

Datenexploration

Beschreiben Sie textuell was Sie in dem Chart sehen und ggfs. eine Interpretation

1. Line-Chart (Körperliche Aktivität in verschiedenen Altersgruppen)

Das Diagramm mit den durch Pfeile verbundenen Punkten zeigt die Abhängigkeit der körperlichen Aktivität vom Alter. Die Werte schwanken sehr stark, bewegen sich aber in einem engen Bereich zwischen 1,5 und 4,5. Die Aktivität steigt bis zum Alter von 30 Jahren von knapp 2,0 Tagen/Woche auf 3,5 Tage/Woche an. Danach schwanken die Werte ohne erkennbaren Trend.

2. Bar-Chart (Fast Food Konsum (Ja/Nein))

Das Diagramm zeigt mit zwei Balken den Anteil der Personen, die Fast Food essen und den Anteil der Personen, die dies nicht tun. Das Diagramm liefert jedoch keine weiteren Informationen, z.B. über die Bedeutung des Konsums. Das Diagramm ist eine anschauliche Visualisierung dieses Merkmals.

3. Histogramm (Verteilung des Alters)

Das Diagramm zeigt die Anzahl der Teilnehmer pro Altersgruppe. Es wurde in 20 Altersgruppen unterteilt, die als Balken dargestellt sind. Es ist zu erkennen, dass die Anzahl der Teilnehmer zwischen 20 und 25 Jahren relativ hoch ist und alle Altersgruppen durch mindestens 20 Personen vertreten sind. Außerdem zeigt die Kernel-Density-Schätzung eine Kurve, die die Wahrscheinlichkeit des Auftretens eines Wertes, in diesem Fall des Alters, darstellt.

4. Box-Plot (Gewichtsklasse Vorkommen und Median je Alter)

Es sind zwei Bereiche dargestellt, einer für jedes Geschlecht. Die Spanne der Werte ist als Striche dargestellt, der Median als blaue Fläche. Man sieht, dass bei beiden Geschlechtern die Gewichtsklasse im Durchschnitt zwischen 2.0 und 3.0 liegt, also normal bis übergewichtig. Allerdings sind bei beiden Geschlechtern alle Klassen vertreten.

5. Scatter-Plot (Gewichtsklasse je Alter)

Die Verteilung der einzelnen Gewichtsklassen nach Alter ist mit Punkten dargestellt. Es fällt auf, dass Übergewicht und Adipositas in allen Altersgruppen vorkommen, Untergewicht dagegen nur bei den unter 45-Jährigen. Das Diagramm hat die Schwäche, dass die 0,5 Zwischenstufen zwar eingezeichnet sind, aber keine Werte haben und somit leer sind.

Dokumentieren Sie jeweils mindestens eine Erkenntnis bzgl.

Datenstruktur, Daten und geschäftliche Domäne

1. Datenstruktur

- Die Darstellung der Daten als Zahlen ist teilweise verwirrend. Daten wie z.B. ob die Person raucht, hätten auch als Boolean Wert dargestellt werden können. So muss man immer in die Erklärung schauen.

2. Daten

- Da das Gewicht der Personen und damit verbundene Daten wie der BMI fehlen, sind einige Analysen kaum möglich. Wichtige Erkenntnisse können so kaum gewonnen werden. Die Einteilung des Gewichts in Klassen kann problematisch sein, da die Befragten eine Selbsteinschätzung vornehmen müssen, die falsch sein kann.

3. geschäftliche Domäne

- Eine solche Analyse kann für Gastronomiebetriebe nützlich sein. Sie können bestimmte Aktivitätstypen bestimmten Merkmalen wie z.B. dem Essverhalten zuordnen und ihr Angebot entsprechend anpassen. Zum Beispiel mit speziellen Produkten für normalgewichtige Sportler mit einer Vorliebe für Fast Food.
- Im Fitnessbereich können die gewonnenen Daten hilfreich sein, da sie zeigen, wie viel sich die Menschen bewegen und welcher Gewichtsklasse sie angehören. Aktive Personen können so motiviert werden, z.B. in ein Fitnessstudio zu gehen und Übergewichtigen können Möglichkeiten zur Gewichtsreduktion aufgezeigt werden. Hier ist es allerdings etwas schwierig, die Zielgruppe zu erreichen, da die Daten keiner Person zugeordnet sind,