|  |  |  |  |
| --- | --- | --- | --- |
|  | IIT ISM DHANBAD Dhanbad, Jharkhand | SPIE Annual Report |  |
| PROJECT rePOrT | | | | |
| Project Summary | | | | |
|  | | | | |

|  |  |  |
| --- | --- | --- |
| Report Date | Project Name | Prepared By |
| Date 06-07-2020 | Project Covid 19 Data Analysis using  Python | Name Akash |

|  |
| --- |
| Status Summary |
|  |

Successfully completed data analysis of **Covid 19** with the data of Covid confirmed in various countries from its starting phase to June and data of happiness report by **Sustainable Development Solutions Network** for the United Nations. In the analysis, I checked if there is any relation between Covid spread and the happiness indices. I merged the required columns of both the tables and checked the correlation of the Covid infection rate and the rest parameters of happiness. After that, plotted the graphs to visualize the results. Used Python for this analysis and its libraries viz. pandas, numpy, matplotlib.pyplot, seaborn.

|  |
| --- |
| Project Overview |
|  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| task | % Done | Due date | driver | notes |
| **Imported Covid 19 dataset, removed unwanted columns, aggregated the rows by the country** | **100%** | **05-07-2020** | **Used panda read\_csv method, drop(), groupby(--).sum(), loc[--].plot(), diff().max(), legend()** | **Visualized the country wise data and added a new column with max infection rate and created a new DataFrame with only indices and max infection rate column** |
| **Imported World happiness report, removed unwanted columns, changed indices to countries to merge easily** | **100%** | **05-07-2020** | **Used panda read\_csv method, drop(), set\_index()** |  |
| **Joined the two datasets we have prepared and created a correction matrix** | **100%** | **06-07-2020** | **Used join(---, how=’inner’), corr()** |  |
| **Visualized the data by plotting graphs between max infection rate and rest of the columns one by one** | **100%** | **06-07-2020** | **Used scatterplot(), regplot(), np.log()** | **Used logarithm of max infection rate in y axis to better visualize the plotting and result.** |

|  |
| --- |
| CONCLUSIONS/RECOMMENDATIONS |
|  |

After the analysis, found that the max. infection rate of each country has positive coeff. of correlation with each of the happiness parameters. After this result, I got on a conclusion that developed countries have higher Covid infection rate and this might be because of following two reasons:

1. People in developed countries have more international exposure and visits and hence might get infected from other infected countries very quickly.
2. Number of testing is much more in developed countries and that’s why getting more no. infected patients daily.