

Week 2: Exploratory Data Analysis

1. KEY QUESTIONS

- What is exploratory data analysis (EDA)?
- Why is EDA important?
- What are different types of data?
- What are the ways in which you might need to clean data?
- What EDA steps should you take in order to formulate a data-driven question and hypothesis?

2. BRIEF AGENDA

- Introductions (again)
- Announcements
- Article Group Discussion
- Step-by-step Demo
- Break
- Flex time for EDA

3. TIMELINE

- **5:50pm - 15 minutes:** Let folks trickle into class
- **6:05pm - 5 minutes:** Introductions
Sejal oversees, Sook-Hee takes attendance
- **6:10pm - 15 minutes:** Article Group Discussion
Sejal calls on people to share, Sook-Hee oversees
- **6:25pm - 10 minutes:** What is Exploratory Data Analysis?
Sook-Hee leads, Sejal oversees
 - Garbage in --> Garbage out
 - Data Cleaning: duplicate data, missing data, inaccurate data, incorrect data
 - Identifying Variables (brainstorm with life expectancy data)
 - Finding Correlations (heatmap)
- **6:35pm - 5 minutes:** Why is Exploratory Data Analysis important?
Sejal leads, Sook-Hee oversees
 - Data is inherently messy
 - Identify errors with the data
 - Begin to understand patterns
 - Find interesting relations
- **6:40pm - 5 minutes:** Data Types
Sejal leads, Sook-Hee oversees
 - Brainstorm: textual, wearable, geographic, temporal, spatial, qualitative, quantitative, emotive
 - Language-based data types: str, int, float, bool

- **6:45pm - 10 minutes:** Pandas Crash Course
Sook-Hee leads, Sejal oversees
- **6:55pm - 25 minutes:** BREAK
Keep Zoom on; Circle around to chat with people in person!
- **7:20pm - 10 minutes:** Small Group Discussion
Sejal leads, Sook-Hee oversees
 - What associations do you think of when you hear the word "data"?
 - Go around in a circle and rattle off 1-2 word associations until you run out of ideas
 - Jot them down as best as you can. Then share some of your best ones with the class
- **7:30pm - 25 minutes:** EDA Demo (using life expectancy data)
Sook-Hee leads, Sejal oversees
- **7:55pm - 5 minutes:** Explain Homework
Sejal leads, Sook-Hee oversees
- **8:00pm - end:** Flex Time
Sejal and Sook-Hee both circulate