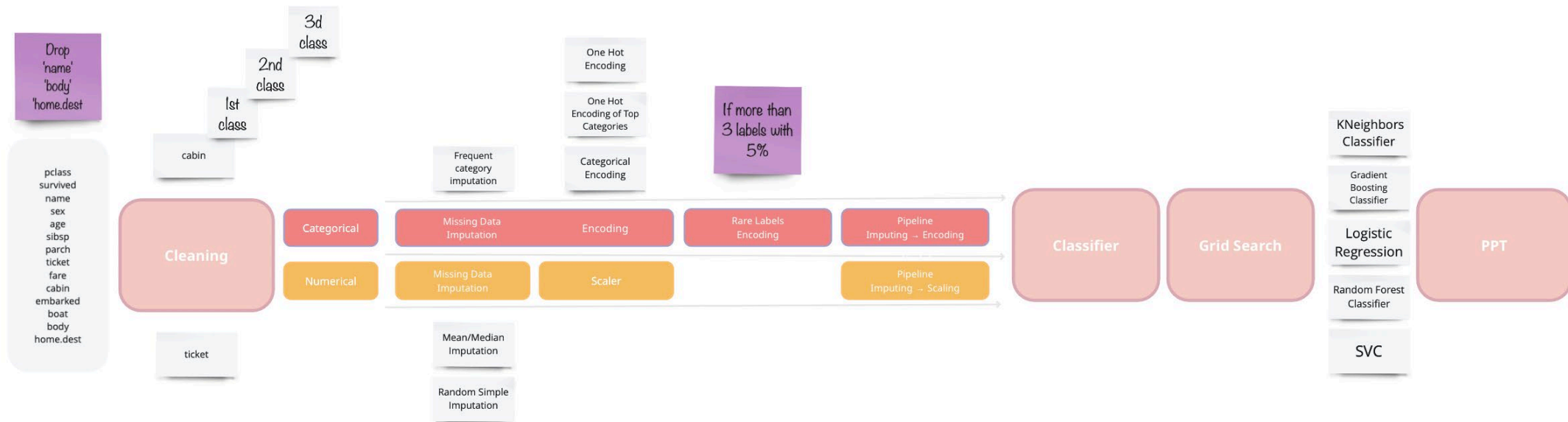


TITANIC

MIRO



Variable	Definition
survival	Survival
pclass	Ticket class
sex	Sex
Age	Age in years
sibsp	# of siblings / spouses aboard the Titanic
parch	# of parents / children aboard the Titanic
ticket	Ticket number
fare	Passenger fare
cabin	Cabin number
embarked	Port of Embarkation



JIRA

Проекты / Omega Korea

Доска OK

🔍 W B P Элик

К ВЫПОЛНЕНИЮ

+ Создать задачу

В РАБОТЕ 1 ЗАДАЧА

ppt preparation

OK-26

ГОТОВО 15 ЗАДАЧ

data interpretation

OK-27

var cabin best imputation method search

OK-25

Group Numerical Variables

IMPUTATION

OK-4

clean cabin column & drop name column

OK-11

Suitable imputer for age and fare

OK-17

data analysis in graphics and diagrams

OK-29

train model

OK-28

Find Best Imputation method

IMPUTATION

OK-6

Build Pipeline

OK-18

StandardScaler

OK-15

search for suitable imputation method (categorical var)

OK-23

data transformation

OK-24

Group Categorical Variables

IMPUTATION

OK-5

clean ticket column

OK-10

missing data analysis

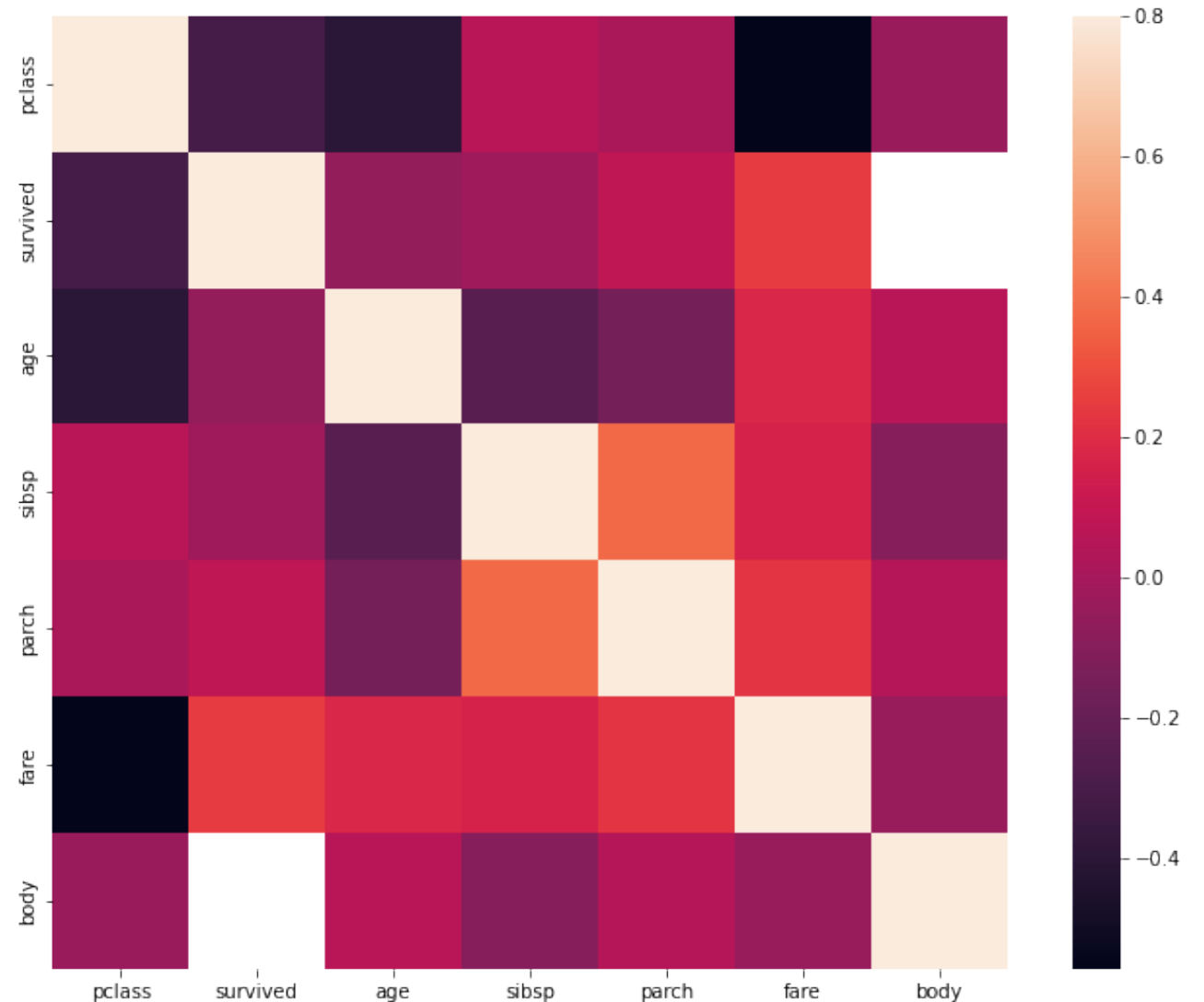
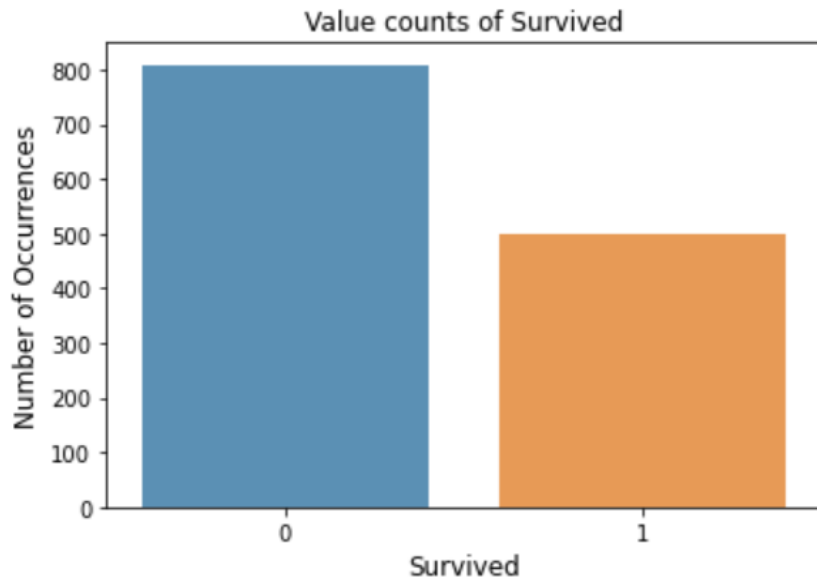
OK-22

PROCESS



CURRENT DATA

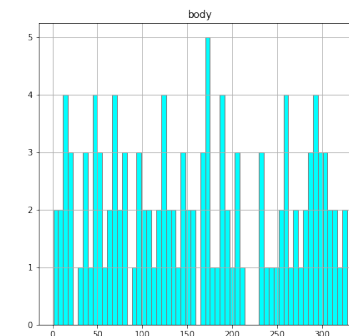
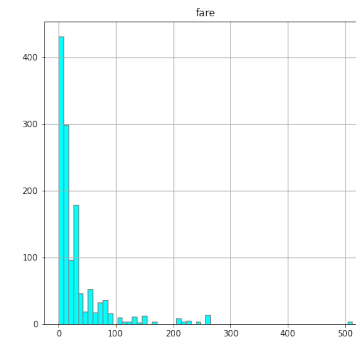
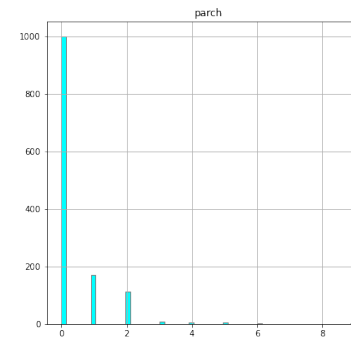
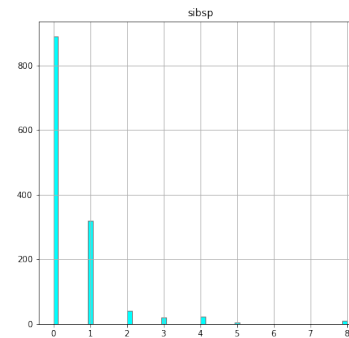
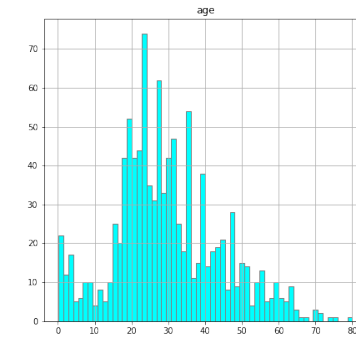
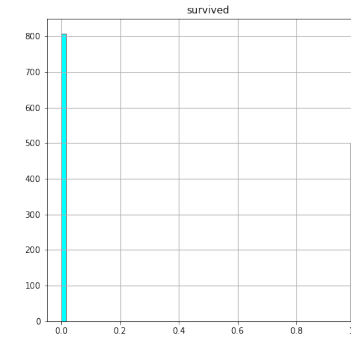
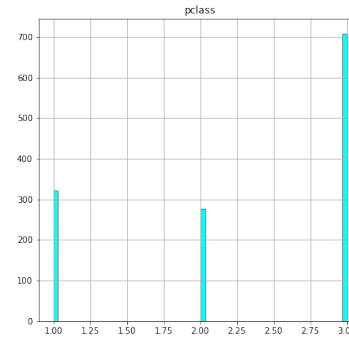
1. pclass
2. name
3. sex
4. age
5. sibsp
6. parch
7. ticket
8. fare
9. cabin:
10. embarked
11. boat
12. body
13. home.dest
14. survival



NUMERICAL DISTRIBUTION

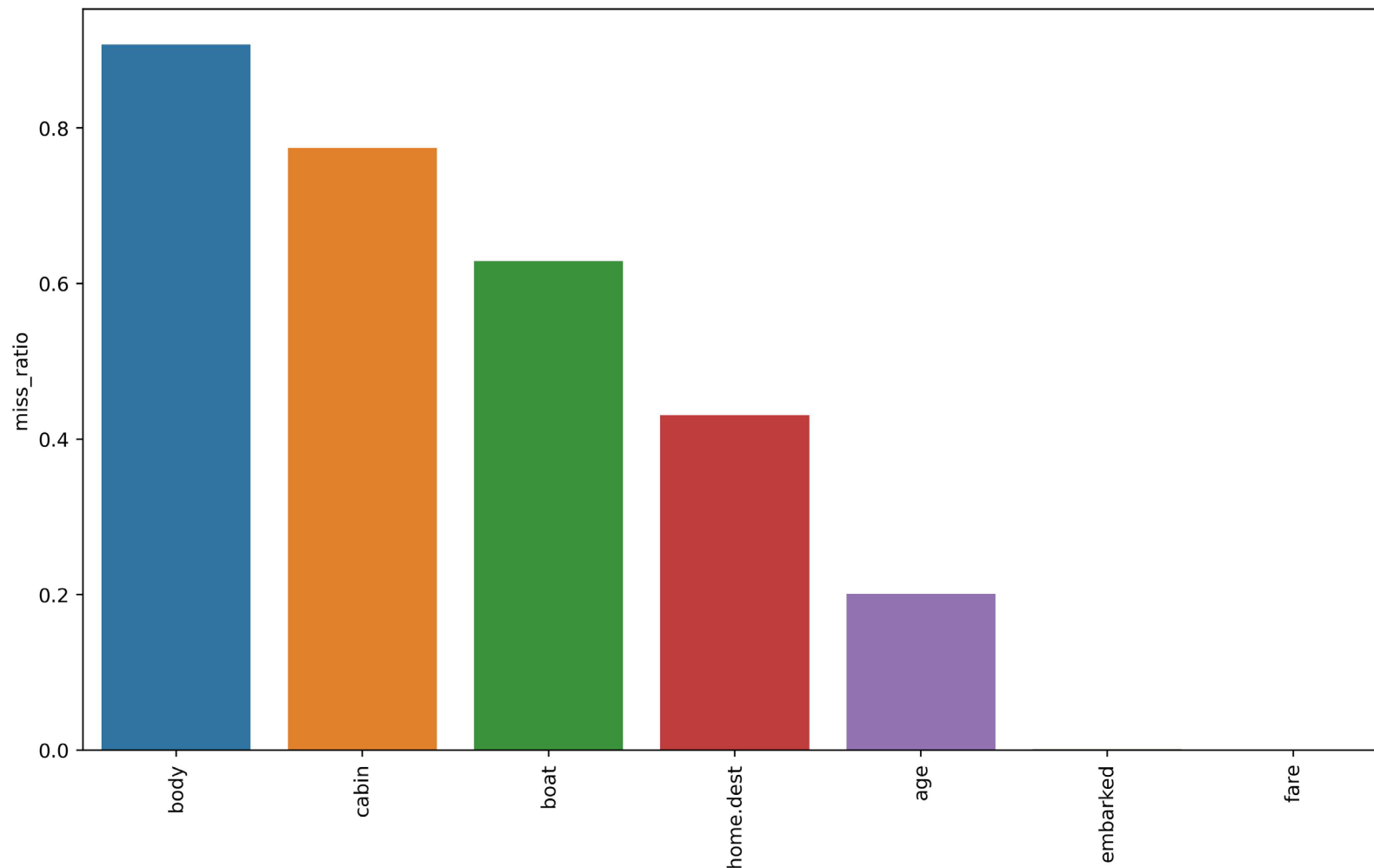
pclass
survived

age
sibsp
parch
fare
body



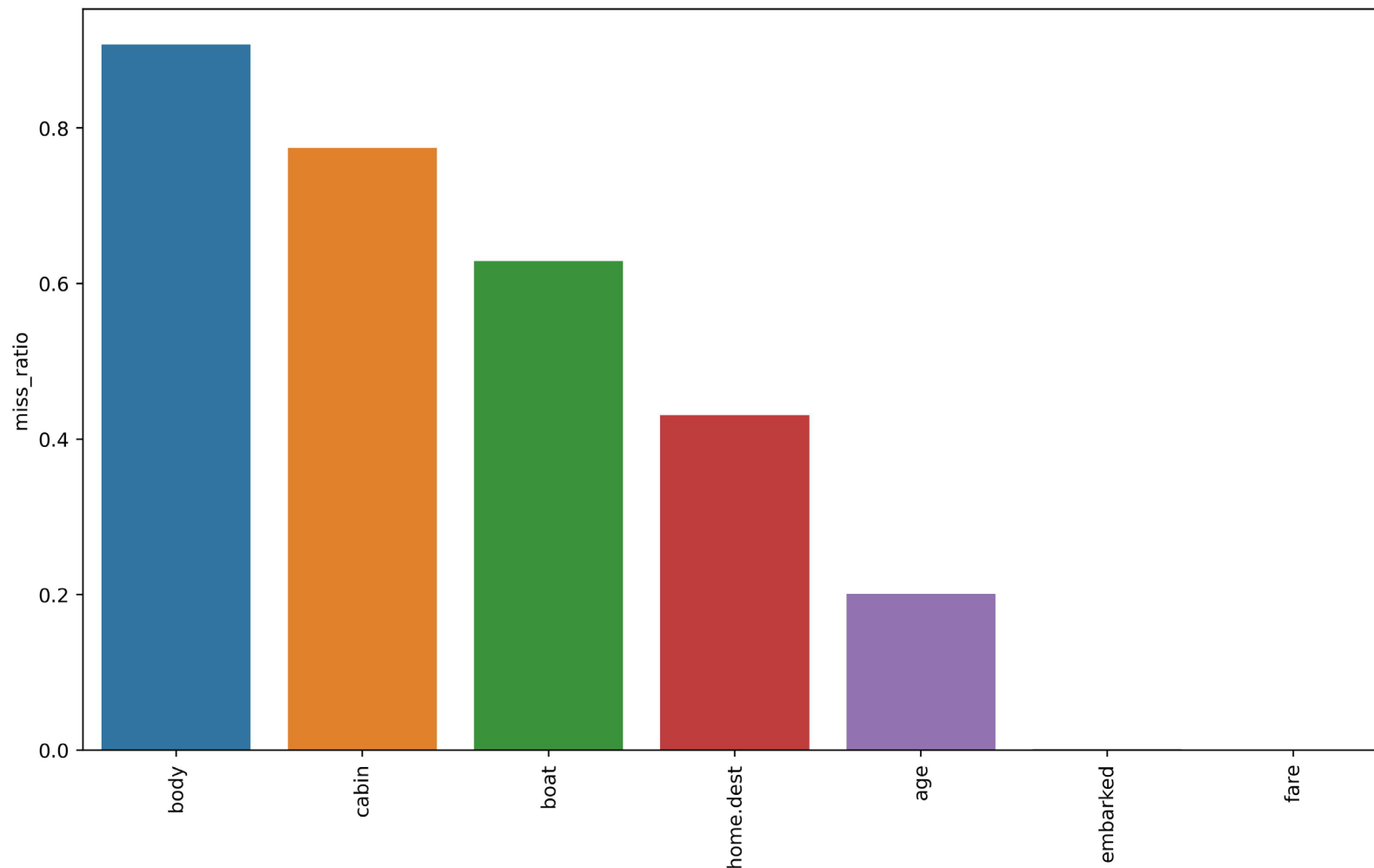
MISSING DATA

pclass	0%
survived	0%
name	0%
sex	0%
age	20%
sibsp	0%
parch	0%
ticket	0%
fare	0.007%
cabin	77%
embarked	0.15%
boat	63%
body	90%
home.dest	43%



MISSING DATA

pclass	0%
survived	0%
name	0%
sex	0%
age	20%
sibsp	0%
parch	0%
ticket	0%
fare	0.007%
cabin	77%
embarked	0.15%
boat	63%
body	90%
home.dest	43%



MIXED VARIABLES



TICKET



TICKET



TICKET



TICKET



TICKET

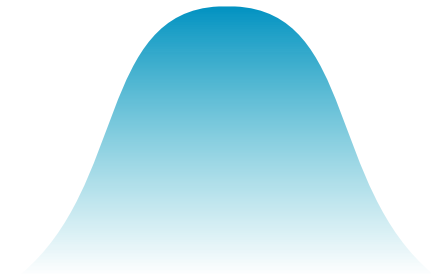


TICKET



Ticket_numerical

97

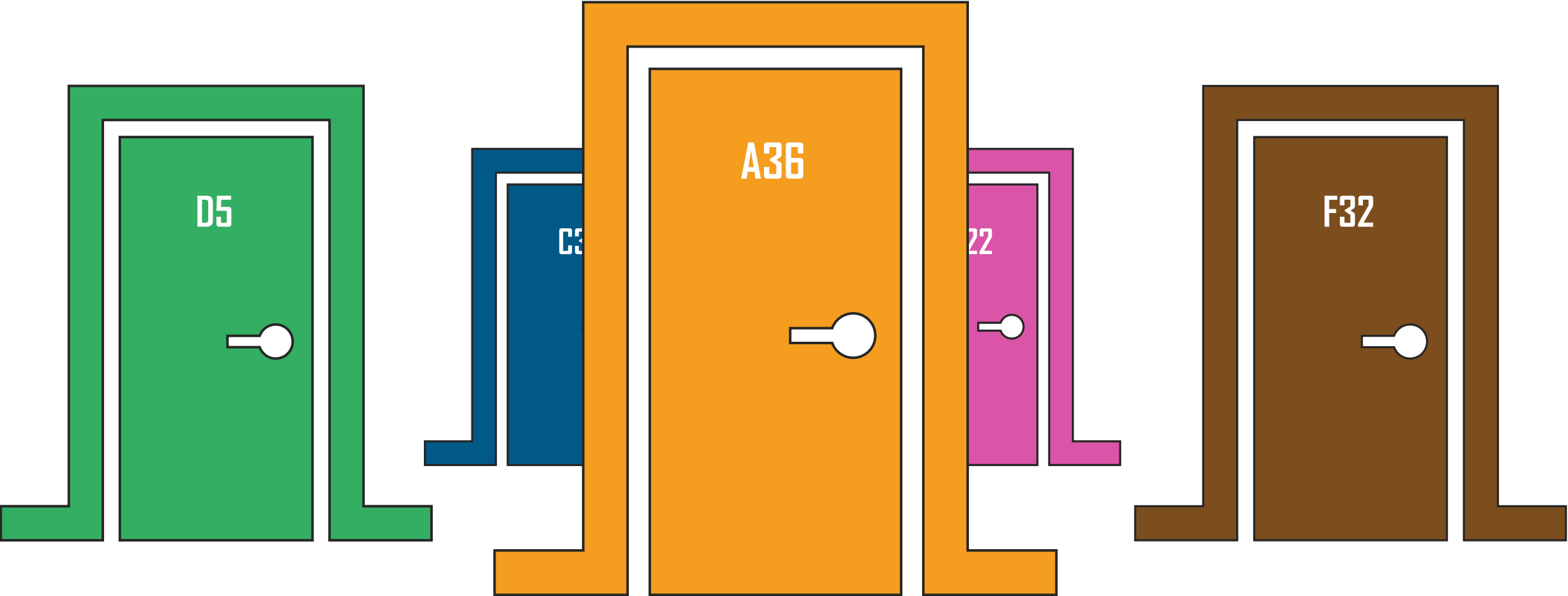


Ticket_categorical

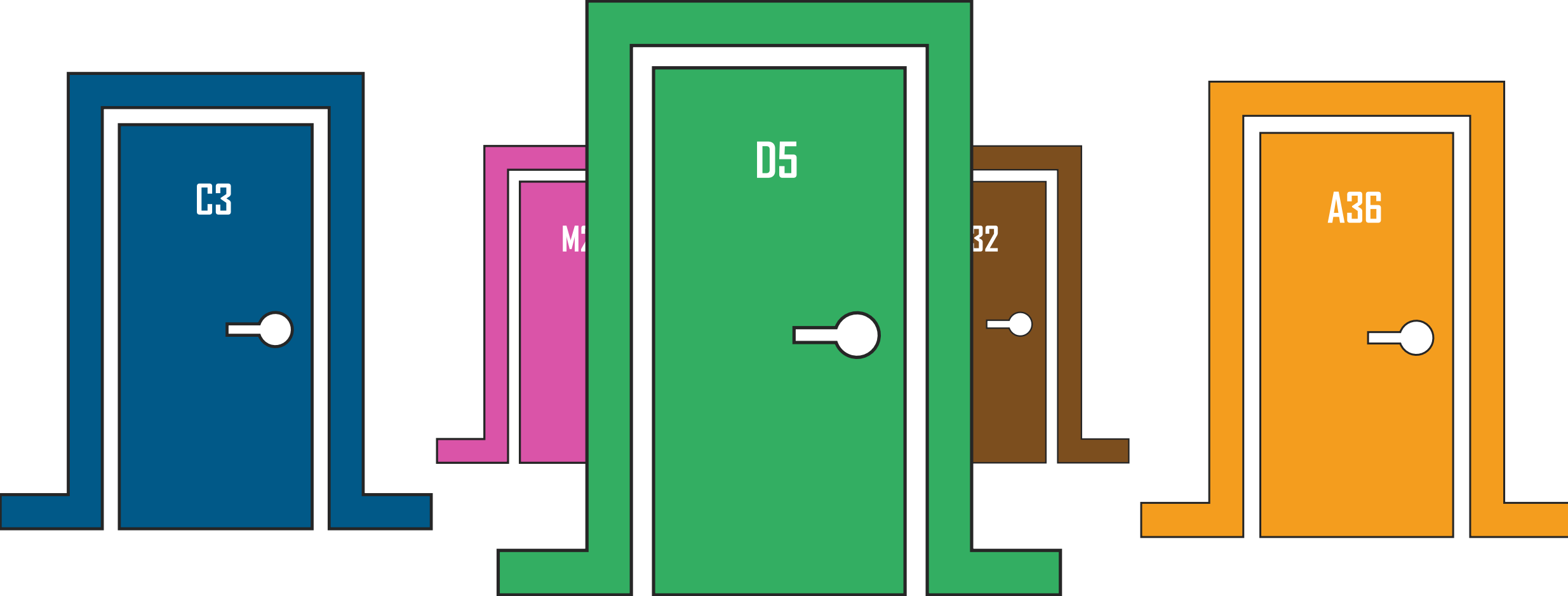
27



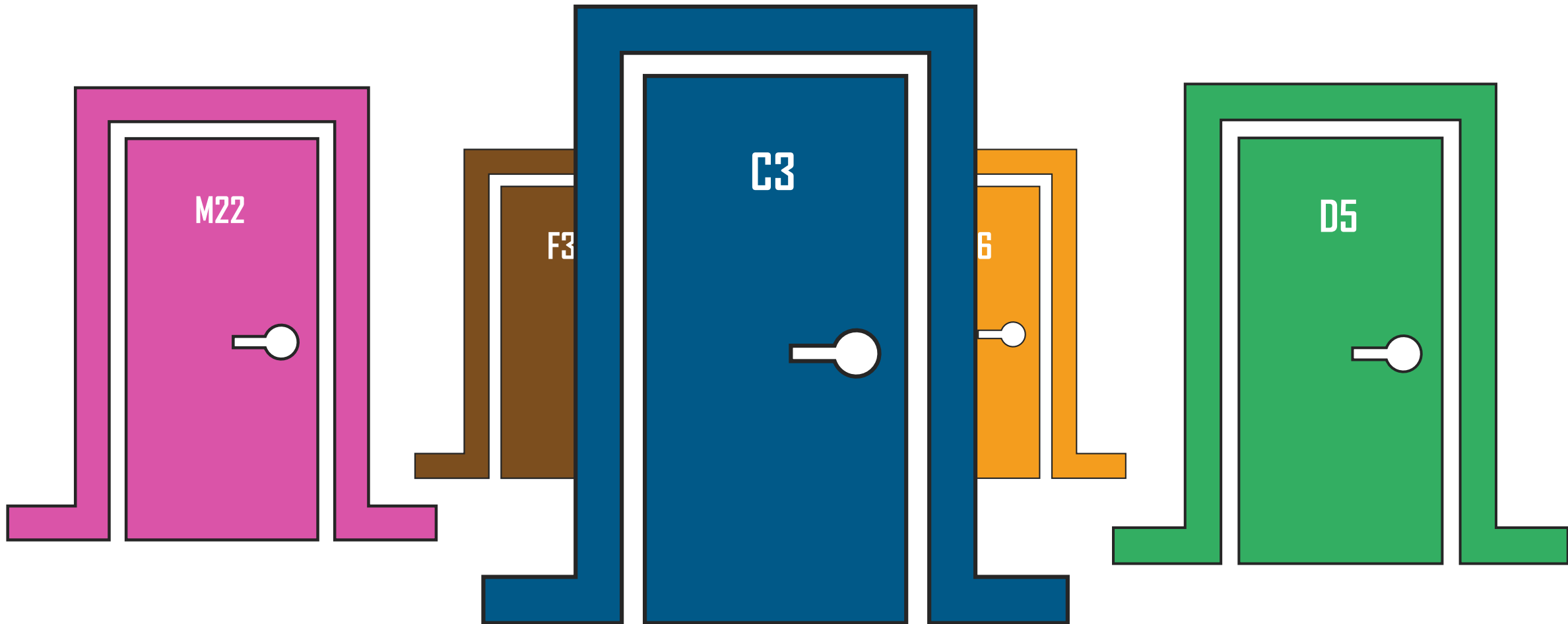
CABIN



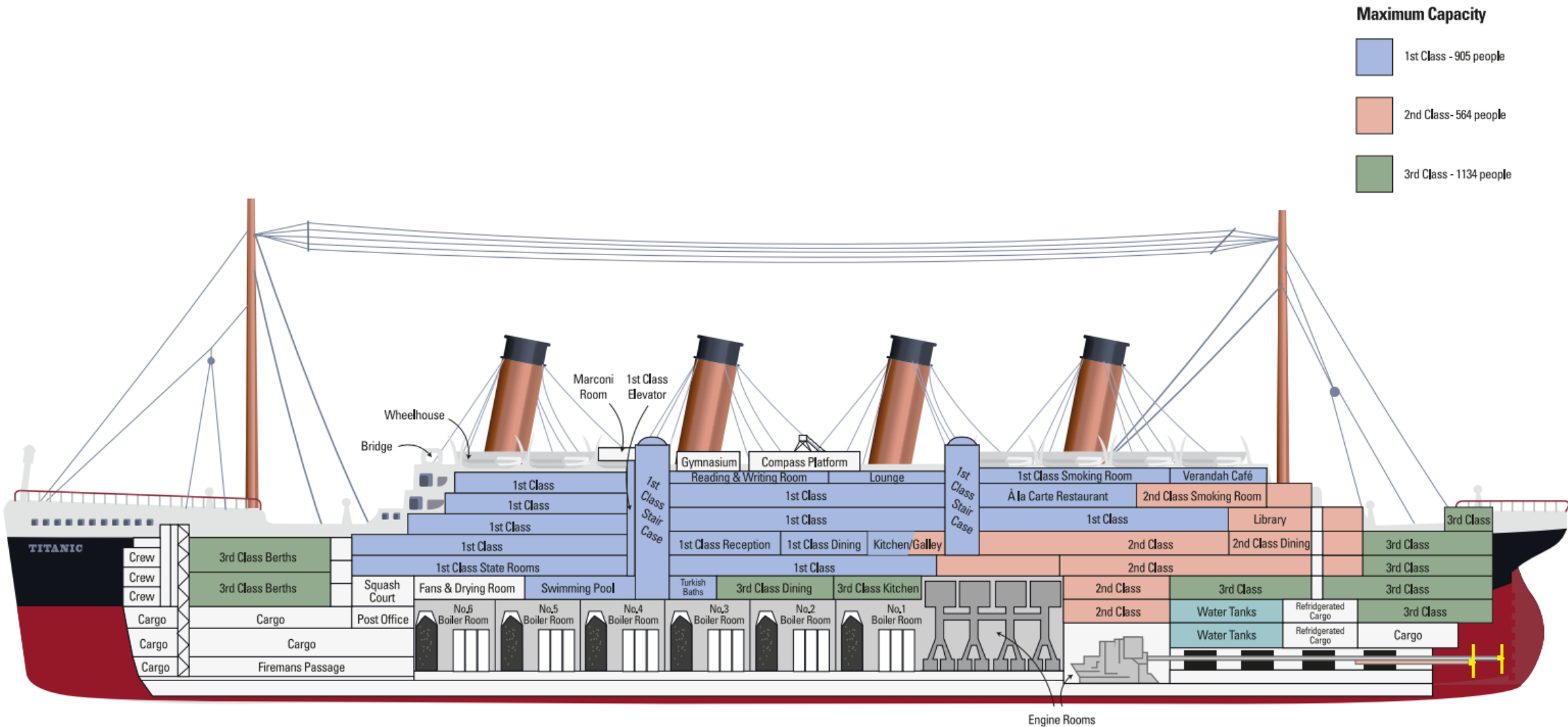
CABIN



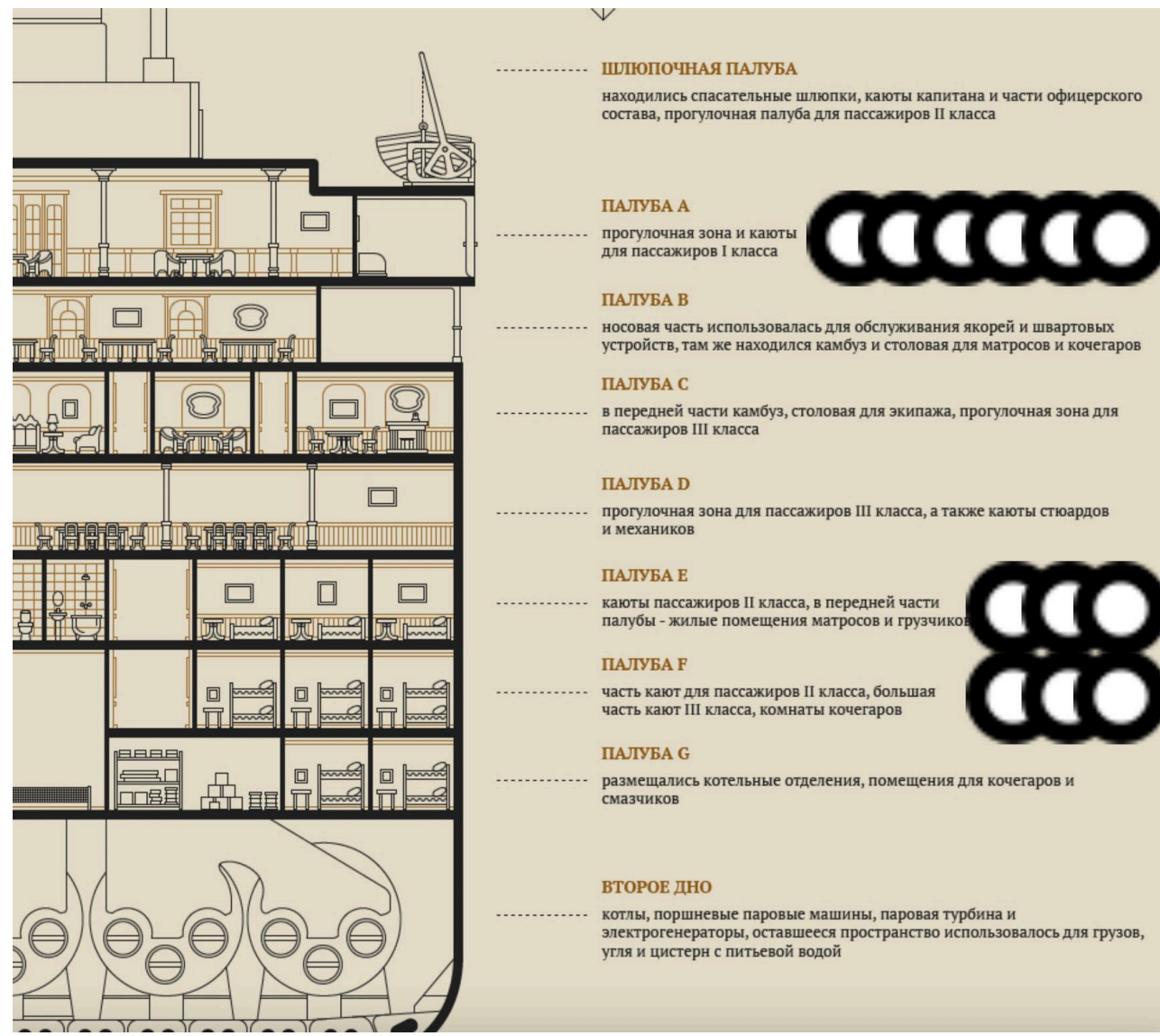
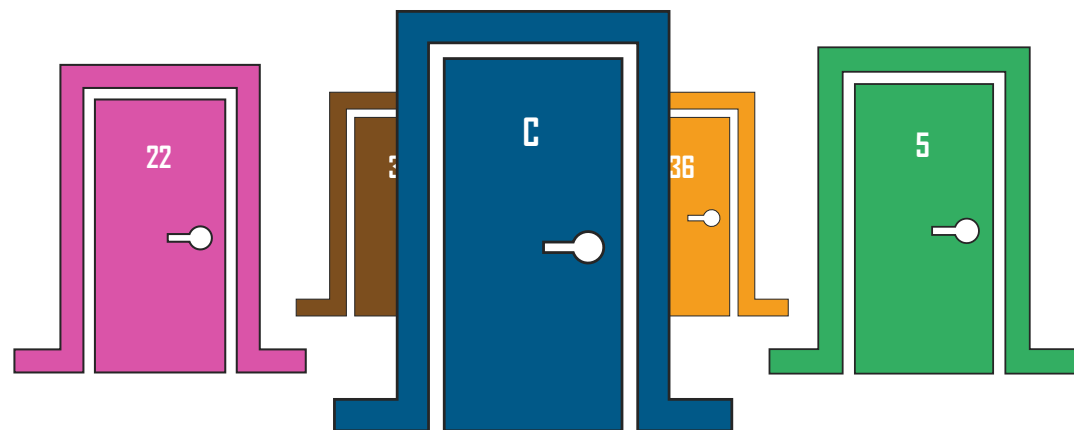
CABIN



CABIN



CABIN



ШЛЮПОЧНАЯ ПАЛУБА

находились спасательные шлюпки, каюты капитана и части офицерского состава, прогулочная палуба для пассажиров II класса

ПАЛУБА А

прогулочная зона и каюты для пассажиров I класса

ПАЛУБА В

носовая часть использовалась для обслуживания якорей и швартовых устройств, там же находился камбуз и столовая для матросов и кочегаров

ПАЛУБА С

в передней части камбуз, столовая для экипажа, прогулочная зона для пассажиров III класса

ПАЛУБА D

прогулочная зона для пассажиров III класса, а также каюты стюардов и механиков

ПАЛУБА Е

каюты пассажиров II класса, в передней части палубы - жилые помещения матросов и грузчиков

ПАЛУБА F

часть кают для пассажиров II класса, большая часть кают III класса, комнаты кочегаров

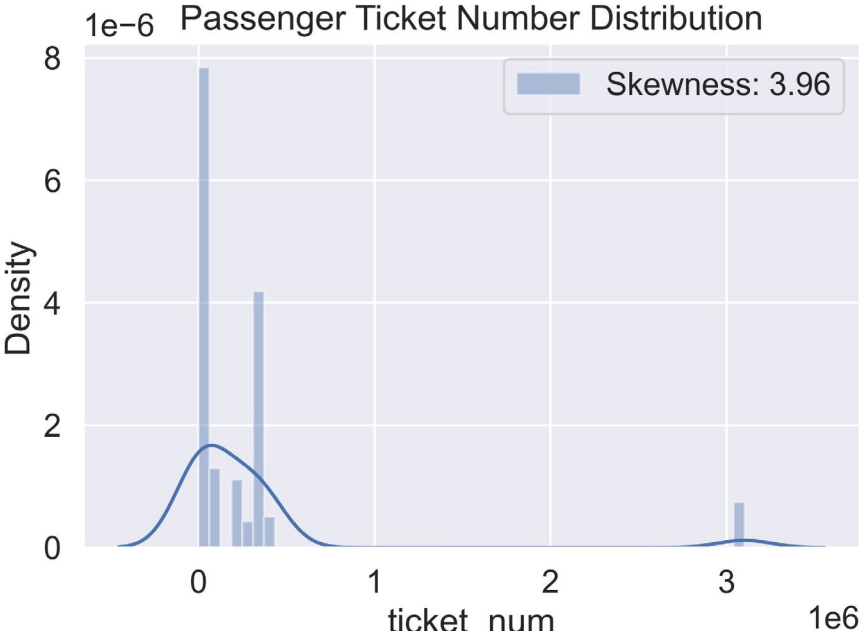
ПАЛУБА G

размещались котельные отделения, помещения для кочегаров и смазчиков

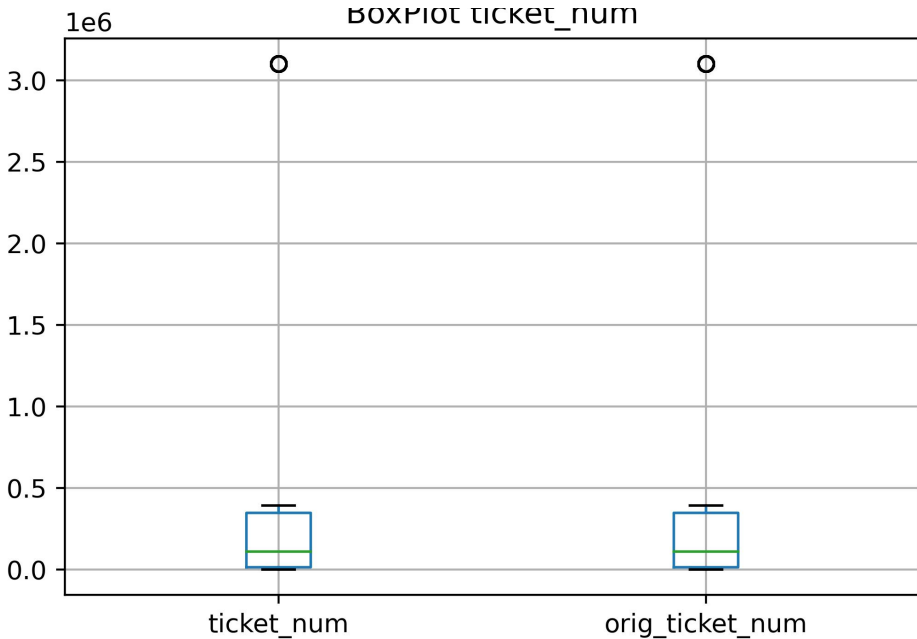
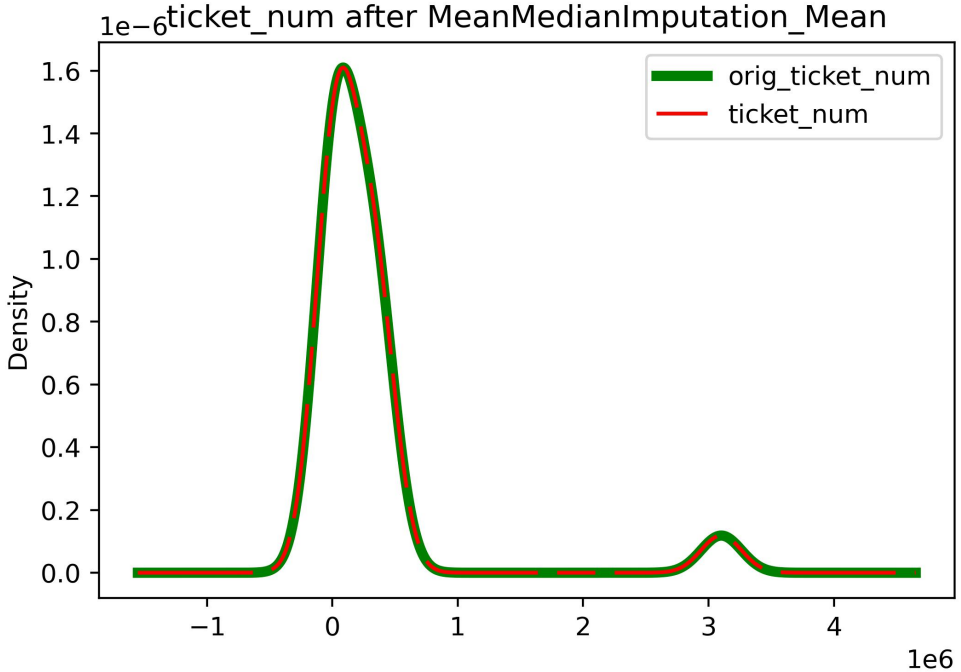
ВТОРОЕ ДНО

котлы, поршневые паровые машины, паровая турбина и электрогенераторы, оставшееся пространство использовалось для грузов, угля и цистерн с питьевой водой

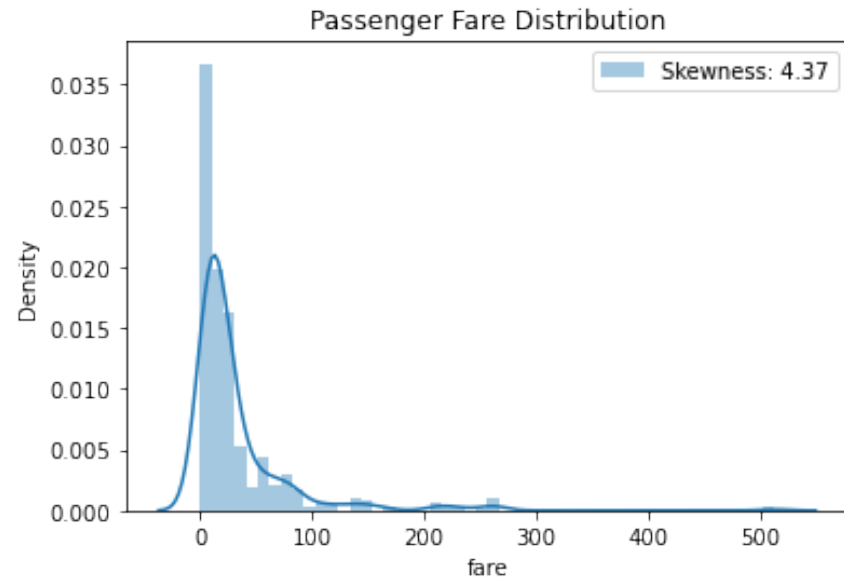
ticket_num



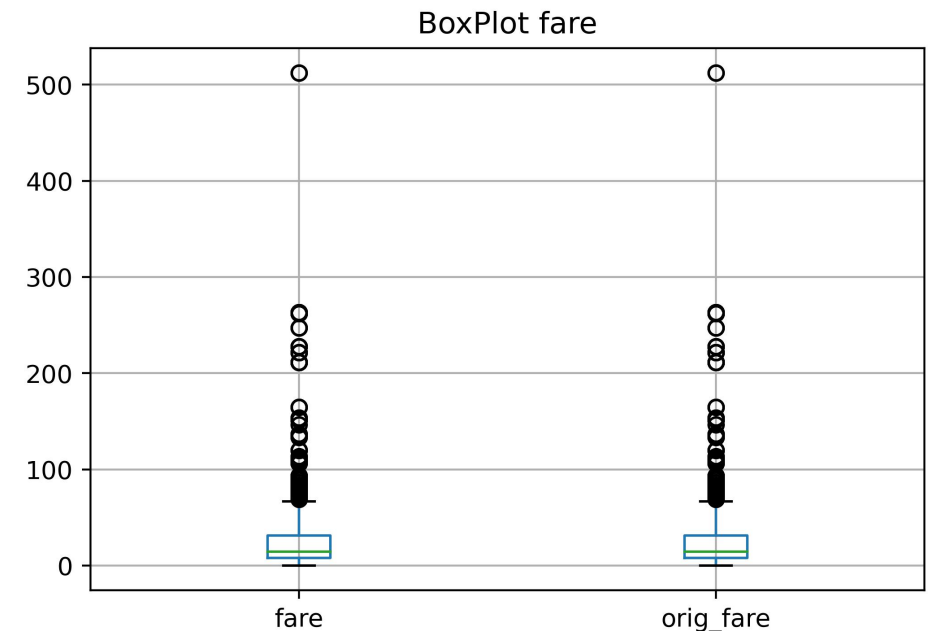
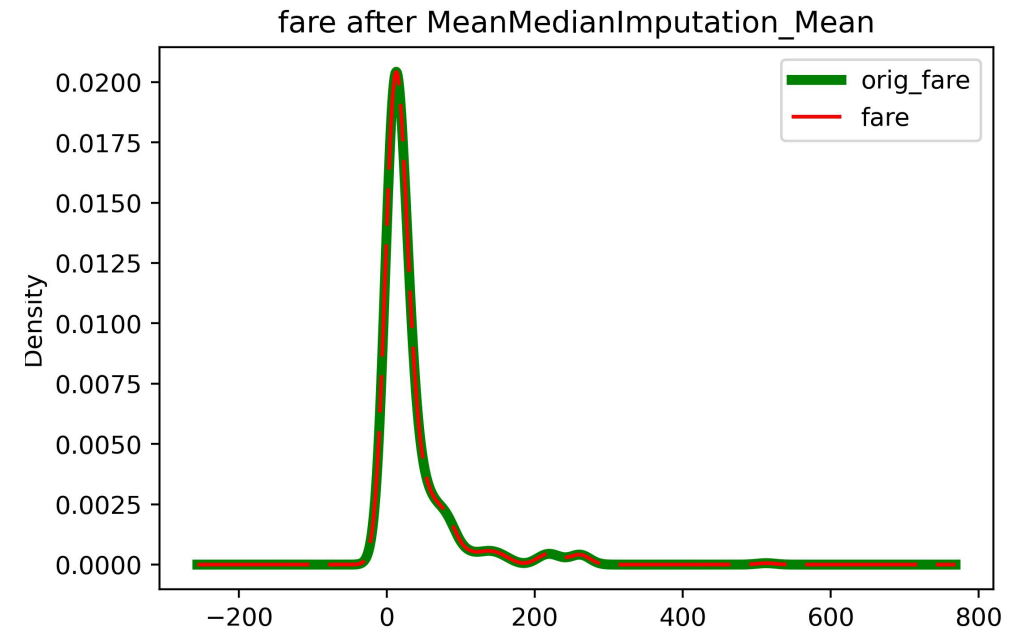
0.002865



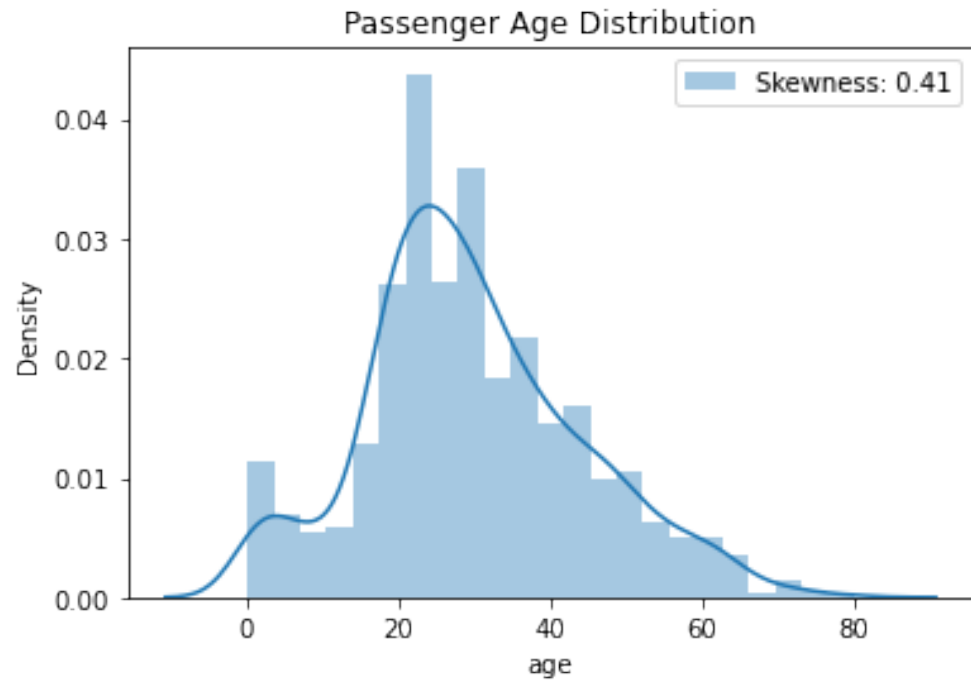
fare



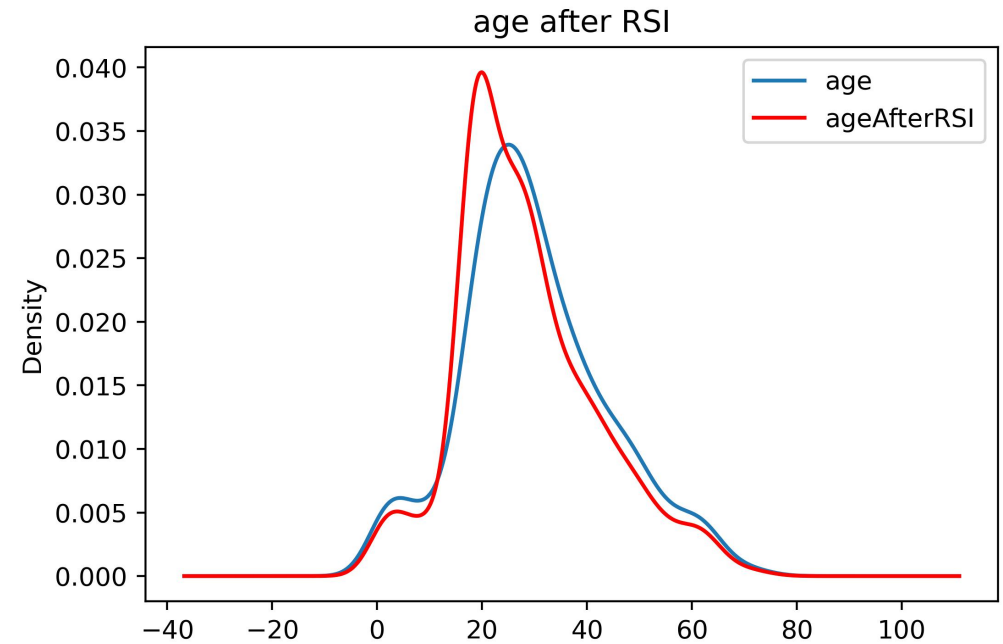
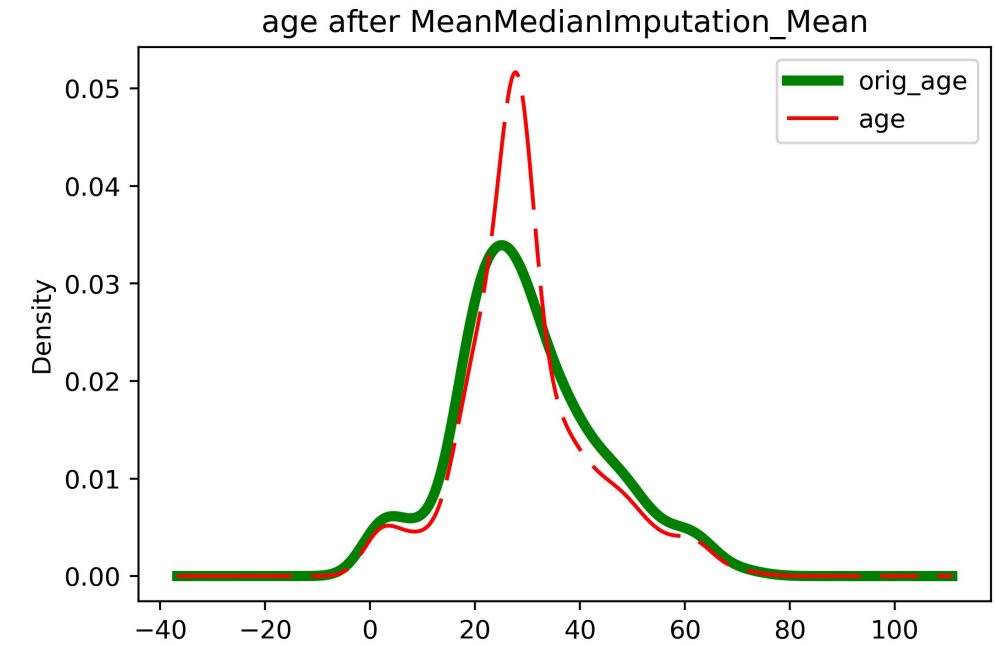
0.000764



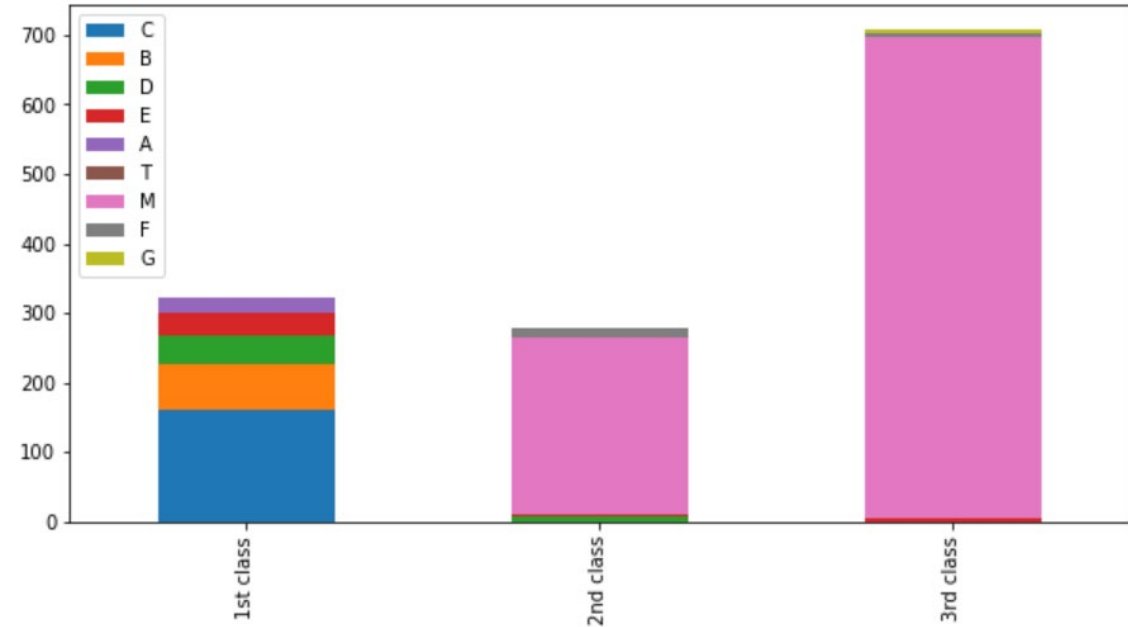
age



0.200917

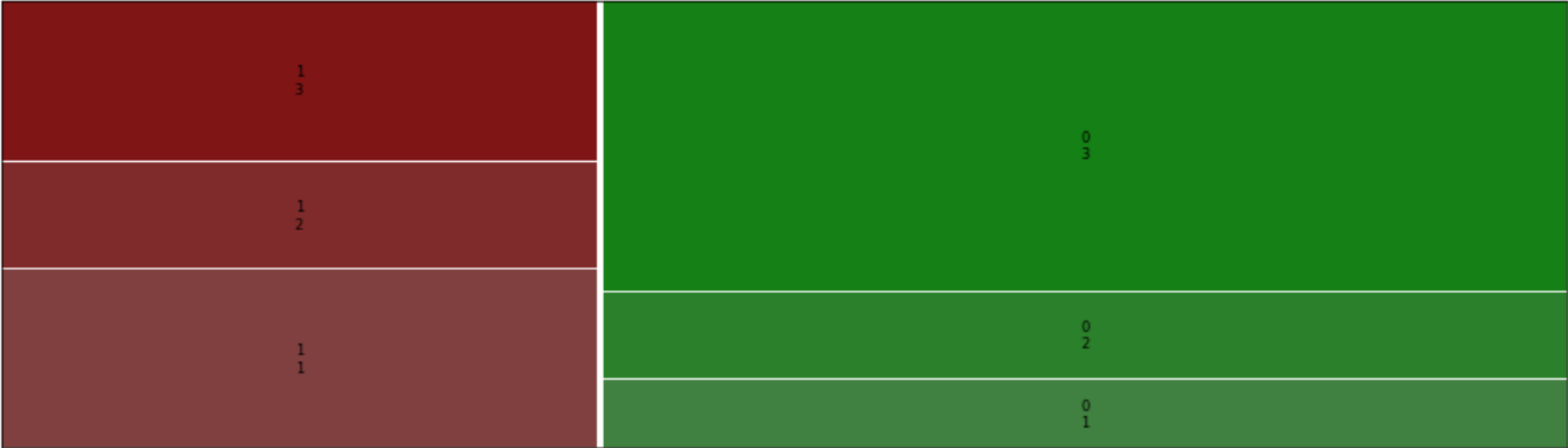


CABIN

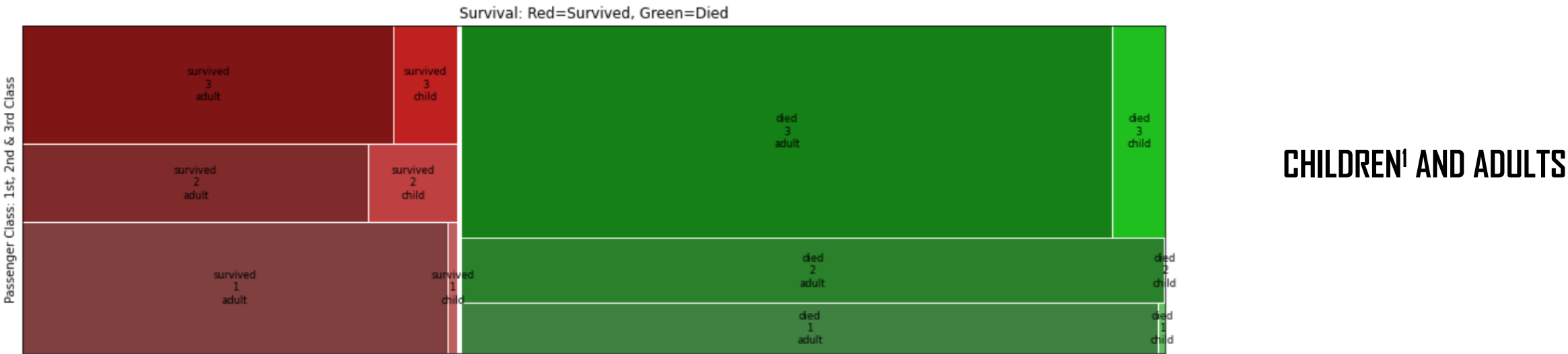


Passenger Class: 1st, 2nd & 3rd Class

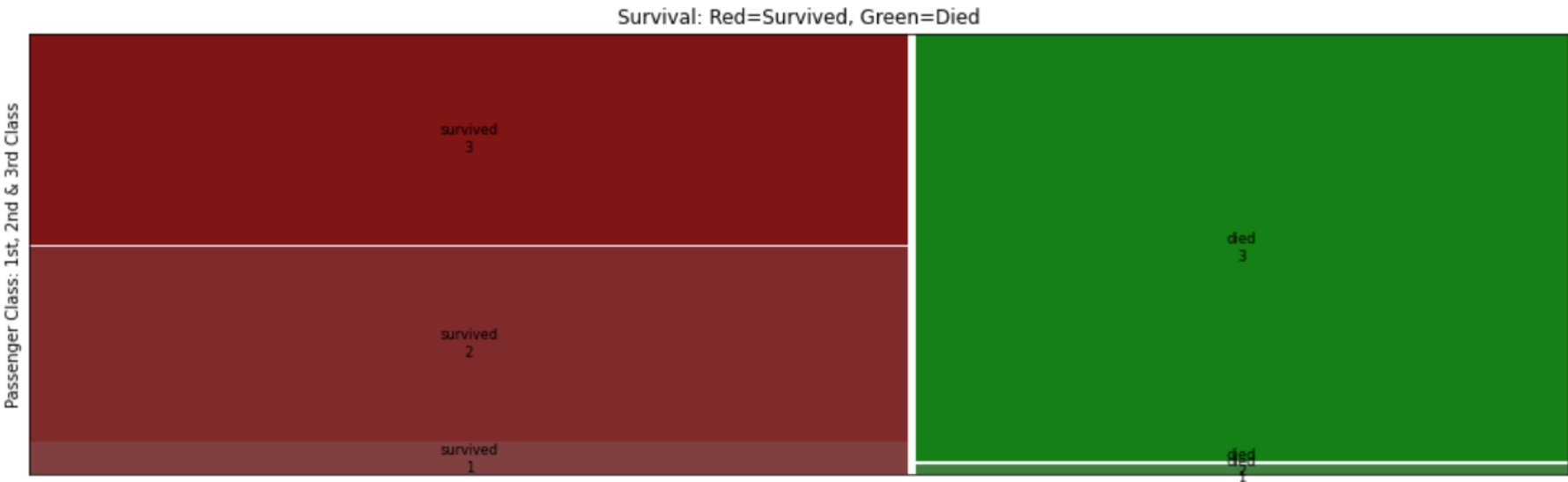
Survival: Red=Survived, Green=Died



LADIES AND CHILDREN FIRST (CHILDREN)

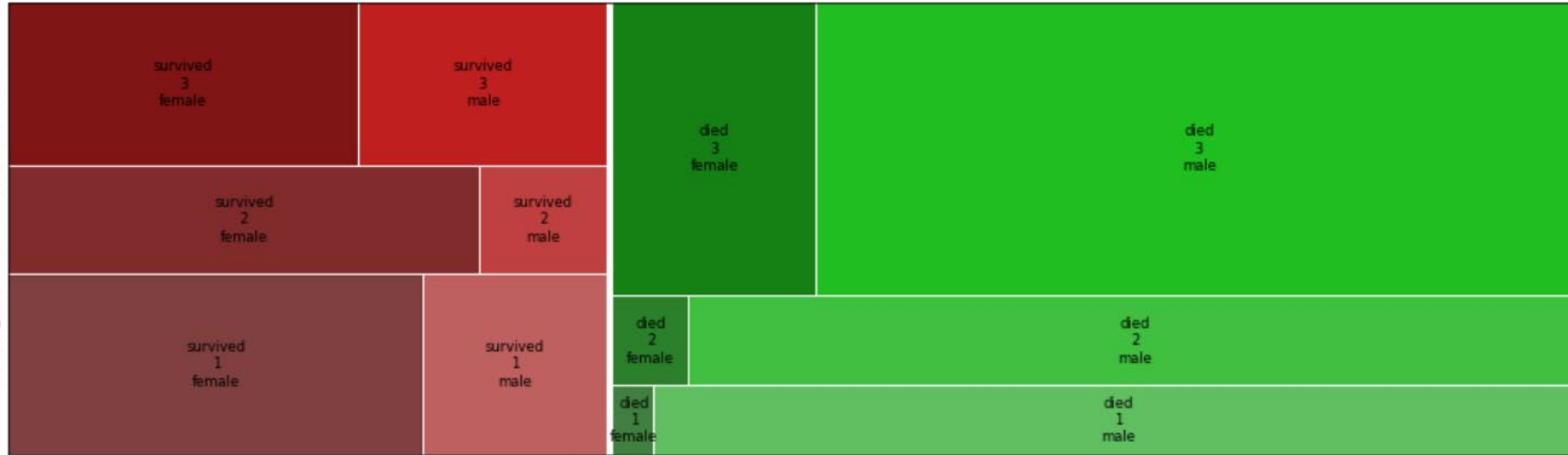


CHILDREN



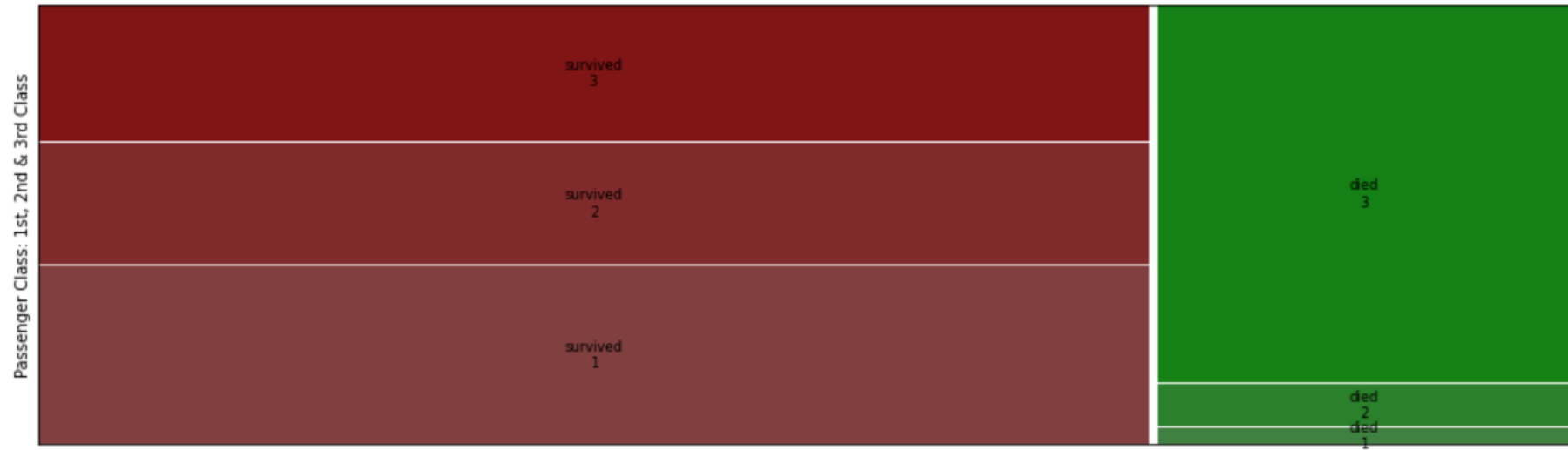
LADIES AND CHILDREN FIRST (LADIES)

Survival: Red=Survived, Green=Died



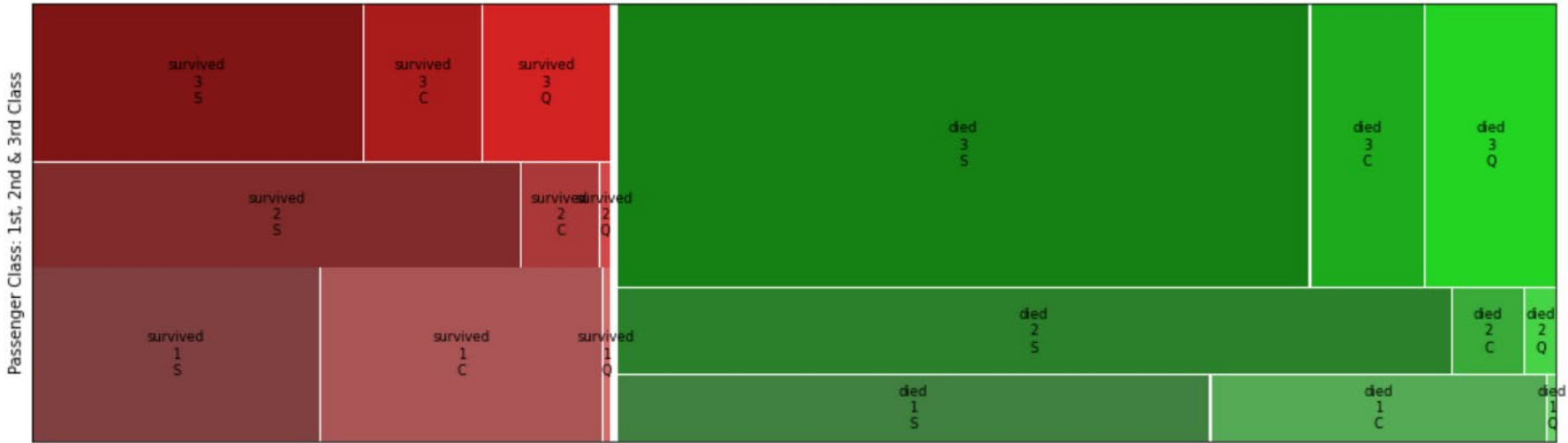
Sex: Female & Male

Survival: Red=Survived, Green=Died



EMBARKED

Survival: Red=Survived, Green=Died



Embarked

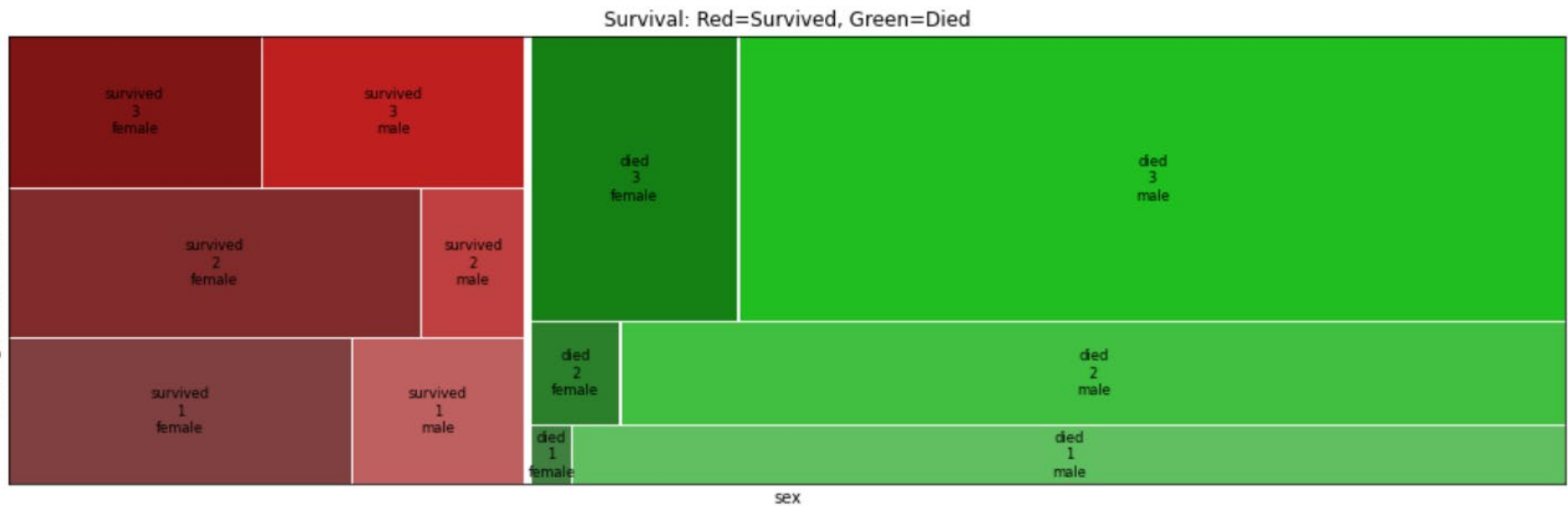
Survival: Red=Survived, Green=Died



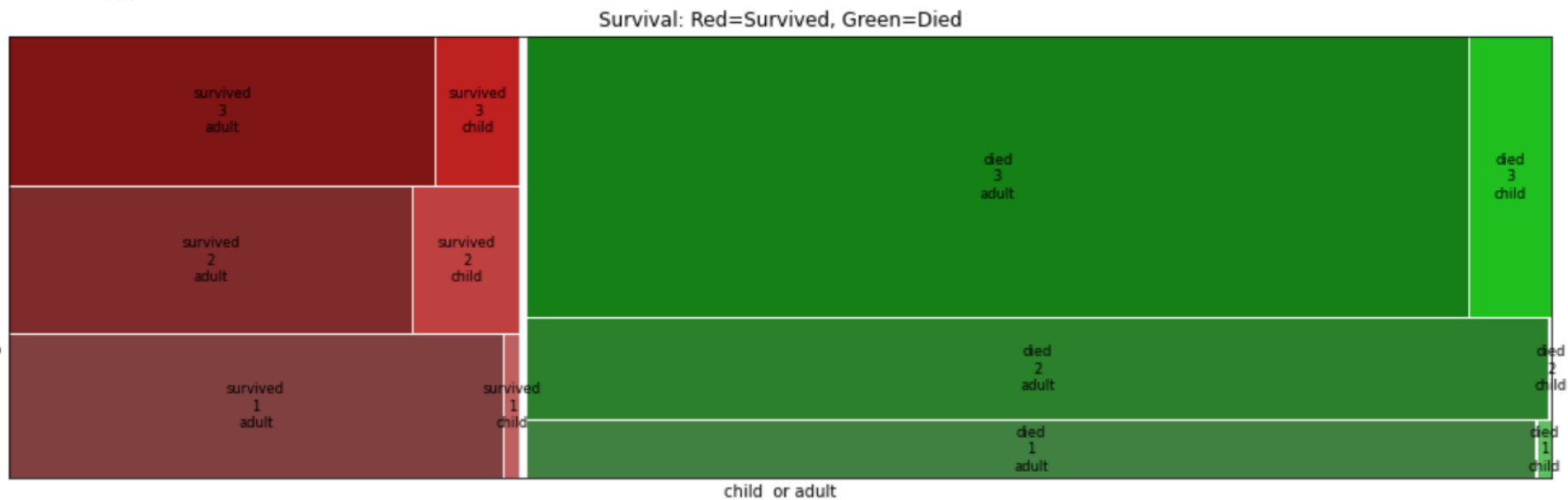
Age: child & non child

EMBARKED S

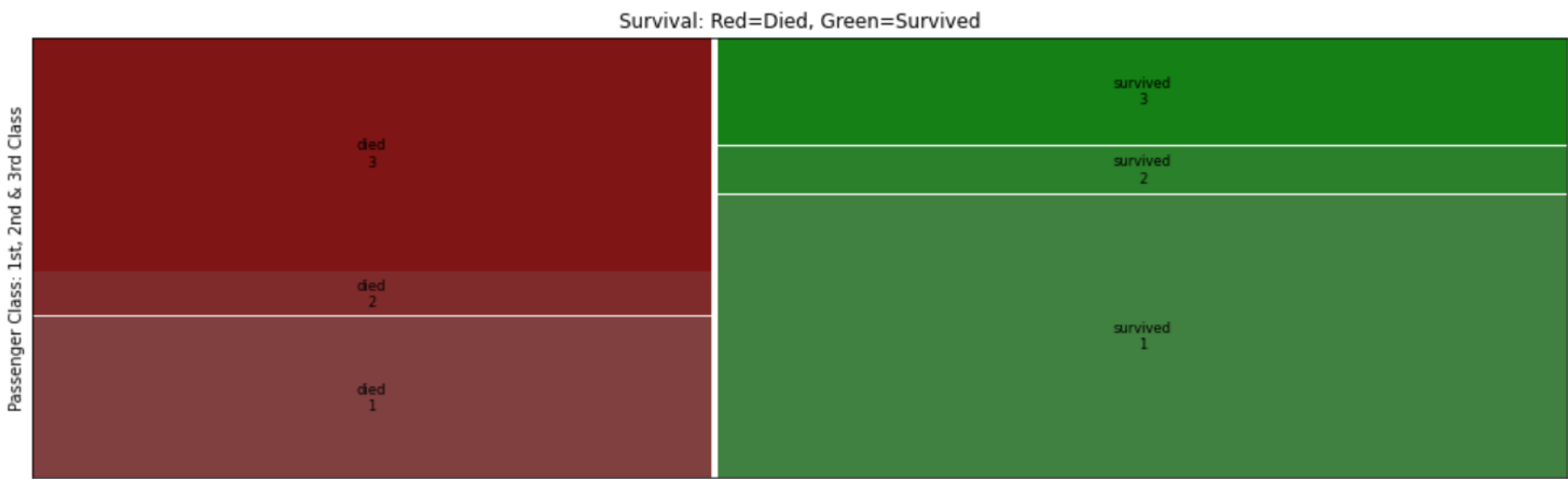
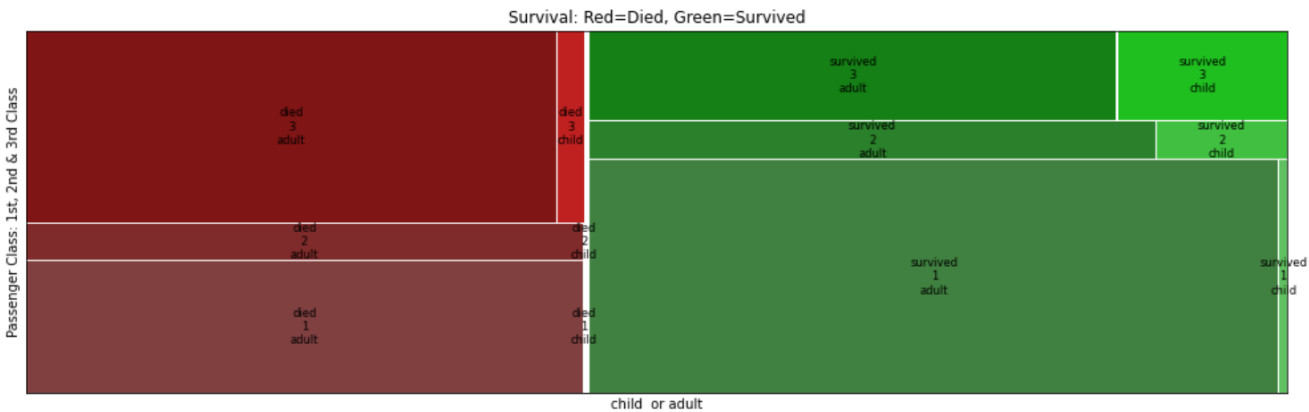
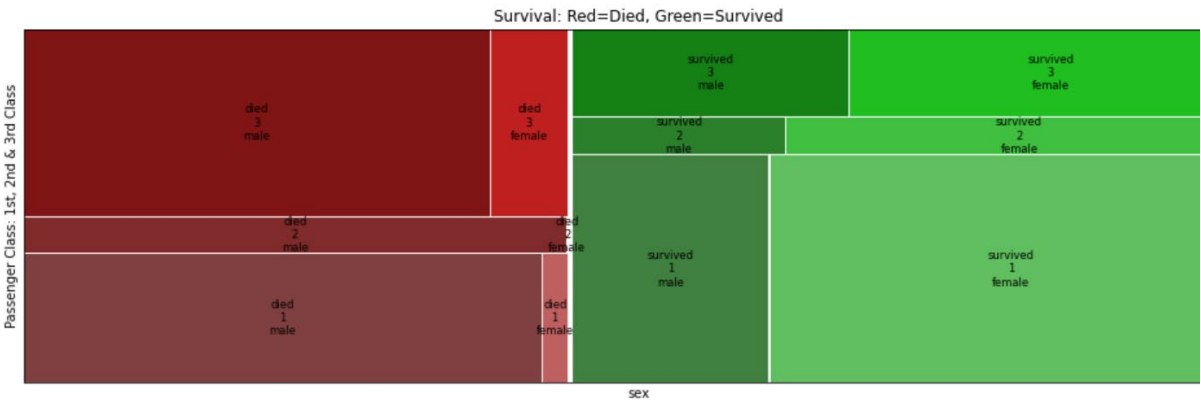
Passenger Class: 1st, 2nd & 3rd Class



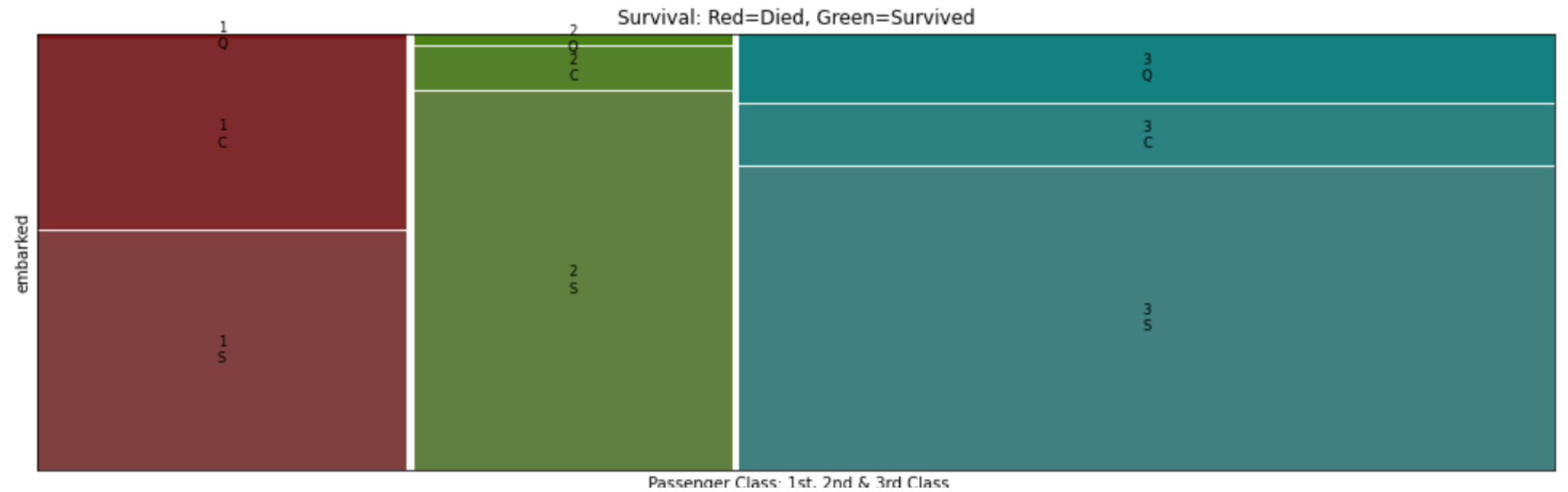
Passenger Class: 1st, 2nd & 3rd Class



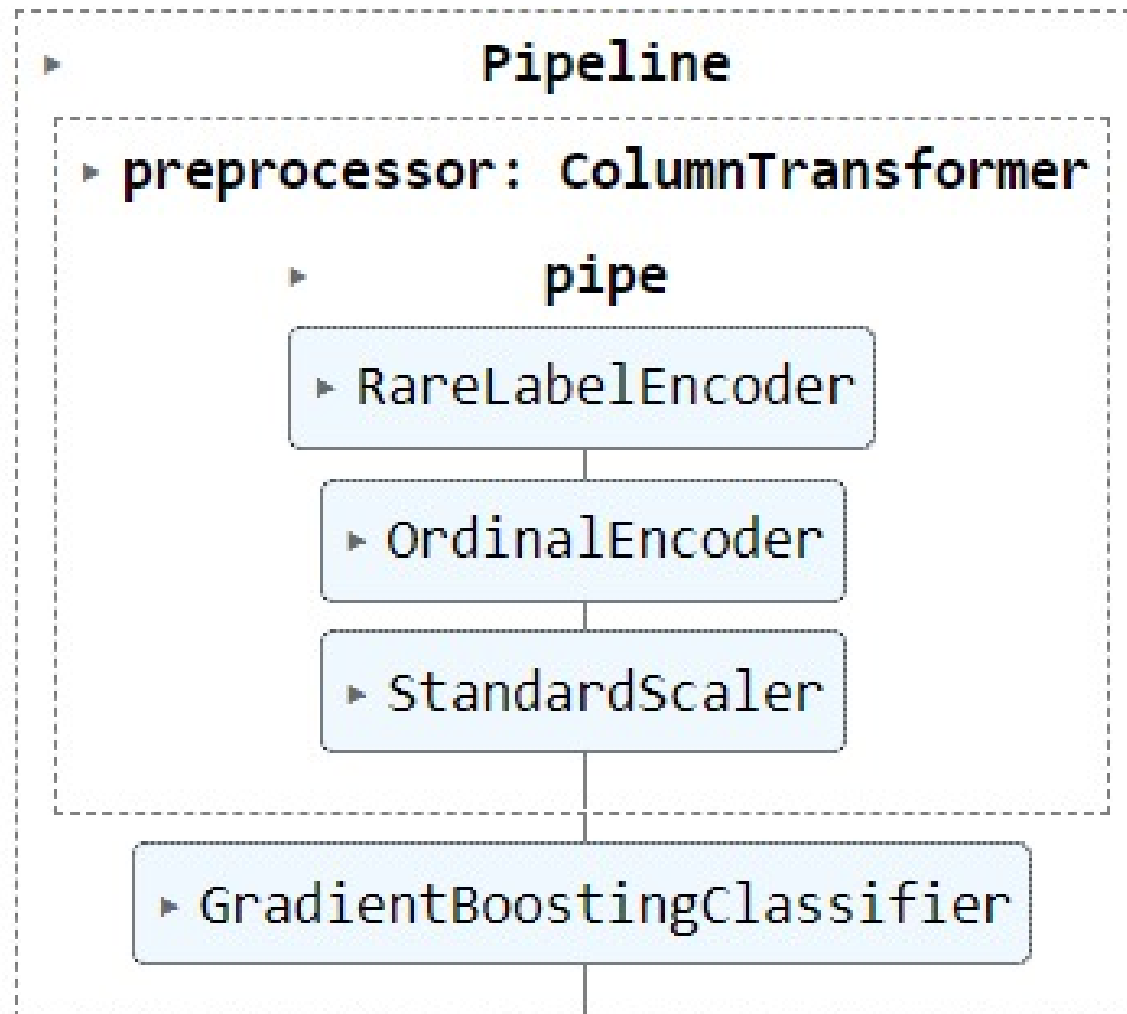
EMBARKED C



EMBARKED



PIPELINE



CLASSIFICATION

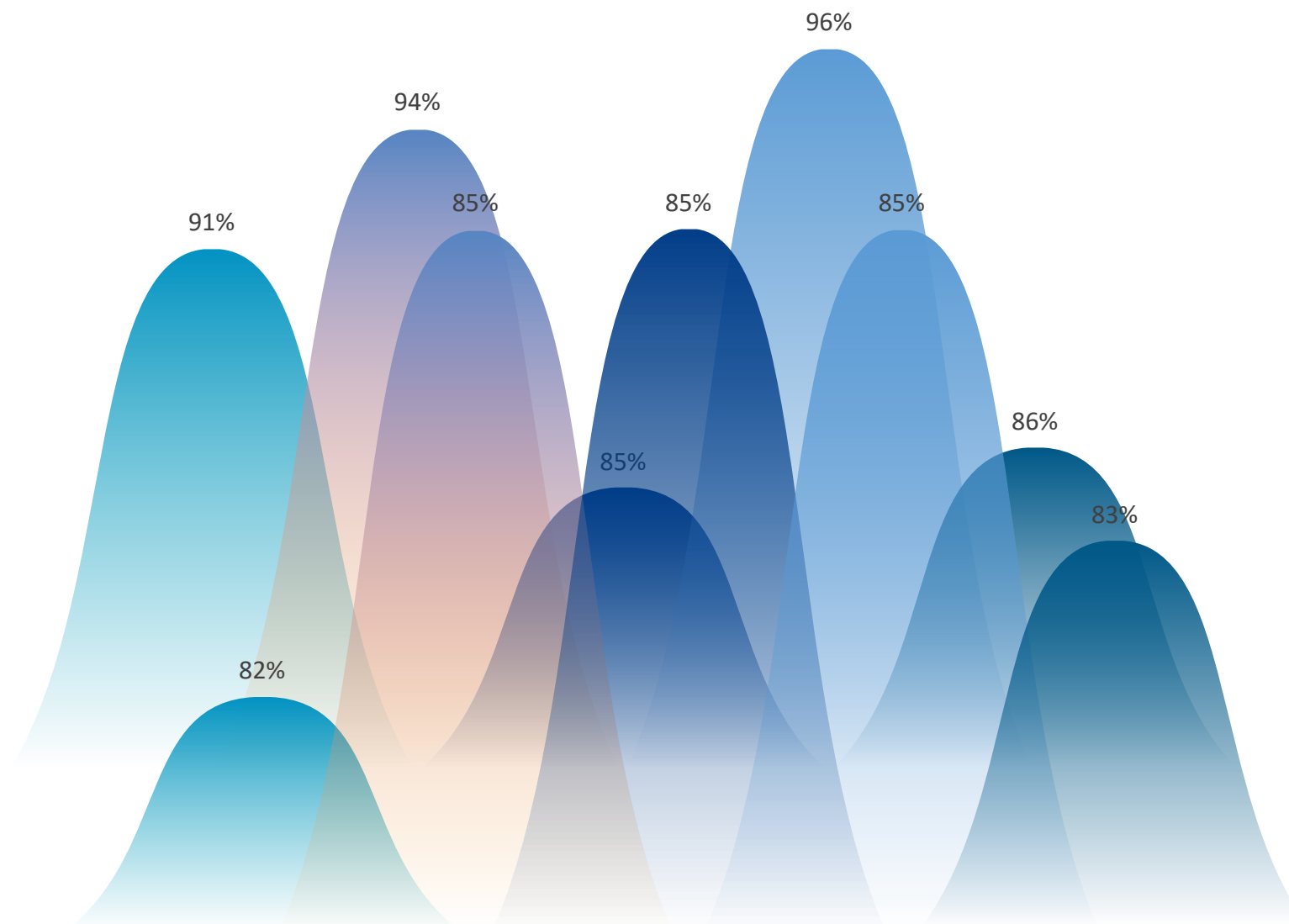
KNN Classifier

Gradient Boosting Classifier

Logistic Regression

Random Forest Classifier

SVC



CLASSIFICATION (TRAIN)

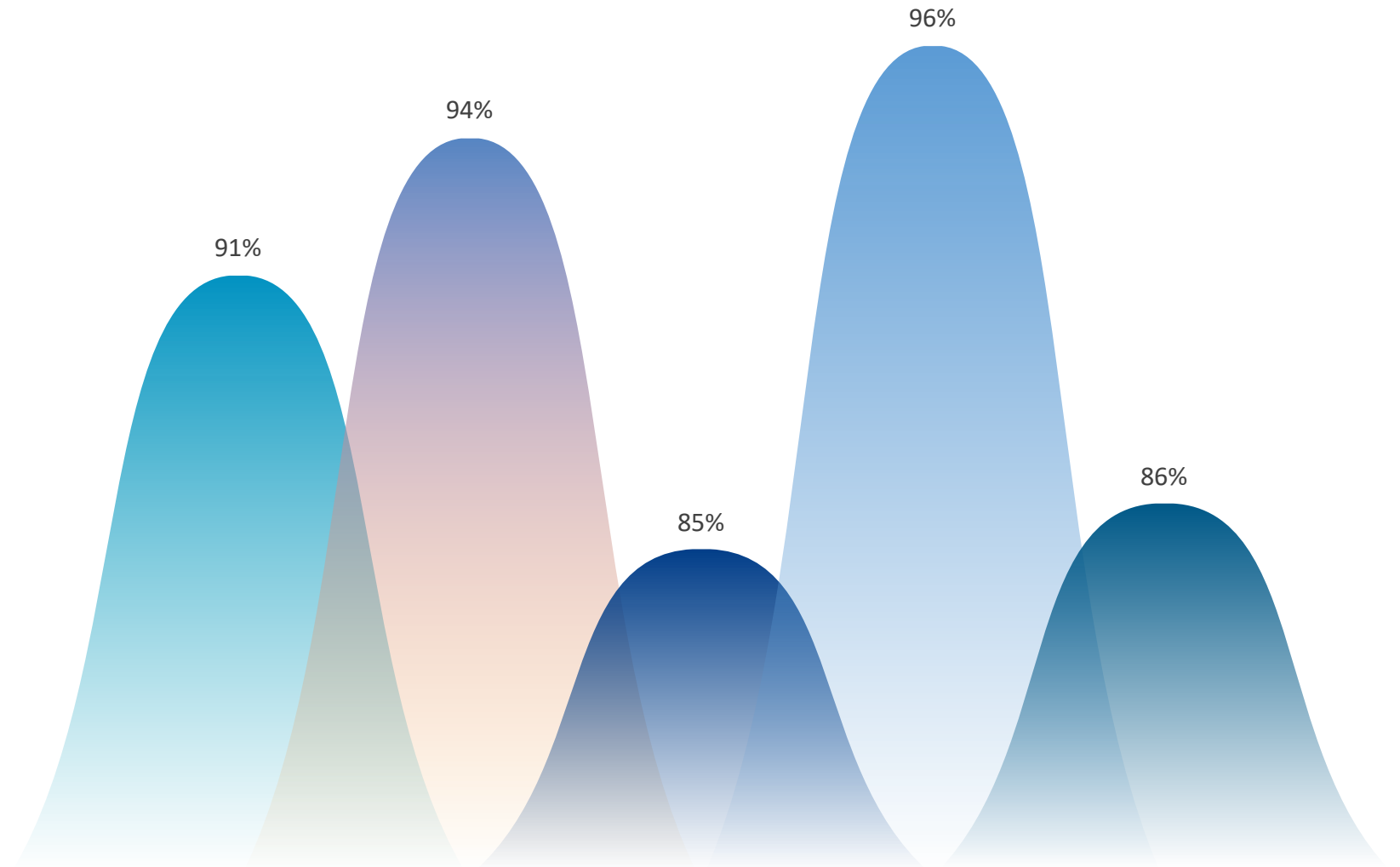
KNN Classifier

Gradient Boosting Classifier

Logistic Regression

Random Forest Classifier

SVC



CLASSIFICATION (TEST)

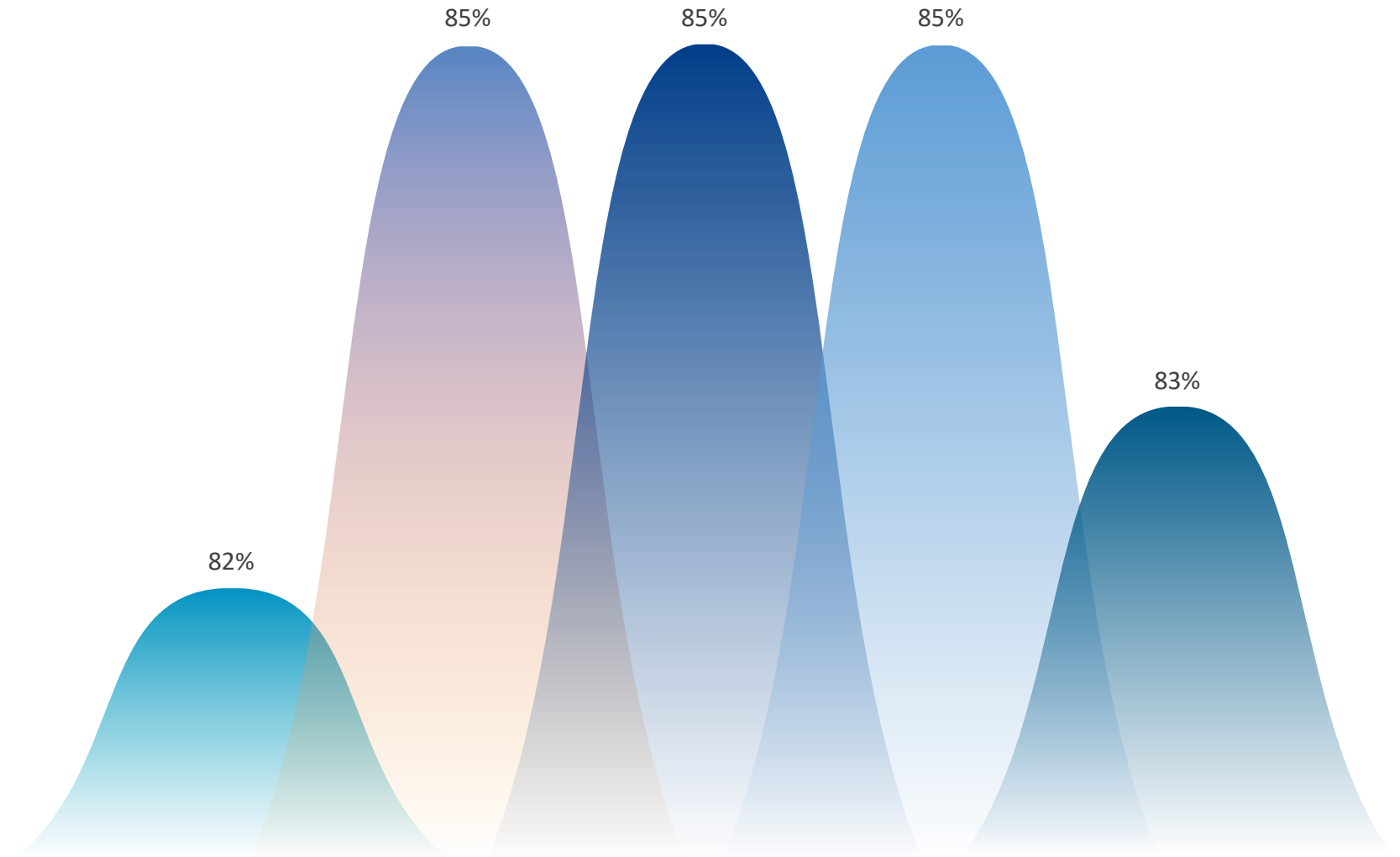
KNN Classifier

Gradient Boosting Classifier

Logistic Regression

Random Forest Classifier

SVC



RESULTS

GRADIENT BOOSTING CLASSIFIER

train roc_auc score: 0.94
test roc_auc score: 0.85

