ZONGO DEVS PROJECT PRESENTATION

- A CASE STUDY ON THE TITANIC Disaster DATASET;

15TH APRIL. 1912
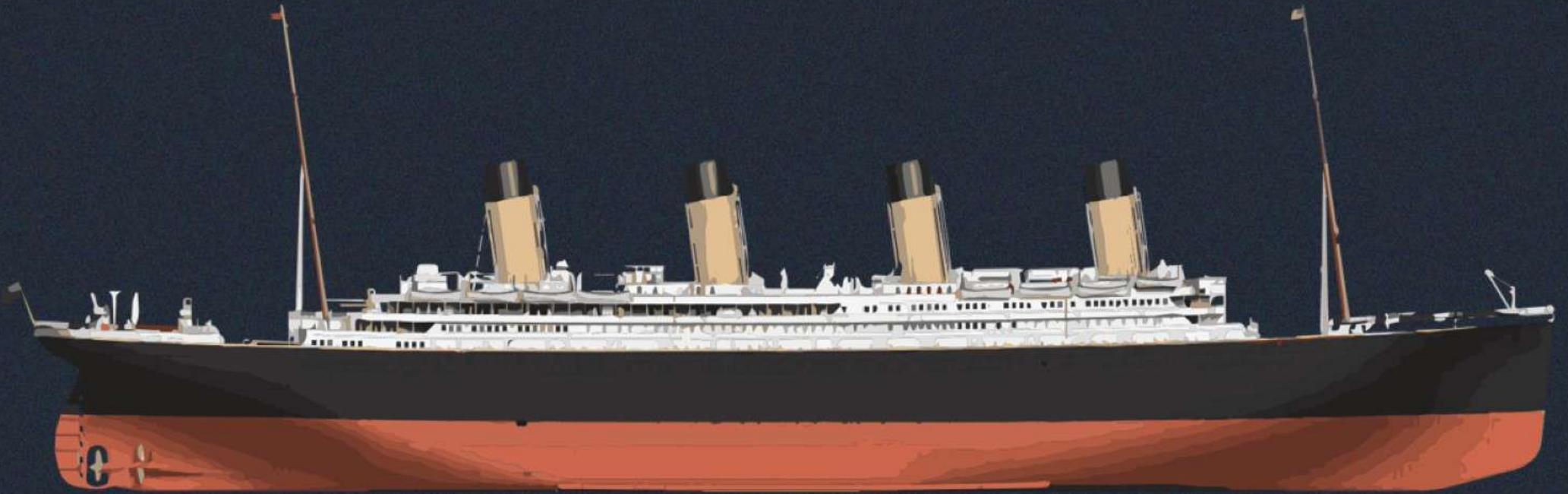
BRIEF HISTORY ON THE TITANIC

882 FEET 6 INCHES (269M)

BRIEF HISTORY ON THE TITANIC…

The RMS Titanic was a British luxury passenger liner that sank on April 15, 1912, during its maiden voyage from Southampton, England, to New York City.

The ship, considered "unsinkable" due to its advanced safety features, collided with an iceberg shortly before midnight on April 14.

The impact caused five of its sixteen watertight compartments to rupture, leading to the ship's sinking

KEY TIMELINES

|	
MAR 31, 1909	CONSTRUCTION BEGINS
MAY 31, 1911	TITANIC IS LAUNCHED
APR 10, 1912	MAIDEN VOYAGE
APR 14, 1912	ICEBERG SPOTTED
APR 15, 1912	FIRST DISTRESS SIGNALS SEND
APR 15, 1912	TITANIC SINKS

BRIEF HISTORY ...

DURATION 2hr and 40min

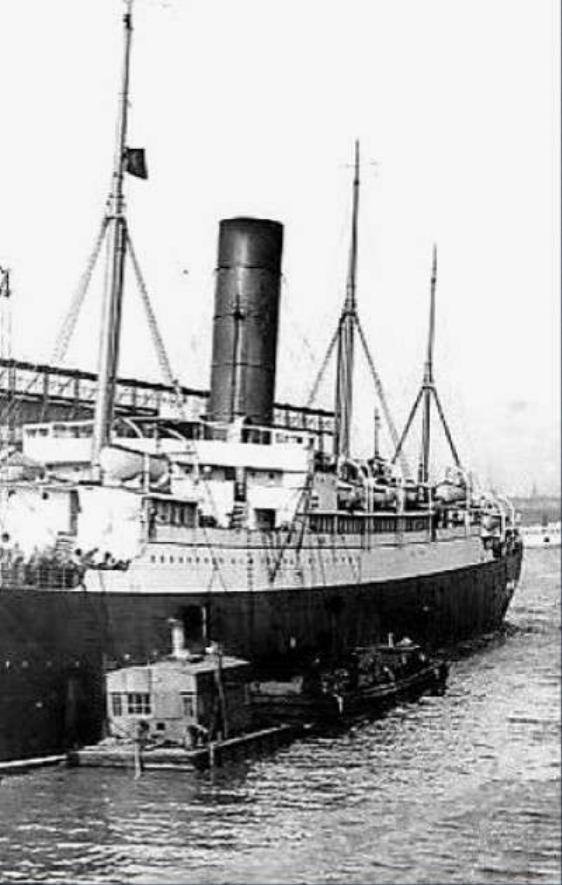**LOCATION** North Atlantic Ocean, 600 km southeast of Newfoundland

In 1985, the wreck of the Titanic was discovered at a depth of about 13,000 feet (4,000 meters) in the North Atlantic Ocean
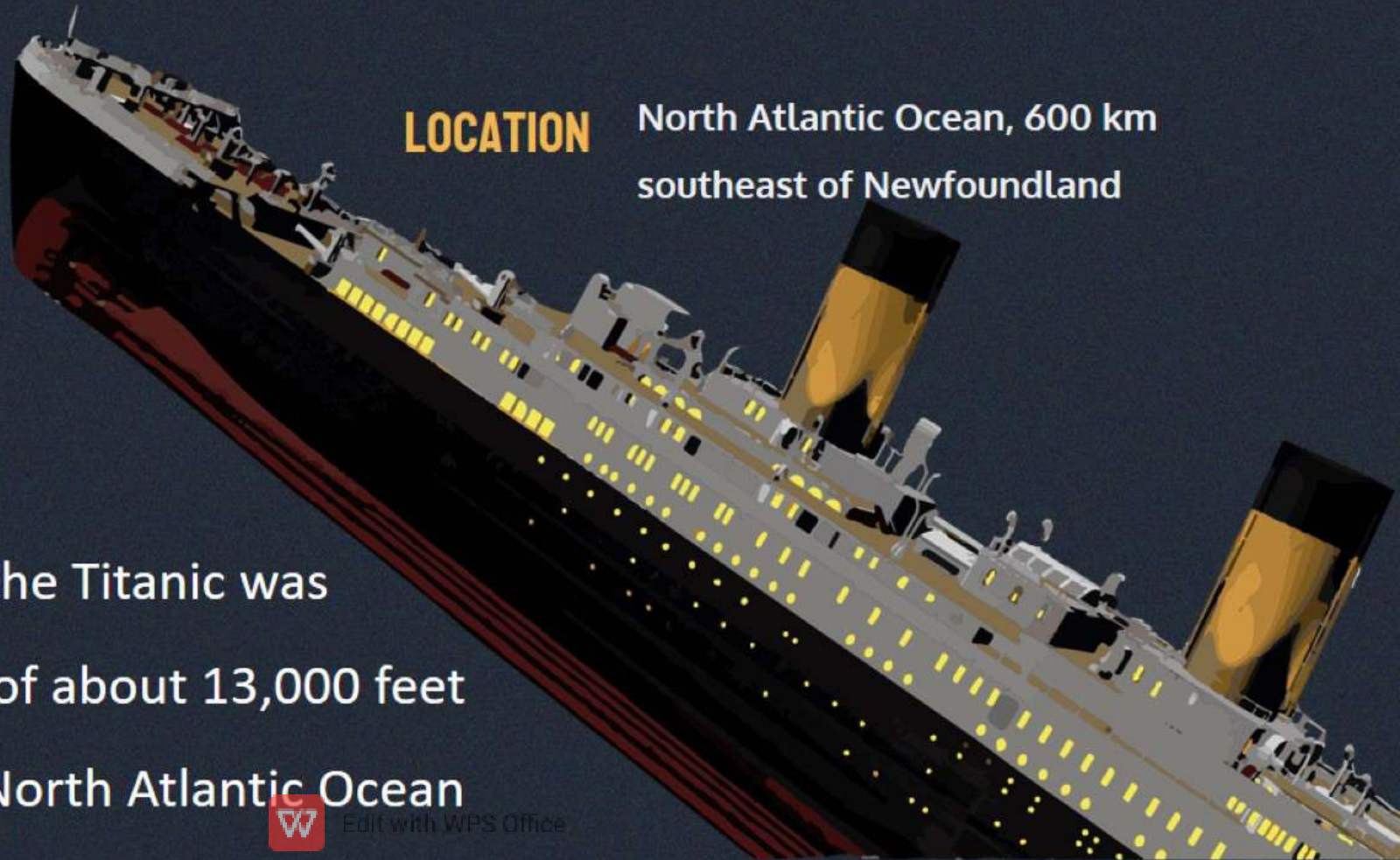
PROJECT OBJECTIVES

The objectives of the Project is as follows;

1. To do a simple cleaning of the Dataset
2. To do an Exploratory Data Analysis (EDA)
3. To do some descriptive statistics
4. To do an analysis of survival patterns of the Disaster.

UPLOADING ESSENTIAL LIBRARIES & AND READING OF THE DATASET

#We firstly;

- import pandas as pd
- import numpy as np
- import matplotlib.pyplot as plt
- import seaborn as sns
- from sklearn.preprocessing import LabelEncoder

- Reading of the   Dataset (Train);
- Train = pd.read_csv('/content/train.csv')

OBJECTIVE ONE (1)

- To do a data cleaning of the Dataset

  I.   We dropped irrelevant variables using the

  train.drop(columns=['Name', 'Ticket', 'Cabin'],inplace=True)

  II.   We encode categorical variables;

  train['Sex']=encode.fit_transform(train["Sex"])

  III. We find missing data; *.isna ( ) . sum( )*

  IV.   We fill in missing data appropriately accordingly using classes;

  train.loc[(train['Pclass'] == 1) & (train['Age'].isna()),'Age'] =37

OBJECTIVE TWO (2)

To do an Exploratory Data Analysis (EDA) on the dataset

Here, we employed the following functions;
➢ .Shape ( )
➢ .info ( )
➢ .head ( )
➢ .describe ( )

We also did some data visualization using;
Pie Chart.

OBJECTIVE TWO (3)

- To do some descriptive statistics

  Here, we focus on the interpretation of our " .describe ( )" to derive interesting understanding of our data.

OBJECTIVE TWO (3); To do some descriptive statistics

```
train.describe()
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

OBJECTIVE TWO (3); To do some descriptive statistics

Considering the above output, it was realized that the Count gives as a complete sum of 891 entries or observation in our dataset. This signifies the total number of individuals onboarded. The mean age of the population was 29 years, with 25% of the population being within 20 years and below, while the median age is 28 years. 75% of the population are within 38years and below. The youngest age was 0.42 years approximately 5months, while the oldest was 80years. It was also observed that 38% of the population survived, whiles 62% of the population loosing their lives. In conclusion, 891 individual onboarded the ship ranging from babies, adults and aged. majority of people who boarded died, which account for 62% of the total population.

OBJECTIVE TWO (4); To do an analysis of survival patterns of the Disaster.

We firstly do an estimate of how many people survived and how many loss their lives.

survival_counts = train['Survived'].value_counts( )

```
print(survival_counts)

Survived
0       549
1       342
Name: count, dtype: int64
```

OBJECTIVE TWO (4); To do an analysis of survival patterns of the Disaster.

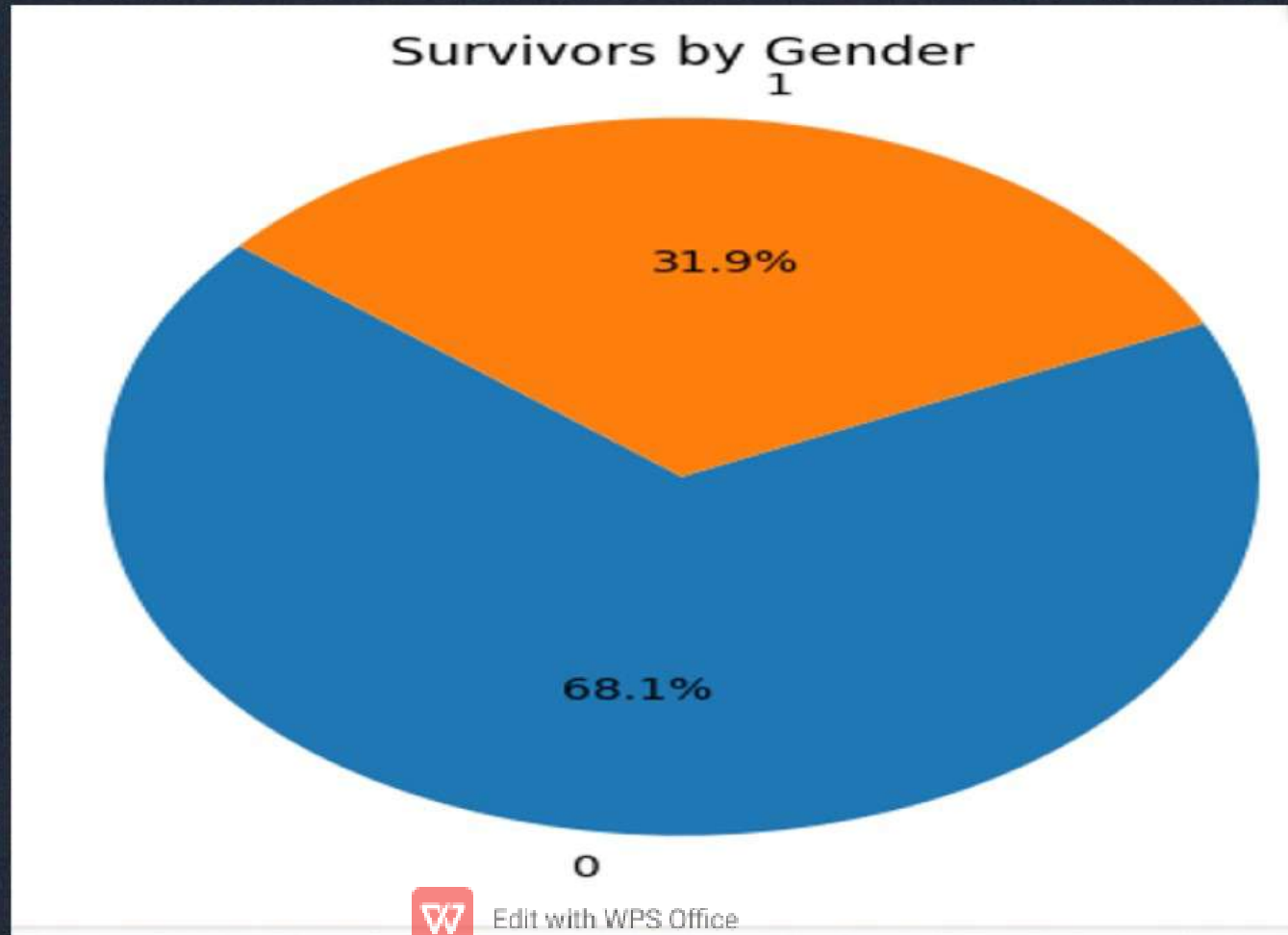We then estimated the patters of 'survived' agains 'Gender' in a pie chart;

```
plt.pie(gsurvival, labels=gsurvival.index, autopct='%1.1f%%', startangle=140)
plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.
plt.title("Survivors by Gender")
plt.show()
```

The Pie chart below indicates that, majority of the death cases where men accounting for 68.1% whiles the female death rate was 38.9%

OBJECTIVE TWO (4); To do an analysis of survival patterns of the Disaster.

Survivors by Gender

1

31.9%

68.1%

0

"I'LL NEVER LET GO, JACK."

THANK YOU