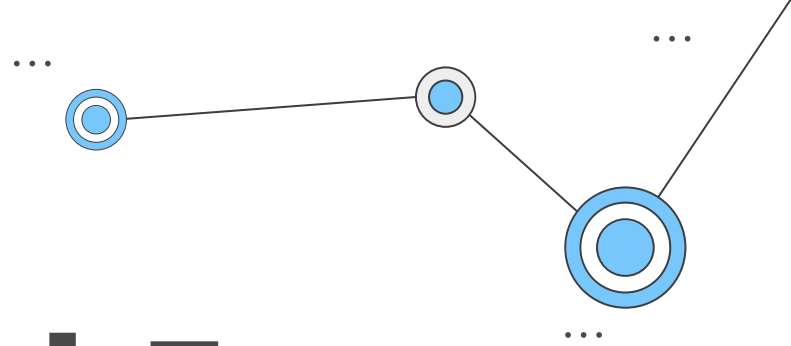


MAISI



CRA Week 7:

Subgroup Work Session

#1 – Brainstorming!

Michigan Data Science Team
Fall 2025

Session 7 Agenda

01

Fun Icebreaker!!

Get to know your projectmates!

...

02

Final Expo Overview

What is our final deliverable going to look like?

...

03

Your Next Steps

Applying what you've learned independently of modules.

...

04

Subgroup Formation

Group up with others who have the same research ideas.

...

05

Brainstorming Time!

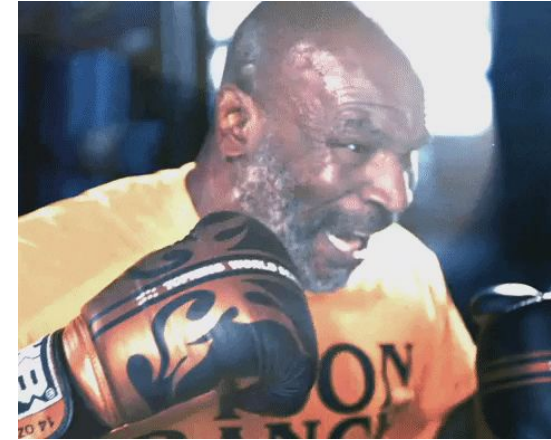
Get the ball rolling on your final presentation for the project.

...

Quick Icebreaker!!

Share with the people around you :)

If you were a boxer/professional athlete, what would your theme song be?





Final Expo Presentation Logistics

- Each semester, MDST hosts a Project Expo for all our teams and their members to show off what they have learned over the past two months!
 - This event hosted in two different sessions and is typically attended by other MDST members, DS enthusiasts on campus, professors, and sometimes even our corporate sponsors!
- General Logistical Notes
 - When: Friday, November 21st (likely around 6:00 PM)
 - Where: CCCB
 - What: Our final presentations!

...



Steps for Project Presentation

Hypothesis: Define the hypothesis or question you are aiming to answer with the COMPAS data. Feel free to use past workbooks as inspiration!

Research Questions: Clearly state some research questions that will guide your analysis. To create a strong research question include the *who*, *what*, *when*, *where*, *why*, and *how*

Tools and Techniques: Mention the data analysis tools your group will be using to solve the research questions. (pandas, confusion matrices, logistic regression, etc.)

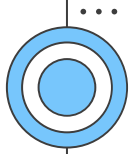


Sample Ideas

- T1: How does COMPAS score middle-aged defendants when compared to (otherwise similar) older/younger defendants?
- T2: Is there a disparity in how COMPAS scores women vs. men that is not reflected in actual recidivism rates?
- T3: Does COMPAS give White defendants with prior convictions high scores at the same rate as Black defendants w/priors?
- T4: To what extent do the data suggest that the COMPAS algorithm has an issue with racial profiling?

...

(If you want to use any of these as your topic, please type “T1”, “T2”, “T3”, or “T4” in the Google Sheets column asking for your choice. We encourage you to be creative!)

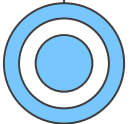


Getting into Subgroups (2-4 People)

- If you have already formed a group by filling out the Google Sheet, get together and start working on your presentation following the steps outlined in the previous slide.
- If you aren't quite sure of what you want to do or who you want to work with, talk with the people around you and find others who have a similar research interest to you!
 - Everyone needs to be in a group! (no solo presentations)
- **When you have your group, come talk to us!!**



...



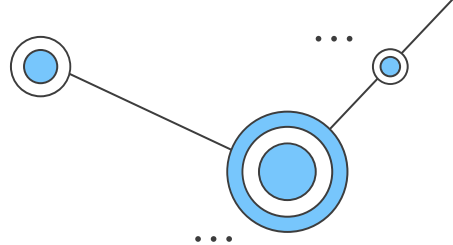
...



Group Work Time!

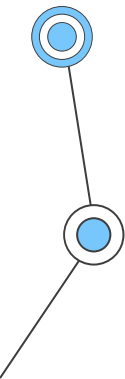
Let's break into our teams and start
researching topics to present on!

Hands-On Data Science!! :0



Next Steps:

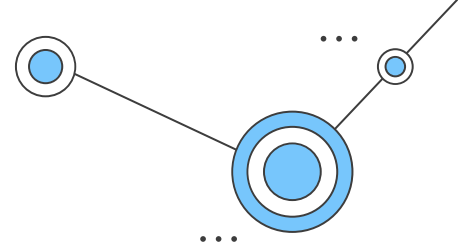
1. Get in your groups and create a shared Google Drive for files (notebooks, slides, Docs, etc.)
2. **Fill out the Google Sheet!!** [Link](#)
3. Create a blank notebook and research a topic of your choice!
4. Ask us if you need any inspiration or help!



[Pandas Cheat Sheet](#)

[Seaborn Cheat Sheet](#)

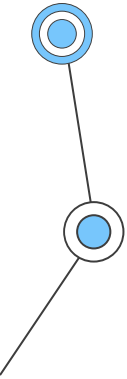
Reminders



- Don't share colab notebooks with teammates if you are working at the same time
- Where to put csv and data files
 - Google Drive
 - Need to include:

```
from google.colab import drive
drive.mount('/content/drive')

pd.read_csv('/content/drive/MyDrive/[FILE NAME]')
```
 - Colab Files
 - See next slides



Reminders

Click on the folder in the sidebar



Q Commands + Code + Text ▶ Run all

Files

- ..
- sample_data

Week 1 - Pandas Practice

Here is where you import the libraries necessary to perform the following tasks!

```
import pandas as pd
import seaborn as sns

# Allows you to provide a path to a Google Drive address rather than a local file path
from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

Load the Google Forms .csv into a Pandas dataframe.

```
df = pd.read_csv('/content/MDST Week 1 - Pandas Practice.csv')
```

Print out the .head() and the datatypes.

```
df.head()
```

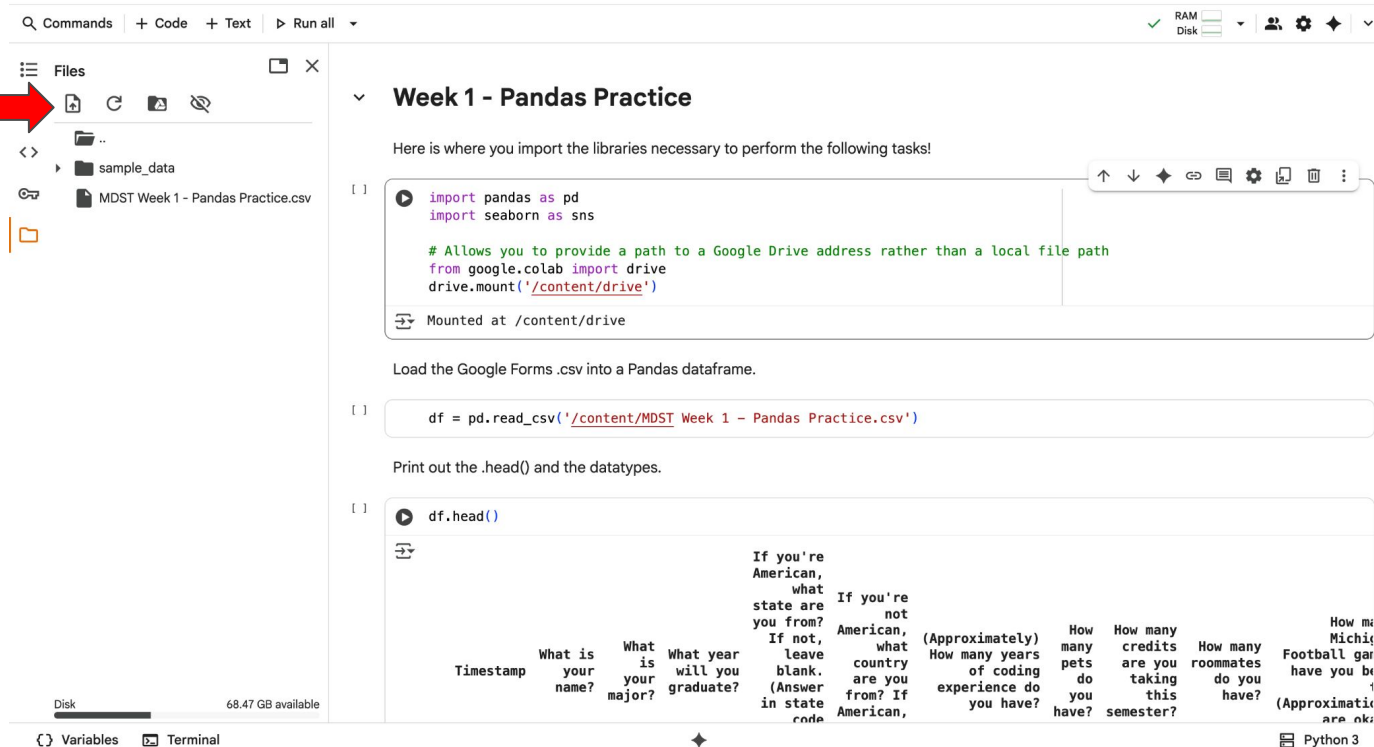

Timestamp	What is your name?	What is your major?	What year will you graduate?	If not, leave blank. (Answer in state code)	If you're American, what state are you from?	If you're not American, what country are you from? If American, (Approximately) How many years of coding experience do you have?	How many pets do you have?	How many credits are you taking this semester?	How many roommates do you have?	How many Football games have you been to?	How many Michigan are you from?

Disk 68.47 GB available

Variables Terminal Python 3

Reminders

Click the upload button and select the file you want to upload



The screenshot shows the Google Colab interface. On the left, the 'Files' pane shows a folder named 'sample_data' and a file named 'MDST Week 1 - Pandas Practice.csv'. A red arrow points to the upload button (a square with a plus sign) in the file explorer. The main area displays a Jupyter Notebook titled 'Week 1 - Pandas Practice'. The notebook contains the following code:

```
import pandas as pd
import seaborn as sns

# Allows you to provide a path to a Google Drive address rather than a local file path
from google.colab import drive
drive.mount('/content/drive')
```

Below the code, it says 'Mounted at /content/drive'. The next cell contains the following code:

```
df = pd.read_csv('/content/MDST Week 1 - Pandas Practice.csv')
```

Below the code, it says 'Load the Google Forms .csv into a Pandas dataframe.' The next cell contains the following code:

```
df.head()
```

Below the code, it shows the output of the `df.head()` command, which is a preview of the data from the CSV file. The output is a table with columns: Timestamp, What is your name?, What is your major?, What year will you graduate?, If you're American, what state are you from?, If you're not American, what country are you from?, (Approximately) How many years of coding experience do you have?, How many pets do you have?, How many credits are you taking this semester?, How many roommates do you have?, and How many Michigan Football games have you been to? (Approximately).

Reminders

Click the three dots and copy the path. Put this in your read function



The screenshot shows a Google Colab environment. On the left, the 'Files' pane displays a directory structure with 'sample_data' and 'MDST'. A context menu is open for the 'MDST' folder, with 'Copy path' highlighted. A red arrow points to this option. The main code editor shows the following code:

```
[ ] import pandas as pd
import seaborn as sns

# Allows you to provide a path to a Google Drive address rather than a local file path
from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive

Load the Google Forms .csv into a Pandas dataframe.

[ ] df = pd.read_csv('/content/MDST Week 1 - Pandas Practice.csv')

Print out the .head() and the datatypes.

[ ] df.head()
```

The output of `df.head()` is displayed as a table with 10 columns and 1 row of data. The columns are: Timestamp, What is your name?, What is your major?, What year will you graduate?, If not, leave blank. (Answer in state code), If you're American, what state are you from?, If you're not American, what country are you from?, (Approximately) How many years of coding experience do you have?, How many pets do you have?, How many credits are you taking this semester?, How many roommates do you have?, and How many Michigan Football games have you been to? (Approximately).

At the bottom of the interface, there are tabs for 'Variables' and 'Terminal', and a status bar indicating 'Python 3' and '68.47 GB available'.