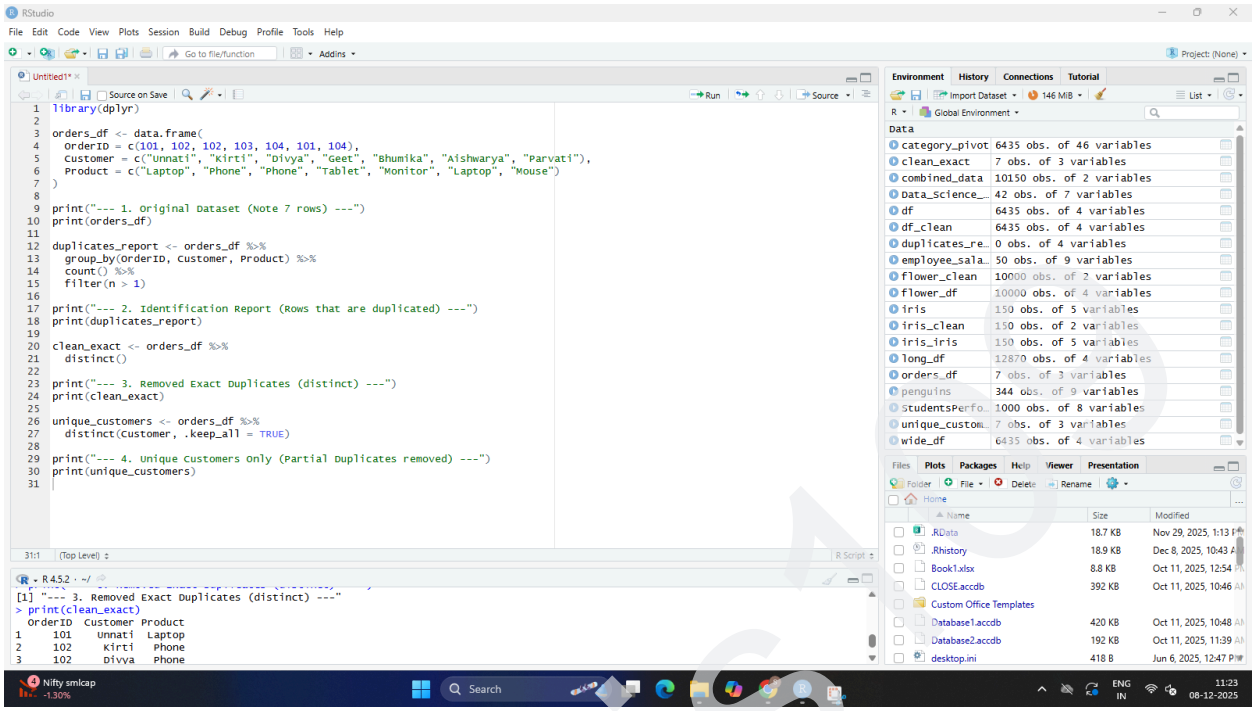
SHETH L.U.J AND M.V COLLEGE
PRACTICAL NO .13
SUBJECT - DATA ANALYSIS

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

```
library(dplyr)

orders_df <- data.frame(
  OrderID = c(101, 102, 102, 103, 104, 101, 104),
  Customer = c("Unnati", "Kirti", "Divya", "Geet", "Bhumika", "Aishwarya", "Parvati"),
  Product = c("Laptop", "Phone", "Phone", "Tablet", "Monitor", "Laptop", "Mouse")
)

print("--- 1. Original Dataset (Note 7 rows) ---")
print(orders_df)

duplicates_report <- orders_df %>%
  group_by(OrderID, Customer, Product) %>%
  count() %>%
  filter(n > 1)

print("--- 2. Identification Report (Rows that are duplicated) ---")
print(duplicates_report)

clean_exact <- orders_df %>%
  distinct()

print("--- 3. Removed Exact Duplicates (distinct) ---")
print(clean_exact)

unique_customers <- orders_df %>%
  distinct(Customer, .keep_all = TRUE)

print("--- 4. Unique Customers Only (Partial Duplicates removed) ---")
print(unique_customers)
```


NAME - UNNATI RATHOD
ROLL NO S109

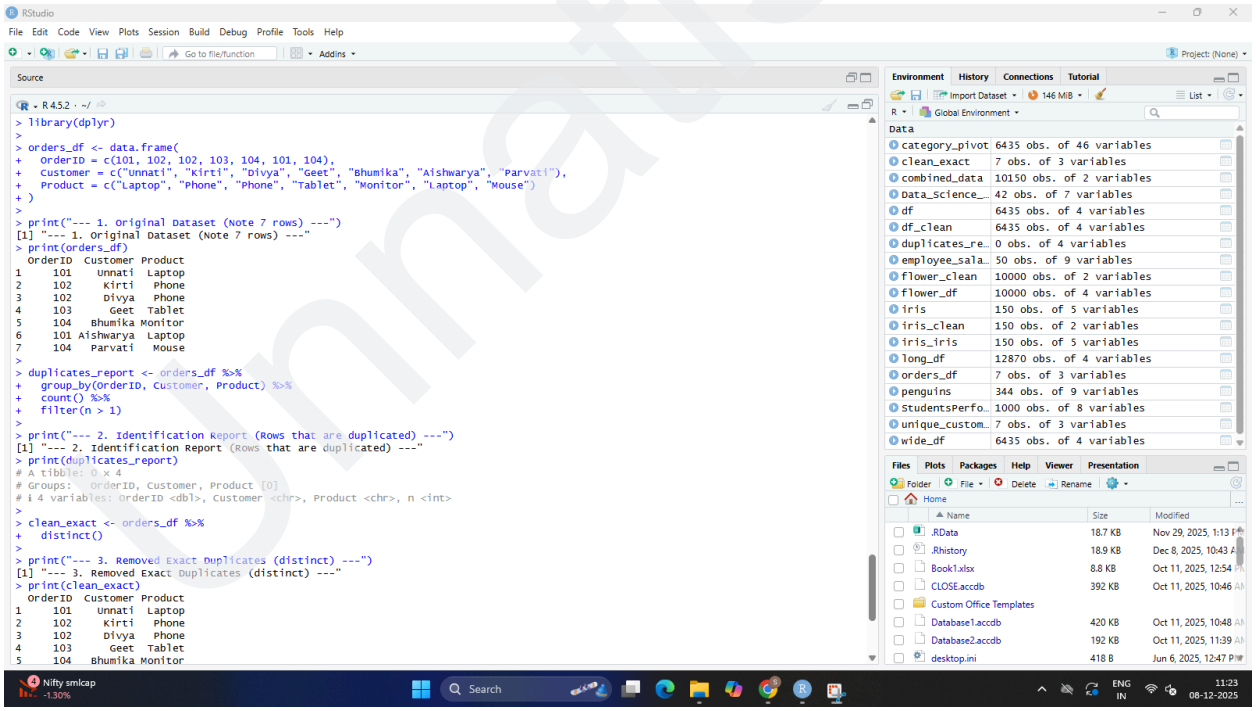
SHETH L.U.J AND M.V COLLEGE
PRACTICAL NO .13
SUBJECT - DATA ANALYSIS



```
library(dplyr)

orders_df <- data.frame(
  OrderID = c(101, 102, 102, 103, 104, 101, 104),
  Customer = c("Unnati", "Kirti", "Divya", "Geet", "Bhumika", "Aishwarya", "Parvati"),
  Product = c("Laptop", "Phone", "Phone", "Tablet", "Monitor", "Laptop", "Mouse")
)

print("--- 1. Original Dataset (Note 7 rows) ---")
print(orders_df)

duplicates_report <- orders_df %>%
  group_by(OrderID, Customer, Product) %>%
  count() %>%
  filter(n > 1)

print("--- 2. Identification Report (Rows that are duplicated) ---")
print(duplicates_report)

clean_exact <- orders_df %>%
  distinct()

print("--- 3. Removed Exact Duplicates (distinct) ---")
print(clean_exact)

unique_customers <- orders_df %>%
  distinct(Customer, .keep_all = TRUE)

print("--- 4. Unique Customers Only (Partial Duplicates removed) ---")
print(unique_customers)
```

```
[1] "--- 3. Removed Exact Duplicates (distinct) ---"
> print(clean_exact)
  OrderID  Customer Product
1     101    Unnati  Laptop
2     102     Kirti   Phone
3     102     Divya   Phone
```

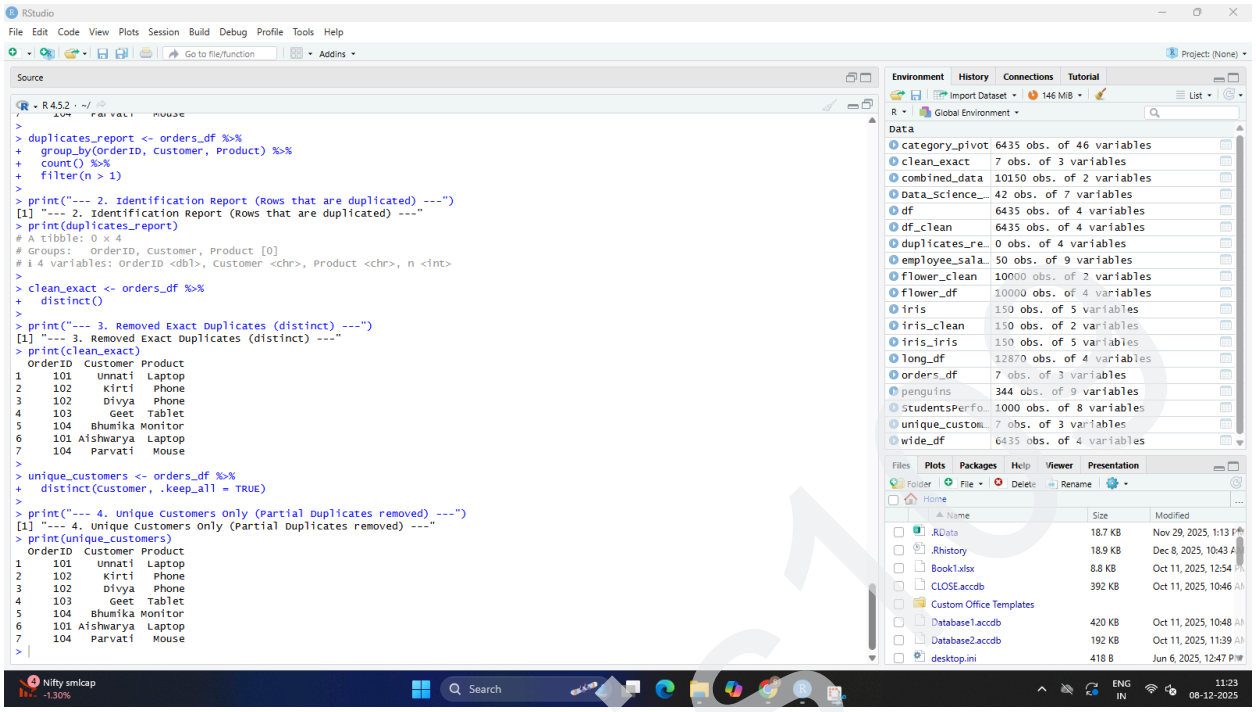
OUTPUT -



```
> library(dplyr)
>
> orders_df <- data.frame(
+   OrderID = c(101, 102, 102, 103, 104, 101, 104),
+   Customer = c("Unnati", "Kirti", "Divya", "Geet", "Bhumika", "Aishwarya", "Parvati"),
+   Product = c("Laptop", "Phone", "Phone", "Tablet", "Monitor", "Laptop", "Mouse")
+ )
>
> print("--- 1. Original Dataset (Note 7 rows) ---")
[1] "--- 1. Original Dataset (Note 7 rows) ---"
> print(orders_df)
  OrderID  Customer Product
1     101    Unnati  Laptop
2     102     Kirti   Phone
3     102     Divya   Phone
4     103      Geet  Tablet
5     104   Bhumika Monitor
6     101 Aishwarya  Laptop
7     104   Parvati   Mouse
>
> 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     101    Unnati  Laptop
2     102     Kirti   Phone
3     102     Divya   Phone
4     103      Geet  Tablet
5     104   Bhumika Monitor
```

NAME - UNNATI RATHOD
ROLL NO S109

SHETH L.U.J AND M.V COLLEGE
PRACTICAL NO .13
SUBJECT - DATA ANALYSIS



```
> 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 × 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     101    Unnati  Laptop
2     102     Kirti   Phone
3     102     Divya   Phone
4     103      Geet  Tablet
5     104   Bhumika Monitor
6     101 Aishwarya  Laptop
7     104   Parvati   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     101    Unnati  Laptop
2     102     Kirti   Phone
3     102     Divya   Phone
4     103      Geet  Tablet
5     104   Bhumika Monitor
6     101 Aishwarya  Laptop
7     104   Parvati   Mouse
>
```

NAME - UNNATI RATHOD
ROLL NO S109