

Advanced Database

Section

Python Project

Deadline 8/4/2021

All team Must attend the Discussion

Students who don't attend the discussion will be graded * Zero and no excuses will be accepted

Discussion will be held at *8/4/2021 12:30pm or 2:30pm

Project Requirements:

- 1) Choose topic and download an excel sheet for it (you can use [www.Kaggle.com](https://www.kaggle.com))
- 2) Must use pandas and matplotlib libraries
- 3) you must use these functions:
 - Head ()
 - Tail ()
 - Shape
 - Columns
 - Info ()
 - Describe ()
 - value_counts()
 - unique ()
- 4) Cleaning data (use drop & fill nulls)
- 5) Analysis: make any 4 analysis you want on your data but you must use
 - At least use 2 for loops
 - At least use 2 functions

6) Histogram:

- Must include a histogram for all your data set
- Must make a histogram for each analysis

7) Outliers + histogram of outlier

- You must clean at least 1 outlier in your data (if your data don't have an outlier, you must copy your data and add an outlier to any column as explained in the outlier video and then solve it)

8) Export the final updated dataset

9) make a documentation word file include the following:

- Cover Page
 - A brief description of your topic
 - Screenshot of your dataset and explain each column in it
 - A screenshot for each code cell and write below it an explanation of what it actually does.
- ,