ABSTRACT: An organization wants to hand out flyers at subway stations. In order to maximize their time, we want to find the busiest subway stations, using MTA raw data of cumulative turnstile entries and exits. Use SQL to import the MTA data into python. Use python to clean the data, convert cumulative to differences, and find the busiest stations at specific times. Use matplotlib to plot the busiest stations. Create a PowerPoint presentation to let the organization know which stations would be best to attend to attempt to hand out fliers.

DESIGN
WTWY wants to hand out flyers that will attract people to their gala. I want to focus on getting busy stations so there's a maximum chance of getting some people who are interested. I also want locals rather than tourists, so I'll focus on the stations that are busy during rush hours. I'm also breaking it out by entries and exits, so I'll only find the top 20 for each of the four options (am rush hour entries, am rush hour exits, pm rush hour entries, pm rush hour exits).

DATA
The data is the MTA's raw data of cumulative turnstiles' entries and exits. I'm looking at data starting the first week of April through September, 2021, for the most up to date results. I'm breaking out by time in order to get rush hours.

ALGORITHMS
Get the data into Python. Look for duplicates. Create columns that contain time and days of the week. Calculate number of entries over a specific time from cumulative counts Sort data into new dataframes, one for the morning, one for the evening. Count the number of entries from each station, and sort by number of entries to find the busiest; repeat for exits.

TOOLS
SQLalchemy is used to import the data from the MTA. Pandas is used to clean the data of duplicates. Pandas is used to separate the data by station and time. Matplotlib and Seaborn is used to display results.

COMMUNICATION
I will use a PowerPoint presentation to describe the problem and present the data. I will include graphs for the organization, and a summary slide that recommends the 20 most promising stations. Additional slides will have more detail if the clients have questions and there is time