

Laboratory Guide: Computational Thinking with the Titanic Dataset using Python

Objective

By the end of this laboratory exercise, students will:

- Apply the four pillars of computational thinking: Decomposition, Pattern Recognition, Abstraction, and Algorithm Design.
- Analyze the Titanic dataset to identify key survival factors.
- Implement a simple rule-based model for predicting survival.

Pre-requisites

- Basic knowledge of Python.
- Familiarity with Pandas for data manipulation.

Dataset

Download the **Titanic dataset** from Kaggle: <https://www.kaggle.com/c/titanic/data>

- Note that when you download the file, the file name to be used is `train.csv`. For this case, you may need to rename your file from `train.csv` to `titanic.csv`

Place `titanic.csv` in the working directory.

Lab Instructions

Step 1: Decomposition

1. Open a Jupyter Notebook or any Python environment.
2. Load the Titanic dataset and inspect its structure.

Load the Dataset

Run the following Python code:

```
import pandas as pd

# Load the dataset
df = pd.read_csv("titanic.csv")

# Display the first few rows
df.head()
```

Question 1:

Up to this point, determine the underlying tasks needed to be done for loading and handling the data based on the steps described in this part? Write this as a text in your Colab notebook

Step 2: Pattern Recognition

Now, let's analyze patterns in the dataset.

Identify Missing Data

Run the following Python code:

```
df.isnull().sum()
```

💡 Question 2:

Using computation thinking, determine the blank values in the data. Write the algorithm as a text in your Google Colab

- Which columns contain missing values? Should we remove or fill them?

Explore Survival Rates

Run the following Python code:

```
df["Survived"].mean()
```

💡 Question 3:

Using computation thinking, determine the survival rate of all passengers in the dataset. Write the algorithm as a text in your Google Colab.

- What can you deduce on the survival rate of the passengers?

Analyze Survival by Age

Run the following Python code:

```
import matplotlib.pyplot as plt

# Plot age distribution of survivors vs. non-survivors
plt.hist(df[df["Survived"] == 1]["Age"].dropna(), bins=20, alpha=0.5, label="Survived")
plt.hist(df[df["Survived"] == 0]["Age"].dropna(), bins=20, alpha=0.5, label="Did Not Survive")
plt.legend()
plt.xlabel("Age")
plt.ylabel("Count")
plt.show()
```

💡 Question 4:

Using computation thinking, determine the average age of the survivors and non-survivors in the dataset. Write the algorithm as a text in your Google Colab.

- Does age impact survival? How?