

PRACTICAL NO 5

AIM :Sorting data using arrange() in R.

Output:

```

R - R 4.5.2 - C:/Users/PRIVANKA/Downloads/
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
> library(dplyr)
> library(readr)
> fraud_data <- read_csv("synthetic_fraud_dataset.csv")
Rows: 10000 Columns: 10
Column specification
Delimiter: ','
chr (3): transaction_type, merchant_category, country
dbl (7): transaction_id, user_id, amount, hour, device_risk_score, ip_risk_score, is_fraud
i Use 'spec()' to retrieve the full column specification for this data.
i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
>
> cat("---- Initial Data Load ----\n")
---- Initial Data Load ----
> print(head(fraud_data))
# A tibble: 6 x 10
  transaction_id user_id amount transaction_type merchant_category country hour device_risk_score ip_risk_score is_fraud
  <dbl> <dbl> <dbl> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl>
1 9608 363 4923. ATM Travel TR 12 0.992 0.948 1
2 456 692 48.0 QR Food US 21 0.169 0.224 0
3 4747 587 137. online Travel TR 14 0.296 0.125 0
4 6934 445 80.5 POS Clothing TR 23 0.125 0.159 0
5 1646 729 120. online Grocery FR 16 0.0981 0.0275 0
6 2183 944 97.1 POS Clothing DE 17 0.235 0.105 0
> cat("\n")
>
> fraud_sorted_amount <- fraud_data |>
+ arrange(amount)
> cat("---- 1. Top 5 Lowest Transaction Amounts ----\n")
---- 1. Top 5 Lowest Transaction Amounts ----
> print(head(fraud_sorted_amount, 5))
# A tibble: 5 x 10
  transaction_id user_id amount transaction_type merchant_category country hour device_risk_score ip_risk_score is_fraud
  <dbl> <dbl> <dbl> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl>
1 3094 200 1 POS Food US 9 0.107 0.180 0
2 6904 310 1 POS Food DE 14 0.111 0.203 0
3 6087 216 1 online Grocery DE 7 0.236 0.0198 0
4 3199 846 1 ATM Electronics US 21 0.0503 0.0781 0
5 4538 330 1 ATM Clothing TO 10 0.215 0.00508 0

```

SHETH L.U.J. AND SIR M.V. COLLEGE OF ARTS SCIENCE AND COMMERCE

Subject : Data Analysis with SAS / SPSS /R

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help

Source

Console Terminal Background Jobs
R - R 4.5.2 - C:/Users/PRIVANKA/Downloads/
5 4538 339 1 ATM Clothing TR 10 0.215 0.00598 0
> fraud_sorted_amount_desc <- fraud_data |>
+ arrange(desc(amount))
> cat("\n--- 2. Top 5 Highest Transaction Amounts ---\n")
--- 2. Top 5 Highest Transaction Amounts ---
> print(head(fraud_sorted_amount_desc, 5))
# A tibble: 5 x 10
  transaction_id user_id amount transaction_type merchant_category country hour device_risk_score ip_risk_score is_fraud
1 9747 423 11628. POS Food TR 23 0.951 0.942 1
2 9956 589 11085. ATM Food DE 7 0.868 0.804 1
3 9657 590 10430. ATM Travel DE 8 0.976 0.727 1
4 8593 201 9648. POS Travel NG 23 0.733 0.778 1
5 9538 144 8894. POS Grocery FR 7 0.868 0.932 1
> fraud_sorted_category_amount <- fraud_data |>
+ arrange(merchant_category, desc(amount))
> cat("\n--- 3. Top 10 Transactions: Sorted by Category, then by highest Amount ---\n")
--- 3. Top 10 Transactions: Sorted by Category, then by highest Amount ---
> print(head(fraud_sorted_category_amount, 10))
# A tibble: 10 x 10
  transaction_id user_id amount transaction_type merchant_category country hour device_risk_score ip_risk_score is_fraud
1 9518 810 2504. POS Clothing TR 12 0.930 0.934 1
2 9997 574 6428. Online Clothing NG 15 0.760 0.973 1
3 9577 797 4765. ATM Clothing US 21 0.769 0.752 1
4 9730 926 4695. ATM Clothing DE 4 0.744 0.735 1
5 9613 874 4340. POS Clothing DE 4 0.846 0.803 1
6 9890 109 3981. QR Clothing NG 16 0.750 0.904 1
7 9753 188 3578. POS Clothing UK 3 0.858 0.901 1
8 9878 567 3541. Online Clothing TR 17 0.724 0.787 1
9 9860 55 3465. ATM Clothing NG 4 0.793 0.857 1
10 9566 461 3445. ATM Clothing FR 4 0.782 0.928 1
> high_risk_sorted_country <- fraud_data |>
+ filter(device_risk_score > 0.95) |>
+ arrange(country)
> cat("\n--- 4. Transactions with Device Risk > 0.95, sorted by Country ---\n")
--- 4. Transactions with Device Risk > 0.95, sorted by Country ---
> print(head(high_risk_sorted_country))
# A tibble: 6 x 10
  transaction_id user_id amount transaction_type merchant_category country hour device_risk_score ip_risk_score is_fraud
1 9888 278 1853. Online Clothing DE 11 0.974 0.727 1
2 9832 86 2356. ATM Grocery DE 4 0.971 0.762 1
3 9800 595 1783. QR Travel DE 12 0.992 0.972 1
4 9674 739 22.2 POS Food DE 6 0.987 0.971 1
5 9821 743 22.2 ATM Travel DE 6 0.983 0.946 1
6 9558 114 22.2 ATM Travel DE 6 0.957 0.887 1
> low_ip_risk_sorted <- fraud_data |>
+ arrange(ip_risk_score)
> cat("\n--- 5. Top 5 Lowest IP Risk Scores ---\n")
--- 5. Top 5 Lowest IP Risk Scores ---
> print(head(low_ip_risk_sorted, 5))
# A tibble: 5 x 10
  transaction_id user_id amount transaction_type merchant_category country hour device_risk_score ip_risk_score is_fraud
1 952 312 104. ATM Grocery US 6 0.279 0.0000880 0
2 1709 827 103. QR Clothing DE 14 0.271 0.0000289 0
3 5133 970 156. Online Electronics US 15 0.143 0.0000536 0
4 6429 926 53.0 ATM Food TR 15 0.0738 0.0000667 0
5 522 720 115. POS Electronics UK 22 0.238 0.0000834 0
```

NAME : SHUBHAM SANJAY KARAPE
ROLL NO : S085