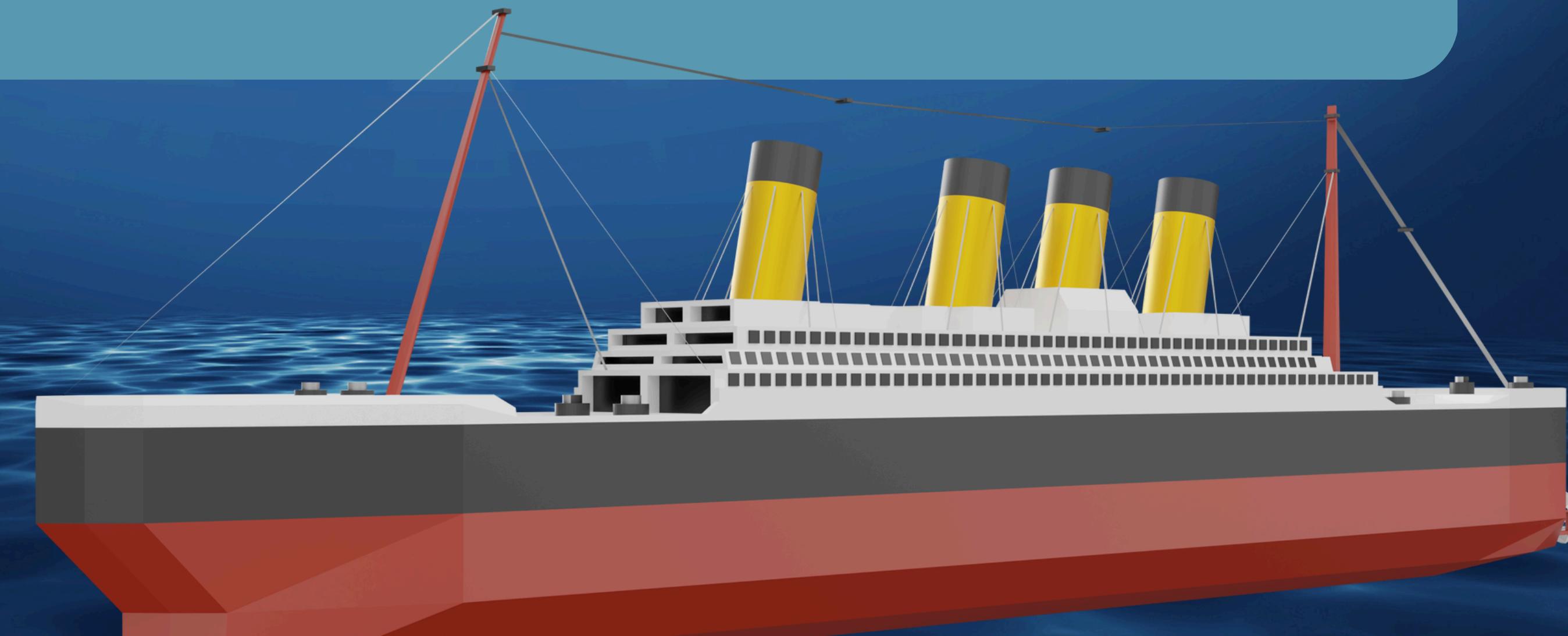


Predicting Titanic Survival

Logistic Regression vs. Random Forest

A Comparative Analysis

By Theran Meadows



Disasterous Overview

March 1912
The Titanic
has been
completed

2240 passengers
board the ship
nicknamed:
“Unsinkable”

April 1912
The ship hits
an iceberg on
its starboard
side and
begins to sink

The crew
begin to fill
the 20
lifeboats on
board.

2 hours and 40
minutes later, the ship
sank to the bottom of
the Atlantic

1570 people
perish from
the freezing
waters

May 2009
The last survivor
of the Titanic dies
at the age of 97.
She was 3 months
old when she
boarded with her
family.

Dataset Overview

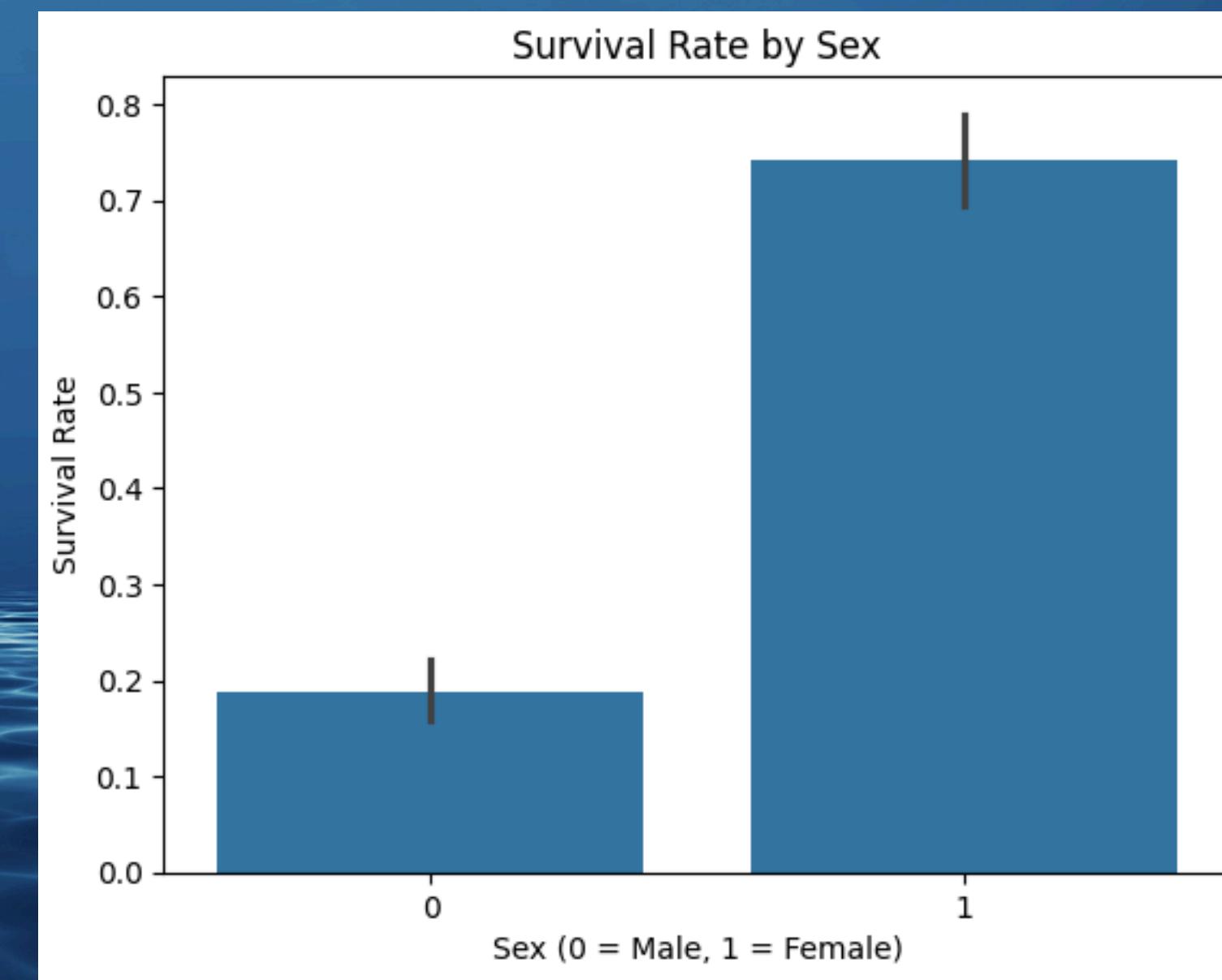
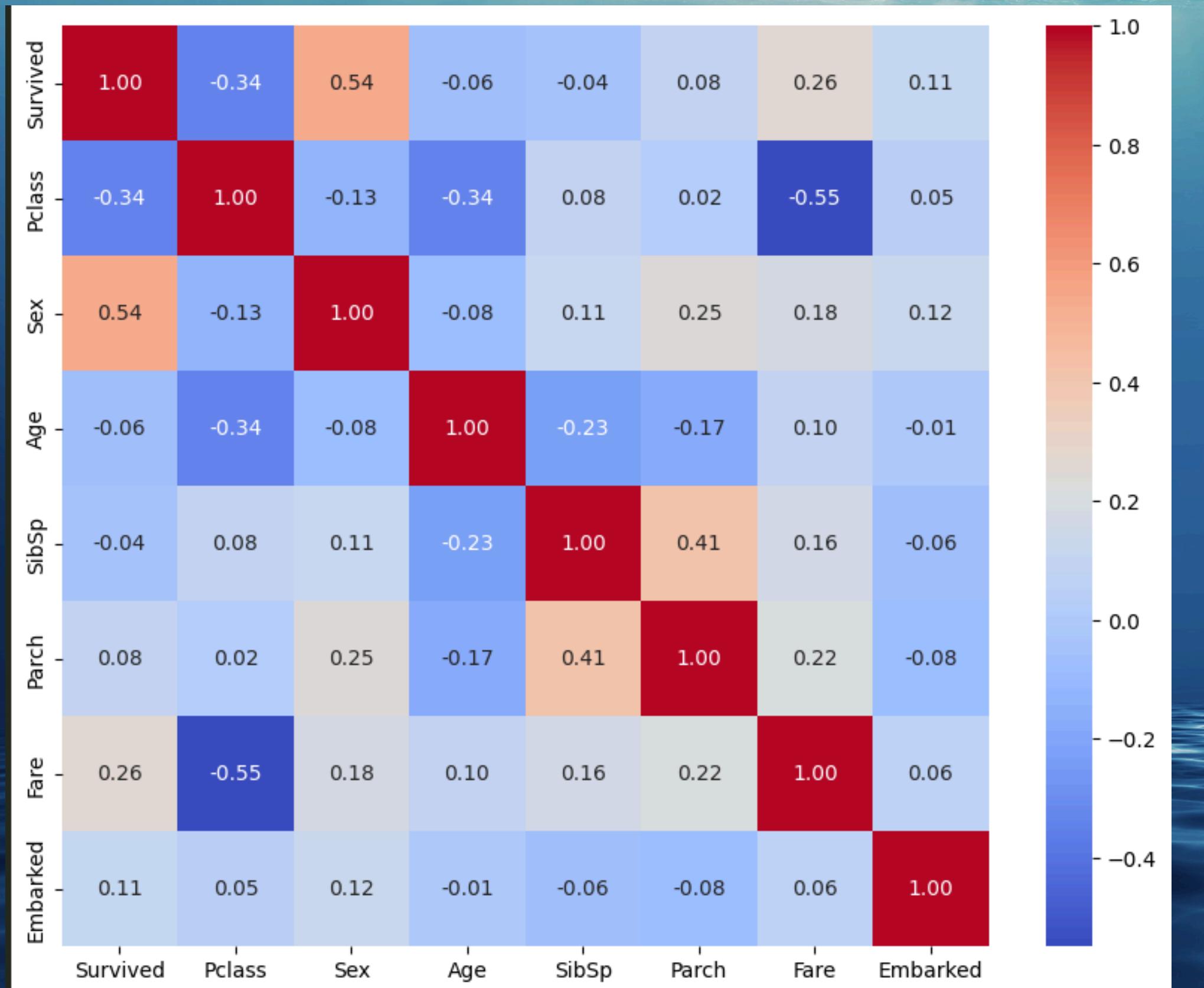
- 891
 - Pclass: 1, 2, 3
 - Age
 - Sibsp
 - Brother, Sister
 - Stepbro./ sis.
 - Husband, Wife
 - Parch
 - Mother, Father
 - Daughter, Son
 - Stepdau./ son
 - Embarked: C, Q, S

- Passenger_id
 - Name
 - Ticket_number
 - Cabin

Target Variable:
Survived: 0, 1

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	Cumings, Mrs. John Bradley (Florence Briggs Th... Heikkinen, Miss. Laina	female	38.0	1	0	PC 17599 STON/O2. 3101282	71.2833	C85	C
2	3	1	3 Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	26.0	0	0	113803	7.9250	NaN	S
3	4	1		female	35.0	1	0		53.1000	C123	S

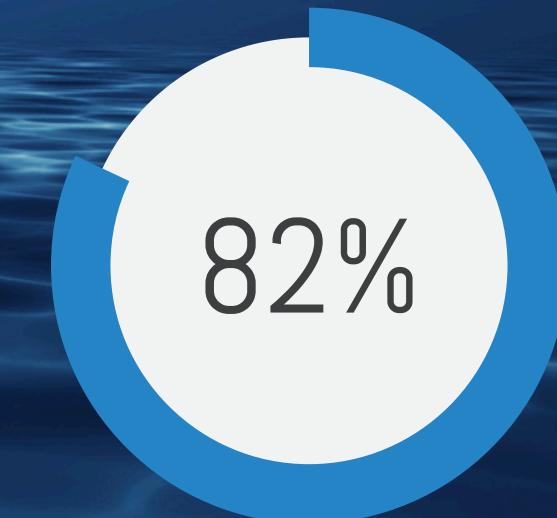
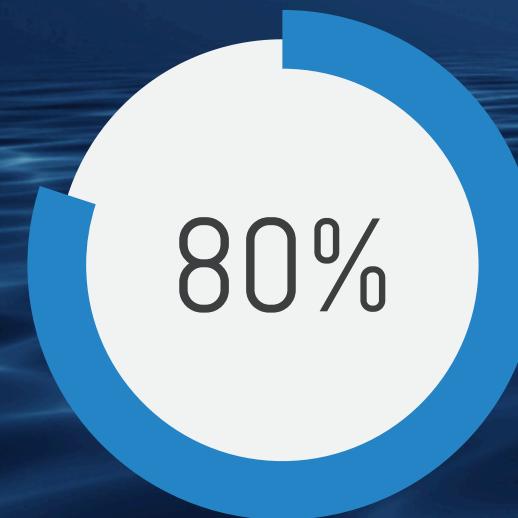
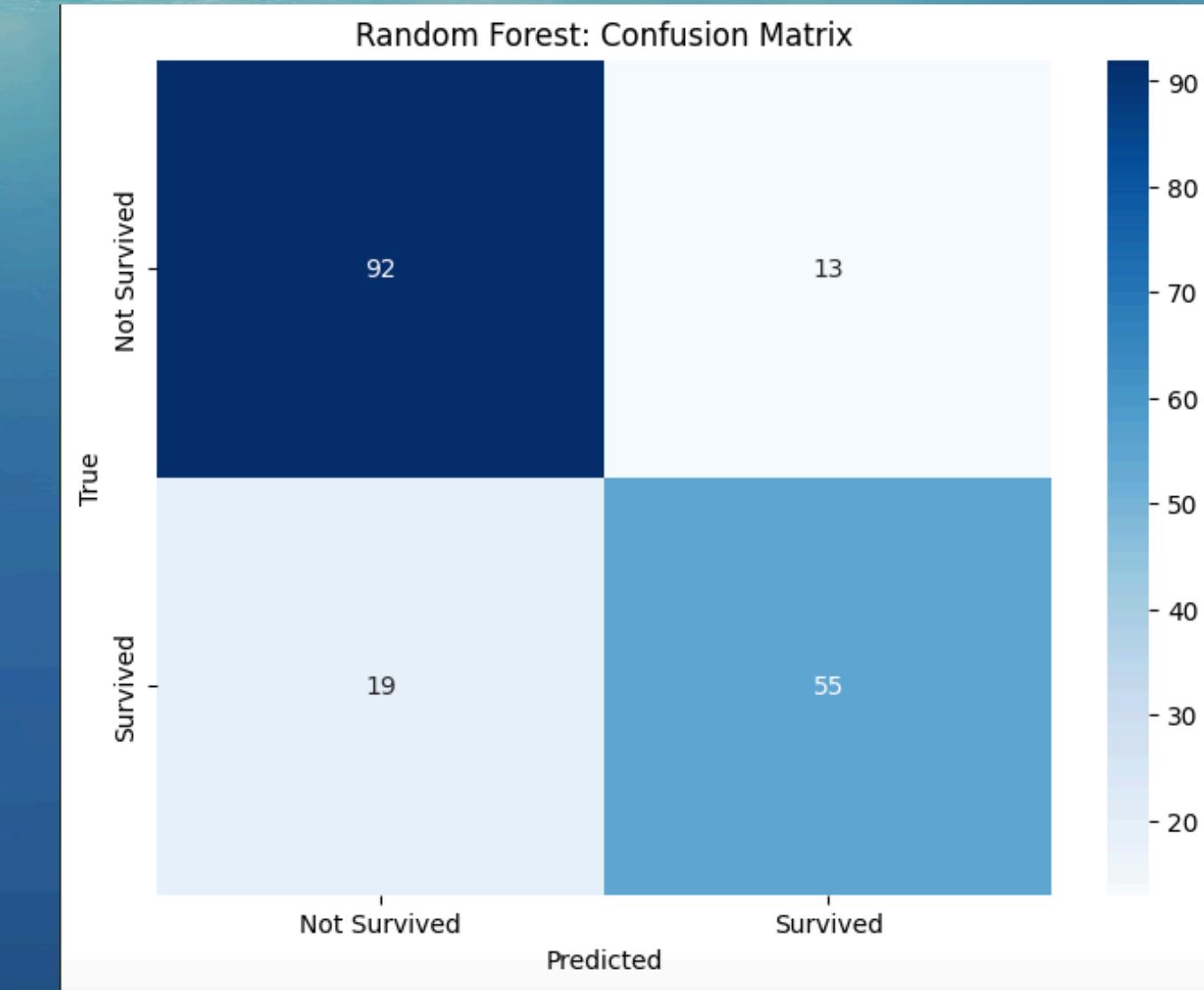
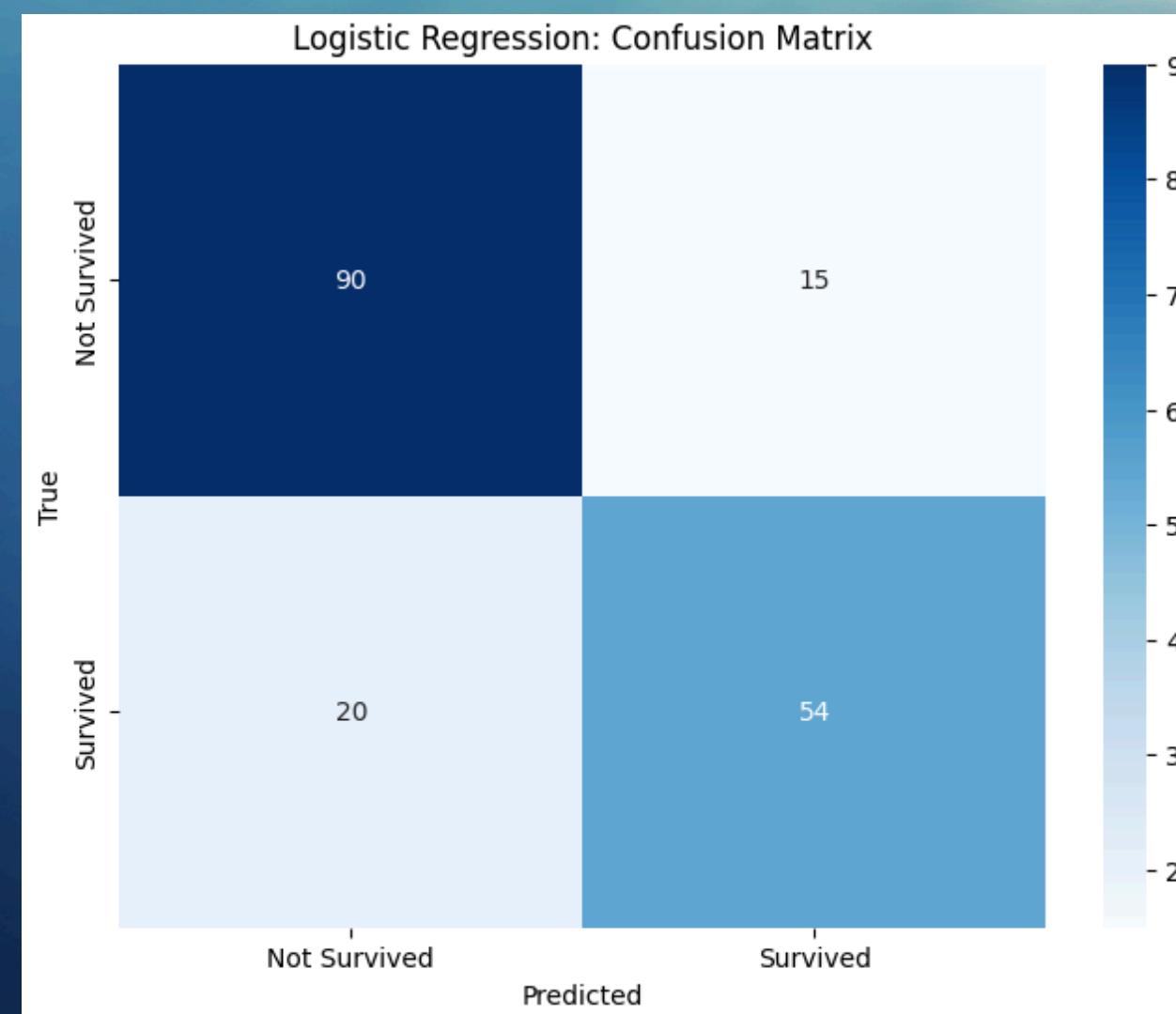
Exploratory Data Analysis



Predictions

- Gender plays a significant role in survival
 - “Women and children only!”
 - Socio-economic class
 - Rich young females
-
- Men of the lower class have the lowest survival rate
 - The more family you have, the more likely you are to stay behind

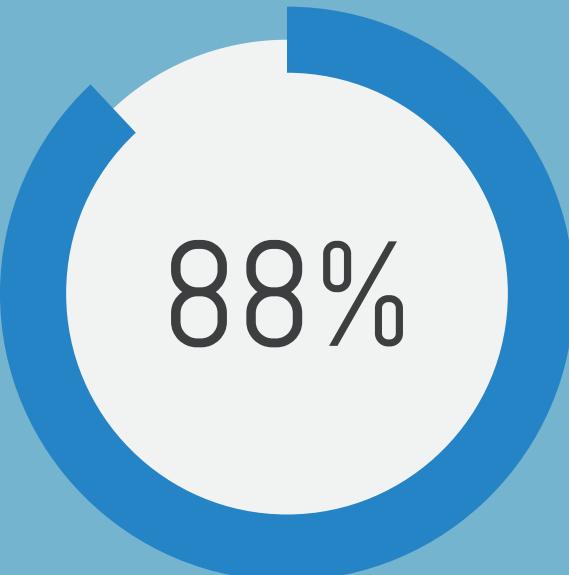
Logistic Regression vs Random Forest



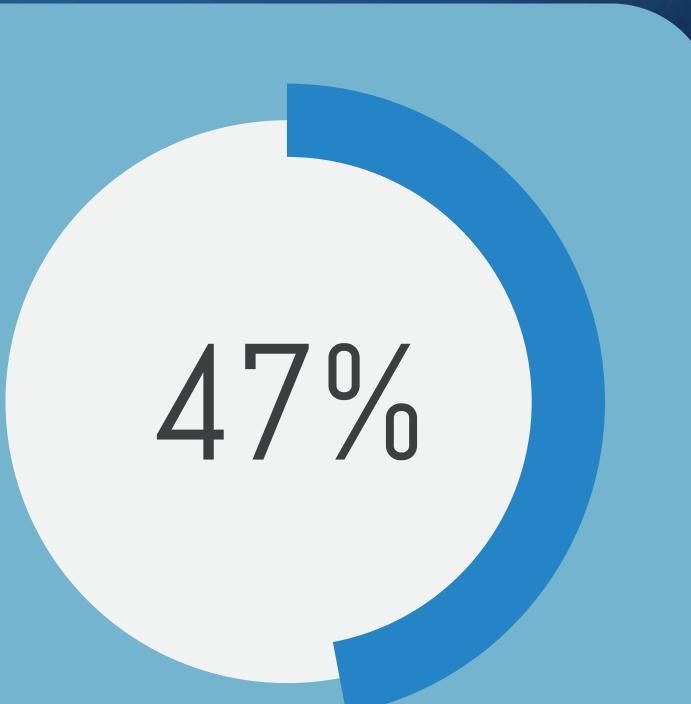
Recommendations

- Some things you can't control
- Be in a higher socio-economic class
- More lifeboats
 - lifeboats with higher capacity
 - Inflatable life-rafts

```
Rich_family_man = {  
    'Pclass': 1,  
    'Sex': 'male',  
    'Age': 30,  
    'SibSp': 1,  
    'Parch': 2,  
    'Fare': 150,  
    'Embarked': 'S'}
```



```
Poor_family_man = {  
    'Pclass': 3,  
    'Sex': 'male',  
    'Age': 30,  
    'SibSp': 1,  
    'Parch': 2,  
    'Fare': 7,  
    'Embarked': 'S'}
```



Conclusion

- Logistic Regression and Random Forest models were both fairly close and accurate
- Gender, class, number of family members play the biggest story of a passengers survival
- Rich, young, lonely females have the highest survival
- Poor, old , family men have the lowest survival