# Cliff Bar NYC Subway Sampling Opportunity Analysis

By: Jenni Hawk

# Abstract

This mock project for Cliff Bars was designed to showcase my exploratory data analysis knowledge, technical knowledge, and strength in storytelling.

With the push in NYC to physically go back to work, Cliff Bar wants to be the top-of-mind brand in NYC associated with giving people the energy they need to get back to normal.

Product sampling has been planned:
- Street teams will be deployed at MTA Stations
- Cliff Bar samples will be distributed with a coupon to be used at Duane Reade

Key Questions to answer: Where + When
- What stations have the most traffic?
- On what days are most people using MTA?
- Where do Duane Read locations fall within top MTA stations?
    - Are they within walking distance?
- How does the reporting period compare against a pre-Covid timeframe?

# Design

**Opportunity Analysis Window:** 1/29/22 - 3/19/22
Rationale: Look at data when Omicron is at it its lowest level.

**Pre-Covid vs emerging Post Covid Analysis Window:** Feb 2020 vs Feb 2022
Rationale: February 2020 No Covid.  February 2022 Omicron at lowest point.

# Data

Utilized the following data sources:

NYC MTA turnstile data http://web.mta.info/developers/turnstile.html
NYC MTA Latitude / Longitude: https://data.ny.gov/widgets/i9wp-a4ja

Duane Read location data: https://www.scrapehero.com/location-reports/Duane%20Reade%20Pharmacy-USA/
CD Covid Case Tracker: https://covid.cdc.gov/covid-data-tracker/#trends_dailycases
CDC Covid Timeline: https://www.cdc.gov/museum/timeline/covid19.html

# Technical Methodology

Three technical workflows were established:

1. MTA Station Volume + Highest Commuter Days
    - Read in MTA csv data via Pandas and SQLAlchemy
    - Utilized Pandas to manipulate data and Matplotlib to visualize data

2. Duane Reade locations filtered by top MTA locations
    - Read in Duane Reed location data into Pandas (contains zip code and lat/long data)
    - Removed non-relevant columns to make it simpler to upload to Google Maps for location plotting
    - Filtered Duane Reed data by the zip codes of the top MTA stations
    - Exported .csv that contained lat/long for all Duane Reades within top MTA footprint
    - Uploaded .csv to Google Maps for plotting

3. Pre-Covid vs Post Covid
- Read in MTA csv data via Pandas and SQLAlchemy
- Utilized Pandas to manipulate data and Matplotlib to visualize data

# Tools

- Pandas: connect to local database via SQLAlchemy, data manipulation, time series / date functionality
- Matplotlib: data visualization

# Communication

Project will be communicated on Github for technical audience and Slideshare for non-technical audience.