

Sports Data Analysis

Week 3 – Sit with your project groups!!!



TABLE OF CONTENTS

01

Slack

02

**Group
Solidification**

03


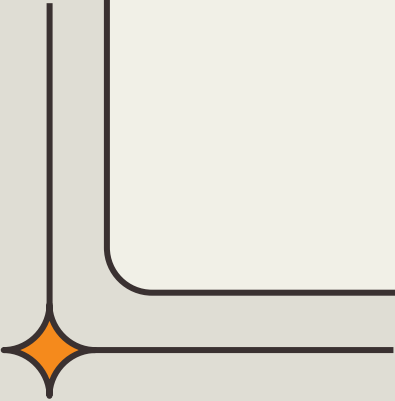
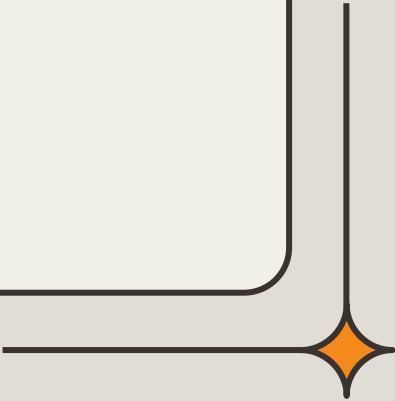
Datasets

04

Data Cleaning



Slack

- You all should have gotten an **email invitation** to join the MDST workspace
 - **Download** the Slack app
 - Invitations will be sent out for our project's channel
- 
- 
- 



Group Solidification

Go to this link: <https://forms.gle/T28Wb7qhFjVb81yo6>

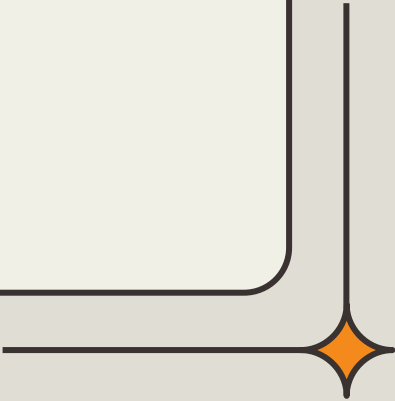
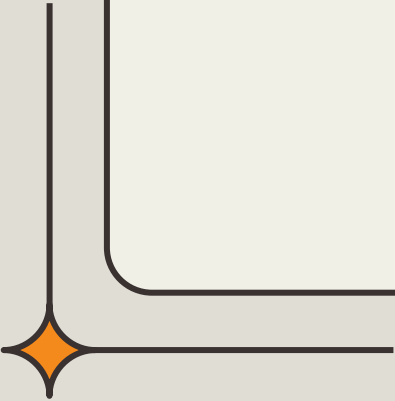
Only one person from each group has to fill it out (and you must do it even if you want to work alone)






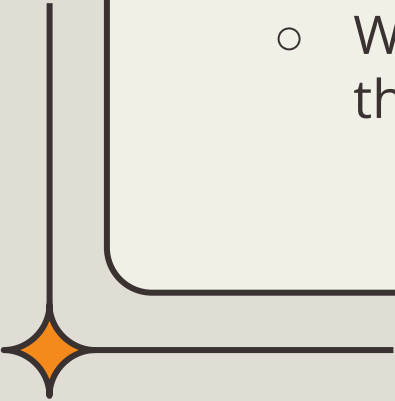
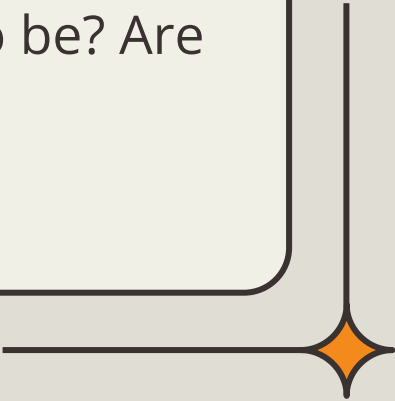
Data Collection

Do we need more time for this?



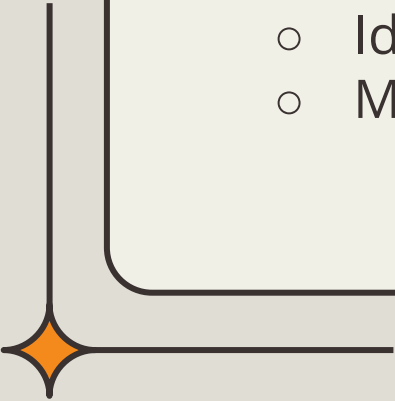
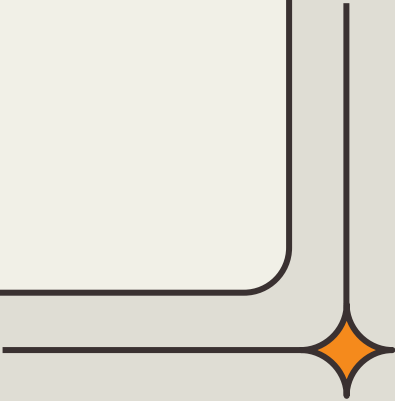


Data Cleaning

- A lot of data comes unclean and requires a little work to ensure **accuracy** and **understandability**
 - Think about:
 - What do you want to ultimately analyze/model?
 - Do you need to create more data outside of what you have (i.e. advanced stats, classes)
 - What data types do you expect each column to be? Are they actually like that?
- 
- 
- 



Data Cleaning

- Things that may need to be done:
 - Changing column names
 - Fixing data types
 - Identifying (and maybe filling) null values
 - Adding/dropping columns
 - Dropping rows (i.e. a header row got duplicated)
 - Identifying (and maybe removing) outliers
 - Merging/splitting tables
- 
- 





Data Cleaning

Download the file at:

<https://github.com/MichiganDataScienceTeam/F25-SportsDataAnalysis/blob/main/Starter%20Code/cleaning.ipynb>

Work on it with your group! I highly encourage you all to do it on your own computers (it will obviously be the same code within groups) to **build confidence** in it



Questions?



CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)

Please keep this slide for attribution

