# Task 1 - Code Documentation

## Introduction

# Training

## Code Cell 1

Code:

import pandas as pd  
import numpy as np  
  
df = pd.read\_csv("train.csv")  
df

Explanation:

This cell executes the above code. (You can expand this explanation as per requirement.)

## Code Cell 2

Code:

df.info()

Explanation:

This cell executes the above code. (You can expand this explanation as per requirement.)

## Code Cell 3

Code:

def fillNaMode(cols):  
 for i in cols:  
 df[i] = df[i].fillna(df[i].mode()[0])  
  
columns = ['Age','Cabin','Embarked']  
fillNaMode(columns)

Explanation:

This cell executes the above code. (You can expand this explanation as per requirement.)

## Code Cell 4

Code:

def changeFloattoInt64(cols):  
 for i in cols:  
 df[i] = df[i].astype('int64')  
  
columns = ['Age','Fare']  
changeFloattoInt64(columns)

Explanation:

This cell executes the above code. (You can expand this explanation as per requirement.)

## Code Cell 5

Code:

from sklearn.preprocessing import LabelEncoder  
  
def encodeCols(cols):  
 for i in cols:  
 temp = pd.DataFrame({i:df[i].unique()})  
 data\_LE = LabelEncoder()  
 data\_LE.fit(np.ravel(temp))  
   
 df[i] = data\_LE.transform(df[i])  
  
columns = ['Name','Sex','Ticket','Cabin','Embarked']  
encodeCols(columns)

Explanation:

This cell executes the above code. (You can expand this explanation as per requirement.)

## Code Cell 6

Code:

df.info()

Explanation:

This cell executes the above code. (You can expand this explanation as per requirement.)

## Code Cell 7

Code:

X = df.drop(columns=['Survived'])  
y = df['Survived']

Explanation:

This cell executes the above code. (You can expand this explanation as per requirement.)

## Code Cell 8

Code:

from sklearn.linear\_model import LogisticRegression  
  
model = LogisticRegression(max\_iter=1000, random\_state=42)  
model.fit(X, y)

Explanation:

This cell executes the above code. (You can expand this explanation as per requirement.)

## Code Cell 9

Code:

import pandas as pd  
import numpy as np  
  
df = pd.read\_csv("test.csv")  
df

Explanation:

This cell executes the above code. (You can expand this explanation as per requirement.)

## Code Cell 10

Code:

df.info()

Explanation:

This cell executes the above code. (You can expand this explanation as per requirement.)

## Code Cell 11

Code:

def fillNaMode(cols):  
 for i in cols:  
 df[i] = df[i].fillna(df[i].mode()[0])  
  
columns = ['Age','Fare','Cabin']  
fillNaMode(columns)

Explanation:

This cell executes the above code. (You can expand this explanation as per requirement.)

## Code Cell 12

Code:

def changeFloattoInt64(cols):  
 for i in cols:  
 df[i] = df[i].astype('int64')  
  
columns = ['Age','Fare']  
changeFloattoInt64(columns)

Explanation:

This cell executes the above code. (You can expand this explanation as per requirement.)

## Code Cell 13

Code:

from sklearn.preprocessing import LabelEncoder  
  
def encodeCols(cols):  
 for i in cols:  
 temp = pd.DataFrame({i:df[i].unique()})  
 data\_LE = LabelEncoder()  
 data\_LE.fit(np.ravel(temp))  
   
 df[i] = data\_LE.transform(df[i])  
  
columns = ['Name','Sex','Ticket','Cabin','Embarked']  
encodeCols(columns)

Explanation:

This cell executes the above code. (You can expand this explanation as per requirement.)

## Code Cell 14

Code:

df.info()

Explanation:

This cell executes the above code. (You can expand this explanation as per requirement.)

## Code Cell 15

Code:

y\_predict = model.predict(df)

Explanation:

This cell executes the above code. (You can expand this explanation as per requirement.)

## Code Cell 16

Code:

y\_predict

Explanation:

This cell executes the above code. (You can expand this explanation as per requirement.)

## Code Cell 17

Code:

submission = pd.read\_csv("gender\_submission.csv")  
submission['Survived'] = y\_predict  
submission.to\_csv("submission.csv", index=False)

