

BUDT 730

Team Assignment 1

Amal Byju

Sreyas Sourav

Tanvi Mehra

Question 1: Bay Wheels's Trip Data

PRELIMINARY STEPS

The Bay wheels's CSV file was loaded into Tableau. A calculated field called Trip Duration was added. The trips with duration greater than 45 mins (2700 seconds) were filtered out.

a) HISTOGRAM FOR TRIP DURATION

The figure below shows the histogram for Trip Duration. The data is right skewed. In order to create the histogram, bins of size 120 were created and added to the columns shelf and count of trip duration was placed in the rows shelf.

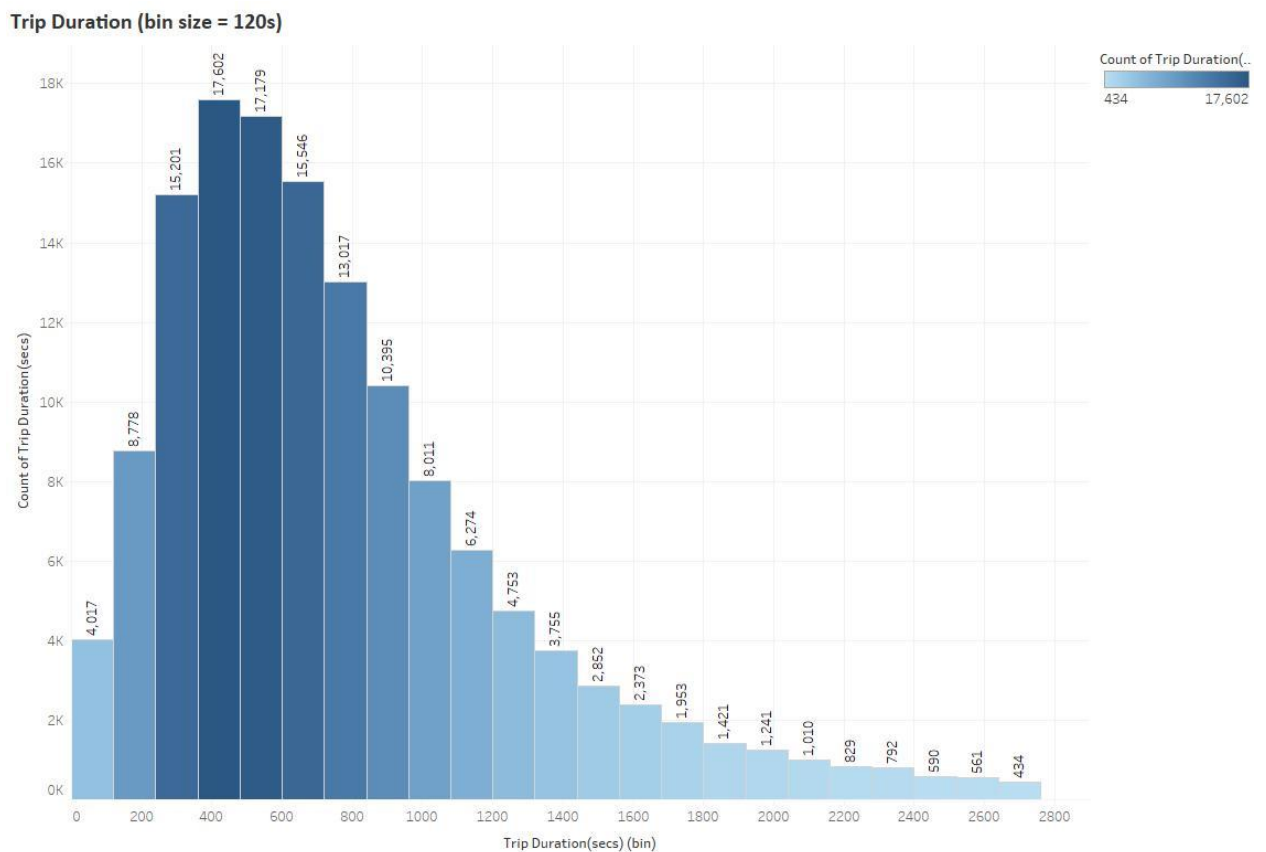


Figure a: Histogram of TripDuration. (Bin size = 120 s)

b) MAP FOR STARTING STATIONS

A map was created with the starting latitudes and longitudes for each sub region in the San Francisco Bay Area . Latitudes and longitudes were converted to dimensions in order to assign a geographic role to them. The map intuitively shows the average trip duration (secs) which was calculated by changing the measure to AVG on the Trip Duration.

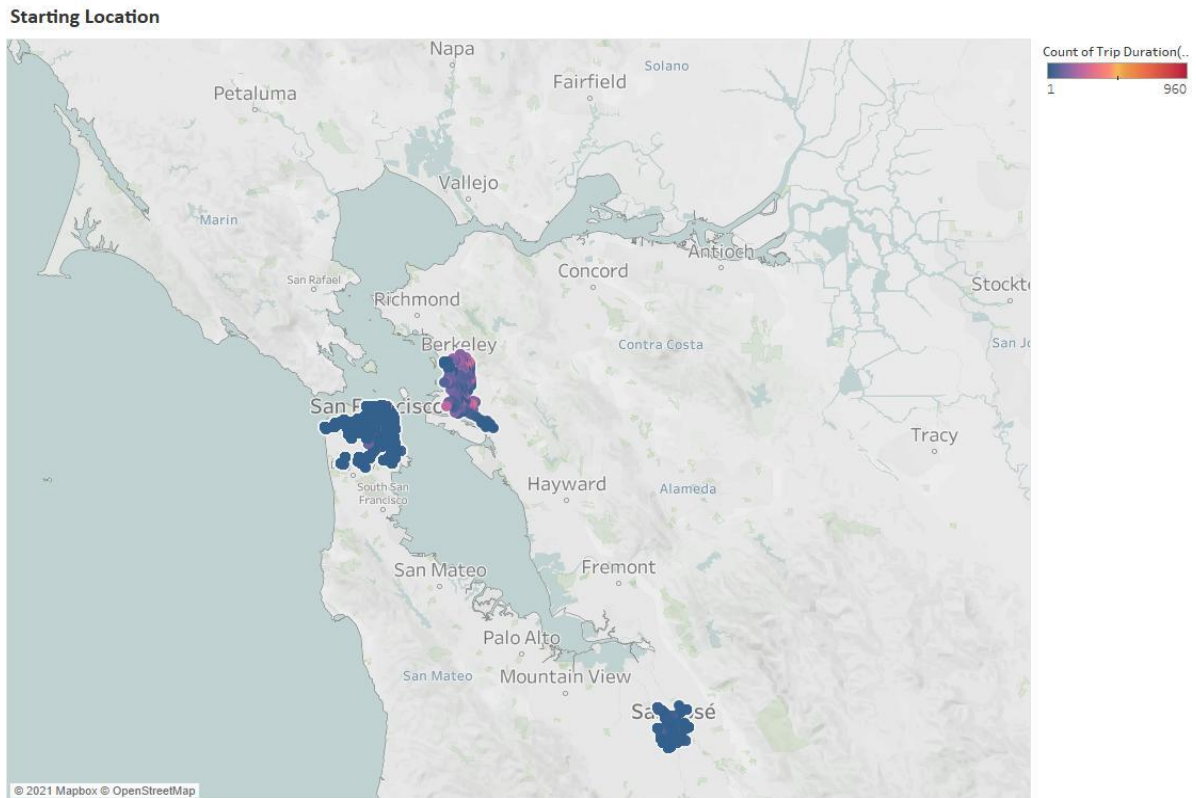


Figure b: Map for Starting Location

c) STATIONS IN SAN FRANCISCO COUNTY

The map below shows the stations in San Francisco county. A group was created with sub regions (San Francisco County, South Bay, East Bay) by applying a group on the Start Station ID (Dimension). Each station was then mapped to the sub region accordingly.

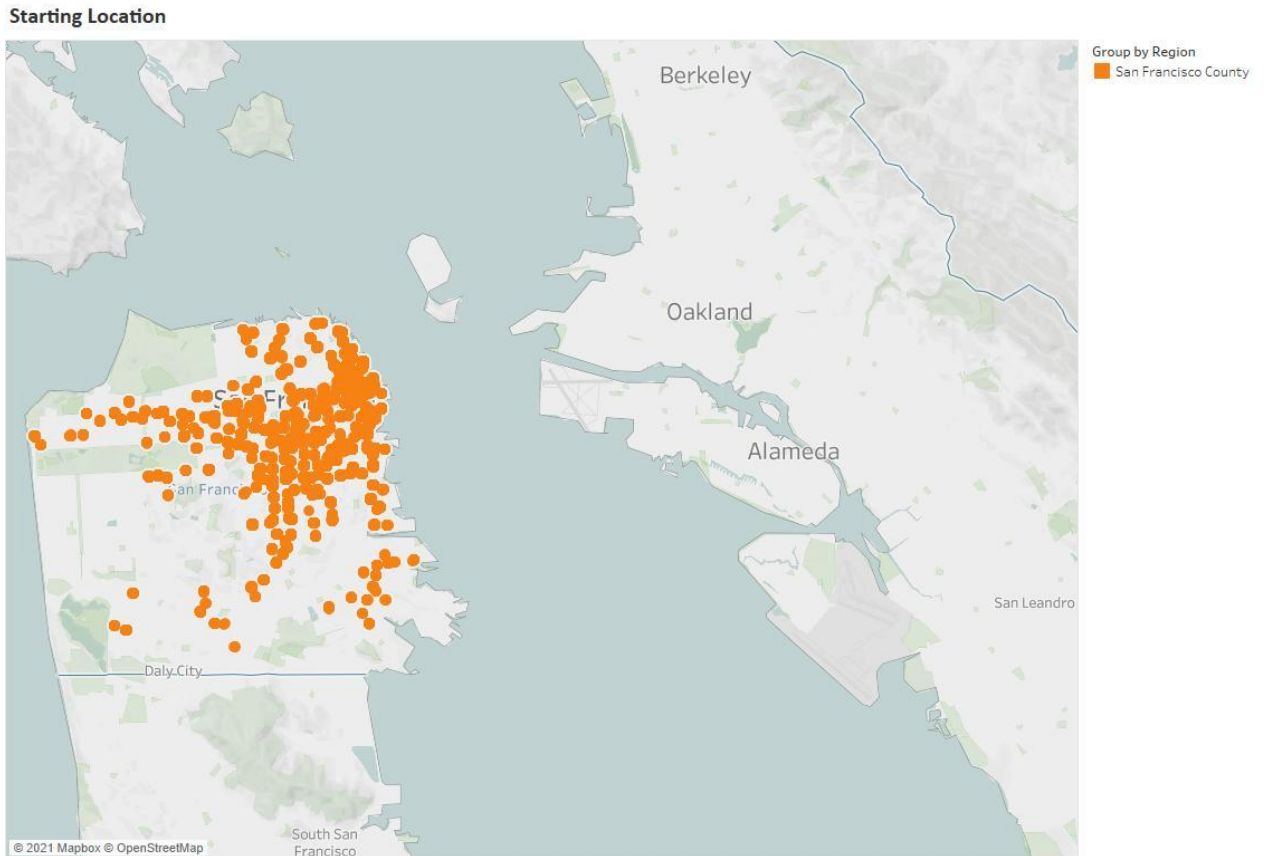


Figure c: Stations in San Francisco County

d) STATIONS AND TOTAL TRIPS BY SUBREGION

The figure below shows the number of stations and number of trips for each sub-region (San Francisco County, South Bay, East Bay) in the San Francisco Bay Area. The distinct count of station ID and the count of ride ID were placed in the column shelf. The stations were grouped by region and the group was placed in the rows shelf. The ride IDs are unique and the total trips would be given by count of ride IDs. Counting the distinct stations gives the number of stations.

San Francisco County has the highest number of stations as well as the number of trips, showing that a positive correlation exists between number of stations and total number of trips.

Number of Stations and Total Number of trips Starting at each Sub Region

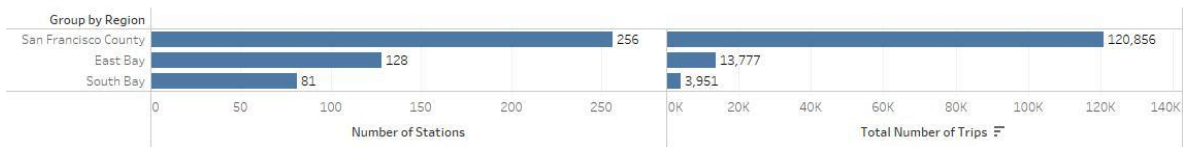


Figure d: Number of Station and Total Number of Trips starting at each sub region.

e) TRIP DURATION ANALYSIS IN SAN FRANCISCO COUNTY, EAST BAY AND SOUTH BAY

The figure below shows the standard deviation and average trip duration (in seconds) for each sub-region (San Francisco County, South Bay and East Bay) in the San Francisco Bay Area. The trip duration (secs) was placed in the columns shelf and the measure was set to Std Dev and Average. The stations were grouped by region and the group was placed in the rows shelf.

Though San Francisco County has the highest number of stations as well as the most number of trips in the Bay Area, people in the East Bay have used the service for a longer time and the least in the South Bay. Moreover we can conclude that there is no significant difference in the Std. Dev. for the three locations . But there is a significant difference between the Std. Dev and Mean for San Francisco County which indicates the variation between the avg. trip duration and actual time for which the trip was taken by the user .

Std. Dev and Avg. Trip Duration (secs) for each Sub Region



Figure e: Standard Deviation and Average Trip duration (secs) for each Sub Region.

f) TRIPS ENDING AT STATION ID Sf-J23-1

The map below shows all the trips that end at Sf-J23-1. Each point in the map describes where trips that end at SF-J23-1 start, average duration of the trip and SF-H26 and SF-L21 (Figure f(a)) have the most number of trips. To generate the figure, place the end station id in filters and select SF-J23-1 as the end station. In order to measure the average duration of trips, place the average measure on trip duration and perform the count on each starting station to find the station with the highest number of trips to the end station SF-J23-1.

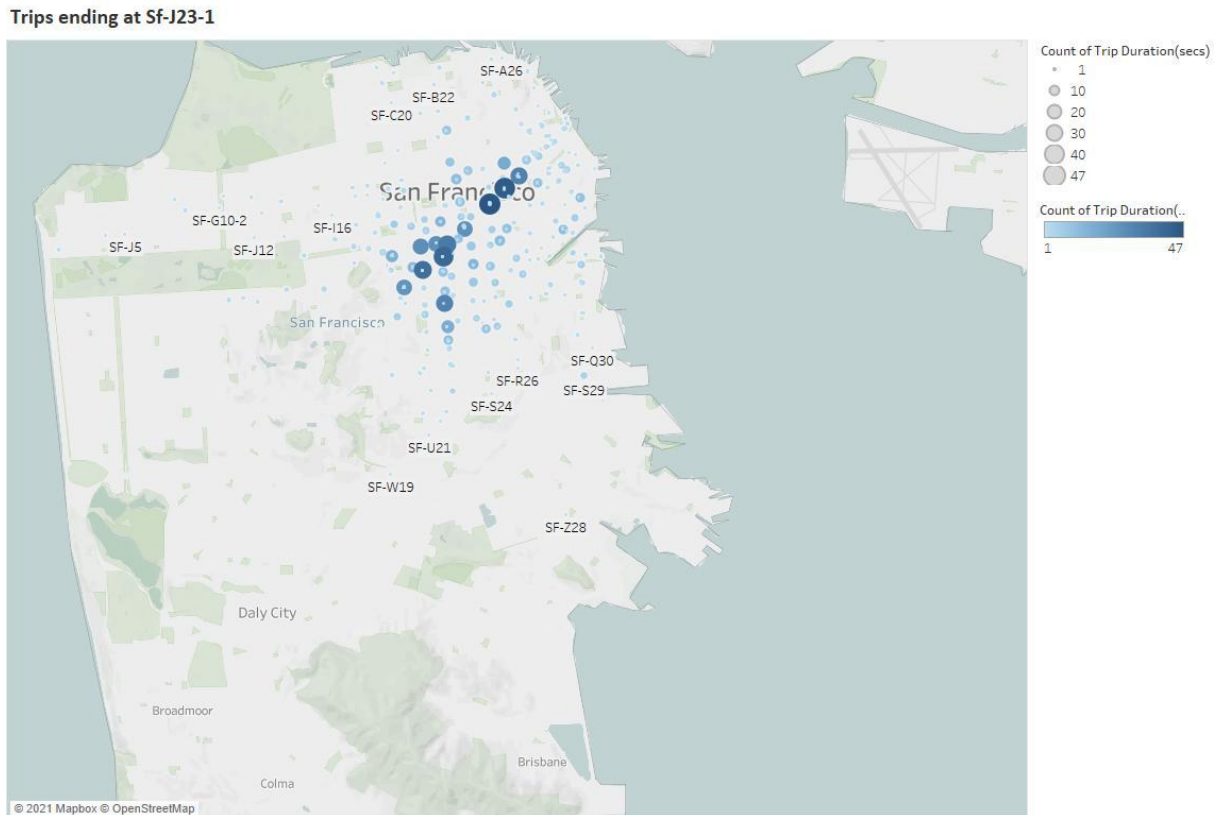


Figure f: Trips ending at SF-J23-1

Top 10 Starting Stations for SF-J23-1

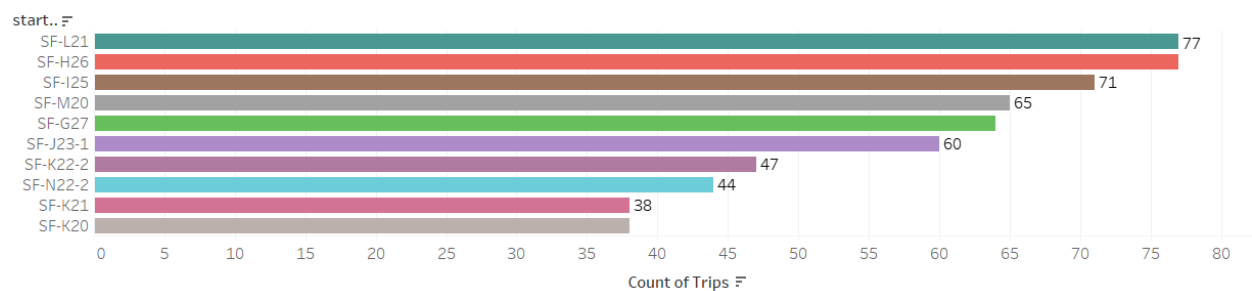


Fig f(a): Top Ten Starting Stations

Question 2: COVID-19 Cases

a) CUMULATIVE COVID-19 CASES BY COUNTRY ON JULY 26, 2021

The map below shows cumulative confirmed cases by country as of Jul 26, 2021. In order to create the map, place the location in the filters and filter out the continents and the location names that were not recognised. Place the date in the filters and select Jul 26, 2021. Place the sum measure on New cases, this would provide the total number of cases in each country till 26 July, 2021.

The Covid-19 cases are widespread in the world. The United States of America has the highest number of cases followed by India and Brazil.

Confirmed Covid-19 Cases as on 07/26/2021

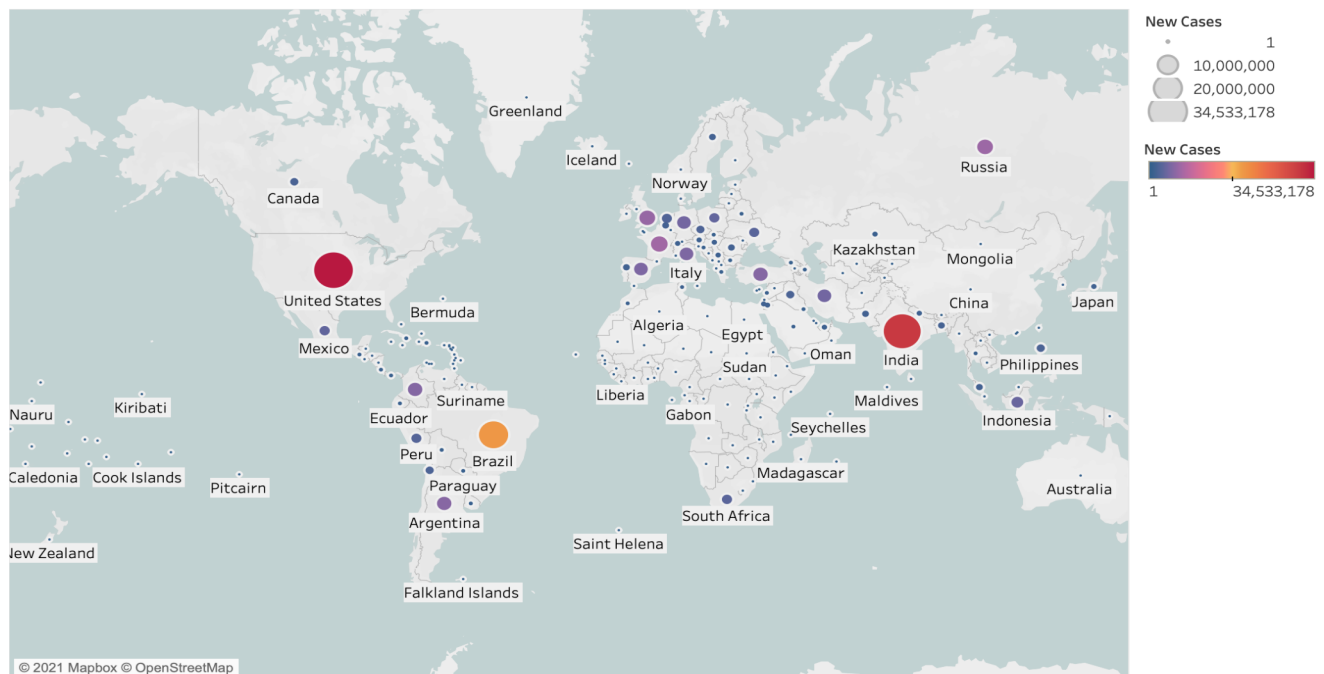


Figure a: Confirmed Cases of Covid-19 as on 07/26/2021

b) TIME SERIES CHART FOR CUMULATIVE COVID-19 CASES BY COUNTRY BETWEEN 2/24/2020 AND 06/24/2021

The figure below shows the cumulative confirmed cases between 2/24/2020 and 06/24/2021 for top 10 countries. To create the map, place the Date in the columns shelf to show a time series representation of the data. Apply the sum measure to the number of cases and place it on the rows shelf. Filter out the values from the sum of total cases and further place the location in filters to get the data for top 10 countries. In order to get the

date range, place the Date in the filters and select the date range 2/24/2020 and 06/24/2021.

Covid-19 Cases (02/24/2020 to 06/24/2021) for Top 10 Countries

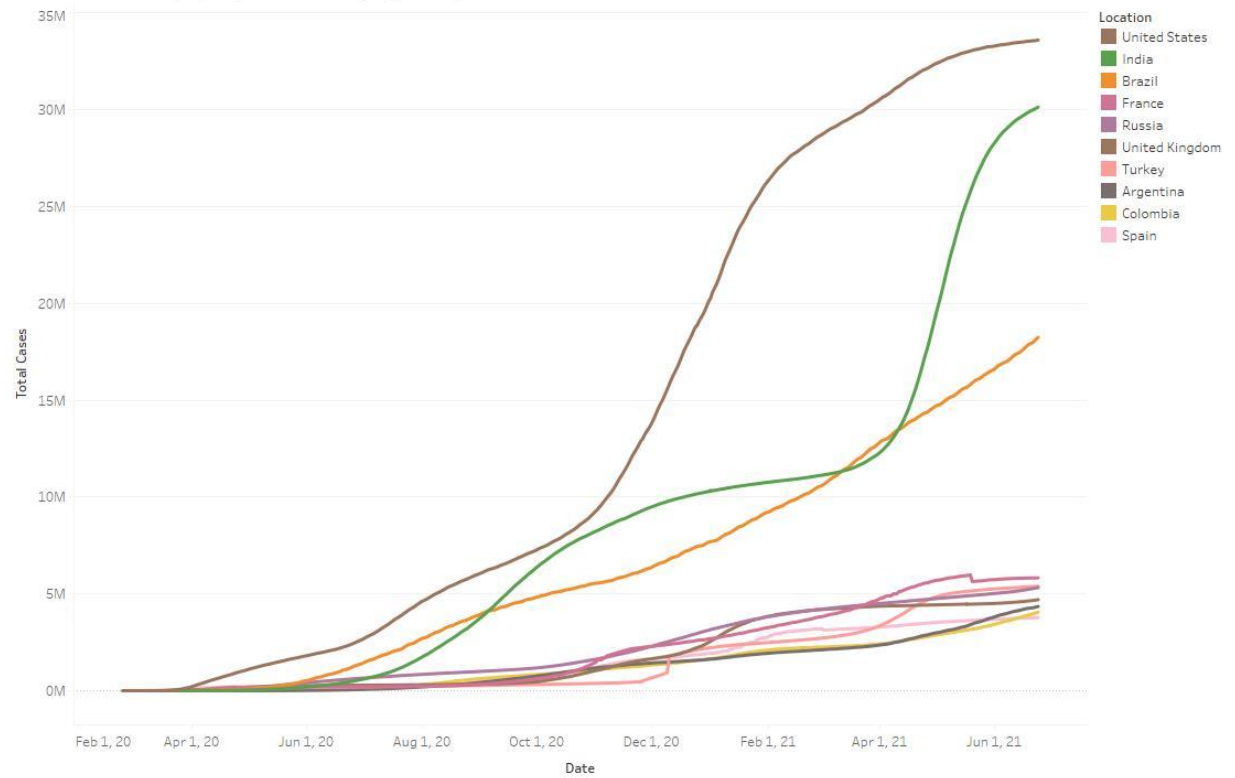


Figure b: Covid-19 Cases (02/24/2020 to 06/24/2021) for top 10 Countries