

# Intro to Data Science - Lab 3

DATA 1501 — Dr. Mihail  
Department of Computer Science  
Valdosta State University

September 8, 2021

## 1 Introduction

In this lab, you will learn how to import, modify and export a dataset using Python. In Part 1, you will download a fictitious dataset of grades from the Internet using Python and create a Pandas dataframe. In Part 2, you will add a new attribute to the dataset, and in Part 3 you will plot a histogram of the average scores.

### 1.1 Part 1

Create a new Colab notebook and run the following code in a code cell:

```
!wget https://cs.valdosta.edu/~rpmihail/DATA1500/lab3/grades.csv
```

Confirm that the file was downloaded and no errors reported from misspellings. You should see something like: *HTTP request sent, awaiting response... 200 OK*.

In the next part, you will write code that creates a Pandas dataframe from the file you just downloaded. Create a new code cell and run the following code:

```
import pandas as pd
df = pd.read_csv('grades.csv')
df
```

### 1.2 Part 2

Now, you will add an attribute to your dataset, initially populated with NaN (Not A Number). Create a code cell and run the following code:

```
import numpy as np
df["Average Exam"] = np.nan
df
```

Notice there is a new attribute now named “Average Exam”. You will now populate it by computing the average of Exam 1 and Exam 2. This operation is done for each record (row), and consists of adding attributes Exam 1 and Exam 2, then dividing the result by 2. Create a new code cell and run the following code:

```
df["Average Exam"] = (df["Exam 1"] + df["Exam 2"]) / 2
df
```

Notice how executing the above code populated values into attribute “Average Exam”. If everything worked, you will now download the newly edited dataset that contains the average of Exam 1 and Exam 2, to your local computer by running the following code:

```
df.to_csv('improved_dataset.csv', index = False, header=True)
from google.colab import files
files.download('improved_dataset.csv')
```

Confirm you can open it in Excel and you see the Average Exam attribute is there. Do not exit out of Excel. Submit this file for Lab 3 on Blazeview.

### 1.3 Part 3

Next, you will use Pandas to draw a boxplot of the average exam:

```
df.boxplot()
```

Create a new Word document and paste a screenshot to demonstrate you’ve created the boxplot. Also in the Word document, answer the following questions:

1. By looking only at the boxplots, what can you say about Exam 1 and Exam 2?
2. What have you learned in this lab? Please be specific as possible, write two complete paragraphs with a minimum of 50 words each.

**Due Date:** Before Midnight on Sunday, September 12th.  
Submit the Word document via Blazeview/Assignments/Lab 3.