UDACITY Google spark

Palestine Launchpad Program

Data Analysis NanoDegree

Our agenda for today

1. Recap about the last session
2. Answer Your Questions on Google form
3. Project (one) overview
4. New Concepts
5. Activity (Study Case)
6. Q\A

Palestine Launchpad
منصة فلسطين التقنية
Google

Recap

What did we talk about last session

UDACITY

Last Session we covered the following:

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

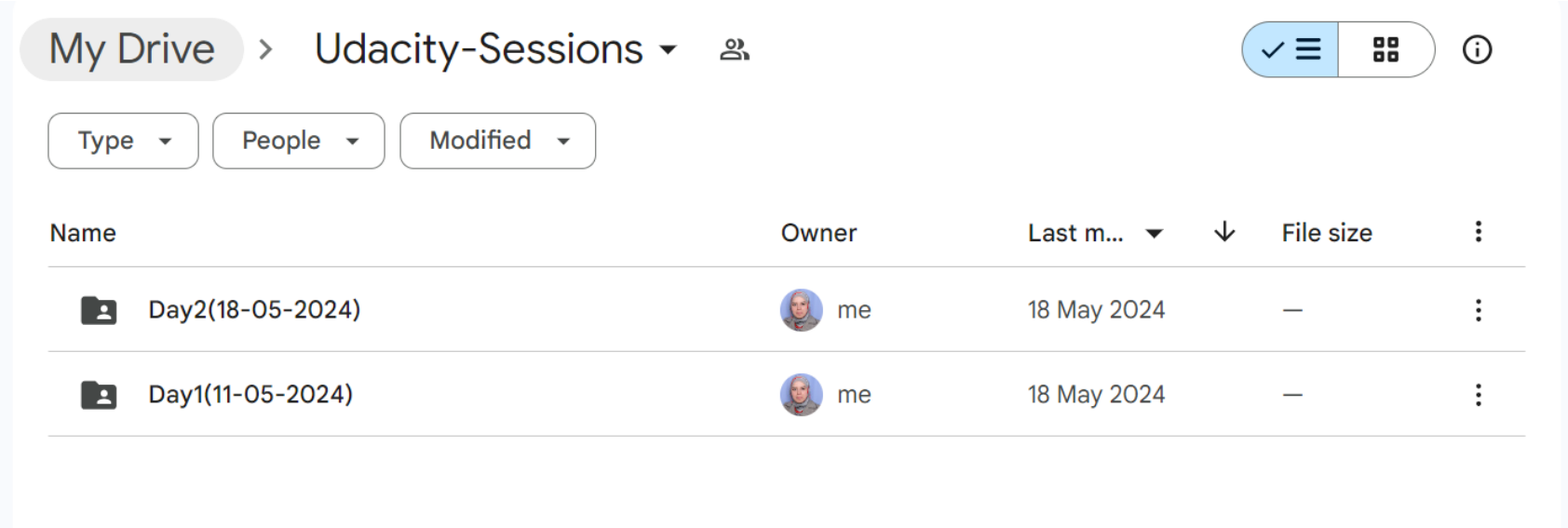
- ✓ **Project (one) overview:** We discussed an overview of the first project. This likely included the project datasets, rubric, and timelines.
- ✓ **Weekly Schedule:** We reviewed the weekly schedule for the class. This may have included upcoming lessons.
- ✓ **Activity (Study Case):** We worked on a study case activity. Exploring a dataset from students' marks.
- ✓ **Writing Report:** We discussed how to write well-documented report.
- ✓ **Q&A:** We answered your questions about the material covered in the session.

Last Session we covered the following:

We also agreed on start working on the first project (investigating a dataset). Here are the next steps you should have taken:

- ❑ **Select a dataset for the first project.**
- ❑ **Gain an understanding of the context of the dataset.**
- ❑ **Clean the data by removing any errors or inconsistencies by following up the steps taken on the case study.**
- ❑ **Document your steps by writing an introduction and a data cleaning section for your project report.**
- ❑ **Begin brainstorming your research questions.**

Material on Google Drive



https://drive.google.com/drive/folders/1cGl979d86JQF2jBZzh6a9mFpXVuLkZC5?usp=sharing

Palestine Launchpad
منصة فلسطين التقنية
Google

Q\A

Answers to your enquieres

UDACITY

I recived the following:

When encountering invalid values in columns, such as negative values in the 'Age' column like -1, what should be the appropriate action? Should I drop these records? (**Answer on Community)**
i didn't understand pandas explode and small stress about start the project, otherwise everything good
its just not easy to me because my background is in programming not so much. because i am working in computer networks and information security
simple question: can we use merge always instead of concatenation?? **(Answer on Community)**
I'm completely clueless when it comes to anything related to Lambda, I can't understand the syntax of it, also, its a bit costing me sometime to get used to the syntax, not sure if that's ok, can we have some resources to practice the usage of functions ? also, I strongly believe that the part of communicating data should've been before the manipulating data using pandas and numpys, because it explained how the functions work. **(Answer on Community)**

Topics to cover

apply_method.ipynb - Colab (google.com)

Variance
should i take lessons in python to be able to continue in this course
more details about project and submission
do you recommend us to learn Power BI??? if yes, when and why?
Lambda concept + syntax
any additional material of practice? do you recommend me checking hackerrank or leetcode to understand Pandas? Learn Pandas | Kaggle

Did you Choose Dataset?

I haven't decided yet, because i was focus on finish the plan for this week, but I took a quick look at the dataset so maybe i will choose the movies datadset
i am not sure yet
I want to read more about them and not rush into making a decision without understanding each dataset. But within the next couple of days, I will have chosen the dataset I want.
I prefer to start the project only when I'm confident I fully understand the topics.
I don’t have enough time because I have a job through Sunday to Thursday so unfortunately, I don’t have enough time. So usually, I look into the course on Friday and Saturday.
Didn't get to it yet, planning to start working on it tomorrow Friday entirely

Palestine Launchpad
منصة فلسطين التقنية
Google

Project one : Invistigating a dataset

What is needed to make a successful submit

UDACITY

Project Rubric

Data Wrangling Phase

Criteria	Submission Requirements
Is the data cleaning well documented?	The project documents any changes that were made to clean the data, such as merging multiple files, handling missing values, etc.

Project Rubric

Exploration Phase

Criteria	Submission Requirements
Is the data explored in many ways?	• The project investigates the stated question(s) from multiple angles. • The project explores at least three variables in relation to the primary question. This can be an exploratory relationship between three variables of interest, or looking at how two independent variables relate to a single dependent variable of interest. • The project performs both single-variable (1d) and multiple-variable (2d) explorations.
Are there a variety of relevant visualizations and statistical summaries?	• The project's visualizations are varied and show multiple comparisons and trends. • At least two kinds of plots should be created as part of the explorations. • Relevant statistics are computed throughout the analysis when an inference is made about the data.

Project Rubric

Conclusions Phase

Criteria	Submission Requirements
Has the student correctly communicated tentativeness of findings?	• The Conclusions have reflected on the steps taken during the data exploration. • The Conclusions have summarized the main findings in relation to the question(s) provided at the beginning of the analysis accurately. • The project has pointed out where additional research can be done or where additional information could be useful. • The conclusion should have at least 1 limitation explained clearly. • The analysis does not state or imply that one change causes another based solely on a correlation.

Project Rubric

Conclusions Phase

Criteria	Submission Requirements
Has the student correctly communicated tentativeness of findings?	• The Conclusions have reflected on the steps taken during the data exploration. • The Conclusions have summarized the main findings in relation to the question(s) provided at the beginning of the analysis accurately. • The project has pointed out where additional research can be done or where additional information could be useful. • The conclusion should have at least 1 limitation explained clearly. • The analysis does not state or imply that one change causes another based solely on a correlation.

Project Rubric

Communication

Criteria	Submission Requirements
Is the flow of the analysis easy to follow?	• The code should have ideally the following sections: Introduction; Questions; Data Wrangling; Exploratory Data Analysis; Conclusions, Limitation. • Reasoning is provided for each analysis decision, plot, and statistical summary. • Interpretation of plots and application of statistical tests should be correct and without error. • Comments are used within the code cells. • Documented the flow of analysis in the mark-down cells.
Is the data visualized using appropriate plots and parameter choices?	Visualizations made in the project depict the data in an appropriate manner (i.e., has appropriate labels, scale, legends, and plot type) that allows plots to be readily interpreted.

Palestine Launchpad
منصة فلسطين التقنية
Google

Case Study

Put all thingd togther

UDACITY

Palestine Launchpad
منصة فلسطين التقنية
Google

QA

Udacity, Students and Session Leads

UDACITY

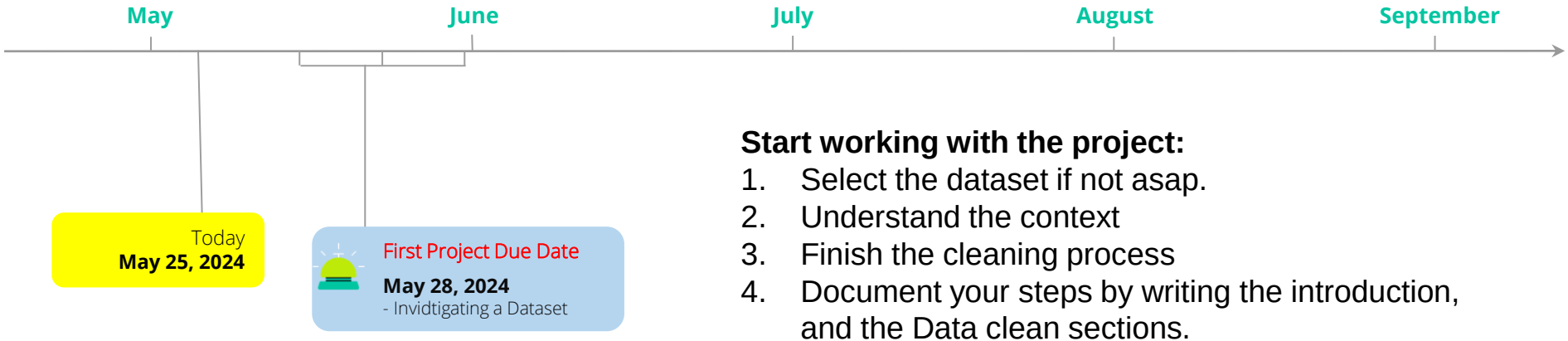
Palestine Launchpad
منصة فلسطين التقنية
Google

Next Steps

Let's get things started...

UDACITY

First Project May 28, 2024



Start working with the project:

1. Select the dataset if not asap.
2. Understand the context
3. Finish the cleaning process
4. Document your steps by writing the introduction, and the Data clean sections.
5. Write the research questions.
6. Draw visualization
7. Write the conclusion
8. **Submit the project on Time.**

Palestine Launchpad
منصة فلسطين التقنية
Google

Thank You!

And Good Luck!

UDACITY