

Titanic analysis

By Krzysztof Kleszcz

Introduction

Explore the poignant story of the Titanic through this dataset. It contains detailed information about the passengers aboard the ill-fated ship, which sank on April 15, 1912, after striking an iceberg. 📄

Columns:

- pclass - Ticket class
- survived - Whether the passenger survived the disaster
- name - Passenger's name
- sex - Passenger's gender
- age - Passenger's age
- sibsp - Number of siblings/spouses aboard
- parch - Number of parents/children aboard
- ticket - Ticket number
- fare - Ticket fare
- cabin - Cabin number
- embarked - Port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton)
- boat - Lifeboat number
- body - Body number (if the passenger did not survive and the body was recovered)
- home.dest - Destination

Just in case, please see table of contents:

1. General Data Overview 📊
2. Analysis of Missing Values 📈
3. Single Variable Analysis 🔍
4. Data Transformation 🔄
5. Analysis of Relationships Between Data 🔍
6. Outlier Analysis 📊
7. Analysis Summary 📄

1. General Data Overview 📊

We can observe that we have very diverse data, including both textual and numerical values. It's important to note that we also have missing parameters. 📄

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked
960	3.0	0.0	Lemberopolous, Mr. Peter L	male	34.5	0.0	0.0	2683	6.4375	NaN	C
902	3.0	0.0	Johnston, Mr. Andrew G	male	NaN	1.0	2.0	W./C. 6607	23.4500	NaN	S
887	3.0	1.0	Johannesen-Bratthammer, Mr. Bernt	male	NaN	0.0	0.0	65306	8.1125	NaN	S
487	2.0	0.0	Lingane, Mr. John	male	61.0	0.0	0.0	235509	12.3500	NaN	Q

We have 1,309 records and 14 columns, though not all columns are well-filled, such as "cabin" and "body."

```

pclass      1309
survived     1309
name         1309
sex          1309
age         1046
sibsp       1309
parch       1309
ticket      1309
fare        1308
cabin        295
embarked    1307
boat         486
body         121
home.dest    745
dtype: int64

```

	pclass	survived	age	sibsp	parch	fare	body
count	1309.000000	1309.000000	1046.000000	1309.000000	1309.000000	1308.000000	121.000000
mean	2.294882	0.381971	29.881135	0.498854	0.385027	33.295479	160.809917
std	0.837836	0.486055	14.413500	1.041658	0.865560	51.758668	97.696922
min	1.000000	0.000000	0.166700	0.000000	0.000000	0.000000	1.000000
25%	2.000000	0.000000	21.000000	0.000000	0.000000	7.895800	72.000000
50%	3.000000	0.000000	28.000000	0.000000	0.000000	14.454200	155.000000
75%	3.000000	1.000000	39.000000	1.000000	0.000000	31.275000	256.000000
max	3.000000	1.000000	80.000000	8.000000	9.000000	512.329200	328.000000

The most unique values that can be grouped are pclass, survived, sex, and embarked. 

```
pclass      3
survived    2
name       1307
sex         2
age        98
sibsp       7
parch       8
ticket     929
fare       281
cabin      186
embarked    3
boat       27
body       121
home.dest   369
dtype: int64
```

```
Unique values: pclass [ 1.  2.  3. nan]
Unique values: survived [ 1.  0. nan]
Unique values: sex ['female' 'male' nan]
Unique values: embarked ['S' 'C' nan 'Q']
```

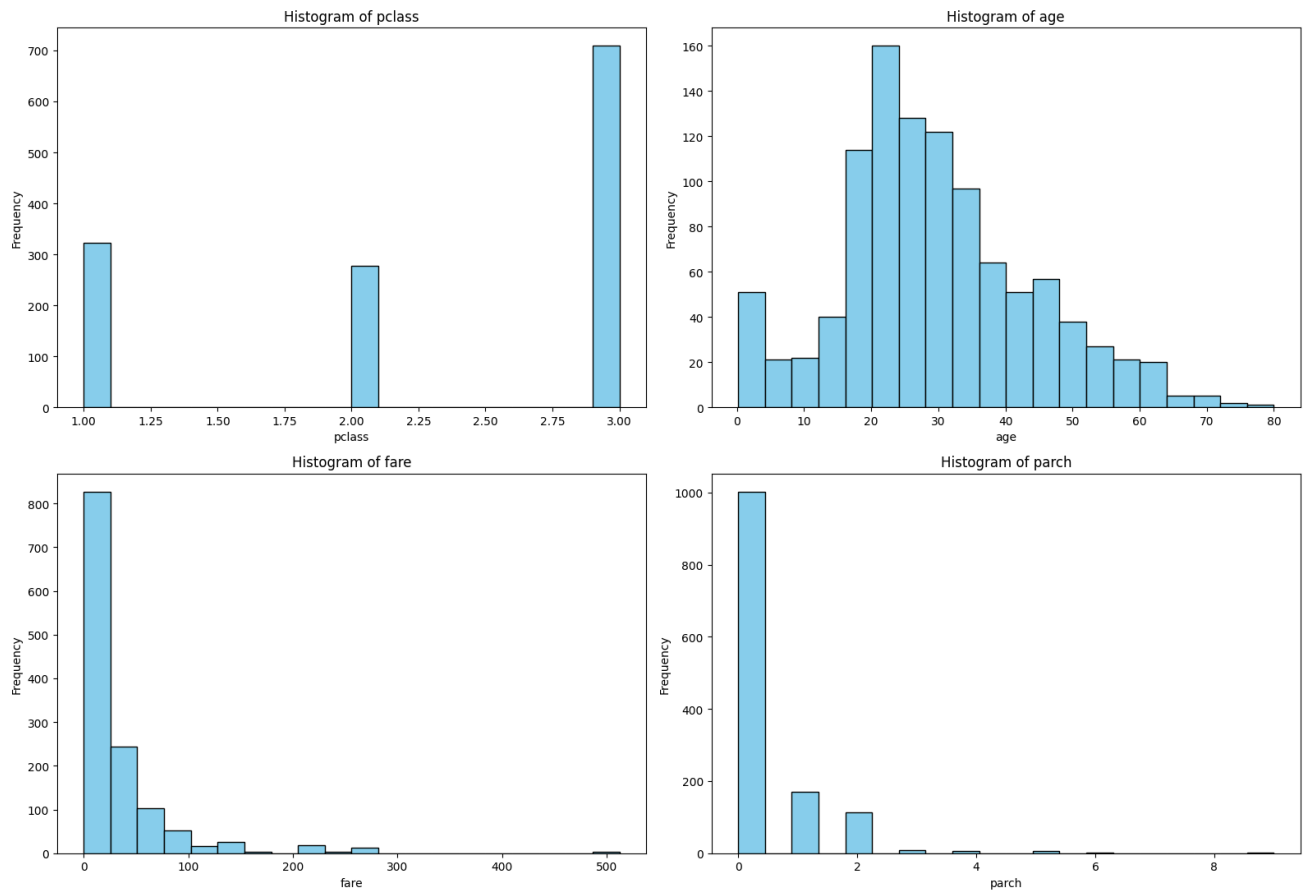
2. Analysis of Missing Values

The majority of missing values pertain to information about cabins, lifeboats, bodies, and future home destinations.

```
pclass      1
survived     1
name         1
sex          1
age        264
sibsp        1
parch        1
ticket       1
fare         2
cabin     1015
embarked     3
boat        824
body       1189
home.dest    565
dtype: int64
```

3. Single Value Analysis

- The majority of passengers traveled in third class.
- Only around 450 passengers survived, while over 800 were lost.
- Most passengers were between 20 and 40 years old.
- Most passengers traveled without family or spouse.



4. Data Transformation

- We will combine data about family and children into one table: family.
- We will remove tables with the most missing data, such as cabin, ticket, and body.
- These data points won't be particularly necessary for our analysis.
- We will replace the missing data with the median, the most frequent value, or the value "Unknown".

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embar
1092	3.0	0.0	Oreskovic, Mr. Luka	male	20.0	0.0	0.0	315094	8.6625	NaN	
699	3.0	0.0	Cacic, Mr. Luka	male	38.0	0.0	0.0	315089	8.6625	NaN	
845	3.0	1.0	Hakkarainen, Mrs. Pekka Pietari (Elin Matilda ...	female	24.0	1.0	0.0	STON/O2. 3101279	15.8500	NaN	
519	2.0	0.0	Norman, Mr. Robert Douglas	male	28.0	0.0	0.0	218629	13.5000	NaN	
909	3.0	1.0	Jussila, Mr. Eiriik	male	32.0	0.0	0.0	STON/O 2. 3101286	7.9250	NaN	
229	1.0	1.0	Penasco y Castellana, Mrs. Victor de Satode (M...	female	17.0	1.0	0.0	PC 17758	108.9000	C65	

```

pclass      0
survived     0
name        0
sex         0
age         0
ticket      0
fare        0
cabin       0
embarked    0
boat        0
body       1189
home.dest   0
family      0
dtype: int64

```

```

pclass  survived  name \
418     2.0      0.0      Gilbert, Mr. William
301     1.0      0.0      Walker, Mr. William Anderson
511     2.0      0.0      Myles, Mr. Thomas Francis
624     3.0      0.0      Andersson, Miss. Ellis Anna Maria
940     3.0      0.0      Kraeff, Mr. Theodor
50      1.0      1.0  Cardeza, Mrs. James Warburton Martinez (Charlo...

```

```

sex  age  ticket  fare  cabin embarked  boat  body \
418  male  47.0  C.A. 30769  10.5000  Unknown  S  Unknown  NaN
301  male  47.0      36967  34.0208      D46      S  Unknown  NaN
511  male  62.0  240276   9.6875  Unknown  Q  Unknown  NaN
624  female  2.0  347082  31.2750  Unknown  S  Unknown  NaN
940  male  28.0  349253   7.8958  Unknown  C  Unknown  NaN
50   female  58.0  PC 17755  512.3292  B51 B53 B55  C      3  NaN

```

```

home.dest  family
418        Cornwall  0.0
301      East Orange, NJ  0.0
511      Cambridge, MA  0.0
624      Sweden Winnipeg, MN  6.0
940      New York, NY  0.0
50  Germantown, Philadelphia, PA  1.0

```

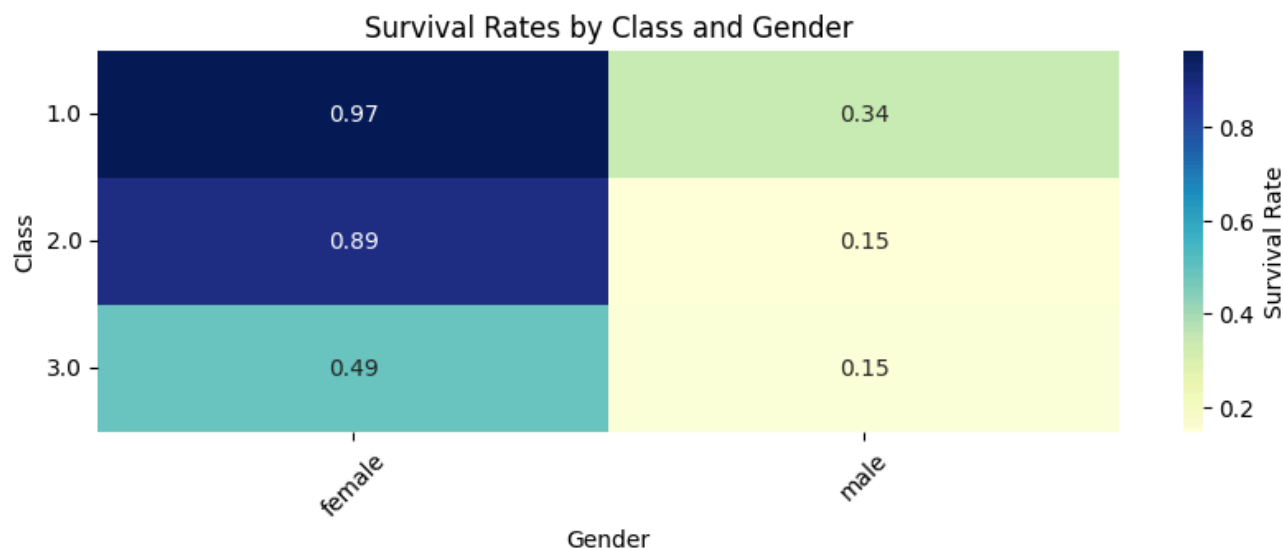
	pclass	survived	name	sex	age	ticket	fare	cabin	embarked	l
0	1.0	1.0	Allen, Miss. Elisabeth Walton	female	29.0000	24160	211.3375	B5	S	
1	1.0	1.0	Allison, Master. Hudson Trevor	male	0.9167	113781	151.5500	C22 C26	S	
2	1.0	0.0	Allison, Miss. Helen Loraine	female	2.0000	113781	151.5500	C22 C26	S	Unkn
3	1.0	0.0	Allison, Mr. Hudson Joshua Creighton	male	30.0000	113781	151.5500	C22 C26	S	Unkn
4	1.0	0.0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.0000	113781	151.5500	C22 C26	S	Unkn
...
1305	3.0	0.0	Zabour, Miss. Thamine	female	28.0000	2665	14.4542	Unknown	C	Unkn
1306	3.0	0.0	Zakarian, Mr. Mapriededer	male	26.5000	2656	7.2250	Unknown	C	Unkn
1307	3.0	0.0	Zakarian, Mr. Ortin	male	27.0000	2670	7.2250	Unknown	C	Unkn
1308	3.0	0.0	Zimmerman, Mr. Leo	male	29.0000	315082	7.8750	Unknown	S	Unkn
1309	3.0	0.0	Unknown	male	28.0000	CA. 2343	14.4542	Unknown	S	Unkn

1310 rows × 13 columns

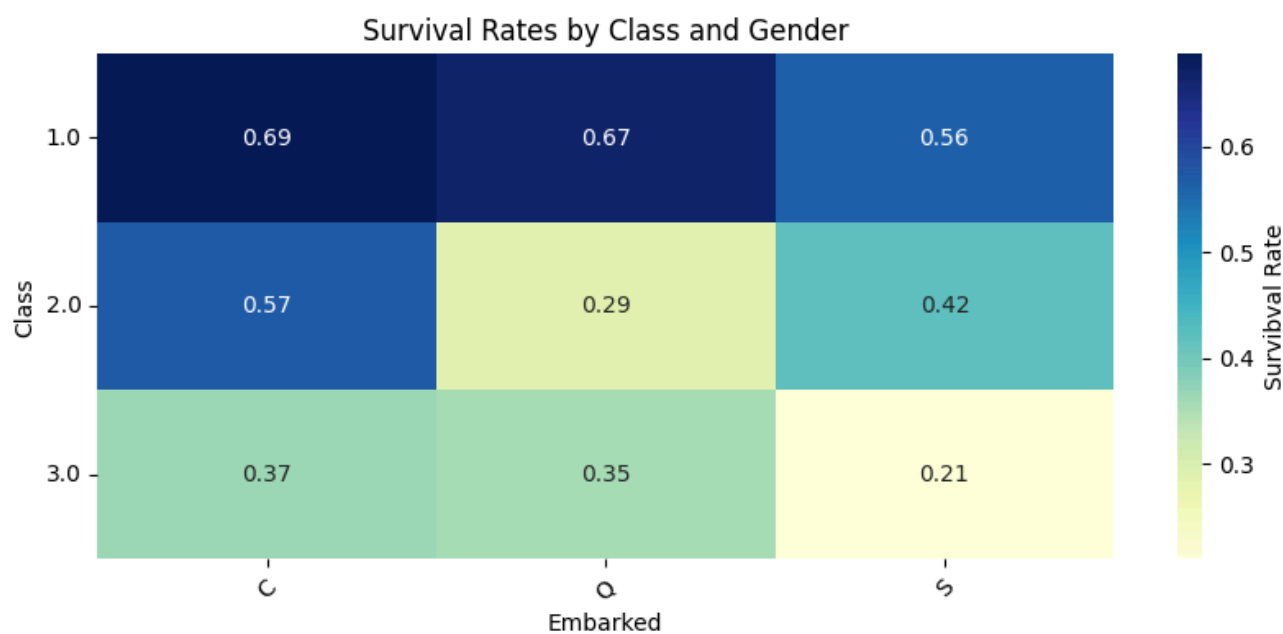


5. Analysis of Relationships Between Data 🔍

We can observe that nearly every woman in first and second class survived. On the other hand, far fewer men survived compared to women.

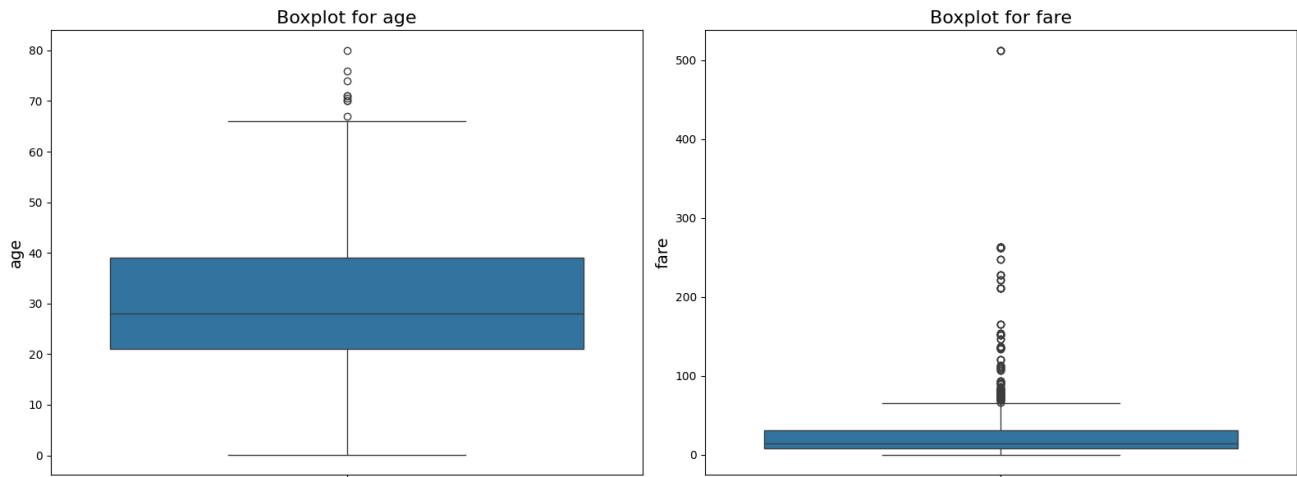


The highest chances of survival were for those who embarked from port C. 🚢




6. Analysis of Outliers 📊

- We can observe that the more expensive the ticket, the higher the chance of survival.
- The most outliers were found in the first class, specifically regarding ticket prices.



7. Analysis Summary

- The data turned out to be somewhat inconvenient for analysis due to many missing values.
- This led to the necessity of data transformation by calculating averages or transforming columns.
- About 33% of the passengers survived.
- We can observe that nearly every woman in first and second class survived.
- The highest chances of survival were for those who embarked from port C.
- We can observe that the more expensive the ticket, the higher the chance of survival.
- The most outliers were found in the first class, specifically regarding ticket prices.

Thank you for your attention!  Your interest and time mean a lot.

```
[NbConvertApp] Converting notebook titanic.ipynb to slides  
[NbConvertApp] WARNING | Alternative text is missing on 4 image(s).  
[NbConvertApp] Writing 470293 bytes to titanic.slides.html
```

```
[NbConvertApp] Converting notebook titanic.ipynb to html  
[NbConvertApp] WARNING | Alternative text is missing on 4 image(s).  
[NbConvertApp] Writing 472646 bytes to titanic2.html
```