

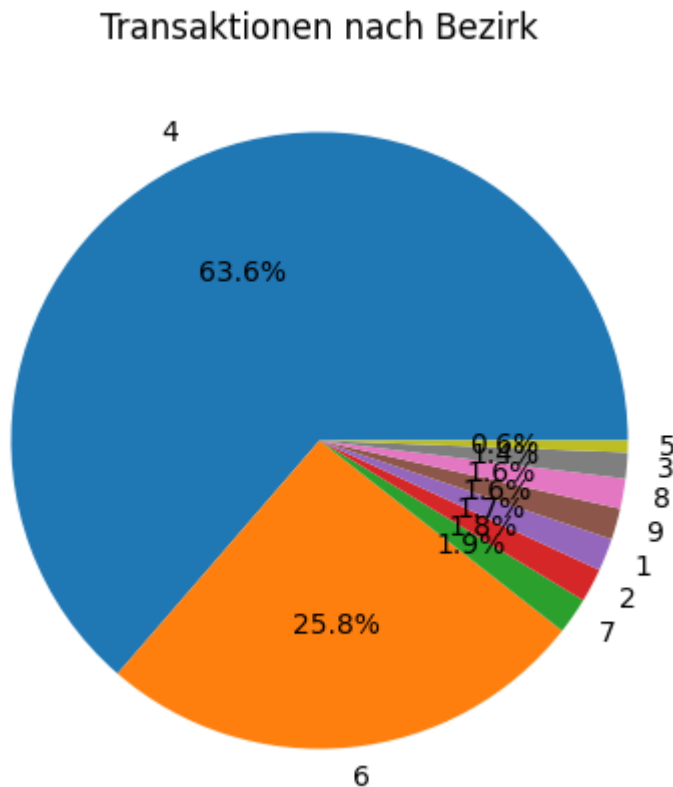
Data Science Hausarbeit

INF21D - Grundlagen Data Science WS 23/24 - Gruppe 9

Philipp Eichwald, Joshua Luipold, Luis Neumeier

1. In welchem Bezirk befindet sich die Buchhandlung?

Unter der Annahme, dass die meisten Kunden in der Nähe der Buchhandlung wohnen, können wir durch die Analyse der Kundenadressen den Bezirk der Buchhandlung herausfinden.

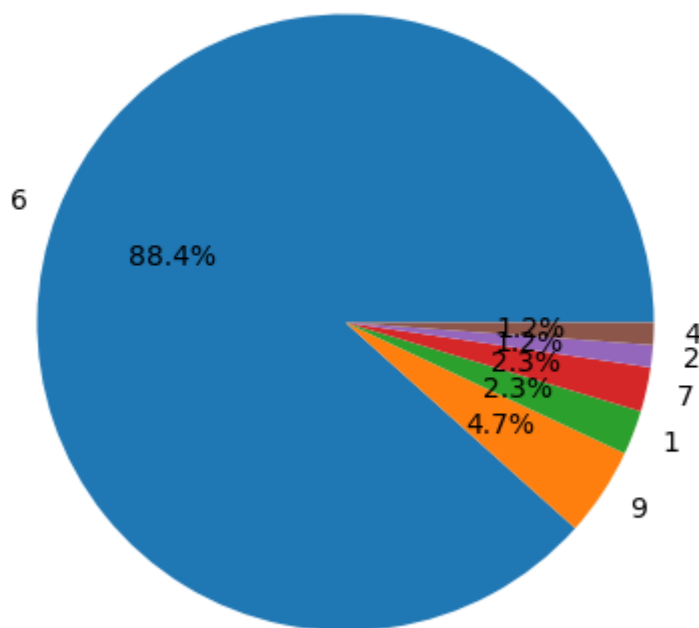


ANTWORT: Die Analyse liegt nahe, dass die **Buchhandlung in Bezirk 4** liegt, da dies in über 63% der Transaktionen der Wohnort der Kunden ist.

2. In welchem Bezirk wohnt die Mehrzahl der Studenten?

Dazu betrachten wir nur die Kunden, die als Beruf "Student" angegeben haben und sortieren sie nach Wohnort.

Studenten nach Bezirk



ANTWORT: Die Analyse zeigt, dass über **88% der Studenten im Bezirk 6** wohnt.

3. Welche Fakultät befindet sich in der Außenstelle der Universität in deinem Viertel?

Dazu können wir die gekauften Bücher der Studenten analysieren.

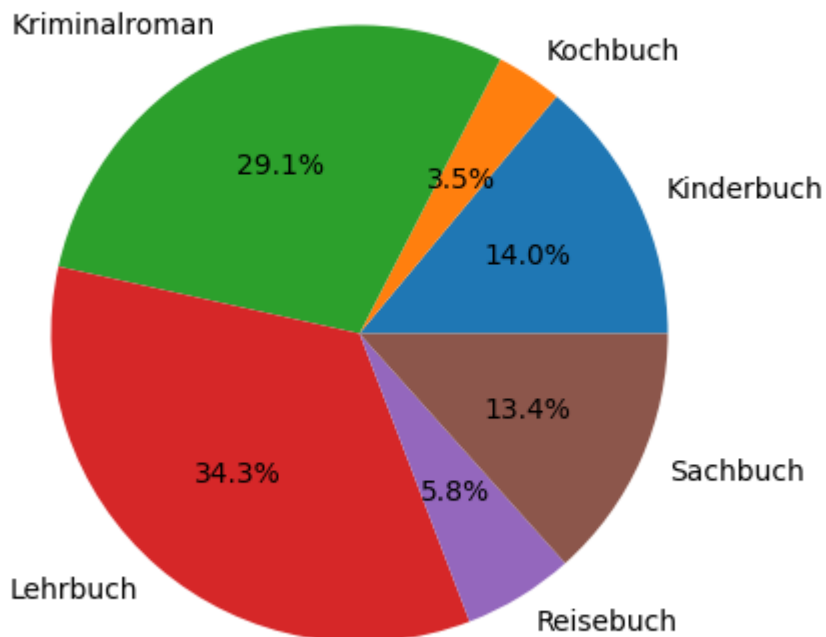
Titel	
Fiskalstrafrecht	98
Grundriss des deutschen Steuerrechts	60
Die Organschaft im Steuerrecht mit Fallbeispielen	52
Erbrecht für die steuerberatenden Beruf	52
Anwalts- und Steuerberaterhaftung	49
EStG Kommentare	49
Die Selbstanzeige im Steuerstrafverfahren	47
Der Steuerstreit	46
Grundlagen des schweizer Steuerrechts	44
Haftung für Steuerschulden	44
Grundlagen des US-amerikanischen Steuerrechts	41
Erbschafts- und Schenkungssteuer	40
Das große Umsatzsteuer-Handbuch	40
Das Steuer-Handbuch	39

ANTWORT: Da die Titel der gekauften Bücher überwiegend das Thema Recht aufgreifen, liegt es nahe dass die **Fakultät für Rechtswissenschaft** nahe der Buchhandlung liegt.

4. Clustere die verkauften Bücher für die Analyse sinnvoll in Genres

Dazu können wir die Buchtitel analysieren und die Bücher in Genres clustern. Für das Clustering wurde GPT4 verwendet.

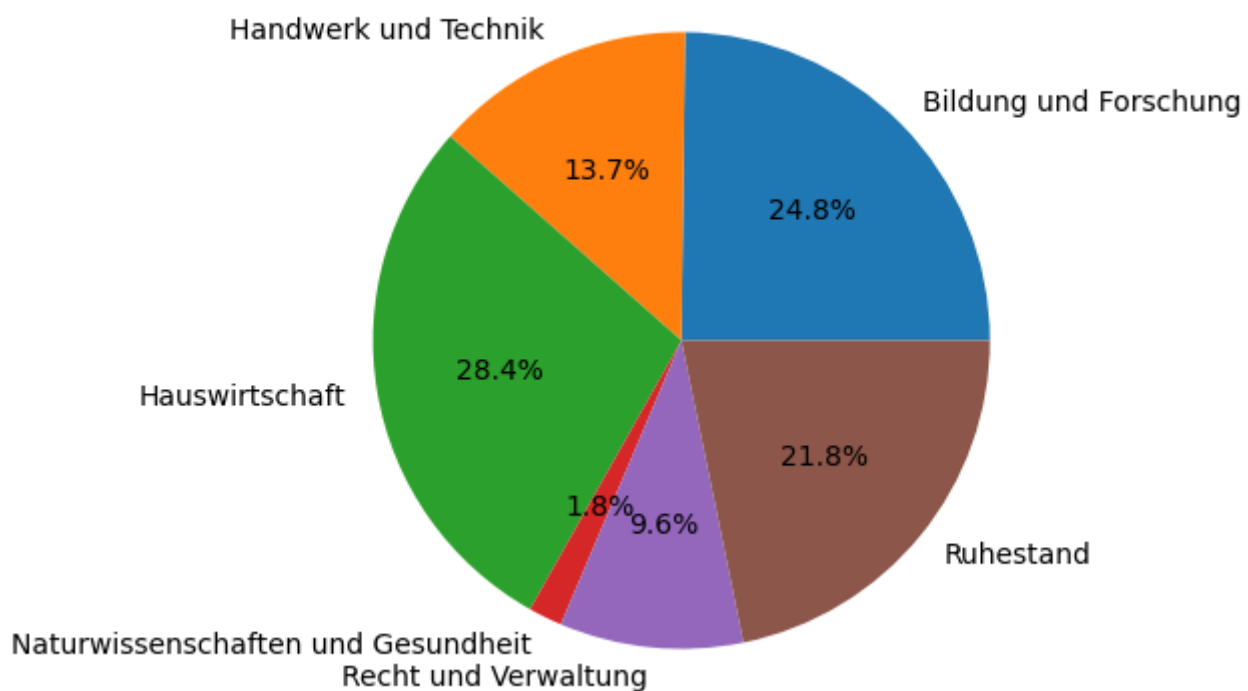
Clustering der Produkte



Aufgabe 5: Clustere die Berufe sinnvoll in Berufsgruppen

Hierzu werden die vorverarbeiteten Daten in drei Kategorien unterteilt: Kind, Angestellt, Arbeitslos, Rentner

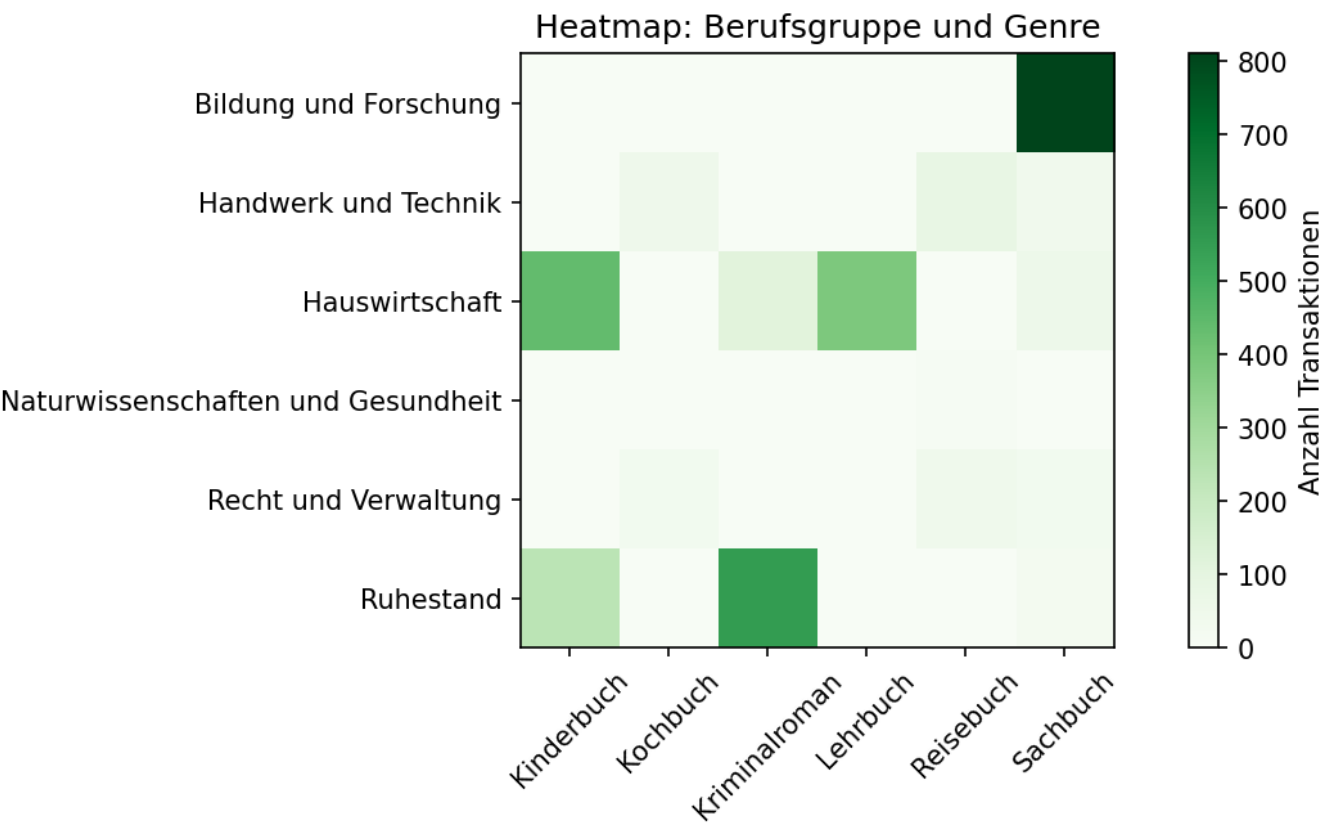
Berufsgruppe der Kunden

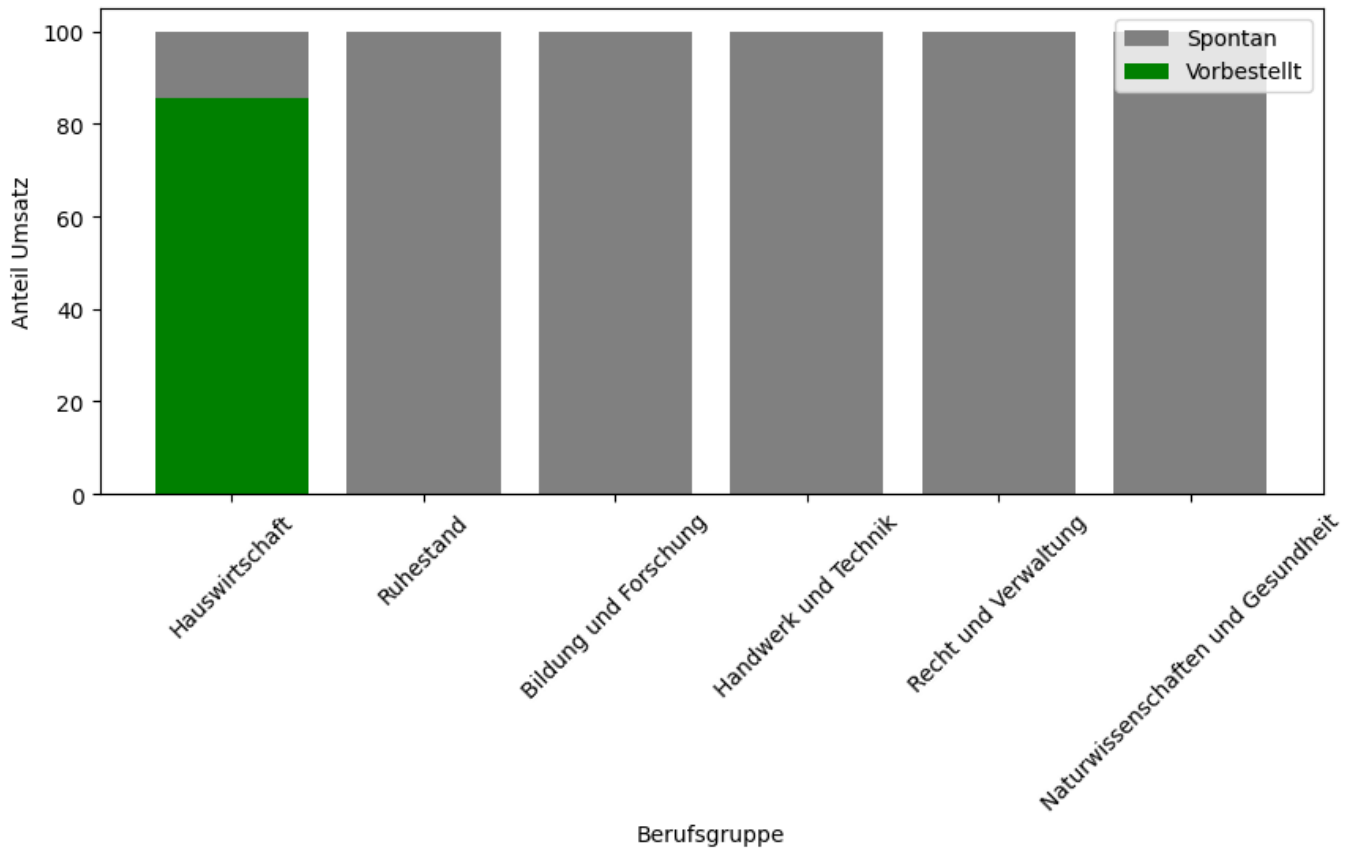


Aufgabe 6: Erstelle Kundenprofile

Welche sind die drei wichtigsten Kundenprofile für den Bücherabsatz in deinem Ladengeschäft?
Charakterisiere die Kundenprofile.

Dazu werden zunächst die Zusammenhänge von Berufsgruppen und Genre analysiert, bevor das
Kauffverhalten (spontan vs. vorbestellt) nach Gruppen untersucht wird.





ANTWORT: Die Analyse lässt die Charakterisierung von drei Kundenprofilen zu.

"Der Lehrling"

- Doktorant oder Student
- Kauft hauptsächlich Sachbücher
- Kauft hauptsächlich spontan

"Der Fürsorgliche"

- Elternteil oder Großelternteil
- Kauft gerne Kinder- oder Lehrbücher
- Bestellt hauptsächlich vor

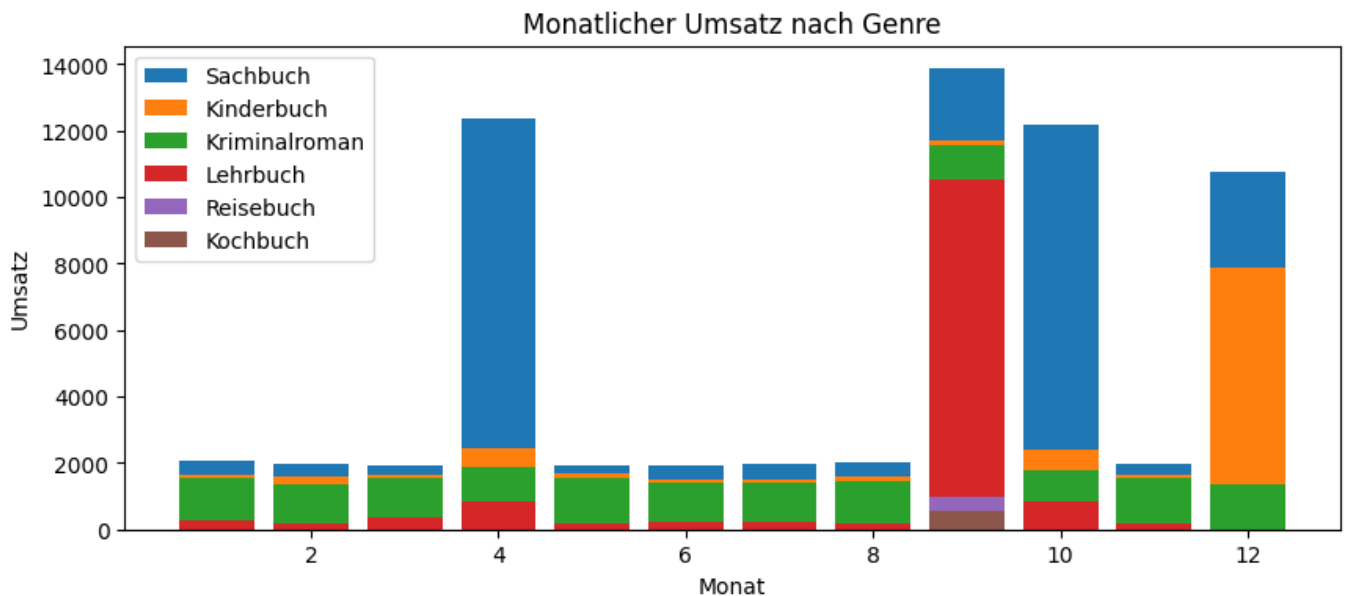
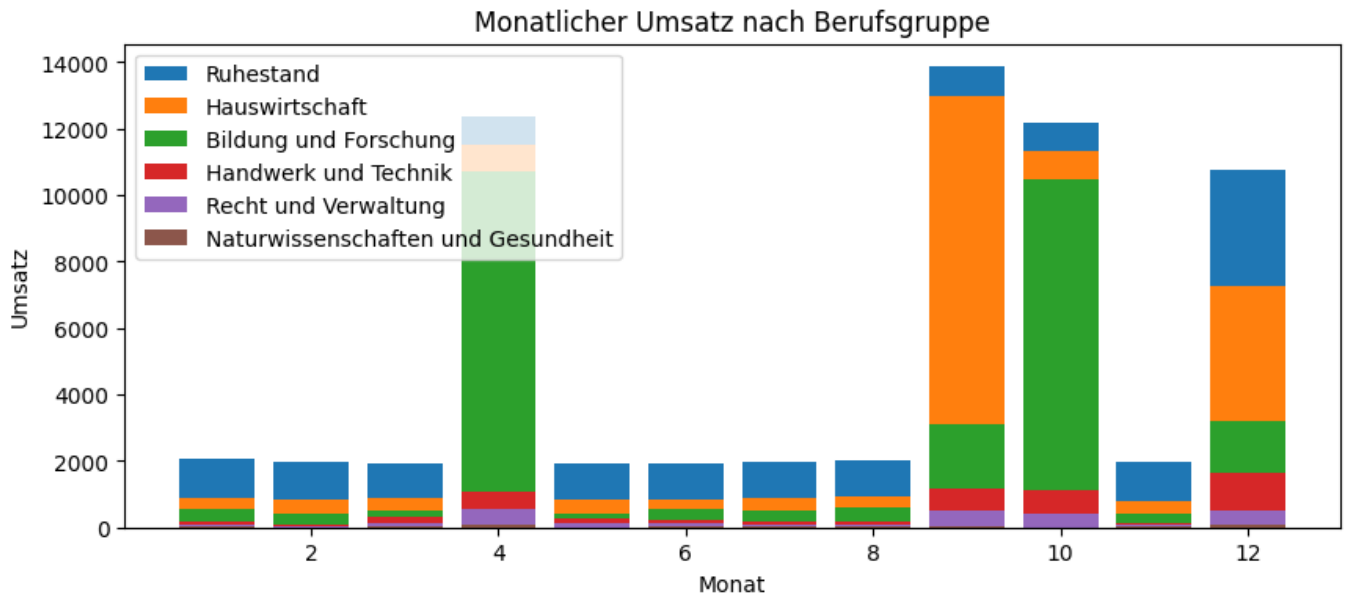
"Der Gelingweilte"

- Rentner, Pensionär oder Arbeitslos
- Kauft gerne Kriminalromane, vereinzelt auch Kinderbücher
- Kauft hauptsächlich spontan

Aufgabe 7: Erstelle eine Umsatzanalyse für 2023 auf Monatsbasis

Was sind die Schwerpunktmonate?

Dazu wird der Umsatz je Monat nach Genre und nach Berufsgruppe visualisiert.



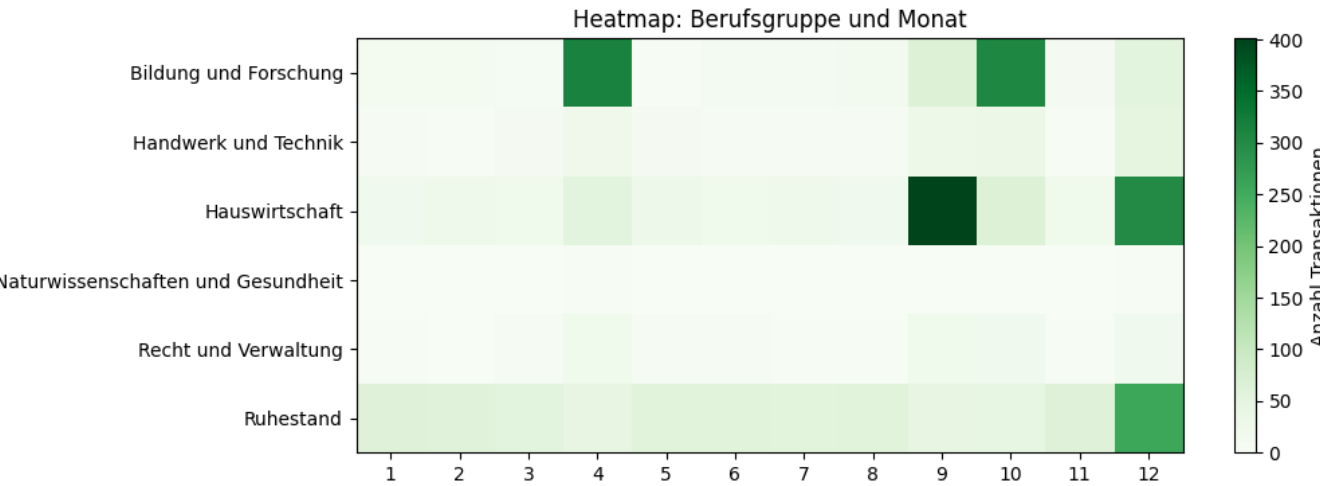
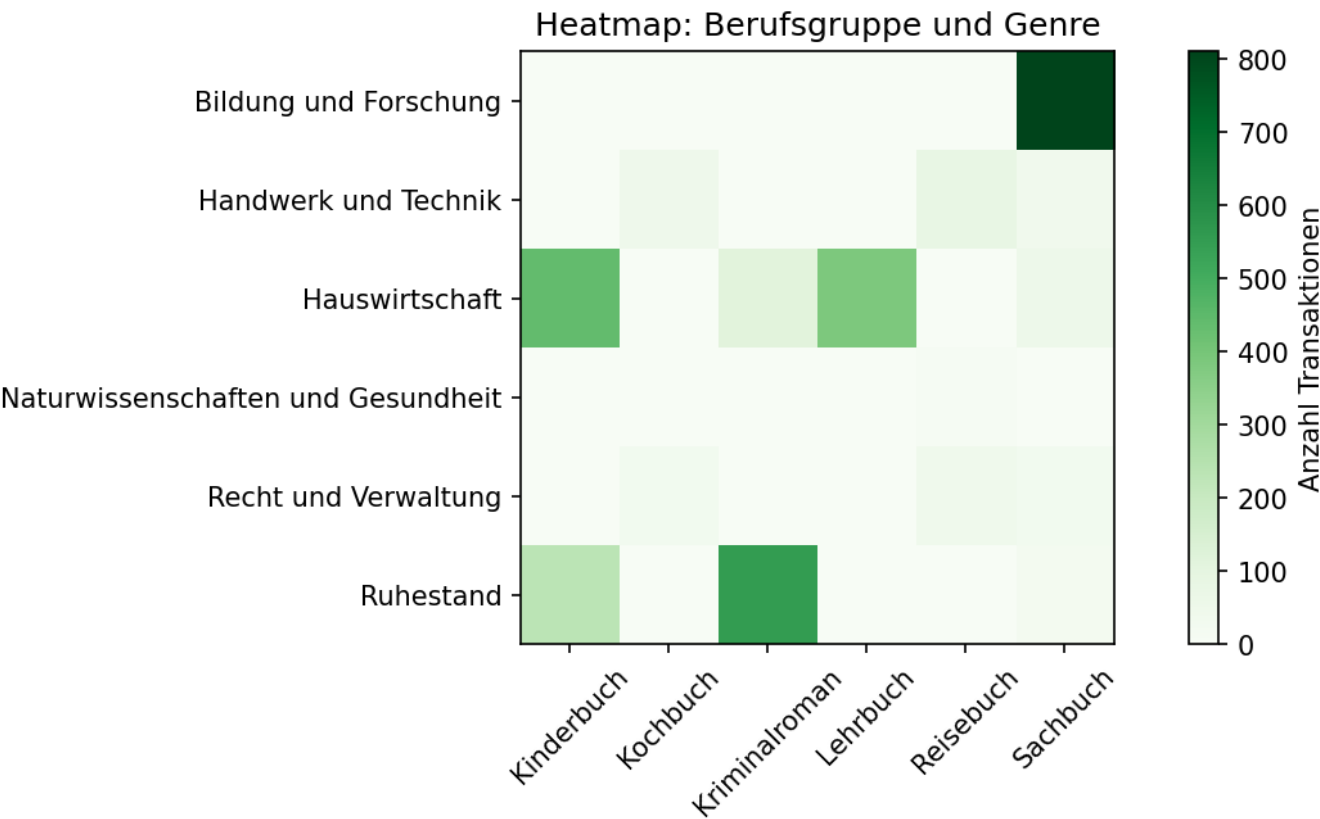
ANTWORT: Die Visualisierungen zeigen, dass die Monate **April, September, Oktober und Dezember die umsatzstärksten sind**. Dabei werden im April und Oktober hauptsächlich Sachbücher von Studenten gekauft (Semesterstarts), während im September die Eltern Lehrbücher kaufen. Im Dezember werden vermehrt Kinderbücher gekauft, vermutlich als Weihnachtsgeschenk.

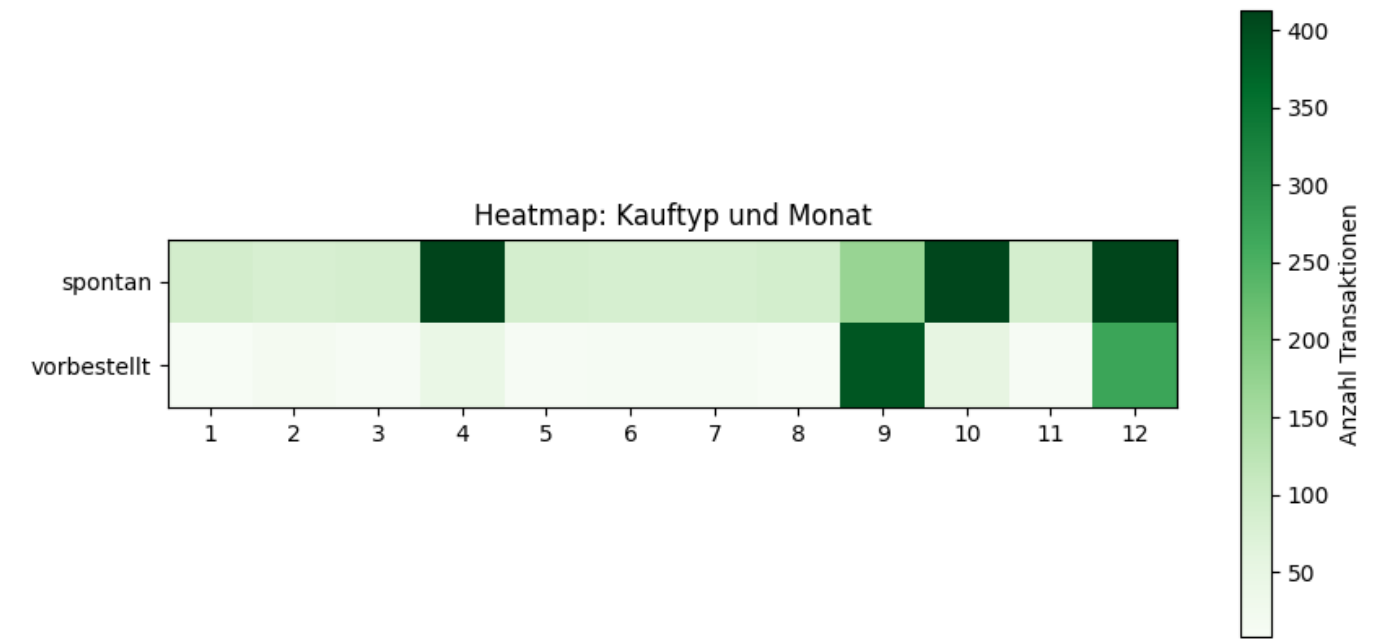
Aufgabe 8: Analysiere die Nichtschwerpunktmonate und die Schwerpunktmonate

Welche Kundenprofile kaufen wann was?

Da die Visualisierung einer dreidimensionalen Beziehung schwierig ist, wird jeweils die Relation zwischen zwei Dimensionen visualisiert.

- Wer kauft was? (Berufsgruppe zu Genre)
- Wer kauft wann? (Berufsgruppe zu Monat)
- Was wird wann gekauft? (Genre zu Monat)



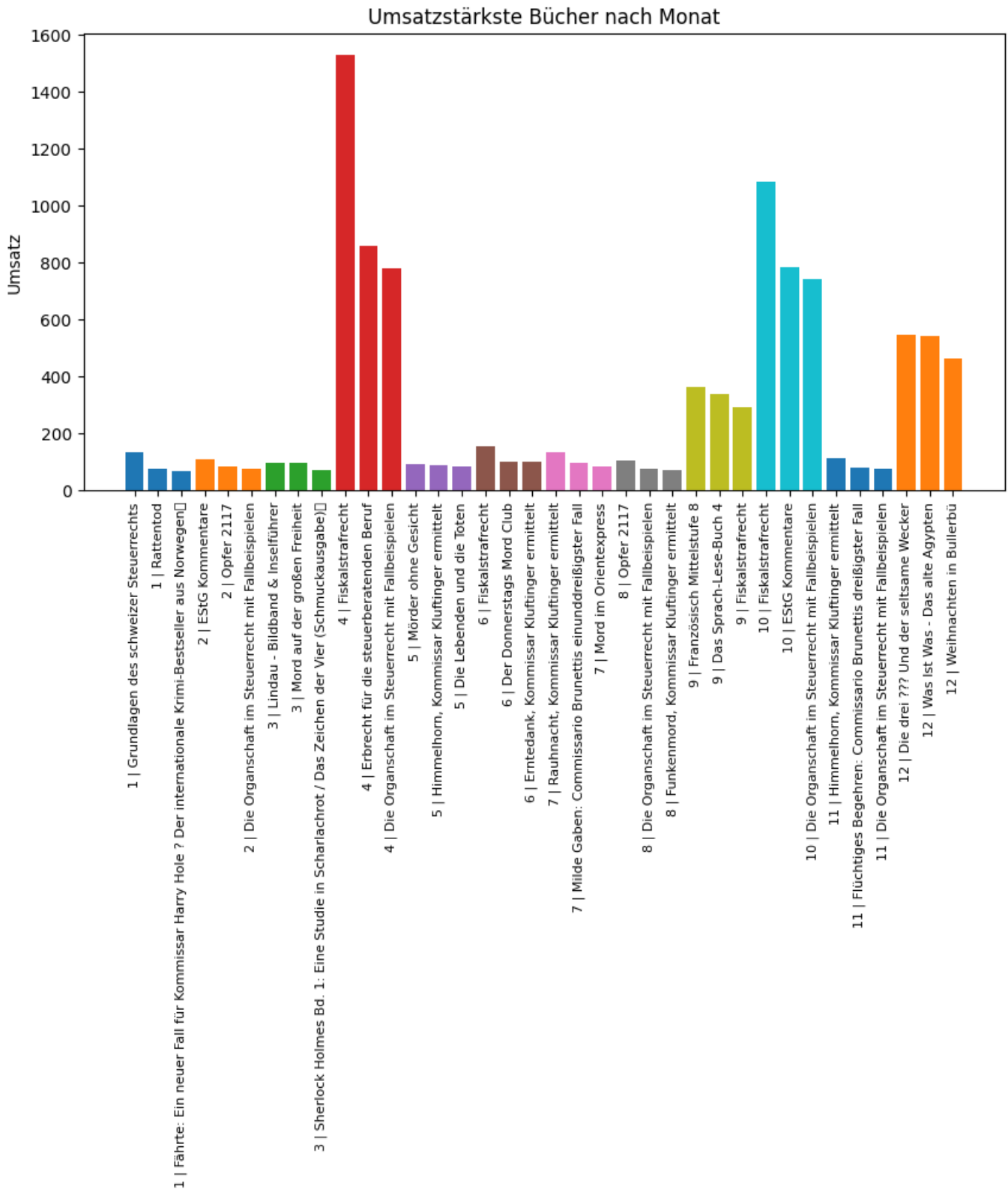


Aufgabe 9: Welche Bücher ins Schaufenster?

Wir nehmen an, dass das Ausstellen von Büchern mit dem maximalen Umsatz im jeweiligen Monat am sinnvollsten ist. Daher suchen wir für jeden Monat die umsatzstärksten Bücher und stellen diese aus.

Monat	Titel
1	Fährte: Ein neuer Fall für Kommissar Harry Hol...
1	Rattentod
1	Grundlagen des schweizer Steuerrechts
2	Die Organschaft im Steuerrecht mit Fallbeispielen
2	Opfer 2117
2	ESTG Kommentare
3	Sherlock Holmes Bd. 1: Eine Studie in Scharlac...
3	Mord auf der großen Freiheit
3	Lindau - Bildband & Inselführer
4	Die Organschaft im Steuerrecht mit Fallbeispielen
4	Erbrecht für die steuerberatenden Beruf
4	Fiskalstrafrecht
5	Die Lebenden und die Toten
5	Himmelhorn, Kommissar Kluftinger ermittelt
5	Mörder ohne Gesicht
6	Der Donnerstags Mord Club
6	Erntedank, Kommissar Kluftinger ermittelt
6	Fiskalstrafrecht
7	Mord im Orientexpress
7	Milde Geben: Commissario Brunetti einunddreißig

7	Milde Gaben: Commissario Brunettis einhundert...
7	Rauhnacht, Kommissar Kluftinger ermittelt
8	Funkenmord, Kommissar Kluftinger ermittelt
8	Die Organschaft im Steuerrecht mit Fallbeispielen
8	Opfer 2117
9	Fiskalstrafrecht
9	Das Sprach-Lese-Buch 4
9	Französisch Mittelstufe 8
10	Die Organschaft im Steuerrecht mit Fallbeispielen
10	EStG Kommentare
10	Fiskalstrafrecht
11	Die Organschaft im Steuerrecht mit Fallbeispielen
11	Flüchtiges Begehren: Commissario Brunettis dre...
11	Himmelhorn, Kommissar Kluftinger ermittelt
12	Weihnachten in Bullerbü
12	Was Ist Was - Das alte Agypten
12	Die drei ??? Und der seltsame Wecker

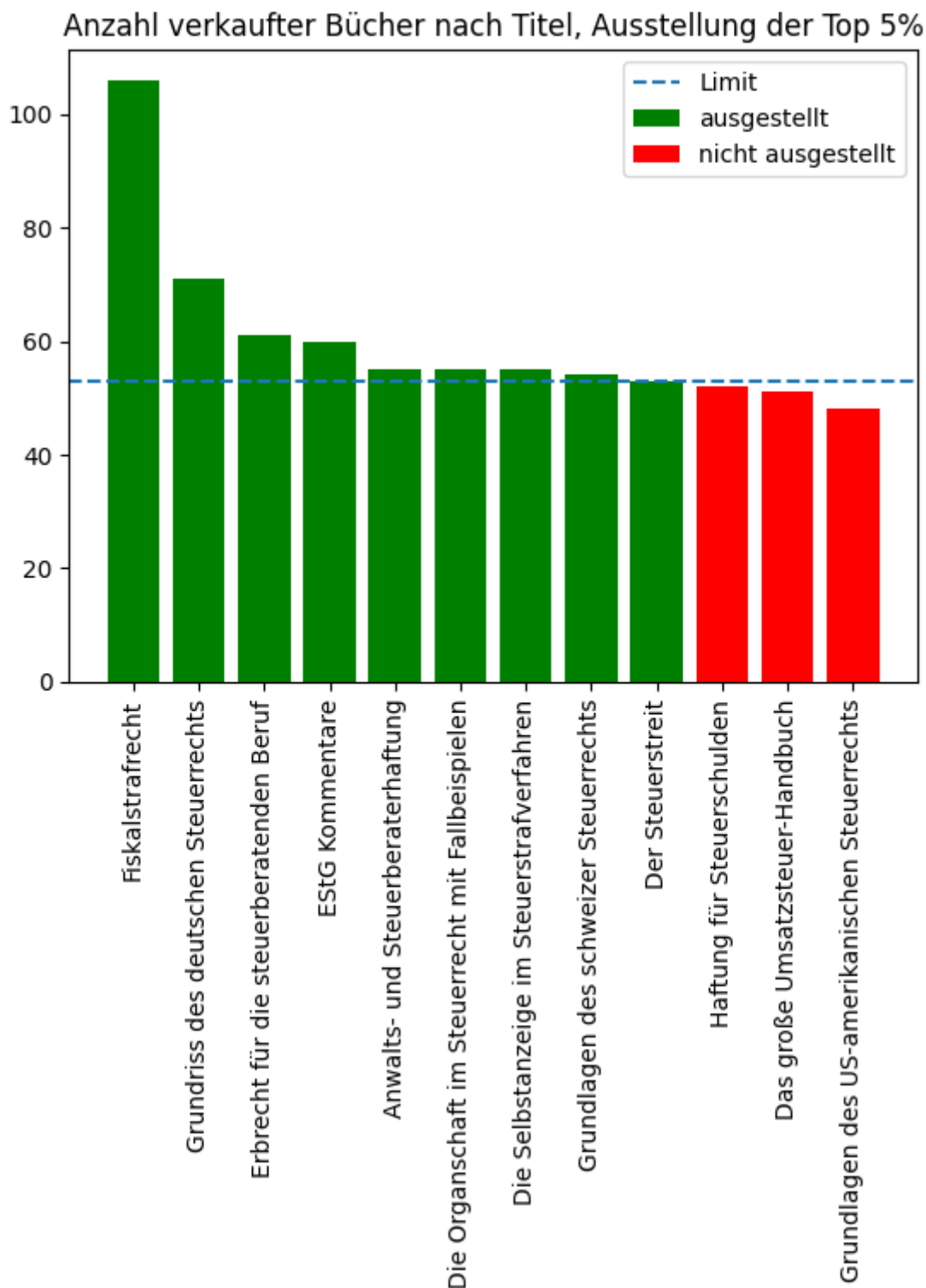


ANTWORT: Die exakten Titel sind der Tabelle unten zu entnehmen. Auffällig ist hierbei wieder das **Angebot an Kinderbüchern im Dezember**, sowie der **Verkauf von Sachbüchern in den Monaten April und Oktober**.

Aufgabe 10: Welche Bücher in die Ladenecke?

Da die Ecke klein ist und die Vorfinanzierung des Buchbestandes Geld kostet, sollten hier nur Bücher ausgestellt werden die auch tatsächlich gekauft werden. Außerdem sollte sichergestellt werden, dass die Nachfrage immer ausreichend gedeckt ist, um keinen Umsatz entgehen zu lassen.

Die Antwort auf diese Aufgabe ist eng mit Aufgabe 09 verknüpft und kann auf die gleichen Informationen zurückgreifen. In die Ladenecke sollten aber nur Bücher, die sich garantiert verkaufen. Dafür kann ein dynamischer Schwellenwert eingeführt werden, der entweder eine minimale Anzahl an Verkäufen vorgibt (z.B. 50 Verkäufe pro Jahr) oder ein Perzentil der am besten verkauften Bücher widerspiegelt (z.B. Top 5%).



ANTWORT: Je nach Festlegung des Schwellenwerts werden unterschiedlich viele Bücher ausgestellt. Bei Wahl der Top 5% sind es 9 Bücher, die alle Verkaufszahlen von über 50 Stück pro Jahr ausweisen. Diese sind in der Grafik oben abgebildet. Je nach Risikobereitschaft sollte die initiale Beschaffung sich an den Verkaufszahlen des letzten Jahres orientieren.