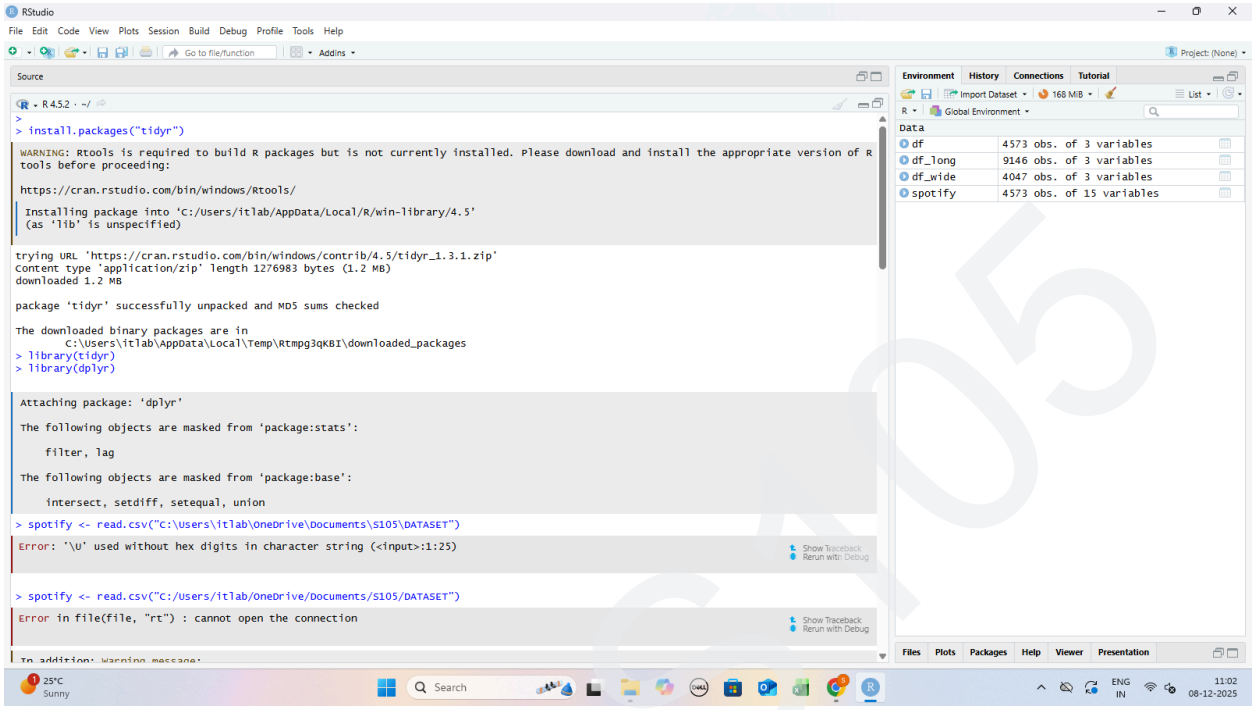
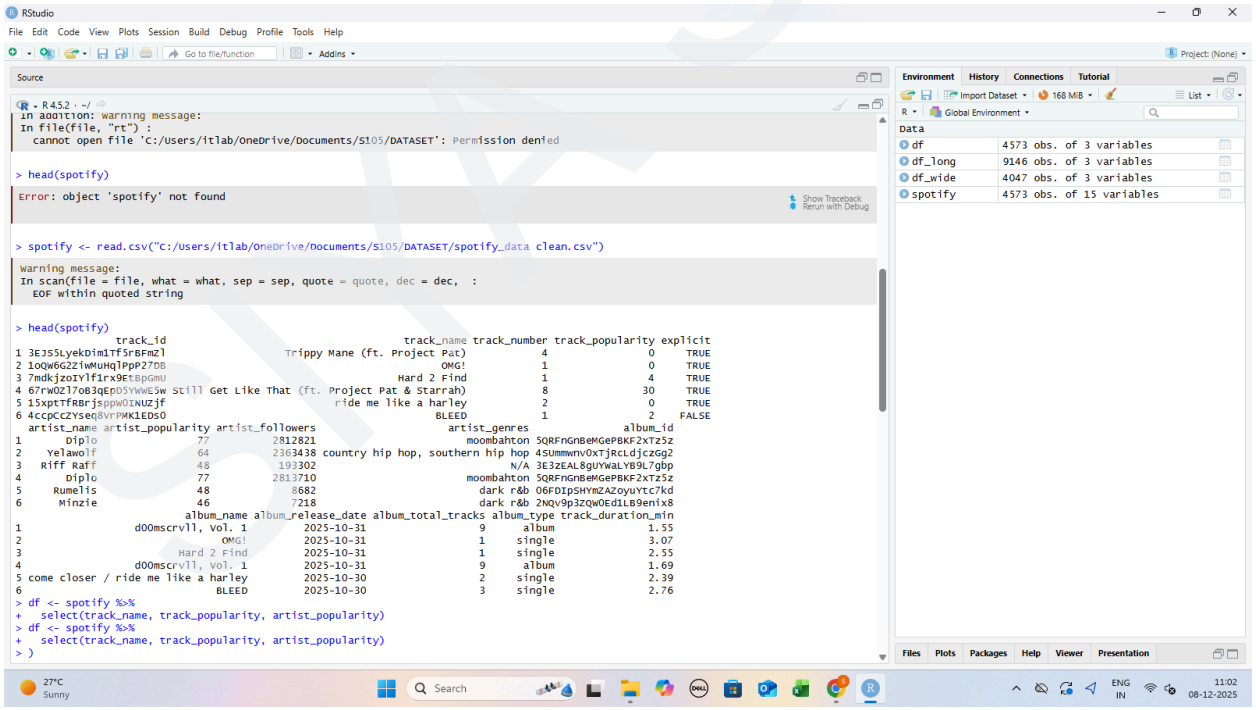
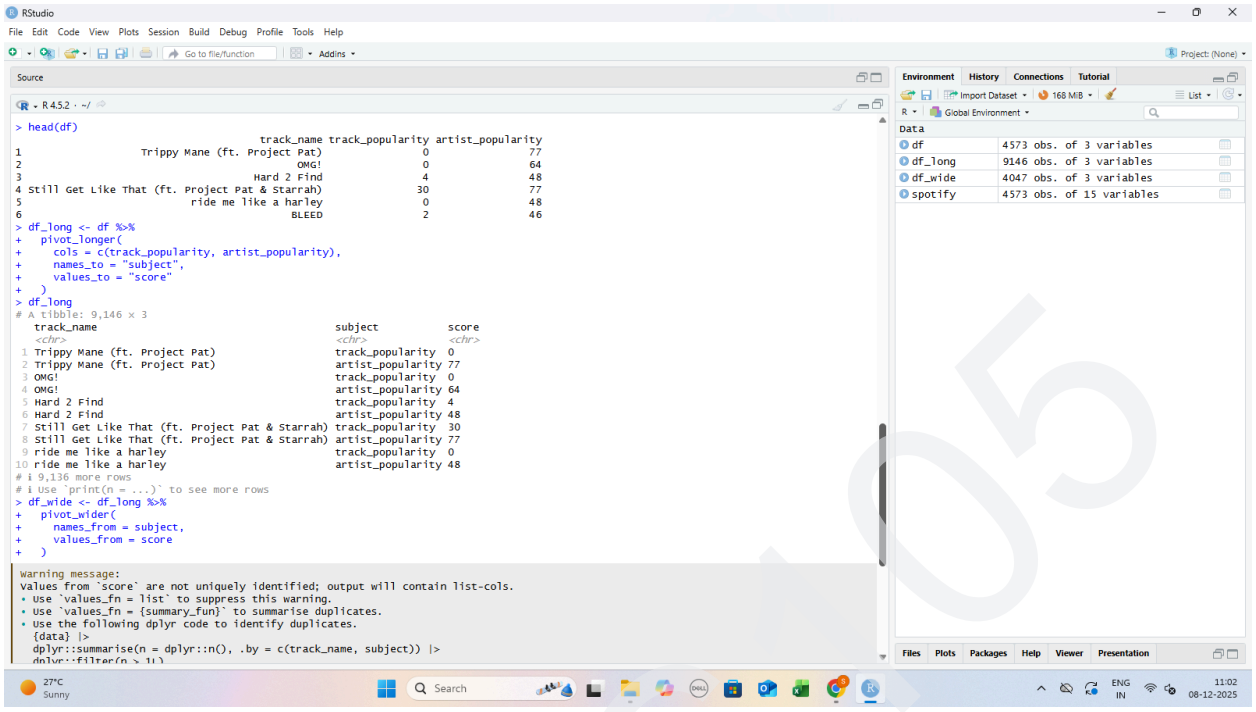
MVLU COLLEGE
DATA ANALYSIS WITH SAS/ SPSS/ R

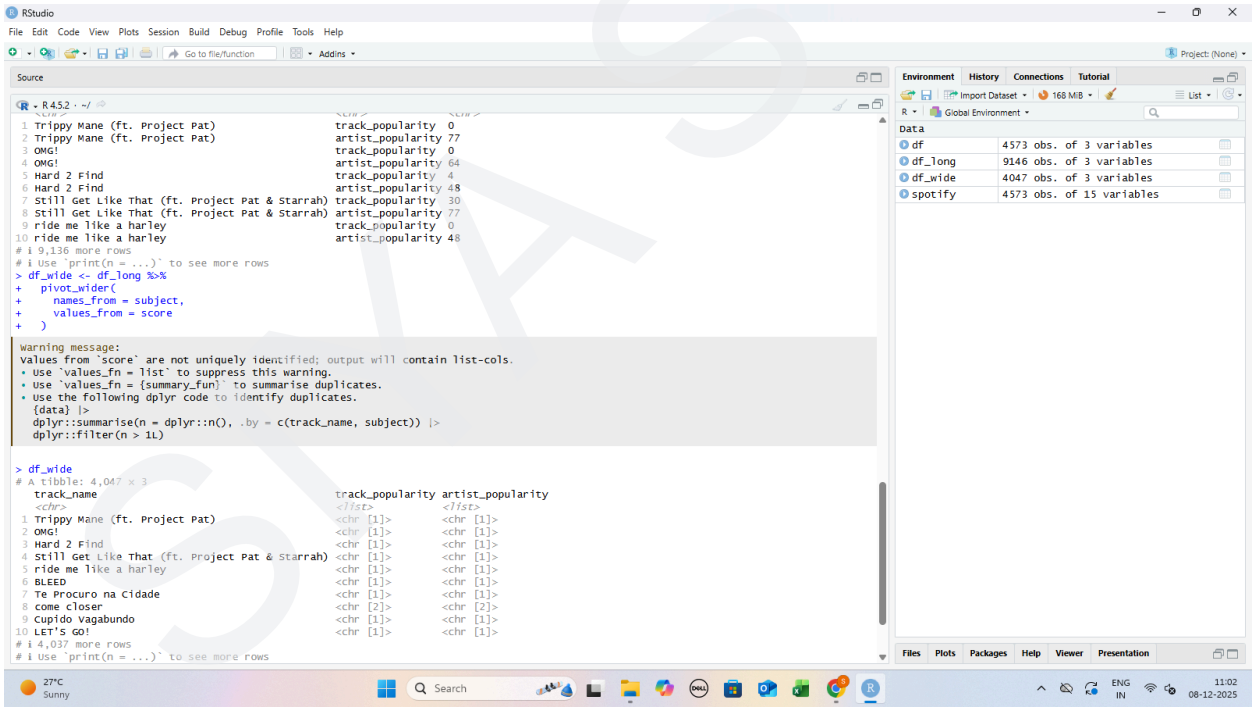
AIM: 11. Reshaping data using pivot_longer() and pivot_wider() (R).





MVLU COLLEGE
DATA ANALYSIS WITH SAS/ SPSS/ R

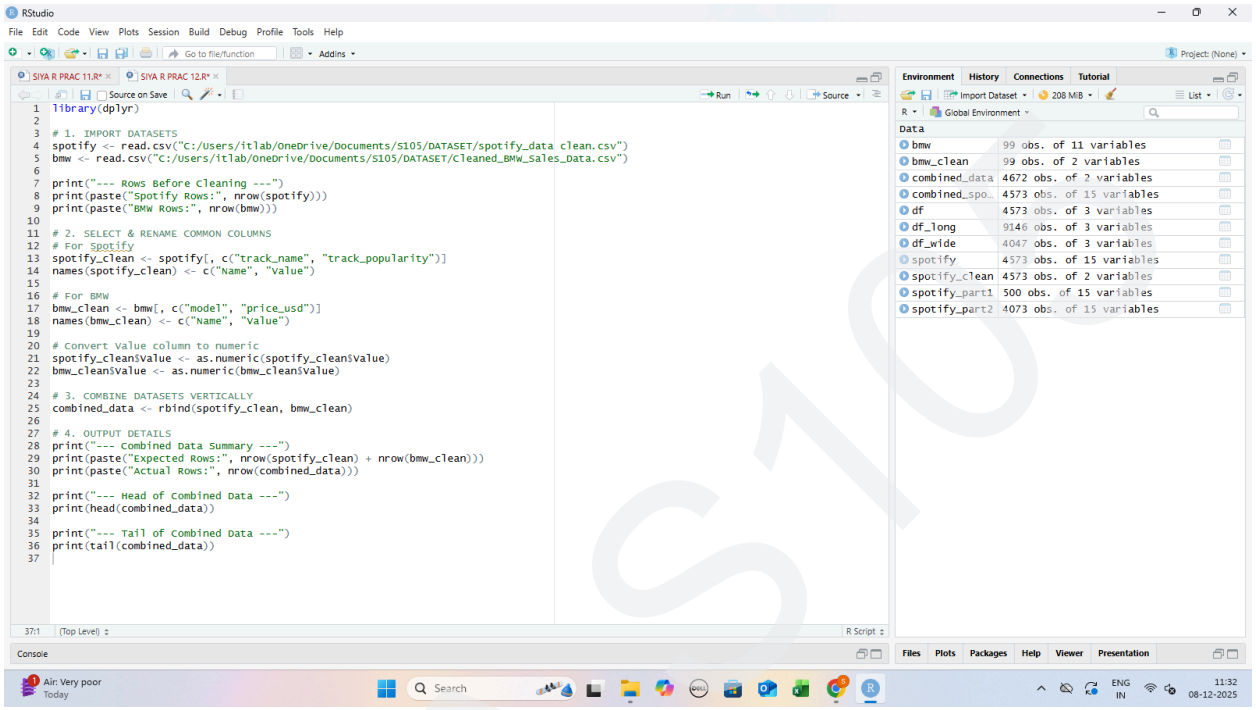




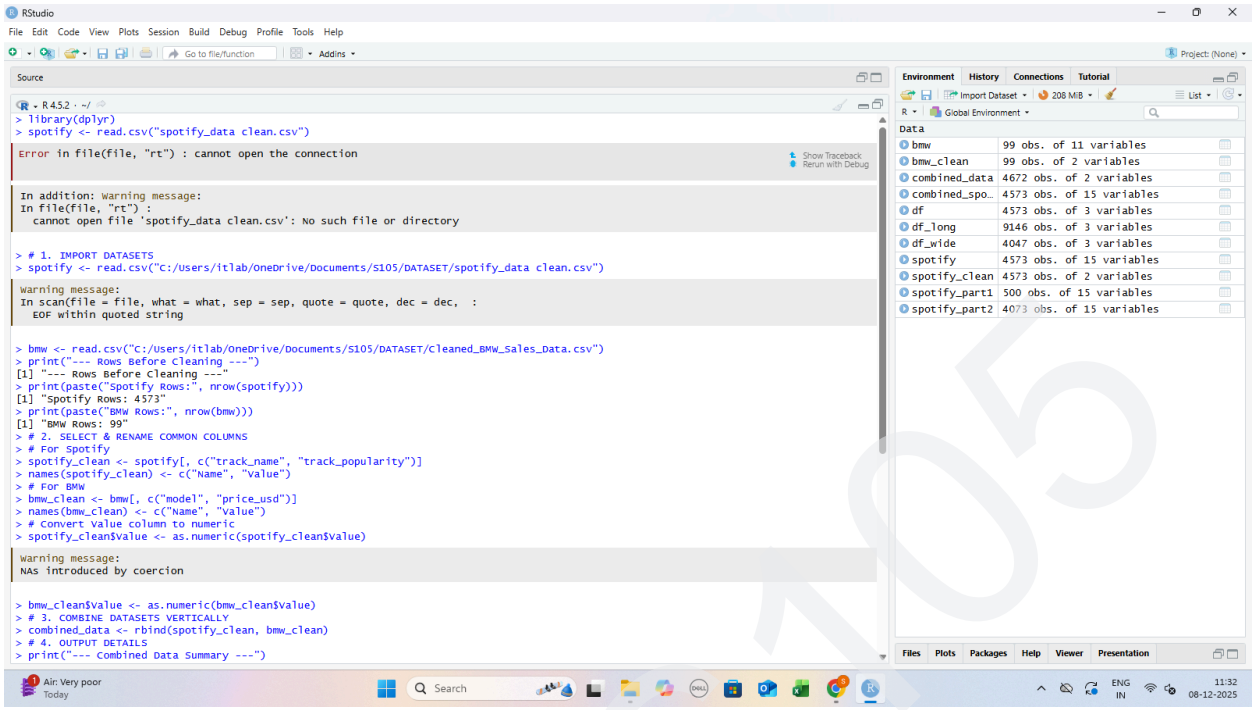
SIYA S105
R PROGRAMMING

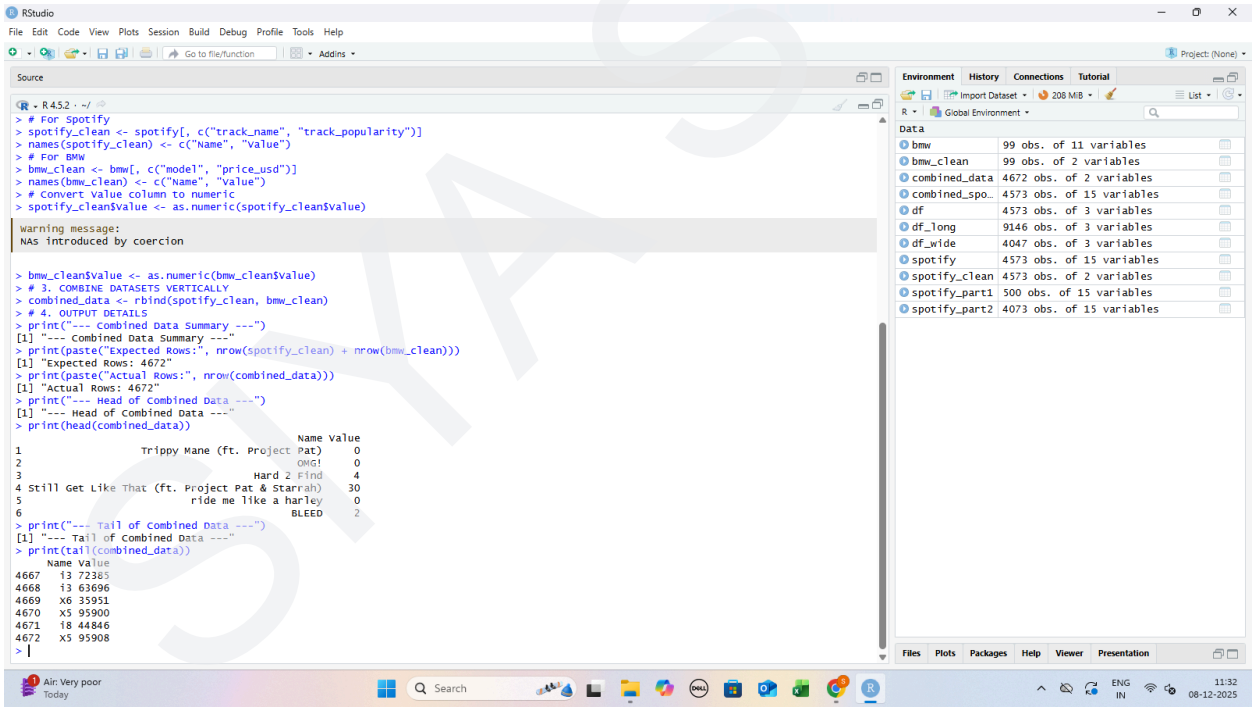
MVLU COLLEGE
DATA ANALYSIS WITH SAS/ SPSS/ R

AIM: 12. Combining datasets vertically (concatenation) using rbind() (R).Write code toCombining datasets vertically (concatenation) using rbind() in R studio.



MVLU COLLEGE
DATA ANALYSIS WITH SAS/ SPSS/ R

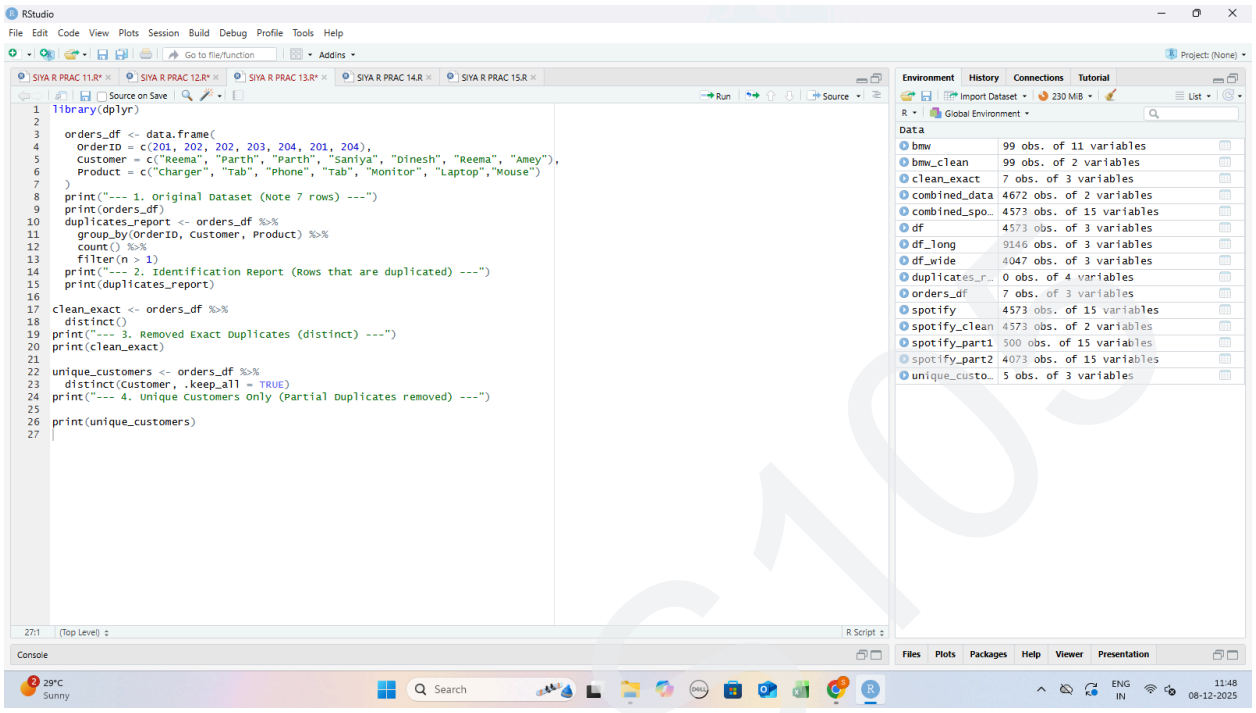




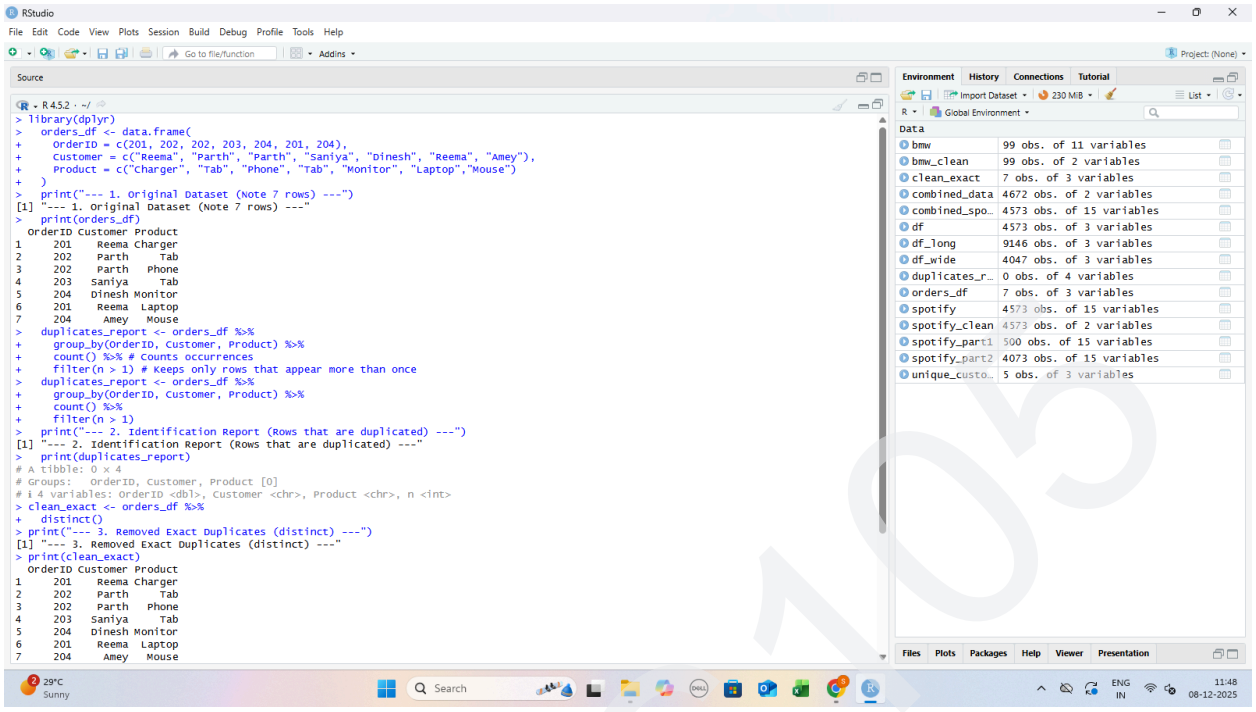
SIYA S105
R PROGRAMMING

MVLU COLLEGE
DATA ANALYSIS WITH SAS/ SPSS/ R

AIM:13 Identifying and handling duplicates using distinct() (R).

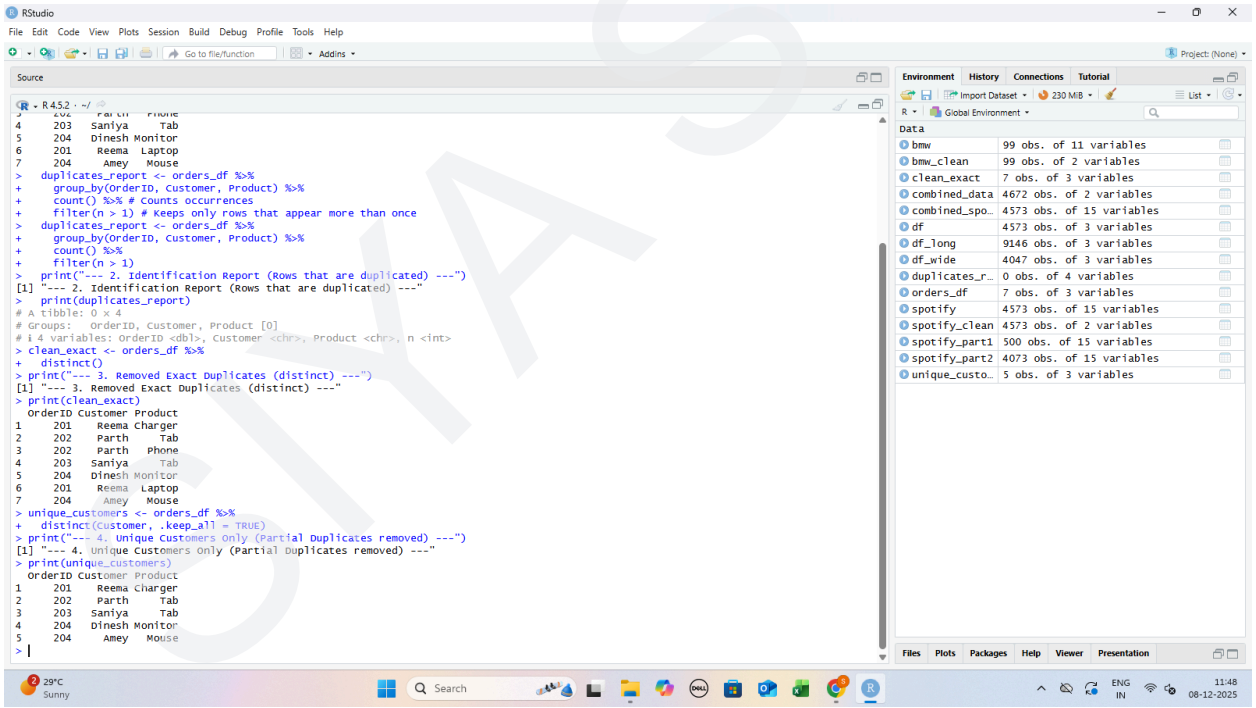


MVLU COLLEGE
DATA ANALYSIS WITH SAS/ SPSS/ R



```
> library(dplyr)
>   orders_df <- data.frame(
+     OrderID = c(201, 202, 202, 203, 204, 201, 204),
+     Customer = c("Reema", "Parth", "Parth", "Saniya", "Dinesh", "Reema", "Amey"),
+     Product = c("Charger", "Tab", "Phone", "Tab", "Monitor", "Laptop","Mouse")
+   )
>   print("--- 1. Original Dataset (Note 7 rows) ---")
[1] "--- 1. Original Dataset (Note 7 rows) ---"
>   print(orders_df)
  OrderID Customer Product
1     201    Reema Charger
2     202    Parth     Tab
3     202    Parth   Phone
4     203   Saniya     Tab
5     204   Dinesh Monitor
6     201    Reema  Laptop
7     204     Amey   Mouse
>   duplicates_report <- orders_df %>%
+     group_by(OrderID, Customer, Product) %>%
+     count() %>% # Counts occurrences
+     filter(n > 1) # Keeps only rows that appear more than once
>   duplicates_report <- orders_df %>%
+     group_by(OrderID, Customer, Product) %>%
+     count() %>%
+     filter(n > 1)
>   print("--- 2. Identification Report (Rows that are duplicated) ---")
[1] "--- 2. Identification Report (Rows that are duplicated) ---"
>   print(duplicates_report)
# A tibble: 0 x 4
# Groups:   OrderID, Customer, Product [0]
# i 4 variables: OrderID <dbl>, Customer <chr>, Product <chr>, n <int>
> clean_exact <- orders_df %>%
+   distinct()
> print("--- 3. Removed Exact Duplicates (distinct) ---")
[1] "--- 3. Removed Exact Duplicates (distinct) ---"
> print(clean_exact)
  OrderID Customer Product
1     201    Reema Charger
2     202    Parth     Tab
3     202    Parth   Phone
4     203   Saniya     Tab
5     204   Dinesh Monitor
6     201    Reema  Laptop
7     204     Amey   Mouse
```

Environment	History	Connections	Tutorial
Data			
bmw	99 obs. of 11 variables		
bmw_clean	99 obs. of 2 variables		
clean_exact	7 obs. of 3 variables		
combined_data	4672 obs. of 2 variables		
combined_spo…	4573 obs. of 15 variables		
df	4573 obs. of 3 variables		
df_long	9146 obs. of 3 variables		
df_wide	4047 obs. of 3 variables		
duplicates_r…	0 obs. of 4 variables		
orders_df	7 obs. of 3 variables		
spotify	4573 obs. of 15 variables		
spotify_clean	4573 obs. of 2 variables		
spotify_part1	500 obs. of 15 variables		
spotify_part2	4073 obs. of 15 variables		
unique_custo…	5 obs. of 3 variables		



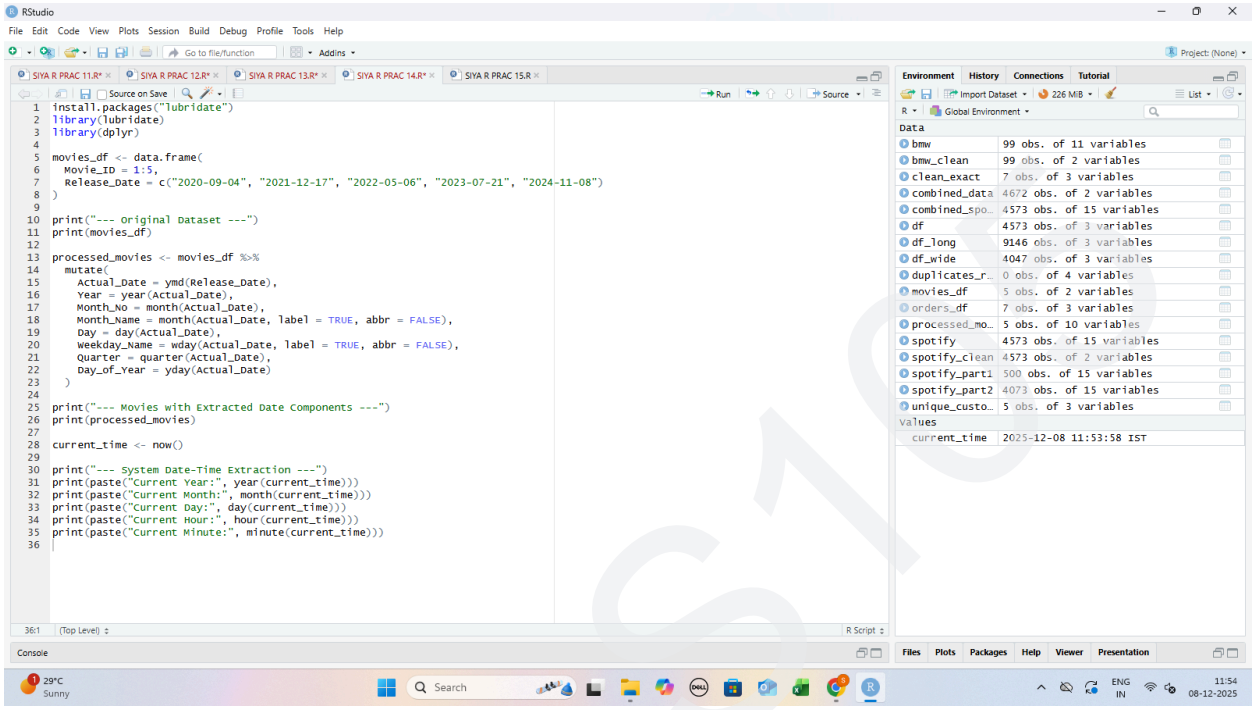
```
4     203   Saniya     Tab
5     204   Dinesh Monitor
6     201    Reema  Laptop
7     204     Amey   Mouse
>   duplicates_report <- orders_df %>%
+     group_by(OrderID, Customer, Product) %>%
+     count() %>% # Counts occurrences
+     filter(n > 1) # Keeps only rows that appear more than once
>   duplicates_report <- orders_df %>%
+     group_by(OrderID, Customer, Product) %>%
+     count() %>%
+     filter(n > 1)
>   print("--- 2. Identification Report (Rows that are duplicated) ---")
[1] "--- 2. Identification Report (Rows that are duplicated) ---"
>   print(duplicates_report)
# A tibble: 0 x 4
# Groups:   OrderID, Customer, Product [0]
# i 4 variables: OrderID <dbl>, Customer <chr>, Product <chr>, n <int>
> clean_exact <- orders_df %>%
+   distinct()
> print("--- 3. Removed Exact Duplicates (distinct) ---")
[1] "--- 3. Removed Exact Duplicates (distinct) ---"
> print(clean_exact)
  OrderID Customer Product
1     201    Reema Charger
2     202    Parth     Tab
3     202    Parth   Phone
4     203   Saniya     Tab
5     204   Dinesh Monitor
6     201    Reema  Laptop
7     204     Amey   Mouse
> unique_customers <- orders_df %>%
+   distinct(Customer, .keep_all = TRUE)
> print("--- 4. Unique Customers Only (Partial Duplicates removed) ---")
[1] "--- 4. Unique Customers Only (Partial Duplicates removed) ---"
> print(unique_customers)
  OrderID Customer Product
1     201    Reema Charger
2     202    Parth     Tab
3     203   Saniya     Tab
4     204   Dinesh Monitor
5     204     Amey   Mouse
>
```

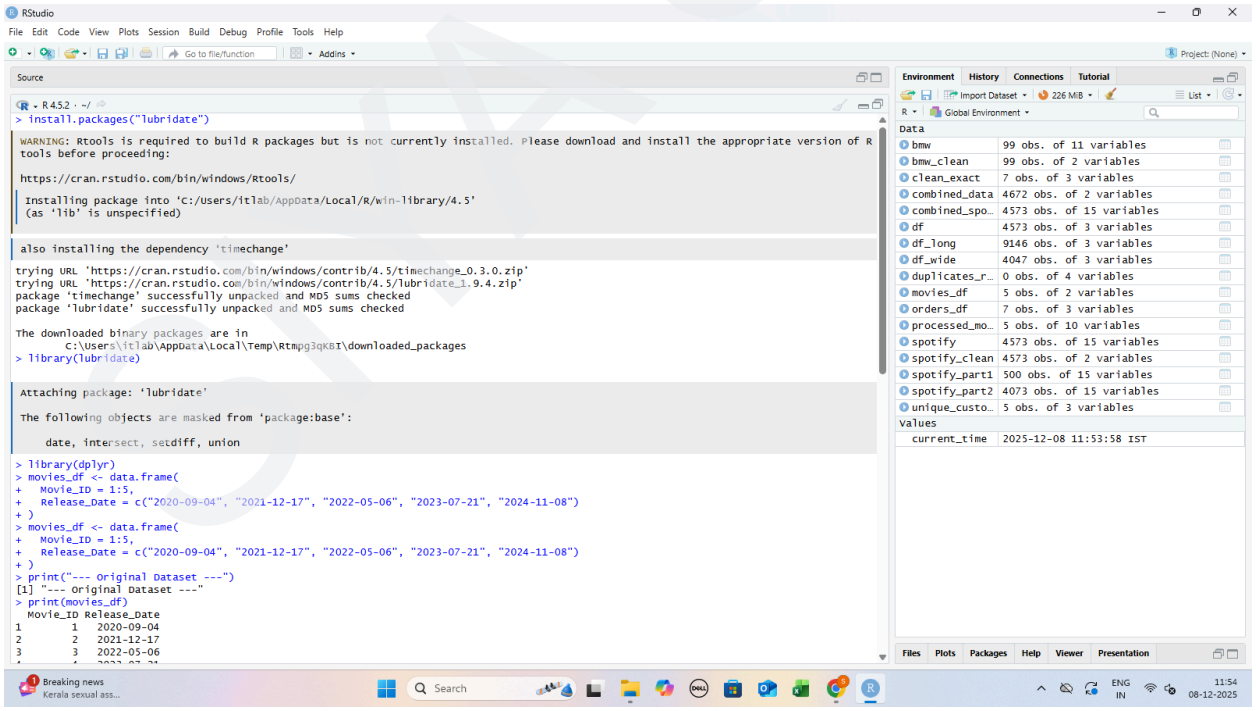
Environment	History	Connections	Tutorial
Data			
bmw	99 obs. of 11 variables		
bmw_clean	99 obs. of 2 variables		
clean_exact	7 obs. of 3 variables		
combined_data	4672 obs. of 2 variables		
combined_spo…	4573 obs. of 15 variables		
df	4573 obs. of 3 variables		
df_long	9146 obs. of 3 variables		
df_wide	4047 obs. of 3 variables		
duplicates_r…	0 obs. of 4 variables		
orders_df	7 obs. of 3 variables		
spotify	4573 obs. of 15 variables		
spotify_clean	4573 obs. of 2 variables		
spotify_part1	500 obs. of 15 variables		
spotify_part2	4073 obs. of 15 variables		
unique_custo…	5 obs. of 3 variables		

SIYA S105
R PROGRAMMING

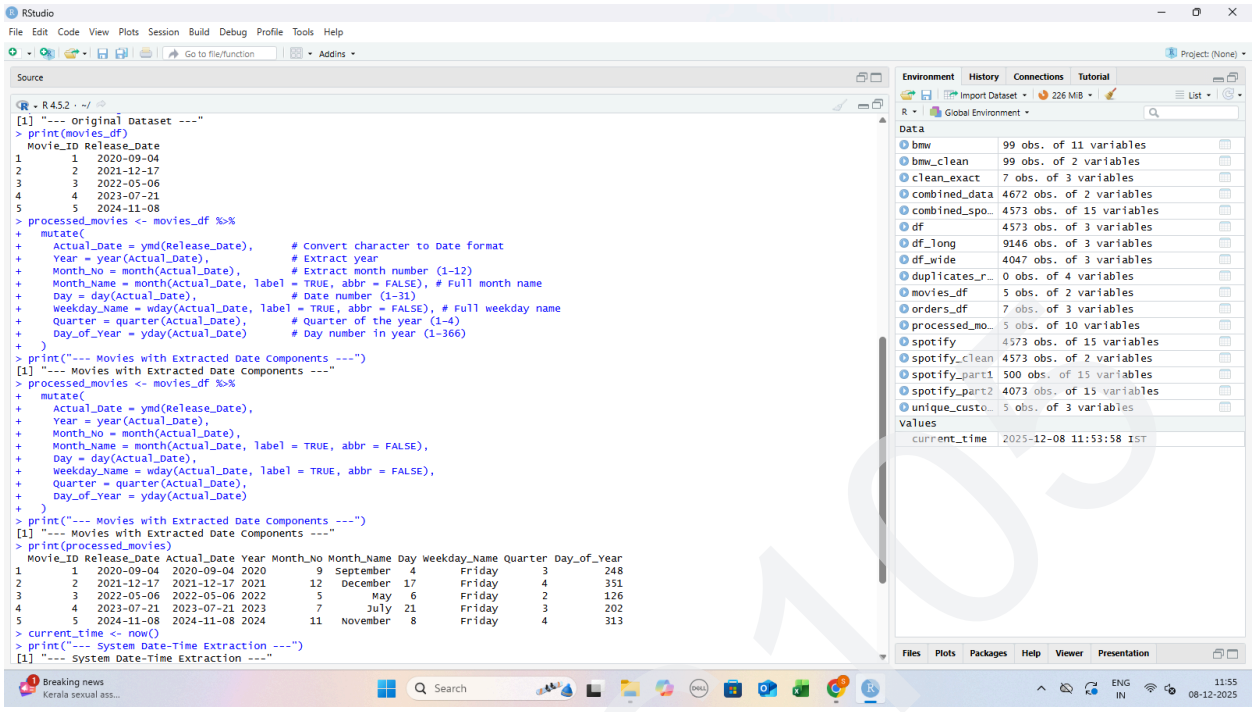
MVLU COLLEGE
DATA ANALYSIS WITH SAS/ SPSS/ R

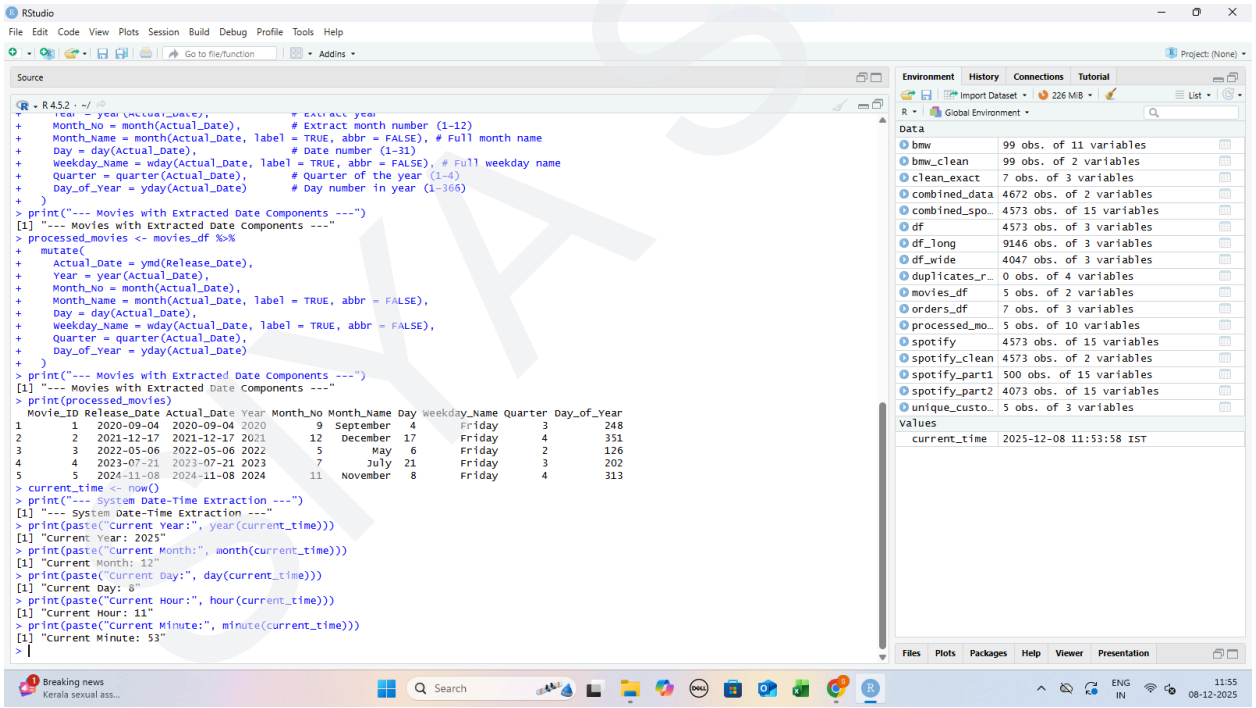
AIM:14 Extracting date components using lubridate:: functions (R).





MVLU COLLEGE
DATA ANALYSIS WITH SAS/ SPSS/ R

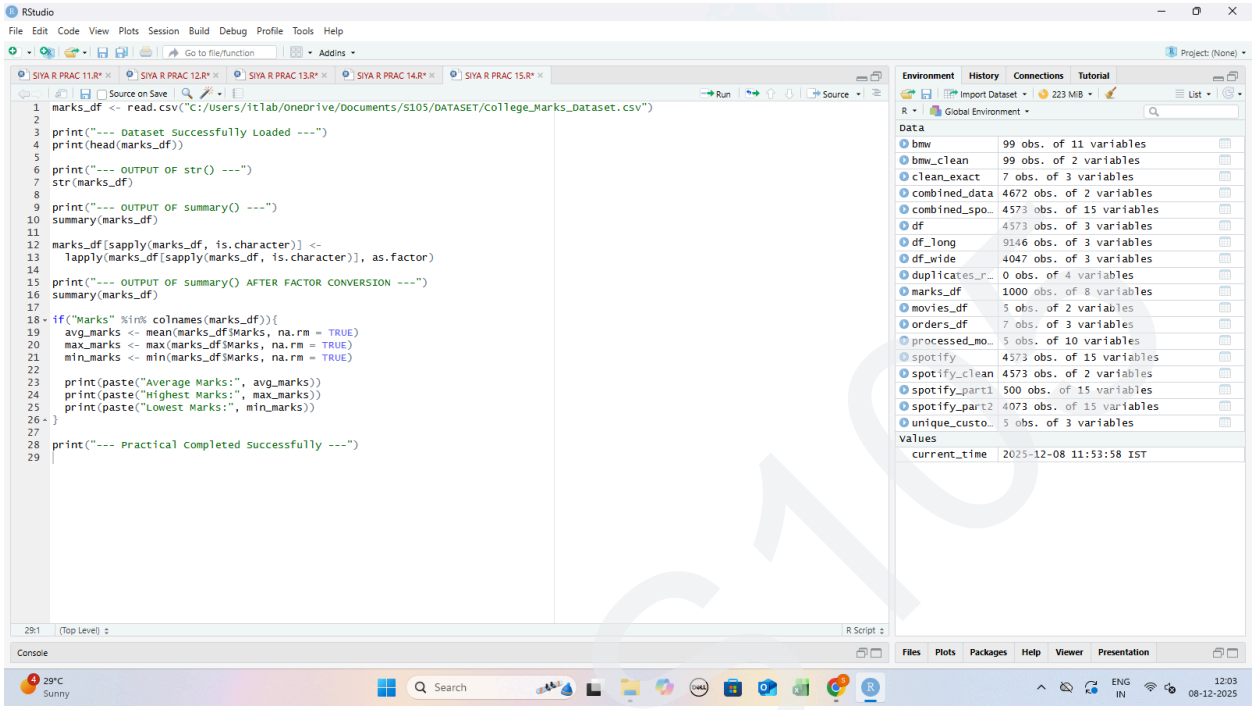




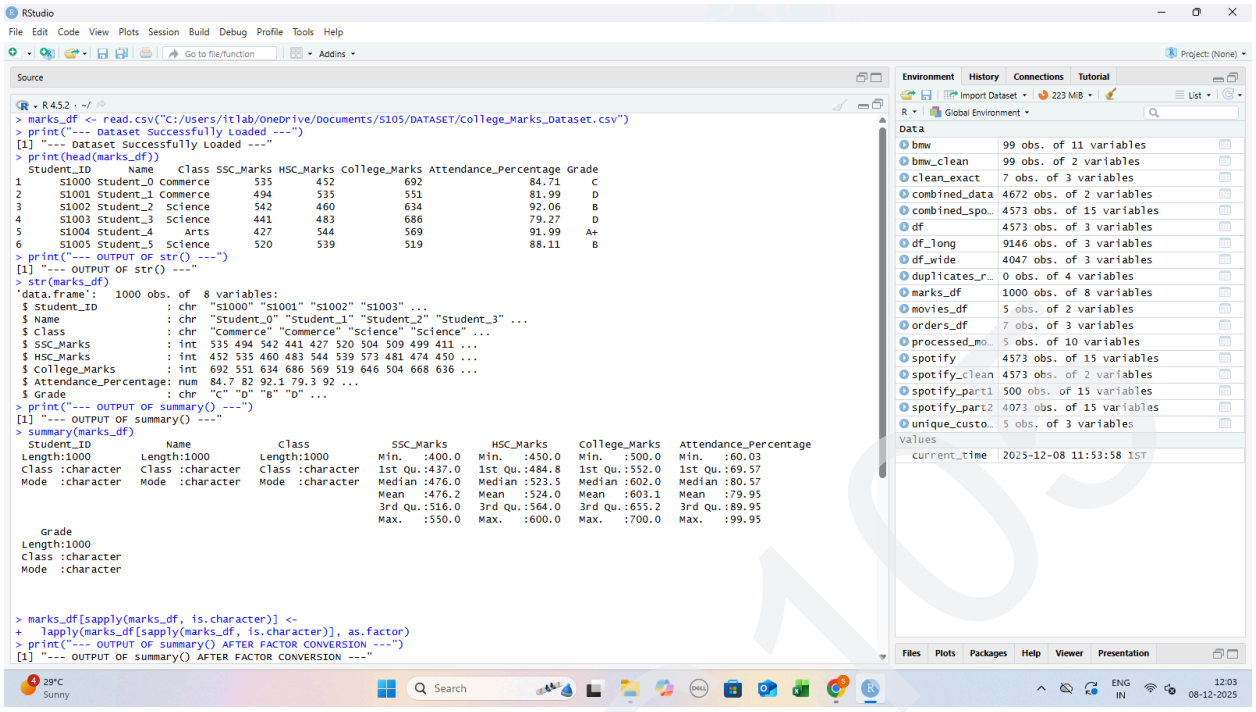
SIYA S105
R PROGRAMMING

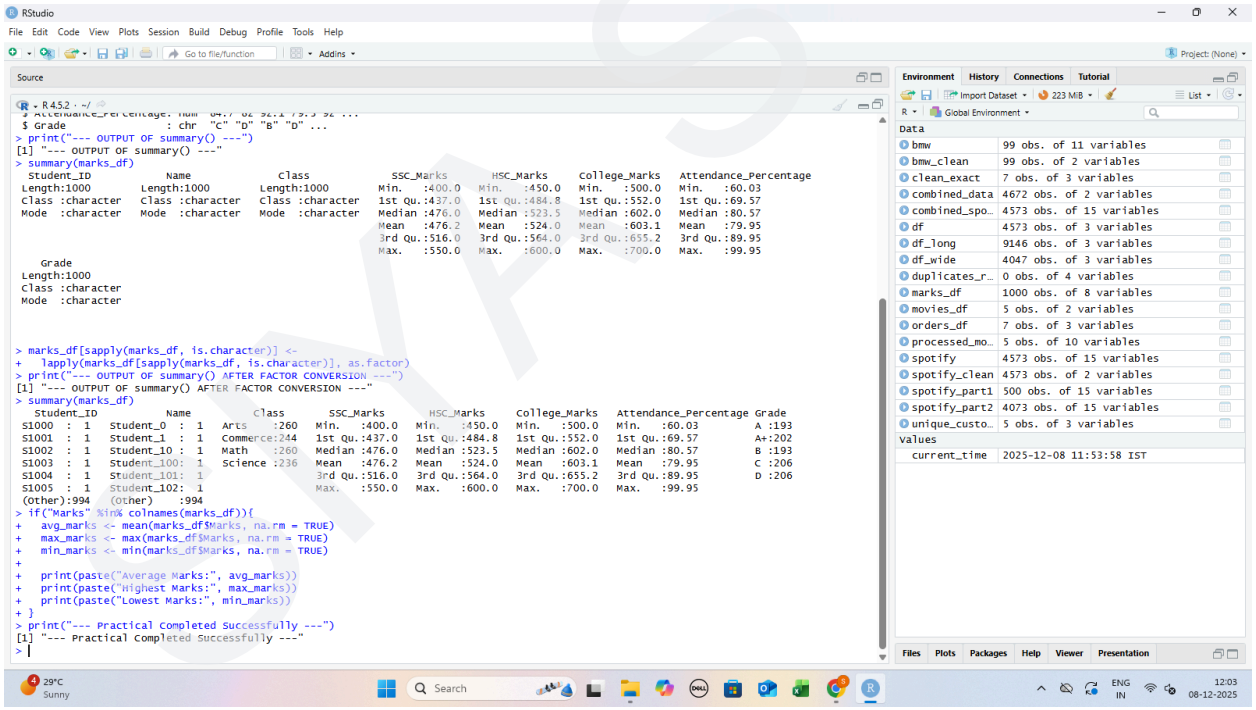
MVLU COLLEGE
DATA ANALYSIS WITH SAS/ SPSS/ R

AIM:15 Generating basic summaries using str() or summary() (R).



MVLU COLLEGE
DATA ANALYSIS WITH SAS/ SPSS/ R





SIYA S105
R PROGRAMMING