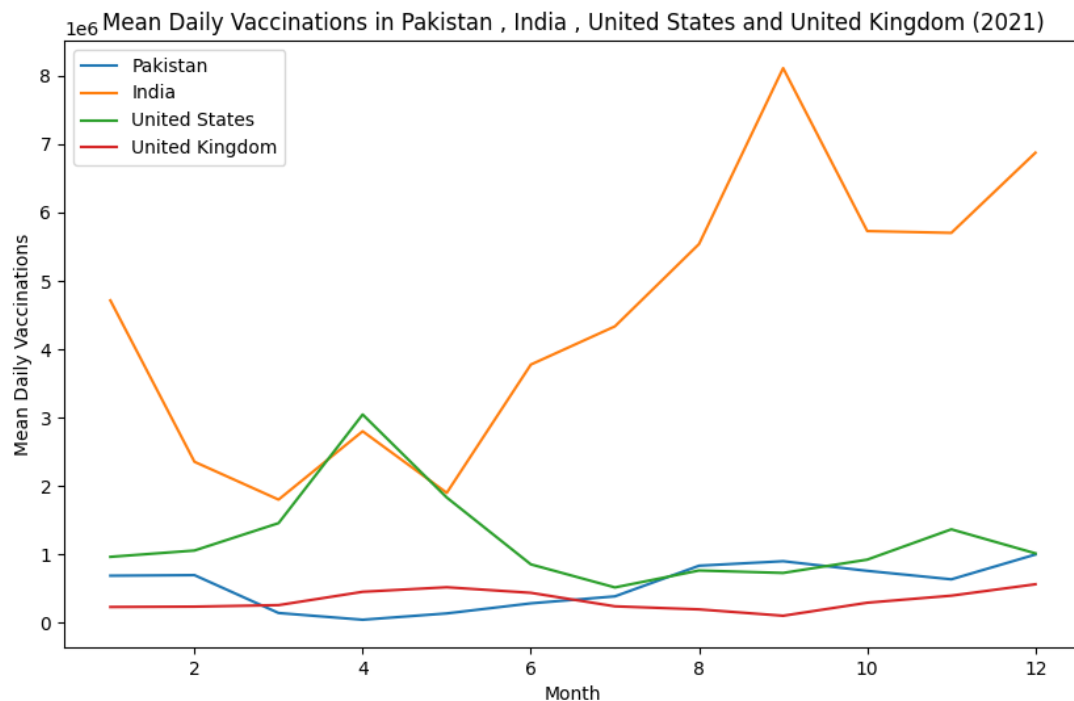# Visualisation

Name: Hassan Faraz Khan
Student ID: 21089475
Subject: Applied Data Science 1
Date: 3/March/2023
Data Source: https://www.kaggle.com/
Repository Link: https://github.com/HassanFrazKhan/Assignment_ADS

# Visualisation 1 : Plotting Line Graph Comparing The Mean Daily Vaccination

Data source:

What we get from this Plot?

From the above line plot, we can see the comparison of the mean daily vaccinations between Pakistan and India, United States and Pakistan in the year 2021. We can observe that India had a higher mean daily vaccination rate than Pakistan and United Kingdom throughout the year. However, we can also see that All the countries experienced an increase in the mean daily vaccination rate from around June to August except United Kingdom 2021, which could be due to increased vaccination efforts and availability of vaccines during that time. Overall, the plot gives us an idea about the vaccination progress in Four countries during the year 2021.

Visualis

ation 2: Scatter Plot for Showing Relationship Between Math Scores and Reading Scores

Data source: https://www.kaggle.com/spscientist/students-performance-in-exams
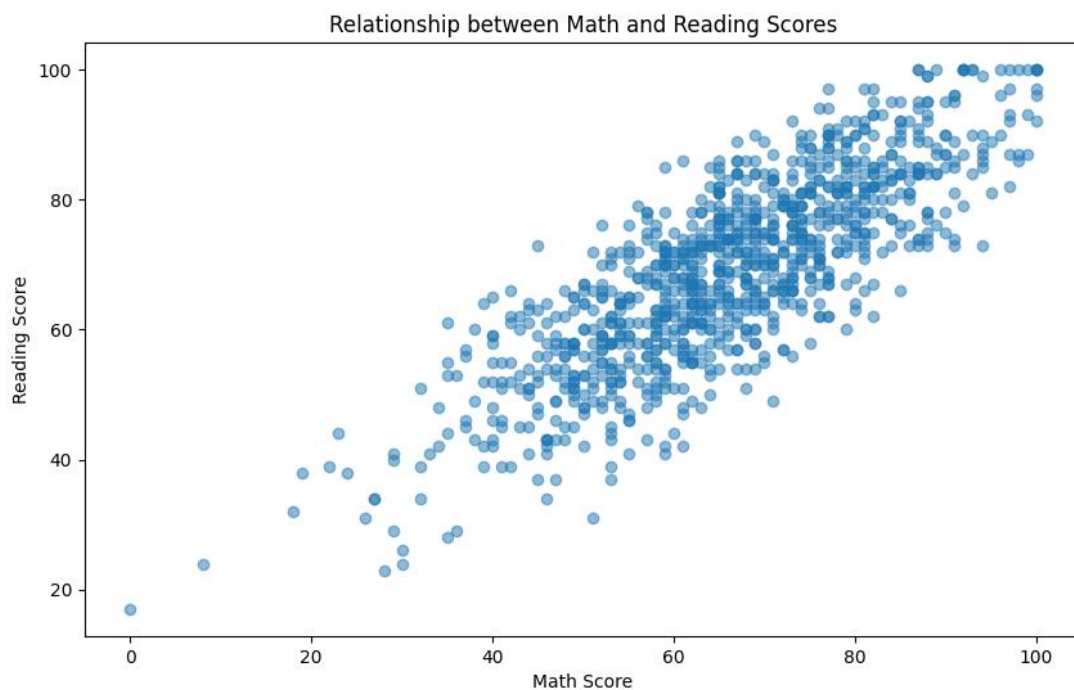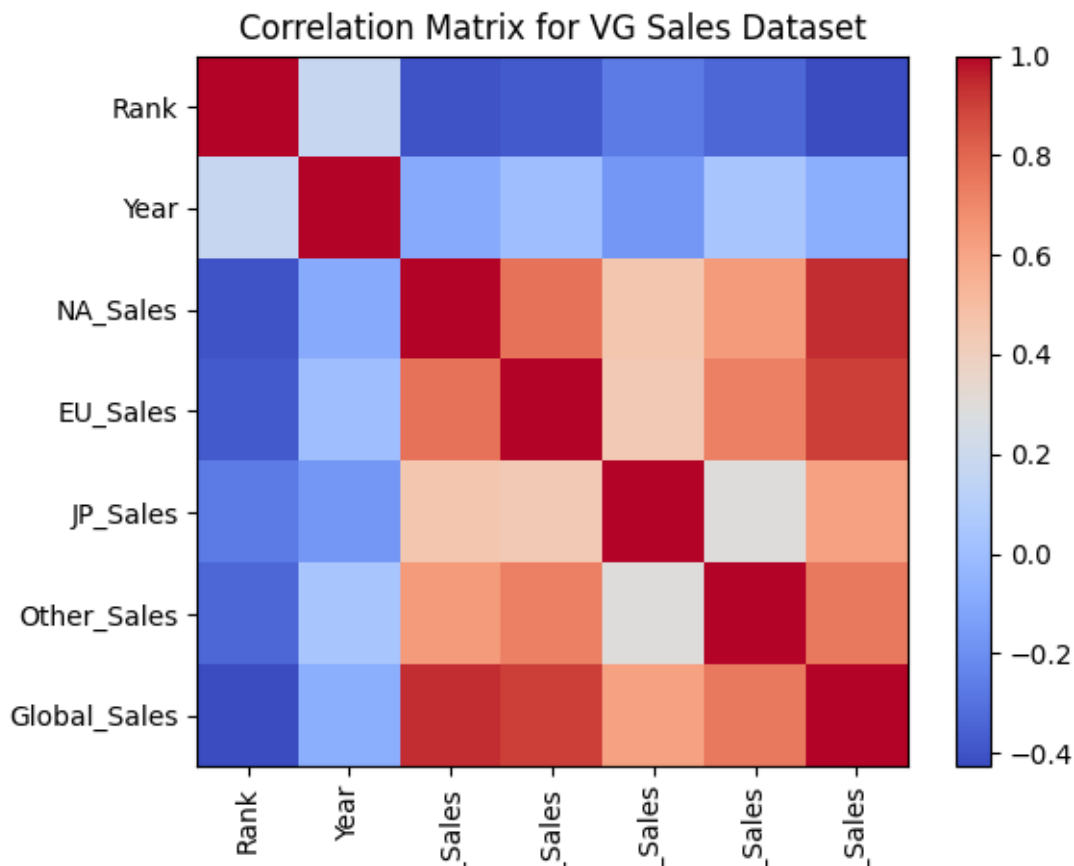

What we get from this plot?

From the scatter plot, we can see that there is a positive correlation between math scores and reading scores. This means that as the math scores increase, the reading scores also tend to increase. The scatter plot also shows that there are some students who scored high in both math and reading, while others scored low in both subjects. However, there are also some students who performed well in one subject but not the other. For example, there are some students who scored high in math but low in reading, and vice versa.

Visualisation 3: Heatmap for video games sales

Data Source:https://www.kaggle.com/datasets/gregorut/videogamesales

What we get from this plot?

The heatmap shows the correlation between the regions and the global sales of video games. We can see that there is a strong positive correlation between the sales in North America and the sales in Europe, as well as between the sales in North America and the global sales. There is also a weaker positive correlation between the sales in Japan and the global sales. On the other hand, there is a negative correlation between the sales in other regions and the sales in Japan, as well as between the sales in Other regions and the global sales. These insights can help video game companies to understand the market trends and tailor their strategies accordingly.

Python Code for Visualisation Using Visual Studio Code.

```python
import pandas as pd
import matplotlib.pyplot as plt
import pylab

def plot_mean_daily_vaccinations(data_path, country1, country2,country3,country4):
    """
    Ploting a line graph comparing the mean daily vaccinations for four countries
or we can increase the by changig paremeters  in a given year.

    Args:
    data_path (str): File path of dta
    country1 (str): The name of the first country to compare.
    country2 (str): The name of the second country to compare.
    country3 (str): The name of the third country to compare.
    country4 (str): The name of the fourth country to compare.

    Returns(displays a line plot)
    None
    """
    # Load the dataset
    data = pd.read_csv(data_path)

    # Filter the data for the specified countries
    filtered_data = data.loc[data['country'].isin([country1, country2,coun-
try3,country4]), ['country', 'date', 'daily_vaccinations']]

    # Convert the date column to datetime format
    filtered_data['date'] = pd.to_datetime(filtered_data['date'])

    # Extract the month from the date column
    filtered_data['month'] = filtered_data['date'].dt.month

    # Group the data by month and country, and calculate the mean daily vaccination
rate for each month
    grouped_data = filtered_data.groupby(['country', 'month']).mean().reset_index()

    # Pivot the data to create separate columns for each country
    pivot_data = grouped_data.pivot(index='month', columns='country', val-
ues='daily_vaccinations')

    # Create the plot
    plt.figure(figsize=(10, 6))
    plt.plot(pivot_data[country1], label=country1)
    plt.plot(pivot_data[country2], label=country2)
```

```python
    plt.plot(pivot_data[country3], label=country3)
    plt.plot(pivot_data[country4], label=country4)

    # Add labels and title
    plt.xlabel('Month')
    plt.ylabel('Mean Daily Vaccinations')
    plt.title(f'Mean Daily Vaccinations in {country1} , {country2} , {country3} and
{country4} (2021)')
    fig = pylab.gcf()
    fig.canvas.manager.set_window_title('Line Plot')

    # Add legend
    plt.legend()

    # Show the plot
    plt.show()



def create_scatterplot(data_path):
    """
    Creates a scatter plot showing the relationship between math scores and reading
scores
    for the "Students Performance" dataset on Kaggle.

    Parameters:
    data_path (str): The path to the CSV file containing the dataset.

    Returns:
    None (displays a scatter plot)
    """

    # Load the dataset
    data = pd.read_csv(data_path)

    # Create the scatter plot
    plt.figure(figsize=(10, 6))
    plt.scatter(data["math score"], data["reading score"], alpha=0.5)

    # Add labels and title
    plt.xlabel("Math Score")
    plt.ylabel("Reading Score")
    plt.title("Relationship between Math and Reading Scores")
    fig = pylab.gcf()
    fig.canvas.manager.set_window_title('Scatter plot')

    # Show the plot
    plt.show()
```

```python
def create_heatmap(data_path):
    """
    Create a heatmap for the "vgsales.csv" dataset from Kaggle using pyplot.

    Returns:
    None (display heatmap)
    """
    # Load the dataset
    df = pd.read_csv(data_path)

    # Compute the correlation matrix
    corr_matrix = df.corr()

    # Create the heat map using Pyplot
    plt.imshow(corr_matrix, cmap='coolwarm')

    # Add labels
    plt.xticks(range(len(corr_matrix)), corr_matrix.columns, rotation=90)
    plt.yticks(range(len(corr_matrix)), corr_matrix.columns)

    # Add colorbar
    plt.colorbar()
    fig = pylab.gcf()
    fig.canvas.manager.set_window_title('HeatMap')

    # Add title
    plt.title('Correlation Matrix for VG Sales Dataset')

    # Show the plot
    plt.show()


if __name__=='__main__':
    plot_mean_daily_vaccinations('country_vaccinations.csv', 'Pakistan', 'In-
dia','United States','United Kingdom')
    create_scatterplot('StudentsPerformance.csv')
    create_heatmap('vgsales.csv')
```