

UNIVERSITY OF TARTU

Petfinder
Team

Today's Agenda



Who we are?

Brief about the project

Problem statement

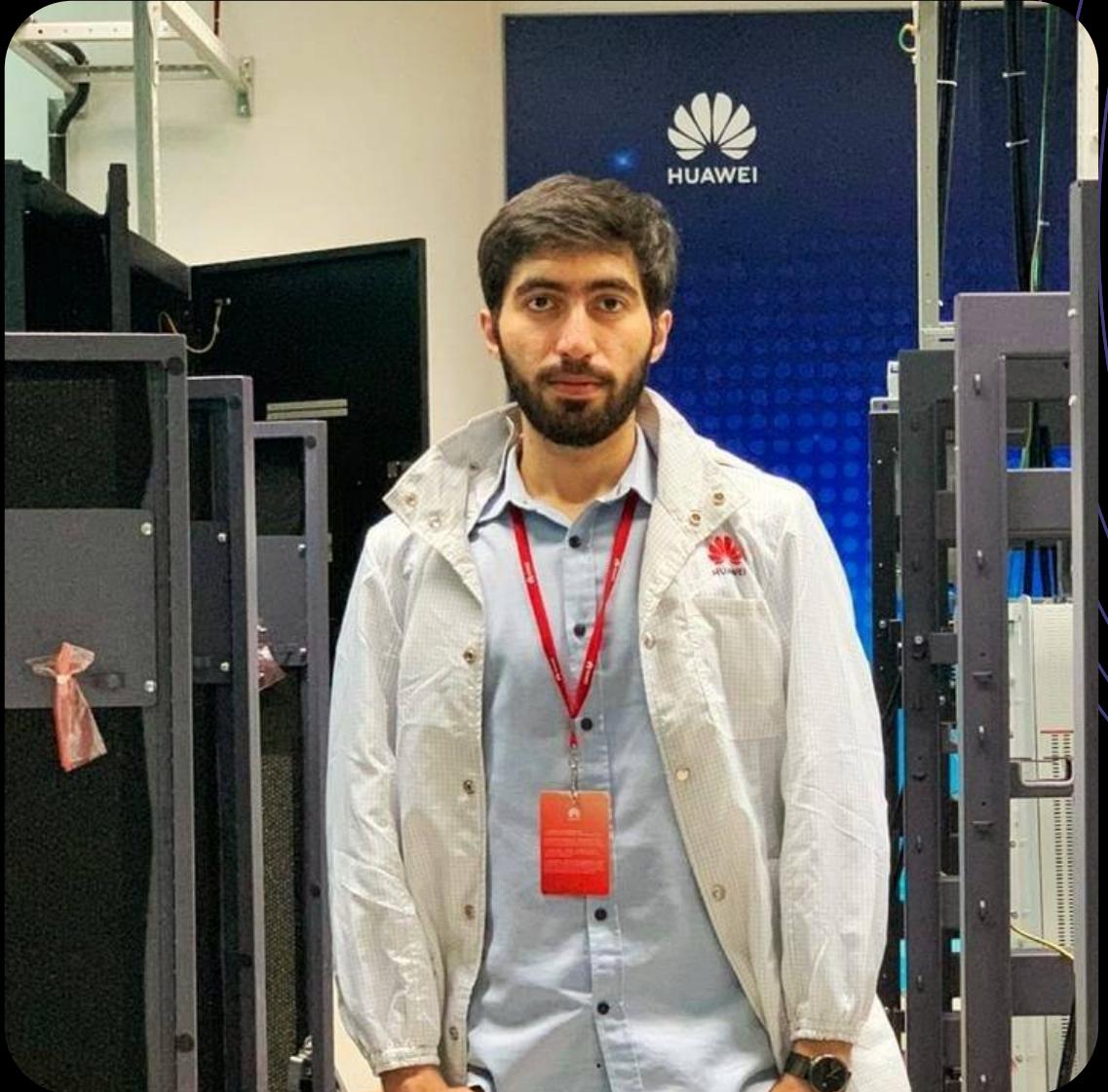
Project

Result of our work

Encountered difficulties

What did we learn?

Who we are?

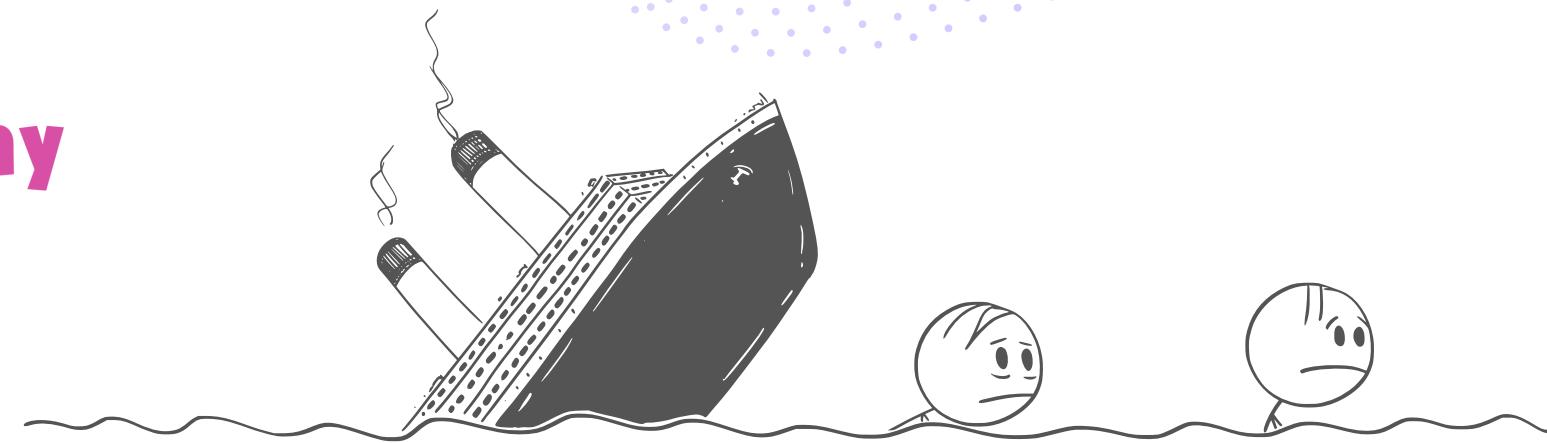
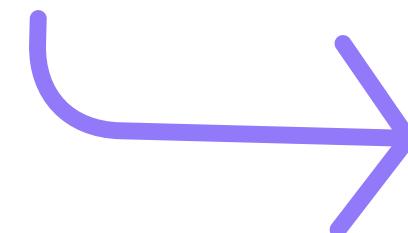


Kamil
Aliyev



Adil
Abdurrazakli

Brief about the project



Titanic

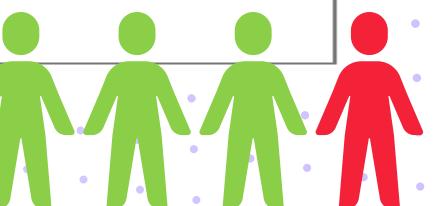
On April 15, 1912, the widely considered “unsinkable” Titanic sank after colliding with an iceberg.

Unfortunately, there weren’t enough lifeboats for everyone on board, resulting in the death of 1502 out of 2224 passengers and crew.

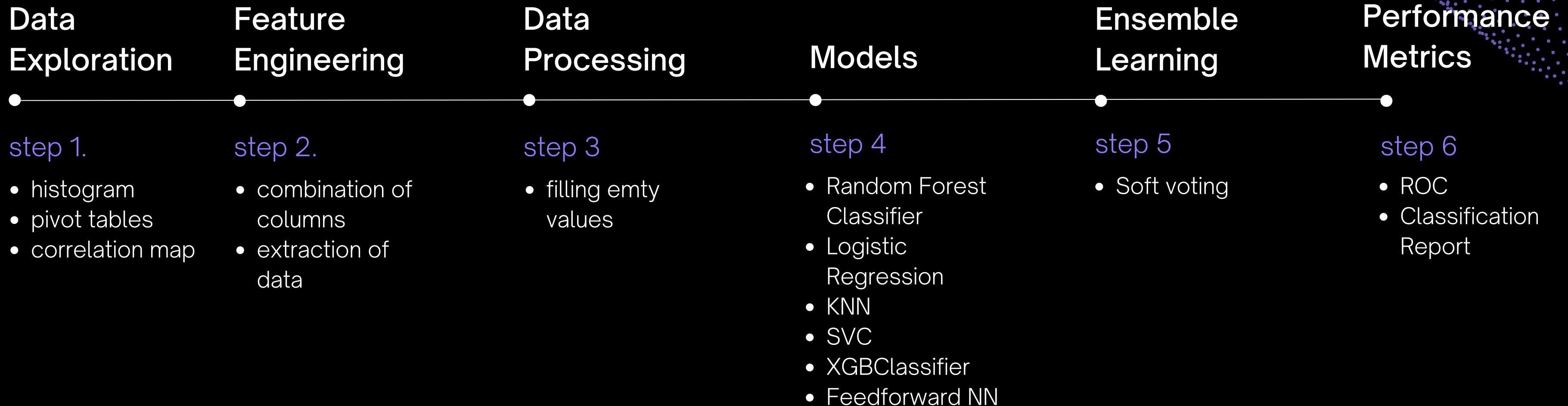
Problem Statement

“what sorts of people were more likely to survive?”

Variable	Definition	Key
survival	Survival	0 = No, 1 = Yes
pclass	Ticket class	1 = 1st, 2 = 2nd, 3 = 3rd
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	C = Cherbourg, Q = Queenstown, S = Southampton



Project



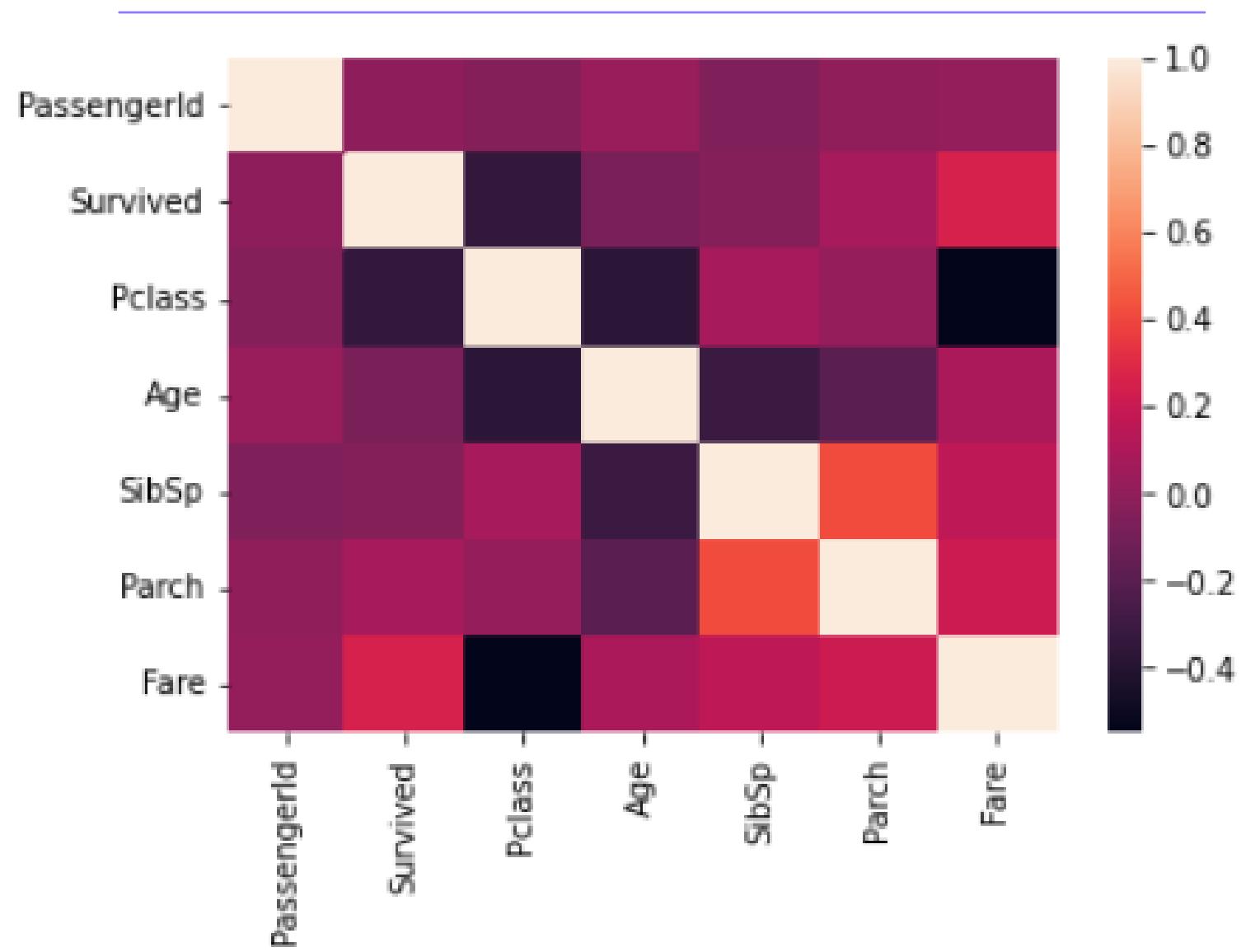
Data Exploration

Empty Values

#	Column	Non-Null Count	Dtype	#	Column	Non-Null Count	Dtype		
0	PassengerId	891	non-null	int64	0	PassengerId	418	non-null	int64
1	Survived	891	non-null	int64	1	Pclass	418	non-null	int64
2	Pclass	891	non-null	int64	2	Name	418	non-null	object
3	Name	891	non-null	object	3	Sex	418	non-null	object
4	Sex	891	non-null	object	4	Age	332	non-null	float64
5	Age	714	non-null	float64	5	SibSp	418	non-null	int64
6	SibSp	891	non-null	int64	6	Parch	418	non-null	int64
7	Parch	891	non-null	int64	7	Ticket	418	non-null	object
8	Ticket	891	non-null	object	8	Fare	417	non-null	float64
9	Fare	891	non-null	float64	9	Cabin	91	non-null	object
10	Cabin	204	non-null	object	10	Embarked	418	non-null	object
11	Embarked	889	non-null	object					

Data Exploration

Correlation map

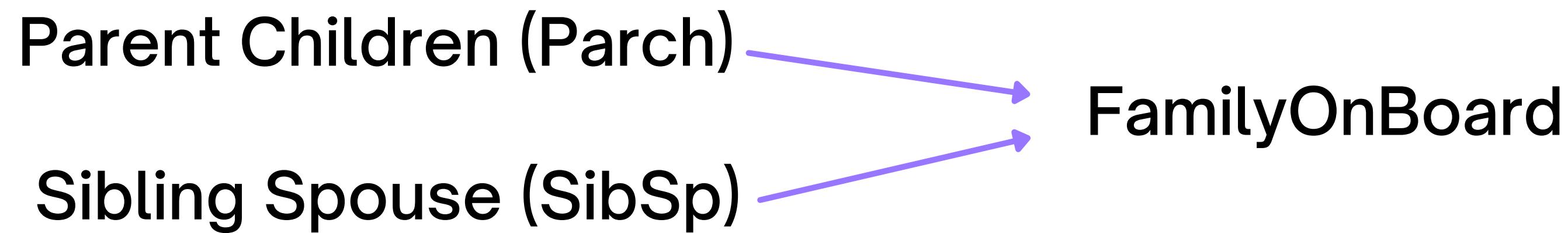


Data Exploration

Some interesting points

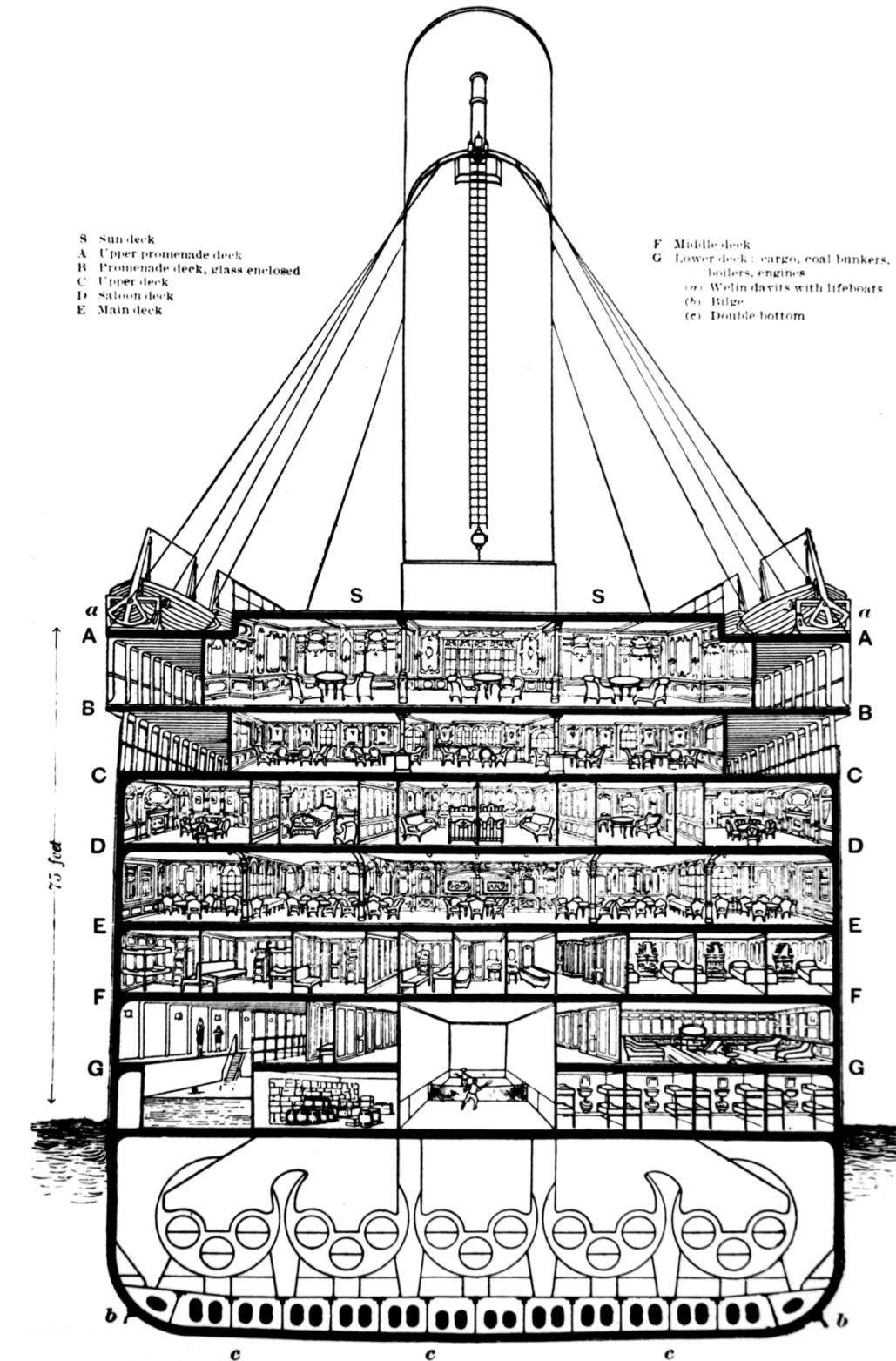
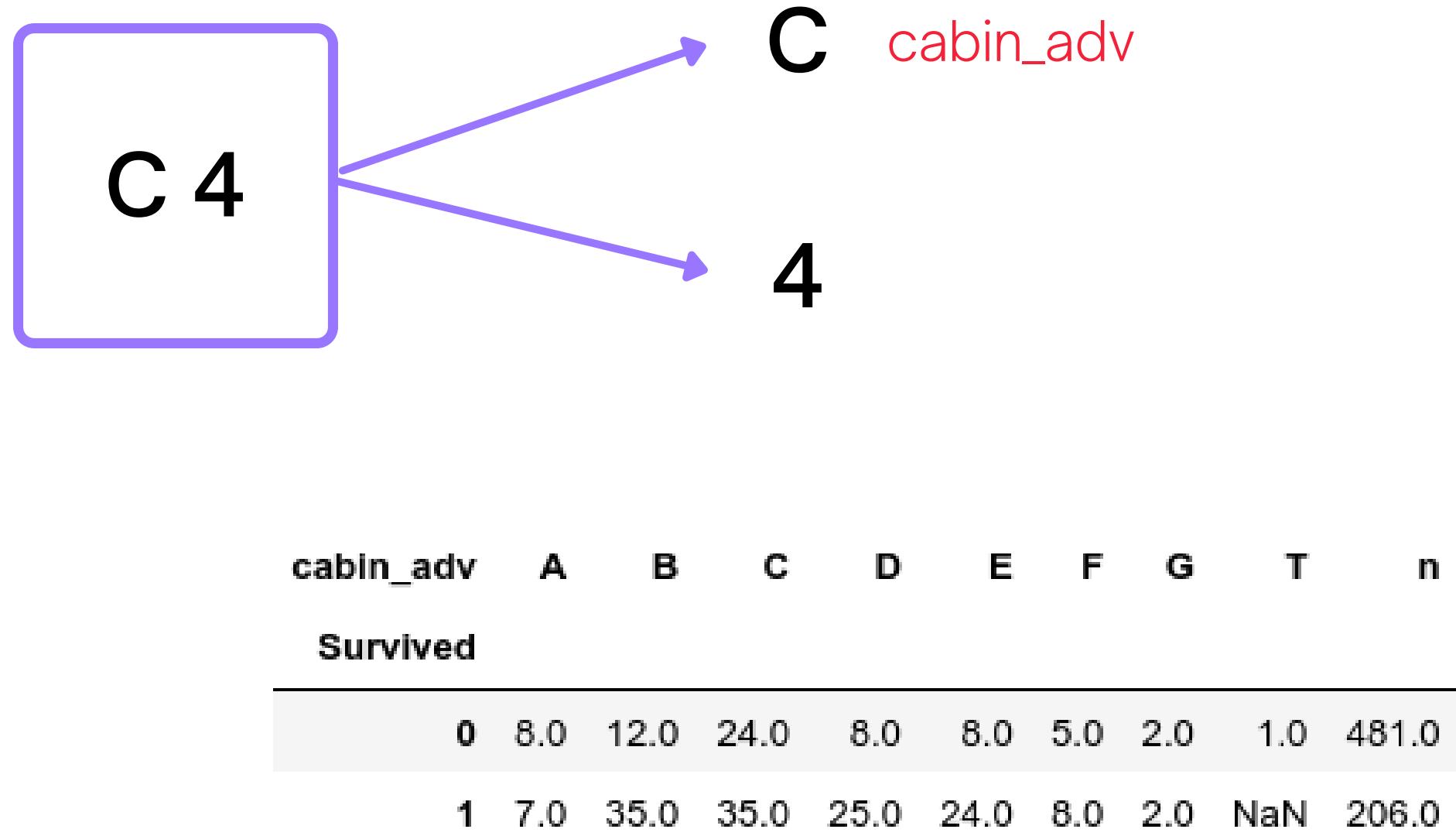
Sex	female	male	Embarked	C	Q	S	Pclass	Survived
Survived			Survived				0	1 0.629630
0	81	468	0	75	47	427	1	2 0.472826
1	233	109	1	93	30	217	2	3 0.242363

Feature Engineering



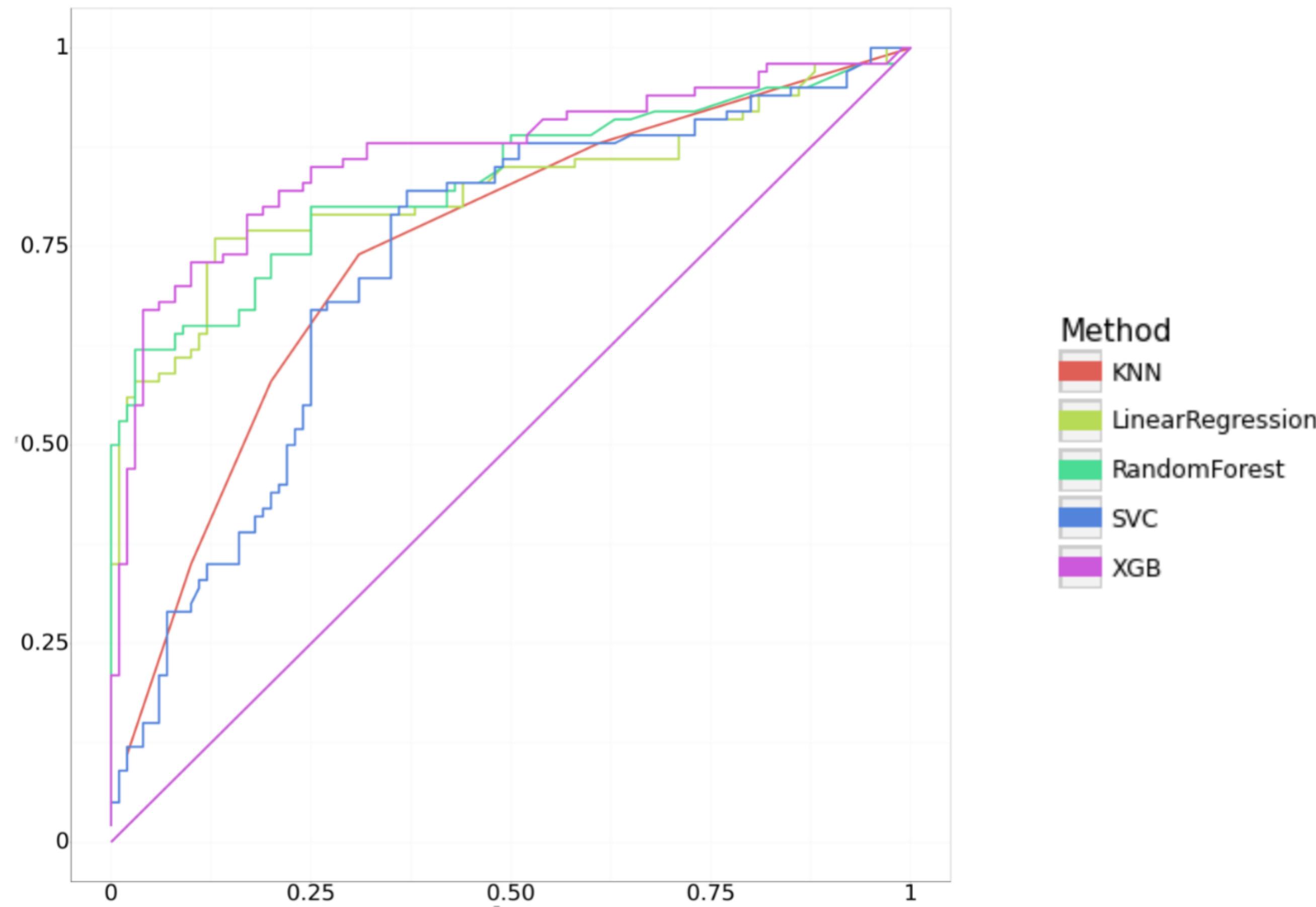
Feature Engineering

Cabin



model	accuracy	
Random Forest	0.82	
KNN	0.72	
Linear Regression	0.82	
SVC	0.68	
XGB	0.82	AUC of RandomForest classifier is 0.836
Soft Voting	0.82	AUC of KNN classifier is 0.748
Feed-forwarding NN	0.8316	AUC of LinearRegression classifier is 0.825 AUC of SVC classifier is 0.733 AUC of XGB classifier is 0.867

Models & Power Metrics



Models & Power Metrics

Result of our
work?

Public score: 0.78229
In top: 21%

Encountered difficulties

- Numpy not getting along with Pandas
- pd.get_dummies() on separate test and train dataframe leads to problem with categorical data
- Filling empty values with only median data of one column makes models biased

pd.get_dummies()

#	Column	Non-Null Count	Dtype
0	Pclass	891 non-null	int64
1	Age	891 non-null	float64
2	Fare	891 non-null	float64
3	FamilyOnBoard	891 non-null	int64
4	Sex_female	891 non-null	uint8
5	Sex_male	891 non-null	uint8
6	Embarked_C	891 non-null	uint8
7	Embarked_Q	891 non-null	uint8
8	Embarked_S	891 non-null	uint8
9	cabin_adv_A	891 non-null	uint8
10	cabin_adv_B	891 non-null	uint8
11	cabin_adv_C	891 non-null	uint8
12	cabin_adv_D	891 non-null	uint8
13	cabin_adv_E	891 non-null	uint8
14	cabin_adv_F	891 non-null	uint8
15	cabin_adv_G	891 non-null	uint8
16	cabin_adv_T	891 non-null	uint8
17	cabin_adv_n	891 non-null	uint8

train_x

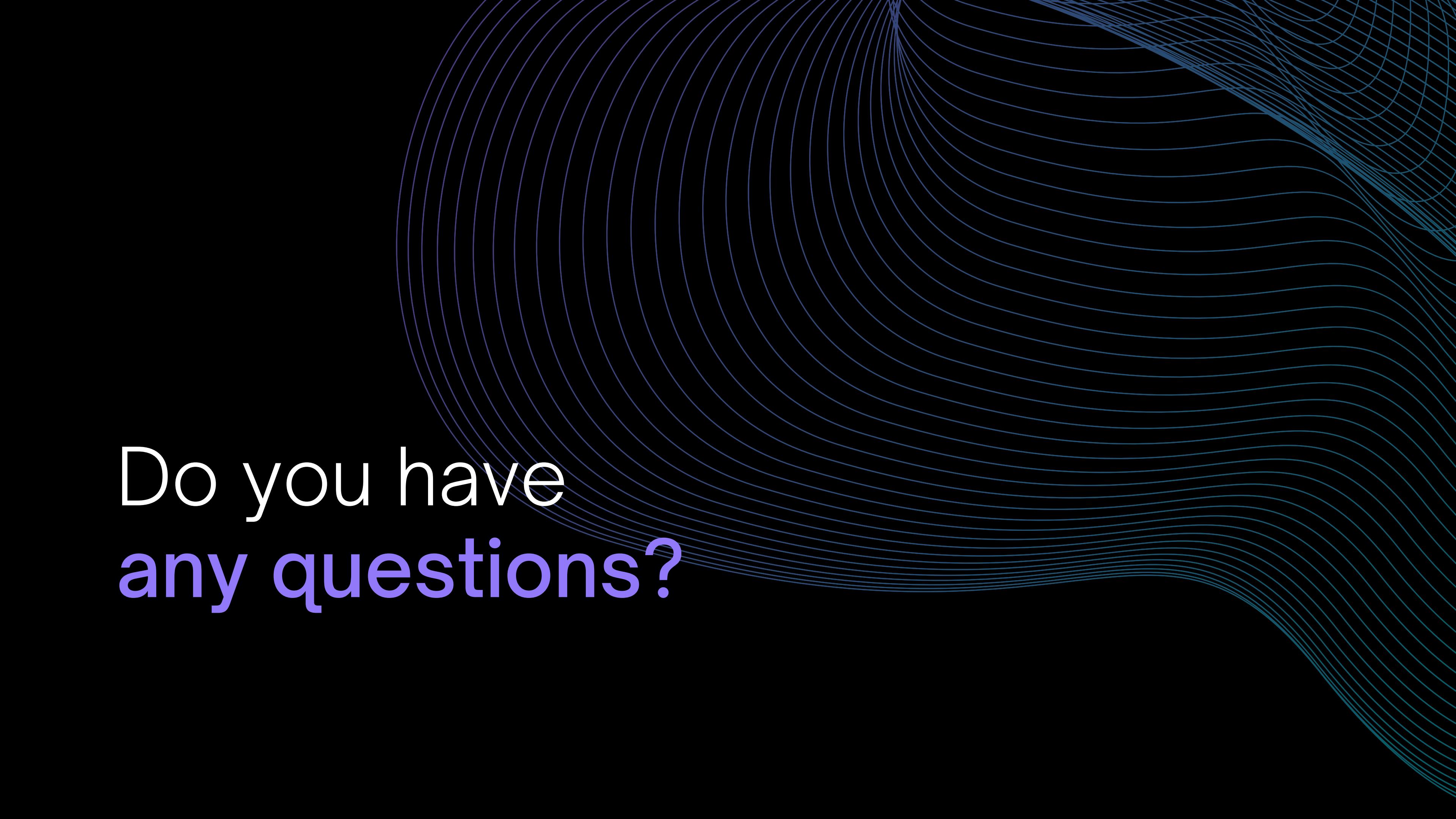


#	Column	Non-Null Count	Dtype
0	Pclass	418 non-null	int64
1	Age	418 non-null	float64
2	Fare	418 non-null	float64
3	FamilyOnBoard	418 non-null	int64
4	Sex_female	418 non-null	uint8
5	Sex_male	418 non-null	uint8
6	Embarked_C	418 non-null	uint8
7	Embarked_Q	418 non-null	uint8
8	Embarked_S	418 non-null	uint8
9	cabin_adv_A	418 non-null	uint8
10	cabin_adv_B	418 non-null	uint8
11	cabin_adv_C	418 non-null	uint8
12	cabin_adv_D	418 non-null	uint8
13	cabin_adv_E	418 non-null	uint8
14	cabin_adv_F	418 non-null	uint8
15	cabin_adv_G	418 non-null	uint8
16	cabin_adv_n	418 non-null	uint8

test_x

What did we learned?

- `sklearn.preprocessing.OneHotEncoder`
- How to fill empty cells by grouping with other columns?
- Models are just simple tools in any project.
The difficult part starts while playing with weights and parameters of them.
- **What AI realistically can and cannot do?**

The background features a dark gray or black surface with a subtle, flowing pattern of thin, light blue lines. These lines are arranged in concentric, undulating arcs that radiate from the bottom right corner towards the top left, creating a sense of depth and motion.

Do you have
any questions?

Project Link
here