

Analyse des Titanic-Datensatzes

Rit Alija Yasin Sabanoglu

4. Februar 2026

1 Einleitung

Der Titanic-Datensatz ist ein häufig verwendeter Beispieldatensatz zur Einführung in Datenanalyse und Statistik. Er enthält Informationen über Passagiere der Titanic, unter anderem zu Geschlecht, Alter, Ticketklasse, Ticketpreis und Überlebensstatus. Ziel dieser Arbeit ist es, den Datensatz aufzubereiten, deskriptiv zu analysieren und erste Zusammenhänge zwischen ausgewählten Variablen zu untersuchen.

2 Datenaufbereitung

Die Rohdaten wurden zunächst bereinigt und für die Analyse vorbereitet. Dabei wurden unter anderem folgende Schritte durchgeführt:

- Extraktion und Vereinheitlichung der Anrede aus dem Namen
- Umkodierung geeigneter Variablen in Faktoren
- Imputation fehlender Alterswerte anhand des Medianalters pro Anrede
- Aufbereitung der Kabineninformation (Deck und Seite)
- Entfernen nicht benötigter Variablen

Der bereinigte Datensatz wurde anschließend als neue CSV-Datei gespeichert und diente als Grundlage für alle weiteren Analysen.

3 Deskriptive Analyse

3.1 Metrische Variablen

Tabelle 1 zeigt die deskriptiven Kennzahlen für das Alter der Passagiere.

Tabelle 1: Deskriptive Statistik des Alters

Kennzahl	Wert
Anzahl	891
Mittelwert	29.7
Median	28.0
Standardabweichung	14.5
Minimum	0.4
Maximum	80.0

Tabelle 2 zeigt die deskriptiven Kennzahlen für den Ticketpreis.

Tabelle 2: Deskriptive Statistik des Ticketpreises

Kennzahl	Wert
Anzahl	891
Mittelwert	32.2
Median	14.5
Standardabweichung	49.7
Minimum	0.0
Maximum	512.3

3.2 Kategoriale Variablen

Tabelle 3 zeigt die Verteilung des Überlebensstatus.

Tabelle 3: Überlebensstatus der Passagiere

Status	Häufigkeit	Anteil
Ja	342	0.384
Nein	549	0.616

4 Bivariate Analyse

Abbildung 1 zeigt die Überlebenshäufigkeit in Abhängigkeit von der Ticketklasse.

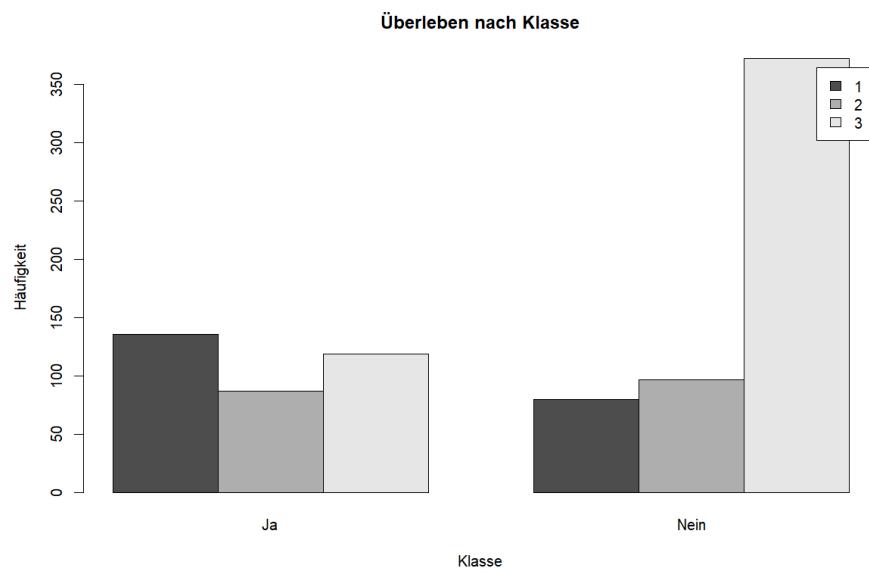


Abbildung 1: Überleben nach Ticketklasse

Abbildung 2 zeigt den Zusammenhang zwischen Geschlecht und Überlebensstatus.

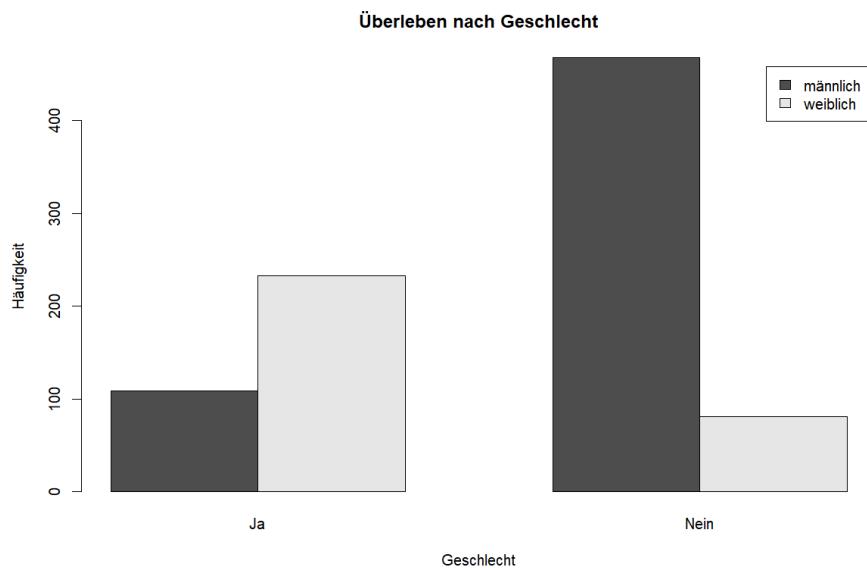


Abbildung 2: Überleben nach Geschlecht

Abbildung 3 zeigt den Überlebensstatus nach Einschiffungshafen.

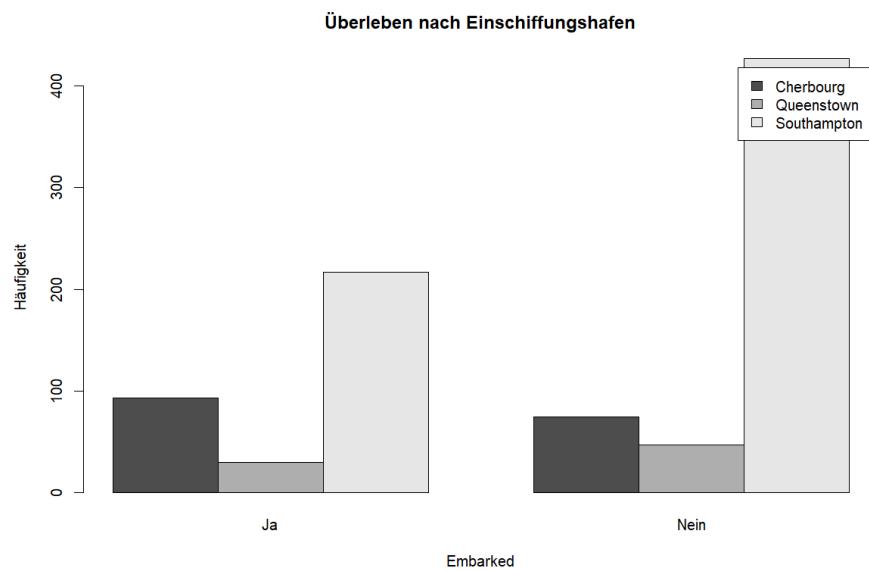


Abbildung 3: Überleben nach Einschiffungshafen

Abbildung 4 zeigt die Altersverteilung getrennt nach Überlebensstatus.

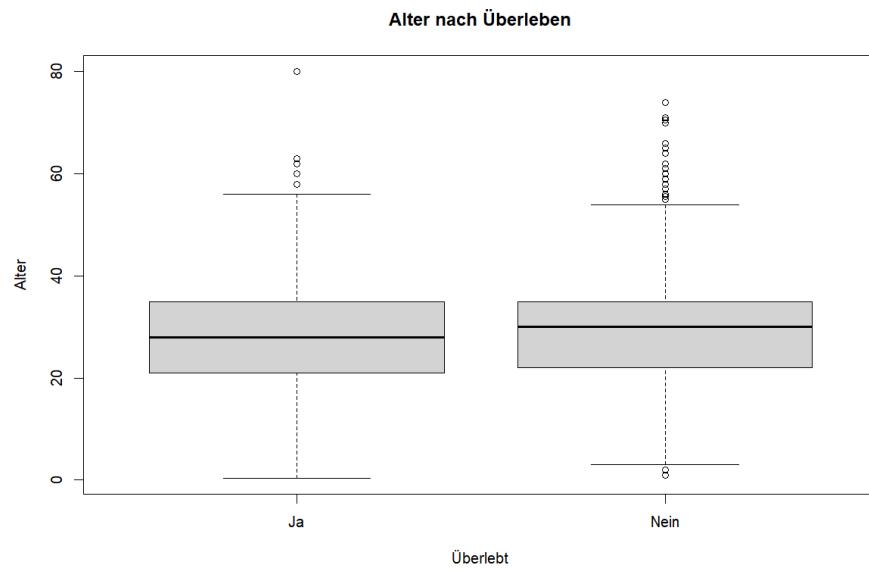


Abbildung 4: Alter nach Überlebensstatus

Abbildung 5 zeigt die Verteilung der Ticketpreise nach Überlebensstatus.

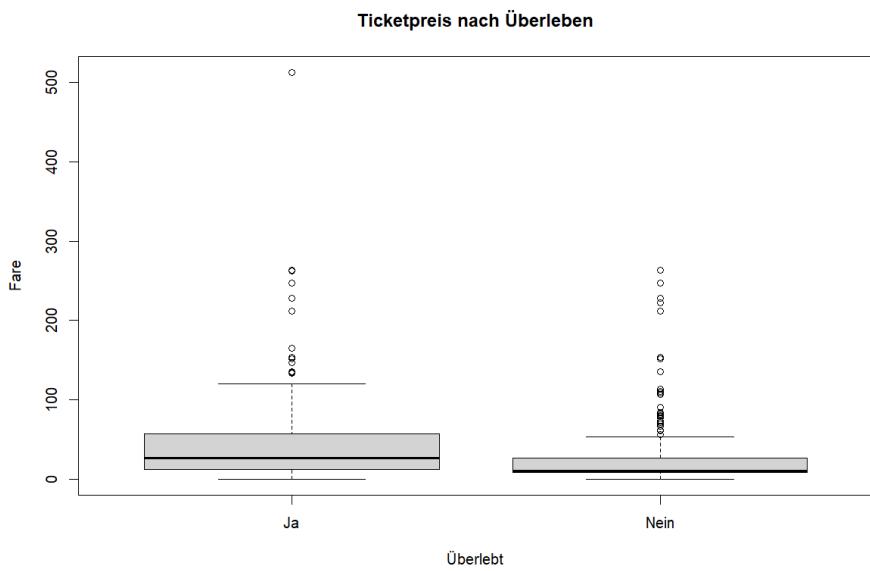


Abbildung 5: Ticketpreis nach Überlebensstatus

Abbildung 6 zeigt die gemeinsame Verteilung von Überlebensstatus, Geschlecht und Ticketklasse.

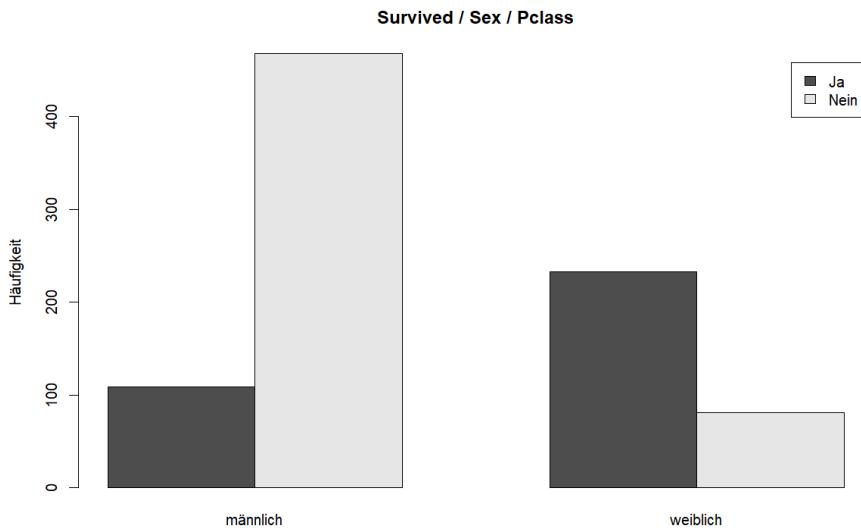


Abbildung 6: Überleben nach Geschlecht und Ticketklasse

5 Diskussion

Die Ergebnisse zeigen deutliche Unterschiede in den Überlebenswahrscheinlichkeiten zwischen verschiedenen Gruppen. Insbesondere weibliche Passagiere sowie Passagie-

re höherer Klassen hatten eine deutlich höhere Überlebenschance. Auch der Ticketpreis unterscheidet sich klar zwischen Überlebenden und Nicht-Überlebenden.

6 Fazit

In dieser Arbeit wurde der Titanic-Datensatz erfolgreich aufbereitet und analysiert. Die Ergebnisse verdeutlichen bekannte Muster und zeigen den Nutzen strukturierter Datenanalyse anhand eines realitätsnahen Beispiels.