

Titanic: Ketika Norma Sosial Menentukan Siapa yang Bertahan!

Without any further ado, lets talk about this with data.

Titanic Tragedy as Representation of Human Vanity

Tragedi tenggelamnya Titanic bukan sekadar kisah kapal mewah yang karam, tapi juga potret nyata betapa manusia bisa terlalu percaya diri pada ciptaannya sendiri. Disebut sebagai kapal “yang tak bisa tenggelam” Titanic membawa harapan, kemewahan, dan rasa aman yang ternyata semu. Di balik gemerlapnya, ada kesombongan bahwa teknologi bisa menaklukkan alam. Tapi nyatanya, satu gunung es cukup untuk menghancurkan semuanya. Dari peristiwa ini, kita diingatkan bahwa sehebat apa pun manusia, kita tetap rapuh di hadapan kenyataan yang tak bisa dikendalikan.



Will the morale stay strong?

Lalu siapa yang benar-benar diselamatkan saat Titanic tenggelam? Wanita? Pria? Atau mereka yang paling kuat dan nekat? Di balik tragedi ini, ada kisah tentang prioritas dan moral yang akan diuji. 🧐

Data yang akan menjadi saksi atas apa yang terjadi pada kejadian Titanic saat itu. 😊

Data that leads us to insights

	survived	name	sex	age
0	1	Allen, Miss. Elisabeth Walton	female	29.0000
1	1	Allison, Master. Hudson Trevor	male	0.9167
2	0	Allison, Miss. Helen Loraine	female	2.0000
3	0	Allison, Mr. Hudson Joshua Creighton	male	30.0000
4	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.0000

Kita cuma punya empat kolom: siapa mereka, usianya, jenis kelaminnya, dan apakah mereka selamat atau nggak. Terlihat simpel, **tapi justru dari data sederhana ini kita bisa menggali cerita besar di balik tragedi Titanic**. Siapa yang paling mungkin selamat? Dan kenapa bisa begitu? Yuk, kita olah datanya dan lihat apa yang bisa terungkap.

Tapi sebelum masuk lebih jauh, kita perlu bersihin dulu datanya alias **data cleaning**. Biar nanti analisisnya nggak ngaco. Oke?

Note:

Data yang digunakan dalam analisis ini berasal dari pelatihan [Dibimbing Digital Fair 38.0](#). Program ini sangat direkomendasikan bagi yang ingin mengenal dunia data analysis dan data science lebih dalam. Worth it banget buat diikuti! 😊

Data Cleaning

Sebelum kita bisa gali lebih dalam cerita dari data Titanic, ada satu hal penting yang harus kita lakukan dulu: data cleaning. Yup, proses ini penting banget supaya analisis kita nanti nggak misleading.

Nah, ada lima langkah pembersihan data yang bakal kita lewati:

1. Cleaning Type
2. Cleaning Missing Value
3. Cleaning Duplicate Value
4. Cleaning Inconsistent Data
5. Cleaning Outlier

Kalau lima langkah ini udah beres, baru deh kita bisa lanjut ke analisis yang lebih dalam. Yuk, mulai dari langkah pertama!

Cleaning Type

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  -
0   survived    500 non-null    int64
1   name        500 non-null    object
2   sex         500 non-null    object
3   age         451 non-null    float64
dtypes: float64(1), int64(1), object(2)
memory usage: 15.8+ KB
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  -
0   survived    500 non-null    int64
1   name        500 non-null    object
2   sex         500 non-null    category
3   age         451 non-null    float64
dtypes: category(1), float64(1), int64(1), object(1)
memory usage: 12.5+ KB
```

Ternyata, kolom Sex di data kita masih bertipe object. Sebenarnya sih nggak masalah kalau dibiarkan begitu aja. Tapi, karena kita mau membiasakan diri dengan praktik yang baik (apalagi kalau ke depannya kita bakal ngurusin big data) ada baiknya kita ubah tipe datanya ke category.

Kenapa? Karena **data bertipe category** itu jauh lebih efisien, baik dari sisi **memori maupun kecepatan saat kita eksplorasi data (EDA)**. Jadi, walau kelihatannya sepele, langkah kecil kayak gini bisa kasih dampak besar nanti.

Cleaning Missing Value

```
df.isnull().sum()
✓ 0.0s
survived    0
name        0
sex         0
age         49
dtype: int64
```

```
count    451.000000
mean     35.917775
std      14.766454
min       0.666700
25%      24.000000
50%      35.000000
75%      47.000000
max       80.000000
Name: age, dtype: float64
```

```
df.isnull().sum()
✓ 0.0s
survived    0
name        0
sex         0
age         0
dtype: int64
```

Ternyata, setelah kita cek, ada kolom Age yang memiliki **missing value sebanyak 49 data**. Di sini, kita nggak bisa sembarangan mengisi nilai yang hilang, karena itu bisa bikin hasil kita jadi bias. Jadi, langkah pertama yang perlu kita lakukan adalah **cek distribusi data ini terlebih dahulu, apakah normal atau justru ekstrem?**

Setelah kita lihat melalui fungsi **describe()**, ternyata distribusinya normal karena nilai mean dan median sangat dekat satu sama lain. Dengan begitu, kita bisa mengisi nilai yang hilang dengan menggunakan **mean()**, karena cara ini paling masuk akal untuk data dengan distribusi yang normal.

Cleaning Duplicate Data

```
duplicates = df[df.duplicated(keep=False)]
```

✓ 0.0s

```
duplicates
```

✓ 0.0s

	survived		name	sex	age
104	1		Eustis, Miss. Elizabeth Mussey	female	54.0
349	1		Eustis, Miss. Elizabeth Mussey	female	54.0

```
df = df.drop_duplicates()
```

✓ 0.0s

```
(len(df.drop_duplicates()))/len(df)
```

✓ 0.0s

1.0

Waktu kita cek data pakai `drop_duplicates()`, ternyata masih ada data yang terduplikasi. Nah, karena data duplikat bisa bikin hasil analisis jadi bias, maka kita putuskan untuk hapus salah satunya.

Setelah proses penghapusan selesai, kita cek ulang dan hasilnya udah aman. Ini juga diperkuat dengan perhitungan rasio `jumlah_tanpa_duplikat / jumlah_total` yang hasilnya 1, artinya udah nggak ada lagi data yang dobel. Oh ya, trik ini juga aku dapetin dari [Dibimbing Digital Fair 38.0](#) (insight kecil yang ternyata sangat berguna). 😊

Cleaning Inconsistent Data

	survived	age
count	499.000000	499.000000
mean	0.539078	35.881538
std	0.498971	14.013337
min	0.000000	0.666700
25%	0.000000	25.500000
50%	1.000000	35.917775
75%	1.000000	45.000000
max	1.000000	80.000000

age	
0.666700	1
0.833300	1
0.916700	1
28.500000	1
32.500000	1
35.917775	49
45.500000	1
Name: count, dtype: int64	

	survived	age
count	499.000000	499.000000
mean	0.539078	35.889780
std	0.498971	14.011967
min	0.000000	1.000000
25%	0.000000	25.500000
50%	1.000000	36.000000
75%	1.000000	45.000000
max	1.000000	80.000000

Nah, di tahap ini baru deh mulai kelihatan “boroknya” data kita, ternyata *banyak banget nilai yang nggak konsisten*. Tapi santai aja, kita bisa perbaiki itu semua.

Kita ambil pendekatan simpel:

1. Kalau ada bilangan float di bawah 1, kita anggap itu salah input dan kita bulatkan ke 1.
2. Sisanya, kita rapihin pakai fungsi round() biar jadi bilangan bulat yang lebih masuk akal.

Setelah kita bersihin dan cek ulang pakai describe(), hasilnya udah jauh lebih rapi dan masuk akal dibanding sebelumnya. Tinggal satu langkah terakhir nih: cek outlier. Let’s go!

Cleaning Outlier

```
Q1 = df['age'].quantile(0.25)
Q3 = df['age'].quantile(0.75)
IQR = Q3 - Q1

batas_atas = Q3 + (1.5 * IQR)
batas_bawah = Q1 - (1.5 * IQR)

outliers = df[(df['age'] > batas_atas) | (df['age'] < batas_bawah)]
```

	survived	name	sex	age
14	1	Barkworth, Mr. Algernon Henry Wilson	male	80.0
61	1	Cavendish, Mrs. Tyrell William (Julia Florence...	female	76.0

Di tahap ini, kita pakai pendekatan yang sederhana aja buat nyari outlier, yaitu dengan metode interquartile range (IQR). Aku nggak akan jelasin detail soal teorinya, karena info tentang ini udah banyak banget tersebar di internet (jadi kalian bisa pelajari sendiri kalau mau lebih dalam, yaa).

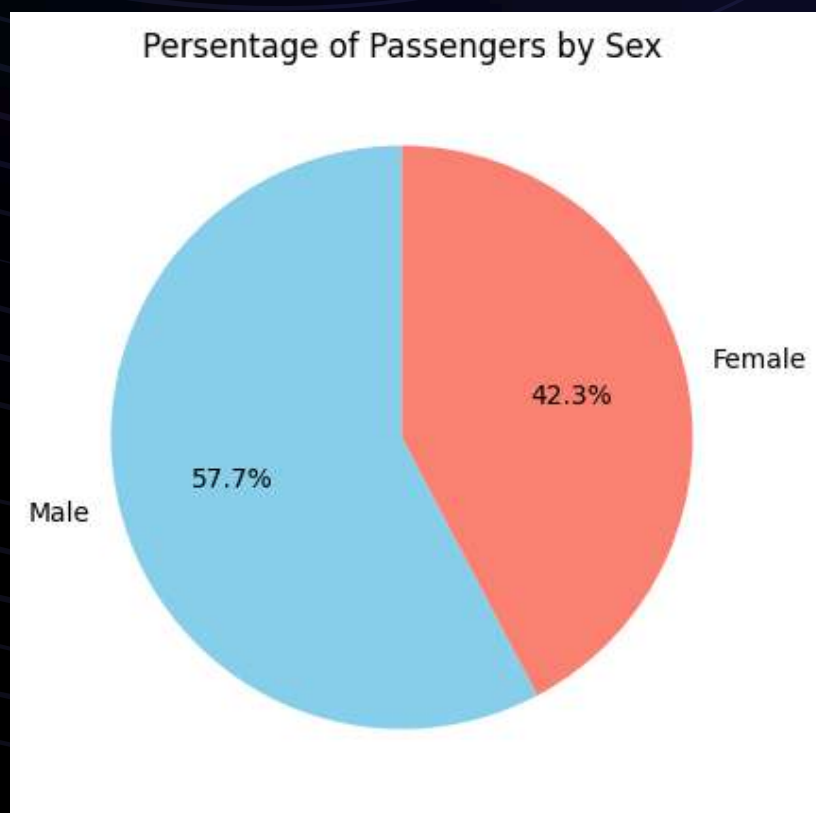
Hasilnya, kita nemuin dua outlier pada kolom umur, yaitu umur 76 dan 80. Sementara untuk kolom Sex, kita nggak lanjutkan pengecekan karena nilai uniknya cuma dua: male dan female. Jadi, nggak ada urgensi untuk cek outlier di sana. Lalu, kenapa outlier-nya nggak kita hapus? Karena dalam kasus ini, outlier tersebut nggak bikin bias. Artinya nilai mean dan distribusi data masih kelihatan representatif. Biasanya **outlier dihapus kalau kehadirannya bikin data jadi “ngaco” atau terlalu tertarik ke satu sisi**. Tapi di sini, aman.

Dan setelah melewati lima langkah pembersihan tadi, sekarang saatnya untuk... 😊

Extract Information from Data

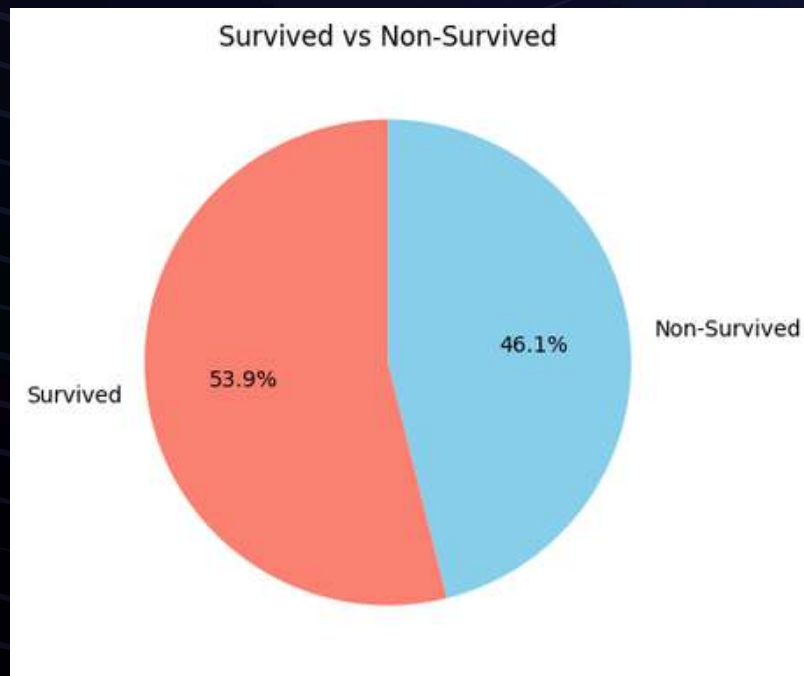
Mari kita coba visualisasikan data yang tadi kita bersihkan untuk mendapatkan insight.

Insights from Visualization 1



Dari visualisasi di samping, kita bisa lihat bahwa jumlah penumpang kapal Titanic yang ada di data kita terdiri dari sekitar 288 laki-laki dan 211 perempuan. Meskipun perbedaannya nggak terlalu mencolok, tapi tetap kelihatan kalau jumlah pria memang sedikit lebih banyak dibanding perempuan di kapal ini.

Insights from Visualization 2



Kalau kita lihat dari data keseluruhan, bisa diperhitungkan bahwa ada sekitar 269 orang yang selamat dan 230 orang yang tidak selamat dari tragedi Titanic.

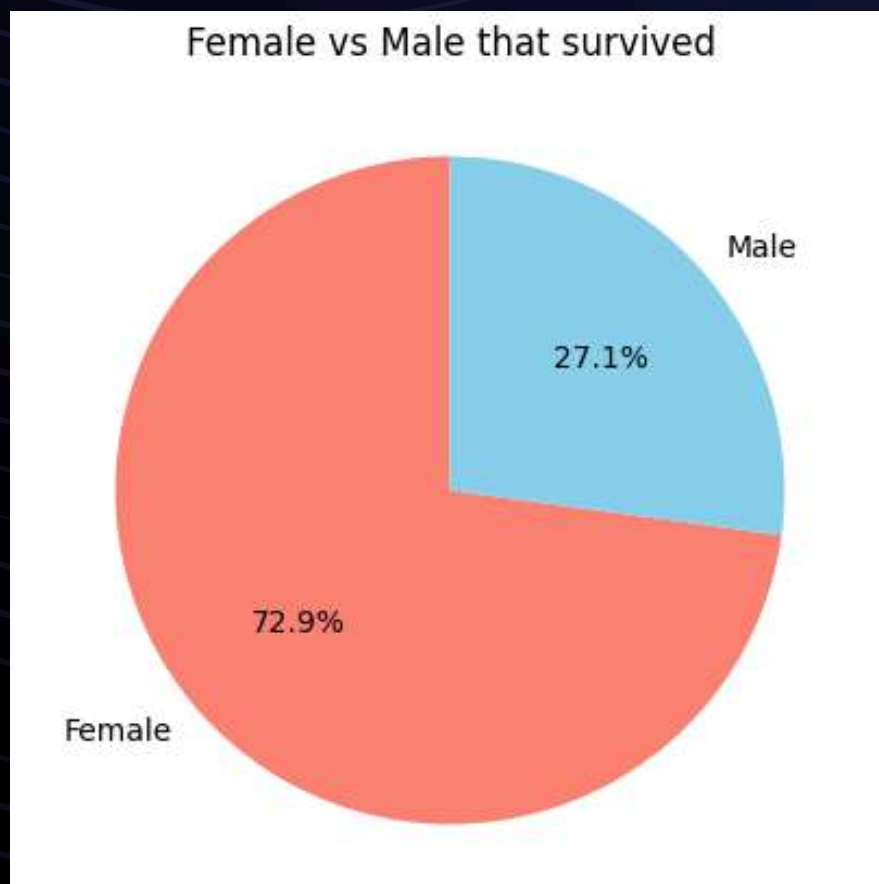
Dari data yang kita miliki, tercatat bahwa 269 orang selamat dari insiden Titanic, sementara 230 orang lainnya tidak berhasil diselamatkan. Secara kasar, itu berarti hanya sekitar 54% dari penumpang yang tercatat di data ini berhasil bertahan hidup.

Tapi angka ini bukan sekadar statistik. Di balikanya, ada pertanyaan besar:

Siapa yang diselamatkan lebih dulu? Dan kenapa, apakah karena mereka perempuan? Atau masih anak-anak?

Mari kita gali lebih dalam, karena dari data sederhana ini, kita bisa menemukan cerita besar tentang **prioritas, keputusan, dan kemanusiaan dalam kondisi darurat.**

Insights from Visualization 3

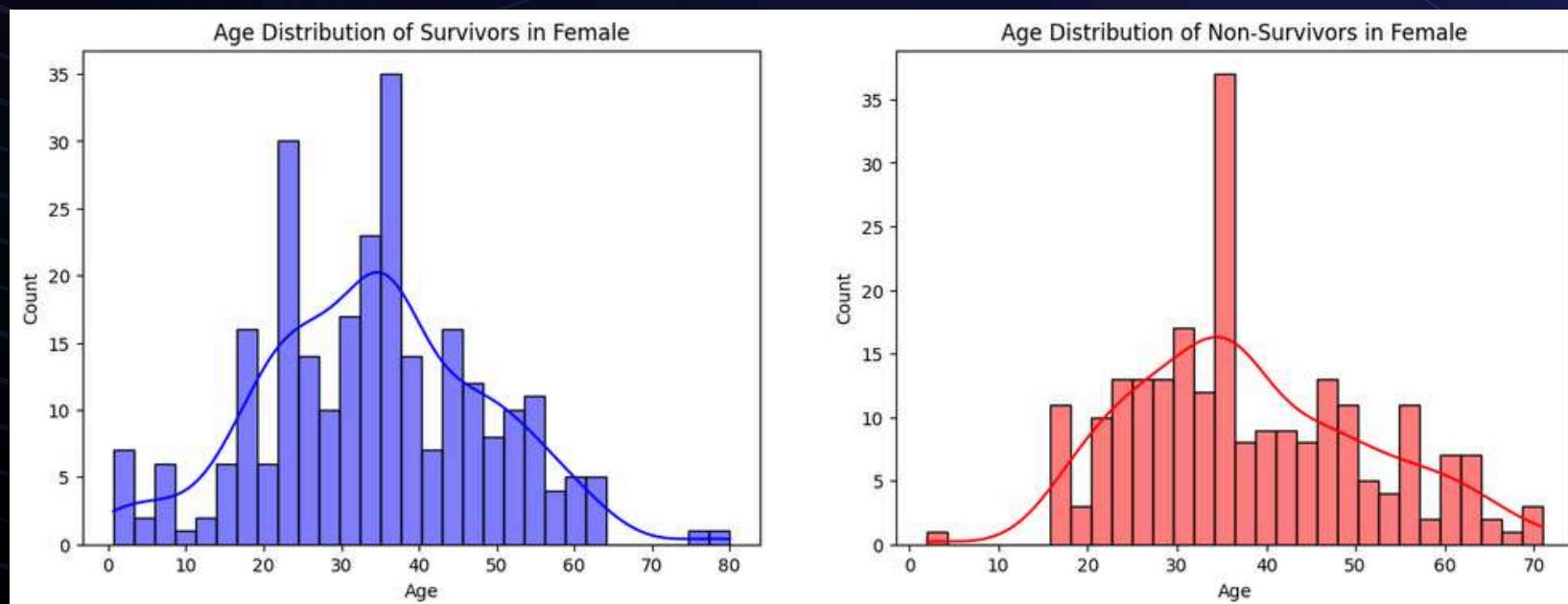


Dari data kelompok yang selamat, terlihat bahwa 73% adalah perempuan, sementara hanya 27% laki-laki. Angka ini memberi gambaran jelas bahwa pada tragedi Titanic, masih berlaku moral kuat: **perempuan didahulukan untuk diselamatkan.**

Di tengah kepanikan, moral sosial tetap terlihat memegang kendali. Prioritas penyelamatan mencerminkan norma saat itu yaitu **perempuan dianggap lebih perlu dilindungi, sementara banyak laki-laki memilih atau terpaksa untuk tetap tinggal, bahkan jika itu berarti mengorbankan nyawa mereka.**

Ini bukan hanya soal data. Ini soal bagaimana nilai, budaya, dan keputusan mendesak bisa terlihat sangat nyata dalam angka-angka yang kita lihat sekarang.

Insights from Visualization 4

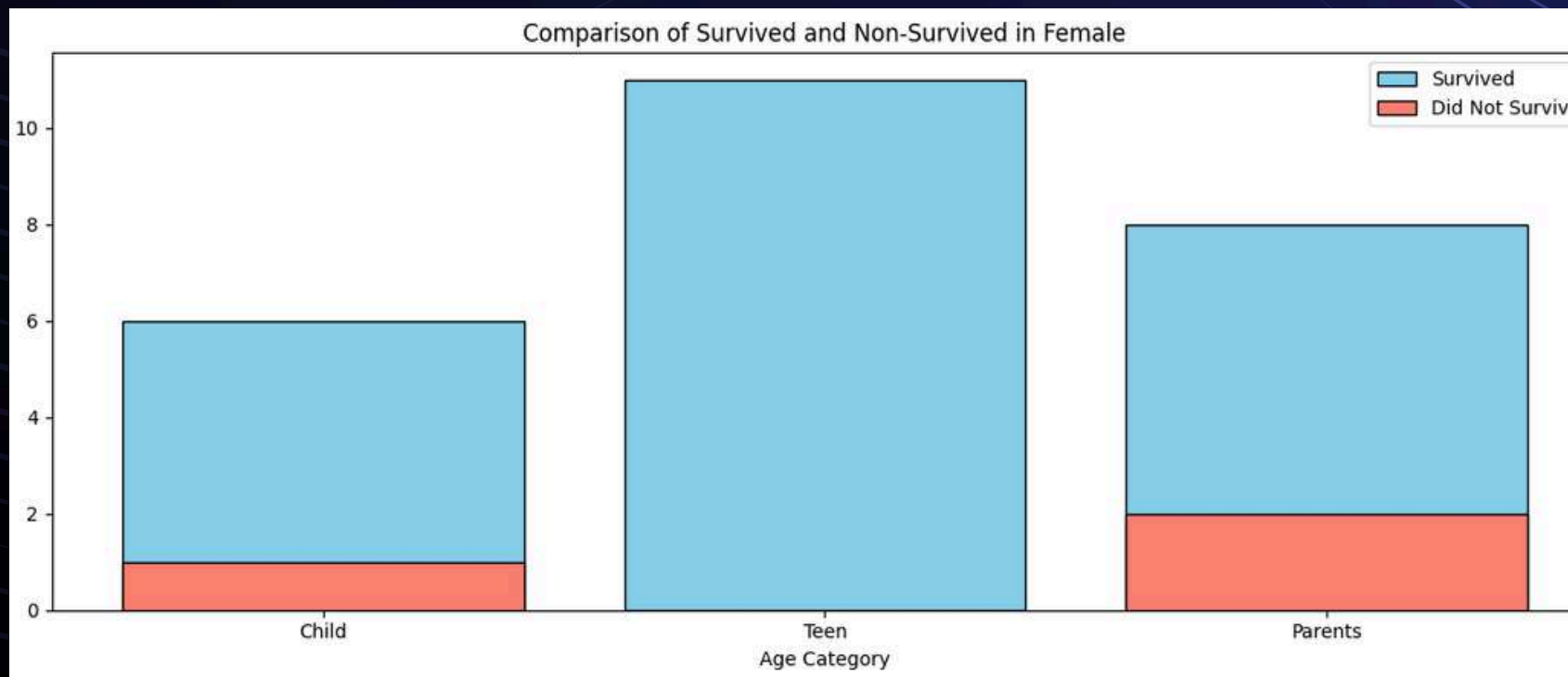


Ketika kita telusuri lebih lanjut berdasarkan rentang usia, terlihat pola yang menarik. **Anak-anak hingga remaja (teen) lebih banyak yang didahulukan untuk diselamatkan, sementara untuk kelompok dewasa hingga lansia, jumlah yang diselamatkan jauh lebih sedikit.**

Ini menegaskan bahwa, selain perempuan, **ada prioritas kuat terhadap anak-anak dalam situasi darurat.** Seolah-olah, dalam peristiwa mengerikan itu, kehidupan yang paling rentan yakni anak-anak mendapatkan perhatian utama. Sementara orang dewasa dan lansia, yang dianggap lebih mampu bertahan atau kurang prioritas, harus menghadapi kenyataan yang lebih pahit.

Ini membuka wawasan tentang bagaimana keputusan moral dan sosial dapat mempengaruhi siapa yang dipilih untuk selamat, bahkan dalam keadaan yang penuh tekanan seperti tragedi Titanic.

Insights from Visualization 5

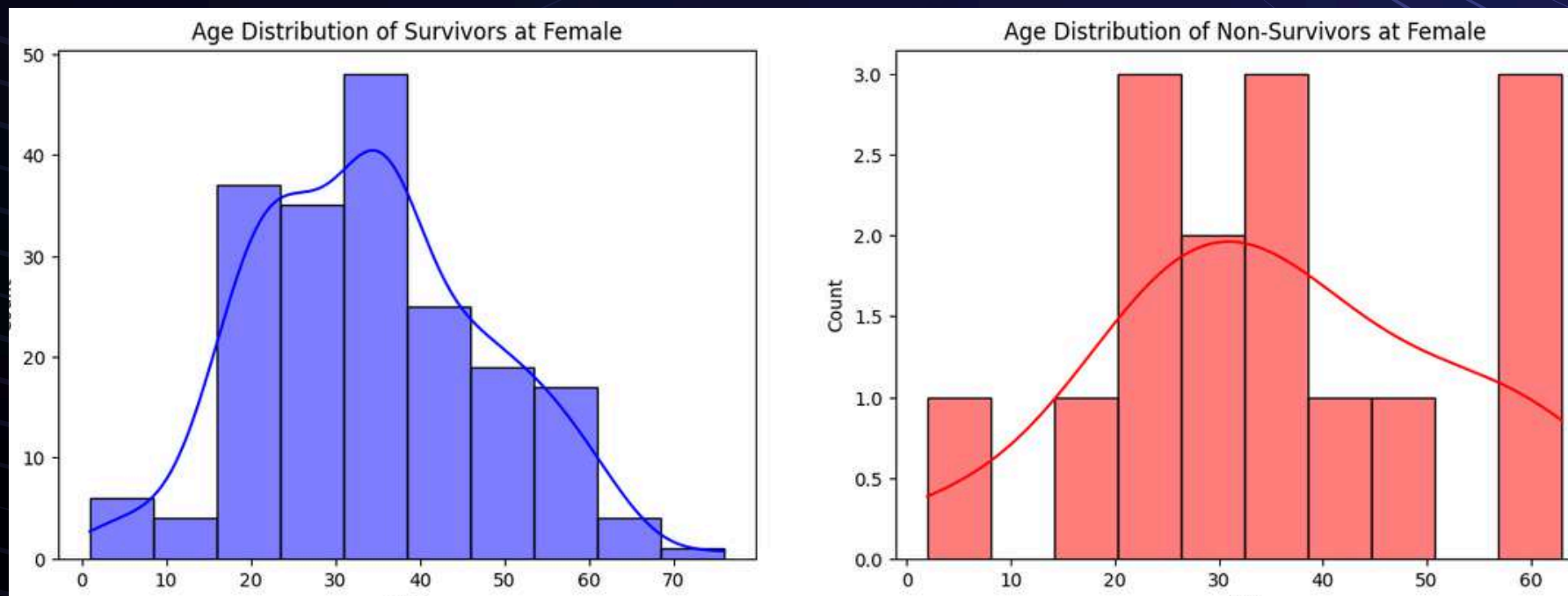


Ketika kita membagi data menjadi empat kategori: Child (0–11), Teen (12–17), Adults (18–59), dan Parents (60–100). Kita bisa melihat bahwa Teen adalah kelompok yang paling banyak diselamatkan. Hal ini tidak mengherankan karena memang **distribusi penumpang Titanic mayoritas terdiri dari kelompok usia tersebut dan kelompok usia Adults.**

Jumlah penumpang yang lebih banyak di kelompok Teen tentunya berpengaruh pada keputusan penyelamatan, mengingat mereka lebih banyak tersedia di area yang bisa dijangkau oleh sekoci. Jadi, secara alami mereka mendapatkan prioritas lebih banyak.

Namun, yang menarik dan tragis adalah anak-anak (Child) yang jumlahnya sedikit, tapi ada di antaranya yang tidak selamat. **Ini mungkin karena waktu yang terbatas atau ketidakteraturan dalam proses evakuasi,** sehingga membuat mereka yang lebih sulit dijangkau akhirnya menjadi korban tragedi tersebut.

Insights from Visualization 6

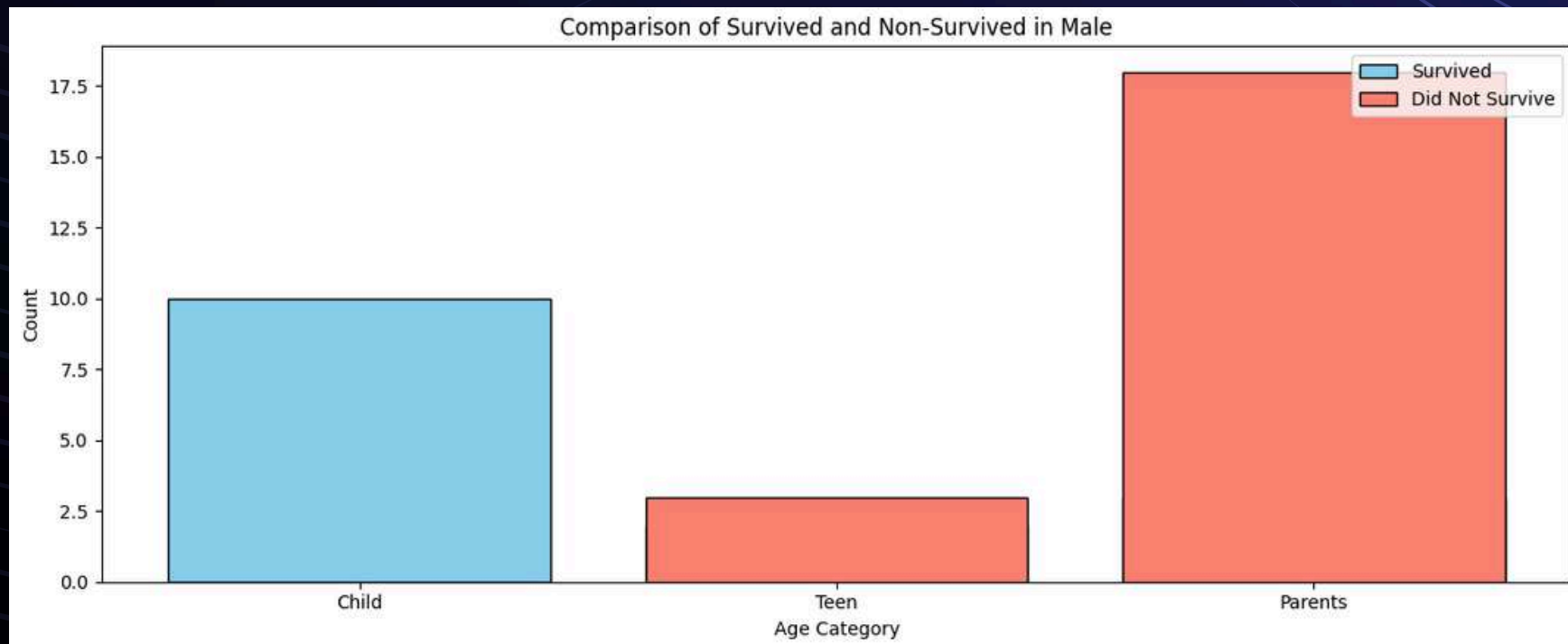


Hal ini semakin terlihat jelas ketika kita memvisualisasikan distribusi perempuan yang selamat dan tidak selamat berdasarkan usia. **Ternyata, kelompok Child dan Teen memang diprioritaskan untuk diselamatkan.** Namun, saat kita naik ke kategori Adults dan Parents, angkanya mulai menurun tajam.

Ada dua kemungkinan di balik penurunan ini. Pertama, bisa jadi memang jumlah perempuan di usia dewasa dan lanjut usia lebih sedikit, sehingga secara statistik angka penyelamatan juga menurun. Kedua, bisa jadi mereka secara sadar memilih untuk memberi jalan kepada yang lebih muda, menyerahkan tempat di sekoci, dan menerima takdir dengan tenang.

Ini menjadi potret yang sangat tragis. Titanic, yang seharusnya menjadi simbol kemewahan dan kebahagiaan, justru berubah menjadi saksi bisu bagaimana banyak orang harus merelakan hidupnya karena kurangnya persiapan dan antisipasi. Di balik angka-angka ini, kita sedang melihat kisah pengorbanan dan kepedihan yang nyata.

Insights from Visualization 7

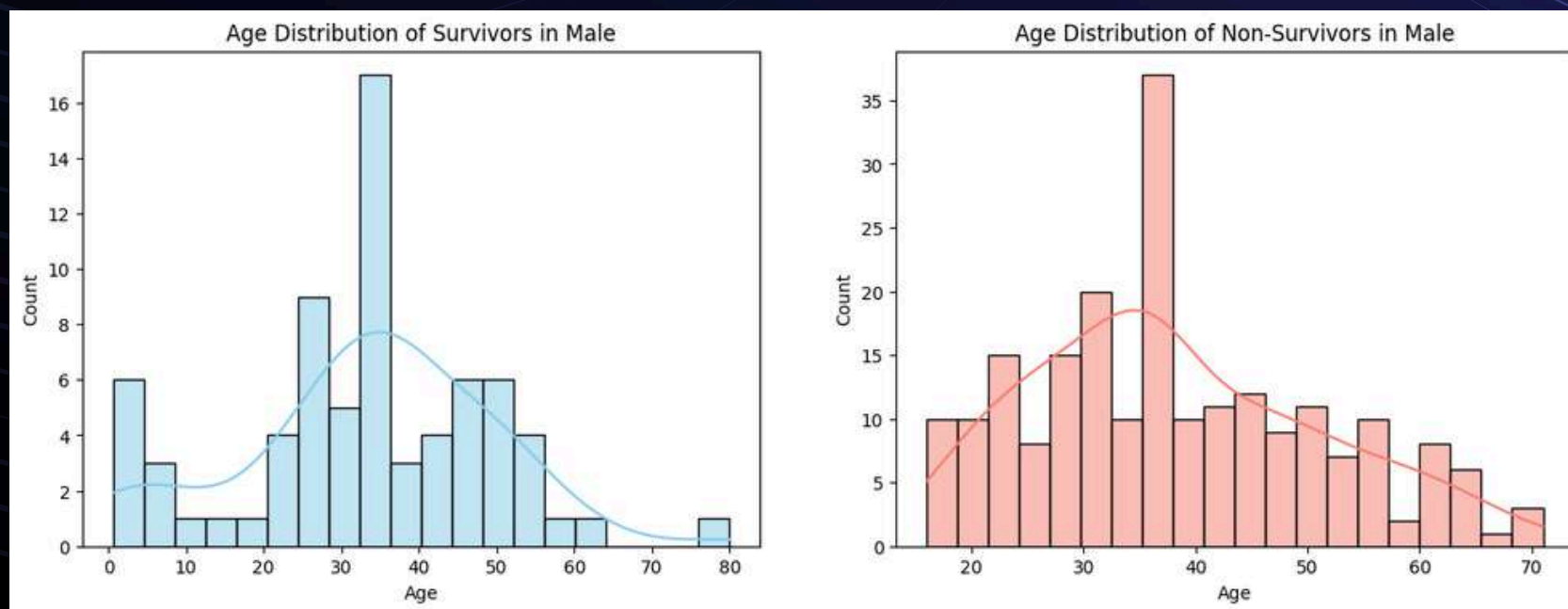


Sekarang kita beralih ke kelompok lelaki. Dari visualisasi dan data yang kita olah, **terlihat bahwa nilai moral dan norma saat itu tetap dijaga.**

Hal yang paling menarik adalah fakta bahwa **kategori Child menjadi satu-satunya kelompok laki-laki yang secara signifikan mendapatkan prioritas penyelamatan.** Sementara itu, Teen dan Parents, yang juga merupakan kelompok rentan, justru tidak mendapatkan perlakuan istimewa. Ini menunjukkan bahwa meskipun dalam keadaan panik dan kacau, **ada semacam konsensus sosial yang tidak tertulis yaitu “selamatkan anak-anak terlebih dahulu”.**

Disini tidak menyertakan kategori Adults dalam perbandingan ini karena distribusi usia sangat didominasi oleh kelompok tersebut, yang bisa menciptakan bias dan menutupi pola penting dari kategori lain. Namun dari situ bisa kita lihat bagaimana lelaki dewasa mungkin mengambil peran sebagai pelindung, dan banyak dari mereka yang memilih berkorban demi yang lebih muda.

Insights from Visualization 8



Lebih terlihat jelas bahwa lelaki benar-benar mengambil peran sebagai pelindung dalam tragedi ini. Hanya sedikit dari mereka yang akhirnya selamat dan **itu seolah menegaskan bahwa keselamatan perempuan dan anak-anak jadi prioritas utama.**

Saat aku bandingkan lebih jauh, distribusi lelaki yang tidak selamat tampak sedikit stabil di berbagai rentang usia. Ini berbeda dengan distribusi perempuan yang lebih berfluktuasi (mungkin karena jumlahnya lebih sedikit).

Namun satu hal yang tidak berubah: angka-angka ini bicara tentang pengorbanan diam-diam. **Tentang bagaimana para lelaki, terutama yang dewasa, memilih untuk tetap tinggal,** mungkin sambil menatap keluarga mereka dari kejauhan dengan satu harapan bahwa mereka yang dicintai bisa hidup lebih lama.

When the data talks about morality in humans!

Dari sekadar empat kolom data (name, age, sex, and survived), aku menemukan lebih dari sekadar angka. Aku melihat cerita. Cerita tentang prioritas moral, tentang siapa yang dipilih untuk hidup, dan siapa yang rela melepaskan kesempatan itu.

Perempuan dan anak-anak, terutama di usia child dan teen, menjadi kelompok yang paling diprioritaskan untuk selamat. Sementara itu, banyak lelaki dewasa memilih tetap tinggal (mungkin bukan karena tak sempat diselamatkan, tapi karena mereka memilih untuk tidak pergi lebih dulu).

Data ini juga menggambarkan bagaimana norma sosial tetap hidup bahkan dalam kekacauan. Seperti memberi kita isyarat bahwa manusia tetap punya hati walaupun berada di tengah kepanikan.

Tragedi Titanic bukan hanya soal kapal megah yang tenggelam. Ini adalah cermin yang menunjukkan bahwa di balik teknologi dan kemewahan, ada kesombongan, ada kelemahan, tapi juga ada keberanian dan cinta.

Dan semua itu bisa ditemukan... hanya dari sebaris data.