

|  |  |  |  |
| --- | --- | --- | --- |
| 1 | 21F-9110 | Syed Fasih Zaidi | BSCS-6C |
| 2 | 21F-9242 | Sufiyan | BSCS-6C |
| 3 | 22F-3661 | Imama kainat | BSSE-4A |
| 4 | 22F-3738 | Zainab Eman | BSSE-4A |
| 5 | 22F-3634 | Noor Fatima | BSSE-4A |

**Statistics and Probability**

**Topic: Corona Virus in Pakistan**

**Course Instructor: Dr. Sana Shahid**

**Group Members:**

**Content** :

[**Corona Virus in Pakistan** 2](#_Toc166184088)

[**Problem Statement** 2](#_Toc166184089)

[**Objective** 3](#_Toc166184090)

[**Data Description** 3](#_Toc166184091)

[**Results** 3](#_Toc166184092)

[**Main Interface** 3](#_Toc166184093)

[**Summary Description of Data** 4](#_Toc166184094)

[**Probabilities of death and recovery** 5](#_Toc166184095)

[Probability of Death and Recovery by province: 5](#_Toc166184096)

[ **Probabilities of Recovery by Province:** 5](#_Toc166184097)

[**Scatter Plot of Total Deaths by Province:** 6](#_Toc166184098)

[**Pie Chart of Cases in Each Province:** 6](#_Toc166184099)

[**Pie Chart of Deaths in Each Province:** 7](#_Toc166184100)

[**Pie Chart of Recovered Cases in Each Province:** 7](#_Toc166184101)

[**Component bar Graph:** 8](#_Toc166184102)

[**Predictions of data using linear model** 9](#_Toc166184103)

[**R codes** 10](#_Toc166184104)

[**Conclusion** 15](#_Toc166184105)

[**Case Summary:** 15](#_Toc166184106)

[**Probabilities:** 15](#_Toc166184107)

[**Graphical Analysis:** 16](#_Toc166184108)

[ **Scatter Plot of Total Deaths by Province**:. 16](#_Toc166184109)

[ **Pie Charts of Distribution:** 16](#_Toc166184110)

[ **Component Bar Graphs:**. 16](#_Toc166184111)

[**Predictions:** 17](#_Toc166184112)

[ **Linear Regression Analysis**: 17](#_Toc166184113)

[ **Insight into Distribution Variability**:. 17](#_Toc166184114)

[ **Utilization of Predictive Models**: 17](#_Toc166184115)

[ **Strategic Planning Support:** . 17](#_Toc166184116)

[**Conclution:** 17](#_Toc166184117)

# **Corona Virus in Pakistan**

## **Problem Statement**

The dataset contains information on the total number of COVID-19 cases, total deaths, and recovery cases for every province of Pakistan. This data provides insights into the spread and impact of the virus at a local level. By analyzing the COVID-19 situation in different cities within each province, we can identify areas with high case counts, assess the severity of the outbreak, and monitor the recovery progress. This dataset serves as a valuable resource for understanding the regional dynamics of the pandemic and guiding targeted interventions to control its spread.

## **Objective**

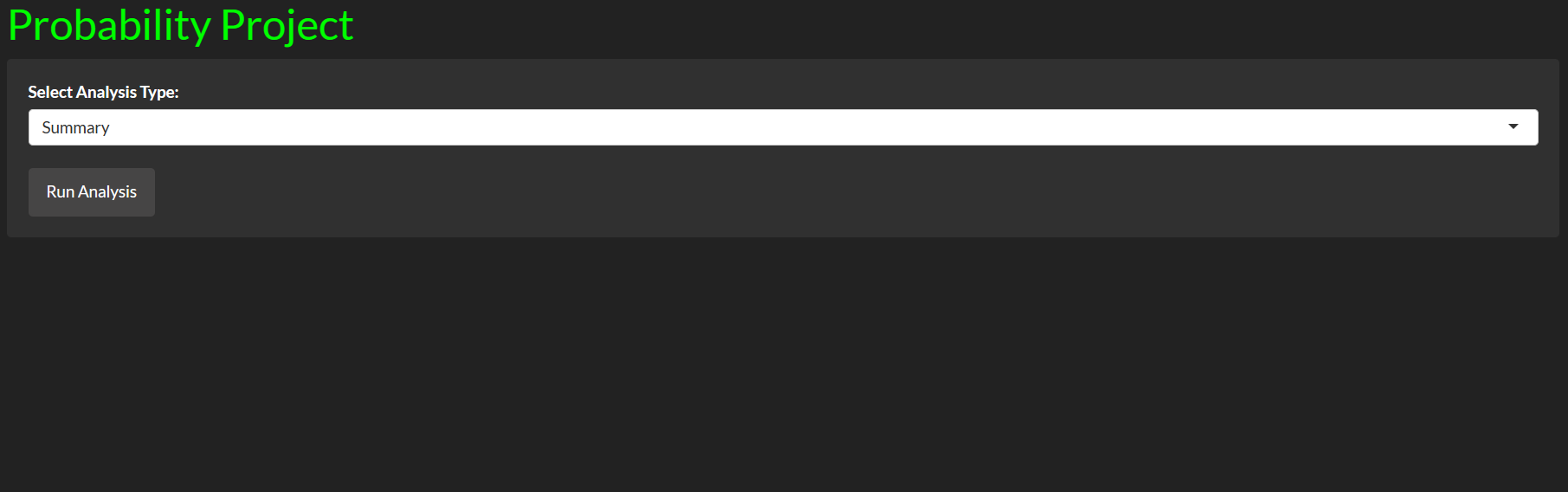
* Analyze city-level COVID-19 data across each province in Pakistan, focusing on total cases, deaths, and recovery rates.
* Provide a comprehensive understanding of the regional impact of the virus on different provinces.
* Identify critical areas with high infection rates or inadequate recovery progress.
* Assess the effectiveness of the measures implemented to control the spread of the virus.
* Facilitate targeted interventions and efficient allocation of resources by understanding pandemic trends at the city level.
* Support decision-making processes for more effective management of COVID-19 through detailed data analysis.

## **Data Description**

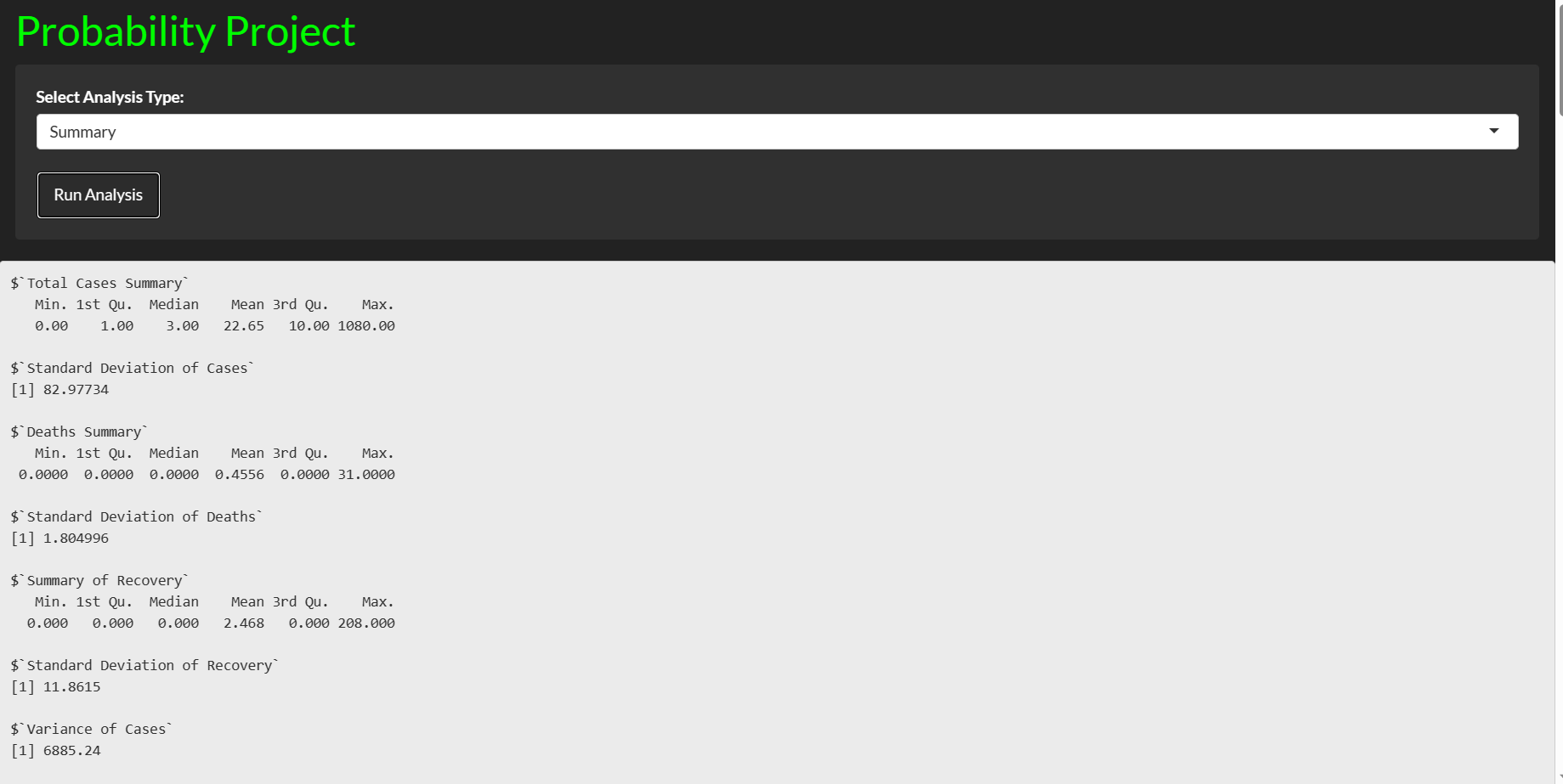
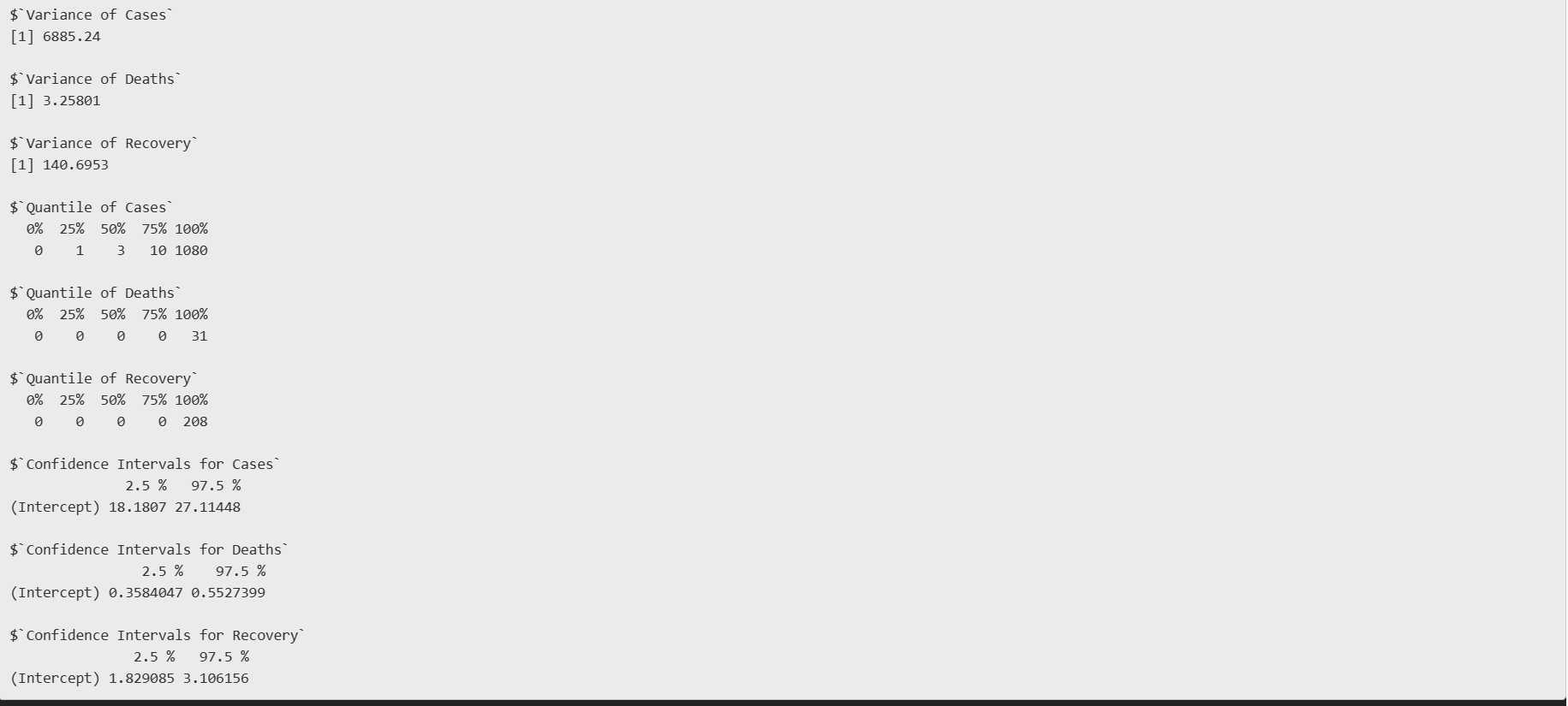
The dataset can be accessed at the following URL: [Kaggle: Pakistan COVID-19 City-Wise Data](https://www.kaggle.com/datasets/zusmani/pakistan-corona-virus-citywise-data). It encompasses data on COVID-19 cases in Pakistan, covering the period from February 26, 2020, to April 19, 2020, and includes city-level details for each province.

## **Results**

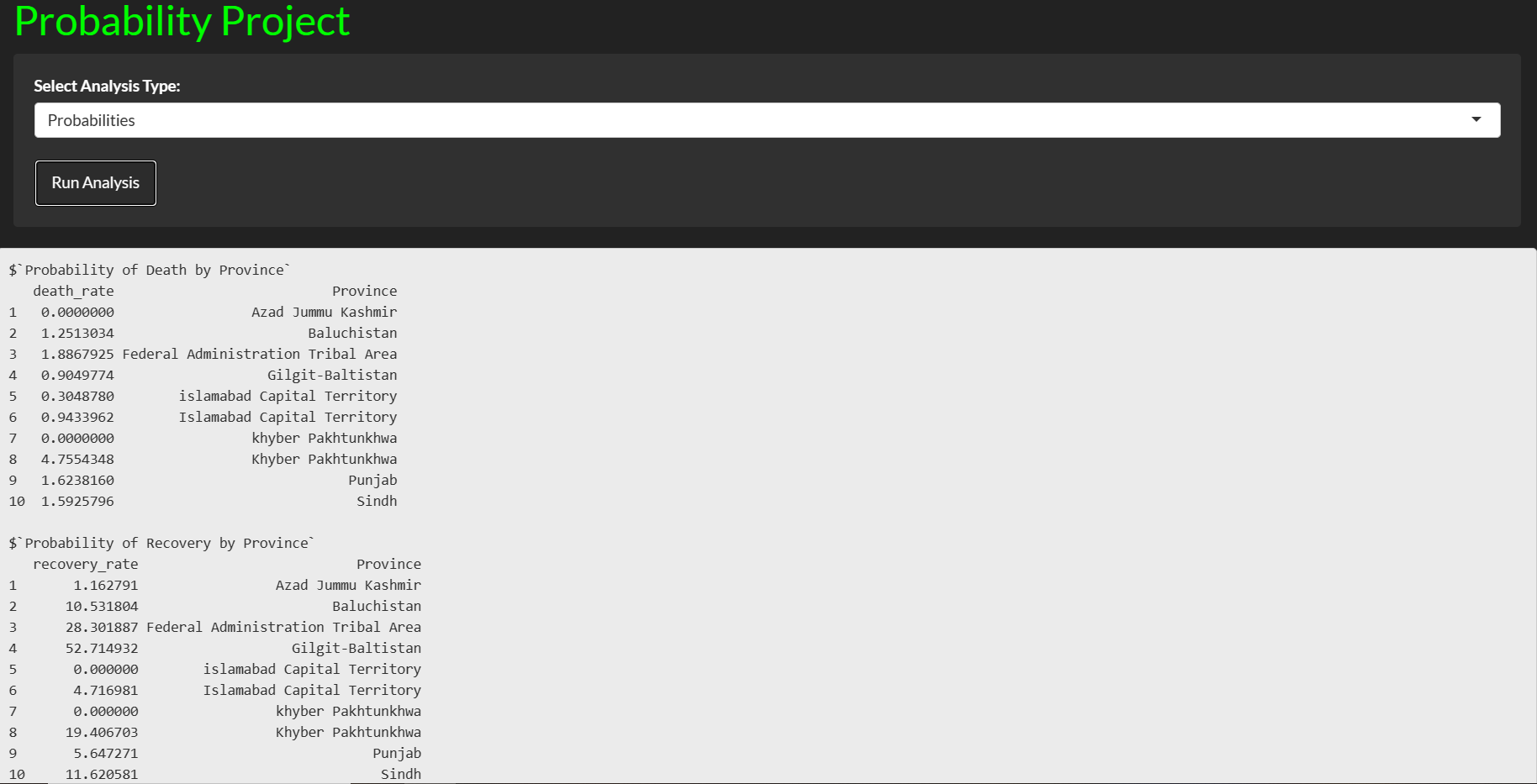
### **Main Interface**



### **Summary Description of Data**

### **Probabilities of death and recovery**



### Probability of Death and Recovery by province:

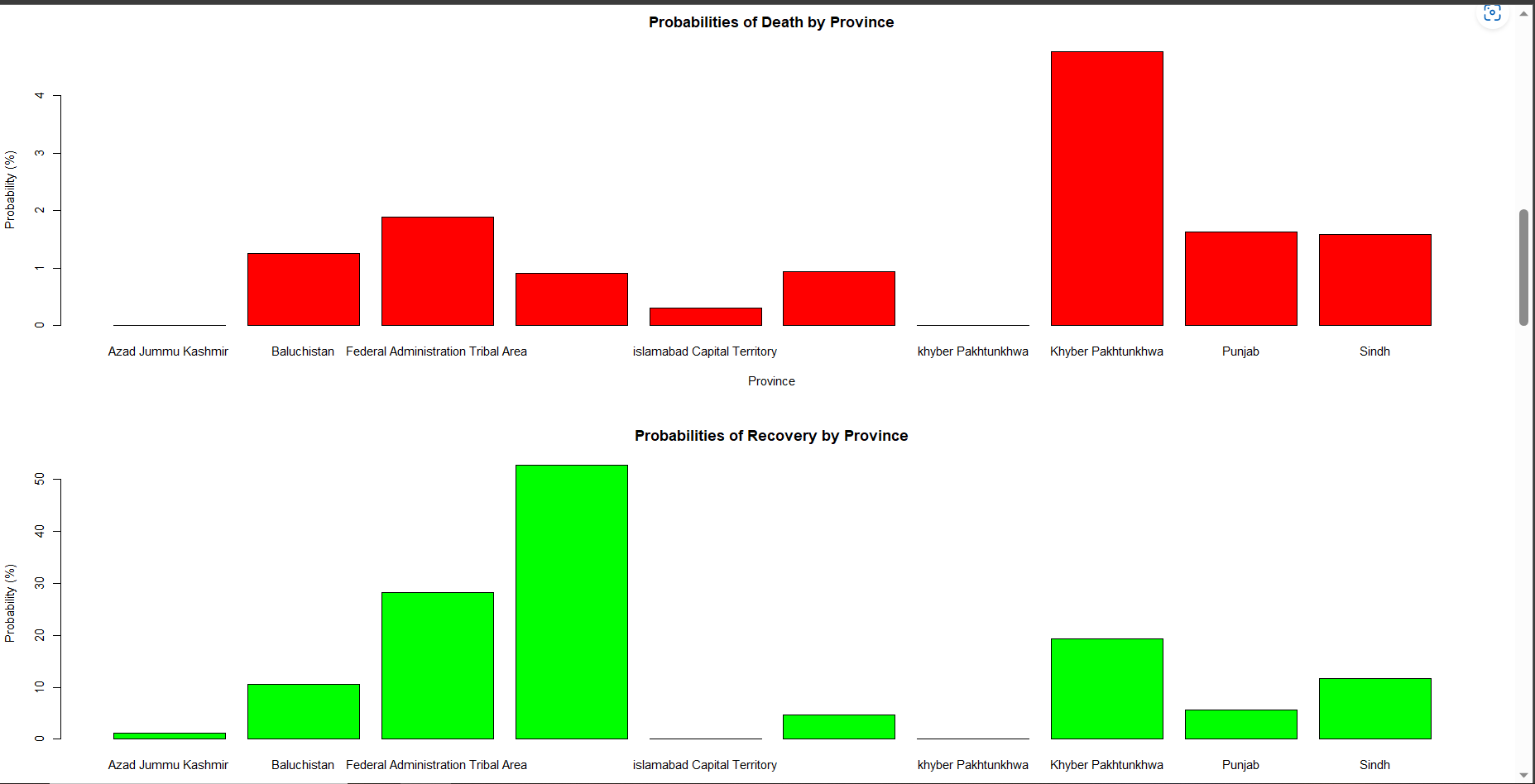
The graph displays two sets of bar charts depicting the:

* **Probabilities of Death by Province:**

This red bar chart represents the likelihood of death from COVID-19 in various provinces. Each bar's height indicates the percentage probability, with some provinces like Punjab and Sindh showing significantly higher probabilities compared to others like Azad Jammu Kashmir or the Federal Administration Tribal Area.

#### **Probabilities of Recovery by Province:**

The green bar chart below shows the recovery probabilities across the same set of provinces. The recovery rates also vary, with provinces like Sindh and Punjab demonstrating relatively higher recovery probabilities, reflecting a higher number of recoveries in relation to total cases in these regions.

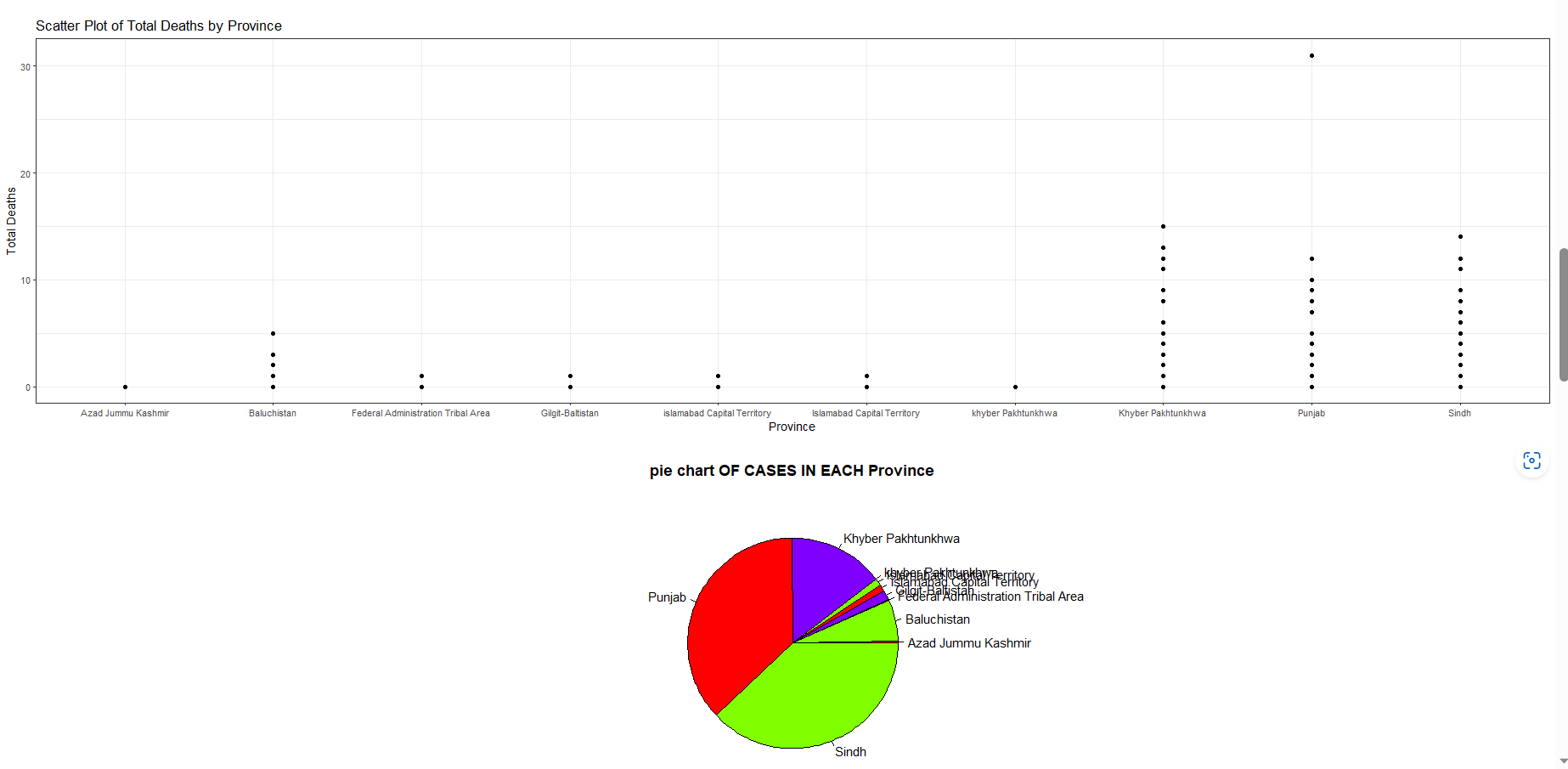


### **Scatter Plot of Total Deaths by Province:**

graph is located in the upper half of the image. It shows a scatter plot where each dot represents the total number of COVID-19 deaths in a specific province. The provinces are labeled along the x-axis, and the y-axis represents the total number of deaths. The plot provides a clear visual of the distribution of deaths across provinces, with the distribution of points indicating variations in mortality rates.

### **Pie Chart of Cases in Each Province:**

The lower half of the image features a pie chart that represents the proportion of COVID-19 cases in each province. Different colors are used to distinguish the provinces. The larger segments indicate provinces with a higher number of cases. This visual allows for an immediate perception of which provinces are most affected by the number of cases, with provinces like Punjab, Sindh, and Khyber Pakhtunkhwa appearing to have a larger share of cases.

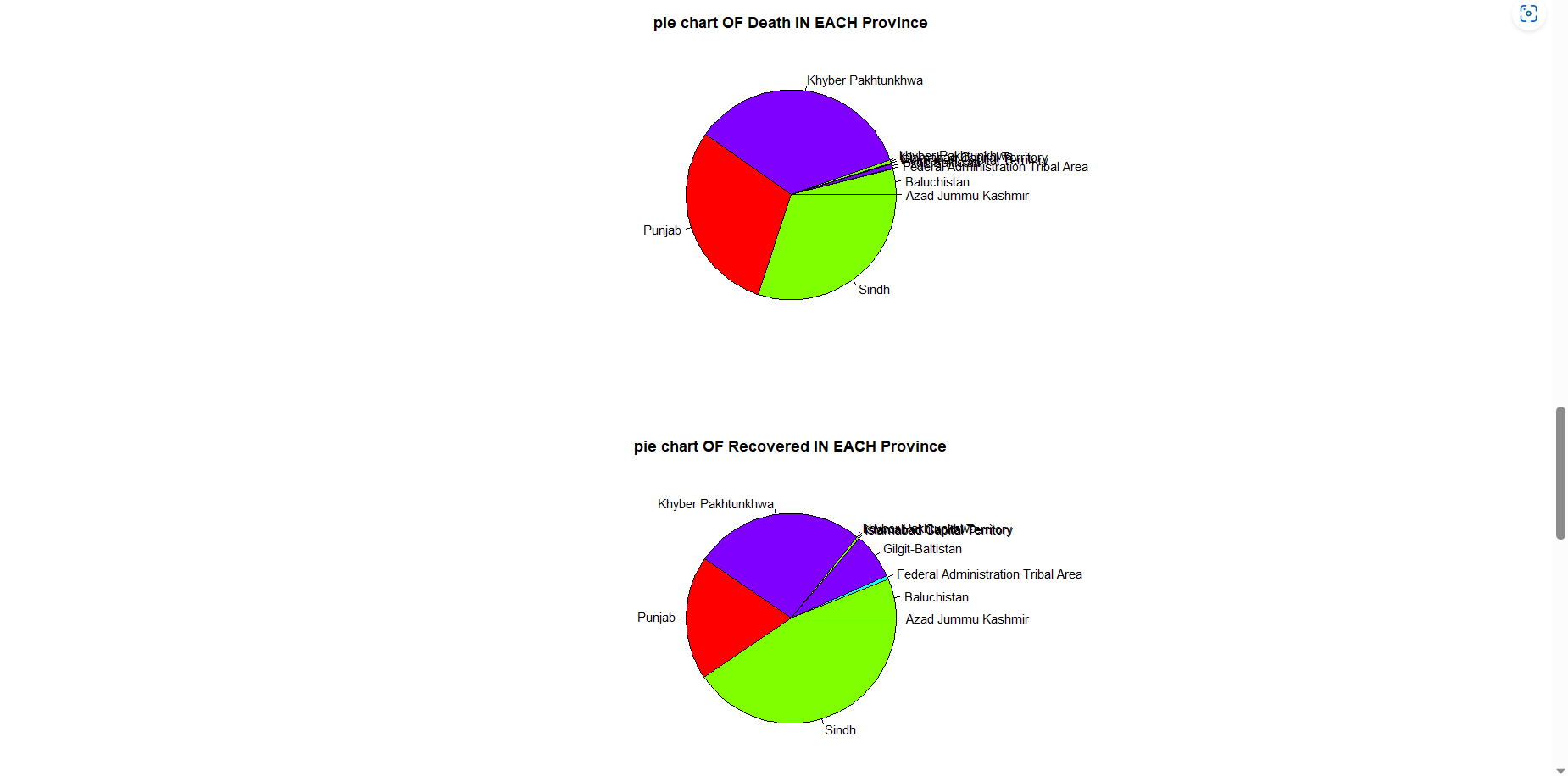


### **Pie Chart of Deaths in Each Province:**

Located in the upper half of the image, this pie chart shows the percentage of total COVID-19 deaths in each province. Each segment is color-coded to represent a different province, making it easy to distinguish between them. Provinces like Sindh, Punjab, and Khyber Pakhtunkhwa appear to have a larger share of deaths, indicating a higher impact of the virus in these regions.

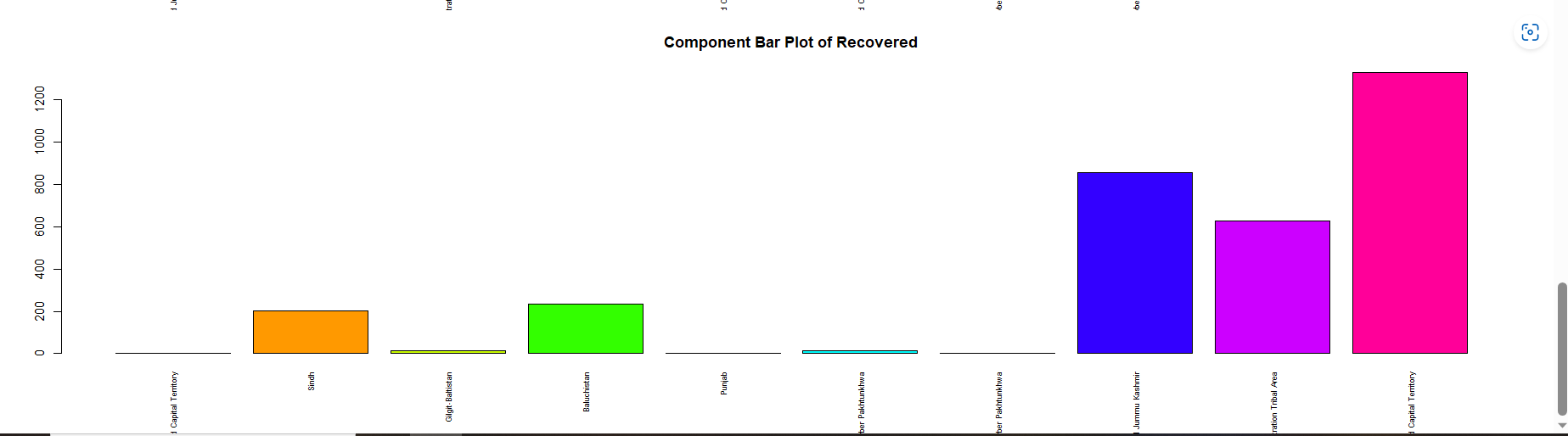
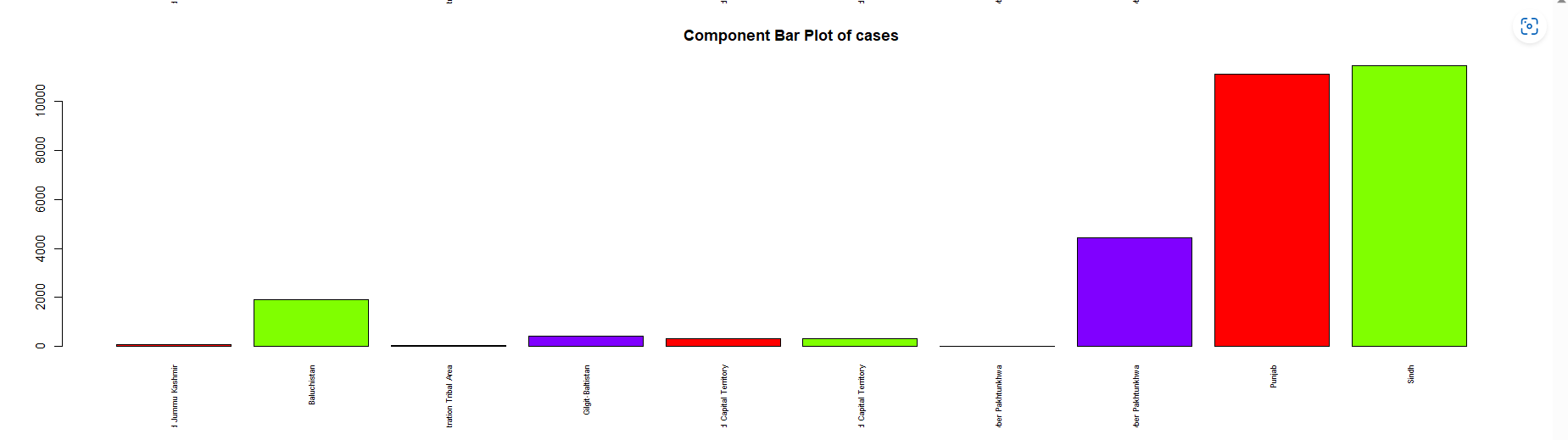
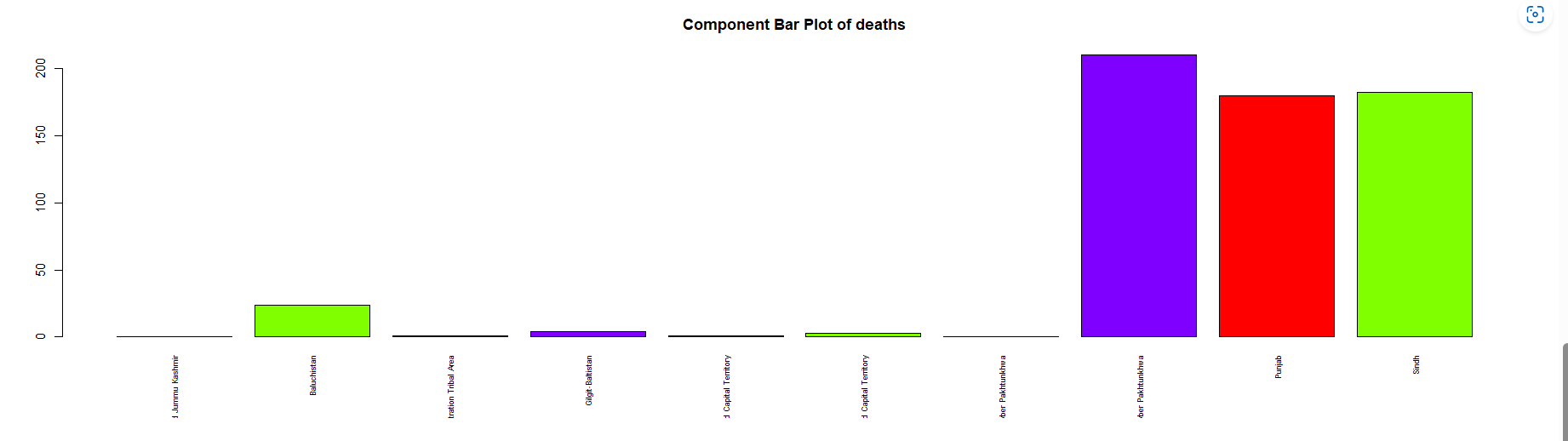
### **Pie Chart of Recovered Cases in Each Province:**

The lower pie chart displays the proportion of COVID-19 recoveries in each province. Similar to the deaths chart, different colors are used for each province to facilitate quick visual comprehension. The segments of the chart highlight the recovery distribution, with provinces like Punjab, Sindh, and Khyber Pakhtunkhwa again showing significant portions, reflecting a large number of recoveries relative to their total cases.



### **Component bar Graph:**

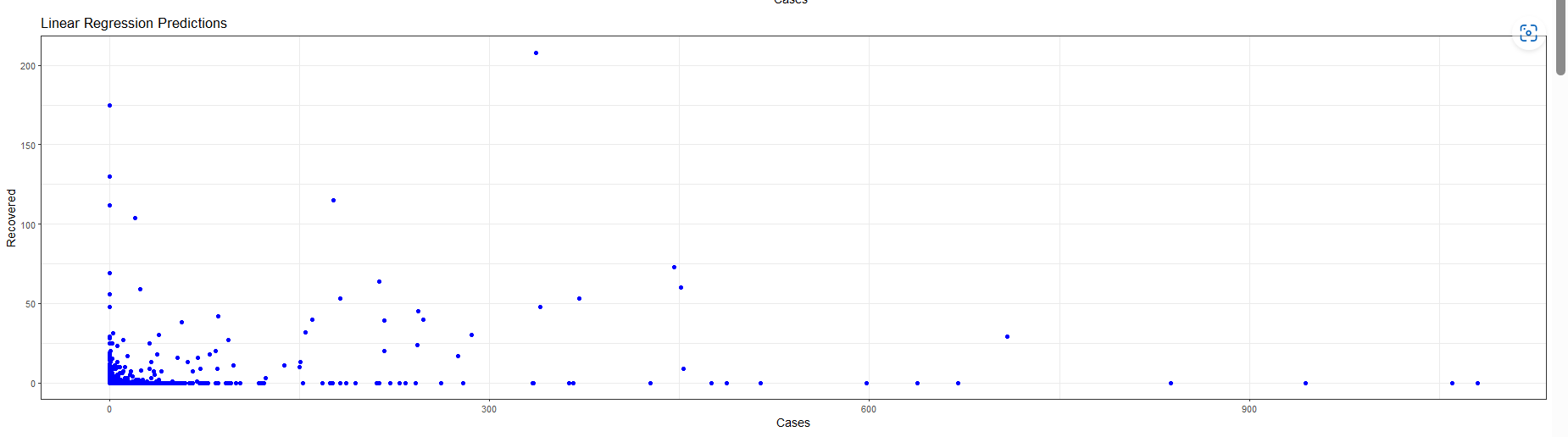
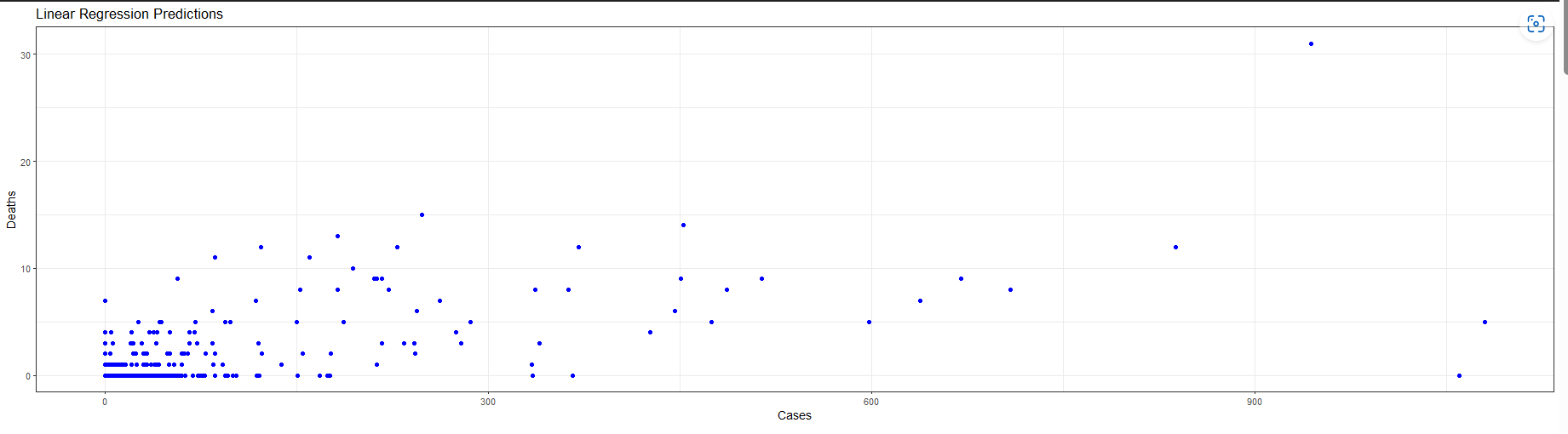
The image displays a component bar plot of COVID-19 deaths across various provinces in Pakistan, with each bar representing a different province and color-coded for clarity. The heights of the bars indicate the relative number of deaths in each province. The chart clearly shows a disparity in death tolls, with some provinces (notably those represented by the tallest purple and red bars) experiencing significantly higher fatalities compared to others. This visualization is useful for quickly identifying the provinces most affected by the pandemic and may assist in directing healthcare resources and interventions more effectively.



### **Predictions of data using linear model**

Predicted number of deaths: Predicted number of deaths for Cases = 15 out of 1000

Predicted number of recoveries: Predicted number of Recovered for Cases = 30 out of 1000



Each of these graphs plays a crucial role in our analysis, helping to paint a comprehensive picture of the COVID-19 situation in Pakistan. By exploring these visualizations, we gain valuable insights into the spread of the virus, the effectiveness of response measures, and potential areas needing more focused intervention.

## **R codes**

library(shiny)

library(readr)

library(ggplot2)

library(shinythemes)

mydata <- read\_csv("Pkcovid.csv")

ui <- fluidPage(

theme = shinytheme("darkly"),

tags$head(

tags$style(HTML("

h1 {

color: lime;

}

h2 {

color: lime;

font-size: 14px;

}

#tabularFormCases td {

color: red;

}

"))

),

tags$h1("Probability Project"),

sidebarLayout(

sidebarPanel(width=12,

selectInput("analysisType", "Select Analysis Type:",

choices = c("Summary",

"Probabilities","Graphs",

"Predictions"),

selected = "Summary of Cases"),

conditionalPanel(

condition = "input.analysisType == 'Predictions'",

numericInput("newCase", "Case Values For Deaths Prediction:", value = 1000),

numericInput("newCase1", "Cases Value FOr Recovery Prediction:", value = 1000)

),

actionButton("runAnalysis", "Run Analysis")

),

mainPanel(

width = 39,

verbatimTextOutput("analysisResult"),

textOutput("txt"),

verbatimTextOutput("regression\_deaths"),

verbatimTextOutput("regression\_deaths\_2"),

plotOutput("predictionPlot"),

plotOutput("plotResult"),

plotOutput("plotR"),

plotOutput("plotResult1"),

plotOutput("plotResult2"),

plotOutput("plotResult3"),

plotOutput("plotResult4"),

plotOutput("componentplot"),

plotOutput("componentplot\_cases"),

plotOutput("componentplot\_recovery")

)

)

)

server <- function(input, output) {

output$txt <- renderText({}) # Clear the output initially

observeEvent(input$runAnalysis, {

output$txt <- renderText({})

analysis <- switch(input$analysisType,

"Summary" = {

case\_summary <- summary(mydata$Cases)

death\_summary <- summary(mydata$Deaths)

case\_summary <- summary(mydata$Cases)

recovery\_summary <- summary(mydata$Recovered)

case\_std <- sd(mydata$Cases)

death\_std <- sd(mydata$Deaths)

recovery\_std <- sd(mydata$Recovered)

var\_cases <- var(mydata$Cases)

var\_death <- var(mydata$Deaths)

var\_recovery <- var(mydata$Recovered)

q\_cases <- quantile(mydata$Cases)

q\_death <- quantile(mydata$Deaths)

q\_recovery <- quantile(mydata$Recovered)

case\_ci <- confint(lm(mydata$Cases ~ 1))

death\_ci <- confint(lm(mydata$Deaths ~ 1))

recovery\_ci <- confint(lm(mydata$Recovered ~ 1))

output$plotResult <- renderPlot(NULL)

output$plotR <- renderPlot(NULL)

output$regression\_deaths<-NULL

output$plotResult1 <- NULL

output$plotResult2 <- NULL

output$plotResult3 <- NULL

output$componentplot <- NULL

output$componentplot\_cases <- NULL

output$componentplot\_recovery <- NULL

output$predictionPlot<-NULL

list(

"Total Cases Summary" = case\_summary,

"Standard Deviation of Cases" = case\_std,

"Deaths Summary" = death\_summary,

"Standard Deviation of Deaths" = death\_std,

"Summary of Recovery" = recovery\_summary,

"Standard Deviation of Recovery" = recovery\_std,

"Variance of Cases" = var\_cases,

"Variance of Deaths" = var\_death,

"Variance of Recovery" = var\_recovery,

"Quantile of Cases" = q\_cases,

"Quantile of Deaths" = q\_death,

"Quantile of Recovery" = q\_recovery,"Confidence Intervals for Cases" = case\_ci,

"Confidence Intervals for Deaths" = death\_ci,

"Confidence Intervals for Recovery" = recovery\_ci

)

},

"Probabilities" = {

output$regression\_deaths\_2<-NULL

output$plotResult<-NULL

output$regression\_deaths<-NULL

output$plotResult1 <- NULL

output$plotResult2 <- NULL

output$plotResult3 <- NULL

output$componentplot <- NULL

output$componentplot\_cases <- NULL

output$componentplot\_recovery <- NULL

output$regression\_deaths<-NULL

output$predictionPlot<-NULL

# Calculate death probabilities for each province

total <- aggregate(mydata$Deaths ~ mydata$Province, data = mydata, sum)

total\_cases <- aggregate(mydata$Cases ~ mydata$Province, data = mydata, sum)

death\_rate <- (total$`mydata$Deaths`) / (total\_cases$`mydata$Cases`) \* 100

result <- data.frame(death\_rate, Province = total$`mydata$Province`)

# Calculate recovery probabilities for each province

r\_total <- aggregate(mydata$Recovered ~ mydata$Province, data = mydata, sum)

r\_total\_cases <- aggregate(mydata$Cases ~ mydata$Province, data = mydata, sum)

recovery\_rate <- (r\_total$`mydata$Recovered`) / (r\_total\_cases$`mydata$Cases`) \* 100

r\_result <- data.frame(recovery\_rate, Province = r\_total$`mydata$Province`)

# Create a list of the results

probabilities <- list(

"Probability of Death by Province" = result,

"Probability of Recovery by Province" = r\_result

)

# Plot the bar chart

output$plotResult <- renderPlot({

barplot(death\_rate, names.arg = total$`mydata$Province`,

main = "Probabilities of Death by Province",

xlab = "Province", ylab = "Probability (%)",col="red")

})

output$plotR <- renderPlot({

barplot(recovery\_rate, names.arg = total$`mydata$Province`,

main = "Probabilities of Recovery by Province",

xlab = "Province", ylab = "Probability (%)",col="green")

})

probabilities

},

"Graphs"={

output$regression\_deaths\_2<-NULL

output$plotResult<-NULL

output$regression\_deaths<-NULL

output$plotResult1 <- NULL

output$plotResult2 <- NULL

output$plotResult3 <- NULL

output$componentplot <- NULL

output$componentplot\_cases <- NULL

output$componentplot\_recovery <- NULL

output$regression\_deaths<-NULL

output$predictionPlot<-NULL

output$plotResult1 <- renderPlot({

ggplot(mydata, aes(x = Province, y = Deaths)) +

geom\_point() +

labs(x = "Province", y = "Total Deaths",

title = "Scatter Plot of Total Deaths by Province") +

theme\_bw()

})

output$plotResult2 <- renderPlot({

province=aggregate(mydata$Cases~mydata$Province,data=mydata,sum)

pie(province$`mydata$Cases`,label=province$`mydata$Province`,main="pie chart OF CASES IN EACH Province",col=rainbow(4))

})

output$plotResult3 <- renderPlot({

city=aggregate(mydata$Deaths~mydata$Province,data=mydata,sum)

pie(city$`mydata$Deaths`,label=city$`mydata$Province`,main="pie chart OF Death IN EACH Province",col=rainbow(4))

})

output$plotResult4 <- renderPlot({

city=aggregate(mydata$Recovered~mydata$Province,data=mydata,sum)

pie(city$`mydata$Recovered`,label=city$`mydata$Province`,main="pie chart OF Recovered IN EACH Province",col=rainbow(4))

})

output$componentplot <- renderPlot({

death\_1=aggregate(mydata$Deaths~mydata$Province,data=mydata,sum)

label\_1=c(death\_1$`mydata$Province`)

barplot(death\_1$`mydata$Deaths`,main="Component Bar Plot of deaths",col=rainbow(4),names.arg = label\_1, las = 3,cex.names=0.6)

})

output$componentplot\_cases <- renderPlot({

death\_1=aggregate(mydata$Cases~mydata$Province,data=mydata,sum)

label\_1=c(death\_1$`mydata$Province`)

barplot(death\_1$`mydata$Cases`,main="Component Bar Plot of cases",col=rainbow(4),names.arg = label\_1, las = 3,cex.names=0.6)

})

output$componentplot\_recovery <- renderPlot({

death\_1 <- aggregate(mydata$Recovered ~ mydata$Province, data = mydata, sum)

label\_1 <- as.character(unique(mydata$Province))

barplot(death\_1$`mydata$Recovered`, main = "Component Bar Plot of Recovered", col = rainbow(length(label\_1)), names.arg = label\_1,las = 3,cex.names=0.6)

})

output$pie<-renderPlot({

total\_cases=c(sum(mydata$Cases),sum(mydata$Recovered),sum(mydata$Deaths))

lables\_name=c("Total Cases","Total Recovered","Total Deaths")

pie(total\_cases,label=lables\_name,main="Corona Cases In Pakistan",col=rainbow(3))

})

},

"Predictions" = {

output$plotResult1 <- NULL

output$plotResult2 <- NULL

output$plotResult3 <- NULL

output$componentplot <- NULL

output$componentplot\_cases <- NULL

output$componentplot\_recovery <- NULL

result <- lm(Deaths ~ Cases, data = mydata)

new\_case <- input$newCase

predicted\_deaths <- predict(result, newdata = data.frame(Cases = new\_case))

output$regression\_deaths <- renderPrint({

paste("Predicted number of deaths for Cases =", new\_case, "is", round(predicted\_deaths))

})

output$predictionPlot <- renderPlot({

new\_data <- data.frame(Cases = new\_case, Deaths = predicted\_deaths)

ggplot() +

geom\_point(data = mydata, aes(x = Cases, y = Deaths), color = "blue") +

geom\_line(data = new\_data, aes(x = Cases, y = Deaths), linetype = "dashed", color = "red") +

labs(x = "Cases", y = "Deaths", title = "Linear Regression Predictions") +

theme\_bw()

})

result1 <- lm(Recovered ~ Cases, data = mydata)

new\_case1 <- input$newCase1

predicted\_recovered <- predict(result1, newdata = data.frame(Cases = new\_case1))

output$regression\_deaths\_2 <- renderPrint({

paste("Predicted number of Recovered for Cases =", new\_case1, "is", round(predicted\_recovered))

})

output$plotResult <- renderPlot({

new\_data1 <- data.frame(Cases = new\_case1, Recovered = predicted\_recovered)

ggplot() +

geom\_point(data = mydata, aes(x = Cases, y = Recovered), color = "blue") +

geom\_line(data = new\_data1, aes(x = Cases, y = Recovered), linetype = "dashed", color = "red") +

labs(x = "Cases", y = "Recovered", title = "Linear Regression Predictions") +

theme\_bw()

})

}

)

output$analysisResult <- renderPrint({

analysis

})

})

}

shinyApp(ui, server)

## **Conclusion**

In this project, we analyzed the cases of COVID-19 in Pakistan and explored various aspects of the data. Here are the main findings:

### **Case Summary:**

The total number of reported cases is 1080 along with recoveries and deaths. The Standard deviation of total cases is 82.97734 which display significant variation in the number of cases in different regions of Pakistan.

### **Probabilities:**

We calculated the probability of death and recovery by province. The probability of death differs in each province according to the number of cases as shown in above results screenshots

### **Graphical Analysis:**

#### **Scatter Plot of Total Deaths by Province**: Provides an overview of the distribution of COVID-19 deaths across various regions, helping to identify provinces with higher mortality rates.

#### **Pie Charts of Distribution:** Illustrate the proportion of cases, deaths, and recoveries in each province, effectively highlighting the relative impact of the virus in different areas.

#### **Component Bar Graphs:** Display the contributions of each province to the national totals of cases, deaths, and recoveries, showcasing the extent of the pandemic's reach within specific regions.

### **Predictions:**

#### **Linear Regression Analysis**: Employed to forecast the number of deaths and recoveries based on the reported cases, enhancing predictive accuracy for planning purposes.

#### **Insight into Distribution Variability**: The analysis shows that the distribution of cases, deaths, and recoveries varies significantly from province to province, underscoring the importance of region-specific responses and resource allocation.

#### **Utilization of Predictive Models**: Predictions from linear regression models assist in estimating future deaths and recoveries, providing valuable data for strategic planning and response optimization.

#### **Strategic Planning Support:** These analytical findings are instrumental in understanding the evolving pandemic landscape, aiding in the development of effective strategies to combat COVID-19 across different provinces in Pakistan.

### **Conclution:**

In conclusion, this analysis has provided crucial insights into the spread and impact of COVID-19 across different provinces in Pakistan, underlining the necessity for region-specific strategies and responses.