

Prompt Guide

Explorative Datenanalyse für Marketing Daten

Informationsklassifikation

öffentlich

Inhalt

1.	Datensammlung	3
2.	Deskriptive Statistik	3
3.	Visualisierungen	3
4.	Relationen	4
5.	Vorhersage	4
6.	Data Storytelling	5
7.	*Bonus* Dashboard	5

1. Datensammlung

- 1.1 Bitte lies das Dataset.
- 1.2 Wie beurteilst Du die Qualität des Datasets?
 - 1.2.1 Gibt es ungewöhnliche Ausprägungen oder Ausreißer? Beachte die Spalten 1-5
- 1.3 Bitte visualisiere mit der Verteilung der Geburtsjahre, damit ich die Ausreißer erkennen kann.
- 1.4 Bitte als Boxplot. Ausreißer als Punkte.
- 1.5 Gestalte das grafisch ansprechender. Jeder Punkt ein Ausreißer, ja ?

2. Deskriptive Statistik

- 2.1 Wie viele Kundeneinträge sind in dem Datensatz?
- 2.2 Wie hoch ist das durchschnittliche Einkommen?
- 2.3 Wieviel gibt ein Single im Durchschnitt für Wein aus?
- 2.4 Kannst Du Alle Felder Addieren, die mit Mnt anfangen? Wie hoch ist die Summe je Kohorte von 10 Jahren (also z.B. 1940-1949). Mache mir eine Tabelle. Ignoriere Werte vor 1940.
- 2.5 Bitte die Summe noch dazu und die Anzahl der Kunden.
- 2.6 Bitte die durchschnittlichen Gesamtausgaben dazu und die Anzahl der Kunden.

Info: Werte gegengerechnet in Excel Datei

3. Visualisierungen

- 3.1 Stelle mir die Ausgaben für die Produktkategorien als Säulendiagramm dar.
- 3.2 Was wäre ein geeignetes Diagramm, um den Familienstand der Kunden darzustellen?
 - 3.2.1 Ich möchte das Kreisdiagramm.

3.3 Kannst Du die Einkommensverteilung nach Familienstand in einem Boxplott darstellen?

3.4 Kannst Du bitte den Ausreißer bei Together, der mehr als 600000 ist, ignorieren, damit die Skalierung besser wird? Außerdem verzichte bitte auf die Gruppen "Absurd" und "YOLO".

3.5 Kannst Du das gleiche noch mal nach Berufsabschluss machen?

3.6 Wie schätzt Du die Verteilung ein?

3.7 Kannst Du noch mal die Anzahl der Personen nach Bildungsabschluss zeigen?

4. Relationen

4.1 Kannst Du mir helfen, Zusammenhänge in den Daten zu erkennen? Wie hoch sind die durchschnittlichen Ausgaben nach Familienstand addiert über alle Warengruppen? Bitte ignoriere Gruppen mit weniger als 10 Personen.

4.2 Okay, Zeige den Zusammenhang zwischen den Geburtsjahren (in 10 Jahre Kohorten) und dem Familienstand, bezogen auf die Gesamtausgaben. Ich hätte gerne eine Heatmap. Ignoriere Gruppen mit weniger als 10 Personen.

4.3 Bitte analysiere welche Kombinationen besonders hervorstechen. Aber vorher erweitere die Heatmap noch um die Zahl der Personen je Gruppe in Klammern.

4.4 Welchen Zusammenhang gibt es zwischen der Höhe des Einkommens und den Gesamtausgaben? Was sind maßgebliche Einflussfaktoren? Erstelle eine Korrelationsmatrix.

4.5 Kannst Du mir helfen, weitere Zusammenhänge zu erkennen? Wie hoch ist der Zusammenhang zwischen den Ausgaben für Wein und dem Familienstand?

4.6 Okay, Zeige den Zusammenhang zwischen den Geburtsjahren (in 10 Jahre Kohorten) und dem Familienstand, bezogen auf die Ausgaben für Wein. Ich hätte gerne eine Heatmap. Ignoriere Gruppen mit weniger als 10 Personen. Kannst Du in Klammern noch die Anzahl der Personen hinzufügen?

5. Vorhersage

5.1 Kannst Du mir helfen, ein Vorhersagemodell zu erstellen? Ich möchte verschiedene demographische Attribute und das Einkommen eingeben und eine Vorhersage der Gesamtausgaben für alle Warengruppen erhalten.

5.2 Mache 3 Test mit zufälligen Werten. Vermeide Werte, die im ursprünglichen Dataset wenig vorkommen, also z.B. Familienstand Yolo.

5.3 Kannst Du mir lokal ausführbaren Code erzeugen, für einen XGBOOST?

5.4 Wie kann ich die Vorhersagequalität noch verbessern?

6. Data Storytelling

6.1 Kannst Du mir sagen, was die Bestandteile einer guten Data Story sind?

6.2 Kannst Du mir helfen, eine Gute Data Story zu erstellen? Bitte berücksichtige die von Dir dargelegten Punkte. Die Story sollte sich um das Thema durchschnittliche Ausgabe pro Person drehen. Dazu hätte ich gerne 10 Personas, die sich aus demografischen Merkmalen, dem Einkommen und den Ausgaben zusammensetzen. Du kannst das Ruhig auch auf einzelne Warengruppen ausprägen. Gib den Personas realistische klingende Namen und eine kurze Bio (1-2 Sätze), um die Persona etwas emotionaler zu machen.

6.3 Wie könnten Inhalte für eine Infografik für die Persona "Carla" aussehen?

6.4 Bitte generiere dazu ein Bild.

7. *Bonus* Dashboard

Kannst Du mir ein Streamlit Dashboard erzeugen, mit dem ich Data Driven Personas erstellen kann? Ich möchte für die Gesamtausgaben, die Aufgaben für die Warengruppe und für das Einkommen Intervalle auswählen können. Für Familienstand, Bildungsgrad hätte ich gerne Dropdowns. Die Geburtsjahre gerne zu Jahrzehnten zusammengefasst. Auf dem Dashboard sollen Summen und Durchschnittswerte ersichtlich sein, und ganz wichtig, die Zahl der Personen. Bitte analysiere vorher noch mal das Dataset und achte auf das Trennzeichen ";"

Kannst Du die extremen Ausreißer noch beim einlesen filtern, sonst ist die Skalierung für die Auswahl der Intervalle so breit ?