanne von 0,30 €**

Durchschnittliche Gewinnrate (%) je Eigenschaft



Beliebteste Süßigkeiten

1 Reese's Peanut Butter Cup

- **Gewinnrate:** 84%
- **Zutaten:** Schokolade und Erdnuss/Mandel
- **Zuckeranteil:** 72%
- **Preis:** 0,32 €
- Kein Riegel

2 Reese's Miniatures

- **Gewinnrate:** 82%
- **Zutaten:** Schokolade und Erdnuss/Mandel
- **Zuckeranteil:** 3,4%
- **Preis:** 0,14 €
- Kein Riegel

3 Twix

- **Gewinnrate:** 82%
- **Zutaten:** Schokolade, Caramel, Reiswaffel
- **Zuckeranteil:** 54,6%
- **Preis:** 0,44 €
- Riegel

Anzahl an Süßigkeiten mit bestimmten Eigenschaften



*Gewinnrate: Bei einem Matchup wurden Probanden jeweils 2 Süßigkeiten vorgelegt, von denen sie sich für eine entscheiden musste. Das Verhältnis von gewonnenen Matchups zu gesamten Matchups der Süßigkeit ergibt die Gewinnrate.

Lineare Regressionsanalyse kombiniert mit Korrelationsanalyse gibt Einblicke in die Signifikanz von Eigenschaften



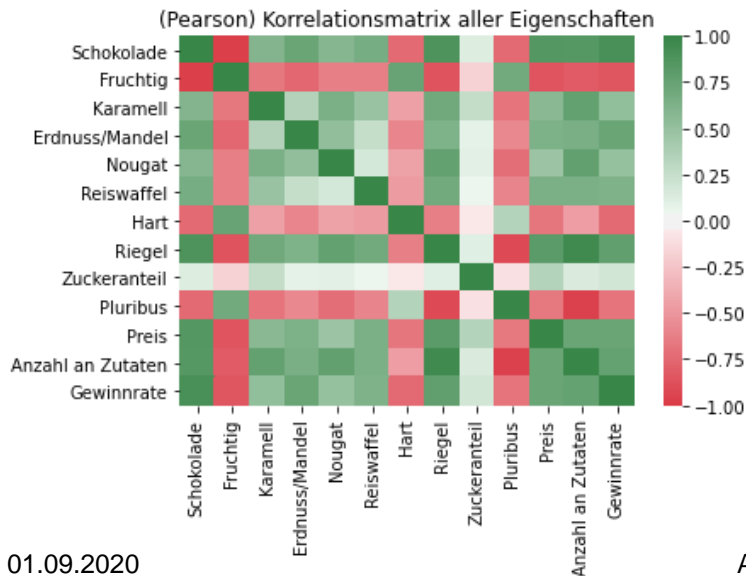
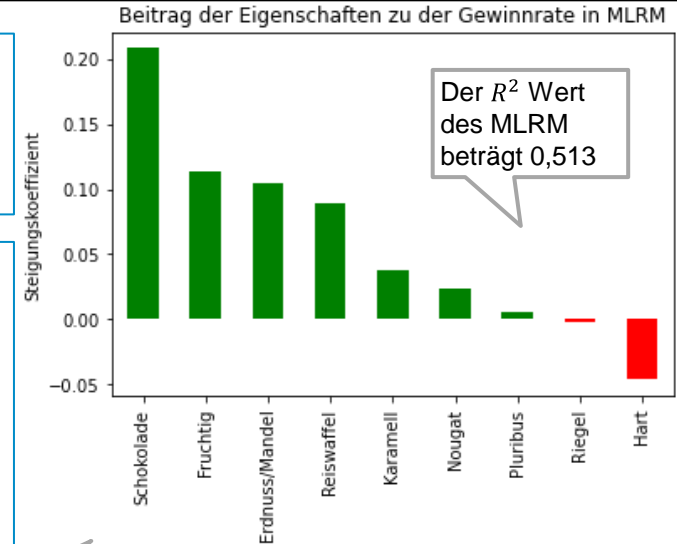
METHODIK

(2) **Multiple Lineares Regressions Modell (MLRM)**: Erstellung eines MLRM mittels binärer Eigenschaften. Interpretation der Steigungskoeffizient d. Eigenschaft in linearer Gleichung als Beitrag der einzelnen Eigenschaften zu der Gewinnrate.

(3) **Pearson Korrelation**: Ermittlung der Pearson Korrelation aller Eigenschaften.

ERKENNTNISSE

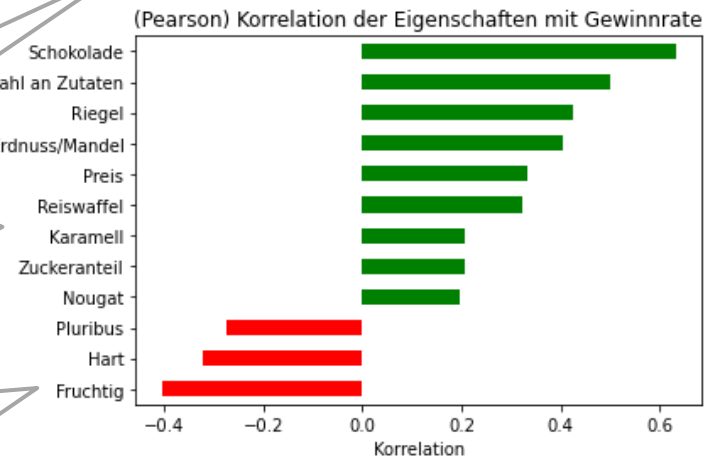
- **Schokolade** hat die **höchste Korrelation mit der Gewinnrate** und hat **größten MLRM Koeffizienten**
- **Schokolade korreliert stark mit Preis** was auf einen hohen „Added Value“ oder auch Kosten hindeuten kann
- **Schokolade, Karamell, Erdnussmandel, Nougat** und **Reiswaffel steigern** die **Gewinnrate im MLRM** und haben **eine positive Korrelation** mit der **Gewinnrate** und zueinander
- **Mehrfachverpackung (Pluribus)** sowie **Harte** und **Fruchtige Süßigkeiten** haben **negative Korrelation** mit der **Gewinnrate**. Die Eigenschaft **Hart** **senkt Gewinnrate in MLRM**.
- **Zuckeranteil** hat **geringste Korrelation** mit **Gewinnrate**



In dem MLRM werden **nur binäre Eigenschaften berücksichtigt** um die Werte nicht zu Verfälschen

Das MLRM muss in **Zusammenhang mit der Korrelation betrachtet werden**, um eine Aussage treffen zu können

Bsp.: Obwohl die Eigenschaft „Fruchtig“ einen positiven Beitrag zu der Gewinnrate leistet, korreliert sie negativ mit der Gewinnrate aufgrund negativer Kovarianz



Eine K-Means Clusteranalyse ermöglicht die Identifizierung der Gruppe mit der größten Gewinnrate und dessen Eigenschaften



METHODIK

(4) K-Means Algorithmus: Teilt Datensatz in eine K Anzahl von Gruppen auf Basis der Ähnlichkeit ihrer Eigenschaften. Der K-Means Algorithmus ist ein Iterativer Algorithmus und eine Unsupervised Machine-Learning Methode.

ERKENNTNISSE

- Silhouette Score impliziert beste Trennbarkeit von Cluster bei einer Anzahl von 3 oder 6 Cluster
- Elbogen-Methode angewendet auf Kumulierte Entfernung der Zentroiden und Anzahl K-Means Cluster impliziert höchste Verbesserung (größte Verringerung des Kurvenabfalls) bei 2-3 Cluster
- Bei Anwendung der K-Means Clusteranalyse auf K = 6 ist das Cluster mit höchster Gewinnrate gut differenzierbar

➔ Schokoladenriegel in Kombination mit Reiswaffel, Erdnuss/Mandel und Karamell

Cluster mit höchster Gewinnrate

K* = 3

Cluster Größe: 18

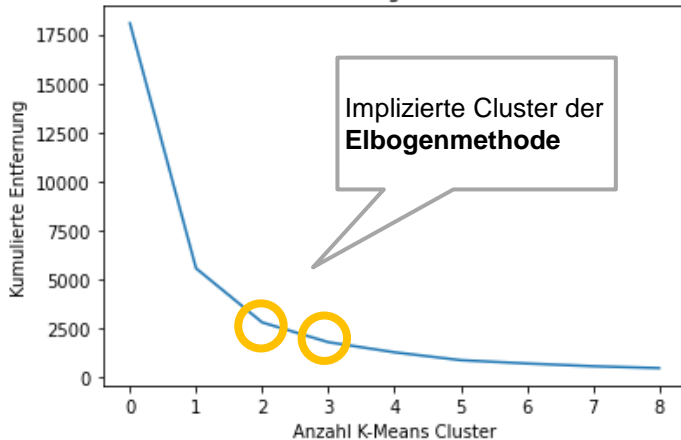
- Ø **Gewinnrate:** 63%
- Ø **Zuckeranteil:** 54%
- Ø **Preis:** 0,34 €
- Ø **Zutaten:** 3 – 4
- **Merkmal:** Einzelne Verpackung aller Süßigkeiten des Cluster
- Hoher Anteil mit Eigenschaft Schokolade und Riegel
- Alle anderen Eigenschaften sind in Cluster enthalten

K* = 6

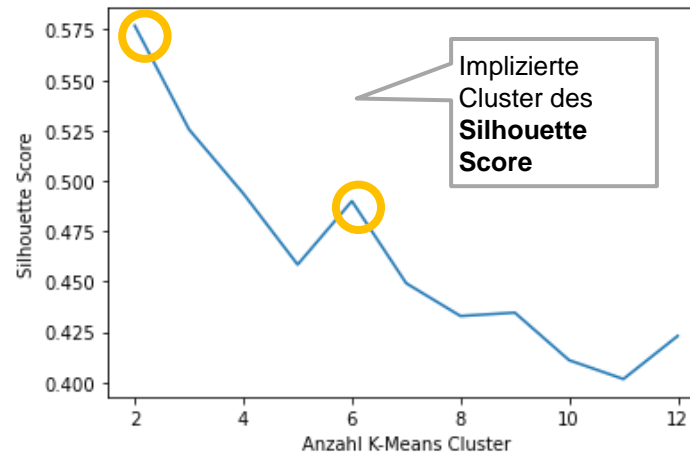
Cluster Größe: 9

- Ø **Gewinnrate:** 69%
- Ø **Zuckeranteil:** 62%
- Ø **Preis:** 0,31 €
- Ø **Zutaten:** 2
- Enthalten alle Schokolade
- Sind alle Riegel
- Enthalten nicht Fruchtig, Nougat und Hart
- Kombination mit anderen Eigenschaften

Kumulierte Entfernung zwischen Zentroiden



Silhouette Score



*Gesamte Anzahl an Cluster in K-Means Clusteranalyse