



Palestine Launchpad Program

Data Analysis NanoDegree



Our agenda for today

- 1 Welcome MSG
- 2 Recap about the last session
- 3 Answer Your Questions on Google form
- 4 Project (Two) Rubric
- 5 Knowledge sharing
- 6 Q\A

Recap

What did we talk about last session

Last Session we covered the following:

Here's a breakdown of the key areas we covered:

- ✓ **Project (two) overview:** We gave an example of project two submission and made clear what is project two about.
- ✓ **New Concepts:** Big Data (ETL, and ELT)
- ✓ **Case Study:** Requests library, BeautifulSoup Examples
- ✓ **Q&A:** We answered your questions about the material covered in the session.

Q\A

Answers to your enquieres



1- After I load the html by api request I don't know how to extract all the data I need...


2- How can I download my datasets on workspace ? is it a csv file or another type?













<https://colab.research.google.com/drive/10Kv-KI9YIsT75pEReOmrGKp9PAPF2TWo?usp=sharing>

<https://drive.google.com/file/d/1qr70IHegK81lfvXe5W-nehekFh-i6rvZ/view?usp=sharing>

2- How can I download my datasets on workspace ? is it a csv file or another type?

 Jupyter Data_Wrangling_Project_Starter Last Checkpoint: 04/29/2024 (autosaved) 

File Edit View Insert Cell Kernel Help Not Trusted | Python 3 (ipykernel) 

          Markdown  

Real-world Data Wrangling

In this project, you will apply the skills you acquired in the course to gather and wrangle real-world data with two datasets of your choice.

You will retrieve and extract the data, assess the data programmatically and visually, accross elements of data quality and structure, and implement a cleaning strategy for the data. You will then store the updated data into your selected database/data store, combine the data, and answer a research question with the datasets.

Throughout the process, you are expected to:

1. Explain your decisions towards methods used for gathering, assessing, cleaning, storing, and answering the research question
2. Write code comments so your code is more readable

1. Gather data

In this section, you will extract data using two different data gathering methods and combine the data. Use at least two different types of data-gathering methods.

1.1. Problem Statement

2- How can I download my datasets on workspace ? is it a csv file or another type?

jupyter

Files Running Clusters

Select items to perform actions on them.

Upload New ↕

	Name ↓	Last Modified	File size
<input type="checkbox"/>	/		
<input type="checkbox"/>	Data_Wrangling_Project_Starter.ipynb	Running 2 months ago	14.9 kB
<input type="checkbox"/>	Test_Notebook.ipynb	2 months ago	5.24 kB

jupyter

Files Running Clusters

Select items to perform actions on them.

Upload New ↕

	Name ↓	Last Modified	File size
<input type="checkbox"/>	proj.xlsx		
<input type="checkbox"/>	Data_Wrangling_Project_Starter.ipynb	Running 2 months ago	14.9 kB
<input type="checkbox"/>	Test_Notebook.ipynb	2 months ago	5.24 kB

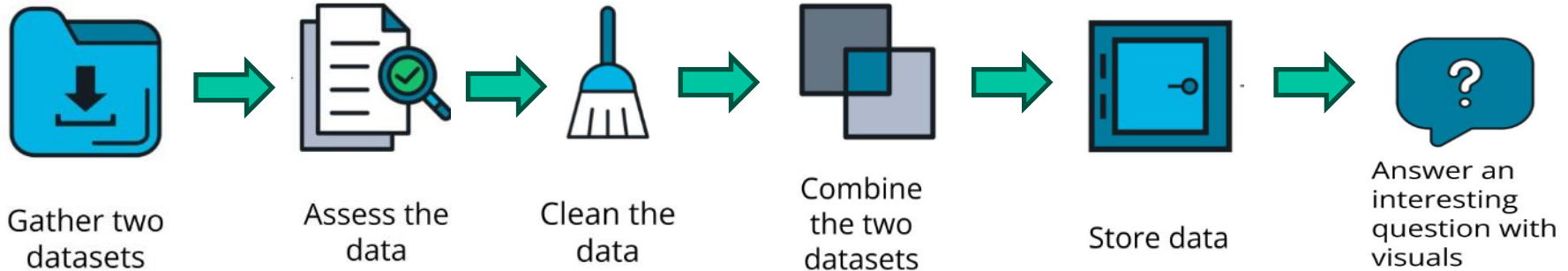
Project Two : Data Wrangling

What is needed to make a successful submit

Project: Real World Data Wrangling with Python



Project: Real World Data Wrangling with Python



[awraq-english](https://www.kaggle.com/awraq/english)

Project Rubric

Code Quality and Submission Phase



The student has uploaded a .zip folder containing their Jupyter Notebook for code review, and their datasets (file/link) for running the code.



The project shows thorough documentation of justification of wrangling decisions.



Does the code work?



Project Rubric

Gathering, Assessment, and Cleaning

✓ The project has a proper explanation of the problem statement.



✓ The student has gathered at least two separate datasets using two different data gathering methods.



✓ The student cleans the data issues they identified with the explanation and justifications.



✓ The student assesses the datasets for quality and tidiness.



✓ Remove unnecessary variables and combine datasets



Project Rubric

Data Storage and Answering the Research Question



Students must update their data store.



Students must identify next steps for the project.



Students will define and answer a short research question.



Case Study

→
Let's get things started...

Scrape Google Play

https://colab.research.google.com/drive/1bFppkgoAkrNN-IFOTrmViUAA8_oOLSrn?usp=sharing

New Concepts

→
Udacity, Students and Session Leads

<https://lookerstudio.google.com/>

Next Steps

→
Let's get things started...

Second Project July 2, 2024

May


June

July

August

September

Today
June 22, 2024

 **Second Project Due Date**
July 2, 2024
- Investigating a Dataset

Start working with the second project:

1. Finish cleaning
2. Writing QA



Program Period
May 9, 2024 - September 14, 2024

Thank You!

→
And Good Luck!