

Inhalt

| | | |
|-----|--|---|
| 1 | Anwendungsfall und Datensatz | 2 |
| 2 | Organisation und einzureichende Dokumente | 2 |
| 3 | Aufgabenstellungen | 3 |
| 3.1 | Explorative Datenanalyse für Marketing und Disposition | 3 |
| 3.2 | Konkrete Fragen der Stakeholder | 3 |
| 3.3 | Erstellung eines prototypischen „Business Dashboards“ | 4 |

1 Anwendungsfall und Datensatz

Ihre Gruppe stellt ein Team einer Data Science-Beratungsgesellschaft dar und Sie beraten einen **Online Lebensmittel-Lieferanten** aus Kalifornien. Sie haben Aufgaben in folgenden Teilgebieten umzusetzen (Details siehe unten):

1. Vorverarbeitung der Daten sowie freie **Explorative Datenanalyse** zur leicht nachvollziehbaren Erläuterung der Daten für unterschiedliche Stakeholder
2. Beantwortung konkreter Fragen sowie Identifikation/Prüfung von Zusammenhängen unter Verwendung **statistischer Methoden** in Kombination mit Visualisierungen
3. Prototypische Umsetzung eines visuellen „**Business Dashboards**“ mit PowerBI

Die Nutzer des Lebensmittel-Lieferdienstes können online bestellen und bekommen die gewünschten Produkte nach Hause geliefert. Der Datensatz besteht aus mehreren Teildatensätzen, die sowohl Stammdaten (bspw. Produktinformationen) sowie Bewegungsdaten (Nutzerbestellungen sowie zugeordnete Bestellpositionen). Eine Kundenbestellung erstreckt sich somit ggf. über mehrere Zeilen, bspw. order_id 31 enthielt 3 Produkte (196, 9434, 19660) und wurde von user_id 21425 getätigt.

Hinweise zu den Daten

- add_to_cart_order: Reihenfolge, in der die Produkte in den Warenkorb gelegt wurden
- days_since_prior_order: Wie viele Tage ist die vorherige Bestellung des Nutzers her.¹
- order_dow: Wochentag der Bestellung
- order_hour_of_day: Stunde der Bestellung
- aisle_id: Informationen zu dem „Gang“, in dem die Produkte in einem Supermarkt stehen
- department_id: Zuordnung der Produkte zu Kategorien
- tip: True, wenn beim Empfang der Lieferung Trinkgeld gegeben wurde

2 Organisation und einzureichende Dokumente

Die Übung ist durch Projektgruppen bestehend aus 4 Studierenden umzusetzen.

- **Abgabe Notebooks: Montag, 20.01.2025, 8 Uhr**
Über ILIAS müssen die Notebooks und PowerBI-Projekt (Hinweise siehe Kapitel 3) als zip-Datei eingereicht werden.
- **Abgabe Screencasts: Montag, 20.01.2025, 14 Uhr**
Über ILIAS ist eine 10-minütiger Screencast einzureichen, in dem die Notebooks und die PowerBI-Lösung „technisch“ erläutert werden. Im Gegensatz zur Präsentation sind hier nicht die Stakeholder die Zielgruppe, sondern Data Scientists aus dem Unternehmen, die Ihren Code übernehmen und die angewendeten Methoden nachvollziehen sollen. Erläutern Sie hier also Hintergründe, technische und statistische Details, verwendete Methoden etc.
- **Präsentation (anhand der Notebooks, keine Powerpoint!): Dienstag, 21.01.2025, ab 14:00 Uhr** (pro Gruppe 15 Minuten Präsentation + Diskussion).
Die Präsentation soll anhand der Notebooks erfolgen. Hier können einige „technische“ Details aber ausgeblendet werden, d.h. es sollen die Inhalte für die beiden Stakeholder-Gruppen leicht verständlich erläutert werden („Pitch“ Ihrer Ergebnisse, insb. durch Visualisierung und Diskussion von aggregierten Ergebnissen)

¹ Der Wert kann „n/a“ sein, wenn es sich um die erste Bestellung eines Nutzers (im zugrundeliegenden Zeitraum) handelt.

3 Aufgabenstellungen

3.1 Explorative Datenanalyse für Marketing und Disposition

Als erstes wird eine explorative Analyse des Datensatzes für die Marketing- und Dispositionsabteilung erwartet. Ihre Aufgabe ist eine zielgruppengerechte Aufbereitung zweier „Data Stories“ (je Abteilung eine) durch die die Stakeholder in der Lage sein sollen, das Geschäftsfeld besser zu durchdringen und Rückschlüsse für die eigenen betrieblichen Handlungsbereiche aus den Daten zu ziehen. Die Stakeholder erwarten hierfür ein **Jupyter-Notebook „EDA für Marketing und Disposition“**.

Für die Analyse sind zunächst einige Vorverarbeitungsschritte nötig. Nutzen Sie Ihre Kenntnisse in Python (bzw. den besprochenen Bibliotheken), um die Daten grundsätzlich zu verstehen, zu kombinieren und ggf. anzureichern.

Die **Marketing-Abteilung** erhofft sich Erkenntnisse, die zur Optimierung des Produktkatalogs und der Produktplatzierung auf der statischen Website verwendet werden können. Wenn viele Kunden bspw. „Flour“ und „Eggs“ zusammen kaufen, könnten beide Produkte nebeneinander auf der Website platziert werden. Der Zusammenhang muss nicht zwingend wie im Beispiel auf Produktebene hergestellt werden, sondern kann auch auf der Ebene Kategorie („department“) oder Gang („aisle“) erfolgen.

Die **Disposition** erwartet von Ihrem Team Erkenntnisse über Unterschiede im Kundenverhalten bezüglich des Trinkgelds. So soll vermieden werden, dass manche Mitarbeiter des Unternehmens regelmäßig mehr Trinkgeld erhalten als andere (bspw. durch eine geänderte Zuordnung der Kunden zu den Kurieren).

In den jeweiligen Analysen sind Sie sehr frei. Nutzen Sie Ihre Fähigkeiten zur tabellarischen und/oder grafischen Aufbereitung der Daten sowie die statistischen Methoden, die Sie im Laufe des Studiums erlernt haben. Strukturieren Sie das Notebook nachvollziehbar und bereiten Sie die Erkenntnisse zielgruppengerecht auf. Belegen Sie dabei Ihre Schlussfolgerungen anhand der Daten in angemessenem Detailgrad. Die Kenntnis gängiger statistischer Kennzahlen und ihrer Interpretation können Sie als bekannt voraussetzen.²

3.2 Konkrete Fragen der Stakeholder

Die Stakeholder haben folgende konkrete Aufgaben bzw. Fragen für Ihren Beratungsauftrag formuliert:

Frage 1: Hängt die Wahrscheinlichkeit, dass bei einer Bestellung Trinkgeld gegeben wird, vom Trinkgeldverhalten bei früheren Bestellungen desselben Bestellers ab?

- a) Gibt es einen Zusammenhang bezüglich der vorhergehenden Bestellung?
- b) Gibt es einen Zusammenhang bezüglich der vor-vorhergehenden Bestellung?
- c) Liefert das Trinkgeldverhalten der vor-vorhergehende Bestellung Informationen auch über das hinaus, was bereits aus der vorhergehenden Bestellung abgelesen werden kann?

Frage 2:

- a) Gibt es einen Zusammenhang zwischen dem Trinkgeldverhalten und den Departments, aus denen bestellt wird?
- b) Gibt es einzelne Produkte, die die Trinkgeldwahrscheinlichkeit besonders stark beeinflussen? (Ermitteln Sie die TOP10 und FLOP10)

² Sie dürfen die Jupyter-Visualisierungen oder -Tabellen auch in einem anderen Tool wie Excel/Visio/PowerPoint oder ähnlichem nachbearbeiten, um Aspekte hervorzuheben (sofern die Methode der Hervorhebung in der Veranstaltung nicht behandelt wurde). Bitte binden Sie diese Ergebnisse als Abbildungen in die Notebooks ein und versehen Sie diese mit einem Kommentar, bspw. „Erstellt in Jupyter-Notebook und nachbearbeitet mit <Tool-Name>“.

- c) Besteht der Zusammenhang aus (b) zusätzlich zu dem aus (a), d.h. beeinflusst das Vorhandensein der Produkte aus (b) die Trinkgeldwahrscheinlichkeit auch über das hinaus, was durch das Vorhandensein der entsprechenden Departments schon absehbar ist?

Frage 3: Gibt es einen Zusammenhang zwischen dem Trinkgeldverhalten und der Tageszeit, dem Wochentag, oder dem Zeitabstand zur vorhergehenden Bestellung?

Beantworten Sie diese Fragen in einem **Jupyter-Notebook** „Fragen der Stakeholder“. Auch in diesem Bereich sind Sie frei in der Gestaltung, insb. der (visuellen) Aufbereitung Ihrer Erkenntnisse. Sofern Fragestellungen Interpretationsspielraum lassen, legen Sie sie im Sinne der Stakeholder sinnvoll aus, und stellen Sie sicher, dass die Stakeholder den Ergebnissen entnehmen können, wie Sie die Frage ausgelegt haben.

Belegen Sie gegenüber den Stakeholdern in geeigneter Weise das Vorhandensein der von Ihnen festgestellten Zusammenhänge. Die Stakeholder interessieren sich vermutlich nicht nur dafür, ob ein Zusammenhang besteht, sondern auch, wie er quantitativ aussieht. Ergänzen Sie die Analysen, sofern möglich und Ihrer Meinung nach geeignet, durch passende Visualisierungen. Diese sollten sich möglichst gut zur Kommunikation Ihrer Erkenntnisse eignen, so dass die in Analytics „ungeschulten Stakeholder“ die Ergebnisse leicht nachvollziehen können.

Soweit dies nicht offensichtlich ist, machen Sie auch nachvollziehbar, ob die Zusammenhänge signifikant sind bzw. inwieweit sie einfach Zufallsprodukte sein könnten.

3.3 Erstellung eines prototypischen „Business Dashboards“

Setzen Sie prototypisch mit **PowerBI** ein Projekt „**Business Dashboard**“ um. Wählen Sie dazu zunächst einen der beiden oben genannten Stakeholder als Nutzer dieses Dashboards aus. Definieren Sie stichpunktartig, welchen Informationsbedarf Sie für den Stakeholder durch das Dashboard decken wollen. Notieren Sie den gewählten Stakeholder und die Informationsbedarfs-Definition bitte formlos in einer **PDF-Datei** „**Business Dashboard - Annahmen**“. Laden Sie die Rohdaten in Power BI und erzeugen Sie auf maximal 2 Seiten entsprechende Visualisierungen, Tabellen, Filter, passende Kennzahlen und/oder Hierarchien etc. Sie dürfen dabei auf den Erkenntnissen aus den o.g. Analysen sowie entsprechenden Visualisierungen in Python aufbauen. Achten Sie in diesem Teil bitte stark auf die Übersichtlichkeit!