SF BART Daily Exits

SkillBuilder

This SkillBuilder uses data about daily ridership on San Francisco's regional rapid transit system, BART, for the year 2021. This data can be found in the **DailyExits** table in the SkillBuilder workbook.

Part 1 - Engineering Data Columns

To the right of the DailyExits table are four blank columns, highlighted in yellow. Use your cleaning skills to engineer the required features.

1. In the Month column (Column I), extract the month from each date (Column E).
2. In the DayOfWeek column (Column J), extract the day of the week from each date (Column E).
3. In the YellowLine column (Column K), extract whether or not a station is on the Yellow Line: "Yes" if it is on the Yellow Line, and "No" if it is not.
    - Use the Lines column (Column G) to find this information. Based on the data dictionary, the lines a station is on is given as an abbreviation of the first letter of each line color in a set order. For example, "YO" means a station is on the Yellow and Orange lines, while "OG" means a station is on the Orange and Green lines.
    - Fortunately, Yellow will always be first, so all we need to do is check if the first letter of the Lines string is a "Y".
    - Engineering the column will require a formula that nests two functions inside one another: an inner function to extract the first letter of the Lines string, and an outer function to check if it is a "Y" and decide on the appropriate output ("Yes" or "No").
4. In the Downtown column, extract whether or not a station is one of the four stations in downtown San Francisco: EM, MT, PL, CC: "Yes" if it is one of those stations, "No" if it is not.
    - You don't need to know the station names for this one, though they are provided in the StationCodes tab of the workbook for reference.
    - What *is* important is that you choose a function that lets you check for multiple logical conditions. One strategy is to check if the station code matches each of the four downtown stations, one by one. Don't forget to include a way to output "No" if a station is not a downtown station!

Part 2 - Summarize the Data

Use the engineered columns to answer the following questions.

5. What is the total number of trips made on the BART for all of 2021?
    - This doesn't actually require an engineered column. But it's a good baseline to compare the next few questions against!
6. How many trips were made in May?
7. How many trips were made during a weekday?
8. How many trips ended on a station on the Yellow line?
9. What proportion of Yellow Line trips ended on a downtown station?
    - This is the number of downtown station exits divided by the number of all trips on the Yellow Line (Question 8).
10. What is the average daily ridership for rides that end at a downtown station?
11. What is the average daily ridership for rides on weekends that end at a downtown station?
    - You should use a function that can select on multiple criteria at the same time.