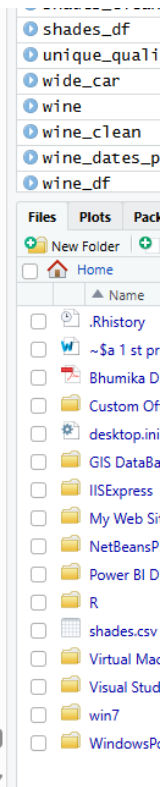


SHETH L.U.J.AND SIR M.V. COLLEGE
Data Analysis with SAS / SPSS /R

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

OUTPUT:-

```
> print("--- 1. Original wide Data ---")
[1] "--- 1. Original Wide Data ---"
> print(head(car_df))
  CarID brand  model  price mileage  fuel horsepower transmission
1     1  Toyota Innova 1800000     11 Diesel        150    Automatic
2     2  Hyundai Creta 1500000     17 Petrol        115     Manual
3     3  Maruti  Swift  700000     22 Petrol         83     Manual
4     4   Honda  City 1400000     18 Petrol        119    Automatic
5     5    Tata  Nexon 1200000     20 Diesel        110     Manual
6     6 Mahindra  Thar 1700000     14 Diesel        130     Manual
>
>
> long_car <- car_df %>%
+   pivot_longer(
+     cols = c(price, mileage, horsepower),
+     names_to = "Metric",
+     values_to = "value"
+   )
>
> print("--- 2. Long Format ---")
[1] "--- 2. Long Format ---"
> print(head(long_car, 10))
# A tibble: 10 x 7
  CarID brand  model  fuel  transmission Metric      value
  <int> <chr>   <chr>   <chr>   <chr>         <chr>    <int>
1     1  Toyota Innova Diesel Automatic price    1800000
2     1  Toyota Innova Diesel Automatic mileage    11
3     1  Toyota Innova Diesel Automatic horsepower 150
4     2  Hyundai Creta Petrol Manual price    1500000
5     2  Hyundai Creta Petrol Manual mileage    17
6     2  Hyundai Creta Petrol Manual horsepower 115
7     3  Maruti  Swift Petrol Manual price    700000
8     3  Maruti  Swift Petrol Manual mileage    22
9     3  Maruti  Swift Petrol Manual horsepower 83
10    4  Honda  City Petrol Automatic price    1400000
>
>
> wide_car <- long_car %>%
+   pivot_wider(
+     names_from = Metric,
```



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

Data Analysis with SAS / SPSS /R

```
Console Terminal Background Jobs
R 4.1.2 · D:/S079_VIBHUTI/ADV PYTHON FOR DATA SCIENCE/

> wide_car <- long_car %>%
+   pivot_wider(
+     names_from = Metric,
+     values_from = value
+   )
>
> print("--- 3. Wide Format ---")
[1] "--- 3. Wide Format ---"
> print(head(wide_car))
# A tibble: 6 x 8
  CarID brand    model fuel transmission price mileage horsepower
  <int> <chr>    <chr> <chr>    <chr>    <int> <int>    <int>
1     1 Toyota  Innova Diesel Automatic 1800000    11     150
2     2 Hyundai  Creta  Petrol  Manual   1500000    17     115
3     3 Maruti   Swift  Petrol  Manual    700000    22      83
4     4 Honda    City   Petrol  Automatic 1400000    18     119
5     5 Tata     Nexon  Diesel  Manual   1200000    20     110
6     6 Mahindra Thar   Diesel  Manual   1700000    14     130
>
>
> fuel_pivot <- car_df %>%
+   select(CarID, fuel, price) %>%
+   pivot_wider(
+     names_from = fuel,
+     values_from = price
+   )
>
> print("--- 4. Fuel Pivot ---")
[1] "--- 4. Fuel Pivot ---"
> print(head(fuel_pivot))
# A tibble: 6 x 3
  CarID Diesel Petrol
  <int> <int> <int>
1     1 1800000 NA
2     2 NA 1500000
3     3 NA 700000
4     4 NA 1400000
5     5 1200000 NA
6     6 1700000 NA
> long_car <- car_df %>%
```

```
# A tibble: 6 x 3
  CarID Diesel Petrol
  <int> <int> <int>
1     1 1800000 NA
2     2 NA 1500000
3     3 NA 700000
4     4 NA 1400000
5     5 1200000 NA
6     6 1700000 NA
> long_car <- car_df %>%
+   pivot_longer(
+     cols = c(price, mileage, horsepower),
+     names_to = "Metric",
+     values_to = "Value"
+   )
> |
```

12) AIM:- Combining datasets vertically (concatenation) using rbind() (R).

OUTPUT:-

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

Data Analysis with SAS / SPSS /R

```
> library(dplyr)
>
> wine_df <- read.csv("D:/S079_VIBHUTI/ADV PYTHON FOR DATA SCIENCE/winequality-red.csv")
> shades_df <- read.csv("D:/S079_VIBHUTI/ADV PYTHON FOR DATA SCIENCE/shades.csv")
>
> print("--- Column Names Before Alignment ---")
[1] "---- Column Names Before Alignment ----"
> print(names(wine_df))
[1] "fixed.acidity"      "volatile.acidity"    "citric.acid"         "residual.sugar"
[5] "chlorides"         "free.sulfur.dioxide" "total.sulfur.dioxide" "density"
[9] "pH"                "sulphates"          "alcohol"             "quality"
> print(names(shades_df))
[1] "brand"              "brand_short"        "product"             "product_short"      "hex"
[6] "H"                  "S"                  "V"                   "L"                  "group"
>
> wine_clean <- wine_df %>%
+   select(quality) %>%
+   mutate(Type = "wine") %>%
+   rename(value = quality)
>
> shades_clean <- shades_df %>%
+   select(product) %>%
+   mutate(Type = "shade") %>%
+   rename(value = product)
>
> combined_data <- rbind(wine_clean, shades_clean)
>
> print("--- Combined Data Preview ---")
[1] "---- Combined Data Preview ----"
> print(head(combined_data))
  value Type
1     5 wine
2     5 wine
3     5 wine
4     6 wine
5     5 wine
6     5 wine
```

```
> wine_clean <- wine_df %>%
+   select(quality) %>%
+   mutate(Type = "wine") %>%
+   rename(value = quality)
>
> shades_clean <- shades_df %>%
+   select(product) %>%
+   mutate(Type = "Shade") %>%
+   rename(value = product)
>
> combined_data <- rbind(wine_clean, shades_clean)
>
> print("--- Combined Data Preview ---")
[1] "---- Combined Data Preview ----"
> print(head(combined_data))
  value Type
1     5 wine
2     5 wine
3     5 wine
4     6 wine
5     5 wine
6     5 wine
> print(tail(combined_data))
  value Type
2219 True Match Shade
2220 True Match Shade
2221 True Match Shade
2222 True Match Shade
2223 True Match Shade
2224 True Match Shade
>
> print("--- Row Summary ---")
[1] "---- Row Summary ----"
> print(paste("wine rows:", nrow(wine_clean)))
[1] "wine rows: 1599"
> print(paste("shades rows:", nrow(shades_clean)))
[1] "shades rows: 625"
> print(paste("Total rows:", nrow(combined_data)))
[1] "Total rows: 2224"
>
```

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

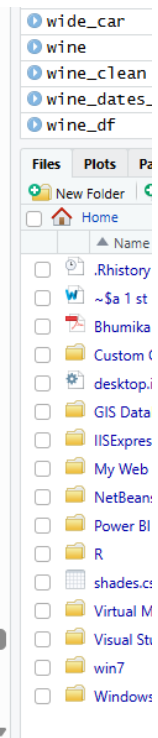
OUTPUT:-

VIBHUTI GAWADE
S079

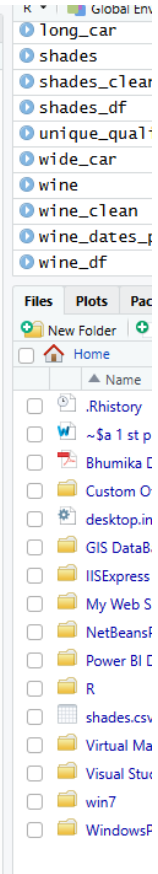
SHETH L.U.J.AND SIR M.V. COLLEGE

Data Analysis with SAS / SPSS /R

```
> print("--- 1. Original wine Dataset ---")
[1] "--- 1. Original wine Dataset ---"
> print(head(wine_df))
  fixed.acidity volatile.acidity citric.acid residual.sugar chlorides free.sulfur.dioxide
1          7.4           0.70         0.00           1.9      0.076             11
2          7.8           0.88         0.00           2.6      0.098             25
3          7.8           0.76         0.04           2.3      0.092             15
4         11.2           0.28         0.56           1.9      0.075             17
5          7.4           0.70         0.00           1.9      0.076             11
6          7.4           0.66         0.00           1.8      0.075             13
  total.sulfur.dioxide density    pH sulphates alcohol quality
1             34 0.9978 3.51      0.56      9.4      5
2             67 0.9968 3.20      0.68      9.8      5
3             54 0.9970 3.26      0.65      9.8      5
4             60 0.9980 3.16      0.58      9.8      6
5             34 0.9978 3.51      0.56      9.4      5
6             40 0.9978 3.51      0.56      9.4      5
>
> duplicates_report <- wine_df %>%
+   group_by(across(everything())) %>%
+   count() %>%
+   filter(n > 1)
>
> print("--- 2. Rows that are duplicated (Full duplicate report) ---")
[1] "--- 2. Rows that are duplicated (Full duplicate report) ---"
> print(duplicates_report)
# A tibble: 220 x 13
# Groups:   fixed.acidity, volatile.acidity, citric.acid, residual.sugar, chlorides,
#   free.sulfur.dioxide, total.sulfur.dioxide, density, pH, sulphates, alcohol, quality [220]
  fixed.acidity volatile.acidity citric.acid residual.sugar chlorides free.sulfur.dioxide
1             5.2           0.34         0           1.8      0.05             27
2             5.6           0.5         0.09         2.3      0.049             17
3             5.6           0.54         0.04         1.7      0.049              5
4             5.6           0.66         0           2.2      0.087              3
5             5.9           0.61         0.08         2.1      0.071             16
6             6.1           0.32         0.25         2.3      0.071             23
7             6.1           0.36         0.24         2.2      0.095             19
8             6.2           0.56         0.09         1.7      0.053             24
9             6.2           0.5         0           1.4      0.057             15
10            6.2           0.51         0           2.1      0.064             40
11            6.7           0.58         0.08         1.8      0.097             15
12            5.6           0.615        0.00         1.6      0.089             16
13            7.8           0.610        0.29         1.6      0.114              9
14            8.9           0.620        0.18         3.8      0.176             52
15            8.5           0.280        0.56         1.8      0.092             35
16            8.1           0.560        0.28         1.7      0.368             16
17            7.4           0.590        0.08         4.4      0.086              6
18            7.9           0.320        0.51         1.8      0.341             17
19            8.9           0.220        0.48         1.8      0.077             29
20            7.6           0.390        0.31         2.3      0.082             23
21            7.9           0.430        0.21         1.6      0.106             10
22            8.5           0.490        0.11         2.3      0.084              9
```



```
3             5.6           0.54         0.04         1.7      0.049              5
4             5.6           0.66         0           2.2      0.087              3
5             5.9           0.61         0.08         2.1      0.071             16
6             6           0.5         0           1.4      0.057             15
7             6           0.51         0           2.1      0.064             40
8             6.1           0.32         0.25         2.3      0.071             23
9             6.2           0.36         0.24         2.2      0.095             19
10            6.2           0.56         0.09         1.7      0.053             24
# i 210 more rows
# i 7 more variables: total.sulfur.dioxide <dbl>, density <dbl>, pH <dbl>, sulphates <dbl>,
#   alcohol <dbl>, quality <int>, n <int>
# i Use 'print(n = ...)' to see more rows
>
> clean_exact <- wine_df %>%
+   distinct()
> print("--- 3. Dataset After Removing Exact Duplicates ---")
[1] "--- 3. Dataset After Removing Exact Duplicates ---"
> print(clean_exact)
  fixed.acidity volatile.acidity citric.acid residual.sugar chlorides free.sulfur.dioxide
1          7.4           0.700         0.00           1.90      0.076             11
2          7.8           0.880         0.00           2.60      0.098             25
3          7.8           0.760         0.04           2.30      0.092             15
4         11.2           0.280         0.56           1.90      0.075             17
5          7.4           0.660         0.00           1.80      0.075             13
6          7.9           0.600         0.06           1.60      0.069             15
7          7.3           0.650         0.00           1.20      0.065             15
8          7.8           0.580         0.02           2.00      0.073              9
9          7.5           0.500         0.36           6.10      0.071             17
10         6.7           0.580         0.08           1.80      0.097             15
11         5.6           0.615         0.00           1.60      0.089             16
12         7.8           0.610         0.29           1.60      0.114              9
13         8.9           0.620         0.18           3.80      0.176             52
14         8.9           0.620         0.19           3.90      0.170             51
15         8.5           0.280         0.56           1.80      0.092             35
16         8.1           0.560         0.28           1.70      0.368             16
17         7.4           0.590         0.08           4.40      0.086              6
18         7.9           0.320         0.51           1.80      0.341             17
19         8.9           0.220         0.48           1.80      0.077             29
20         7.6           0.390         0.31           2.30      0.082             23
21         7.9           0.430         0.21           1.60      0.106             10
22         8.5           0.490         0.11           2.30      0.084              9
```



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

Data Analysis with SAS / SPSS /R

R 4.1.2 · D:/S079_VIBHUTI/ADV PYTHON FOR DATA SCIENCE/

63	6.6	0.705	0.07	1.60	0.076	6
64	9.3	0.320	0.57	2.00	0.074	27
65	8.0	0.705	0.05	1.90	0.074	8
66	7.7	0.630	0.08	1.90	0.076	15
67	7.7	0.670	0.23	2.10	0.088	17
68	7.7	0.690	0.22	1.90	0.084	18
69	8.3	0.675	0.26	2.10	0.084	11
70	9.7	0.320	0.54	2.50	0.094	28
71	8.8	0.410	0.64	2.20	0.093	9
72	6.8	0.785	0.00	2.40	0.104	14
73	6.7	0.750	0.12	2.00	0.086	12
74	8.3	0.625	0.20	1.50	0.080	27
75	6.2	0.450	0.20	1.60	0.069	3
76	7.8	0.430	0.70	1.90	0.464	22
77	7.4	0.500	0.47	2.00	0.086	21
78	7.3	0.670	0.26	1.80	0.401	16
79	6.3	0.300	0.48	1.80	0.069	18
80	6.9	0.550	0.15	2.20	0.076	19
81	8.6	0.490	0.28	1.90	0.110	20
82	7.7	0.490	0.26	1.90	0.062	9
83	9.3	0.390	0.44	2.10	0.107	34
total.sulfur.dioxide density pH sulphates alcohol quality						
1	34	0.9978	3.51	0.56	9.4	5
2	67	0.9968	3.20	0.68	9.8	5
3	54	0.9970	3.26	0.65	9.8	5
4	60	0.9980	3.16	0.58	9.8	6
5	40	0.9978	3.51	0.56	9.4	5
6	59	0.9964	3.30	0.46	9.4	5
7	21	0.9946	3.39	0.47	10.0	7
8	18	0.9968	3.36	0.57	9.5	7
9	102	0.9978	3.35	0.80	10.5	5
10	65	0.9959	3.28	0.54	9.2	5
11	59	0.9943	3.58	0.52	9.9	5
12	29	0.9974	3.26	1.56	9.1	5
13	145	0.9986	3.16	0.88	9.2	5
14	148	0.9986	3.17	0.93	9.2	5
15	103	0.9969	3.30	0.75	10.5	7
16	56	0.9968	3.11	1.28	9.3	5
17	29	0.9974	3.38	0.50	9.0	4
18	56	0.9969	3.04	1.08	9.2	6
19	60	0.9968	3.39	0.53	9.4	6

R 4.1.2 · D:/S079_VIBHUTI/ADV PYTHON FOR DATA SCIENCE/

67	6.6	0.705	0.07	1.60	0.076	6
68	9.3	0.320	0.57	2.00	0.074	27
69	8.0	0.705	0.05	1.90	0.074	8
70	7.7	0.630	0.08	1.90	0.076	15
71	7.7	0.670	0.23	2.10	0.088	17
72	7.7	0.690	0.22	1.90	0.084	18
73	8.3	0.675	0.26	2.10	0.084	11
74	9.7	0.320	0.54	2.50	0.094	28
75	8.8	0.410	0.64	2.20	0.093	9
76	6.8	0.785	0.00	2.40	0.104	14
77	6.7	0.750	0.12	2.00	0.086	12
78	8.3	0.625	0.20	1.50	0.080	27
79	6.2	0.450	0.20	1.60	0.069	3
80	7.8	0.430	0.70	1.90	0.464	22
81	7.4	0.500	0.47	2.00	0.086	21
82	7.3	0.670	0.26	1.80	0.401	16
83	6.3	0.300	0.48	1.80	0.069	18
84	6.9	0.550	0.15	2.20	0.076	19
85	8.6	0.490	0.28	1.90	0.110	20
86	7.7	0.490	0.26	1.90	0.062	9
87	9.3	0.390	0.44	2.10	0.107	34
total.sulfur.dioxide density pH sulphates alcohol quality						
1	34	0.9978	3.51	0.56	9.4	5
2	67	0.9968	3.20	0.68	9.8	5
3	54	0.9970	3.26	0.65	9.8	5
4	60	0.9980	3.16	0.58	9.8	6
5	40	0.9978	3.51	0.56	9.4	5
6	59	0.9964	3.30	0.46	9.4	5
7	21	0.9946	3.39	0.47	10.0	7
8	18	0.9968	3.36	0.57	9.5	7
9	102	0.9978	3.35	0.80	10.5	5
10	65	0.9959	3.28	0.54	9.2	5
11	59	0.9943	3.58	0.52	9.9	5
12	29	0.9974	3.26	1.56	9.1	5
13	145	0.9986	3.16	0.88	9.2	5
14	148	0.9986	3.17	0.93	9.2	5
15	103	0.9969	3.30	0.75	10.5	7
16	56	0.9968	3.11	1.28	9.3	5
17	29	0.9974	3.38	0.50	9.0	4
18	56	0.9969	3.04	1.08	9.2	6
19	60	0.9968	3.39	0.53	9.4	6

```

> unique_quality <- wine_df %>%
+   distinct(quality, .keep_all = TRUE)
> print("--- 4. Unique Quality values (only first appearance kept) ---")
[1] "--- 4. Unique Quality values (only first appearance kept) ---"
> print(unique_quality)
fixed.acidity volatile.acidity citric.acid residual.sugar chlorides free.sulfur.dioxide
1      7.4          0.70         0.00          1.9      0.076          11
2     11.2          0.28         0.56          1.9      0.075          17
3      7.3          0.65         0.00          1.2      0.065          15
4      7.4          0.59         0.08          4.4      0.086           6
5      7.9          0.35         0.46          3.6      0.078          15
6     11.6          0.58         0.66          2.2      0.074          10
total.sulfur.dioxide density pH sulphates alcohol quality
1      34 0.9978 3.51 0.56 9.4 5
2     67 0.9968 3.16 0.58 9.8 6
3     21 0.9946 3.39 0.47 10.0 7
4     29 0.9974 3.38 0.50 9.0 4
5     37 0.9973 3.35 0.86 12.8 8
6     47 1.0008 3.25 0.57 9.0 3

```

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

Data Analysis with SAS / SPSS /R

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

OUTPUT:-

```
> install.packages("lubridate")
> library(lubridate)
> library(dplyr)
> setwd("D:/S079_VIBHUTI/ADV PYTHON FOR DATA SCIENCE")
> print(list.files())
[1] "cardata.csv" "cleaned_cardata.csv"
[3] "cleaned_data.csv" "cleaned_dataset_summary.csv"
[5] "cleaned_IRIS.csv" "cleaned_shades.csv"
[7] "cleaned_Student_Mental_Health.csv" "cleaned_winequality-red.csv"
[9] "dataset_summary.csv" "day.csv"
[11] "IRIS.csv" "Processed_dataset_summary.csv"
[13] "Processed_IRIS.csv" "Processed_Shades.csv"
[15] "Processed_Shades_Summary.csv" "Processed_Student_Mental_Health.csv"
[17] "Processed_WineQuality_Red.csv" "S079_PRAC1.py"
[19] "S079_VIBHUTI_PRAC1.docx" "S079_VIBHUTI_PRAC1.pdf"
[21] "S079vibhu.py" "S079VIBHUTI.PY"
[23] "S079VIBHUTI_PRAC11.pdf" "S079VIBHUTI_PRAC12.pdf"
[25] "S079VIBHUTI_PRAC13.pdf" "S079VIBHUTI_PRAC14.pdf"
[27] "S079VIBHUTI_PRAC15.pdf" "S079VIBHUTI_PRAC2.pdf"
[29] "S079VIBHUTI_PRAC3.pdf" "S079VIBHUTI_PRAC4.pdf"
[31] "S079VIBHUTI_PRAC6.pdf" "sales_data.csv"
[33] "shades.csv" "Social_Network_Ads.csv"
[35] "Student_Mental_health.csv" "ua_passengers.csv"
[37] "vibh.ipynb" "VIBHUTI_PRAC11.R"
[39] "VIBHUTI_PRAC12.R" "VIBHUTI_PRAC13.R"
[41] "VIBHUTI_PRAC14.R" "VIBHUTI_PRAC15.R"
[43] "vibhuti_prac3.py" "vibhuti_prac6.py"
[45] "VIBHUTI_PRAC6.R" "VIBHUTIS079.ipynb"
[47] "VIBHUTIS079.PY" "winequality-red.csv"
>
> wine <- read.csv("winequality-red.csv")
> head(wine)
  fixed.acidity volatile.acidity citric.acid residual.sugar chlorides free.sulfur.dioxide
1         7.4             0.70         0.00           1.9      0.076              11
2         7.8             0.88         0.00           2.6      0.098              25
3         7.8             0.76         0.04           2.3      0.092              15
```

```
total.sulfur.dioxide density pH sulphates alcohol quality
1          34 0.9978 3.51    0.56    9.4      5
2          67 0.9968 3.20    0.68    9.8      5
3          54 0.9970 3.26    0.65    9.8      5
4          60 0.9980 3.16    0.58    9.8      6
5          34 0.9978 3.51    0.56    9.4      5
6          40 0.9978 3.51    0.56    9.4      5
>
> wine$date_recorded <- seq(
+   from = as.Date("2023-01-01"),
+   length.out = nrow(wine),
+   by = "day"
+ )
>
> wine_dates_processed <- wine %>%
+   mutate(
+     Actual_Date = ymd(date_recorded),
+     Year_Num = year(Actual_Date),
+     Month_Num = month(Actual_Date),
+     Month_Name = month(Actual_Date, label = TRUE),
+     Day_Num = day(Actual_Date),
+     Weekday_Num = wday(Actual_Date),
+     Weekday_Name = wday(Actual_Date, label = TRUE, abbr = FALSE),
+     Quarter = quarter(Actual_Date),
+     Day_of_Year = yday(Actual_Date)
+   )
>
> print(head(wine_dates_processed))
  fixed.acidity volatile.acidity citric.acid residual.sugar chlorides free.sulfur.dioxide
1         7.4             0.70         0.00           1.9      0.076              11
2         7.8             0.88         0.00           2.6      0.098              25
3         7.8             0.76         0.04           2.3      0.092              15
4         11.2            0.28         0.56           1.9      0.075              17
5         7.4             0.70         0.00           1.9      0.076              11
6         7.4             0.66         0.00           1.8      0.075              13
```

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

Data Analysis with SAS / SPSS /R

```

43:1 (Top Level) R Script
Console Terminal Background Jobs
R v 4.1.2 - D:/S079_VIBHUTI/ADV PYTHON FOR DATA SCIENCE/
+ )
> print(head(wine_data_processed))
fixed.acidity volatile.acidity citric.acid residual.sugar chlorides free.sulfur.dioxide
1 7.4 0.70 0.00 1.9 0.076 11
2 7.8 0.88 0.00 2.6 0.098 25
3 7.8 0.76 0.04 2.3 0.092 15
4 11.2 0.28 0.56 1.9 0.075 17
5 7.4 0.70 0.00 1.9 0.076 11
6 7.4 0.66 0.00 1.8 0.075 13
total.sulfur.dioxide density pH sulphates alcohol quality date_recorded Actual_Date Year_Num
1 34 0.9978 3.51 0.56 9.4 5 2023-01-01 2023-01-01 2023
2 67 0.9968 3.20 0.68 9.8 5 2023-01-02 2023-01-02 2023
3 54 0.9970 3.26 0.65 9.8 5 2023-01-03 2023-01-03 2023
4 60 0.9980 3.16 0.58 9.8 6 2023-01-04 2023-01-04 2023
5 34 0.9978 3.51 0.56 9.4 5 2023-01-05 2023-01-05 2023
6 40 0.9978 3.51 0.56 9.4 5 2023-01-06 2023-01-06 2023
Month_Num Month_Name Day_Num Weekday_Num Weekday_Name Quarter Day_of_Year
1 1 Jan 1 1 Sunday 1 1
2 1 Jan 2 2 Monday 1 2
3 1 Jan 3 3 Tuesday 1 3
4 1 Jan 4 4 Wednesday 1 4
5 1 Jan 5 5 Thursday 1 5
6 1 Jan 6 6 Friday 1 6
>
> current_time <- now()
> print(paste("Current Year:", year(current_time)))
[1] "Current Year: 2025"
> print(paste("Current Month:", month(current_time)))
[1] "Current Month: 12"
> print(paste("Current Day:", day(current_time)))
[1] "Current Day: 8"
> print(paste("Current Hour:", hour(current_time)))
[1] "Current Hour: 12"
> print(paste("Current Minute:", minute(current_time)))
[1] "Current Minute: 20"

```

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

OUTPUT:-

```

> setwd("D:/S079_VIBHUTI/ADV PYTHON FOR DATA SCIENCE")
>
> car_df <- read.csv("cardata.csv")
> print("--- Car Data Loaded Successfully ---")
[1] "--- Car Data Loaded Successfully ---"
>
> head(car_df)
  Car_Name Year Selling_Price Present_Price Kms_Driven Fuel_Type Seller_Type Transmission
1    ritz 2014         3.35         5.59      27000    Petrol    Dealer      Manual
2    sx4 2013         4.75         9.54      43000    Diesel    Dealer      Manual
3    ciaz 2017         7.25         9.85       6900    Petrol    Dealer      Manual
4  wagon r 2011         2.85         4.15       5200    Petrol    Dealer      Manual
5  swift 2014         4.60         6.87      42450    Diesel    Dealer      Manual
6 vitara brezza 2018         9.25         9.83       2071    Diesel    Dealer      Manual
owner
1 0
2 0
3 0
4 0
5 0
6 0
>
> print("--- OUTPUT OF str() ---")
[1] "--- OUTPUT OF str() ---"
> str(car_df)
'data.frame':   301 obs. of  9 variables:
 $ Car_Name   : chr  "ritz" "sx4" "ciaz" "wagon r" ...
 $ Year       : int   2014 2013 2017 2011 2014 2018 2015 2016 2015 ...
 $ Selling_Price: num   3.35 4.75 7.25 2.85 4.6 9.25 6.75 6.5 8.75 7.45 ...
 $ Present_Price: num   5.59 9.54 9.85 4.15 6.87 9.83 8.12 8.61 8.89 8.92 ...
 $ Kms_Driven  : int   27000 43000 6900 5200 42450 2071 18796 33429 20273 42367 ...
 $ Fuel_Type   : chr   "Petrol" "Diesel" "Petrol" "Petrol" ...
 $ Seller_Type : chr   "Dealer" "Dealer" "Dealer" "Dealer" ...
 $ Transmission: chr   "Manual" "Manual" "Manual" "Manual" ...
 $ owner       : int   0 0 0 0 0 0 0 0 0 0 ...
>

```

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

Data Analysis with SAS / SPSS / R

```

Console Terminal Background Jobs
R 4.1.2 · D:/S079_VIBHUTI/ADV PYTHON FOR DATA SCIENCE/

$ Transmission : chr "Manual" "Manual" "Manual" "Manual" ...
$ Owner : int 0 0 0 0 0 0 0 0 0 ...

> print("--- OUTPUT OF summary() [Before Factor Conversion] ---")
[1] "--- OUTPUT OF summary() [Before Factor Conversion] ---"
> summary(car_df)
  Car_Name      Year  Selling_Price Present_Price  Kms_Driven
Length:301  Min. :2003  Min. : 0.100  Min. : 0.320  Min. : 500
Class :character 1st Qu.:2012 1st Qu.: 0.900 1st Qu.: 1.200 1st Qu.: 15000
Mode :character  Median :2014 Median : 3.600 Median : 6.400 Median : 32000
                Mean :2014  Mean : 4.661 Mean : 7.628 Mean : 36947
                3rd Qu.:2016 3rd Qu.: 6.000 3rd Qu.: 9.900 3rd Qu.: 48767
                Max. :2018  Max. :35.000 Max. :92.600 Max. :500000

  Fuel_Type  Seller_Type  Transmission  Owner
Length:301  Length:301  Length:301  Min. :0.00000
Class :character  Class :character  Class :character 1st Qu.:0.00000
Mode :character  Mode :character  Mode :character  Median :0.00000
                Mean :0.04319
                3rd Qu.:0.00000
                Max. :3.00000

> factor_cols <- c("FuelType", "Transmission", "Owner", "Seller_Type")
> factor_cols <- factor_cols[factor_cols %in% names(car_df)]
> car_df[factor_cols] <- lapply(car_df[factor_cols], as.factor)
> print("--- OUTPUT OF summary() [After Factor Conversion] ---")
[1] "--- OUTPUT OF summary() [After Factor Conversion] ---"
> summary(car_df)
  Car_Name      Year  Selling_Price Present_Price  Kms_Driven
Length:301  Min. :2003  Min. : 0.100  Min. : 0.320  Min. : 500
Class :character 1st Qu.:2012 1st Qu.: 0.900 1st Qu.: 1.200 1st Qu.: 15000
Mode :character  Median :2014 Median : 3.600 Median : 6.400 Median : 32000
                Mean :2014  Mean : 4.661 Mean : 7.628 Mean : 36947
                3rd Qu.:2016 3rd Qu.: 6.000 3rd Qu.: 9.900 3rd Qu.: 48767
                Max. :2018  Max. :35.000 Max. :92.600 Max. :500000

  Fuel_Type  Seller_Type  Transmission  Owner
Length:301  Dealer :195 Automatic: 40 0:290
Class :character Individual:106 Manual :261 1: 10
Mode :character                               3: 1

```

```

Console Terminal Background Jobs
R 4.1.2 · D:/S079_VIBHUTI/ADV PYTHON FOR DATA SCIENCE/

  3rd Qu.: 6.00000
  Max. :3.00000

> factor_cols <- c("FuelType", "Transmission", "Owner", "Seller_Type")
> factor_cols <- factor_cols[factor_cols %in% names(car_df)]
> car_df[factor_cols] <- lapply(car_df[factor_cols], as.factor)
> print("--- OUTPUT OF summary() [After Factor Conversion] ---")
[1] "--- OUTPUT OF summary() [After Factor Conversion] ---"
> summary(car_df)
  Car_Name      Year  Selling_Price Present_Price  Kms_Driven
Length:301  Min. :2003  Min. : 0.100  Min. : 0.320  Min. : 500
Class :character 1st Qu.:2012 1st Qu.: 0.900 1st Qu.: 1.200 1st Qu.: 15000
Mode :character  Median :2014 Median : 3.600 Median : 6.400 Median : 32000
                Mean :2014  Mean : 4.661 Mean : 7.628 Mean : 36947
                3rd Qu.:2016 3rd Qu.: 6.000 3rd Qu.: 9.900 3rd Qu.: 48767
                Max. :2018  Max. :35.000 Max. :92.600 Max. :500000

  Fuel_Type  Seller_Type  Transmission  Owner
Length:301  Dealer :195 Automatic: 40 0:290
Class :character Individual:106 Manual :261 1: 10
Mode :character                               3: 1

> avg_kms <- mean(car_df$Kms_Driven, na.rm = TRUE)
> max_price <- max(car_df$Selling_Price, na.rm = TRUE)
> min_year <- min(car_df$Year, na.rm = TRUE)
> max_year <- max(car_df$Year, na.rm = TRUE)
> print(paste("Average KM Driven:", avg_kms))
[1] "Average KM Driven: 36947.2059800664"
> print(paste("Highest Selling Price:", max_price))
[1] "Highest Selling Price: 35"
> print(paste("Oldest Car Year:", min_year))
[1] "Oldest Car Year: 2003"
> print(paste("Newest Car Year:", max_year))
[1] "Newest Car Year: 2018"
>

```