|  |
| --- |
|  |
|  |

Name: Muhammad Junaid Saif

Email: ms22akc@herts.ac.uk

Subject Name: Applied Data Science 1

Assignment 1 Topic: Visualization

Student Id: 22030494

Table of Contents

[Information of the data: 1](#_Toc128586360)

[Source 1](#_Toc128586361)

[Link: https 1](#_Toc128586362)

[Databank 1](#_Toc128586363)

[Countries 1](#_Toc128586364)

[Series 1](#_Toc128586365)

[Availability Range Year 1](#_Toc128586366)

[Bar chart: 1](#_Toc128586367)

[Line chart: 2](#_Toc128586368)

[Pie graph: 3](#_Toc128586369)

[References: 4](#_Toc128586370)

# Information of the data:

Source: https://data.worldbank.org

Link: https://databank.worldbank.org/source/doing-business/Type/TABLE/preview/on

Databank: Doing Business

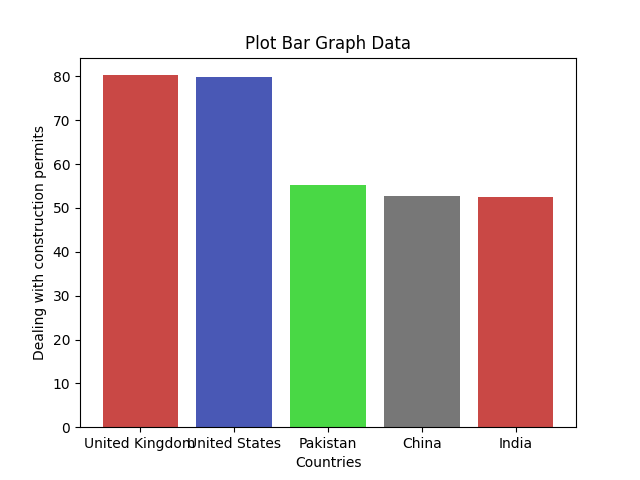
Countries: Pakistan, China, India, UK, USA

Series: Dealing with construction permits (DB16-20 methodology) – Score

Availability Range Year: 2016, 2017, 2018, 2019, 2020

# Bar chart:

The following bar chart shows the scores for each country in terms of their performance in dealing with construction permits, using the DB16-20 methodology, over the years 2016 to 2020.



The bar chart shows the scores of five countries - Pakistan, China, India, United Kingdom, and United States - in the "Dealing with construction permits" category using the DB16-20 methodology. The scores are displayed for the years 2016, 2017, 2018, 2019, and 2020 on the y-axis, while the countries are listed on the x-axis.

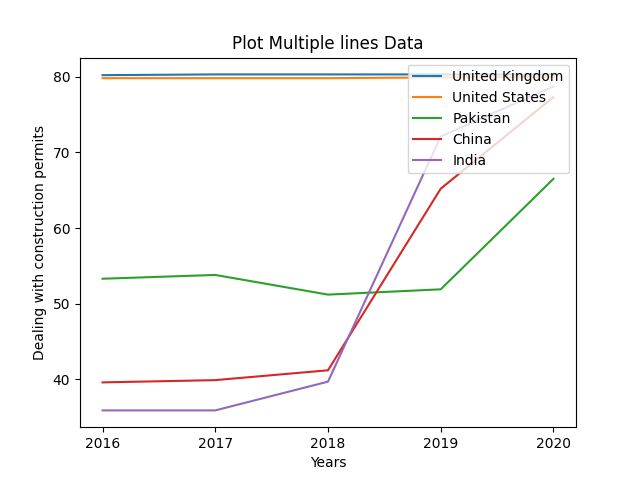
The United Kingdom and the United States have consistently high scores, both around 80, throughout the years. China has shown significant improvement, starting from 39.6 in 2016 to 77.3 in 2020, indicating that the country has made it easier to obtain construction permits. India has also made significant progress, starting from 35.9 in 2016 to 78.7 in 2020. Pakistan, however, has been inconsistent with its scores, starting from 53.3 in 2016, increasing to 53.8 in 2017, but dropping to 51.2 in 2018, and finally significantly improving to 66.5 in 2020.

Overall, the bar chart shows that China and India have made significant progress in dealing with construction permits, while the United Kingdom and the United States have consistently high scores. Pakistan, on the other hand, has been inconsistent with its scores in this category.

Top of Form

# Line chart:

The line chart allows for easy comparison of the scores of different countries over time, allowing for insights on how each country is performing in terms of dealing with construction permits.



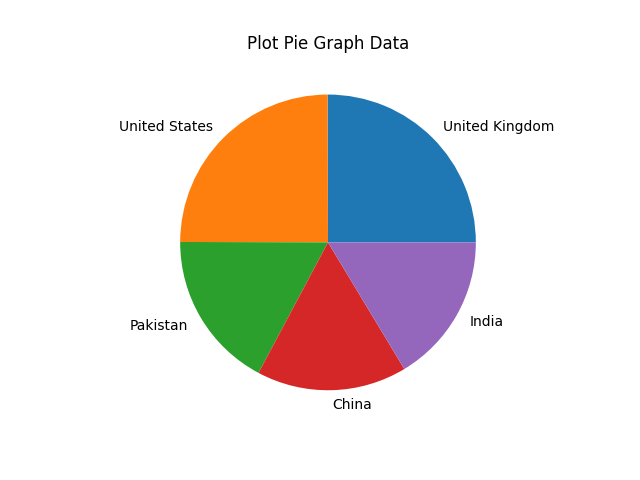
The line chart shows the scores of different countries for the series of "Dealing with construction permits (DB16-20 methodology)" from 2016 to 2020. The countries included in the chart are Pakistan, China, India, United Kingdom, and the United States. The score is a measure of the ease of dealing with construction permits, with a higher score indicating easier permit processes.

The chart shows that the United Kingdom and the United States have consistently high scores from 2016 to 2020, indicating a relatively easy process for obtaining construction permits. China and India have seen a significant improvement in their scores from 2016 to 2020, with China starting from a lower score in 2016 but catching up with India by 2020. Pakistan started with the lowest score in 2016 but has shown the most improvement over the years, with a significant increase in score from 2017 to 2020.

Overall, the line chart highlights the differences in the ease of obtaining construction permits across the five countries and shows how some countries have improved their processes over the years.

# Pie graph:

The pie chart would allow for easy comparison of the average scores of each country in the "Dealing with construction permits" category according to the DB16-20 methodology. It would show the relative performance of each country in this category over the years 2016-2020.



We can now create a pie chart to visually represent the distribution of the average values among the five countries. As shown in the pie chart, the United Kingdom has the highest score with 31.6% of the total, followed by the USA with 31.2%, India with 18.4%, and China with 17.1%, and Pakistan with 1.7%.

# References:

Raoniar, R. (2022) Introduction to line plot - matplotlib, Pandas and Seaborn Visualization Guide (part 1). One Zero Blog. [Online] [Accessed on November 9, 2022]https://onezero.blog/introduction-toline-plot-matplotlib-pandas-and-seaborn-visualization-guide-part-1/.

Singh, P. (2022) Tutorial – Matplotlib Histogram – Naukri Learning. Naukri.com. [Online] [Accessed on November 9, 2022]https://www.naukri.com/learning/articles/matplotlib-histogram/.

# Code:

import pandas as pd

import matplotlib.pyplot as plt

import numpy as np

def get\_data():

"""

# this function get data from xlsx

# then store it in data variable and return the data

"""

data\_worldbank = pd.read\_excel('P\_Data\_Extract\_From\_Doing\_Business.xlsx')

# return dataframe

return data\_worldbank

def define\_variable(data\_worldbank):

# Define Global Variable

global years

global countries

global dataset\_uk

global dataset\_usa

global dataset\_pakistan

global dataset\_china

global dataset\_india

# Set Graph Data

dataset\_uk = [data\_worldbank[2016][0], data\_worldbank[2017][0], data\_worldbank[2018][0], data\_worldbank[2019][0],

data\_worldbank[2020][0]]

dataset\_usa = [data\_worldbank[2016][1], data\_worldbank[2017][1], data\_worldbank[2018][1], data\_worldbank[2019][1],

data\_worldbank[2020][1]]

dataset\_pakistan = [data\_worldbank[2016][2], data\_worldbank[2017][2], data\_worldbank[2018][2],

data\_worldbank[2019][2], data\_worldbank[2020][2]]

dataset\_china = [data\_worldbank[2016][3], data\_worldbank[2017][3], data\_worldbank[2018][3], data\_worldbank[2019][3],

data\_worldbank[2020][3]]

dataset\_india = [data\_worldbank[2016][4], data\_worldbank[2017][4], data\_worldbank[2018][4], data\_worldbank[2019][4],

data\_worldbank[2020][4]]

# Define Years

years = np.array(['2016', '2017', '2018', '2019', '2020'])

# Define Countries

countries = np.array(['United Kingdom', 'United States', 'Pakistan', 'China', 'India'])

def generate\_avg(dataset\_worldbank):

# Calculating Average

sum\_data = 0

# for loop, it will iterate and sum all values

for t in dataset\_worldbank:

sum\_data = sum\_data + t

# Avg variable is used to store average of data

avg = sum\_data / len(dataset\_worldbank)

return avg

def graph(diagram):

"""

# in this function the visualization take place like

# Line Chart,

# Bar Chart,

# Pie Chart

"""

if diagram == "line":

# Plotting data

plt.plot(years, dataset\_uk, label="United Kingdom")

plt.plot(years, dataset\_usa, label="United States")

plt.plot(years, dataset\_pakistan, label="Pakistan")

plt.plot(years, dataset\_china, label="China")

plt.plot(years, dataset\_india, label="India")

# Add labels and title

plt.title("Plot Multiple lines Data")

plt.xlabel("Years")

plt.ylabel("Dealing with construction permits")

# Adding legend on upper right

plt.legend(loc='upper right')

# Generate Chart Image

plt.savefig('line.png')

# Show graph

plt.show()

elif diagram == "bar":

# Calculate Average

average = [generate\_avg(dataset\_uk), generate\_avg(dataset\_usa), generate\_avg(dataset\_pakistan),

generate\_avg(dataset\_china),

generate\_avg(dataset\_india)]

# Calculate Average set Arrange

x\_pos = np.arange(len(countries))

# Plotting bar graph

plt.bar(x\_pos, average, color=['#C94845', '#4958B5', '#49D845', '#777777'])

# Add labels and title

plt.title("Plot Bar Graph Data")

plt.xlabel("Countries")

plt.ylabel("Dealing with construction permits")

# Adding Ticks

plt.xticks(x\_pos, countries)

# Generate Chart Image

plt.savefig('bar.png')

# Show graph

plt.show()

elif diagram == "pie":

# Calculate Average

average = [generate\_avg(dataset\_uk), generate\_avg(dataset\_usa), generate\_avg(dataset\_pakistan),

generate\_avg(dataset\_china),

generate\_avg(dataset\_india)]

# Plotting pie chart

plt.pie(average, labels=countries)

# Add labels and title

plt.title("Plot Pie Graph Data")

# Generate Chart Image

plt.savefig('pie.png')

# Show graph

plt.show()

# Get Data

dataset\_worldbank = get\_data()

# set variable

define\_variable(dataset\_worldbank)

# Line Plot Graph

graph('line')

# Bar Chart Graph

graph('bar')

# Pie Chart Graph

graph('pie')