

Project 1 Collaboration Worksheet

Group Members

Are you in Lasso or Ridge?

Ridge

Who is in your group?

Jesus (Aaron) Barela
Amy Butler
Ko-Fan (Katie) Huang
William Moore

The Project

What are the goals of your analysis? Be as specific as possible (Hint: what WTWY's goals?)

Advise WTWY as helpfully as possible where to place their limited resources given their goals to (1) fill the gala event space and (2) find people who are interested in what WTWY is doing.

Ok, now look at those goals and take your thoughts a step further. Can you make those goals even more detailed?

(1) Filling the gala event space:

- Identify stations in high income areas with high volume of exits in the evening commute hours.

(2) Targeting people interested in WTWY:

- Professional interests of businesses interested in women in technology
- Academic interests of education institutions interested in women in technology
- Identify stations near tech companies and universities with high volume of exits in the morning commute hours.

How will you address those goals? Write down the specific steps and/or analytical tools you will use for each goal.

1. Determine weeks and years to use from MTA NY subway data source
2. Obtain data dictionary explaining MTA data fields

3. Explore and clean data using Python packages pandas and numpy
4. Use Python to conduct basic statistical analysis including calculating subway station exit volumes
5. Visualize data using Python packages matplotlib and seaborn

Brainstorm what else you could do for this project. One example might be adding another data source, but what else can you come up with? What additional steps would enhance your analysis?

This website has several possible sources for data enrichment: <https://data.world/datasets/mta>
For example, we'd like to include a map of the subway lines and stations in our presentation.

We'd like to associate household income ranges with subway station location. We will explore possible sources of this data.

We might use DataStudio to create visualizations.

What are steps you will take while you go through this project? What is your process? For example, one step would be to make a presentation. What might come before and after that?

We will follow the Project process steps outlined at
https://github.com/thisismetis/onl_ds5/blob/main/curriculum/project-01/project-01-introduction/project_01.md.

Setting up Group Work

How often will your group check in? Set a schedule.

We will check in during the designated project work time each day and meet after class in the evenings if needed. We will schedule any additional meetings in ET.

How will your group check in?

We will check in using Zoom. The Metis Zoom if meeting during class hours or from 7:30pm to 8:30pm ET. Our personal Zoom if otherwise.

We created a DM Slack Channel for our team members to discuss this project.

How will you share what you're working on? (e.g. through a GitHub repo, verbally through check-ins, etc)

We created a Shared GitHub repo for code and presentation:
https://github.com/abarela13/onl_ds5_project_1

We will discuss what we're each working on during our check-in team meetings.

Look above at what you designated as the steps of the project. Who will be responsible for which steps or tasks? How will you divide up your work?

Don't neglect the communication part of this project: think about how each part of your work with data will flow into a presentation and who will be doing which part of the presenting.

Task	Team Member
Create GitHub repository	Aaron
Obtain MTA data for last two weeks in May and first two weeks in June for both 2018 and 2019 and save files to GitHub repository	Aaron
Find and share MTA data dictionary	Katie
Explore availability of income data	Amy
Obtain supplementary MTA data	Aaron
Identify stations near tech businesses and universities	Aaron and Will
Review MTA data to understand fields, contents, and identify any cleaning needed	ALL
Compare findings of reviewing MTA data	ALL
Decide on basic statistics to include with results and visualizations to create	ALL
Divide presentation sections responsibilities and prepare slides for own section	ALL
<i>Tentative:</i> Create visualization(s) in DataStudio	Aaron

When should each task be completed?

MON-TUES: Data exploration, cleaning, discussing our notes and ideas for analysis and visualizations.

WED: Conduct analysis and create visualizations.

THURS: Create and compile presentation slides. Practice presentation.

FRI: Practice presentation in the morning.

Establish group norms. Are there any ground rules you should all follow? How will you fruitfully address any conflict that might arise?

Same rules as [Bootcamp for working with one another](#).

A Few Tips

You probably want to start off by looking at the project starter for this project. It will help you to clean your data and start EDA.

If you have a favorite slide template, definitely use that, but if you don't, check out Slide Carnival. They have lots of professional templates.

Think about what holes someone could poke in your analysis. You don't necessarily have to address every potential issue in your work, but you should be aware of potential drawbacks and should be ready to discuss them!