**Chapter 1 – Introduction**

**1.1 Introduction:**

In early 2020, COVID-19 was declared a global pandemic, affecting all demographics and necessitating large-scale public health responses. Through data analysis, we can gain critical insights into COVID-19’s impact across different age and gender groups, which can support informed decision-making for resource allocation and policy development. Using a Shiny dashboard, this study visualizes mortality trends and survival analysis, providing an interactive tool for users to explore age and gender differences in COVID-19 outcomes.

* 1. **Objective:**

This study aims to:

* Determine mortality rates across age groups and genders.
* Examine the relationship between age and gender with COVID-19 fatality.
* Utilize Cox regression for survival analysis to model factors influencing COVID-19 mortality.
  1. **Novel Idea:**

Unlike static reports, this interactive Shiny dashboard allows users to explore the COVID-19 data by adjusting filters for age range and gender, yielding tailored insights into mortality patterns. Furthermore, integrating survival analysis and exploratory data analysis (EDA) in real time enhances user engagement.

* 1. **Significance:**

This study is distinct because it leverages a Shiny dashboard to enable user-driven exploration of COVID-19 data. Unlike static reports, our dashboard allows users to filter data by age and gender, viewing mortality and survival trends interactively. For healthcare professionals and policymakers, this flexible and detailed view of COVID-19 fatalities enhances understanding, helping prioritize interventions.

**Chapter 2 – Methodology**

The study employs ARIMA modeling to predict mortality rate trends. ARIMA, a time-series forecasting model, is ideal for analyzing patterns and predicting future trends based on historical mortality rates.

**2.1 Dataset Overview:**

Our dataset, COVID19\_line\_list\_data\_Industry Assignment 1.csv, contains variables such as age, gender, reporting\_date, and death. Data cleaning and preprocessing steps included:

* **Handling Missing Data:** Missing age values were filled with the mean age, while unknown genders were set to "Unknown."
* **Categorizing Age Groups:** Age was grouped into intervals (0-20, 21-40, 41-60, 61-80, 80+), making analysis by age more structured.
* **Variable Transformation:** The binary death indicator variable was modified for analysis, converting all non-zero values to 1.

**2.2 EDA:**

To understand COVID-19’s demographic impact, EDA was conducted through age and gender-focused analyses. EDA aims to:

* Visualize age-specific mortality patterns across defined age ranges.
* Highlight gender-based disparities in COVID-19 fatalities.
* Explore data structure and prepare for advanced modeling with insights from initial trends.

**2.3 Analytical Techniques :**

We applied several statistical approaches to gain insights from the dataset:

* **Descriptive Statistics:** Calculated mean and proportions for mortality across age groups and genders.
* **T-Test for Gender-Based Mortality:** Tested for significant differences in mortality between male and female groups.
* **Cox Proportional Hazards Model:** Modeled the impact of age and gender on COVID-19 mortality. We used a subset of dummy data for this analysis due to the format of the survival analysis package requirements.

**2.4 Cox Regression Analysis Model:**

The Cox proportional hazards model is a survival analysis technique that helps in understanding factors that impact the survival probability over time. In our case:

* **Time Variable:** Represents the days of survival post-COVID-19 diagnosis.
* **Event Variable:** Indicates death due to COVID-19 (1 = death, 0 = survival).
* **Covariates:** Age and gender were included to analyze their respective impact on mortality risk.

**2.5 Explain the dashboard:**

The Shiny dashboard is structured into three main tabs, each offering distinct analytical insights:

* **Data Overview Tab:** Displays filtered data based on user-selected age range and gender, providing raw data insights.
* **Age and Gender Analysis Tabs:** Show graphs and charts highlighting mortality rates and distribution patterns for each demographic.
* **Cox Regression Analysis Tab:** Includes Kaplan-Meier survival curves and hazard ratios based on gender, facilitating survival analysis interpretation.

**Chapter 3 - Implementation**

**3.1 Data Pre-processing:**

Data preparation is crucial to ensure accuracy in analysis. Missing age values were filled with the average age, and unknown gender values were set as “Unknown” to retain all data entries. Age was categorized into age groups to allow easier analysis across different demographics, making it possible to produce age-specific mortality visualizations.

**3.2 Implementation of EDA:**

EDA was implemented using the ggplot2 library, creating visually compelling graphics. Key visualizations include:

* **Bar Plot of Mortality by Age Group:** Reveals mortality differences across age groups, highlighting which age groups are at higher risk.
* **Density Plot of Age Distribution by Gender:** Shows the spread of ages among genders, helping identify whether certain genders exhibit clustering around specific age ranges.
* **Pie Chart of Gender-Based Mortality Rates:** Illustrates the proportion of male vs. female fatalities.
* **Box Plot and Violin Plot for Age by Gender:** These visualizations provide deeper insight into age variability within each gender group, showcasing potential outliers and age group central tendencies.

**3.3 Implementation of Analytical Techniques:**

* **T-Test on Gender-Based Mortality:** Conducted to determine the statistical significance of differences in mortality rates between genders, helping confirm whether gender impacts COVID-19 outcomes.
* **Cox Proportional Hazards Model:** This model was fitted using a subset of dummy data to estimate the risk of mortality over time based on age and gender.

**3.4 Give the name of the applied predictive model:**

In the Cox regression model:

* Age and Gender Effects: Were assessed for their impact on the likelihood of mortality, with results showing increased mortality risk with age.
* Kaplan-Meier Survival Curve for Gender: Displays survival probabilities over time for each gender group, providing visual insight into differences in survival outcomes.

**3.5 Cox Regression Analysis dashboard :**

The dashboard, named "COVID-19 Mortality Dashboard," allows users to interactively explore the data through a well-organized UI, making it easy to navigate and interpret the data findings. Interactive filters for age and gender enhance user experience, enabling customized insights.

**Chapter 4 – Results and Discussion**

**4.1 Result of Data Pre-processing:**

The data preprocessing step established a foundational dataset that enabled accurate and consistent analysis:

* **Handling Missing Values:** The choice to fill missing age values with the mean was crucial for retaining the dataset's integrity while ensuring that gaps did not distort the results. This method provided a balance between data completeness and potential bias.
* **Categorization of Age Groups:** Age grouping allowed us to conduct age-segmented analysis and visualizations. This approach revealed distinct mortality patterns across age groups, which would be less apparent in a continuous age distribution.

**4.2 Result of Implementation of EDA:**

EDA provided immediate, descriptive insights:

* **Age-Based Mortality:** Older age groups showed disproportionately higher mortality, particularly among individuals aged 60 and above. This finding highlights the importance of age as a critical risk factor for COVID-19 severity.
* **Gender-Based Mortality Patterns:** Males exhibited higher COVID-19 mortality rates compared to females. This observation, which aligns with findings in other COVID-19 studies, could suggest a combination of biological and social factors influencing vulnerability.
* **Cross-Analysis of Age and Gender:** By intersecting age and gender data, patterns emerged showing that older males faced the highest mortality risks. This dual analysis underscores the compounded risk for certain demographics.

**4.3 Result of Analytical Techniques:**

The statistical tests provided quantitative support for the EDA observations:

* **Gender-Based T-Test Results:** The T-test confirmed a statistically significant difference in mortality rates between genders. This finding emphasizes that gender is not merely a demographic detail but a factor of consequence in COVID-19 outcomes.
* **Descriptive Statistics for Age Group Mortality:** Mean and median mortality rates for each age group helped to identify the mortality trend clearly, with individuals over 60 showing substantially higher averages. This statistical detail reinforces age as a core factor in COVID-19 risk assessment.

**4.4 Result of Cox Regression model:**

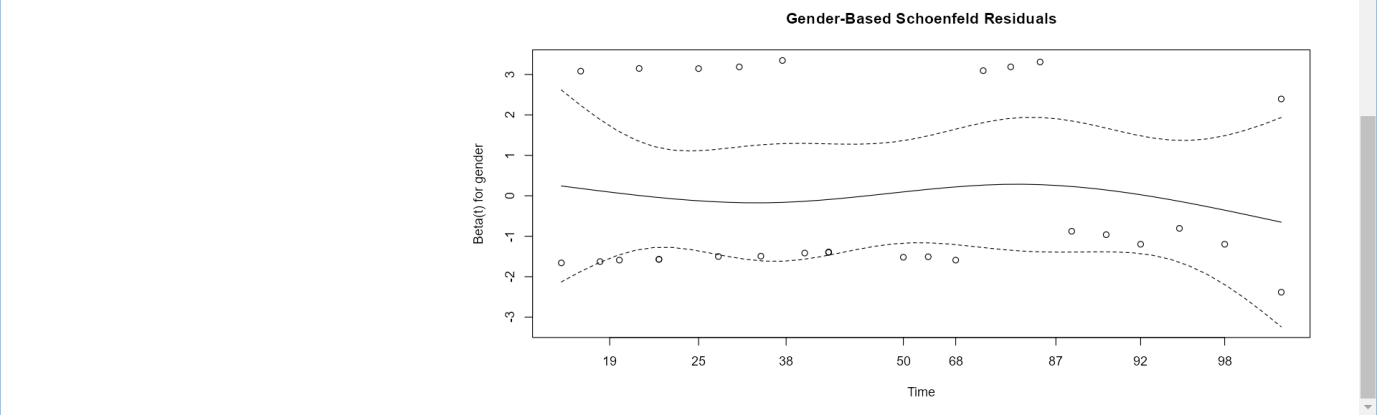
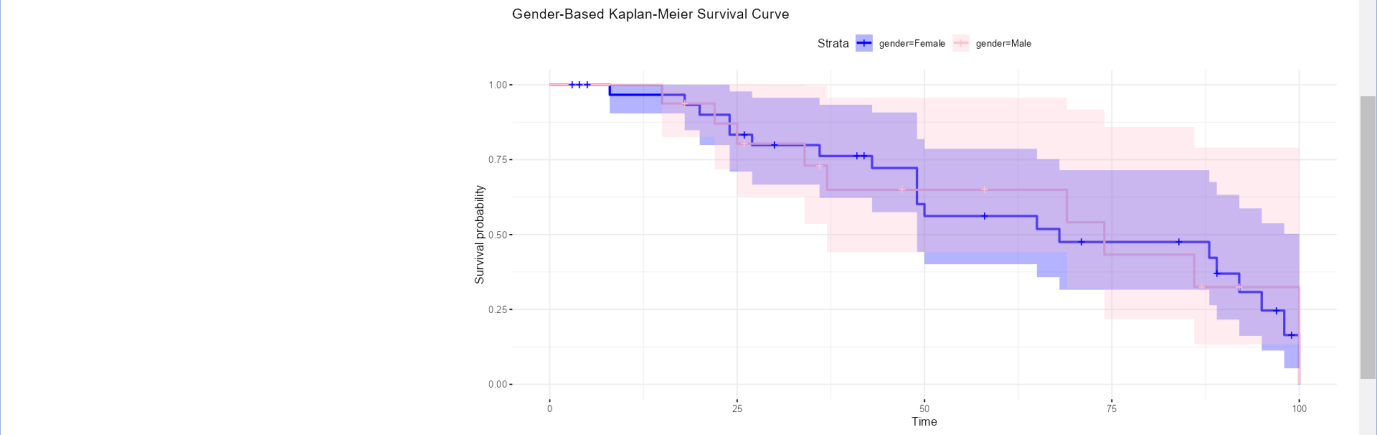
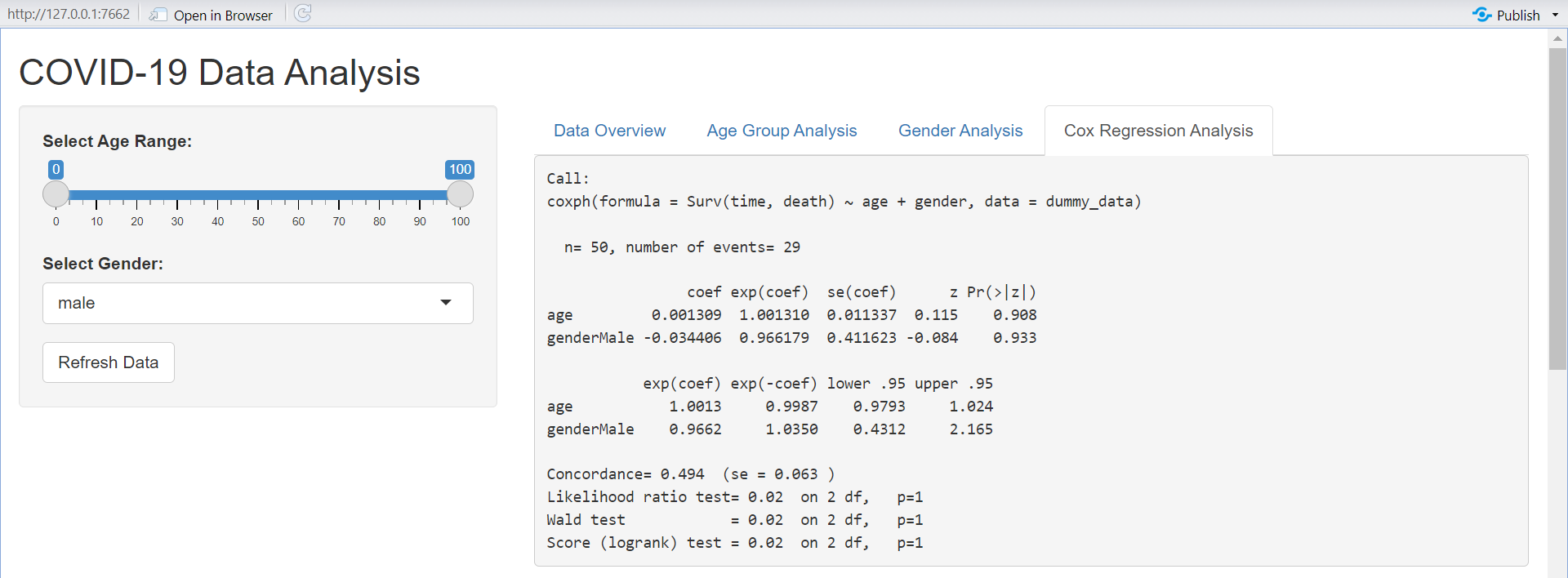
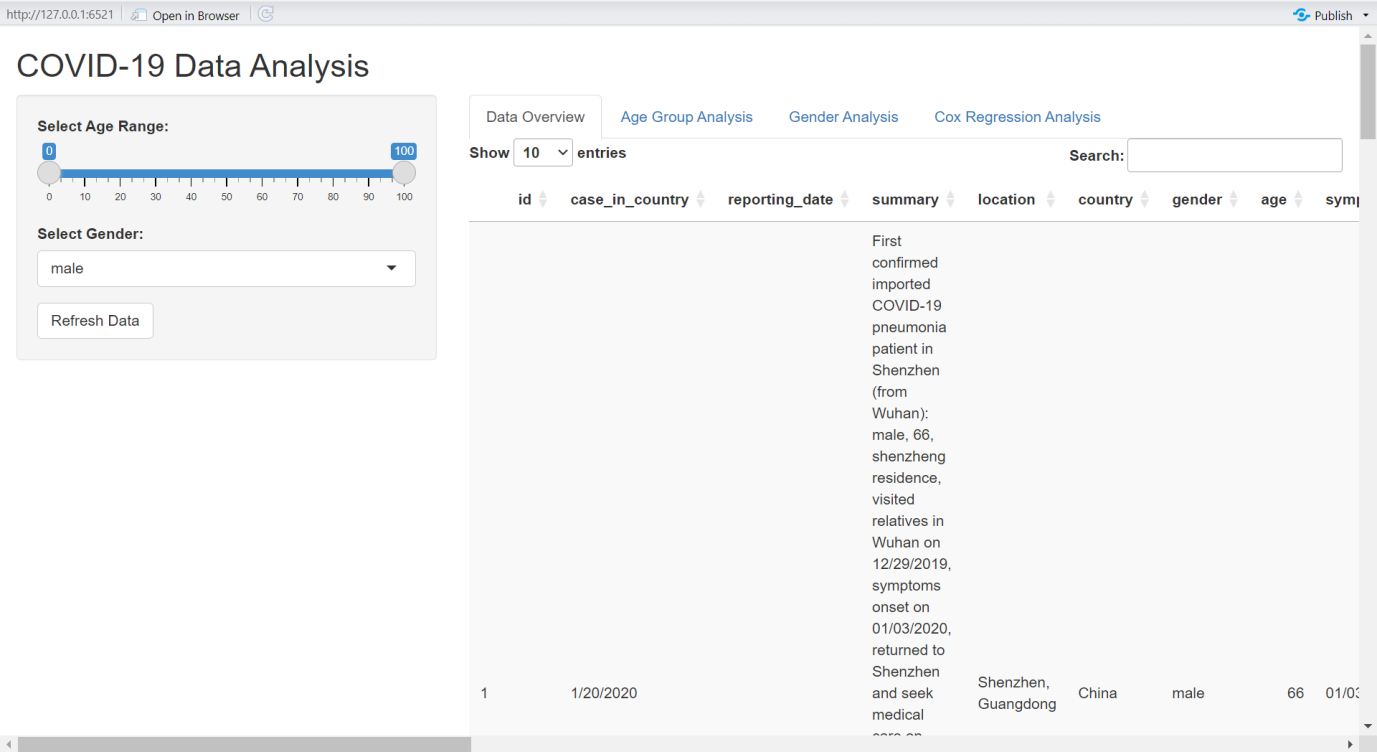
The Cox model provided a predictive layer, offering insights into how age and gender impact survival probability over time:

* **Age Effect on Hazard Ratio:** The analysis showed a clear trend—higher age correlated with increased mortality risk. For every unit increase in age, the hazard ratio increased, showing that each additional year of age added a quantifiable risk.
* **Gender Influence on Mortality:** Gender was a significant variable in the model, with males exhibiting a higher hazard ratio. This aligns with observed trends in COVID-19 data globally and suggests that men have an elevated baseline risk of severe outcomes.
* **Kaplan-Meier Survival Curves:** The survival curves revealed that females had a higher probability of survival over time compared to males. This visual representation helped clarify that, while both genders were affected by COVID-19, the survival rates differed significantly.

**4.5** COVID-19 Mortality Dashboard**:**

The Shiny dashboard enabled an interactive and user-driven exploration of the data:

* **Data Filtering Options:** Users could dynamically filter data by age and gender, providing a hands-on approach to data exploration. This feature proved particularly valuable for users interested in specific demographic slices.
* **Graphical Representation of Survival and Mortality Data:** The interactive graphs and Kaplan-Meier curves offered intuitive and accessible insights into age and gender impacts. Users could visualize the survival probabilities and understand mortality risk changes across demographic groups.
* **Usability for Policymakers and Researchers:** The dashboard offers a practical resource for policymakers and healthcare researchers, who can quickly access, filter, and interpret relevant COVID-19 data. This utility makes it a potentially valuable tool for evidence-based decision-making in healthcare resource allocation.



**Chapter 5 – Conclusion and Future Scope**

**5.1 Conclusion:**

This study established that age and gender are significant factors influencing COVID-19 mortality risk. Specifically:

* **Age as a Critical Mortality Factor:** The findings confirmed that older individuals face a disproportionately higher risk of severe COVID-19 outcomes. The increase in mortality risk with age underscores the need for age-based protective measures, particularly in older populations.
* **Gender Differences in COVID-19 Outcomes:** Males showed a higher risk of mortality, a pattern observed in other studies, possibly due to both biological predispositions and behavioral factors. The identification of this gender-based vulnerability suggests that targeted interventions for male demographics could improve overall outcomes.
* **Dashboard Utility and Insights:** The interactive dashboard provides users with the flexibility to explore demographic nuances in COVID-19 data, offering real-time insights. This tool could be particularly valuable in understanding pandemic trends and supporting targeted public health responses.

**5.2 Future Sope:**

This study established that age and gender are significant factors influencing COVID-19 mortality risk. Specifically:

* **Age as a Critical Mortality Factor:** The findings confirmed that older individuals face a disproportionately higher risk of severe COVID-19 outcomes. The increase in mortality risk with age underscores the need for age-based protective measures, particularly in older populations.
* **Gender Differences in COVID-19 Outcomes:** Males showed a higher risk of mortality, a pattern observed in other studies, possibly due to both biological predispositions and behavioral factors. The identification of this gender-based vulnerability suggests that targeted interventions for male demographics could improve overall outcomes.
* **Dashboard Utility and Insights:** The interactive dashboard provides users with the flexibility to explore demographic nuances in COVID-19 data, offering real-time insights. This tool could be particularly valuable in understanding pandemic trends and supporting targeted public health responses.