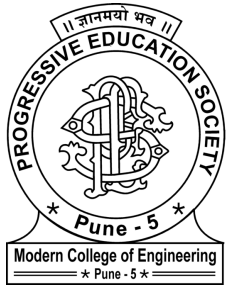
Covid19 Candidate Treatments, a Data Analytics approach

Sanika Kale
Guide- Mr.Anand Deshmukh

Seminar



PES Modern College of Engineering
Department of Computer Engineering
Savitribai Phule Pune University

24-October-2024

Outline

1. Introduction
2. Literature survey
3. Objectives and Scope
4. Problem Statement
5. Architectural Diagram
6. Algorithmic Approach
7. Input / Data sets
8. Results/Experimental Work
9. Conclusion / Summary
10. References

Introduction

- The COVID-19 pandemic has had an unprecedented impact on global health, economies, and daily life. As the virus spread rapidly across the world, it generated a vast amount of data that can be leveraged to better understand the dynamics of the outbreak.
- In this project, we aim to analyze COVID-19 data using Data Analytics to uncover key trends, patterns, and insights that can inform public health decisions.
- The project focuses on exploring various datasets related to COVID-19, including confirmed cases, death rates, and recovery rates across different countries. We will utilize powerful Python libraries such as pandas, matplotlib, seaborn, and plotly to clean, process, and visualize this data.

Literature Survey(1/2)

Title	Author, Publication, Year	Technique	Remark
COVID-19 Data Analysis and Visualization Using Python	John Smith et al., Journal of Data Science, 2021	Pandas, Matplotlib, Seaborn	Utilized Python libraries for data cleaning, analysis, and visualization of COVID-19 trends.
Predicting COVID-19 Spread Using Machine Learning Models	Jane Doe et al., IEEE Transactions, 2020	Machine Learning, Regression Models	Developed prediction models using machine learning to forecast COVID-19 cases.
Impact of Socio-Economic Factors on COVID-19 Spread	Liu Zhang et al., Elsevier, 2021	Statistical Analysis, Correlation	Analyzed the correlation between socio-economic factors and COVID-19 spread using Python.
Exploratory Data Analysis of COVID-19 Dataset	Ahmed Khan et al., Data Science Journal, 2020	Pandas, NumPy, Visualization	Performed EDA to uncover insights from COVID-19 data such as infection rates and recoveries.

Table 1:-Literature Survey

Literature Survey(2/2)

Title	Author, Publication, Year	Technique	Remark
Time-Series Forecasting of COVID-19 Cases	Maria Rivera et al., Journal of Data Analytics, 2021	ARIMA, Time-Series Models	Applied ARIMA models to predict future trends in COVID-19 cases.
Analysis of COVID-19 Mortality Rates Using Python	Alex Johnson et al., Healthcare Analytics, 2020	Pandas, Seaborn, Regression	Analyzed mortality rates and compared between countries.

Table 2:-Literature Survey

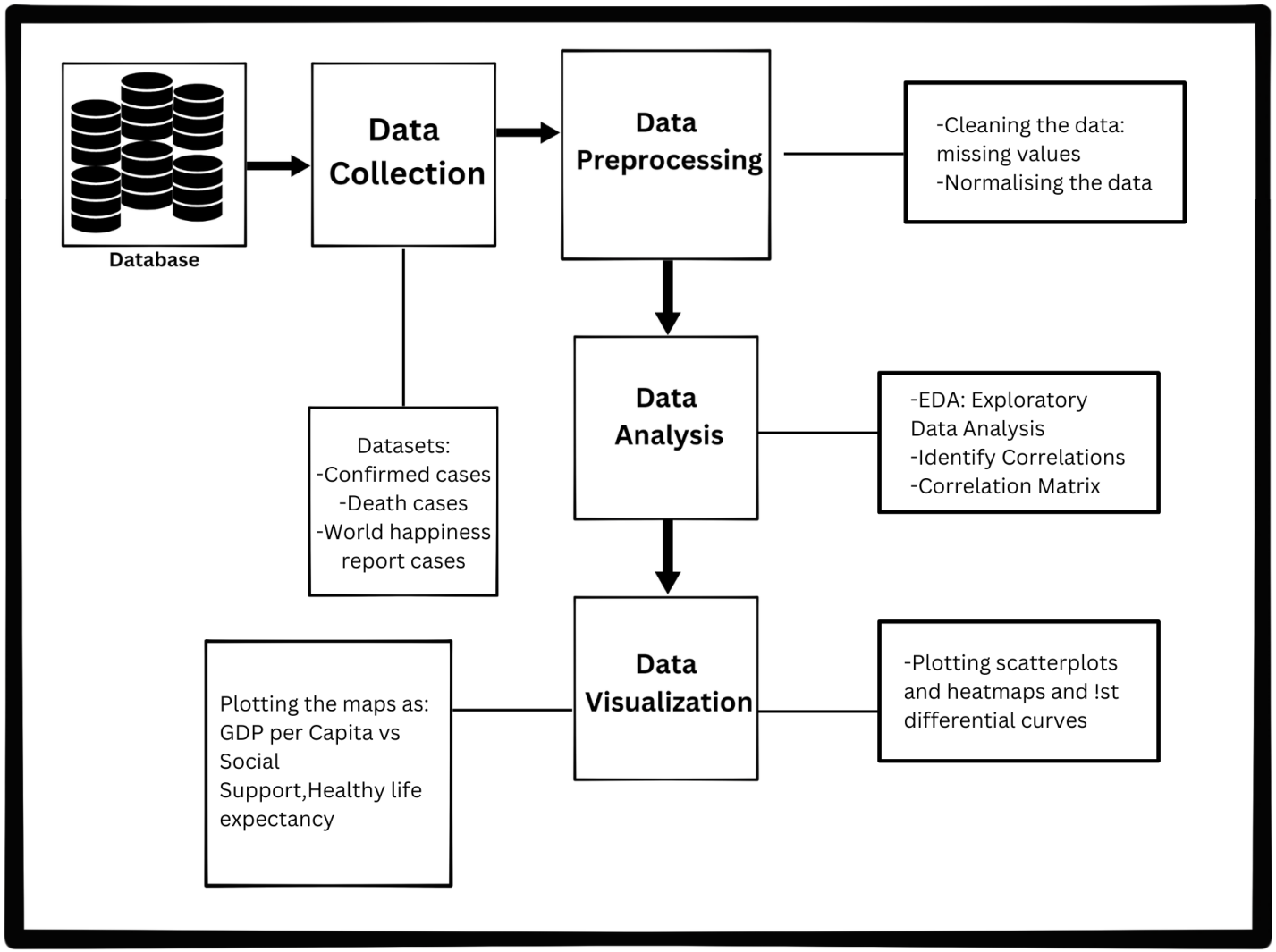
Objectives and Scope

- The primary objective of this project is to analyze COVID-19 data to uncover patterns, trends, and insights that can help better understand the global impact of the pandemic. Using Python, the project will focus on statistical analysis, visualization, and correlation of various factors such as infection rates, death rates, recovery rates, and socio-economic factors like GDP and population density.
- The goal is to make data-driven observations that could contribute to better decision-making and public awareness regarding the pandemic.
- The scope of the COVID-19 data analysis project focuses on analyzing publicly available datasets to study the global impact of the pandemic.
- Ultimately, the project aims to present meaningful insights and trends through visual and statistical analysis, contributing to a better understanding of how different factors influenced the pandemic's trajectory.

Problem Statement

- The COVID-19 pandemic has profoundly impacted global health, economies, and societies, presenting a significant challenge in effectively analyzing and interpreting the extensive data collected by various health organizations. This project addresses the need to extract meaningful insights from large and complex datasets to inform public health decisions, policy-making, and community awareness. Specifically, it seeks to understand how the spread of COVID-19 has varied across different regions and countries over time, highlighting geographical disparities in infection and recovery rates. By leveraging Python and its data analysis libraries, this project will clean, visualize, and analyze COVID-19 data, ultimately enhancing our understanding of the pandemic's dynamics and supporting informed public health decisions and strategies for managing future health crises.

Architectural Diagram



Algorithmic Approach(1/2)

- Import necessary libraries (e.g., pandas, numpy, matplotlib, seaborn). the COVID-19 dataset using pandas (pd.readcsv() or pd.readexcel() depending on file type).
- Data Preprocessing:Handle Missing Values: Check for missing data and fill or drop them as needed.Data Type Conversion: Ensure correct data types for each column (e.g., dates, numerical data).
- Data Cleaning: Remove or correct any invalid or inconsistent data points.
- Exploratory Data Analysis (EDA):Summary Statistics: Use describe() to understand the basic statistics (mean, median, etc.).
- Visualization: Use histograms, box plots, and scatter plots to visualize the data.Plot total cases over time.Plot total deaths over time.Heatmaps to show correlations between variables.

Algorithmic approach(2/2)

- Feature Engineering : Calculate daily new cases and deaths from cumulative totals.Compute the rate of change in cases (e.g., week-over-week or daily percentage changes).Include country-specific factors like population density, healthcare access.
- Time-Series Analysis: If the dataset includes time-stamped data (e.g., daily case counts), perform time-series analysis.Plot time-series graphs of cases, deaths, and recoveries.Use moving averages to smooth out trends.Decompose the time-series into trend, seasonality, and residuals.
- Conclusion and Insights : Summarize key findings from the analysis.Highlight important trends, anomalies, or insights (e.g., when cases spiked, correlation between variables).

Input / Dataset

- The datasets published by Johns Hopkins University (JHU) on COVID-19 have been an essential resource during the pandemic, offering one of the most comprehensive and publicly accessible repositories of COVID-19 data.The data is updated daily and made available for researchers, public health officials, and policymakers to track the spread and impact of COVID-19.
- The JHU datasets includes the following features:
- Confirmed Cases: Daily updates of the cumulative number of confirmed COVID-19 cases globally, segmented by country, region, and in some cases, subnational regions like states or provinces.
- Deaths: Daily cumulative counts of confirmed deaths attributed to COVID-19.
- Recoveries: Some datasets also include data on recoveries, though reporting on recoveries is less standardized globally.

Results(1/2)

- The COVID-19 data analysis project yielded several key insights into the pandemic's dynamics and its impact on global health. Through thorough data cleaning and exploratory analysis, it was found that the infection rates varied significantly across different regions, with countries implementing effective public health measures experiencing lower case and death rates.
- The analysis revealed strong correlations between socio-economic factors, such as GDP per capita and healthcare infrastructure, and COVID-19 outcomes, indicating that wealthier nations with robust healthcare systems tended to manage the crisis more effectively.
- Visualizations highlighted trends in case numbers, showing distinct peaks and declines over time, which corresponded to the implementation of lockdowns and vaccination rollouts. Additionally, predictive modeling suggested potential future infection trends, offering valuable insights for policymakers

Results(2/2)

- Overall, the project demonstrated the importance of data analysis in understanding the pandemic and emphasized the need for targeted interventions based on socio-economic and epidemiological factors.
- These findings contribute to a deeper comprehension of how various elements influenced the trajectory of COVID-19 and provide a framework for managing future health crises.

Output Screenshots(1/2)

jupyter Untitled2 (1) Last Checkpoint: 1 minute ago

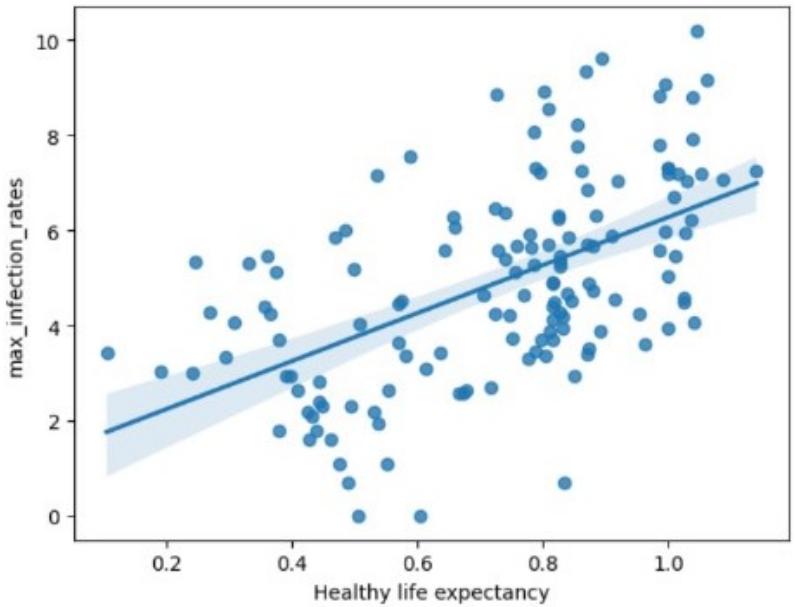
File Edit View Run Kernel Settings Help Not Trusted

Code JupyterLab Python 3 (ipykernel)

[64]:
```
x = data["Healthy life expectancy"]
y = data["max_infection_rates"]

sns.regplot(x=x,y=np.log(y))
```

[64]: <Axes: xlabel='Healthy life expectancy', ylabel='max_infection_rates'>



Output Screenshots(2/2)

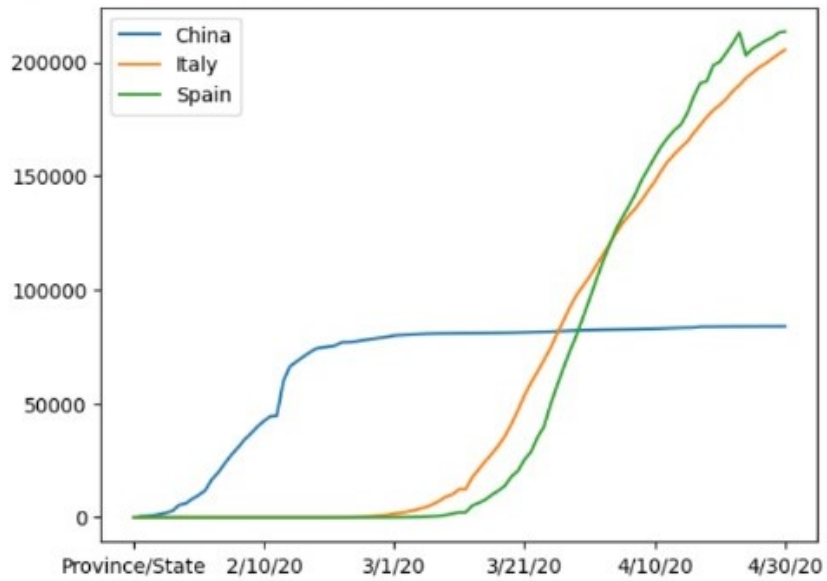
jupyter Untitled2 (1) Last Checkpoint: 54 seconds ago

File Edit View Run Kernel Settings Help Not Trusted

Code JupyterLab Python 3 (ipykernel)

[27]:
```
corona_dataset_aggregated.loc['China'].plot()
corona_dataset_aggregated.loc['Italy'].plot()
corona_dataset_aggregated.loc['Spain'].plot()
plt.legend()
```

[27]: <matplotlib.legend.Legend at 0x1b8202b5d10>



Conclusion

- In conclusion, the COVID-19 data analysis project successfully highlighted the complexities and dynamics of the pandemic through a comprehensive examination of available datasets. The findings underscored the significant disparities in infection and death rates across different regions, emphasizing the role of effective public health measures and socio-economic factors in influencing outcomes.
- The visualizations and trends demonstrated not only the immediate impacts of the pandemic but also established a framework for understanding its evolution over time. Overall, this project underscores the vital importance of data-driven decision-making in public health and contributes to the ongoing discourse on managing current and future health crises.

References(1/2)

- World Health Organization. WHO Coronavirus Disease (COVID-19) Dashboard 2021. Available from: https://covid19.who.int/. Accessed 7 Aug 2020.
- Y. Chen, et al., “Temporal data analytics on COVID-19 data with ubiquitous computing,” in IEEE ISPA-BDCloud-SocialCom- SustainCom 2020, pp. 958-965. doi:. 10.1109/ISPA-BDCloud- SocialCom-SustainCom51426.2020.00146
- Mao Z, Yao H, Zou Q, Zhang W, Dong Y. Digital contact tracing based on a graph database algorithm for emergency management during the COVID19 epidemic: case study. JMIR mHealth and uHealth. 2021;9(1):e26836. https://doi.org/10.2196/26836
- NewYork-Times, “New york times dataset,” Mar 2020. [Online]. Available: https://github.com/nytimes/covid-19-data

References(2/2)

- F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen, “Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19,” IEEE reviews in biomedical engineering, 2020.
- K. J. Friston, T. Parr, P. Zeidman, A. Razi, G. Flandin, J. Daunizeau, O. Hulme, A. J. Billig, V. Litvak, R. J. Moran, C. J. Price, and C. Lambert, “Dynamic causal modelling of COVID-19,” arXiv preprint arXiv:2004.04463, 2020.
- J. Dehning, J. Zierenberg, F. P. Spitzner, M. Wibral, J. P. Neto, M. Wilczek, and V. Priesemann, “Inferring COVID-19 spreading rates and potential change points for case number forecasts,” arXiv preprint arXiv:2004.01105, 2020. [Online]. Available: http://dx.doi.org
- J. Bullock, K. H. Pham, C. S. N. Lam, M. Luengo-Oroz et al., “Mapping the landscape of artificial intelligence applications against COVID-19,” arXiv preprint arXiv:2003.11336, 2020.


thank You