

# **SHETH L.U.J. AND SIR M.V. COLLEGE**

## **SUBJECT:Data Analysis with R**

Aim: Identifying and handling duplicates using distinct() (R studio )

# SHETH L.U.J. AND SIR M.V. COLLEGE

## SUBJECT:Data Analysis with R

```

<-- unique_companies <- cacao_df %>
+ distinct(company_maker_if_known, .keep_all = TRUE)
> print("---- 4. Unique Companies only (Partial Duplicates Removed) ----")
[1] "---- 4. Unique Companies only (Partial Duplicates Removed) ----"
> print(head(unique_companies, 10))
  company_maker_if_known specific_beans_origin_or_bar_name ref review_date cocoa_percent company_location rating
1 A. Morin Agua Grande 1876 2016 63% France 3.75
2 A. Morin Acallí chulucanas, El Platanal 1462 2015 70% U.S.A. 3.75
3 A. Morin Acají Varietal 155 2011 60% France 2.75
4 Aequare (Gianduia) Los Ríos, Quevedo, Tabasco 370 2009 55% Ecuador 2.75
5 Ah Cacao 316 2009 70% Mexico 3.00
6 Akesson's (Pralus) Bali (west), Sukrama Family, Melaya area 636 2011 75% Switzerland 3.75
7 Alain bucasse Trinité 1215 2014 65% France 2.75
8 Alexandre Winak Coop, Napo 1944 2017 70% Netherlands 3.50
9 Altus aka Cao Artisan Momotombo 1728 2016 60% U.S.A. 2.75
10 Amano Morobe 725 2011 70% U.S.A. 4.00
  bean_type broad_beans_origin
1 Sao Tome
2 Peru
3 Trinitario Fiji
4 Forastero (Arriba) Ecuador
5 Criollo Mexico
6 Trinitario Indonesia
7 Trinitario Trinidad
8 Forastero (Nacional) Ecuador
9 Nicaragua
10 Papua New Guinea
> # scenario: keep unique chocolate bars based on Company + Specific Bean origin
> unique_bars <- cacao_df %>
+ distinct(company_maker_if_known, specific_beans_origin_or_bar_name, .keep_all = TRUE)
> print("---- 5. Unique Bars (Company + Specific Bean Origin) ----")
[1] "---- 5. Unique Bars (Company + Specific Bean Origin) ----"
> print(head(unique_bars, 10))
  company_maker_if_known specific_beans_origin_or_bar_name ref review_date cocoa_percent company_location rating bean_type
1 A. Morin Agua Grande 1876 2016 63% France 3.75
2 A. Morin Kpime 1676 2015 70% France 2.75
3 A. Morin Atsane 1676 2015 70% France 3.00
4 A. Morin Akata 1680 2015 70% France 3.50
5 A. Morin Quilla 1704 2015 70% France 3.50
6 A. Morin Carenero 1315 2014 70% France 2.75 criollo
7 A. Morin Cuba 1315 2014 70% France 3.50
8 A. Morin Sur del Lago 1315 2014 70% France 3.50 criollo
9 A. Morin Puerto Cabello 1319 2014 70% France 3.75 criollo
10 A. Morin Pabilio 1319 2014 70% France 4.00
  broad_beans_origin
1 Sao Tome
2 Togo
3 Togo
4 Togo
5 Peru
6 Venezuela
7 Cuba
8 Venezuela
9 Venezuela
10 Peru
> |
```

Breaking news  
Kerala sexual ass...

File Plot Packages Help Viewer Presentation

My Pictures My Videos My Web Sites NetBeansProjects Power BI Desktop scala for DS Sound Recordings T050\_AngaliT vgsales.csv Virtual Machines Visual Studio 18 Visual Studio 2022 WindowsPowerShell Student.csv car\_price\_prediction\_csv disease\_diagnosis.csv flavors\_of\_cacao.csv

1.3 MB Nov 163 KB Dec 156.3 KB Nov 189.1 KB Dec 126.5 KB Nov

ENG IN 11:48 08-12-2025

  

```

<-- unique_companies <- cacao_df %>
+ distinct(company_maker_if_known, specific_beans_origin_or_bar_name, .keep_all = TRUE)
> print("---- 5. Unique Bars (Company + Specific Bean Origin) ----")
[1] "---- 5. Unique Bars (Company + Specific Bean Origin) ----"
> print(head(unique_bars, 10))
  company_maker_if_known specific_beans_origin_or_bar_name ref review_date cocoa_percent company_location rating bean_type
1 A. Morin Agua Grande 1876 2016 63% France 3.75
2 A. Morin Kpime 1676 2015 70% France 2.75
3 A. Morin Atsane 1676 2015 70% France 3.00
4 A. Morin Akata 1680 2015 70% France 3.50
5 A. Morin Quilla 1704 2015 70% France 3.50
6 A. Morin Carenero 1315 2014 70% France 2.75 criollo
7 A. Morin Cuba 1315 2014 70% France 3.50
8 A. Morin Sur del Lago 1315 2014 70% France 3.50 criollo
9 A. Morin Puerto Cabello 1319 2014 70% France 3.75 criollo
10 A. Morin Pabilio 1319 2014 70% France 4.00
  broad_beans_origin
1 Sao Tome
2 Togo
3 Togo
4 Togo
5 Peru
6 Venezuela
7 Cuba
8 Venezuela
9 Venezuela
10 Peru
> |
```

Breaking news  
Kerala sexual ass...

File Plot Packages Help Viewer Presentation

My Pictures My Videos My Web Sites NetBeansProjects Power BI Desktop scala for DS Sound Recordings T050\_AngaliT vgsales.csv Virtual Machines Visual Studio 18 Visual Studio 2022 WindowsPowerShell Student.csv car\_price\_prediction\_csv disease\_diagnosis.csv flavors\_of\_cacao.csv

1.3 MB Nov 163 KB Dec 156.3 KB Nov 189.1 KB Dec 126.5 KB Nov

ENG IN 11:48 08-12-2025

  

```

<-- unique_companies <- cacao_df %>
+ distinct(company_maker_if_known, specific_beans_origin_or_bar_name, .keep_all = TRUE)
> print("---- 5. Unique Bars (Company + Specific Bean Origin) ----")
[1] "---- 5. Unique Bars (Company + Specific Bean Origin) ----"
> print(head(unique_bars, 10))
  company_maker_if_known specific_beans_origin_or_bar_name ref review_date cocoa_percent company_location rating bean_type
1 A. Morin Agua Grande 1876 2016 63% France 3.75
2 A. Morin Kpime 1676 2015 70% France 2.75
3 A. Morin Atsane 1676 2015 70% France 3.00
4 A. Morin Akata 1680 2015 70% France 3.50
5 A. Morin Quilla 1704 2015 70% France 3.50
6 A. Morin Carenero 1315 2014 70% France 2.75 criollo
7 A. Morin Cuba 1315 2014 70% France 3.50
8 A. Morin Sur del Lago 1315 2014 70% France 3.50 criollo
9 A. Morin Puerto Cabello 1319 2014 70% France 3.75 criollo
10 A. Morin Pabilio 1319 2014 70% France 4.00
  broad_beans_origin
1 Sao Tome
2 Togo
3 Togo
4 Togo
5 Peru
6 Venezuela
7 Cuba
8 Venezuela
9 Venezuela
10 Peru
> |
```

Console Terminal Background Jobs

R + R 4.52 - ~ /

8 Venezuela
9 Venezuela
10 Peru
> View(unique\_companies)
> View(cacao\_df)

BAJFINANCE -1.85%

File Plot Packages Help Viewer Presentation

My Pictures My Videos My Web Sites NetBeansProjects Power BI Desktop scala for DS Sound Recordings T050\_AngaliT vgsales.csv Virtual Machines Visual Studio 18 Visual Studio 2022 WindowsPowerShell Student.csv car\_price\_prediction\_csv disease\_diagnosis.csv flavors\_of\_cacao.csv

1.3 MB Nov 163 KB Dec 156.3 KB Nov 189.1 KB Dec 126.5 KB Nov

ENG IN 11:48 08-12-2025

# SHETH L.U.J. AND SIR M.V. COLLEGE

## SUBJECT:Data Analysis with R

The image displays two side-by-side screenshots of the RStudio interface, showing the process of data analysis for two different datasets.

**Session 1 (Top Left):**

- Data View:** Shows a data frame named `cacao_df` with 1795 observations and 9 variables. The columns include `company_maker_if_known`, `specific_beans_origin_or_bar_name`, `ref`, `review_date`, `cocoa_percent`, `company_location`, `rating`, `bean_type`, and `broad_beans_origin`.
- Console View:** Displays R code and its output. The user has run commands like `View(unique_companies)`, `View(cacao_df)`, and `View(clean_exact)`. The output shows the names of companies and their locations.
- Environment View:** Shows the global environment with objects like `cacao_df`, `clean_exact`, etc.
- File Explorer:** Shows files such as `vgsales.csv`, `Virtual Machines`, and various Microsoft Office documents.
- System Status Bar:** Shows the system language as ENG IN, the date as 08-12-2025, and the time as 11:48.

**Session 2 (Bottom Left):**

- Data View:** Shows a data frame named `cacao_df` with 1677 observations and 9 variables, similar to Session 1.
- Console View:** Displays R code and its output, showing the same commands and company names as Session 1.
- Environment View:** Shows the global environment with objects like `cacao_df`, `clean_exact`, etc.
- File Explorer:** Shows files such as `vgsales.csv`, `Virtual Machines`, and various Microsoft Office documents.
- System Status Bar:** Shows the system language as ENG IN, the date as 08-12-2025, and the time as 11:49.

# SHETH L.U.J. AND SIR M.V. COLLEGE

## SUBJECT:Data Analysis with R

The screenshot shows a Windows desktop environment. In the foreground, the RStudio interface is open, displaying a data frame titled "cacao\_df" with 1795 observations and 9 variables. The columns include company\_maker\_if\_known, specific\_bean\_origin\_or\_bar\_name, ref, review\_date, cocoa\_percent, company\_location, rating, bean\_type, and broad\_bean\_origin. The data frame contains entries for various cacao companies like A. Morin, Chulucanas, Vanu Levu, etc., with their respective details. Below the RStudio window, the taskbar shows several pinned icons including Microsoft Edge, File Explorer, and the Start button.

company_maker_if_known	specific_bean_origin_or_bar_name	ref	review_date	cocoa_percent	company_location	rating	bean_type	broad_bean_origin
1 A. Morin	Aguia Grande	1876	2016	63%	France	3.75		Sao Tome
2 Acalí	Chulucanas, El Platanal	1462	2015	70%	U.S.A.	3.75		Peru
3 Adil	Vanua Levu	705	2011	60%	Fiji	2.75	Trinitario	Fiji
4 Aequate (Ganduja)	Los Rios, Quevedo, Arriba	370	2009	55%	Ecuador	2.75	Forastero (Arriba)	Ecuador
5 Ah Cacao	Tabasco	316	2009	70%	Mexico	3.00	Criollo	Mexico
6 Akesson's (Palus)	Bali (west), Sukrama Family, Meaya area	636	2011	75%	Switzerland	3.75	Trinitario	Indonesia
7 Alain Ducasse	Trinite	1215	2014	65%	France	2.75	Trinitario	Trinidad
8 Alexandre	Winak Coop, Napo	1944	2017	70%	Netherlands	3.50	Forastero (Nacional)	Ecuador
9 Altus aka Cao Artisan	Momotombo	1728	2016	60%	U.S.A.	2.75		Nicaragua
10 Amano	Morobe	725	2011	70%	U.S.A.	4.00		Papua New Guinea
11 Amstel (Simon Coll)	Ghana	322	2009	70%	Spain	3.00	Forastero	Ghana
12 Amazona	LamsadelChanka, San Martin, Oro Verde coop	1145	2013	72%	Peru	3.25		Peru
13 Ambrosia	Belize	1494	2015	64%	Canada	3.00	Trinitario	Belize
14 Amedei	Plura, Blanco de Criollo	979	2012	70%	Italy	3.75		Peru
15 AMMA	Catongo	1065	2013	75%	Brazil	3.25	Forastero (Catongo)	Brazil
16 Anahata	Elvesia	1239	2014	75%	U.S.A.	3.00		Dominican Republic
17 Animas	Alto Beni	1852	2016	75%	U.S.A.	3.50		Bolivia
18 Ara	Madagascar	1375	2014	75%	France	3.00	Trinitario	Madagascar
19 Arete	San Juan	1724	2016	70%	U.S.A.	3.50	Trinitario	Trinidad
20 Artisan du Chocolat	Trinidad, Heritage, Limited ed.	1193	2013	72%	UK	3.25	Trinitario	Trinidad
21 Artisan du Chocolat (Casa Luker)	Orinoquia Region, Arauca	1181	2013	72%	UK	2.75	Trinitario	Colombia

Showing 1 to 21 of 416 entries, 9 total columns

Console | Terminal | Background Jobs |

```
R - 4.5.2 - ~ - ✓
> View(unique_companies)
> View(cacao_df)
> View(cacao_clean)
> View(clean_exact)
> View(unique_bars)
> View(unique_companies)
> |
```

File Explorer:

- Home
- My Pictures
- My Videos
- My Web Sites
- NetBeansProjects
- Power BI Desktop
- scala for D5
- Sound Recordings
- T050\_AgnilT
- vgsales.csv
- Virtual Machines
- Visual Studio 18
- Visual Studio 2022
- WindowsPowerShell
- Student.csv
- car\_price\_prediction\_csv
- disease\_diagnosis.csv
- flavors\_of\_cacao.csv

Taskbar:

- BAJFINANCE -1.85%
- Search bar
- Icons for Microsoft Edge, File Explorer, Task View, Taskbar settings, and Start button
- System tray: ENG IN, 11:49, 08-12-2025