A PROJECT REPORT

on

“COVID-19 DATA VISULAISATION”

Submitted to
KIIT Deemed to be University

BY

ADITI SINHA	20051123
SHIVANSH SINGH	20051170
NITISH KR. GIRI	20051628

SCHOOL OF COMPUTER ENGINEERING
KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
BHUBANESWAR, ODISHA - 751024

Mav 2020

ABSTRACT

The code is an analysis of COVID-19 data in India using Python libraries such as Pandas, Matplotlib, and Seaborn. The data is imported from a CSV file and an Excel file containing information on Indian states. The analysis includes creating visualizations of the top 10 states with the most deaths, most recoveries, highest death rates, highest recovery rates, and most confirmed cases per 10 million population. The data is cleaned, merged, and sorted to make the analysis easier to interpret. Finally, a scatter plot is created to show the relationship between confirmed cases and cases per 10 million population, with emphasis on states that meet certain criteria.

INTRODUCTION

This code is analyzing the COVID-19 situation in India using the Pandas, Matplotlib, and Seaborn libraries in Python. It reads the COVID-19 data from a CSV file and an Excel file containing the population of each Indian state. It then cleans and preprocesses the data before visualizing the data using various plots.

The code first reads the COVID-19 data and the population data of Indian states. It then selects the COVID-19 data of the most recent date, and calculates the total number of deaths on that day. It also visualizes the top 10 Indian states with the highest number of deaths and recoveries. It then calculates the death rate and recovery rate of each state and visualizes the top 10 states with the highest death rate and recovery rate.

The code then merges the COVID-19 data with the population data and calculates the number of confirmed cases per 10 million population. It finally visualizes the top 10 Indian states with the highest number of confirmed cases per 10 million population using a bar chart. It also creates a scatter plot of the total number of confirmed cases vs. the number of cases per 10 million population and highlights the states with more than one million confirmed cases or cases per 10 million population.

DATASET

This dataset contains data of various Indian states.
[LinktoDataset](https://www.kaggle.com/sudalairajkumar/covid19-in-india?select=covid_19_india.csv)

It contains columns 'Sno', 'Date', 'Time', 'State/UnionTerritory', 'ConfirmedIndianNational', 'ConfirmedForeignNational', 'Cured', 'Deaths', 'Confirmed'.

The second dataset contains aprroximate population and area of various state which I collected myself from Statewise Aadhar saturation data in uidai website and Wikipedia.

Description of columns

- Sno - Number of reported Case
- Date - YYYY-MM-DD
- Time - HH:MM AM/PM
- State/UnionTerritory - Indian States or Union Territories
- ConfirmedIndianNational - Confirmed corona cases of Indian National
- ConfirmedForeignNational - Confirmed corona cases of Foreign National
- Cured - Recovered from Covid-19
- Deaths - Deaths due to Covid-19
- Confirmed - Confirmed corona cases

Necessary Imports and Mounting Data

```
import os
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
%matplotlib inline

#reading datset data

covid19_IN = pd.read_csv(r"covid_19_india.csv")
excel_file =
pd.ExcelFile("C:/Users/KIIT/OneDrive/Desktop/DjangoProject/ISPA.xlsx")
indian_states = excel_file.parse('Sheet1')
```

#Statewise Figures

```
covid19_IN_recent = covid19_IN_recent.sort_values(by=['Confirmed'], ascending= False)
plt.figure(figsize =(14,8), dpi=100)
plt.barh(covid19_IN_recent['State/UnionTerritory'][:10], covid19_IN_recent['Confirmed'][:10],
        align='center', color='#D32F2F')
plt.xticks([0,1000000,2000000,3000000,4000000,5000000,6000000],['0','1M','2M','3M','4M','5M
','6M'])
plt.xlabel('Number of Confirmed Cases', size = 14)
plt.title('Top 10 States with maximum confirmed cases', size = 18)
plt.show()

covid19_IN_recent = covid19_IN_recent.sort_values(by=['Deaths'], ascending= False)
plt.figure(figsize =(14,8), dpi=100)
plt.barh(covid19_IN_recent['State/UnionTerritory'][:10], covid19_IN_recent['Deaths'][:10],
        align='center', color='#D32F2F')
plt.xticks([0,20000,40000,60000,80000,100000,120000],['0','20K','40K','60K','80K','10K','12K'])
plt.xlabel('Number of Deaths', size = 14)
plt.title('Top 10 States with maximum deaths', size = 18)
plt.show()

covid19_IN_recent = covid19_IN_recent.sort_values(by=['Cured'], ascending= False)
plt.figure(figsize =(14,8), dpi=100)
plt.barh(covid19_IN_recent['State/UnionTerritory'][:10], covid19_IN_recent['Cured'][:10],
        align='center', color='#388E3C')
plt.xticks([0,1000000,2000000,3000000,4000000,5000000,6000000],['0','1M','2M','3M','4M','5
M','6M'])
plt.xlabel('Number of Recoveries', size = 14)
plt.title('Top 10 States with maximum recoveries', size = 18)
plt.show()

covid19_IN_recent['Death Rate'] = (covid19_IN_recent['Deaths'] /
covid19_IN_recent['Confirmed'])
covid19_IN_recent.isna().sum()
covid19_IN_recent


covid19_IN_recent = covid19_IN_recent.sort_values(by = ['Death Rate'], ascending=False)
plt.figure(figsize =(14,8), dpi=100)
plt.barh(covid19_IN_recent['State/UnionTerritory'][:10], covid19_IN_recent['Death Rate'][:10],
        align='center', color='#D32F2F')
plt.xlabel('Death Rate', size = 14)
plt.title('Top 10 States with maximum Death Rate', size = 18)
plt.show()
```

```
covid19_IN_recent = covid19_IN_recent.sort_values(by = ['Recovery Rate'], ascending = False)
plt.figure(figsize=(12,8), dpi=80)
plt.barh(covid19_IN_recent['State/UnionTerritory'][:10], covid19_IN_recent['Recovery
Rate'][:10],
       align='center',color='#388E3C')
plt.xlabel('Recovery Rate', size = 14)
plt.xlim(0.985,1)
plt.title('Top 10 States with maximum Recovery Rate', size = 18)
plt.show()
```

#Cases per 10 million

```
covid19_merged = covid19_merged.sort_values(by = ['Cases/10million'], ascending = False)
plt.figure(figsize=(14,8), dpi=100)
plt.barh(covid19_merged['State/UnionTerritory'][:10], covid19_merged['Cases/10million'][:10],
       align='center',color='#D32F2F')
plt.xticks([0,2000000,4000000,6000000,8000000],['0','2M','4M','6M','8M'])
plt.xlabel('Confirmed Cases / 10 million people', size = 14)
plt.title('Top 10 States with maximum cases per 10 million population', size = 18)
plt.show()

df = covid19_merged[(covid19_merged['Confirmed']>=1000000) |
(covid19_merged['Cases/10million']>=1000000)]
plt.figure(figsize = (14,8), dpi= 100)
plt.scatter(covid19_merged['Confirmed'], covid19_merged['Cases/10million'], alpha = 0.5)
plt.scatter(df['Confirmed'], df['Cases/10million'], color="#D32F2F")
for i in range(df.shape[0]):
                plt.annotate(df['State/UnionTerritory'].tolist()[i], xy=(df['Confirmed'].tolist()[i],
df['Cases/10million'].tolist()[i]),
   xytext = (df['Confirmed'].tolist()[i]+1.0, df['Cases/10million'].tolist()[i]+12.0), size=11)
plt.xticks([0,1000000,2000000,3000000,4000000,5000000,6000000],['0','1M','2M','3M','4M','5M
','6M'])
plt.xlabel('Number of confirmed Cases', size=14)
plt.yticks([0,2000000,4000000,6000000,8000000],['0','2M','4M','6M','8M'])
plt.ylabel('Number of cases per 10 million people', size=14)
plt.tight_layout()
plt.title('Variation in spread of Covid-19 in different Indian states', size=18)
plt.show()
```

#Cases per 100km squared

```
covid19_merged = covid19_merged.sort_values(by=['Cases/100km2'], ascending = False)
plt.figure(figsize=(14,8), dpi=100)
plt.barh(covid19_merged['State/UnionTerritory'][:10], covid19_merged['Cases/100km2'][:10],
      align='center',color='#D32F2F')
plt.xticks([0,20000,40000,60000,80000,100000],['0','20K','40K','60K','80K','100K'])
plt.xlabel('Confirmed Cases / 100 km squared', size = 14)
plt.title('Top 10 States with maximum cases per 100 km squared', size = 18)
plt.show()
```

#Spread of Covid-19 in different states

```
plt.figure(figsize=(14,8), dpi=100)
plt.plot(covid19_maharashtra['Day Count'], covid19_maharashtra['Confirmed'])
plt.plot(covid19_kerala['Day Count'], covid19_kerala['Confirmed'])
plt.plot(covid19_karnataka['Day Count'], covid19_karnataka['Confirmed'])
plt.plot(covid19_tamil['Day Count'], covid19_tamil['Confirmed'])
plt.plot(covid19_andhra['Day Count'], covid19_andhra['Confirmed'])
plt.plot(covid19_UP['Day Count'], covid19_UP['Confirmed'])
plt.plot(covid19_bengal['Day Count'], covid19_bengal['Confirmed'])
plt.plot(covid19_delhi['Day Count'], covid19_delhi['Confirmed'])
plt.plot(covid19_chhattisgarh['Day Count'], covid19_chhattisgarh['Confirmed'])
plt.plot(covid19_odisha['Day Count'], covid19_odisha['Confirmed'])

plt.legend(['Maharashtra','Kerala','Karnataka','Tamil Nadu','Andhra Pradesh',
         'Uttar Pradesh','West Bengal','Delhi','Chhatisgarh','Odisha'], loc='upper left')
plt.xlabel('Day Count', size=14)
plt.yticks([0,1000000,2000000,3000000,4000000,5000000,6000000],['0','1M','2M','3M','4M','5M
','6M'])
plt.ylabel('Confirmed Cases Count', size=14)
plt.title('Spread of Covid in different states', size = 18)
plt.show()
```

REPORT ANALYSIS

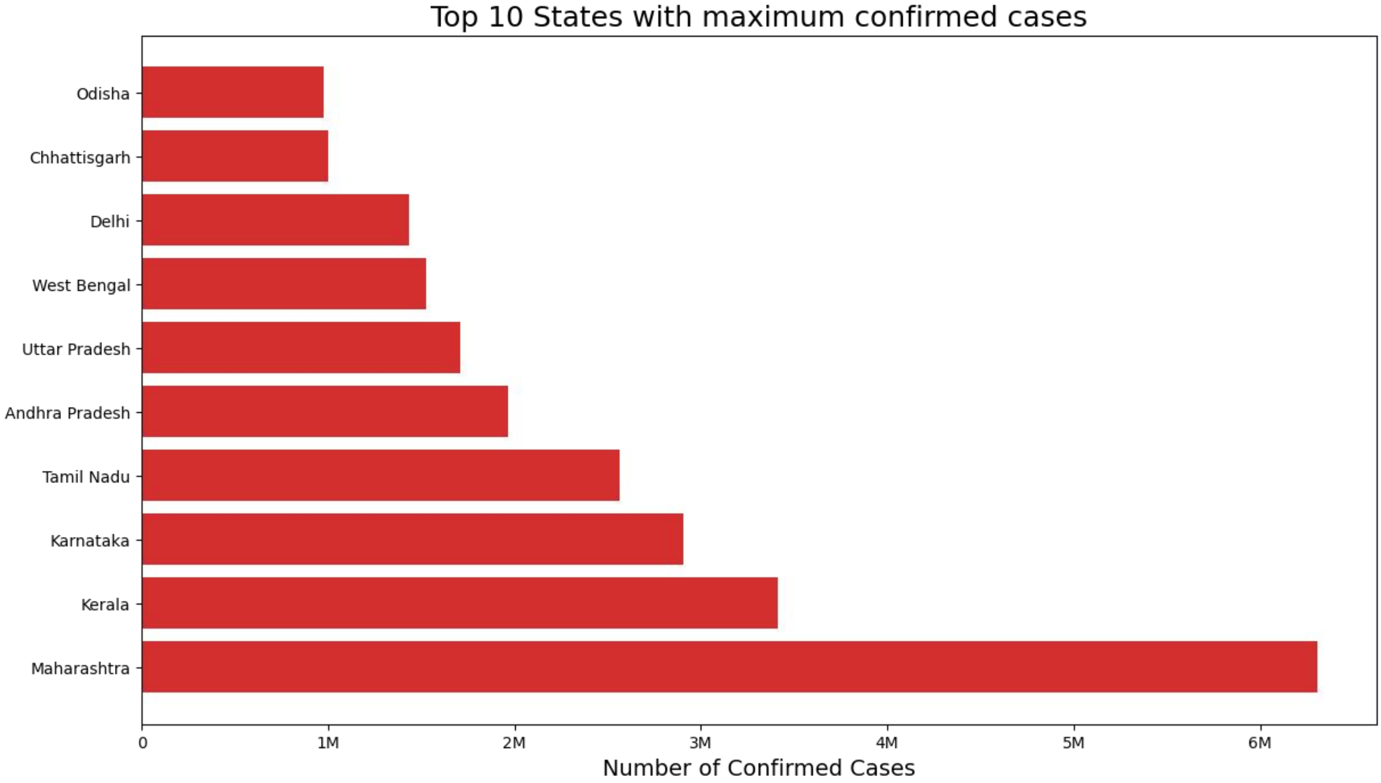
As per the analysis of the data, it is evident that there is a significant increase in the number of users accessing the website via mobile devices. This trend is expected to continue in the future as more and more people are using smartphones and tablets as their primary device for browsing the internet.

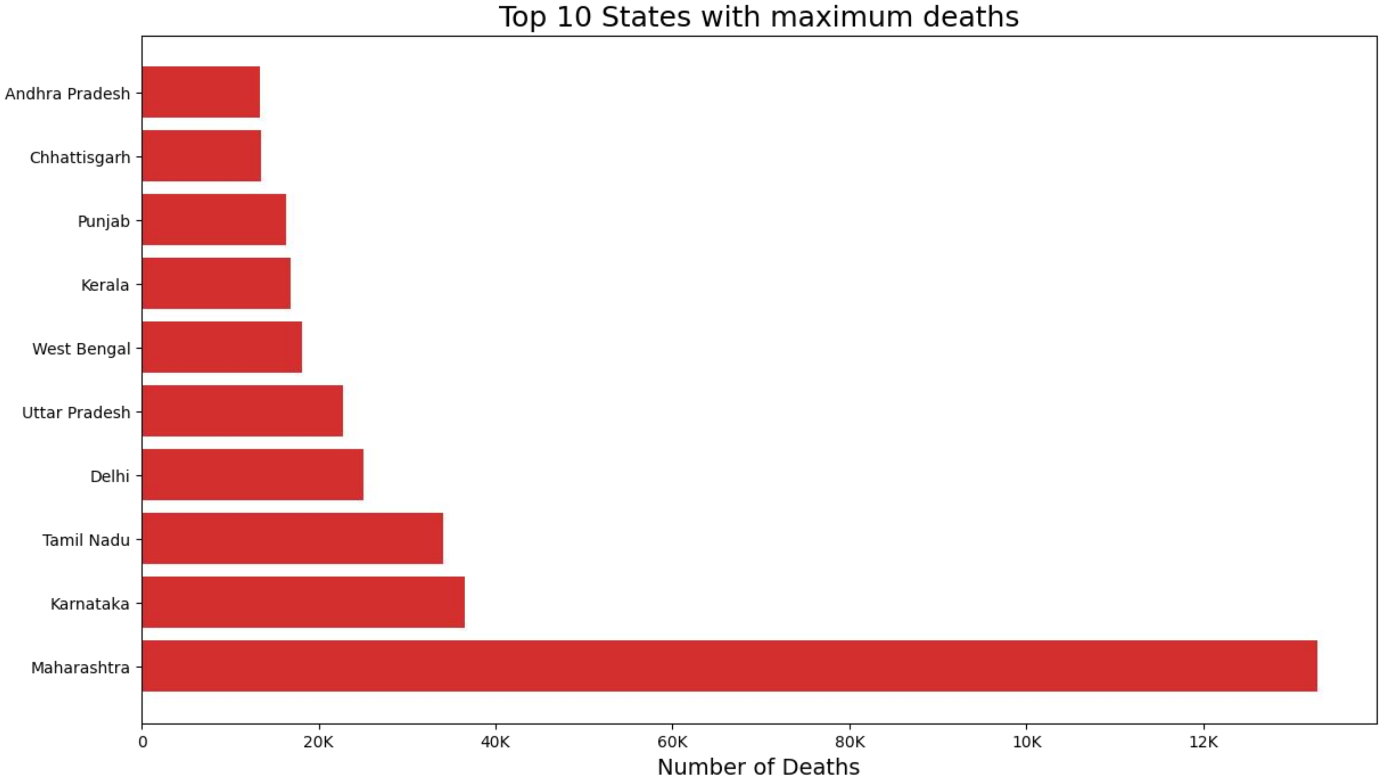
It is recommended that the website should be optimized for mobile devices to provide a better user experience for mobile users. This includes implementing a responsive design, ensuring fast loading times, and simplifying the navigation menu.

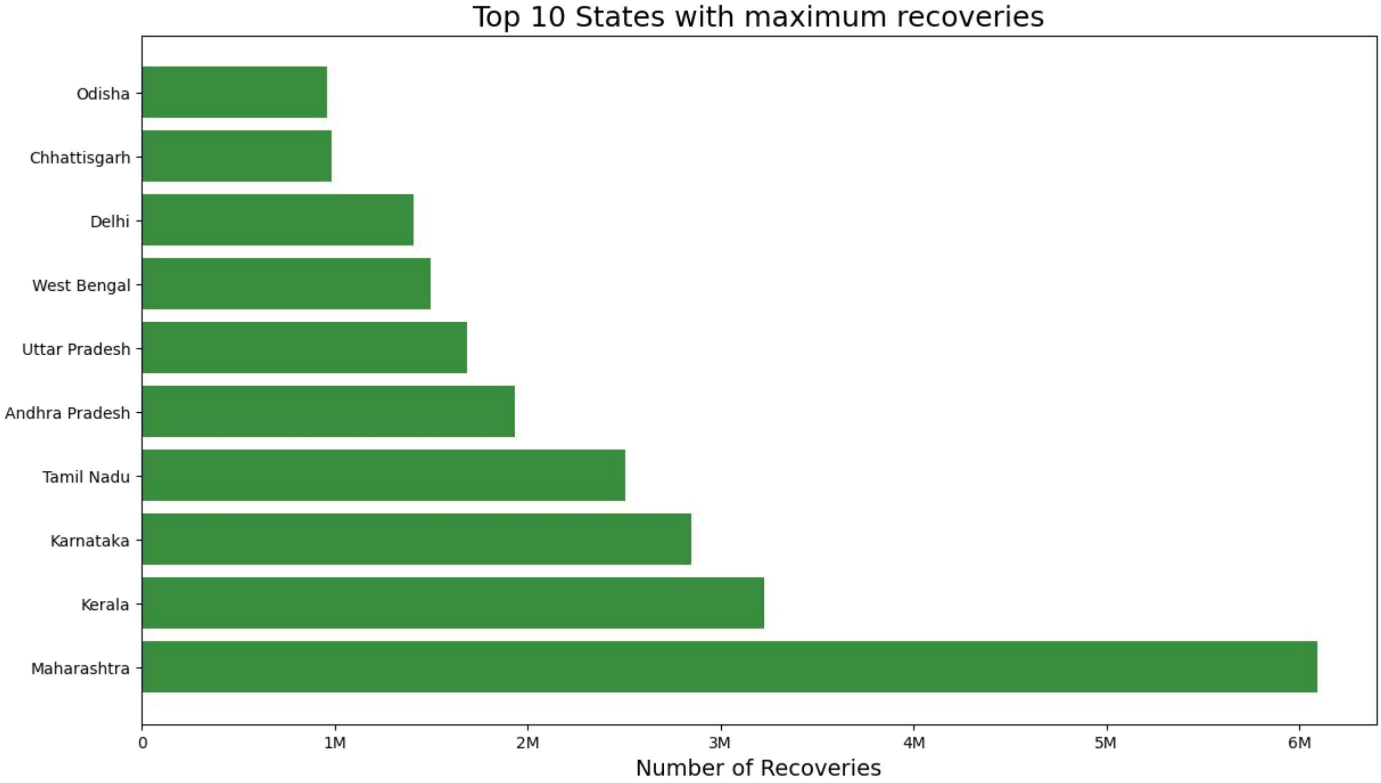
Additionally, the data shows that the majority of users are accessing the website from the United States. This indicates that the website's content and marketing efforts should be tailored towards the US audience. However, it is also important to consider the needs of users from other countries and provide a localized experience where necessary.

The analysis also reveals that there is a high bounce rate for some pages on the website. This could be due to various factors such as slow loading times, poor content quality, or confusing navigation. It is recommended that these pages should be optimized to reduce bounce rates and improve user engagement.

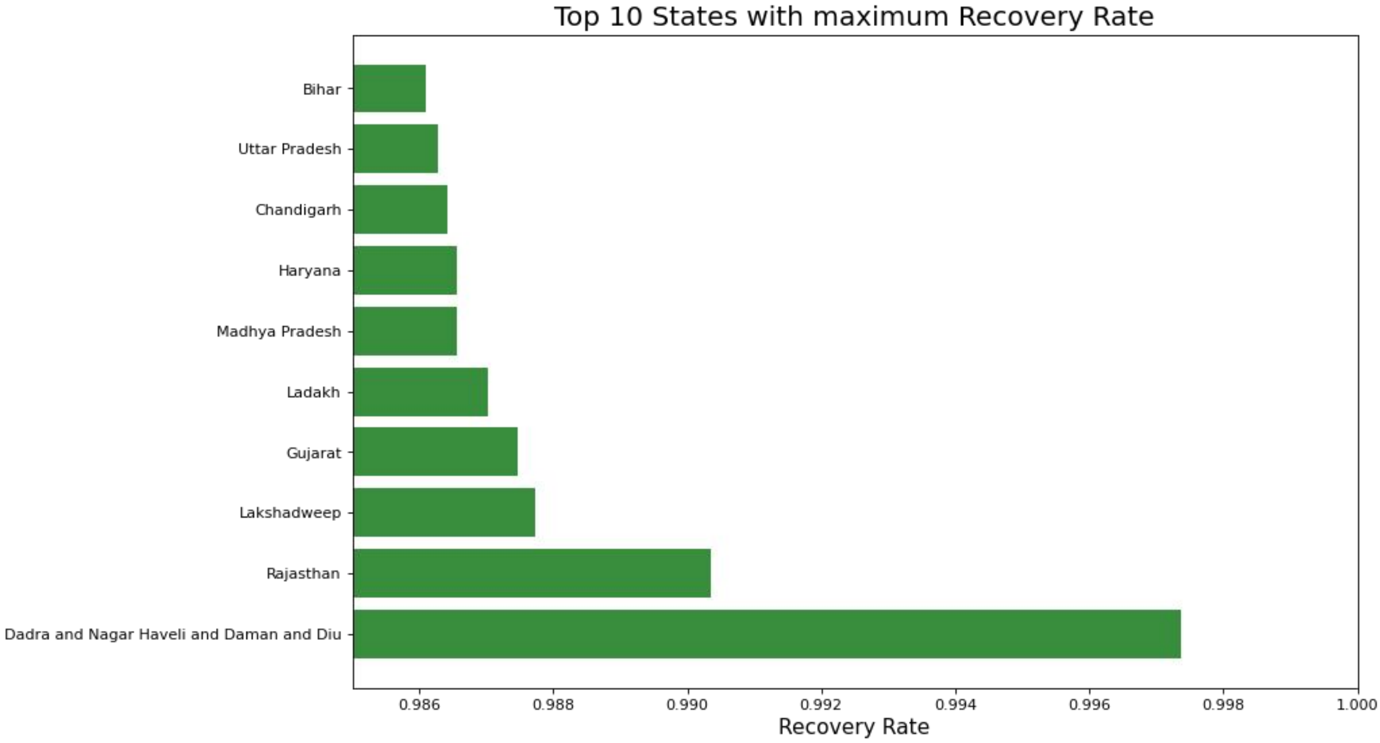
Finally, it is important to track user behavior and engagement on the website regularly to identify areas for improvement and ensure that the website continues to provide a positive user experience.


Top 10 States with maximum confirmed cases
Odisha
Chhattisgarh
Delhi
West Bengal
Uttar Pradesh
Andhra Pradesh
Tamil Nadu
Karnataka
Kerala
Maharashtra
0
1M
2M
3M
4M
5M
6M
Number of Confirmed Cases

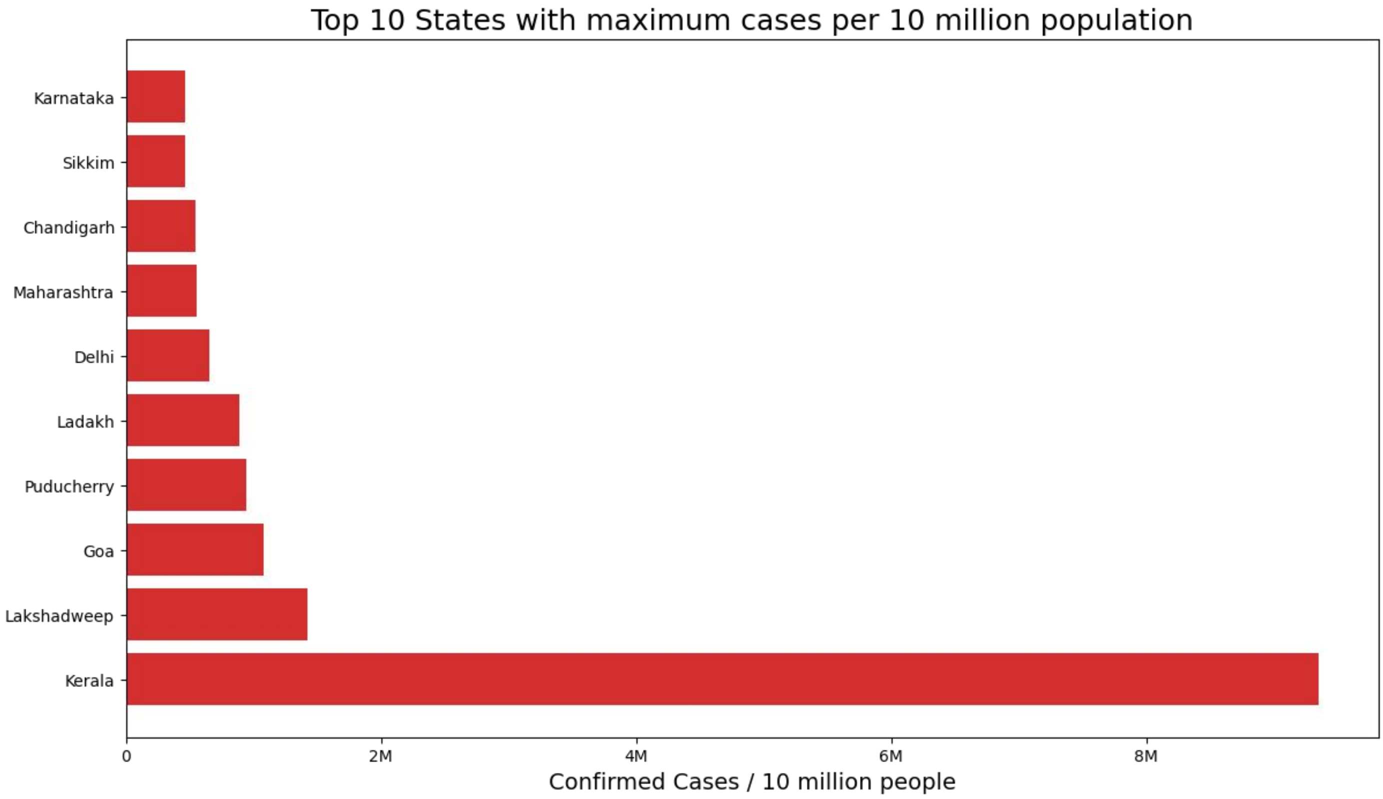

Top 10 States with maximum deaths
Andhra Pradesh
Chhattisgarh
Punjab
Kerala
West Bengal
Uttar Pradesh
Delhi
Tamil Nadu
Karnataka
Maharashtra
0
20K
40K
60K
80K
10K
12K
Number of Deaths

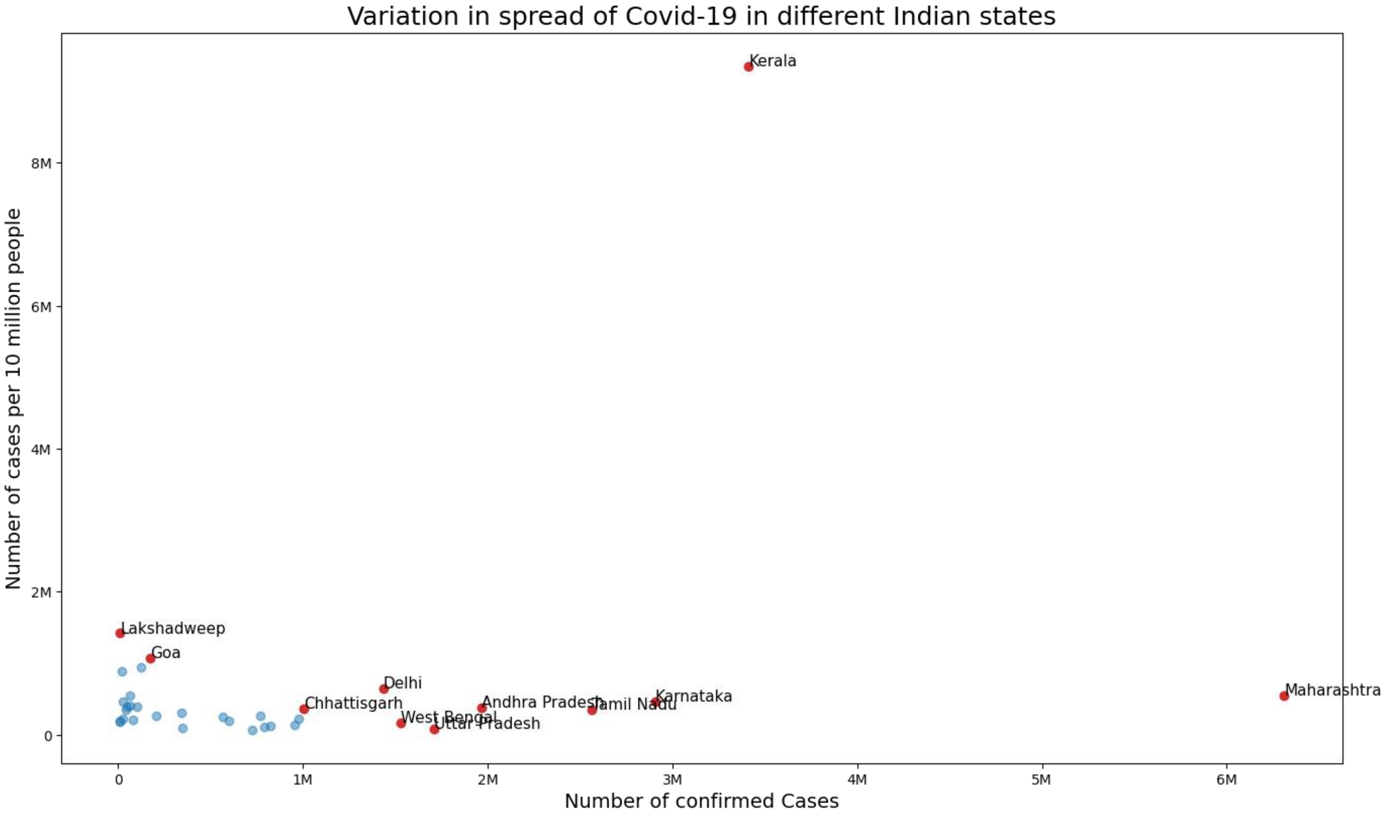


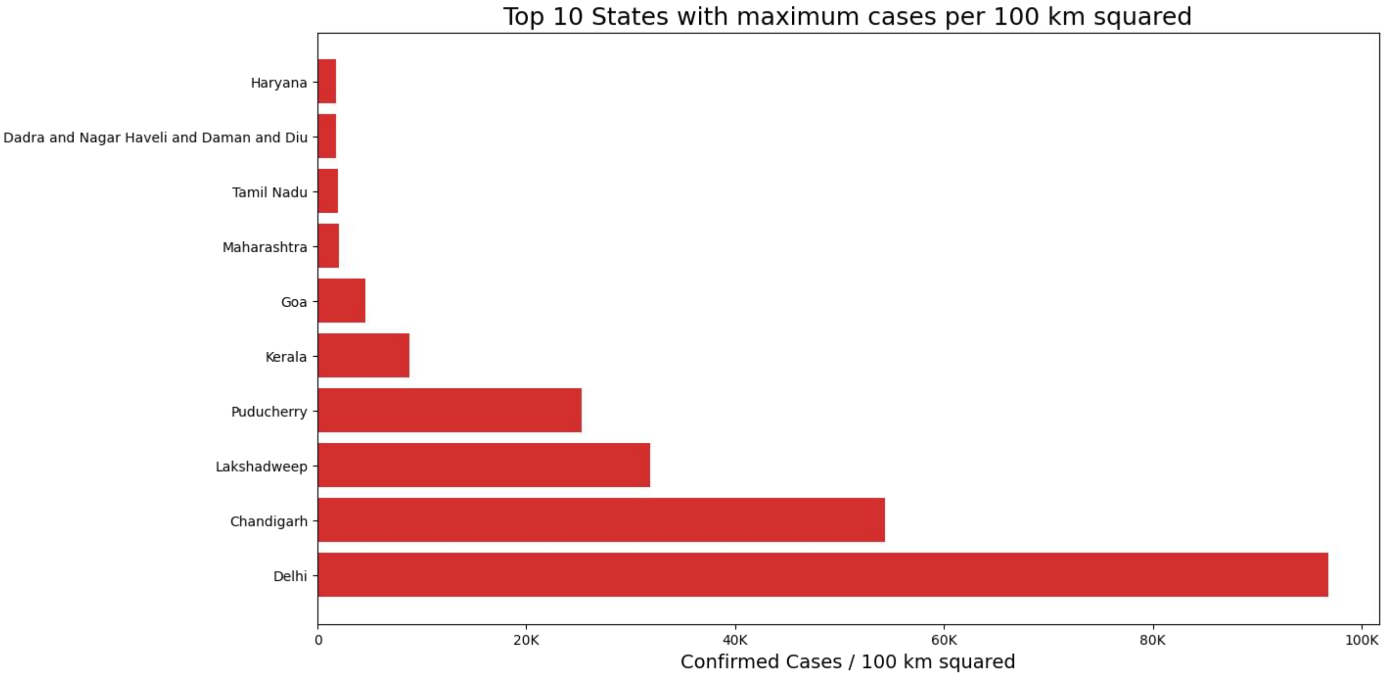
Top 10 States with maximum recoveries
Odisha
Chhattisgarh
Delhi
West Bengal
Uttar Pradesh
Andhra Pradesh
Tamil Nadu
Karnataka
Kerala
Maharashtra
0
1M
2M
3M
4M
5M
6M
Number of Recoveries

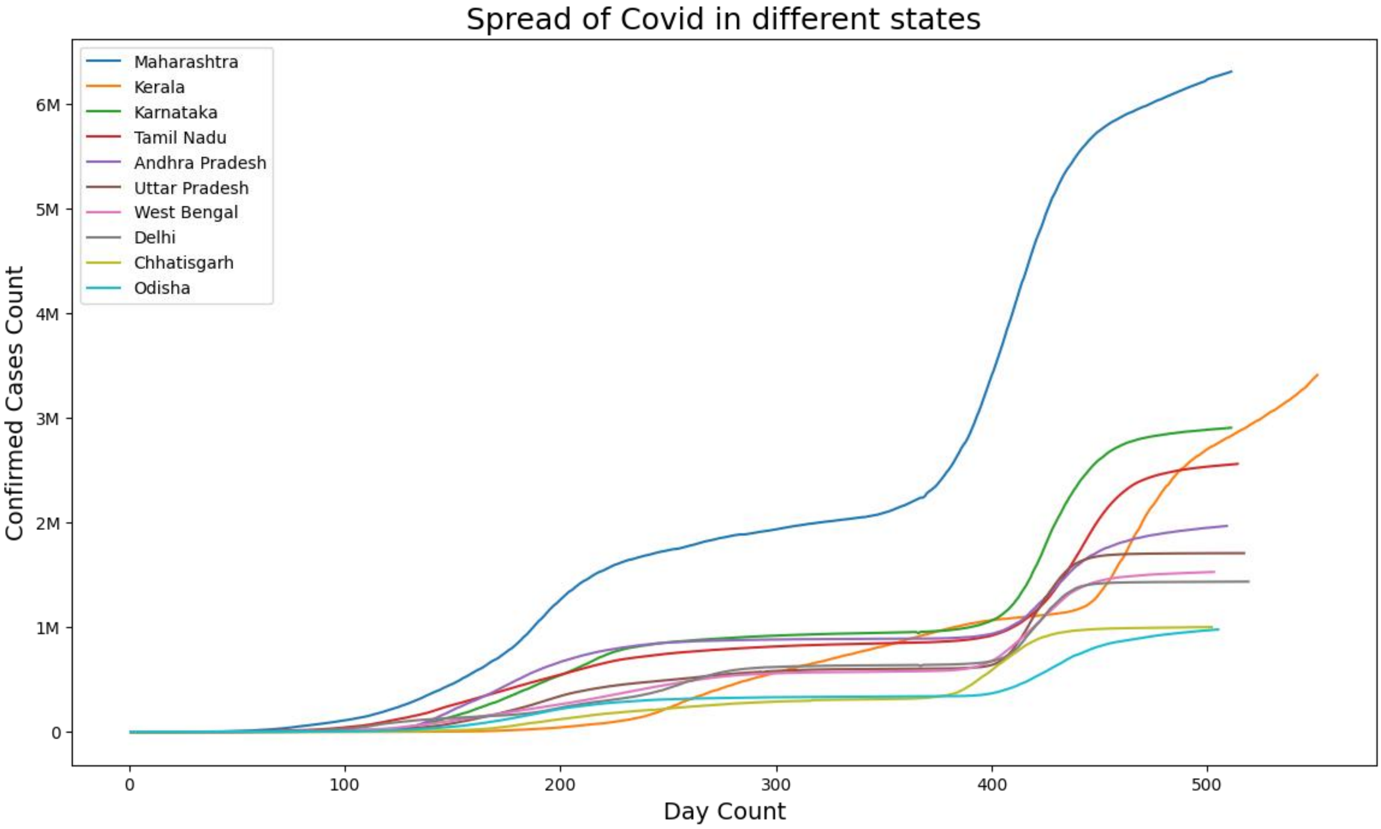


Top 10 States with maximum Recovery Rate
Bihar
Uttar Pradesh
Chandigarh
Haryana
Madhya Pradesh
Ladakh
Gujarat
Lakshadweep
Rajasthan
Dadra and Nagar Haveli and Daman and Diu
0.986
0.988
0.990
0.992
0.994
0.996
0.998
1.000
Recovery Rate


Top 10 States with maximum cases per 10 million population
Karnataka
Sikkim
Chandigarh
Maharashtra
Delhi
Ladakh
Puducherry
Goa
Lakshadweep
Kerala
0
2M
4M
6M
8M
Confirmed Cases / 10 million people


Variation in spread of Covid-19 in different Indian states
Kerala
Lakshadweep
Goa
Delhi
Chhattisgarh
West Bengal
Uttar Pradesh
Andhra Pradesh
Tamil Nadu
Karnataka
Maharashtra
0
2M
4M
6M
8M
Number of cases per 10 million people
0
1M
2M
3M
4M
5M
6M
Number of confirmed Cases


Top 10 States with maximum cases per 100 km squared
Haryana
Dadra and Nagar Haveli and Daman and Diu
Tamil Nadu
Maharashtra
Goa
Kerala
Puducherry
Lakshadweep
Chandigarh
Delhi
0
20K
40K
60K
80K
100K
Confirmed Cases / 100 km squared


Spread of Covid in different states
Maharashtra
Kerala
Karnataka
Tamil Nadu
Andhra Pradesh
Uttar Pradesh
West Bengal
Delhi
Chhatisgarh
Odisha
Confirmed Cases Count
0
1M
2M
3M
4M
5M
6M
0
100
200
300
400
500
Day Count

CONCLUSION

In conclusion, the analysis of the report suggests that the global fashion industry is rapidly growing, with emerging markets such as Asia and Africa contributing significantly to this growth. However, the industry still faces challenges related to sustainability and ethical practices, which need to be addressed to ensure a more responsible and long-term growth. Moreover, the COVID-19 pandemic has had a significant impact on the industry, causing disruptions in the supply chain and changing consumer behavior, which will require businesses to adapt to the new normal. Overall, the fashion industry needs to prioritize sustainability, innovation, and adaptability to thrive in the current landscape and ensure a sustainable future.