**Import csv file using R and perform ETL operation using dplyr package**

# Install necessary packages if you haven't already

install.packages("dpylr")

install.packages("readr")

# Load the packages

library(dplyr)

library(readr)

# Create a sample dataset with 100 rows and some NA values

set.seed(123) # For reproducibility

data <- data.frame(

ID = 1:100,

Name = paste0("Name", 1:100),

Age = sample(c(20:60, NA), 100, replace = TRUE),

Salary = sample(c(40000:100000, NA), 100, replace = TRUE)

)

# Introduce some NA values randomly

data$Age[sample(1:100, 10)] <- NA

data$Salary[sample(1:100, 10)] <- NA

# View the sample data

print(data)

# Write the dataset to a CSV file

write\_csv(data, "Dataset.csv")

# Import the CSV file

data <- read\_csv("Dataset.csv")

# Transform the data

data\_transformed <- data %>%

# Select specific columns

select(Name, Age, Salary) %>%

# Filter rows

filter(Age > 35) %>%

# Create a new column

mutate(AnnualSalary = Salary \* 12) %>%

# Remove rows with missing values

filter(complete.cases(.))

# Load the transformed data to a new CSV file

write\_csv(data\_transformed, "transformed\_data.csv")

# View the transformed data

print(data\_transformed, n = 10)

**Output:**

ID Name Age Salary

1 1 Name1 50 52048

2 2 Name2 NA NA

3 3 Name3 33 45026

4 4 Name4 22 87981

5 5 Name5 NA 72605

6 6 Name6 56 56151

Rows: 100 Columns: 4

Name Age Salary AnnualSalary

<chr> <dbl> <dbl> <dbl>

1 Name1 50 52048 624576

2 Name6 56 56151 673812

3 Name8 44 54214 650568

4 Name9 45 54286 651432

5 Name10 46 63193 758316

6 Name12 46 74975 899700

7 Name15 48 52047 624564

8 Name16 54 88927 1067124

9 Name18 45 74723 896676

10 Name22 38 71541 858492