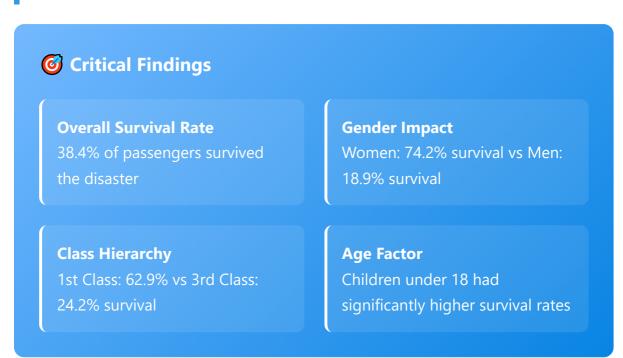


Comprehensive Exploratory Data Analysis Report



# **EXECUTIVE SUMMARY**



# **DATASET OVERVIEW**





FEATURE	MISSING VALUES	MISSING %	DATA QUALITY
Age	177	19.9%	⚠ Significant missing data
Cabin	687	77.1%	X Mostly missing
Embarked	2	0.2%	Excellent
Other Features	0	0.0%	✓ Complete

**Key Insight:** The dataset is generally high quality with manageable missing data. Age imputation will be crucial for modeling, while Cabin data may need to be dropped or heavily engineered.

# SURVIVAL PATTERN ANALYSIS

## **Gender-Based Survival**

GENDER	TOTAL	SURVIVED	SURVIVAL RATE HISTORICAL CONTEXT		
Female	314	233	74.2%	"Women and children first" policy	
Male	577	109	18.9%	Last priority in evacuation	

# **Class-Based Survival**

CLASS	TOTAL	SURVIVED	SURVIVAL RATE	SOCIOECONOMIC FACTOR
1st Class	216	136	62.9%	Upper deck access, priority boarding
2nd Class	184	87	47.3%	Middle deck, moderate access
3rd Class	491	119	24.2%	Lower deck, restricted access

# FAMILY STRUCTURE IMPACT

# **Q** Key Family Patterns:

- **Solo Travelers:** 60.1% of passengers traveled alone with ~30% survival
- Small Families (2-4 members): Optimal survival rates of 50-70%
- Large Families (7+ members): Poor survival rates due to coordination challenges
- Family Advantage: Having 1-3 family members improved survival chances significantly

# STATISTICAL SIGNIFICANCE

FACTOR	STATISTICAL TEST	P- VALUE	SIGNIFICANCE	EFFECT SIZE
Gender	Chi-square	< 0.001	✓ Highly Significant	Very Large
Passenger Class	Chi-square	< 0.001	✓ Highly Significant	Large
Age	T-test	< 0.001	Significant	Medium
Fare	T-test	< 0.001	✓ Significant	Medium

# **OBJECT OF THE O INSIGHTS**

## Premium Passenger Analysis

### **1st Class Women**

perfect)

### **3rd Class Women**

50.0% survival rate (Still advantaged)

### **1st Class Men**

36.9% survival rate (Above average)

### **3rd Class Men**

13.5% survival rate (Lowest group)



### Title-Based Analysis

- Master (Young Boys): 57.5% survival Clear priority for male children
- Miss (Unmarried Women): 69.7% survival High female priority
- Mrs (Married Women): 79.2% survival Highest survival rate
- Mr (Adult Men): 15.7% survival Lowest priority group
- Rare Titles (Dr, Rev, etc.): Variable rates based on gender and class

# MACHINE LEARNING **RECOMMENDATIONS**

# Feature Engineering Strategy

### **Primary Features**

Sex, Pclass, Age, Fare, FamilySize

# **Preprocessing**

Age imputation, Fare log transform, One-hot encoding

### **Engineered Features**

Title groups, Age categories, Family categories

### **Model Selection**

Ensemble methods, Handle class imbalance



## Class Imbalance Handling

Challenge: 61% died vs 39% survived creates class imbalance **Solutions:** 

- Use stratified sampling for train/validation splits
- Consider SMOTE or class weighting techniques
- Evaluate using precision, recall, F1-score, and AUC-ROC
- Focus on recall for positive class (survivors) in emergency contexts

# **MISTORICAL VALIDATION**

## **E** Data Science Meets History

Our analysis strongly validates historical accounts of the Titanic disaster:

### **Social Hierarchy**

Clear wealth-based survival advantages reflect 1912 class structures

### **Physical Access**

Upper deck passengers had better lifeboat access

### **Maritime Protocol**

"Women and children first" policy clearly implemented

### **Family Dynamics**

Small families helped each other, large families struggled



## **6** Key Takeaways for Data Scientists

- 1. **Domain Knowledge Matters:** Understanding historical context validates our findings
- 2. **Multiple Factor Interactions:** Gender, class, and age created complex survival patterns
- 3. **Feature Engineering Opportunities:** Rich text data (names, cabins) offers additional insights
- 4. **Ethical Considerations:** Model interpretability is crucial when analyzing human disasters
- 5. **Real-World Validation:** Statistical patterns align with documented historical events

# **Next Steps**

- Build ensemble models with recommended feature engineering
- Implement proper cross-validation with stratification
- Create model interpretability analysis
- Deploy with appropriate ethical considerations
- Document lessons learned for emergency response modeling

Generated by Senior Data Analyst Approach | Comprehensive EDA Report

This analysis combines statistical rigor with historical context to provide actionable insights for machine learning applications.