> library(dplyr)

Attaching package: ‘dplyr’

The following objects are masked from ‘package:stats’:

filter, lag

The following objects are masked from ‘package:base’:

intersect, setdiff, setequal, union

#**EXPLORE THE DATASET (OUTPUT)**

> **summary(DATA)**

order\_id payment\_sequential payment\_type payment\_installments

Length:103886 Min. : 1.000 Length:103886 Min. : 0.000

Class :character 1st Qu.: 1.000 Class :character 1st Qu.: 1.000

Mode :character Median : 1.000 Mode :character Median : 1.000

Mean : 1.093 Mean : 2.853

3rd Qu.: 1.000 3rd Qu.: 4.000

Max. :29.000 Max. :24.000

payment\_value

Min. : 0.00

1st Qu.: 56.79

Median : 100.00

Mean : 154.10

3rd Qu.: 171.84

Max. :13664.08

**> head(DATA)**

order\_id payment\_sequential payment\_type

1 b81ef226f3fe1789b1e8b2acac839d17 1 credit\_card

2 a9810da82917af2d9aefd1278f1dcfa0 1 credit\_card

3 25e8ea4e93396b6fa0d3dd708e76c1bd 1 credit\_card

4 ba78997921bbcdc1373bb41e913ab953 1 credit\_card

5 42fdf880ba16b47b59251dd489d4441a 1 credit\_card

6 298fcdf1f73eb413e4d26d01b25bc1cd 1 credit\_card

payment\_installments payment\_value

1 8 99.33

2 1 24.39

3 1 65.71

4 8 107.78

5 2 128.45

6 2 96.12

**#DIMENSIONS**

> dim(DATA)

[1] 103886 5

> nrow(DATA)

[1] 103886

> ncol(DATA)

[1] 5

> **colnames(DATA)**

[1] "order\_id" "payment\_sequential" "payment\_type"

[4] "payment\_installments" "payment\_value"

> **summary(DATA$payment\_type)**

Length Class Mode

103886 character character

> **str(DATA)**

'data.frame': 103886 obs. of 5 variables:

$ order\_id : chr "b81ef226f3fe1789b1e8b2acac839d17" "a9810da82917af2d9aefd1278f1dcfa0" "25e8ea4e93396b6fa0d3dd708e76c1bd" "ba78997921bbcdc1373bb41e913ab953" ...

$ payment\_sequential : int 1 1 1 1 1 1 1 1 1 1 ...

$ payment\_type : chr "credit\_card" "credit\_card" "credit\_card" "credit\_card" ...

$ payment\_installments: int 8 1 1 8 2 2 1 3 6 1 ...

$ payment\_value : num 99.3 24.4 65.7 107.8 128.4 ...

> **DATA%>%count(payment\_type)**

payment\_type n

1 boleto 19784

2 credit\_card 76795

3 debit\_card 1529

4 not\_defined 3

5 voucher 5775

> **DATA%>%count(payment\_installments)**

payment\_installments n

1 0 2

2 1 52546

3 2 12413

4 3 10461

5 4 7098

6 5 5239

7 6 3920

8 7 1626

9 8 4268

10 9 644

11 10 5328

12 11 23

13 12 133

14 13 16

15 14 15

16 15 74

17 16 5

18 17 8

19 18 27

20 20 17

21 21 3

22 22 1

23 23 1

24 24 18

> **DATA%>%glimpse()**

Rows: 103,886

Columns: 5

$ order\_id <chr> "b81ef226f3fe1789b1e8b2acac839d17", "a9810da82917~

$ payment\_sequential <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~

$ payment\_type <chr> "credit\_card", "credit\_card", "credit\_card", "cre~

$ payment\_installments <int> 8, 1, 1, 8, 2, 2, 1, 3, 6, 1, 8, 1, 1, 5, 4, 10,

$ payment\_value <dbl> 99.33, 24.39, 65.71, 107.78, 128.45, 96.12, 81.16~