R Notebook

ASSIGMENT SQL

PACKAGES AND LIB

```
install.packages("RJDBC")

system("apt-get install -y default-jdk")
```

JDBC driver (ex: MySQL, PostgreSQL..)

```
library(RJDBC)

install.packages("DBI")

install.packages("RSQLite")

# Cài đặt RJDBC
install.packages("RJDBC")

# Tải thư viện
library(RJDBC)

# Đường dẫn đến JDBC driver
drv <- JDBC("com.ibm.db2.jcc.DB2Driver", "D:/Kì
5/DSR301m/jdbc_sqlj/db2jcc4.jar")


install.packages("RMySQL")

library(RSQLite)
```

CONNECT DATABASE

```
library(RJDBC)

# Đường dẫn đến tệp jar của MySQL Connector/J
jdbc_driver <- JDBC("com.mysql.jdbc.Driver", "D:/Kì 5/DSR301m/mysql-
connector-j-9.0.0/mysql-connector-j-9.0.0.jar", identifier.quote="`")

# Cung cấp thông tin kết nối
hostname <- "localhost"
port <- "3306"
dbname <- "maianh"
username <- "maianh"
password <- "******"

# Tạo chuỗi kết nối
```

```
url <- paste0("jdbc:mysql://", hostname, ":", port, "/", dbname)

# Kết nối đến cơ sở dữ liệu
conn <- dbConnect(jdbc_driver, url, username, password)

# Kiểm tra kết nối
if (dbIsValid(conn)) {
  print("Kết nối thành công!")
} else {
  print("Không thể kết nối!")
}
```

READ DATA

```
# Đọc dữ liệu từ tệp CSV
crime_data <- read.csv("D:/Kì 5/DSR301m/ChicagoCrimeData.csv", header
= TRUE, sep = ",")


# Hiển thị một vài dòng dữ liệu để kiểm tra
head(crime_data)

# Hiển thị tổng quan dữ liệu (summary)
summary(crime_data)
```

IMPORT ChicagoCrimeData.CSV TO DATABADE (SQL)

```
# Nhập dữ liệu trực tiếp vào MySQL
dbWriteTable(conn, name = "crime_data", value = crime_data, row.names
= FALSE, overwrite = TRUE)
```

TEST DATABASE (SQL)

```
# Kiểm tra dữ liệu trong bảng MySQL
query <- "SELECT * FROM crime_data LIMIT 10;"

crime_data_preview <- dbGetQuery(conn, query)

print(crime_data_preview)
```

Problem 1

Total number of cases

```
query <- query <- "
  SELECT COUNT(*) AS total_cases
  FROM crime_data;
"

# Thực thi truy vấn
result <- dbGetQuery(conn, query)
```

```
# In kết quả
print(result)
```

Problem 2

Total number of cases by crime type

```
query <- "
  SELECT
    PRIMARY_TYPE,
    COUNT(*) AS total_cases
  FROM crime_data
  GROUP BY PRIMARY_TYPE
  ORDER BY total_cases DESC;
"

# Thực thi truy vấn
result <- dbGetQuery(conn, query)

# In kết quả
print(result)
```

Problem 3 Total number of cases by year

```
query <- "
  SELECT YEAR(date) AS year, COUNT(*) AS total_cases
  FROM crime_data
  GROUP BY YEAR(date)
  ORDER BY year;
"

# Thực thi truy vấn
result <- dbGetQuery(conn, query)

# In kết quả
print(result)
```

Problem 4

Number of cases without arrests (Non-Arrest)

```
query <- "
  SELECT COUNT(*) AS non_arrest_cases
  FROM crime_data
  WHERE arrest = 'FALSE';
"

# Thực thi truy vấn
result <- dbGetQuery(conn, query)
```

```
# In kết quả
print(result)
```

Problem 5

Total number of cases by community area

```
query <- "
  SELECT community_area_number, COUNT(*) AS total_cases
  FROM crime_data
  GROUP BY community_area_number
  ORDER BY community_area_number;
"

# Thực thi truy vấn
result <- dbGetQuery(conn, query)

# In kết quả
print(result)
```

Problem 6

Number of cases with missing location information

```
query <- "
  SELECT COUNT(*) AS missing_location_cases
  FROM crime_data
  WHERE location IS NULL;
"

# Thực thi truy vấn
result <- dbGetQuery(conn, query)

# In kết quả
print(result)
```

Problem 7

How does the trend of total crime incidents change over the years? Is there an increase or decrease?

```
query <- "
  SELECT YEAR(date) AS year, COUNT(*) AS total_cases
  FROM crime_data
  GROUP BY YEAR(date)
  ORDER BY year;
"

# Thực thi truy vấn
result <- dbGetQuery(conn, query)
```

```
# In kết quả
print(result)
```

Problem 8

Which types of crimes are increasing or decreasing over time? Is there a specific type that is particularly prevalent in recent years?

```
query <- "
  SELECT YEAR(date) AS year, primary_type, COUNT(*) AS total_cases
  FROM crime_data
  GROUP BY YEAR(date), primary_type
  ORDER BY year, primary_type;
"

# Thực thi truy vấn
result <- dbGetQuery(conn, query)

# In kết quả
print(result)
```

Problem 9

Which areas have a higher crime trend compared to others? Does this trend change over time?

```
query <-  "
  SELECT YEAR(date) AS year, block, COUNT(*) AS total_cases
  FROM crime_data
  GROUP BY YEAR(date), block
  ORDER BY block;
"

# Thực thi truy vấn
result <- dbGetQuery(conn, query)

# In kết quả
print(result)
```

Problem 10

Is there a particular time of day (morning, afternoon, evening) when crimes are more likely to occur?

```
query <- "
SELECT
  CASE
    WHEN HOUR(DATE) < 12 THEN 'Morning'
    WHEN HOUR(DATE) < 18 THEN 'Afternoon'
    ELSE 'Evening'
  END AS time_of_day,
```

```
  COUNT(*) AS total_cases
FROM crime_data
GROUP BY time_of_day
ORDER BY FIELD(time_of_day, 'Morning', 'Afternoon', 'Evening')"

# Thực thi truy vấn
result <- dbGetQuery(conn, query)

# In kết quả
print(result)
```

Problem 11

Is the crime trend affected by the season of the year? Which season typically has the highest number of incidents?

```
query <- "
SELECT
  CASE
    WHEN MONTH(DATE) IN (12, 1, 2) THEN 'Winter'
    WHEN MONTH(DATE) IN (3, 4, 5) THEN 'Spring'
    WHEN MONTH(DATE) IN (6, 7, 8) THEN 'Summer'
    WHEN MONTH(DATE) IN (9, 10, 11) THEN 'Fall'
  END AS Season,
  COUNT(*) AS total_cases
FROM crime_data
WHERE DATE IS NOT NULL
GROUP BY Season
ORDER BY total_cases DESC
"

# Thực thi truy vấn
result <- dbGetQuery(conn, query)

# In kết quả
print(result)
```

Problem 12

Which type of crime is the most prevalent in the entire dataset? What is the frequency of that type over the years?

```
query <- "
  SELECT
    PRIMARY_TYPE,
    COUNT(*) AS total_cases
  FROM crime_data
  GROUP BY crime_type
  ORDER BY total_cases DESC
  LIMIT 1;
"
```

```
# Thực thi truy vấn
result <- dbGetQuery(conn, query)

# In kết quả
print(result)
```

Problem 13

Which locations have the highest frequency of crimes? Are there any notable patterns or hotspots?

```
query <- "
SELECT
  LOCATION_DESCRIPTION,
  COUNT(*) AS total_cases
FROM crime_data
GROUP BY LOCATION_DESCRIPTION
ORDER BY total_cases DESC
LIMIT 10
"

# Thực thi truy vấn
result <- dbGetQuery(conn, query)

# In kết quả
print(result)
```

Problem 14

Is there a correlation between total incidents and arrest frequency? What percentage of incidents result in arrests?

```
query <- "
  SELECT
    COUNT(*) AS total_incidents,
    SUM(CASE WHEN arrest = 'TRUE' THEN 1 ELSE 0 END) AS total_arrests
  FROM crime_data;
"

# Thực thi truy vấn
result <- dbGetQuery(conn, query)

# In kết quả
print(result)

total_incidents <- as.numeric(result$total_incidents)
total_arrests <- as.numeric(result$total_arrests)

if (!is.na(total_incidents) && !is.na(total_arrests)) {
```

```
  if (total_incidents > 0) {
    percentage_arrests <- (total_arrests / total_incidents) * 100
    cat(sprintf("Percentage of total arrests compared to total
incidents: %.2f%%\n", percentage_arrests))
  } else {
    cat("No incidents found.\n")
  }
} else {
  cat("Error: Non-numeric data encountered.\n")
}

# Optionally, you can also print the total values for clarity
cat(sprintf("Total Incidents: %d\n", total_incidents))
cat(sprintf("Total Arrests: %d\n", total_arrests))
```

Problem 15

Which community areas have the highest frequency of crimes? Does this frequency change over time?

```
query <- "
SELECT
  COMMUNITY_AREA_NUMBER,
  YEAR,
  COUNT(*) AS total_cases
FROM crime_data
GROUP BY COMMUNITY_AREA_NUMBER, YEAR
ORDER BY total_cases DESC

"

# Thực thi truy vấn
result <- dbGetQuery(conn, query)

# In kết quả
print(result)
```

Problem 16

Which location descriptions have the highest frequency of crimes? Is there a difference between various location descriptions?

```
query <- "
SELECT
  LOCATION_DESCRIPTION,
  COUNT(*) AS total_cases
FROM crime_data
GROUP BY LOCATION_DESCRIPTION
```

```
ORDER BY total_cases DESC
LIMIT 10
"

# Thực thi truy vấn
result <- dbGetQuery(conn, query)

# In kết quả
print(result)
```

Problem 17

Which types of crimes are more prevalent in specific locations? Does this trend change over time?

```
query <-  "
  SELECT PRIMARY_TYPE, COUNT(*) AS total_cases, location
  FROM crime_data
  GROUP BY PRIMARY_TYPE
  ORDER BY total_cases DESC;
"

# Thực thi truy vấn
result <- dbGetQuery(conn, query)

# In kết quả
print(result)
```