

## Datenexploration

### Line-Chart: Körperliche Bewegung nach Alter

Das Liniendiagramm zeigt die durchschnittliche Häufigkeit körperlicher Aktivität (in Tagen pro Woche) in Abhängigkeit vom Alter. Daraus geht hervor, dass die körperliche Aktivität mit zunehmendem Alter von ca. 2 Tagen pro Woche im Alter von 20 Jahren auf ca. 5 Tage pro Woche im Alter von 27 Jahren ansteigt. Danach nimmt die körperliche Aktivität wieder ab, bis sie im Alter von 38 Jahren auf etwa 6 Tage pro Woche ansteigt. Bis zum Alter von 54 Jahren bleibt die körperliche Aktivität etwa zwischen 5 und 6 Tagen pro Woche.

Mögliche Erklärungen könnten sein, dass das Sportbewusstsein mit dem Alter zunimmt. Die Stagnation mit Anfang 30 könnte mit der familiären Situation in diesem Alter zusammenhängen. Die Stagnation im hohen Alter könnte auf gesundheitliche Einschränkungen hinweisen.

### Bar-Chart

Das Balkendiagramm zeigt die Verteilung der Personen, die angaben, regelmäßig Fast Food zu konsumieren. Etwa ein Viertel der Befragten gab an, regelmäßig Fast Food zu konsumieren, während drei Viertel keinen regelmäßigen Konsum angaben. Das Diagramm liefert jedoch keine weiteren Informationen, z.B. über die Bedeutung des Konsums.

### Bar-Chart

Das Histogramm zeigt die Anzahl der Teilnehmer pro Altersgruppe. Es wurde in 6 Altersgruppen zu je 6 Jahren unterteilt, die als Balken dargestellt sind. Es ist zu erkennen, dass die Anzahl der Teilnehmer im Alter von Mitte 20 am höchsten ist. Je höher das Alter, desto weniger Personen haben an der Umfrage teilgenommen. Darüber hinaus zeigt die Kernel-Density-Schätzung eine Kurve, die die Wahrscheinlichkeit des Auftretens eines Wertes, in diesem Fall des Alters, darstellt. Die Altersverteilung könnte durch das Durchschnittsalter der Internetnutzer erklärt werden.

## Boxplot

Der Boxplot zeigt das Alter in Abhängigkeit von der Menge der täglichen Hauptmahlzeiten. Es ist zu erkennen, dass die Anzahl der Hauptmahlzeiten mit zunehmendem Alter ansteigt. Während die durchschnittliche Person, die 1 bis 2 Mahlzeiten pro Tag zu sich nimmt, etwa 27 Jahre alt ist, liegt das Durchschnittsalter der Personen, die 3 oder mehr Mahlzeiten pro Tag zu sich nehmen, bei 40 Jahren. Aufgrund der Altersverteilung in der Befragung kann davon ausgegangen werden, dass viele ältere Menschen 3 oder mehr Mahlzeiten pro Tag zu sich nehmen.

## Scatter-Plot

Die Streudiagramme zeigen die Gewichtsklassen nach Alter. Aufgrund der geringen Anzahl an Spots, die durch die vier Gewichtsklassen entstehen, sind die meisten Werte ausgefüllt. Auffällig ist jedoch das Fehlen von Personen mit Unter- oder Normalgewicht. Dies deutet darauf hin, dass es mit zunehmendem Alter schwieriger wird, ein niedriges Gewicht zu halten.

## Sortieren

Die Excel-Datei ist nach Alter sortiert. Im Allgemeinen sollte die Sortierung für die Zwecke des Projekts keinen Unterschied machen. Insbesondere hat die Reihenfolge keinen Einfluss auf die Diagramme oder die Datenmodellierung. Die Vorteile bei einer einstufigen Dateninspektion werden dadurch negiert, da diese nicht vorgesehen ist.

## Zusammenfassung einer Spalte

Die Spalte "Alter" enthält 1.610 Einträge. Das Durchschnittsalter (Mean) beträgt 33,12 Jahre mit einer Standardabweichung von 9,84 Jahren, was auf eine moderate Streuung der Alterswerte hinweist. Das Mindestalter liegt bei 18 Jahren, das Höchstalter bei 54 Jahren. Die Verteilung zeigt, dass 25 % der Werte unter 25 Jahren liegen, der Median (50. Perzentil) liegt bei 32 Jahren und 75 % der Werte liegen unter 41 Jahren.

## Gruppierung einer Spalte

Bei der Gruppierung des Durchschnittsalters nach Gewichtsklassen fällt auf, dass es Unterschiede zwischen den einzelnen Klassen gibt. Diese Gruppierung ist sinnvoll,

um mögliche Zusammenhänge zwischen Alter und Gewichtsklasse zu erkennen. Es zeigt sich, dass das Durchschnittsalter in den höheren Gewichtsklassen (Klassen 3 und 4) signifikant höher ist als in den niedrigeren Klassen (Klassen 1 und 2). Dies könnte darauf hindeuten, dass ältere Personen eher in den höheren Gewichtsklassen vertreten sind oder das Gewicht und Alter in einem bestimmten Verhältnis zueinander stehen.

## Dokumentation einer Erkenntnis

### Datenstruktur

Die Darstellung der Daten als Zahlen ist teilweise verwirrend. Daten wie z.B. ob die Person raucht, hätten auch als Boolescher Wert dargestellt werden können. Die Erläuterungen in der Excel-Datei werden regelmäßig aufgerufen.

### 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.

### Geschäftliche Domäne

Ein medizinischer Mehrwert der Analyse könnte darin bestehen, dass die einzelnen Faktoren der Adipositas genauer untersucht werden können. Durch die mögliche Isolierung der einzelnen Faktoren voneinander könnten übergewichtige Menschen ein genaueres Bild davon bekommen, an welchen Stellschrauben sie drehen müssen, um am effektivsten abzunehmen. Dies könnte eine Verbesserung gegenüber dem Status quo darstellen, bei dem oft zu viel auf einmal versucht wird und das Ziel an der schieren Menge der verschiedenen Herausforderungen scheitert.