In [ ]:

```
Task 1: PERFORM DATA CLEANING
clean a dataset by removing missing values and outliners
By SHURUTHI R S
```

In [ ]:

```
#IMPLEMENTING THE DEPENDENCIES
```

In [1]:

```
import pandas as pd
import numpy as np
iTask 1: PERFORM DATA CLEANING
clean a dataset by removing missing values and outliners
By SHURUTHI import seaborn as sns
```

In [12]:

```
#DATA READING
```

In [2]:

```
gender_data = pd.read_csv("gender_submission.csv")
print(gender_data)
```

```
     PassengerId  Survived
0            892         0
1            893         1
2            894         0
3            895         0
4            896         1
..           ...       ...
413         1305         0
414         1306         1
415         1307         0
416         1308         0
417         1309         0

[418 rows x 2 columns]
```

In [ ]:

```
#DATA CLEANING
#Fill the missing values for passenger id and survival columns.In order to fill the miss
#will fill the missing values of both the columns by taking the mean of all columns
```

In [8]:

```
#fill passengerID column
gender_data["PassengerId"].fillna(gender_data["PassengerId"].mean(),inplace = True)
gender_data["PassengerId"].isna().sum()
```

Out[8]:

```
0
```

In [9]:

```
#fill survived column
gender_data["Survived"].fillna(gender_data["Survived"].mean(), inplace=True)
gender_data["Survived"].isna().sum()
```

Out[9]:

```
0
```

In [ ]:

```
#Alternatively we will visualize the null value using heatmap
#we will use heatmap method by passing only records which are null
```

In [10]:

```
sns.heatmap(gender_data.isna())
```

Out[10]:

```
<AxesSubplot:>
```

In [ ]:

```
#we can conclude from the above heatmap that there is no null value left in our dataset
```