**PROJECT DOCUMENTATION AND SUBMISSION**

|  |  |
| --- | --- |
| **Date** | **01.11.2023** |
| **Team ID** | **3925** |
| **Project Name** | **COVID 19 Case Analysis** |

**Project Title: COVID 19 Case Analysis**

**Table of Content:**

|  |  |  |
| --- | --- | --- |
| **S.NO.** | **Table of Content** | **P.NO.** |
| **1.** | **PROJECT OVERVIEW** | **2** |
| **2.** | **PROBLEM STATEMENT** | **2** |
| **3.** | **PROBLEM IDENTIFIED** | **2** |
| **4.** | **INTRODUCTION** | **3** |
| **5.** | **LITERATURE SURVEY** | **3** |
| **6.** | **DESIGN THINKINH APPROACH** | **10** |
| **7.** | **STEPS INVOLVED IN MODEL EVALUATION** | **12** |
| **8.** | **TOOLS AND LIBRARIES** | **15** |
| **9.** | **MODEL SELECTION** | **16** |
| **10.** | **MODEL TRAINING AND EVALUATION** | **16** |
| **11.** | **MEAN AND STANDARD DEVIATION OF CASES AND DEATHS** | **17** |
| **12.** | **VISUALIZATION USING PYTHON CODE** | **17** |
| **13.** | **VISUALIZATION USING IBM COGNOS** | **24** |
| **14.** | **HOW INSIGHT IMPROVE USER EXPERIENCE** | **28** |
| **15.** | **CONCLUSION** | **29** |

**1.PROJECT OVERVIEW:**

* **Covid 19 Case Analysis**: Perform a comprehensive analysis of the covid 19 data to understand trends , patterns and factors influencing the spread of diesease.
* **Visualization**: Create data visualizations using IBM cognos as well as python to analyse the data and present the insights effectively.
* **Compare and Contrast Mean and Standard Deviation**: The objective is to compare and contrast the mean values and standard deviations of cases and deaths.
* **Predictive Modeling**: Develop a predictive model to estimate cases and death rate in future

**2. PROBLEM STATEMENT:**

**Objective :** Designing a project to analyze COVID-19 cases and deaths using IBM cognos, The objective is to compare and contrast the mean and standard deviation of cases and deaths, which is a valuable undertaking. This project will involve data analysis, visualization, and deriving insights from the data.

**Data:** Gather data from reliable sources, such as government health departments, research institutions, and international organizations like the World Health Organization (WHO).Collect data on the number of cases, deaths, recoveries, testing rates, vaccination rates, and other relevant variables.

**3.PROBLEM IDENTIFIED:**

The COVID-19 pandemic has been a significant global public health concern. Understanding the spread of cases and the resulting deaths is critical to managing and mitigating its effects.

There is an abundance of COVID-19 data available, including the number of cases and deaths. due to regional Variations, resource allocation, scientific understanding of diesease, policy evaluation etc Decision-makers, including policymakers and healthcare administrators, require data-based information to make informed decisions. Analyzing the data can aid in making evidence-based choices.

**4.INTRODUCTION**:

The COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2, has emerged as one of the most significant global challenges of our time. Since its inception, this infectious disease has had profound consequences on public health, economies, and daily life worldwide. As the pandemic continues to evolve, understanding the spread of COVID-19 cases and the resulting deaths is of paramount importance. This project aims to address this crucial need by leveraging the power of IBM Cognos for data analysis, visualization, and insight derivation.

IBM Cognos offers robust data analytics and reporting tools that enable us to unlock the potential of COVID-19 data. Its features allow for in-depth analysis, real-time updates, and dynamic visualizations.

**Analysing COVID 19 Trends and Patterns:**

The "Covid 19 case analysis" project is a comprehensive initiative to address the pressing global issue. The project collects, analyzes, and visualizes covid 19 data from various sources across the globe, aiming to gain insights into trends, identify areas with high case rates, and develop a predictive model for estimating future cases and deaths.

**Identifying Areas with High infection rates:**

The second objective is to identify areas that consistently experience high levels of infection rates and death rates. By geospatially analyzing the data, the project can create maps and visualizations to pinpoint hotspots areas. This spatial analysis is critical for targeted intervention and mitigation strategies. It helps authorities allocate resources and take precise actions to reduce spread in the areas most affected, thus protecting the people from danger.

**Developing a Predictive Model for cases and death rate:**

A key component of this project is the development of a predictive model. This model will use the features such as date, cases, deaths, and countries as input variables to estimate the future predictions. This predictive model will be a valuable tool for forecasting . It enables authorities to anticipate spread level and take preventive measures to protect public health. It also offers a means of early warning and preparation for residents, especially those in high-risk areas.

**5.LITERATURE SURVEY:**

**“Analysis and Prediction of COVID-19 pandemic in india”**

**Objective:**

The objective of this analysis is to assess the trajectory of the covid-19 pandemic in india, including key trends, factors influencing its spread, and potential future scenarios. this analysis aims to inform public health policies, vaccination strategies, and other interventions to control and mitigate the impact of the pandemic.

**Findings:**

India experienced multiple covid-19 waves, with the second peaking in april-may 2021. factors contributing to the severity included more transmissible variants, large gatherings, and healthcare infrastructure strain. public health measures included testing, contact tracing, social distancing, and targeted lockdowns. the effectiveness of these measures varied based on population adherence. india initiated one of the largest global vaccination campaigns, aiming to immunize a significant portion of its population, but faced challenges like vaccine supply issues, logistical hurdles, and vaccine hesitancy.

**Results:**

* Impact on healthcare system: the waves of the pandemic strained india's healthcare infrastructure, leading to shortages of hospital beds, oxygen, and medical supplies.
* Vaccination progress: india made progress in vaccinating its population, but challenges in vaccine distribution and hesitancy influenced the pace of the vaccination campaign.
* Economic and social impact: the pandemic had significant economic and social consequences, including disruptions to livelihoods, education, and overall well-being.

**Summary:**

India has experienced multiple waves of the Covid-19 pandemic, with the second being severe. Public health measures like testing, contact tracing, and lockdowns have helped mitigate the spread. However, the emergence of new variants raises concerns for future efforts, emphasizing vaccination, monitoring, and evidence-based public health measures.

# “Predicting the number of new cases of COVID-19 in India using Survival Analysis and LSTM”

# Objective:

# The objective of this study is to develop a predictive model for estimating the number of new covid-19 cases in india using a combination of survival analysis and lstm. survival analysis will be used to model the time until a specific event (in this case, the occurrence of new cases), while lstm will capture temporal patterns and dependencies in the data.

# Findings:

# 1.Survival Analysis:

# Survival Analysis provides insights into the time-to-event data, allowing us to estimate the probability distribution of when new cases will occur.

# Factors such as population density, healthcare infrastructure, vaccination rates, and public health measures will be considered as covariates influencing the hazard rate.

# 2.LSTM Model:

# LSTM models are well-suited for capturing complex temporal dependencies in time series data.

# Features such as previous case counts, vaccination data, mobility trends, and other relevant covariates will be used as input to the LSTM.

# Results:

# 1.Survival Analysis Results:

# The Survival Analysis model provides estimates of the hazard function, cumulative hazard, and survival probabilities over time.

# It identifies significant covariates influencing the hazard rate of new COVID-19 cases.

# 2.LSTM Model Performance:

# The LSTM model will be trained and evaluated using historical COVID-19 data.

# Performance metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared (R2) will be used to assess the model's accuracy.

# 3.Combined Prediction:

# The predictions from the Survival Analysis and LSTM models will be combined to provide an ensemble forecast of new COVID-19 cases in India.

# Summary:

# The combination of Survival Analysis and LSTM provides a robust framework for forecasting new COVID-19 cases in India. This approach enhances accuracy and reliability, enabling informed decision-making for public health interventions and mitigation strategies. However, model performance depends on data quality and availability.

# “COVID-19 Candidate Treatments, a Data Analytics Approach”

**Objective:**

The objective of this study is to employ a data analytics approach to evaluate and prioritize potential candidate treatments for COVID-19. This involves analyzing available clinical and experimental data to identify promising therapeutic options, with the aim of informing clinical decision-making and guiding further research efforts.

**Findings:**

**1.Literature Review and Data Collection:**

* Comprehensive literature review and data collection are conducted to gather information on candidate treatments, including antiviral drugs, monoclonal antibodies, repurposed drugs, and investigational therapies.

**2.Data Preprocessing and Feature Selection:**

* Rawdata from clinical trials, experimental studies, and real-world evidence are processed and cleaned. Relevant features such as treatment efficacy, safety profiles, and patient demographics are selected for analysis.

**3.Descriptive Analytics:**

* Initial descriptive analysis provides an overview of the characteristics of each candidate treatment, including summary statistics, distribution of outcomes, and adverse events.

**4.Effectiveness Assessment:**

* Statistical techniques, such as meta-analysis or Bayesian inference, are applied to assess the effectiveness of each treatment option based on available data. This includes measures like odds ratios, hazard ratios, or other relevant metrics.

**5.Safety Profiling:**

* Adverse event data are analyzed to evaluate the safety profiles of candidate treatments. This includes the frequency, severity, and type of adverse events associated with each treatment.

**Results:**

**1.Effectiveness Ranking:**

* Candidate treatments are ranked based on their estimated effectiveness, with accompanying confidence intervals or uncertainty estimates.

**2.Safety Assessment:**

* Treatments are evaluated in terms of their safety profiles, highlighting any significant adverse events or safety concerns associated with specific optio**ns.**

**3.Identification of Promising Candidates:**

* Based on the effectiveness and safety assessments, a shortlist of the most promising candidate treatments is generated. These are treatments that demonstrate a favorable balance between efficacy and safety.

**Summary:**

This study uses data analytics to evaluate COVID-19 treatments, aiming to inform clinical decision-making and guide further research. The analysis reveals the effectiveness and safety profiles of different treatments, prioritizing interventions with the greatest patient benefit. This information is crucial in a rapidly evolving pandemic. The results are based on available data and may be updated as new evidence emerges. Continuous monitoring and reevaluation are necessary to adapt to scientific understanding and clinical practice. The study's conclusions should be considered alongside clinical expertise and other contextual factors.

# “Analyzing and Forecasting COVID-19 Outbreak in India”

# Objective:

# The objective of this analysis is to examine the trajectory of the COVID-19 outbreak in India, employing statistical modeling and forecasting techniques. The study aims to provide insights into past trends, assess the current situation, and project potential future scenarios. This information will assist in making informed decisions related to public health interventions, resource allocation, and healthcare planning.

# Findings:

# 1.Epidemiological Trends:

# Analysis of historical data reveals patterns in the spread of COVID-19, including waves of infections, peaks, and troughs.

# Factors influencing transmission rates, such as public health measures, population density, and vaccination rates, are considered.

# 2.Statistical Modeling:

# Time-series models, such as ARIMA (AutoRegressive Integrated Moving Average) or SEIR (Susceptible-Exposed-Infected-Recovered), are used to model and understand the dynamics of the outbreak.

# 3.Forecasting Scenarios:

# Multiple scenarios are generated based on different assumptions, including variations in vaccination rates, adherence to public health measures, and potential emergence of new variants.

# 4.Sensitivity Analysis:

# Sensitivity analyses are conducted to assess the impact of different parameters and assumptions on the projected outcomes.

# Results:

# 1.Short-term Forecast:

# Short-term forecasts provide estimates of expected new cases, hospitalizations, and potential strain on healthcare infrastructure in the coming weeks.

# 2.Medium to Long-term Projections:

# Medium to long-term projections outline potential trends in the spread of the virus, allowing for better preparedness and resource allocation.

# 3.Scenario Analysis:

# Different scenarios (e.g., high vaccination rates, low vaccination rates, emergence of new variants) are considered, providing a range of potential outcomes and associated uncertainties.

# Summary:

# The analysis and forecasting of the COVID-19 outbreak in India uses historical data, statistical models, and scenario analyses to provide a comprehensive understanding of the current situation and future trajectories. The findings emphasize the importance of vaccination efforts, adherence to public health measures, and preparedness for potential changes. Sensitivity analysis helps understand the impact of variables on projected outcomes. The accuracy of forecasts depends on data quality, assumptions validity, and population responsiveness.

# “Analysis and predictions of spread, recovery, and death caused by COVID-19 in India”

# Objective:

# The objective of this analysis is to examine the spread, recovery, and mortality patterns of COVID-19 in India. This includes investigating the trends in new cases, recoveries, and deaths, identifying factors influencing these trends, and making predictions about future scenarios. The aim is to provide valuable insights for public health planning, resource allocation, and policy-making.

# Findings:

# 1.Epidemiological Trends:

# Analysis of historical data reveals the progression of the pandemic in India, including waves of infections, peaks, and troughs.

# Factors such as public health interventions, vaccination rates, population density, and healthcare infrastructure are considered.

# 2.Recovery Rates:

# Examination of recovery rates over time helps understand the effectiveness of treatment protocols, healthcare system capacity, and patient outcomes.

# 3.Mortality Rates:

# Analysis of mortality rates provides insights into the severity of the pandemic, the impact on vulnerable populations, and the effectiveness of healthcare interventions.

# 4.Factors Influencing Spread and Outcomes:

# Identification of factors, such as vaccination coverage, adherence to preventive measures, emergence of new variants, and healthcare capacity, that significantly influence the spread, recovery, and mortality rates.

# Results:

# 1.Spread Predictions:

# Short-term and medium-term forecasts provide estimates of expected new cases, allowing for better resource allocation and healthcare planning.

# 2.Recovery and Mortality Projections:

# Projections for recovery rates and mortality rates help in anticipating the burden on the healthcare system and guiding interventions to improve patient outcomes.

# 3.Scenario Analysis:

# Different scenarios (e.g., varying vaccination rates, changes in public behavior) are considered, providing a range of potential outcomes and associated uncertainties.

# Summary:

# The analysis of COVID-19 spread, recovery, and mortality in India uses historical data, statistical models, and scenario analyses. It emphasizes vaccination efforts, public health measures, and preparedness for changes. The accuracy of forecasts depends on data quality, assumptions validity, and population responsiveness. Continuous monitoring and updates are crucial for adapting to the pandemic's dynamic nature.

**6.DESIGN THINKINH APPROACH:**

**Empathize:**

Before diving into solving the problem, it's crucial to empathize with the users and understand their needs. In this case, our primary users are common peoples ,healthcare professionals and government officials who take decision of lockdowns and measures . We need to gather insights into what factors are most important to them and what are the information they need regarding covid crisis.

**Actions:**

* **Empathizing:** Conduct surveys or interviews with potential users to gather their perspectives.
* Data Collection: Collect data on the number of cases, deaths, recoveries, testing rates, vaccination rates, and other relevant variables .
* **Data Cleaning**: Clean and preprocess the data to handle missing values, outliers, and inconsistencies.Standardize data formats and units

**Define:**

Based on our understanding of the problem and the users' needs, we will define clear objectives and success criteria for our project.

**Action:**

* Clearly define the problem or challenge related to COVID-19 that you want to address.
* Use the insights gathered during the empathize stage to articulate the needs and pain points of the people affected by COVID-19.
* The problem is the lack of reliable and consistent information about the covid trends which may affect proper decision making.

**Ideate:**

Brainstorm potential solutions and approaches to address the problem. This phase involves thinking creatively and considering various algorithms and techniques for covid 19 case analysis

**Action:**

* **Visualization:** Create visualizations (charts, maps, dashboards) to make the data more understandable.
* **Spatial Analysis**: Identify hotspots and areas with higher transmission rates.
* **Healthcare Resource Allocation**: Use the analysis to recommend resource allocation strategies, such as where to set up additional testing centers etc.

**Prototype:**

create a prototype of the covid 19 case analysis project and the user interface for obtaining necessary information.

**Action:**

* Create a prototypes or mockups of dashboards or reports to get feedback from stakeholders or users.
* Utilize features like drag-and-drop interface, data modeling, and interactive filtering.
* Design an intuitive and user-friendly interface for users to interact with the system, explore data, and gain insights.

**Test :**

Evaluate the project using appropriate metrics and gather feedback from users.

**Action:**

* Share your prototypes with stakeholders and gather their feedback.
* Pay attention to how well the visualizations meet their needs and whether they find the insights actionable.
* Be prepared to make adjustments and refinements based on their input.

**Implement:**

Once the prototype meets the defined objectives and receives positive feedback, proceed with full implementation.

**Action:**

* Once iteration on prototype is completed and received approval from stakeholders, proceed to build the final visualizations and analytical models using IBM Cognos.
* Ensure that the data is up-to-date and properly integrated into the system.

**Iterate:**

Continuous improvement is essential. Gather user feedback and iterate on the model and interface to enhance accuracy and usability.

**Action:**

* Set up a system for ongoing monitoring of COVID-19 data and updating the visualizations as new data becomes available.
* Regularly communicate your findings and insights to stakeholders. Consider automated reports or scheduled updates.
* Continuously gather feedback and make improvements based on user experience and changing requirements.

**7.STEPS INVOLVED IN MODEL EVALUATION:**

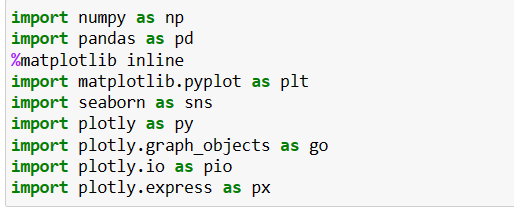
**Data Collection:**

First, ensure you have access to COVID 19 case data ,Gather data from reliable sources such as government health agencies, the World Health Organization (WHO), and reputable research institutions. Collect data on the number of cases, deaths, recoveries, vaccination rates, and other relevant variables .

**Import Libraries:**

Start by importing the necessary libraries such as numpy, pandas for data manipulations, matplotlib and seaborn for visualisations etc

IMPORT LIBRARIES

****

**Load the Dataset:**

* This step involves loading your COVID 19 dataset into your Python environment. The dataset should be in a format that Pandas can easily handle, such as a CSV file.

LOADING DATASET



* The read\_csv () function is used to load a CSV (Comma-Separated Values) file into a Pandas DataFrame. You specify the file path within the parentheses.
* The result of this operation is a DataFrame, which is a tabular data structure that's similar to a spreadsheet. It allows you to work with your data in a structured and flexible way.

**Explore the Dataset**

Before diving into data preprocessing, it's important to understand your dataset. You can use various Pandas functions to explore it:

**data.head ():**

* This function displays the first few rows of your dataset, giving you a glimpse of its structure.



**Data.describe():**

* It provides basic statistical information about your data, including measures like mean, standard deviation, and quartiles for numerical columns.



**Data.columns():**

* This helps you see the names of all the columns in your dataset



**Data.info():**

* This method prints information about a DataFrame including the index dtype and columns, non-null values and memory usage.

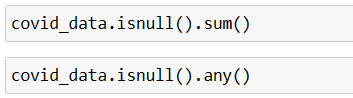


**Data Pre-proccessing:**

* Data preprocessing is crucial for ensuring the quality and usability of your data:

**Handle Missing Values:**

* Check for missing values in your dataset and decide on an appropriate strategy to handle them. You can fill missing values using methods like forward-fill, backward-fill, mean, median, or simply remove rows with missing values.



**Data Transformation:**

* If your dataset contains date or time columns, convert them to the datetime data type for time-based analysis.

|  |
| --- |
| “# Example: Convert a date column to datetime  data['Date'] = pd.to\_datetime (data['Date'])” |

* To convert categorical values such as countries and territories into a numerical values using “one\_heart\_encoding” technique.



**Data Cleaning:**

* Inspect your data for inconsistencies, outliers, or irregularities. Ensure that the data is clean and standardized. This may include dealing with irregular units, correcting typos, or removing duplicates.

**8.TOOLS AND LIBRARIES:**

**Python:**

* Python is a versatile, high-level programming language known for its simplicity and readability. It's widely used in data analysis, machine learning, and scientific computing due to its extensive libraries and frameworks.

**Pandas:**

* Pandas is a Python library for data manipulation and analysis. It provides data structures like DataFrames and Series, making it easy to work with structured data. Pandas is essential for data loading, cleaning, and transformation.

**NumPy:**

* NumPy is another Python library that focuses on numerical computing. It provides support for large, multi-dimensional arrays and matrices, as well as a variety of mathematical functions to operate on these arrays. It's fundamental for numerical data processing.

**Matplotlib:**

* Matplotlib is a popular data visualization library in Python. It allows you to create static, animated, or interactive visualizations in a wide range of formats, including line plots, bar charts, scatter plots, and more. It's excellent for visualizing data and trends.

**Seaborn:**

* Seaborn is a Python data visualization library built on top of Matplotlib. It provides a high-level interface for creating attractive and informative statistical graphics. Seaborn simplifies the creation of complex visualizations and is often used for creating aesthetically pleasing charts.

**Scikit-Learn:**

* Scikit-Learn, also known as sklearn, is a powerful machine learning library in Python. It offers a wide range of tools for data preprocessing, model selection, training, and evaluation. It's especially useful for building predictive models and performing machine learning tasks.

In the context of our project, these tools and libraries play the following roles:

* Python serves as the programming language for the project, providing a flexible and accessible environment for data analysis and modeling.
* Pandas is used for data manipulation and analysis, including loading, cleaning, and organizing the covid19 case data.
* NumPy complements Pandas by providing fundamental support for numerical operations and handling multi-dimensional arrays, which are often used in data analysis.
* Matplotlib is employed for creating visualizations that help in understanding covid trends and conveying insights effectively.
* Seaborn enhances the visualization process by providing a higher-level interface to create aesthetically pleasing and informative statistical graphics.
* Scikit-Learn plays a crucial role in building the predictive model that estimates cases and deaths in future.

**9. MODEL SELECTION:**

* Choose Support Vector Machine (SVM) for regression and classification tasks, handling complex data relationships.
* Train the model using training data and target variables.
* We may consider using an alternative algorithm for this model in order to enhance our ability to predict values.
* The choice of the best algorithm depends on the specific characteristics of the dataset and the problem we are trying to address.

**10. MODEL TRAINING AND EVALUATION :**

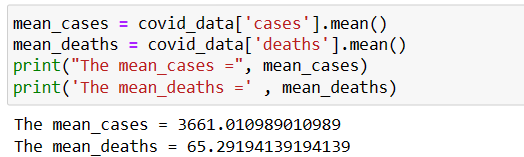
**Support Vector Machine(SVM):**

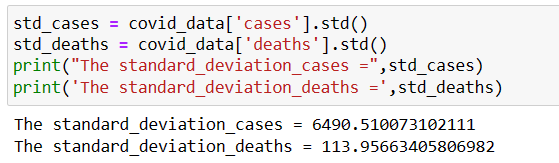
The accuracy of a Support Vector Machine (SVM) model for a COVID-19 dataset depends on various factors, including the specific dataset you're using, the features you've selected, the kernel type (linear, radial basis function, etc.), and the hyperparameters chosen for the SVM.

To determine the accuracy of SVM model, perform the following steps:

* **Data Splitting:** Split the dataset into a training set and a testing set. This allows to train the model on one subset of the data and evaluate its accuracy on another.
* **Feature Selection/Engineering**: choose the relevant features , like deaths, cases, year, month, day, and possibly countries/territories. Make sure the features have a significant impact on predicting COVID-19 cases.
* **Preprocessing:** Preprocess the data, which includes handling missing values, scaling, encoding categorical variables such as countries/Teritorries, and dealing with outliers.
* **Model Selection**: Select the appropriate kernel type (e.g., linear, radial basis function) and hyperparameters for your SVM model.
* **Training:** Train the SVM model using the training data. Make sure to set the random seed or random state to ensure reproducibility.
* **Evaluation:** After training, use the testing dataset to evaluate the model's performance. Common evaluation metrics for regression tasks (like predicting COVID-19 cases) include Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2).

**11.MEAN AND STANDARD DEVIATION OF CASES AND DEATHS :**

****

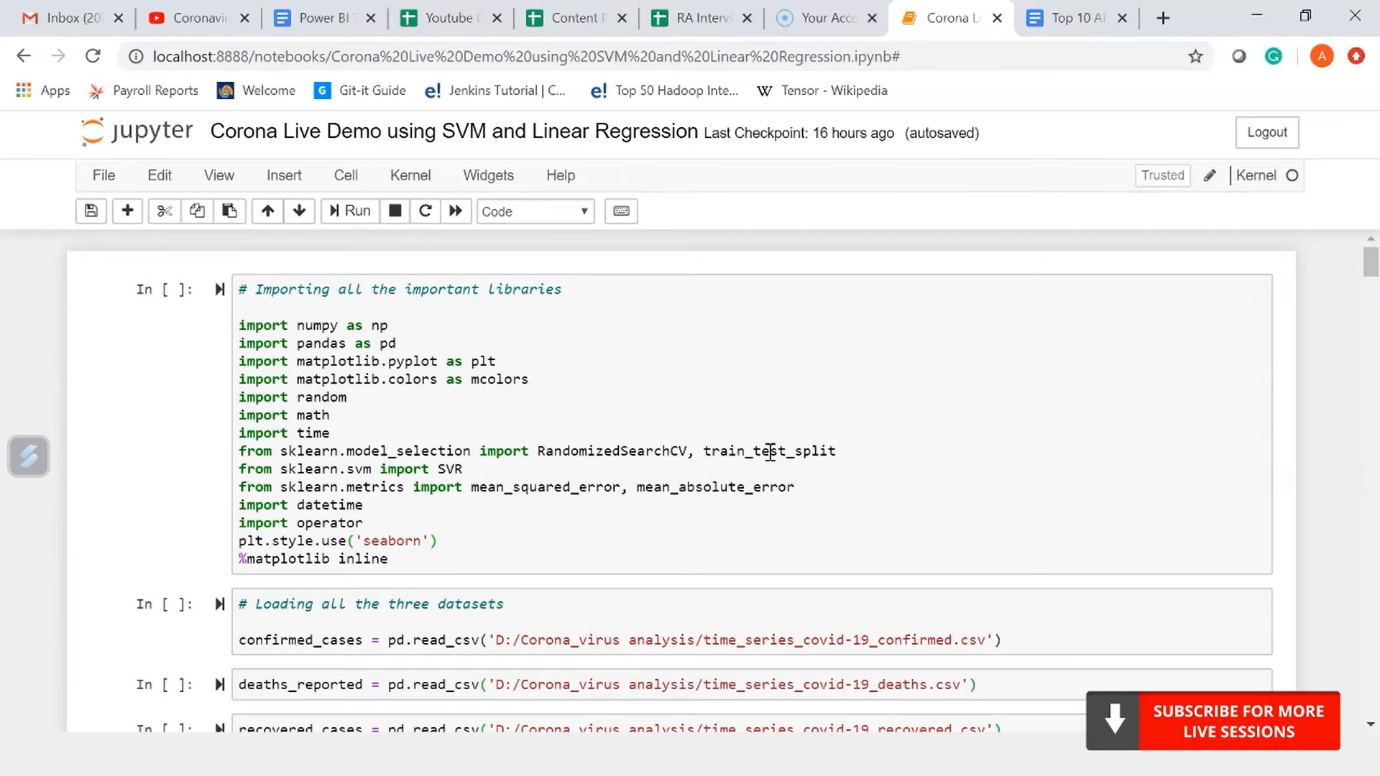
****

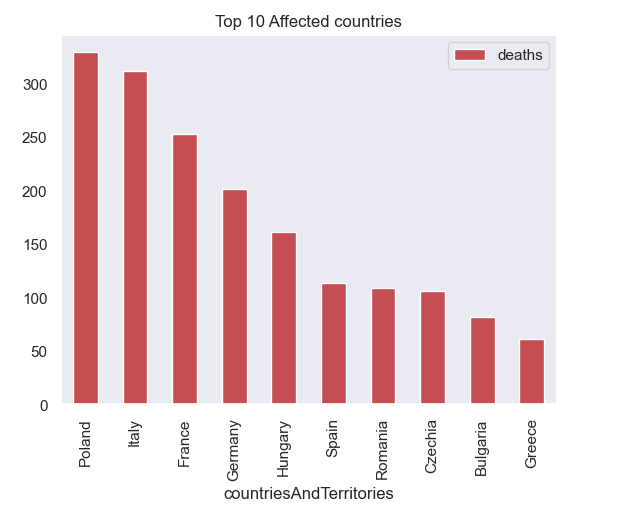
The mean (average) represents the central tendency of the data. In this context, it shows that, on average, there were approximately 3,661 reported COVID-19 cases and 65.29 reported deaths per day in the dataset.

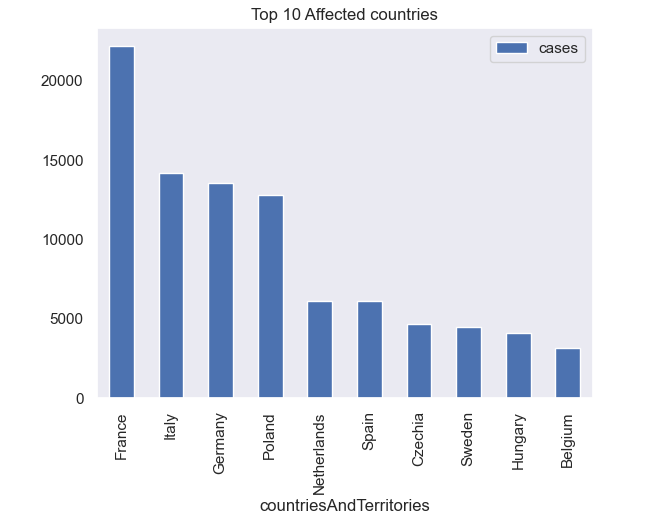
For cases: The standard deviation of 6,490.51 suggests that the number of daily COVID-19 cases varies considerably around the mean. Some days may have significantly higher case numbers, while others may have lower numbers.

For deaths: The standard deviation of 113.96 implies that the daily COVID-19 death counts also exhibit variation around the mean. There may be days with significantly more or fewer deaths.

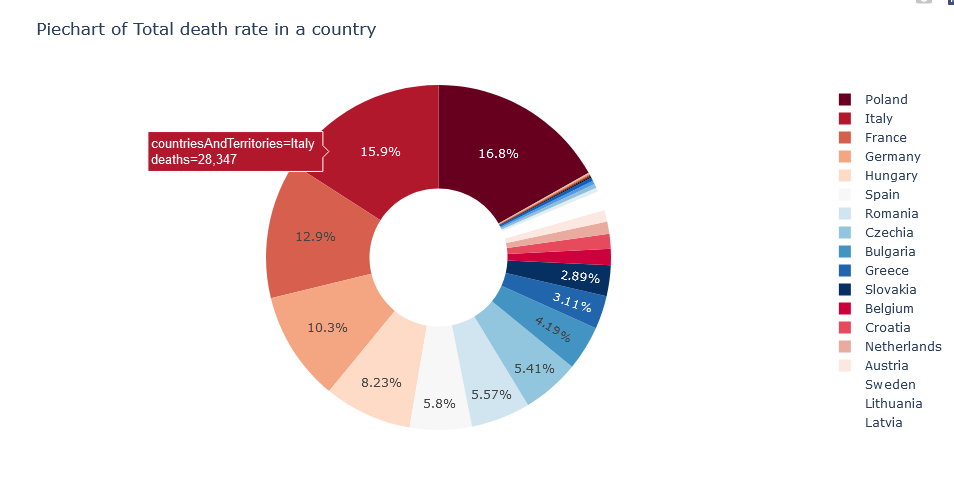
**12.VISUALIZATION USING PYTHON CODE:**



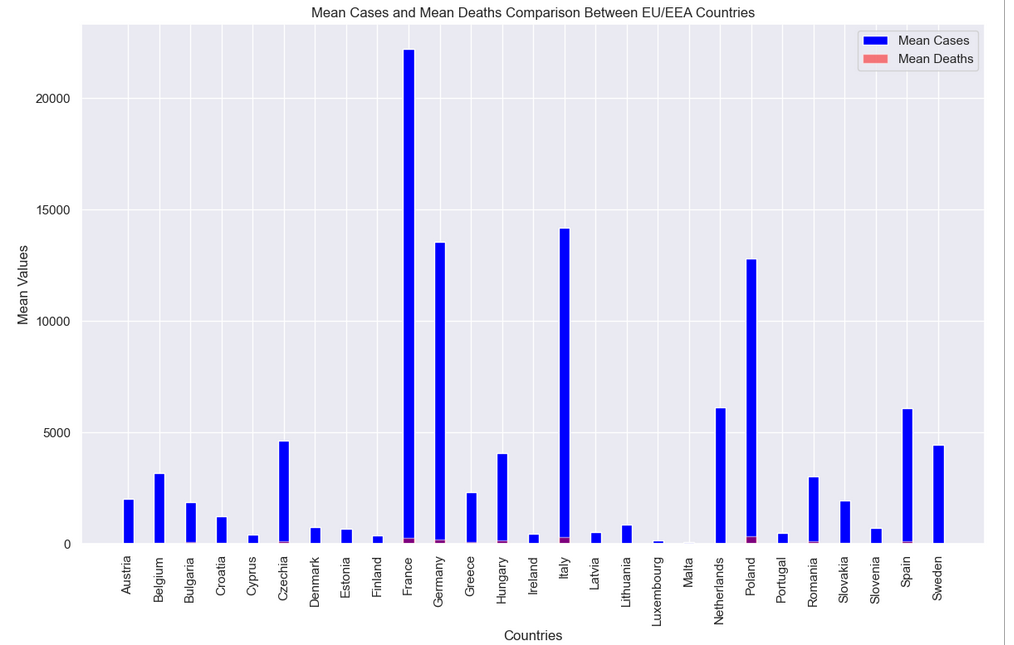


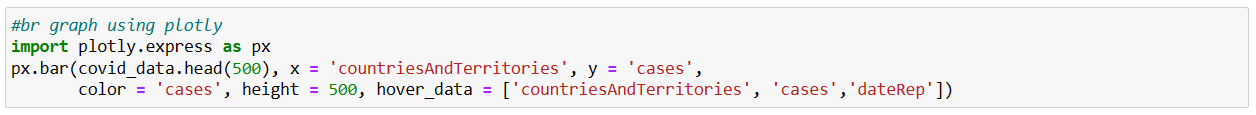


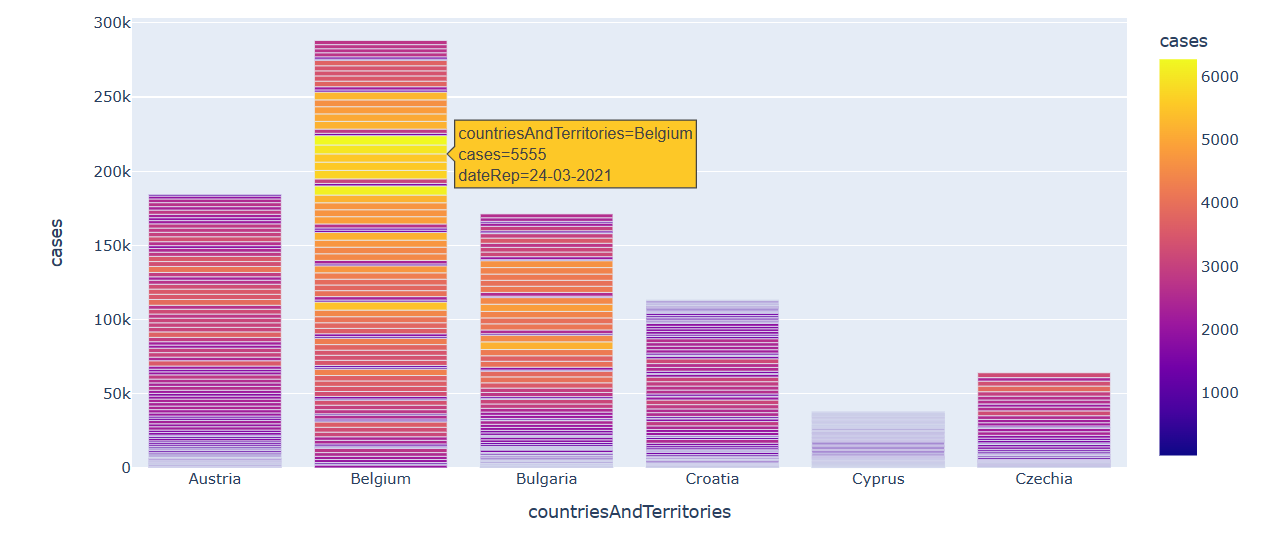
****

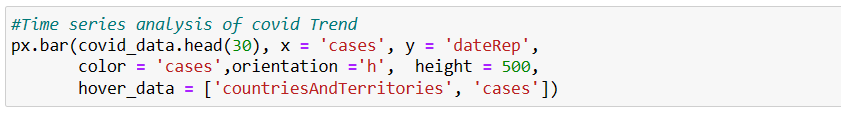
****

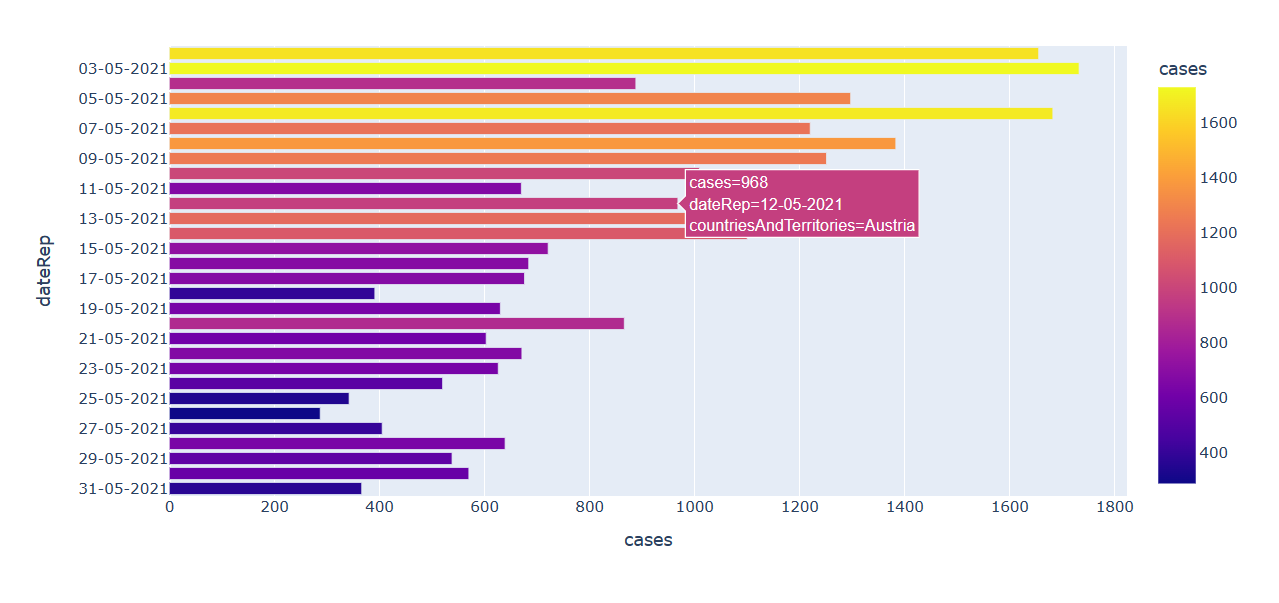
****

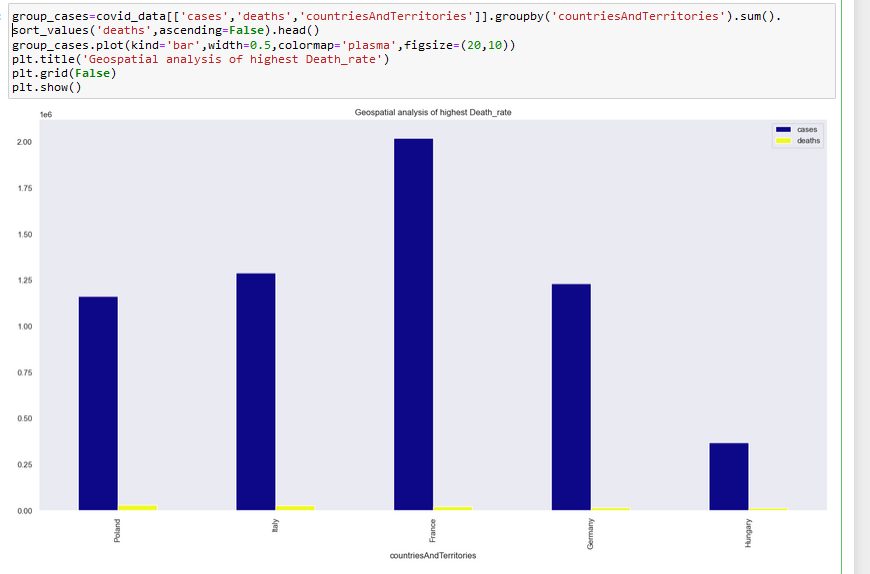
****

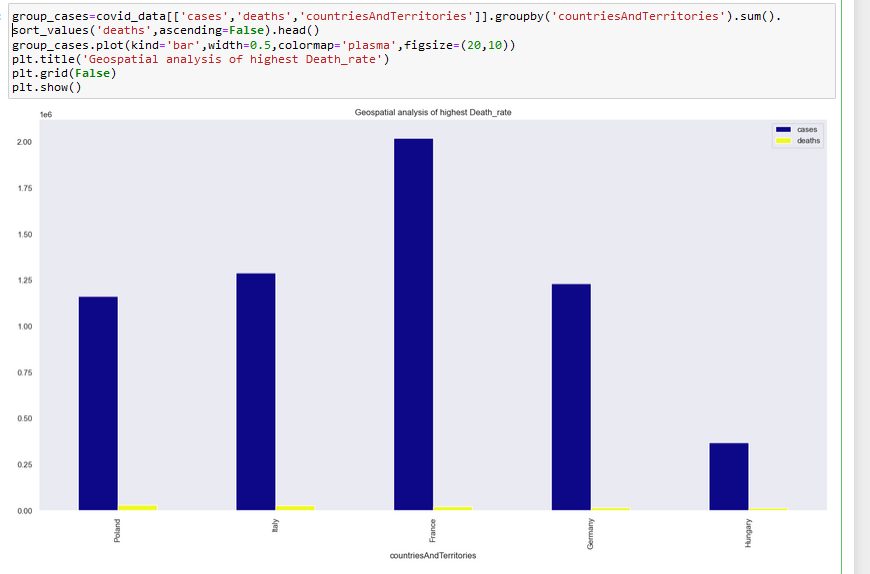








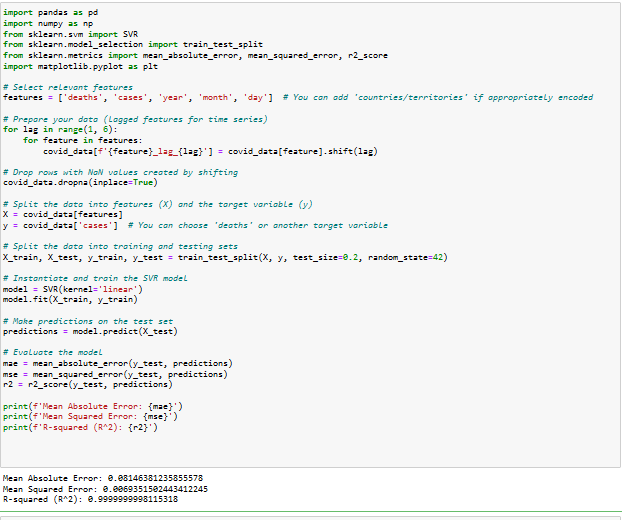






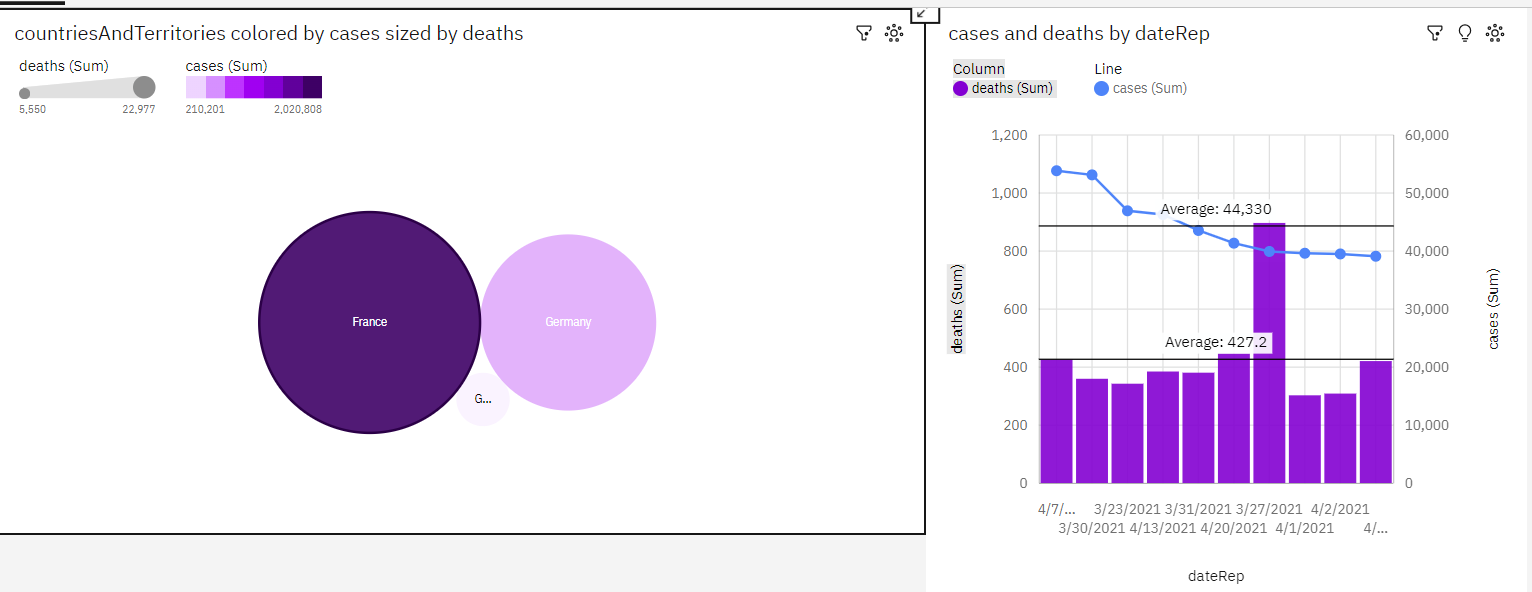


**Machine Learning Model(SVM):**

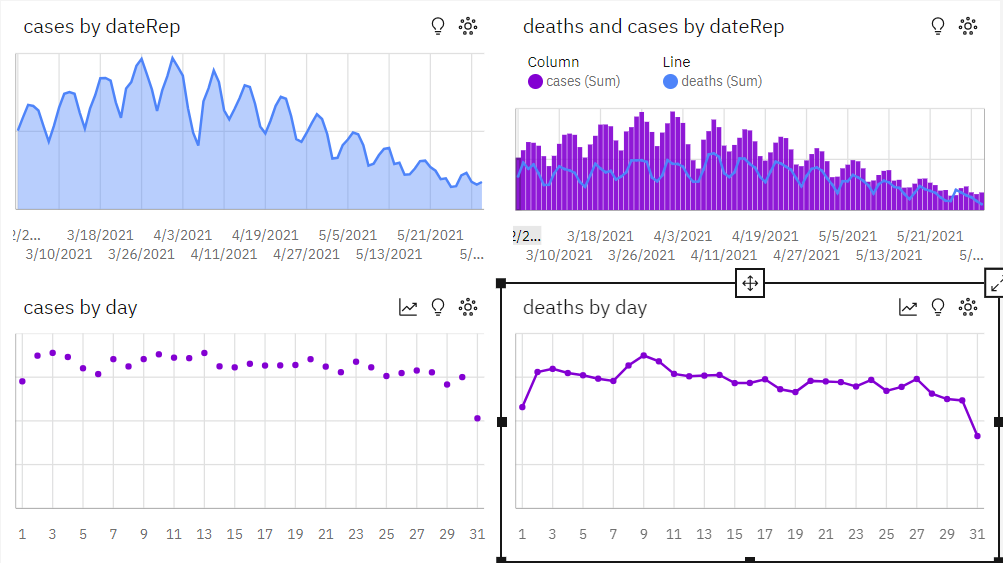
****

* MAE is a measure of the average absolute difference between the predicted values and the actual values.. A lower MAE indicates a better model fit. In our case, the MAE is approximately 0.0815, which means, on average, the predictions are off by about 0.0815 units from the actual values.
* MSE ,it squares the errors, which gives more weight to larger errors. A lower MSE is desirable, indicating a better model fit.
* R-squared is a measure of how well a model explains the variability in the data. It ranges from 0 to 1,In the svm model, the R-squared value is approximately 0.9999999998115318, which is very close to 1.
* In this analysis, it seems like the support vector machine model is performing very well.

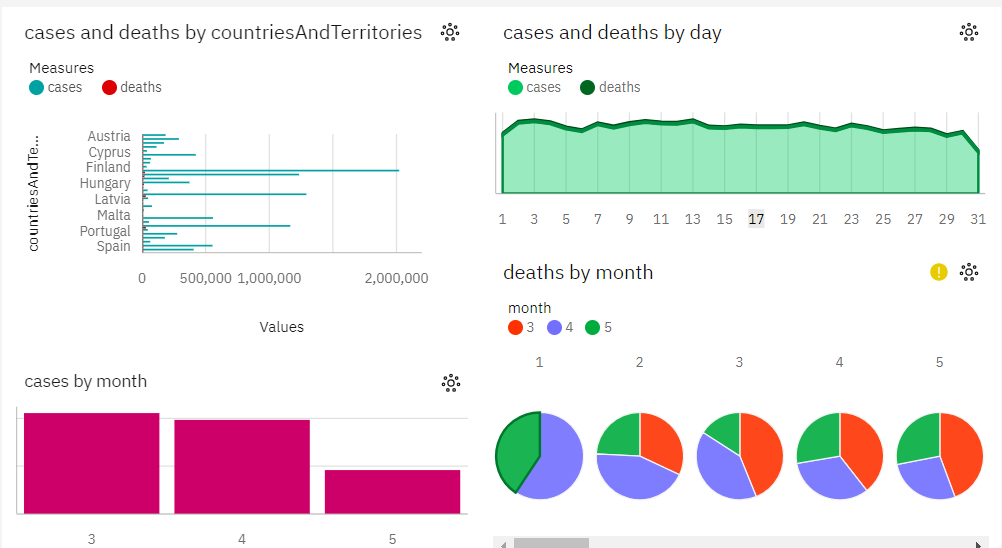
**13.VISUALIZATION USING IBM COGNOS:**



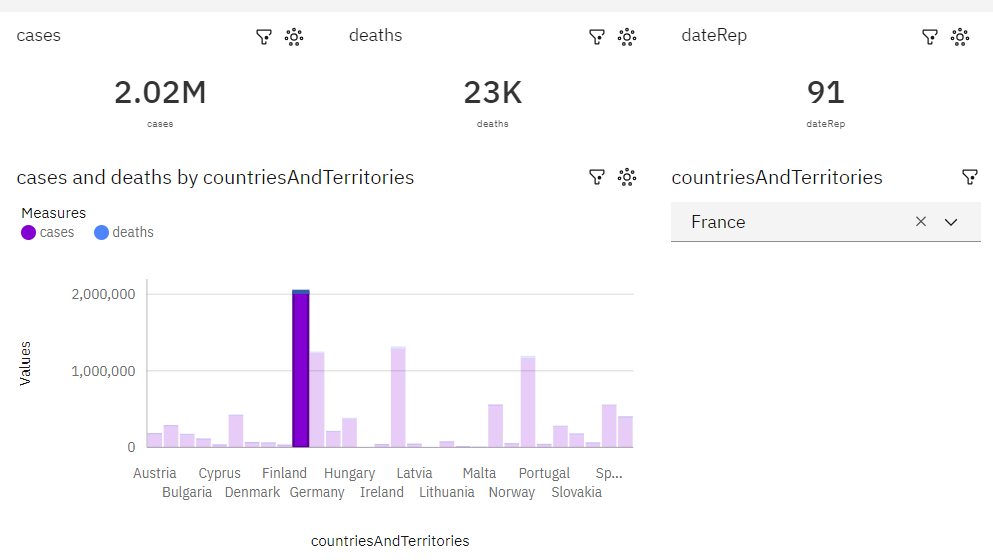
From the above visualization CountriesAndTerritories France has the highest values of both deaths and cases. Over all values of countriesAndTerritories, the sum of deaths is nearly 47 thousand.deaths ranges from over 5500, when countriesAndTerritories is GreeceFor deaths, the most significant values of countriesAndTerritories are France and Germany, whose respective deaths values add up to over 41 thousand, or 88.2 % of the total.

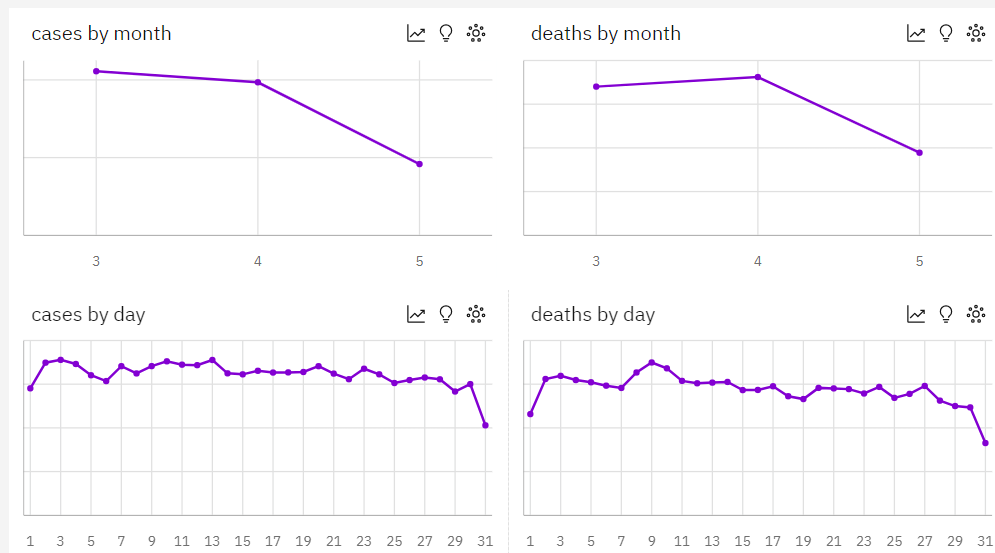


From the above visualization, Based on the current forecasting, deaths may reach nearly 5500 by day 38.Across all days, the sum of deaths is over 178 thousand.deaths ranges from almost 3500, when day is 31, to nearly seven thousand, when day is 9.

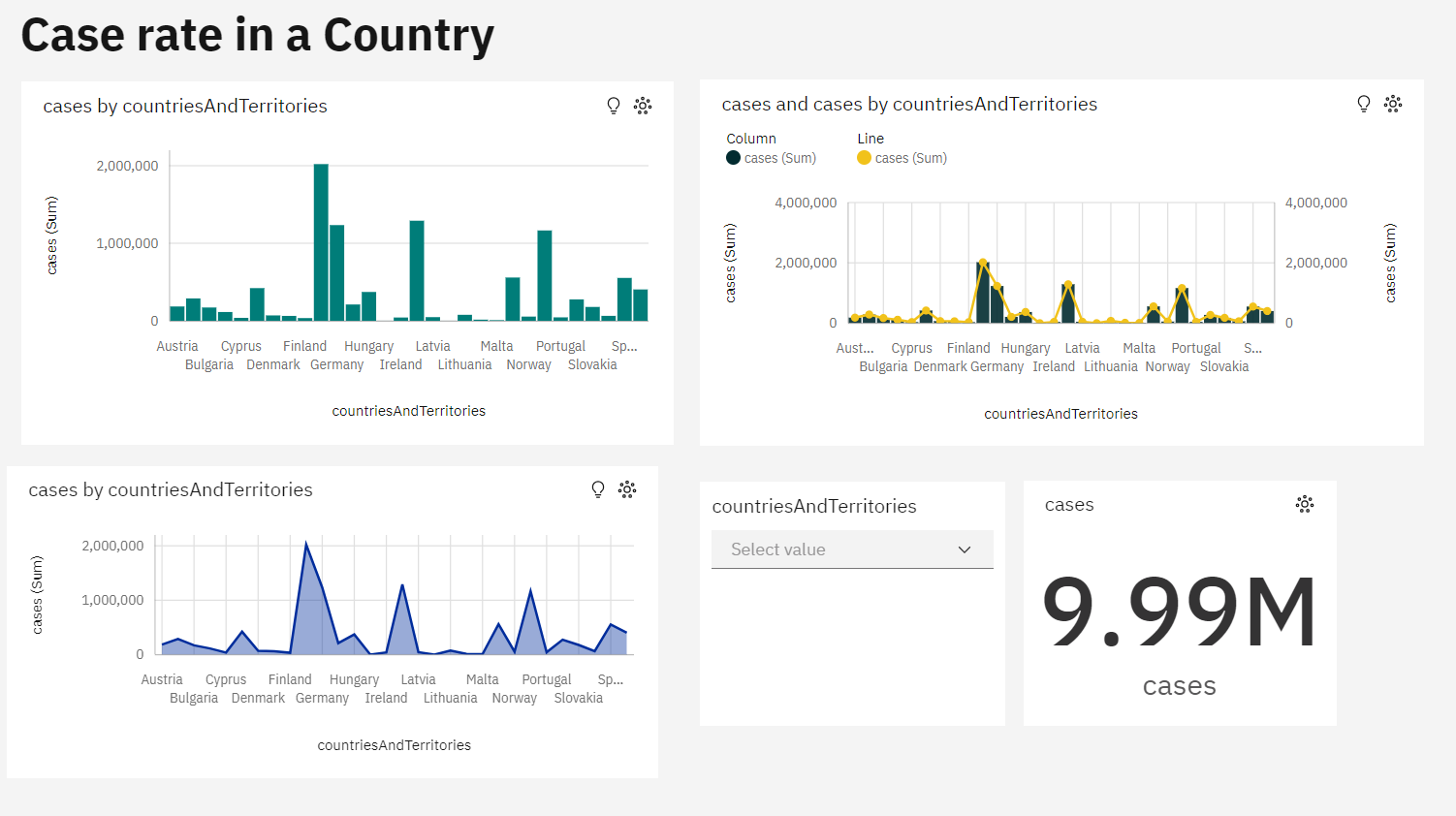


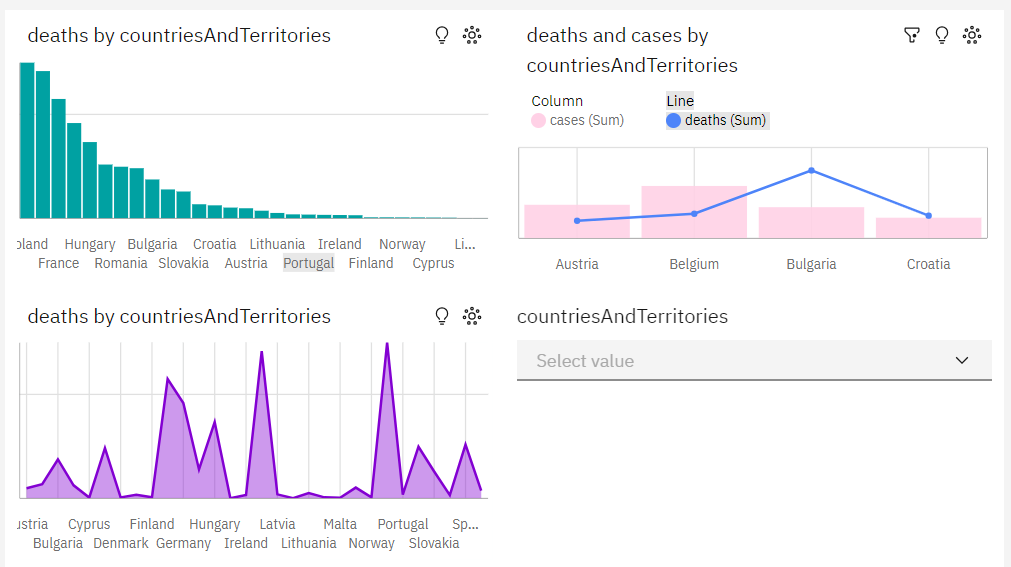
From the above visualization, Over all days, the average of deaths is 65.29.Add insight to favorites .The total number of results for cases, across all days, is over 2500.Cases is unusually high when countriesAndTerritories is France. CountriesAndTerritories Poland has the highest Total deaths but is ranked 4 in Total cases.











From analysing the above visualizations, we know that France has the highest Total cases but is ranked 3 in Total deaths. Poland has the highest Total deaths but is ranked 4 in Total cases. The hierarchy of highest death rate in countries/teritorries follows as Poland, Italy, France, Germany etc. The hierarchy of highest cases rate follows as France, Italy, Germany, Poland etc.

**14. HOW INSIGHT IMPROVE USER EXPERIENCE:**

**Understanding covid 19 Trends:**

* + Through exploratory data analysis, the analysis helps reveal trends in air spread of covid cases. This includes identifying whether case levels have been increasing, decreasing, or remaining stable over time.
  + Insights into trends can inform policymakers and environmental agencies about the effectiveness of disease control measures and the need for future actions.

**Spotting Pollution Hotspots:**

* Geospatial analysis and visualization pinpoint areas with consistently high case and death levels. These hotspots can be indicative of specific sources or regions that require immediate attention.
* Knowing where the danger is concentrated helps allocate resources and implement localized measures.

**Predictive Modeling for Future Trends:**

* The predictive model estimates cases and death rate over time in near future, offering forecasts for the near and long term.
* Predictive insights enable proactive measures, such as issuing early warnings and planning lockdowns and other preventive strategies based on future trends.

**Assessing the factors influencing the spread of covid19:**

* Comparative Analysis, Comparing Countries , allow users to select multiple countries and compare their COVID-19 metrics. Visualizing different countries on the same graph helps users understand the differences in the spread and impact of the virus.
* growth rates for cases and deaths, Insights into exponential growth or decline can help users make informed decisions.

**Recommendations for Mitigation:**

* Based on the insights obtained from the analysis, informed steps can be made by the government to control the spread of diesease. These might include implementing lockdown, social distancing, public awareness campaigns, vaccination etc.
* Recommendations aim to mitigate spread and improve life in the region.

**15.CONCLUSION:**

A COVID-19 case analysis project involves analysing , visualizing the complex dataset and insights derived from the analysis, these exploratory data analysis may include many ways such as statistical analysis where the highest death rate is over 50,000 , geospatial analysis to findout the hotspot/most affected areas such as poland in terms of death rate whereas france in terms of case rate. France has the highest Total cases but is ranked 3 in Total deaths. Poland has the highest Total deaths but is ranked 4 in Total cases. The hierarchy of highest death rate in countries/teritorries follows as Poland, Italy, France, Germany etc. The hierarchy of highest cases rate follows as France, Italy, Germany, Poland etc.

In this project svm model have been used for regression and future prediction . The svm model has significantly lower MAE amd MSE values, indicating that it is better at predicting the target variable. The model R-squared value is close to 1, hence it has a strong fit. This model is essential for futures prediction of cases and death rate over time that would help the policy makers in descision making .