

Analyse des Kunden-Einkaufsverhaltens

1. Projektübersicht

Dieses Projekt analysiert das Einkaufsverhalten von Kunden anhand von Transaktionsdaten aus 3.900 Käufen in verschiedenen Produktkategorien. Das Ziel ist es, Erkenntnisse über Ausgabemuster, Kundensegmente, Produktvorlieben und das Abonnement-Verhalten zu gewinnen, um strategische Geschäftsentscheidungen zu unterstützen.

2. Zusammenfassung des Datensatzes

Zeilen: 3.900

Spalten: 18

Hauptmerkmale:

- Kundendemografie (Alter, Geschlecht, Standort, Abonnement-Status).
- Kaufdetails (Gekaufter Artikel, Kategorie, Kaufbetrag, Jahreszeit, Größe, Farbe).
- Einkaufsverhalten (Rabatt angewendet, Promo-Code genutzt, frühere Käufe, Häufigkeit der Käufe, Bewertung, Versandart).

Fehlende Daten: 37 Werte in der Spalte für Bewertungen.

3. Explorative Datenanalyse mit Python

Wir haben mit der Datenvorbereitung und Datenreinigung in Python begonnen:

Daten laden: Import des Datensatzes mit der Bibliothek „pandas“.

Erste Untersuchung: Nutzung von `df.info()` zur Prüfung der Struktur und `describe()` für statistische Zusammenfassungen.

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season	Review Rating	Subscription Status	Shipping Type	Discount Applied	Promo Code Used	Previous Purchases	Payment Method	Frequency of Purchases
count	3900.000000	3900.000000	3900	3900	3900	3900.000000	3900	3900	3900	3900	3863.000000	3900	3900	3900	3900	3900.000000	3900	3900
unique	NaN	NaN	2	25	4	NaN	50	4	25	4	NaN	2	6	2	2	NaN	6	7
top	NaN	NaN	Male	Blouse	Clothing	NaN	Montana	M	Olive	Spring	NaN	No	Free Shipping	No	No	NaN	PayPal	Every 3 Months
freq	NaN	NaN	2652	171	1737	NaN	96	1755	177	999	NaN	2847	675	2223	2223	NaN	677	584
mean	1950.500000	44.068462	NaN	NaN	NaN	59.764359	NaN	NaN	NaN	NaN	3.750065	NaN	NaN	NaN	NaN	25.351538	NaN	NaN
std	1125.977353	15.207589	NaN	NaN	NaN	23.685392	NaN	NaN	NaN	NaN	0.716983	NaN	NaN	NaN	NaN	14.447125	NaN	NaN
min	1.000000	18.000000	NaN	NaN	NaN	20.000000	NaN	NaN	NaN	NaN	2.500000	NaN	NaN	NaN	NaN	1.000000	NaN	NaN
25%	975.750000	31.000000	NaN	NaN	NaN	39.000000	NaN	NaN	NaN	NaN	3.100000	NaN	NaN	NaN	NaN	13.000000	NaN	NaN
50%	1950.500000	44.000000	NaN	NaN	NaN	60.000000	NaN	NaN	NaN	NaN	3.800000	NaN	NaN	NaN	NaN	25.000000	NaN	NaN
75%	2925.250000	57.000000	NaN	NaN	NaN	81.000000	NaN	NaN	NaN	NaN	4.400000	NaN	NaN	NaN	NaN	38.000000	NaN	NaN
max	3900.000000	70.000000	NaN	NaN	NaN	100.000000	NaN	NaN	NaN	NaN	5.000000	NaN	NaN	NaN	NaN	50.000000	NaN	NaN

Umgang mit fehlenden Daten: Prüfung auf Nullwerte und Ersetzung fehlender Werte in der Review Rating durch den Median der jeweiligen Produktkategorie

Standardisierung: Umbenennung der Spalten in das **snake_case** Format für bessere Lesbarkeit .

Merkmalserstellung (Feature Engineering):

- Erstellung einer Spalte für **age_group**
- Erstellung einer Spalte für **purchase_frequency_days**

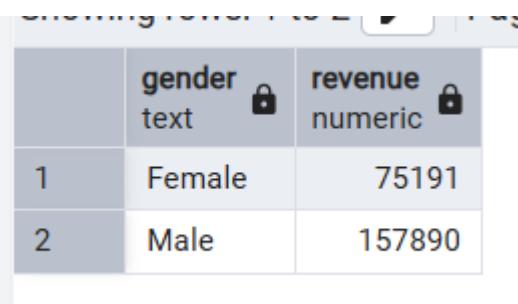
Konsistenzprüfung: Überprüfung, ob *discount_applied* und *promo_code_used* redundant waren; Löschung der Spalte für *promo_code_used*

Datenbank-Integration: Verbindung des Python-Skripts mit PostgreSQL und Laden der bereinigten Daten für die SQL-Analyse.

4. Datenanalyse mit SQL (Geschäftstransaktionen)

Wir haben strukturierte Analysen in PostgreSQL durchgeführt, um Geschäftsfragen zu beantworten:

1. Umsatz nach Geschlecht: Vergleich des Gesamtumsatzes von weiblichen und männlichen Kunden.



	gender	revenue
	text	numeric
1	Female	75191
2	Male	157890

2. Kunden mit hohem Konsum trotz Rabatt: Identifizierung von Kunden, die Rabatte nutzten, aber über dem Durchschnitt einkauften

	customer_id bigint	purchase_amount bigint
1	2	64
2	3	73
3	4	90
4	7	85
5	9	97
6	12	68
7	13	72
8	16	81
9	20	90
10	22	62
11	24	88
12	29	94
13	32	79
14	33	67
15	35	91
16	37	69

Total rows: 839 Query complete 00:00:00.207

3. Top 5 Produkte nach Bewertung: Suche nach Produkten mit den höchsten Durchschnittsbewertungen (z. B. Handschuhe: 3,86; Sandalen: 3,84)

	item_purchased text	avg_product_rating numeric
1	Gloves	3.86
2	Sandals	3.84
3	Boots	3.82
4	Hat	3.80
5	Skirt	3.78

4. Vergleich der Versandarten: Vergleich der Kaufbeträge zwischen Standard und Express-Versand

	shipping_type 	avg_purchase_amount 
1	Express	60.48
2	Standard	58.46

5. Abonnenten vs. Nicht-Abonnenten: Vergleich von Ausgaben und Umsatz nach Abo-Status.

	subscription_status 	count 	avg_spending 	total_revenue 
1	Yes	1053	59.49	62645.00
2	No	2847	59.87	170436.00

6. Rabatt-abhängige Produkte: 5 Produkte mit dem höchsten Anteil an Rabatt-Käufen.

	item_purchased 	total_revenue 	discount_rate 
1	Blouse	10410.00	33.00
2	Shirt	10332.00	42.00
3	Dress	10320.00	45.00
4	Pants	10090.00	47.00
5	Jewelry	10010.00	44.00

7. Kundensegmentierung: Einteilung der Kunden in Loyal (3.116), Wiederkehrend (701) und Neu (83).

	customer_segment 	number_of_customer 
1	Loyal	3116
2	New	83
3	Returning	701

8. Top 3 Produkte pro Kategorie: Auflistung der meistverkauften Artikel (z. B. Schmuck in Accessoires, Blusen in Kleidung).

	item_rank bigint	category text	item_purchased text	total_orders bigint
1	1	Accessori...	Jewelry	171
2	2	Accessori...	Sunglasses	161
3	3	Accessori...	Belt	161
4	1	Clothing	Blouse	171
5	2	Clothing	Pants	171
6	3	Clothing	Shirt	169
7	1	Footwear	Sandals	160
8	2	Footwear	Shoes	150
9	3	Footwear	Sneakers	145
10	1	Outerwear	Jacket	163
11	2	Outerwear	Coat	161

9. Wiederholungskäufer & Abonnements: Prüfung, ob Kunden mit mehr als 5 Käufen eher abonnieren (958 mit Abo vs. 2.518 ohne Abo)

	subscription_status text	total_customer bigint
1	No	2518
2	Yes	958

10. Umsatz nach Altersgruppe:

	age_group text	total_revenue numeric
1	Young Adult	62143
2	Middle-Aged	59197
3	Adult	55978
4	Senoir	55763

5. Dashboard in Power BI Wir haben ein interaktives Dashboard zur visuellen Darstellung erstellt



6. Geschäftsempfehlungen

- **Abonnements fördern:** Bewerbung exklusiver Vorteile für Abonnenten.
- **Kundenbindungsprogramme:** Belohnung von Wiederholungskäufern, um sie in das „loyale“ Segment zu führen.
- **Rabattpolitik prüfen:** Ausgewogenheit zwischen Umsatzsteigerung und Gewinnmargen finden.
- **Produktpositionierung:** Hervorhebung der am besten bewerteten und meistverkauften Produkte in Kampagnen.
- **Gezieltes Marketing:** Konzentration auf umsatzstarke Altersgruppen und Nutzer von Express-Versand.

