

# Data Warehousing and Mining Lab

## Assignment - 9

### CSE 326

13 April 2022

Gyanendra Kumar Shukla  
*CSE 1*  
191112040



*Department of Computer Science*  
*NIT, Bhopal*

# Contents

<b>1</b>	<b>Uploading Files in Google Colab</b>	<b>2</b>
1.1	CSV Upload . . . . .	2
1.1.1	CODE . . . . .	2
1.1.2	Screenshot . . . . .	2
1.2	Excel Upload . . . . .	3
1.2.1	CODE . . . . .	3
1.2.2	Screenshot . . . . .	3
1.3	CSV Git . . . . .	4
1.3.1	CODE . . . . .	4
1.3.2	Screenshot . . . . .	4
1.4	Excel Git . . . . .	5
1.4.1	CODE . . . . .	5
1.4.2	Screenshot . . . . .	5

# 1 Uploading Files in Google Colab

## 1.1 CSV Upload

### 1.1.1 CODE

```
from google.colab import files
uploaded = files.upload()

import pandas as pd
df = pd.read_csv('titanic.csv')
print(df.head())
```

### 1.1.2 Screenshot

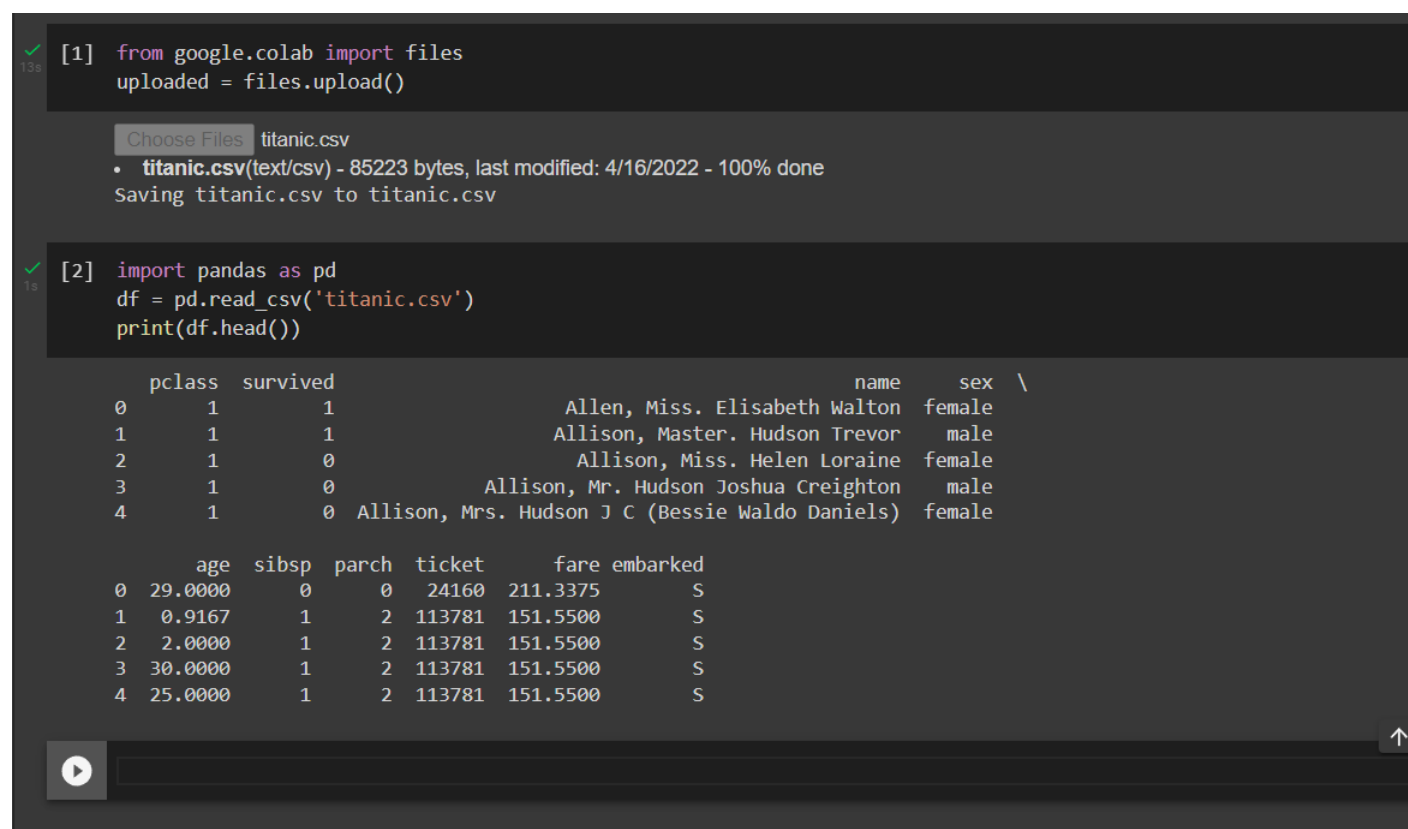


Figure 1: CSV file, uploaded using colab

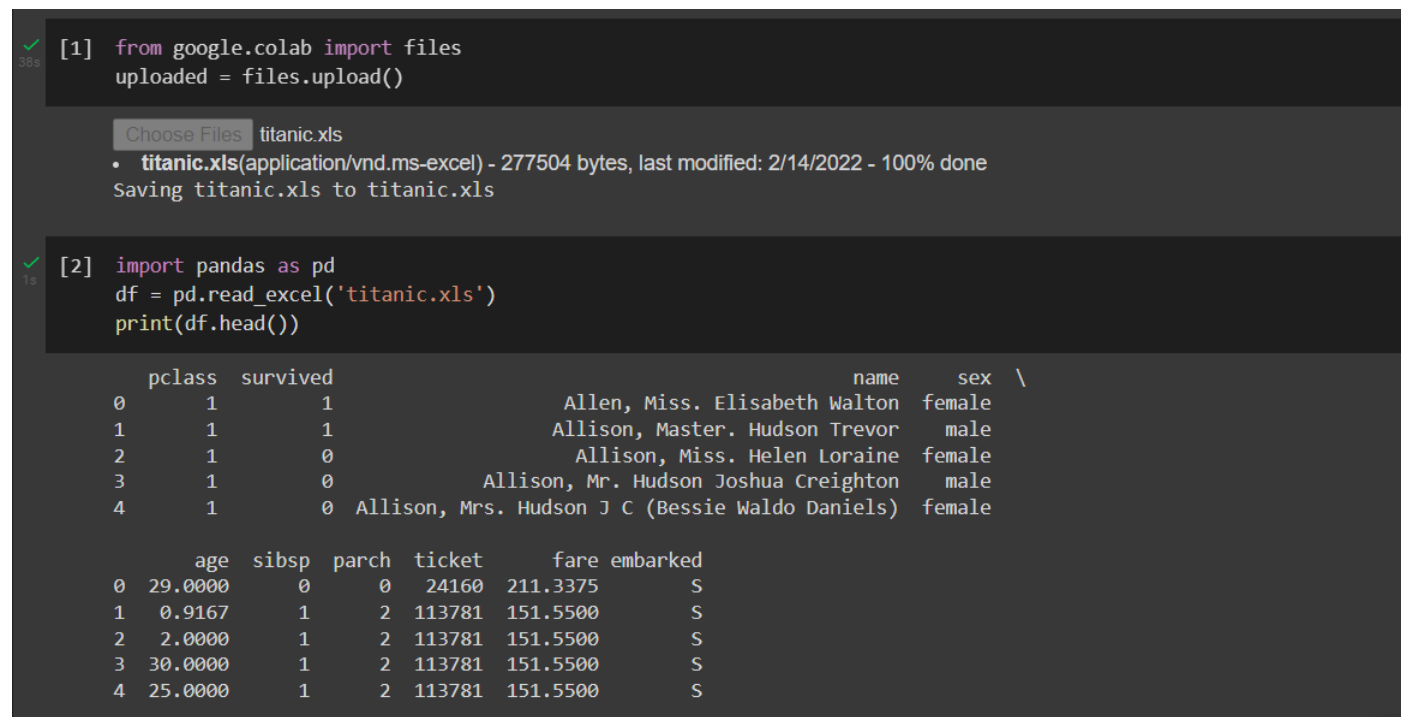
## 1.2 Excel Upload

### 1.2.1 CODE

```
from google.colab import files
uploaded = files.upload()

import pandas as pd
df = pd.read_excel('titanic.xls')
print(df.head())
```

### 1.2.2 Screenshot



The screenshot shows the Google Colab interface. In the first code cell, the file 'titanic.xls' is uploaded. A file selection dialog is shown with 'titanic.xls' selected. Below the dialog, a list of files is displayed: 'titanic.xls(application/vnd.ms-excel) - 277504 bytes, last modified: 2/14/2022 - 100% done'. The second code cell shows the data being read into a pandas DataFrame and the first five rows being printed.

```
[1] from google.colab import files
    uploaded = files.upload()

Choose Files titanic.xls
• titanic.xls(application/vnd.ms-excel) - 277504 bytes, last modified: 2/14/2022 - 100% done
Saving titanic.xls to titanic.xls

[2] import pandas as pd
    df = pd.read_excel('titanic.xls')
    print(df.head())
```

	pclass	survived	name	sex
0	1	1	Allen, Miss. Elisabeth Walton	female
1	1	1	Allison, Master. Hudson Trevor	male
2	1	0	Allison, Miss. Helen Loraine	female
3	1	0	Allison, Mr. Hudson Joshua Creighton	male
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female

	age	sibsp	parch	ticket	fare	embarked
0	29.0000	0	0	24160	211.3375	S
1	0.9167	1	2	113781	151.5500	S
2	2.0000	1	2	113781	151.5500	S
3	30.0000	1	2	113781	151.5500	S
4	25.0000	1	2	113781	151.5500	S

Figure 2: Excel file, uploaded using colab

## 1.3 CSV Git

### 1.3.1 CODE

```
!git clone https://github.com/singhkhushi25/6semdata.git
```

```
import pandas as pd
df = pd.read_csv('/content/6semdata/titanic.csv')
print(df.head())
```

### 1.3.2 Screenshot

The screenshot shows a Jupyter Notebook with two cells. Cell [1] executes the command to clone a GitHub repository, and Cell [2] imports pandas and loads a CSV file. The output of Cell [2] displays the first five rows of the Titanic dataset, showing passenger details and their survival status.

```
[1] !git clone https://github.com/singhkhushi25/6semdata.git
```

Cloning into '6semdata'...  
remote: Enumerating objects: 6, done.  
remote: Counting objects: 100% (6/6), done.  
remote: Compressing objects: 100% (5/5), done.  
remote: Total 6 (delta 0), reused 0 (delta 0), pack-reused 0  
Unpacking objects: 100% (6/6), done.

```
[2] import pandas as pd
df = pd.read_csv('/content/6semdata/titanic.csv')
print(df.head())
```

	pclass	survived	name	sex
0	1.0	1.0	Allen, Miss. Elisabeth Walton	female
1	1.0	1.0	Allison, Master. Hudson Trevor	male
2	1.0	0.0	Allison, Miss. Helen Loraine	female
3	1.0	0.0	Allison, Mr. Hudson Joshua Creighton	male
4	1.0	0.0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female

	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body
0	29.0000	0.0	0.0	24160	211.3375	B5	S	2	NaN
1	0.9167	1.0	2.0	113781	151.5500	C22 C26	S	11	NaN
2	2.0000	1.0	2.0	113781	151.5500	C22 C26	S	NaN	NaN
3	30.0000	1.0	2.0	113781	151.5500	C22 C26	S	NaN	135.0
4	25.0000	1.0	2.0	113781	151.5500	C22 C26	S	NaN	NaN

	home.dest
0	St Louis, MO
1	Montreal, PQ / Chesterville, ON
2	Montreal, PQ / Chesterville, ON
3	Montreal, PQ / Chesterville, ON
4	Montreal, PQ / Chesterville, ON

Figure 3: CSV file, uploaded using git cloning

## 1.4 Excel Git

### 1.4.1 CODE

```
!git clone https://github.com/singhkhushi25/6semdata.git
```

```
import pandas as pd
df = pd.read_excel('/content/6semdata/titanic (3).xls')
print(df.head())
```

### 1.4.2 Screenshot

```
[1] !git clone https://github.com/singhkhushi25/6semdata.git

Cloning into '6semdata'...
remote: Enumerating objects: 6, done.
remote: Counting objects: 100% (6/6), done.
remote: Compressing objects: 100% (5/5), done.
remote: Total 6 (delta 0), reused 0 (delta 0), pack-reused 0
Unpacking objects: 100% (6/6), done.

[3] import pandas as pd
df = pd.read_excel('/content/6semdata/titanic (3).xls')
print(df.head())
```

	pclass	survived	name	sex
0	1	1	Allen, Miss. Elisabeth Walton	female
1	1	1	Allison, Master. Hudson Trevor	male
2	1	0	Allison, Miss. Helen Loraine	female
3	1	0	Allison, Mr. Hudson Joshua Creighton	male
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female

	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body
0	29.0000	0	0	24160	211.3375	B5	S	2	NaN
1	0.9167	1	2	113781	151.5500	C22 C26	S	11	NaN
2	2.0000	1	2	113781	151.5500	C22 C26	S	NaN	NaN
3	30.0000	1	2	113781	151.5500	C22 C26	S	NaN	135.0
4	25.0000	1	2	113781	151.5500	C22 C26	S	NaN	NaN

	home.dest
0	St Louis, MO
1	Montreal, PQ / Chesterville, ON
2	Montreal, PQ / Chesterville, ON
3	Montreal, PQ / Chesterville, ON
4	Montreal, PQ / Chesterville, ON

Figure 4: Excel file, uploaded using git cloning