



AROWWAI INDUSTRIES

CLASS 12 IP PROJECT

COVID-19 DATA ANALYSIS

Prepared by :
Kanishka Sarangi
Class 12 A
Roll No. 42

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all those who have helped and supported me in completing this project on COVID-19 Data Analysis.

First and foremost, I would like to thank my project supervisor, for providing me with invaluable guidance and encouragement throughout the course of this project. Their support, insights, and constructive feedback helped shape this report into its final form.

I would also like to extend my heartfelt thanks to Sanjeev sir, for their continuous support and inspiration during my studies, which laid the foundation for my understanding of the subject and this project.

I would like to acknowledge the contribution of various online resources, such as Databases and websites, especially **Kaggle, World Health Organization, And Johns Hopkins University** for their sample data on India.

CERTIFICATE

This is to certify that

Kanishka Sarangi, a student of 12th Standard at Firayalal Public School, has successfully completed the project titled "COVID-19 Data Analysis" as a part of their academic requirements.

The project demonstrates a thorough analysis of COVID-19 data, reflecting the student's research skills, analytical abilities, and understanding of the topic. Issued on: 17/12/2025

Sanjeev Kumar Srivastava
Project Supervisor

REQUIREMENTS

contourpy==1.3.3

cycler==0.12.1

fonttools==4.61.0

Jinja2==3.1.6

kiwisolver==1.4.9

MarkupSafe==3.0.3

matplotlib==3.10.7

numpy==2.3.5

packaging==25.0

pandas==2.3.3

pandas-stubs==2.3.2.250926

pillow==12.0.0

pyparsing==3.2.5

python-dateutil==2.9.0.post0

pytz==2025.2

seaborn==0.13.2

six==1.17.0

tzdata==2025.2

PYTHON:main.py

```
#import the necessary packages
import pandas as pd #csv file IO
import matplotlib.pyplot as plt
import seaborn as sns
import datetime as dt
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
pd.set_option('display.width', None)

#import the csv file "covid_19_india.csv"
covid_df = pd.read_csv("covid_19_india.csv")
print("Dataset Information")
print(covid_df.info())
print("\nMissing Data Information")
print(covid_df.isna().sum())
print("\n-----")
print("To get the number of deaths, confirmed, and cured cases state wise")
covid_df['Active'] = covid_df['Confirmed'] - covid_df['Cured'] - covid_df['Deaths']
count = covid_df.groupby("State/UnionTerritory")[['Confirmed','Deaths','Cured','Active']].sum().reset_index()
print(count)
print("\n-----")
print("To get the mortality rates and recovery rates of all states and display them in a pivot table analysis")
state_df = pd.pivot_table(covid_df, values=['Confirmed', 'Deaths', 'Cured', 'Active'], index=['State/UnionTerritory'], aggfunc="max")
state_df['Recovery Rate'] = state_df['Cured']*100/state_df['Confirmed']
state_df['Mortality Rate'] = state_df['Deaths']*100/state_df['Confirmed']
state_df=state_df.sort_values(by="Confirmed",ascending=False)
print(state_df)
print("\n-----")
print("To display Top 10 states with Active cases")
top10ActiveCases=covid_df.groupby(by="State/UnionTerritory").max()[['Active','Date']].sort_values(by=['Active'],ascending=False).reset_index()
fig=plt.figure(figsize=(16, 10))
plt.title("Top 10 states with most Active Cases in India",size=22)
ax=sns.barplot(data=top10ActiveCases.iloc[:10],y='Active',x='State/UnionTerritory',linewidth=2,edgecolor='black', hue='State/UnionTerritory')
plt.xlabel("States")
plt.ylabel("Total Active Cases")
plt.show()

print("\n-----")
# print("To display the growth trend in as line graph for top 5 states ")
fig=plt.figure(figsize=(12,6))
ax=sns.lineplot(data=covid_df[covid_df['State/UnionTerritory'].isin(['Maharashtra','Karnataka','Kerala','Tamil Nadu','Uttar Pradesh'])],x='Date',y='Active',hue='State/UnionTerritory')
ax.set_title("Top 5 Affected States in India",size=16)
plt.show()

print("\n-----")
print("To show top Months of 2020 for spike in covid cases")
covid_df['Date'] = pd.to_datetime(covid_df['Date'], format="%Y-%m-%d")
covid_df_new = covid_df[covid_df['Date'].dt.year == 2020]
covid_df_new['Date'] = pd.to_datetime(covid_df_new['Date'], format="%Y-%m-%d")
covid_df_new['Month'] = covid_df_new['Date'].dt.month_name()
topmonths = covid_df_new.groupby(by='Month').max()[['Confirmed','Date']].sort_values(by=['Confirmed'],ascending=False).reset_index()
fig=plt.figure(figsize=(12,6))
ax=sns.lineplot(data=topmonths.iloc[:5],x='Month',y='Confirmed')
ax.set_title("Top 5 Affected Months in 2020",size=16)
plt.show()

print("\n-----")
print("To show top Months of 2021 for spike in covid cases")
covid_df['Date'] = pd.to_datetime(covid_df['Date'], format="%Y-%m-%d")
covid_df_new = covid_df[covid_df['Date'].dt.year == 2021]
covid_df_new['Date'] = pd.to_datetime(covid_df_new['Date'], format="%Y-%m-%d")
covid_df_new['Month'] = covid_df_new['Date'].dt.month_name()
topmonths = covid_df_new.groupby(by='Month').max()[['Confirmed','Date']].sort_values(by=['Confirmed'],ascending=False).reset_index()
fig=plt.figure(figsize=(12,6))
ax=sns.lineplot(data=topmonths.iloc[:5],x='Month',y='Confirmed')
ax.set_title("Top 5 Affected Months in 2021",size=16)
plt.show()
```

OUTPUT: CONSOLE

```
Dataset Information
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18047 entries, 0 to 18046
Data columns (total 6 columns):
#   Column                      Non-Null Count  Dtype
---  -
0   Sno                          18047 non-null  int64
1   Date                         18047 non-null  object
2   State/UnionTerritory        18047 non-null  object
3   Cured                        18047 non-null  int64
4   Deaths                      18047 non-null  int64
5   Confirmed                   18047 non-null  int64
dtypes: int64(4), object(2)
memory usage: 846.1+ KB
None
```

Missing Data Information

```
Sno          0
Date         0
State/UnionTerritory  0
Cured        0
Deaths       0
Confirmed    0
dtype: int64
```

To get the number of deaths, confirmed, and cured cases state wise

	State/UnionTerritory	Confirmed	Deaths	Cured	Active
0	Andaman and Nicobar Islands	1938498	27136	1848286	63076
1	Andhra Pradesh	392432753	2939367	370426530	19066856
2	Arunachal Pradesh	7176907	26799	6588149	561959
3	Assam	99837011	638323	92678680	6520008
4	Bihar	133662075	1112347	126525370	6024358
5	Chandigarh	10858627	147694	10117035	593898
6	Chhattisgarh	163776262	2063920	151609364	10102978
7	Dadra and Nagar Haveli	20722	8	20352	362
8	Dadra and Nagar Haveli and Daman and Diu	1938632	1014	1841750	95868
9	Daman & Diu	2	0	0	2
10	Delhi	287227765	4943294	273419887	8864584
11	Goa	28240159	447801	26027201	1765157
12	Gujarat	143420082	2219448	132487127	8713507
13	Haryana	134347285	1502799	126585342	6259144
14	Himachal Pradesh	30033289	491348	27501110	2040831
15	Himanchal Pradesh	204516	3507	200040	969
16	Jammu and Kashmir	58117726	839694	53297341	3980691

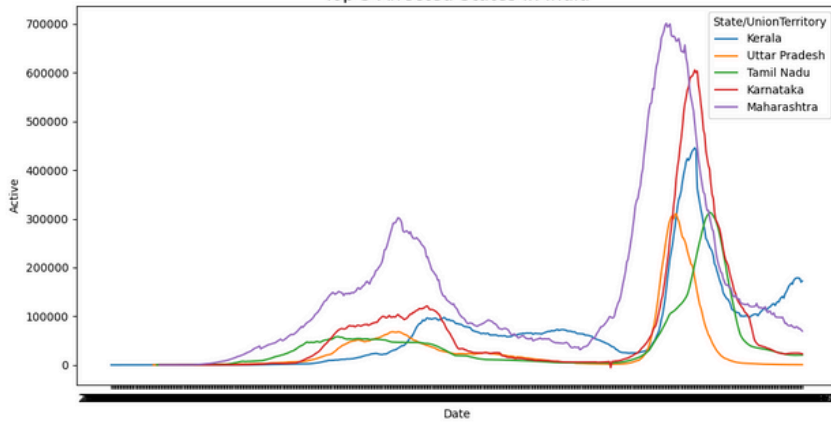
To get the mortality rates and recovery rates of all states and display them in a pivot table analysis

State/UnionTerritory	Active	Confirmed	Cured	Deaths	Recovery Rate	Mortality Rate
Maharashtra	701614	6363442	6159676	134201	96.797865	2.108937
Kerala	445692	3586693	3396184	18004	94.688450	0.501967
Karnataka	605515	2921049	2861499	36848	97.961349	1.261465
Tamil Nadu	313048	2579130	2524400	34367	97.877967	1.332504
Andhra Pradesh	211554	1985182	1952736	13564	98.365591	0.683262
Uttar Pradesh	310783	1708812	1685492	22775	98.635309	1.332797
West Bengal	132181	1534999	1506532	18252	98.145471	1.189056
Delhi	103424	1436852	1411280	25068	98.220276	1.744647
Chhattisgarh	131245	1003356	988189	13544	98.488373	1.349870
Odisha	106493	988997	972710	6565	98.353180	0.663804
Rajasthan	212753	953851	944700	8954	99.040626	0.938721
Gujarat	148297	825085	814802	10077	98.753704	1.221329
Madhya Pradesh	111366	791980	781330	10514	98.655269	1.327559
Haryana	116867	770114	759790	9652	98.659419	1.253321
Bihar	115152	725279	715352	9646	98.631285	1.329971
Telangana	80695	650353	638410	3831	98.163613	0.589065
Punjab	79963	599573	582791	16322	97.201008	2.722271
Assam	56295	576149	559684	5420	97.142232	0.940729
Telengana	78888	443360	362160	2312	81.685312	0.521472
Jharkhand	61195	347440	342102	5130	98.463620	1.476514
Uttarakhand	80000	342462	334650	7368	97.718871	2.151480

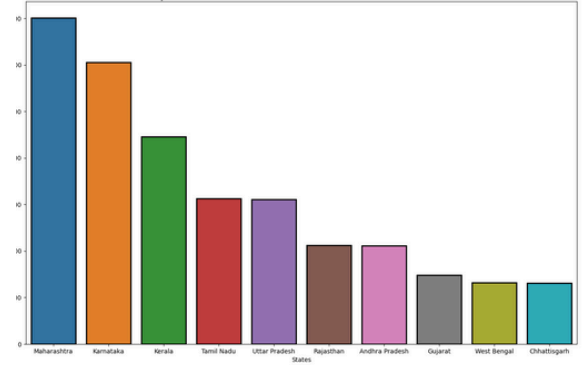
Himachal Pradesh	40008	208616	202761	3537	97.193408	1.695460
Himanchal Pradesh	969	204516	200040	3507	97.811418	1.714780
Goa	32953	172085	167978	3164	97.613389	1.838626
Puducherry	18277	121766	119115	1800	97.822873	1.478245
Manipur	10922	105424	96776	1664	91.796934	1.578388
Tripura	8302	80660	77811	773	96.467890	0.958344
Meghalaya	8255	69769	64157	1185	91.956313	1.698462
Chandigarh	8653	61992	61150	811	98.641760	1.308233
Arunachal Pradesh	4465	50605	47821	248	94.498567	0.490070
Mizoram	13101	46320	33722	171	72.802245	0.369171
Nagaland	5049	28811	26852	585	93.200514	2.030474
Sikkim	4306	28018	25095	356	89.567421	1.270612
Ladakh	2041	20411	20130	207	98.623291	1.014159
Dadra and Nagar Haveli and Daman and Diu	2081	10654	10646	4	99.924911	0.037545
Dadra and Nagar Haveli	250	10377	10261	4	98.882143	0.038547
Lakshadweep	2320	10263	10165	51	99.045114	0.496931
Andaman and Nicobar Islands	1154	7548	7412	129	98.198198	1.709062

OUTPUT: PLOTS

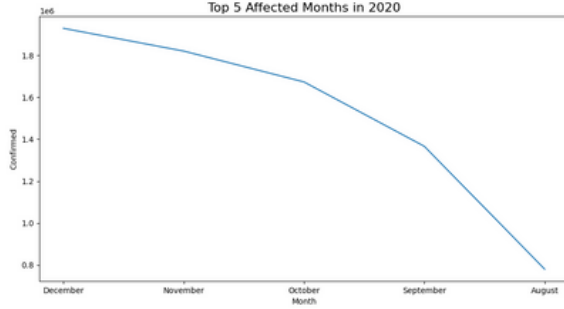
Top 5 Affected States in India



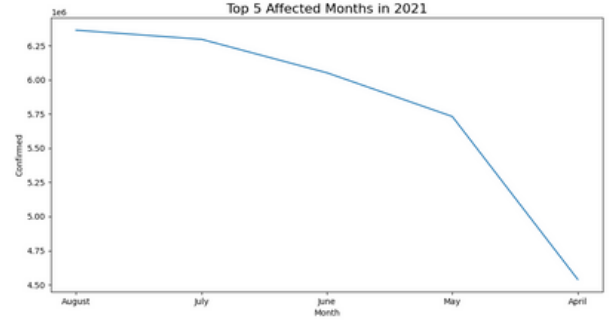
Top 10 states with most Active Cases in India



Top 5 Affected Months in 2020



Top 5 Affected Months in 2021



CITATIONS

- [Kaggle Covid-19 India \(Sample Data from Johns Hopkins University\)](#) (Edited to fit Use Case)
- [World Health Organization](#)
- [Matplotlib Documentation](#)
- [Pandas Documentation](#)
- [Project Code Link](#)