

# Analyse Datensatz Titanic Bericht

## 1 Einleitung

Unser Ziel war es herauszustellen welche Passagiere den Untergang am häufigsten überlebt haben.

AuSSerdem wollten wir noch schauen ob es Unterschiede bei den Ticketpreisen gab.

## 2 Daten

### 2.1 Aufzählung

Zur Verfügung standen uns 11 verschiedene Daten über jeden Passagier auf dem Schiff

- **Survived:** Ob die Person überlebt hat oder nicht
- **Pclass:** In welcher Preisklasse die Person untergebracht war
- **Sex:** Das Geschlecht der Person
- **Age:** Das Alter der Person
- **SibSp:** Die Anzahl der Geschwister und Ehepartner an Bord
- **Parch:** Die Anzahl der Eltern und Kinder an Bord
- **Fare:** Der Preis des Tickets
- **Embarked:** Der Zustiegshafen
- **Anrede:** Die Anrede der Person
- **Bord:** An welchem Bord sich die Kabine befindet
- **Deck:** An welchem Deck sich die Kabine befindet

## 2.2 Bewertung

Wir haben einige dieser Daten als unbrauchbar empfunden

*Anrede* ist in den meisten Fällen das gleiche wie *Sex*

*Deck* und *Bord* hat in über 75 Prozent der Daten ein NA stehen

*SibSp* und *Parch* beinhalten 2 verschiedene Daten die zu einem vereint wurden und es ist schwierig daraus vernünftige Schlussfolgerungen zu ziehen.

Außerdem haben wir die Daten von *Age* und von *Fare* klassiert damit sie sich anschaulicher darstellen lassen

Als Klassen haben wir bei *Age* gewählt:

- [0-12] = **Kinder**
- (12-21] = **Jugendlich**
- (21-50] = **Erwachsene**
- (50+) = **Ältere**

Und bei *Fare*:

- [0-30] = **Low**
- (30-100] = **Medium**
- (100+) = **High**

## 3 Überlebensrate

### 3.1 Hypothese

Unsere Hypothese ist, dass einerseits Frauen und Kinder wahrscheinlicher überlebt haben und andererseits Passagiere der ersten Klasse wahrscheinlich auch.

### 3.2 Auswertung

Als erstes schauen wir uns den Zusammenhang von Geschlecht und Überlebensrate an:

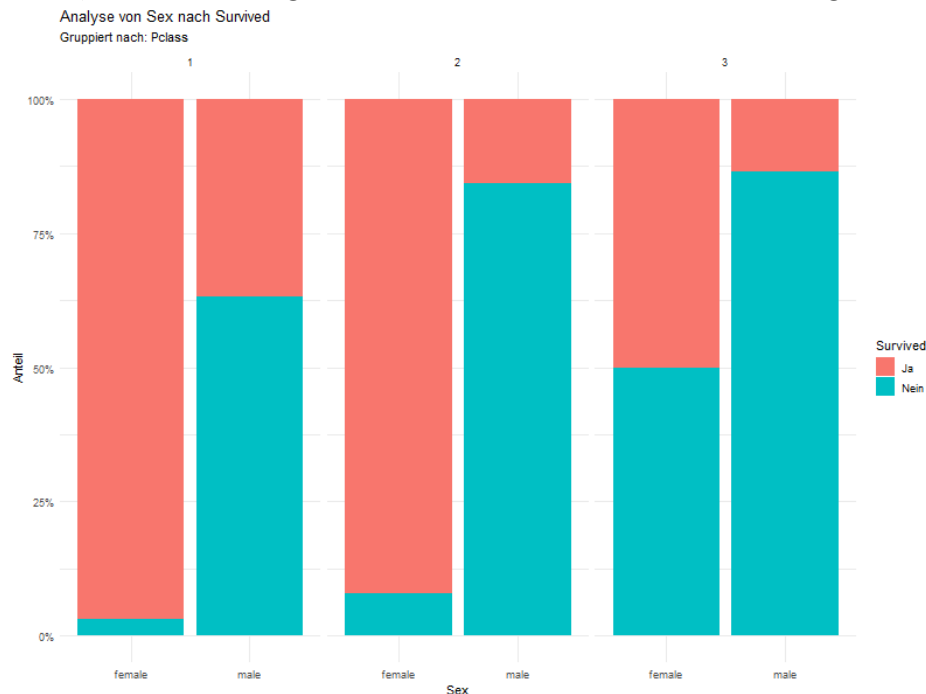
	female	male
Ja	233	109
Nein	81	468

Es sieht ganz danach aus als ob das Geschlecht einen sehr groSSen Einfluss auf das Überleben hat.

Nun das gleiche mit der Klasse:

	1	2	3
Ja	136	87	119
Nein	80	97	372

Das sieht ebenfalls so aus als würde unsere These stimmen, aber es könnte ja sein, dass die meisten Frauen erste Klasse waren und die meisten Männer 3. Klasse; von daher vergleichen wir die beiden Daten nochmal gleichzeitig:



Man sieht deutlich, dass die Überlebensrate in allen 3 Klassen bei den weiblichen Personen höher ist und bei beiden Geschlechtern je nach Klasse höher ist.

Das durchschnittliche Alter von überlebenden Personen ist 28, während das von den verstorbenen Personen bei 30 liegt.

Es könnte aber auch sein, dass zum Beispiel ältere Leute eher sterben, weil sie es zum Beispiel schlechter zu den Rettungsbooten schaffen. Deswegen testen wir es nochmal mit den klassierten Daten.

	Kinder	Jugendliche	Erwachsene	Ältere
Ja	42	69	209	22
Nein	31	102	374	42

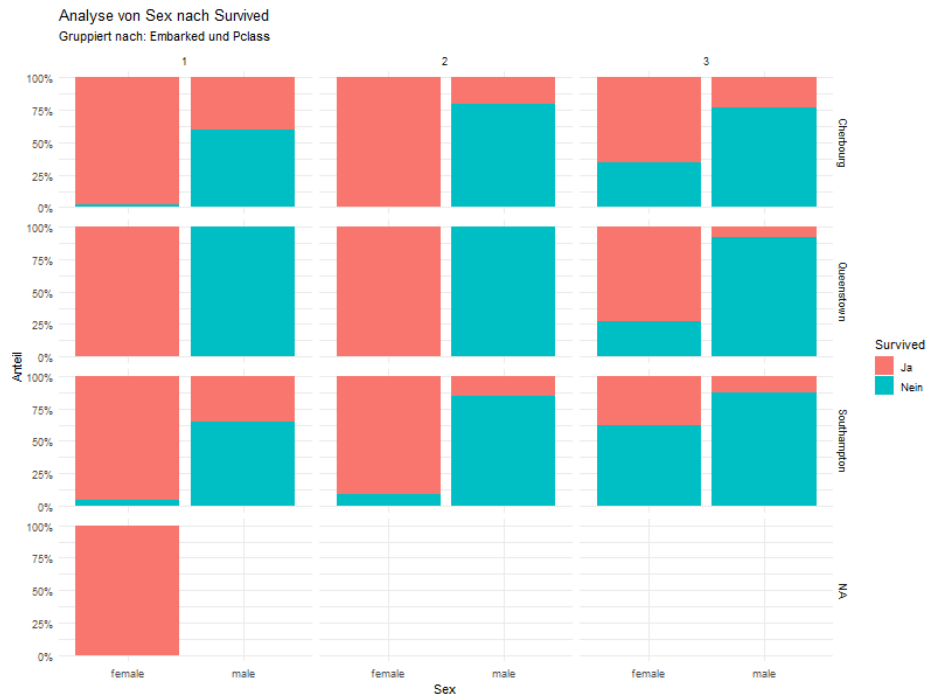
Man kann sehen, es sind tatsächlich die Kinder, die eher überleben.  
 Nun stellen wir noch einmal alle Daten zusammen in einer Grafik dar:



Das meiste sieht wie erwartet aus. Bei den Kindern scheint das Geschlecht kaum einen Einfluss zu haben.

Das einzige was raussticht sind die Mädchen der ersten Klasse, die eigentlich die höchste Überlebensrate haben müssten, aber 0 Prozent haben. Das liegt allerdings daran, dass es nur eine Person gibt auf die alle 3 Eigenschaften zutreffen, von daher ist es ein Einzelfall und eher nicht relevant.

Nun haben wir noch die Zustiegshäfen. Da in den verschiedenen Häfen unterschiedliche Verteilungen von Leuten zusteigen, schauen wir uns die Daten zusammen mit der Klasse, dem Geschlecht und dem Überleben an:



In den meisten Fällen sticht nichts raus, nur vielleicht die Frauen aus Southampton scheinen etwas weniger zu überleben.

### 3.3 Fazit

Wie erwartet sind Frauen und Kinder diejenigen, die am meisten überlebt haben, und die Klasse scheint auch einen großen Einfluss zu haben.

## 4 Ticketpreis

### 4.1 Hypothese

Wir vermuten, dass die Klasse bei weitem den größten Einfluss auf den Preis haben wird und eventuell haben es Kinder noch ein wenig günstiger.

### 4.2 Auswertung

Der durchschnittliche Ticketpreis der ersten Klasse beträgt ungefähr 84 (Median 60).

In der zweiten Klasse 21 (14).

Und in der dritten 14 (8).

Es gibt also eindeutige Preisunterschiede bei den Klassen, wobei 2 und 3 noch relativ nah beieinander liegen im Vergleich zu 1.

Bei den Kindern zahlt der Durchschnitt 32 (26).

Bei den Jugendlichen 27 (9).

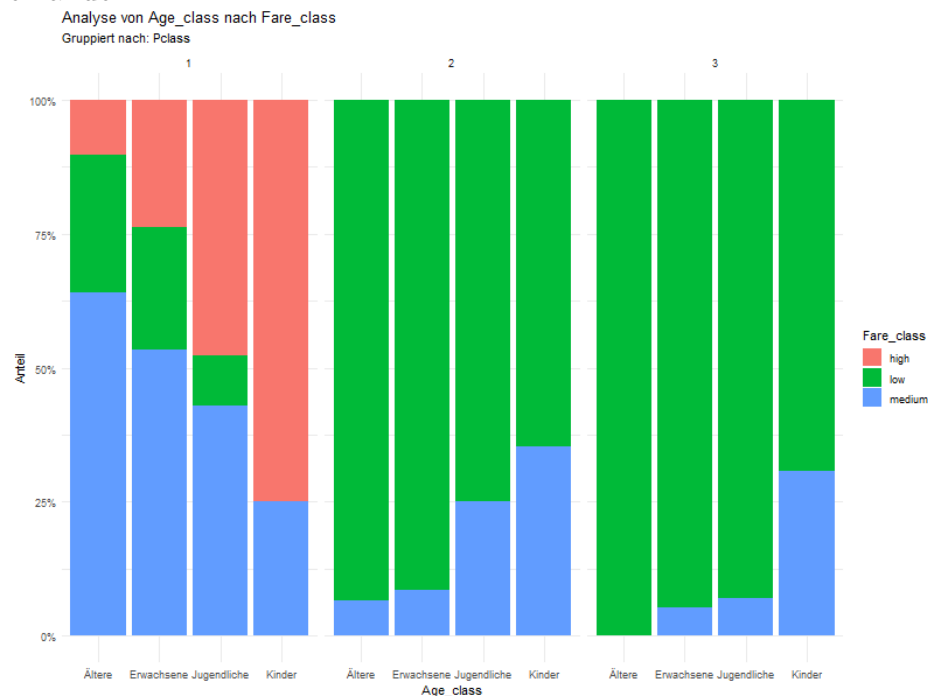
Bei den Erwachsenen 32 (13).

Und bei den Älteren 44 (28).

Da zwischen den Preisen von 1. Klasse und den anderen Klassen ein sehr großer Unterschied liegt, halten wir den Median für aussagekräftiger.

Da liegen die Kinder deutlich über den Erwachsenen, also ist genau das Gegenteil von unserer Hypothese der Fall.

Um es noch einmal zu überprüfen, vergleichen wir es nochmal unabhängig voneinander:



Tatsächlich zahlen Kinder in allen 3 Klassen am meisten und je älter, desto weniger wird bezahlt.

Die Logik hierbei ist zu hinterfragen, da zum einen Kinder normalerweise nicht zu teureren Preisen Tickets bekommen sollten als Erwachsene und diese meistens auch nicht alleine solche Fahrten antreten würden.

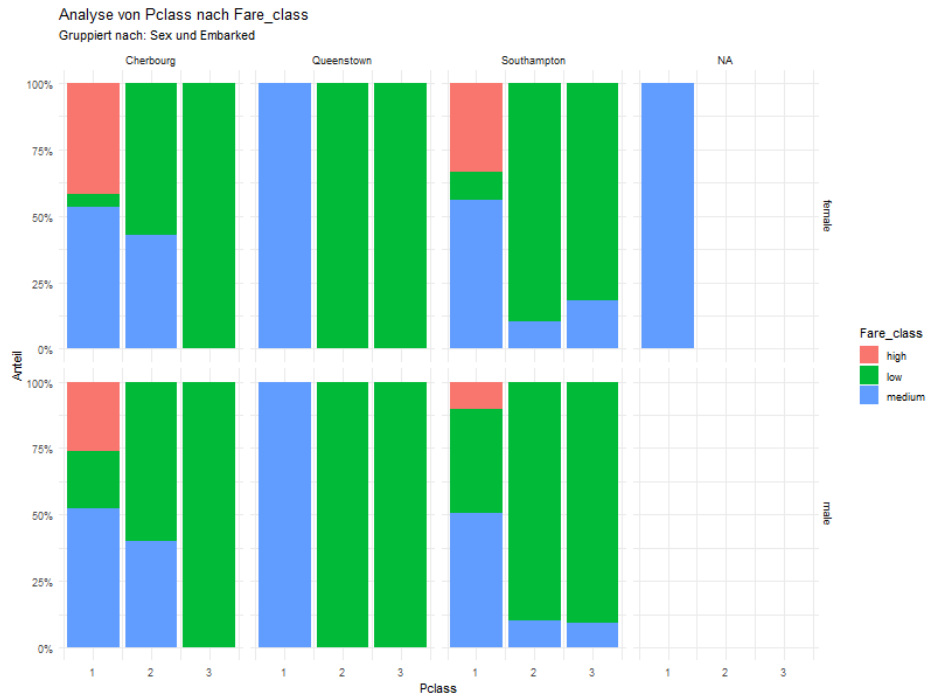
Zu erörtern ist daher, ob sich die Preise von Kindern überwiegend mit denen von Erwachsenen decken, was auf **Gruppentickets** hindeuten würde.

Eine Analyse der Preishäufigkeiten im Skript bestätigt, dass viele Preise identisch sind.

Die hohen Werte bei Kindern entstehen also vermutlich dadurch, dass

ihnen der Gesamtpreis einer Reisebegleitung oder Familie zugeordnet wurde.

Nun schauen wir uns nochmal die übrigen Daten an:



Männer zahlen in der Regel etwas weniger in der ersten Klasse.

Und Passagiere aus Cherbourg zahlen in erster und zweiter Klasse mehr und Passagiere aus Southampton zahlen etwas mehr in der dritten Klasse.

In Queenstown ist kaum jemand etwas anderes als 3. Klasse gefahren, von daher sollte man das weg lassen.

Zum Schluss testen wir noch ob mehr bezahlen unabhängig von der Klasse und Geschlecht die Überlebensrate erhöht:



Tatsächlich sieht es eher so aus, als würde in der dritten Klasse mehr zu bezahlen die Überlebenswahrscheinlichkeit von Frauen eher verringern.

Es sind jedoch nur 16 Personen, auf die das zutrifft, von daher nicht so repräsentativ wie gewünscht.

### 4.3 Fazit

Wie erwartet hat die Klasse den höchsten Einfluss, aber entgegen unserer Erwartungen zahlen Kinder mehr als Erwachsene.

Dies lässt sich jedoch durch die Verwendung von Gruppentickets erklären. Dazu gibt es noch einige leichte Unterschiede beim Geschlecht und beim Hafen.