DATA ANALYTICS ON COVID-19 DATASET USING HIVE

JISHNU P
CB.EN.P2AID19017

INTRODUCTION

- The covid-19 pandemic is the greatest humanitarian challenge the world has faced since world war II.
- Analyzing the situation and taking proper countermeasure is important to reduce the losses due to this covid-19 Pandemic.
- Data analytics can be performed on different datasets using HIVE regardless of its size

OBJECTIVE

- To perform data analytics on the COVID-19 dataset using Hive.
- To visualize the worldwide Impact of COVID-19

WHY HIVE?

- Hive is an ETL and Data warehousing tool developed on top of Hadoop Distributed File System (HDFS).
- Hive makes job easy for performing Analysis of huge datasets.
- HQL syntax is similar to the SQL syntax.

DATASET DESCRIPTION

- Downloaded from Kaggle
- Covid_19_data.csv
- Semi- structured data
- 21220 Rows
- 8 Columns

A	B	C	D	E	F	G	H	I
Sno	Observatior	Province_or_State	Country_or_Region	Last_Update	Confirmed	Deaths	Recovered	
1	1/22/2020	Anhui	Mainland China	1/22/2020 17:00	1	0	0	
2	1/22/2020	Beijing	Mainland China	1/22/2020 17:00	14	0	0	
3	1/22/2020	Chongqing	Mainland China	1/22/2020 17:00	6	0	0	
4	1/22/2020	Fujian	Mainland China	1/22/2020 17:00	1	0	0	
5	1/22/2020	Gansu	Mainland China	1/22/2020 17:00	0	0	0	
6	1/22/2020	Guangdong	Mainland China	1/22/2020 17:00	26	0	0	
7	1/22/2020	Guangxi	Mainland China	1/22/2020 17:00	2	0	0	
8	1/22/2020	Guizhou	Mainland China	1/22/2020 17:00	1	0	0	

INFERENCES

- it is clear that the Italy was badly affected by COVID 19 in this period (Q11) [28884 deaths as of 05/03/2020]

- Most of the European countries were in a helpless situation because Death Rate >=10* recovery rate.[Q17] It also indicates that the social distancing was not efficiently happened in these Regions.

- In May, COVID spreading started in US.Large number of cases where reported in this month.(Q19).

INFERENCES

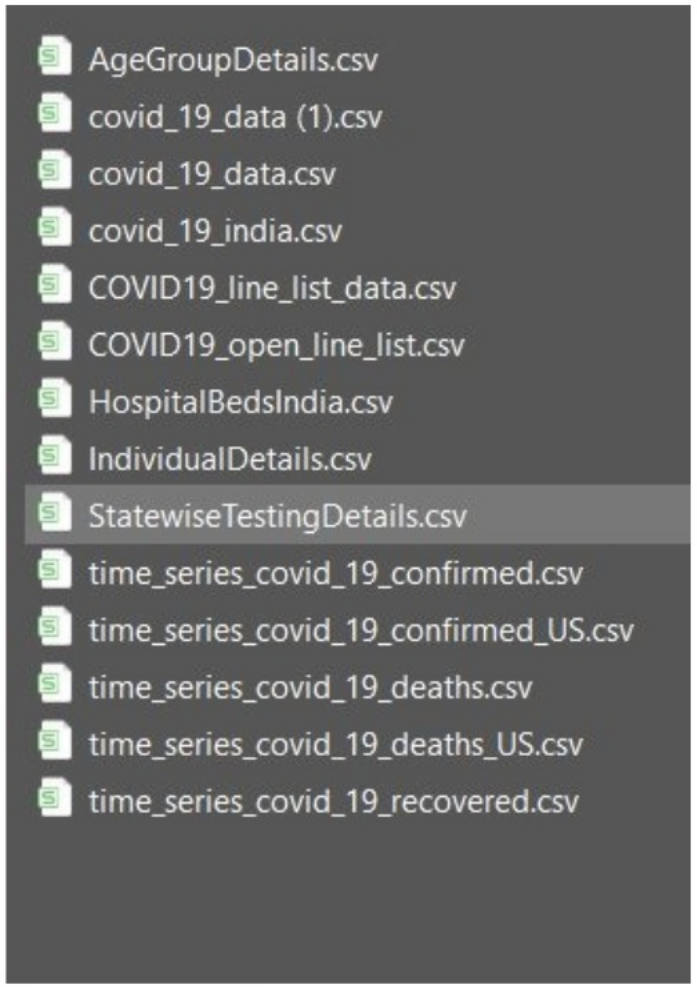
- In this period, Large number of deaths Reported in ITALY ,UK.
  It was the 3rd stage of COVID_19 pandemic in these countries.(Q18)

- INDIA is no where in the TOP 5 list of countries with highest Deaths, Confirmed cases.

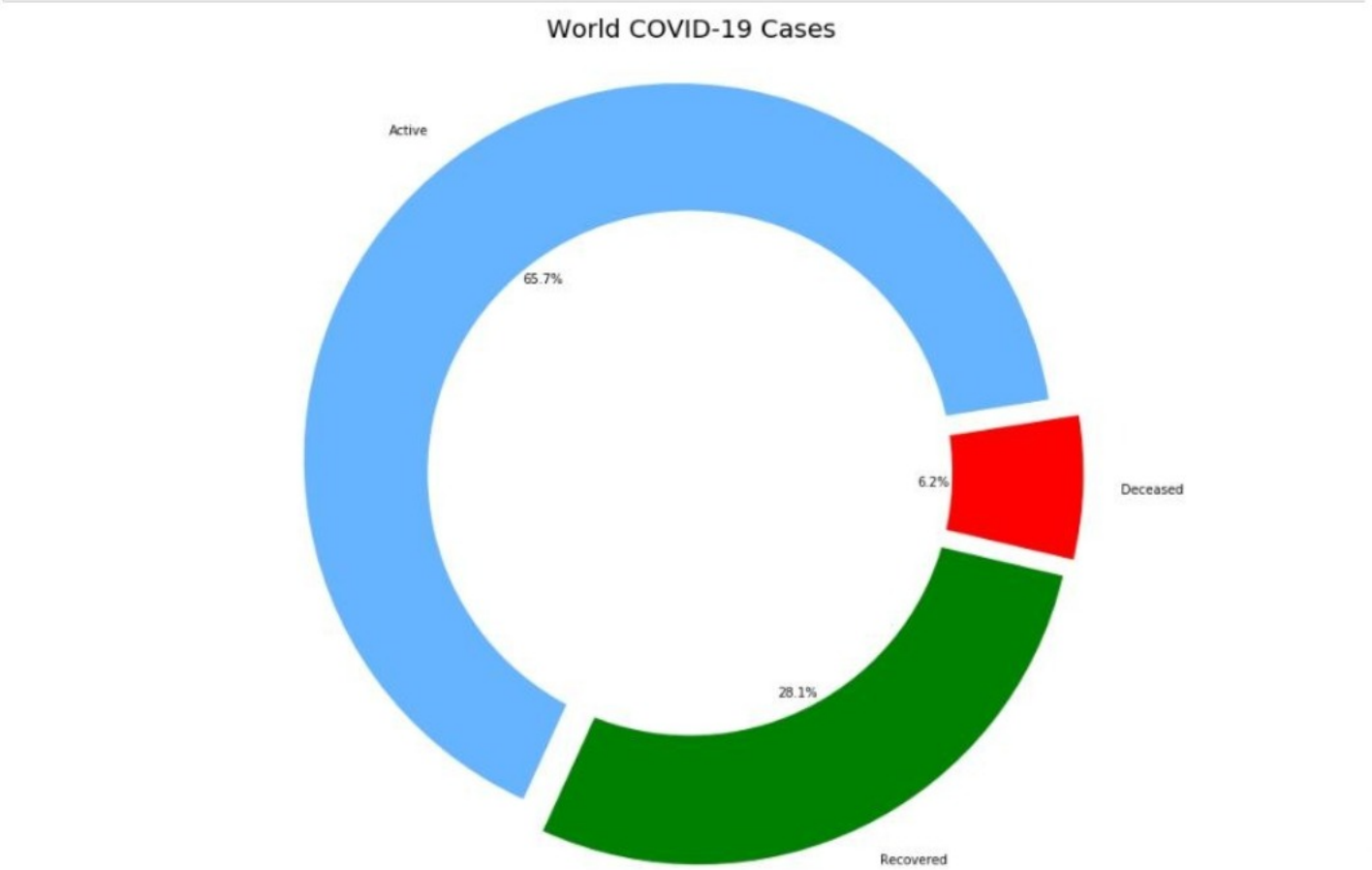
- IN this period LOCK DOWN in INDIA was very helpful to achieve social diatancing and there by reduce the spread.

VISUALIZATION

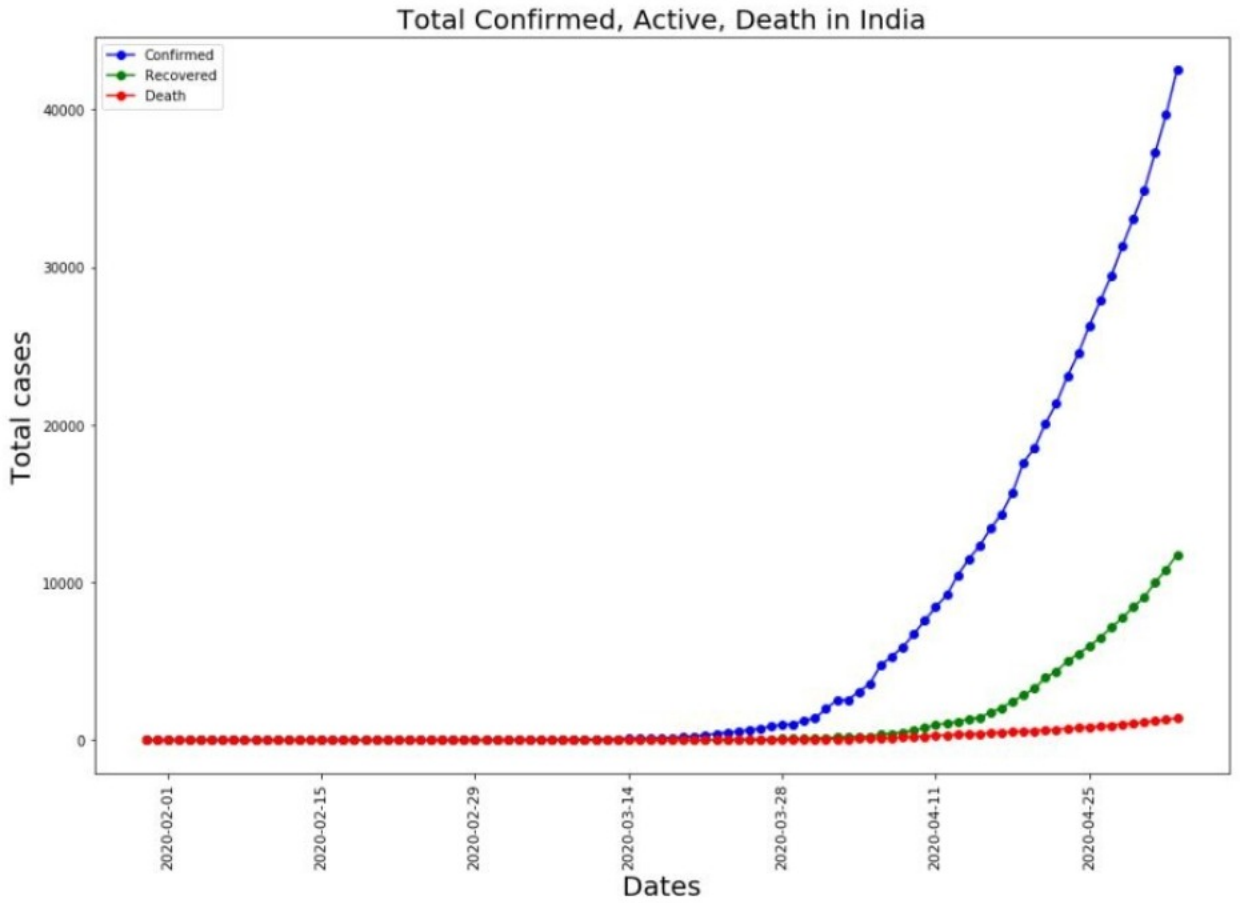
- Datasets



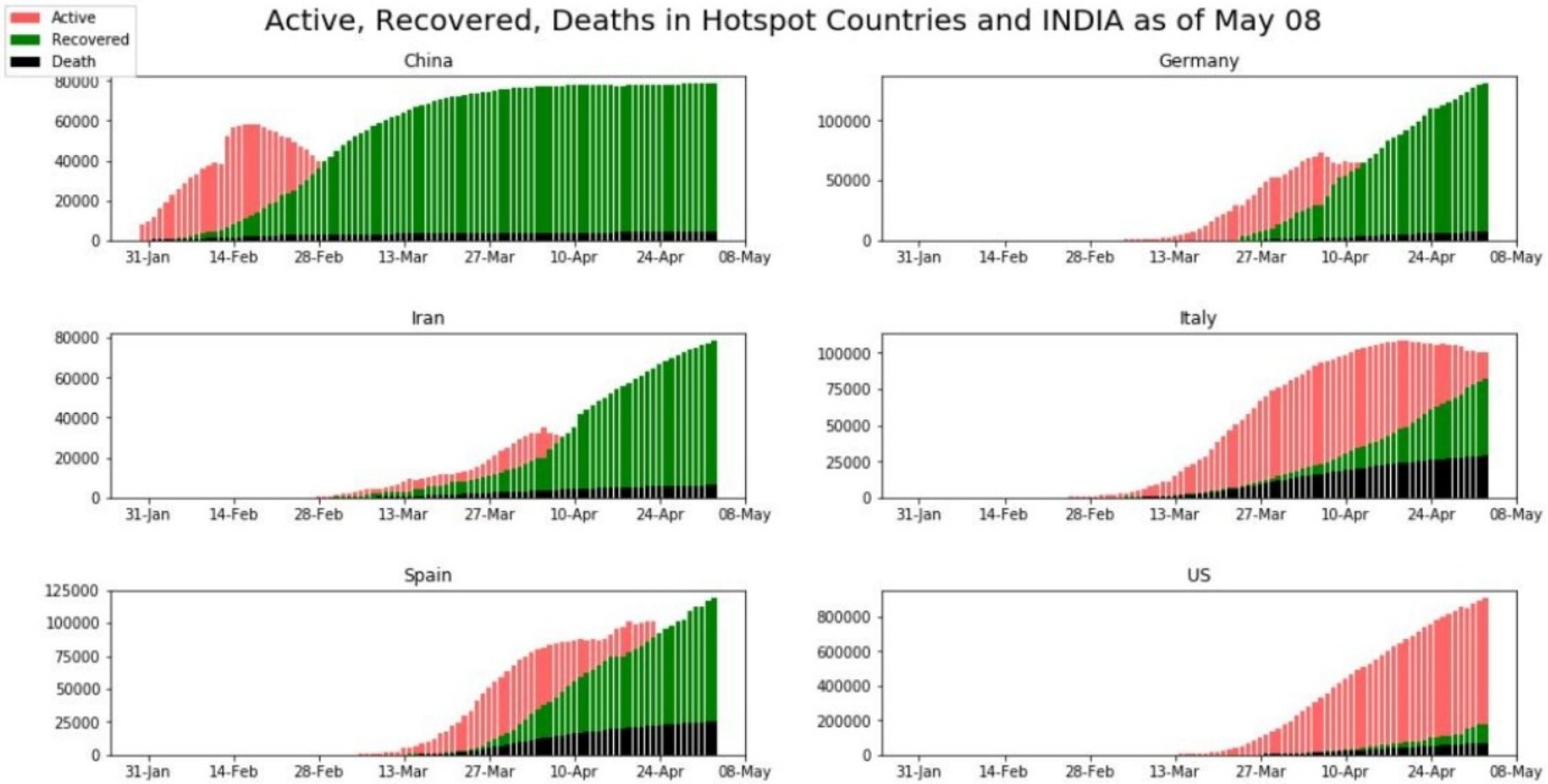
VISUALIZATION



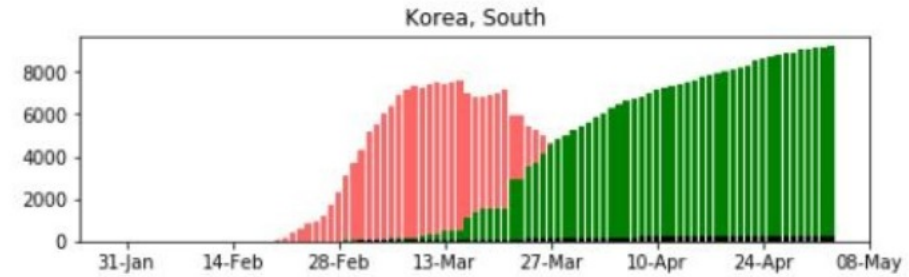
SITUATION IN INDIA

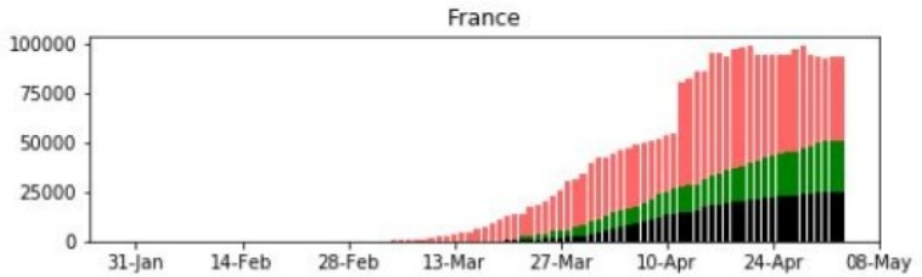


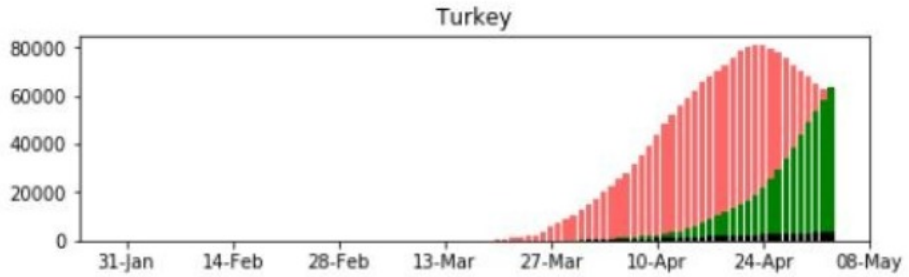
COMPARISON AMONG COUNTRIES

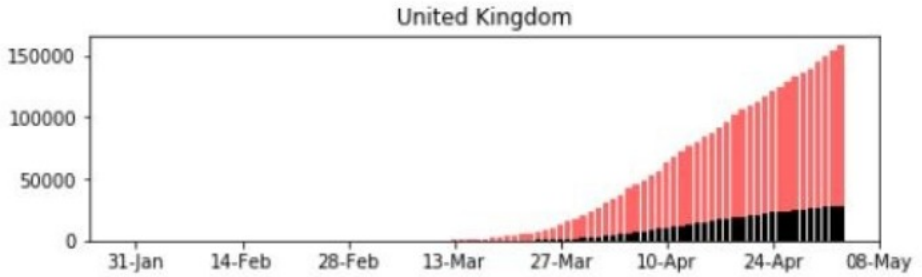


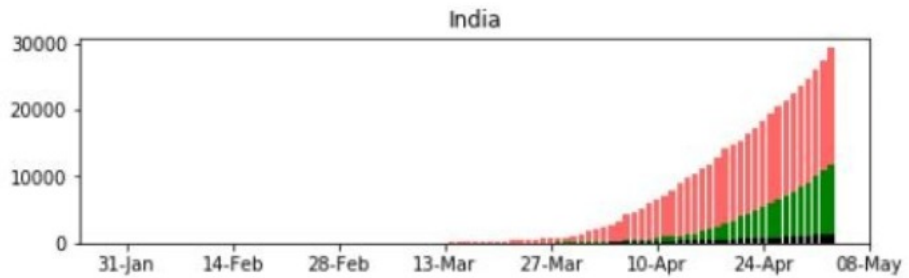
COMPARISON AMONG COUNTRIES



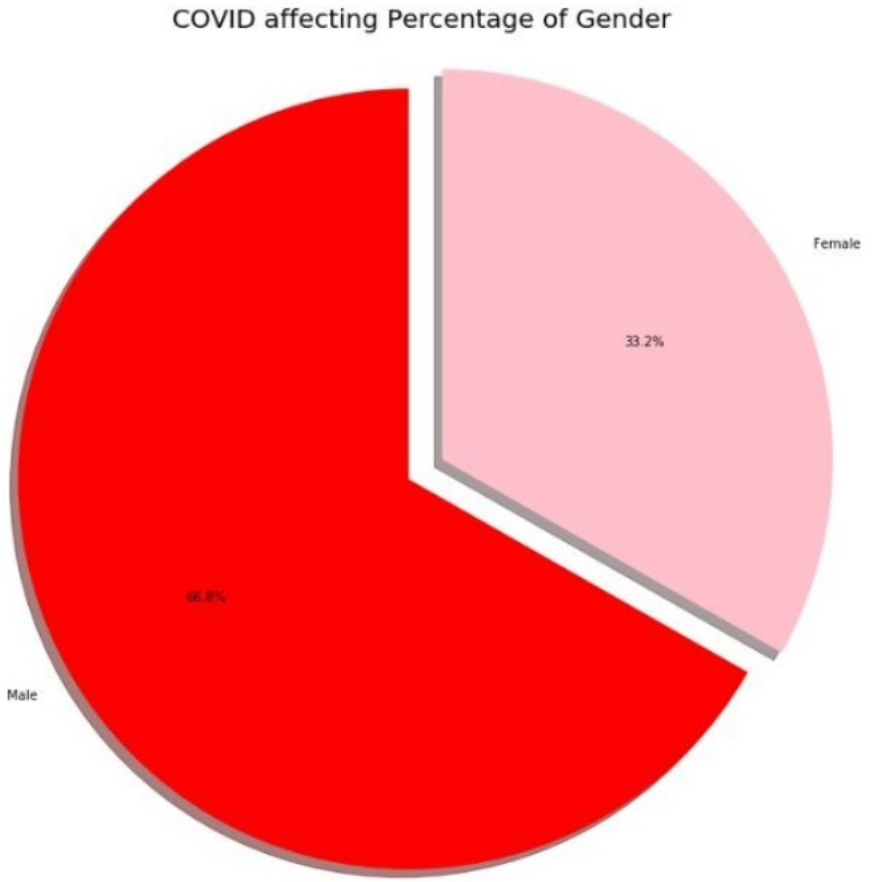




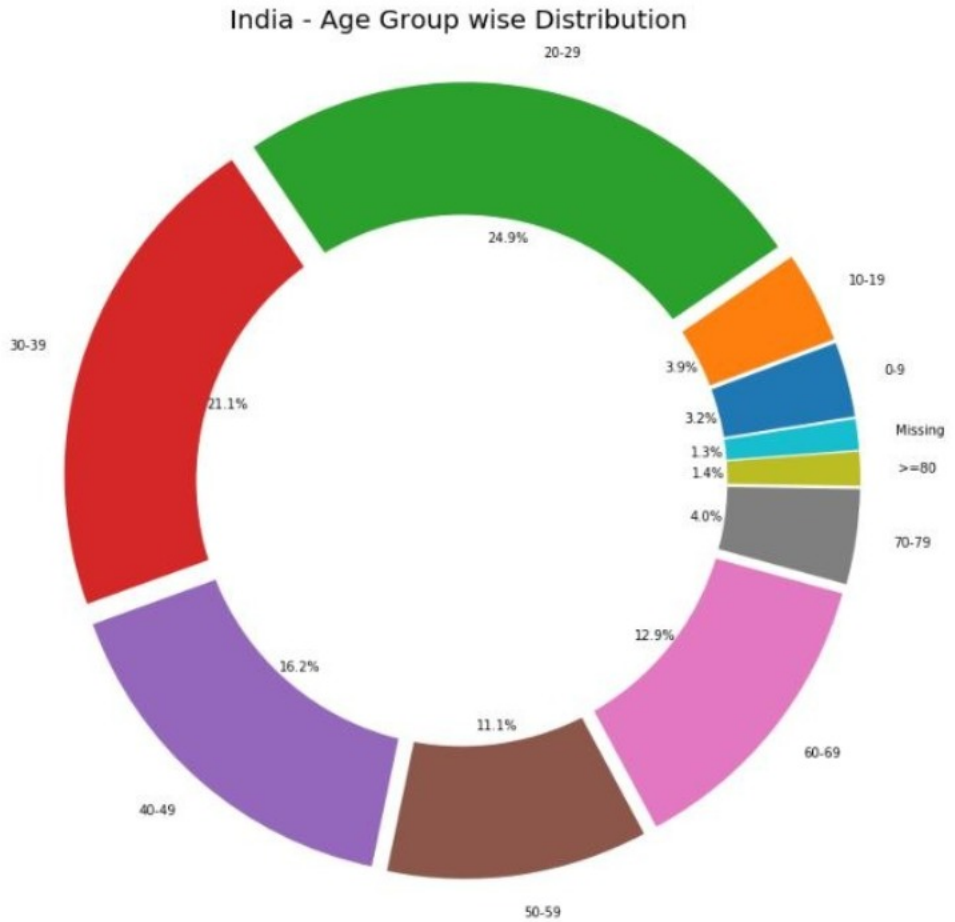


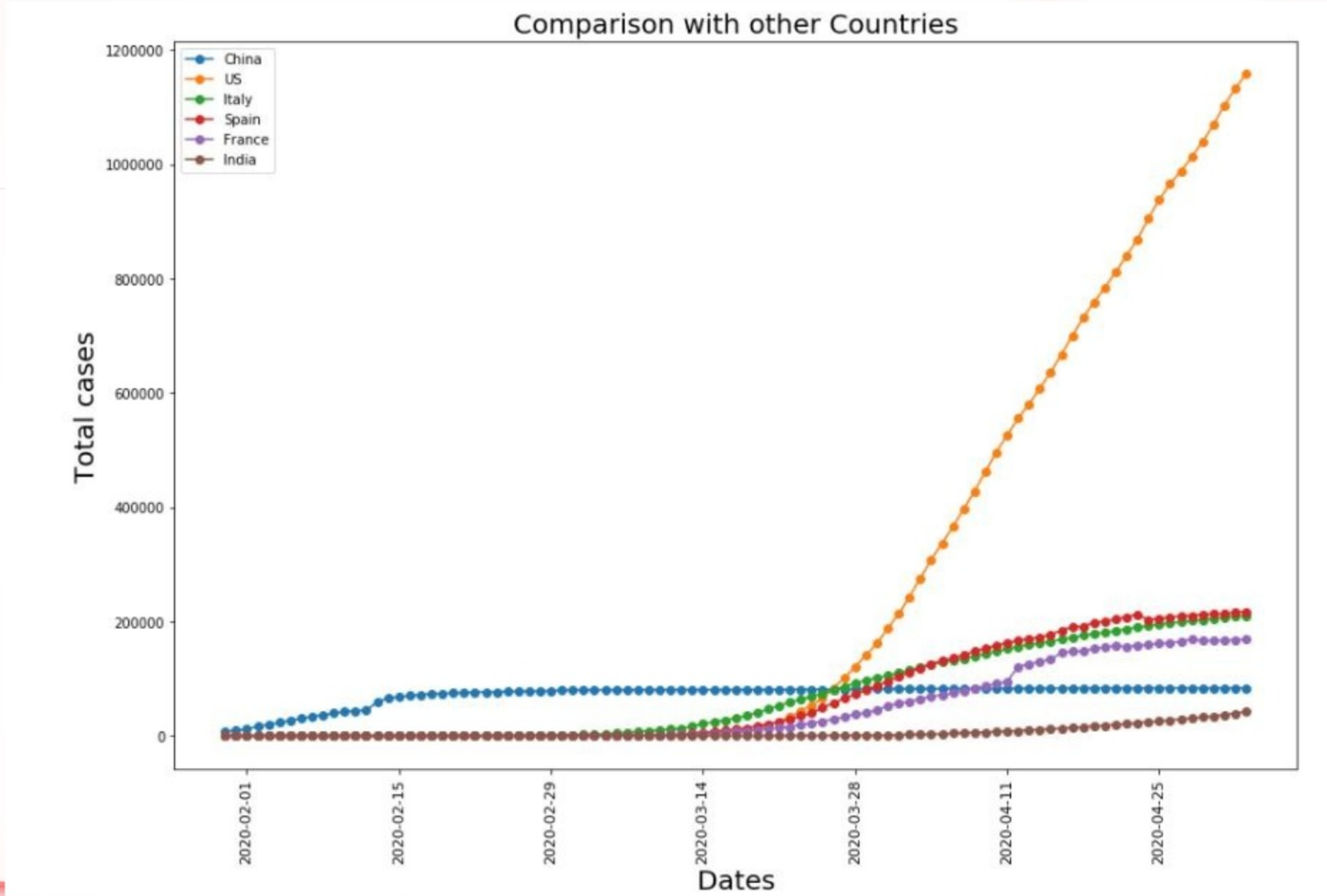


GENDERWISE CASES



AGE GROUPWISE DISTRIBUTION INDIA




Comparison with other Countries
China
US
Italy
Spain
France
India
1200000
1000000
800000
600000
400000
200000
0
Total cases
2020-02-01
2020-02-15
2020-02-29
2020-03-14
2020-03-28
2020-04-11
2020-04-25
Dates

Statewise records in India

	State/UnionTerritory	Confirmed	Deaths	Cured	Active	Death Rate (per 100)	Cure Rate (per 100)
20	Maharashtra	25922	975	5547	19400	3.76	21.4
10	Gujarat	9267	566	3562	5139	6.11	38.44
30	Tamil Nadu	9227	64	2176	6987	0.69	23.58
8	Delhi	7998	106	2858	5034	1.33	35.73
29	Rajasthan	4328	121	2459	1748	2.8	56.82
19	Madhya Pradesh	4173	232	2004	1937	5.56	48.02
34	Uttar Pradesh	3729	83	1902	1744	2.23	51.01
36	West Bengal	2290	207	702	1381	9.04	30.66
1	Andhra Pradesh	2137	47	1142	948	2.2	53.44
28	Punjab	1924	32	200	1692	1.66	10.4
31	Telengana	1367	34	940	393	2.49	68.76
13	Jammu and Kashmir	971	11	466	494	1.13	47.99
16	Karnataka	959	33	451	475	3.44	47.03
4	Bihar	940	7	388	545	0.74	41.28
11	Haryana	793	11	418	364	1.39	52.71
26	Odisha	538	3	143	392	0.56	26.58
17	Kerala	534	4	490	40	0.75	91.76
5	Chandigarh	187	3	28	156	1.6	14.97
14	Jharkhand	173	3	79	91	1.73	45.66
32	Tripura	155	0	16	139	0	10.32
3	Assam	80	2	39	39	2.5	48.75

Text

3	Assam	80	2	39	39	2.5	48.75
33	Unassigned	77	0	0	77	0	0
35	Uttarakhand	72	1	46	25	1.39	63.89
12	Himachal Pradesh	66	2	39	25	3.03	59.09
6	Chhattisgarh	59	0	55	4	0	93.22
15	Jharkhand#	45	3	0	42	6.67	0
18	Ladakh	43	0	22	21	0	51.16
0	Andaman and Nicobar Islands	33	0	33	0	0	100
27	Puducherry	13	1	9	3	7.69	69.23
22	Meghalaya	13	1	10	2	7.69	76.92
9	Goa	7	0	7	0	0	100
21	Manipur	2	0	2	0	0	100
24	Nagaland	1	0	0	1	0	0
23	Mizoram	1	0	1	0	0	100
7	Dadar Nagar Haveli	1	0	0	1	0	0
2	Arunachal Pradesh	1	0	1	0	0	100
25	Nagaland#	0	0	0	0	0	0

FUTURE SCOPE

- Predicting the COVID-19 report of a country for Next 7 days.

- Developing a recommender system to recommend countermeasures taken by other Countries where the entire situation is under control.

STAY HOME ! STAY SAFE !