

File Edit View Insert Cell Kernel Help Trusted Python 3

Save Add Split Copy Paste Undo Redo Run Stop Restart Markdown Keyboard

Information about the dataset:

The dataset contains cases from a study that was conducted between 1958 and 1970 at the University of Chicago's Billings Hospital on the survival of patients who had undergone surgery for breast cancer.

Number of Instances: 306 Number of Attributes: 4 (including the class attribute) Attribute Information:

1. Age = Age of patient at time of operation (numerical)
2. Year = Patient's year of operation (year - 1900, numerical)
3. Nodes = Number of positive axillary nodes detected (numerical)
4. Status = Survival status (class attribute) 1 = the patient survived 5 years or longer 2 = the patient died within 5 year

In [83]: *# Importing the Libraries*

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
import warnings
warnings.filterwarnings('ignore')
```

File Location:

In [111]: `FILE_PATH = "https://raw.githubusercontent.com/Rahul-Kumaran/Exploratory_Data_Analysis_Haberman/master/Input/haberma`

Importing the dataset " Haberman.csv "

In [112]: *# Using the pandas library the csv file is imported to analyse the data*
`haberman = pd.read_csv(FILE_PATH)`

In [113]: *# This command shows the first five rows and information in the dataset*
`haberman.head()`

Out [113]: