

# 510-Capstone-Code-ML

Rochelle Rafn

2022-07-26

```
## 'data.frame': 9267992 obs. of 14 variables:
## $ X : int 1 2 3 4 5 6 7 8 9 10 ...
## $ customer_id : chr "00000dbacae5abe5e23885899a1fa44253a17956c6d1c3d25f88aa139fdcf657"
## $ Active : int 0 0 0 0 0 0 0 0 0 0 ...
## $ club_member_status : chr "ACTIVE" "ACTIVE" "ACTIVE" "ACTIVE" ...
## $ postal_code : chr "52043ee2162cf5aa7ee79974281641c6f11a68d276429a91f8ca0d4b6efa8100"
## $ t_dat : chr "2018-12-27" "2018-12-27" "2018-12-27" "2019-05-02" ...
## $ article_id : int 625548001 176209023 627759010 697138006 568601006 568601006 60764...
## $ price : num 0.0441 0.0356 0.0305 0.0102 0.0508 ...
## $ sales_channel_id : int 1 1 1 2 2 2 1 1 2 2 ...
## $ age_group : chr "mature" "mature" "mature" "mature" ...
## $ s_age : num 1.01 1.01 1.01 1.01 1.01 ...
## $ words : chr "bb chris puff jkt tp jacket garment upper body solid dark blue d...
## $ graphical_appearance_name: chr "Solid" "Solid" "Solid" "All over pattern" ...
## $ age : int 49 49 49 49 49 49 49 49 49 49 ...

## grouped_df [9,267,992 x 4] (S3: grouped_df/tbl_df/tbl/data.frame)
## $ customer_id: chr [1:9267992] "00000dbacae5abe5e23885899a1fa44253a17956c6d1c3d25f88aa139fdcf657" "
## $ t_dat : chr [1:9267992] "2018-12-27" "2018-12-27" "2018-12-27" "2019-05-02" ...
## $ age : int [1:9267992] 49 49 49 49 49 49 49 49 49 49 ...
## $ age_group : chr [1:9267992] "mature" "mature" "mature" "mature" ...
## - attr(*, "groups")= tibble [2,651,972 x 3] (S3: tbl_df/tbl/data.frame)
## ..$ customer_id: chr [1:2651972] "00000dbacae5abe5e23885899a1fa44253a17956c6d1c3d25f88aa139fdcf657"
## ..$ t_dat : chr [1:2651972] "2018-12-27" "2019-05-02" "2019-05-25" "2019-07-25" ...
## ..$ .rows : list<int> [1:2651972]
## .. ..$ : int [1:3] 1 2 3
## .. ..$ : int 4
## .. ..$ : int [1:2] 5 6
## .. ..$ : int [1:2] 7 8
## .. ..$ : int 9
## .. ..$ : int [1:2] 10 11
## .. ..$ : int 12
## .. ..$ : int [1:3] 13 14 15
## .. ..$ : int [1:5] 16 17 18 19 20
## .. ..$ : int 21
## .. ..$ : int [1:2] 22 23
## .. ..$ : int [1:2] 24 25
## .. ..$ : int 26
## .. ..$ : int 27
## .. ..$ : int [1:3] 28 29 30
## .. ..$ : int [1:4] 31 32 33 34
## .. ..$ : int [1:2] 35 36
```

```

## .. ..$ : int [1:2] 37 38
## .. ..$ : int [1:2] 39 40
## .. ..$ : int [1:2] 41 42
## .. ..$ : int [1:4] 43 44 45 46
## .. ..$ : int 47
## .. ..$ : int 48
## .. ..$ : int 49
## .. ..$ : int [1:7] 50 51 52 53 54 55 56
## .. ..$ : int [1:4] 57 58 59 60
## .. ..$ : int 61
## .. ..$ : int 62
## .. ..$ : int [1:4] 63 64 65 66
## .. ..$ : int 67
## .. ..$ : int [1:7] 68 69 70 71 72 73 74
## .. ..$ : int [1:14] 75 76 77 78 79 80 81 82 83 84 ...
## .. ..$ : int 89
## .. ..$ : int [1:21] 90 91 92 93 94 95 96 97 98 99 ...
## .. ..$ : int 111
## .. ..$ : int 112
## .. ..$ : int [1:2] 113 114
## .. ..$ : int [1:4] 115 116 117 118
## .. ..$ : int [1:4] 119 120 121 122
## .. ..$ : int [1:6] 123 124 125 126 127 128
## .. ..$ : int [1:6] 129 130 131 132 133 134
## .. ..$ : int [1:4] 135 136 137 138
## .. ..$ : int [1:4] 139 140 141 142
## .. ..$ : int [1:10] 143 144 145 146 147 148 149 150 151 152
## .. ..$ : int [1:3] 153 154 155
## .. ..$ : int 156
## .. ..$ : int [1:3] 157 158 159
## .. ..$ : int [1:5] 160 161 162 163 164
## .. ..$ : int [1:4] 165 166 167 168
## .. ..$ : int [1:3] 169 170 171
## .. ..$ : int [1:2] 172 173
## .. ..$ : int [1:4] 174 175 176 177
## .. ..$ : int [1:7] 178 179 180 181 182 183 184
## .. ..$ : int [1:6] 185 186 187 188 189 190
## .. ..$ : int [1:8] 191 192 193 194 195 196 197 198
## .. ..$ : int [1:4] 199 200 201 202
## .. ..$ : int [1:3] 203 204 205
## .. ..$ : int [1:4] 206 207 208 209
## .. ..$ : int 210
## .. ..$ : int [1:2] 211 212
## .. ..$ : int 213
## .. ..$ : int [1:5] 214 215 216 217 218
## .. ..$ : int [1:4] 219 220 221 222
## .. ..$ : int [1:4] 223 224 225 226
## .. ..$ : int [1:5] 227 228 229 230 231
## .. ..$ : int [1:3] 232 233 234
## .. ..$ : int [1:6] 235 236 237 238 239 240
## .. ..$ : int [1:2] 241 242
## .. ..$ : int [1:5] 243 244 245 246 247
## .. ..$ : int [1:2] 248 249
## .. ..$ : int [1:2] 250 251

```

```
## ..$ : int 252
## ..$ : int [1:4] 253 254 255 256
## ..$ : int [1:5] 257 258 259 260 261
## ..$ : int [1:4] 262 263 264 265
## ..$ : int [1:2] 266 267
## ..$ : int 268
## ..$ : int [1:8] 269 270 271 272 273 274 275 276
## ..$ : int [1:4] 277 278 279 280
## ..$ : int [1:7] 281 282 283 284 285 286 287
## ..$ : int [1:3] 288 289 290
## ..$ : int [1:19] 291 292 293 294 295 296 297 298 299 300 ...
## ..$ : int [1:6] 310 311 312 313 314 315
## ..$ : int [1:2] 316 317
## ..$ : int 318
## ..$ : int 319
## ..$ : int 320
## ..$ : int [1:5] 321 322 323 324 325
## ..$ : int 326
## ..$ : int [1:4] 327 328 329 330
## ..$ : int 331
## ..$ : int 332
## ..$ : int [1:3] 333 334 335
## ..$ : int 336
## ..$ : int [1:2] 337 338
## ..$ : int [1:2] 339 340
## ..$ : int 341
## ..$ : int [1:2] 342 343
## ..$ : int [1:4] 344 345 346 347
## .. [list output truncated]
## ..@ ptype: int(0)
## ..- attr(*, ".drop")= logi TRUE
```

```
## # A tibble: 6 x 3
## # Groups:   customer_id, t_dat [6]
##   customer_id                t_dat      n
##   <chr>                <chr>    <int>
## 1 d00063b94dcb1342869d4994844a2742b5d62927f36843164fb3f818f630bca9 2018-1~    570
## 2 8465e2822ea250ecdd28081a8a09170f68d047cee45778fb7084b7aa264fe8c9 2020-0~    178
## 3 5af6d447ad97b87d8c70a08542ca8283753e3178f019395e023e61e3b92e09e2 2019-0~    171
## 4 a4eaac96b7da2fb0cd9210df432eadf9de240f7df5e03b8db718c813f67fad5 2019-0~    156
## 5 689e541449c3eebe0cb305100a03221f1bdf69824af961f595e11b1c637df5ce 2020-0~    153
## 6 ed0cb6fd4649f038efd9d1c028ccf44497a0881c3248000b0edf428d984b56d4 2019-1~    151
```

```
## grouped_df [9,267,992 x 5] (S3: grouped_df/tbl_df/tbl/data.frame)
## $ customer_id: chr [1:9267992] "d00063b94dcb1342869d4994844a2742b5d62927f36843164fb3f818f630bca9" "
## $ t_dat      : chr [1:9267992] "2018-10-14" "2018-10-14" "2018-10-14" "2018-10-14" ...
## $ n          : int [1:9267992] 570 570 570 570 570 570 570 570 570 570 ...
## $ age        : int [1:9267992] 27 27 27 27 27 27 27 27 27 27 ...
## $ age_group  : chr [1:9267992] "young" "young" "young" "young" ...
## - attr(*, "groups")= tibble [2,651,972 x 3] (S3: tbl_df/tbl/data.frame)
## ..$ customer_id: chr [1:2651972] "00000dbacae5abe5e23885899a1fa44253a17956c6d1c3d25f88aa139fd5c657
## ..$ t_dat      : chr [1:2651972] "2018-12-27" "2019-05-02" "2019-05-25" "2019-07-25" ...
## ..$ .rows      : list<int> [1:2651972]
## ..$ : int [1:3] 6097164 6097165 6097166
```

```

## .. ..$ : int 8480268
## .. ..$ : int [1:2] 7242012 7242013
## .. ..$ : int [1:2] 7242014 7242015
## .. ..$ : int 8480269
## .. ..$ : int [1:2] 7242016 7242017
## .. ..$ : int 8480270
## .. ..$ : int [1:3] 6097167 6097168 6097169
## .. ..$ : int [1:5] 4258190 4258191 4258192 4258193 4258194
## .. ..$ : int 8480271
## .. ..$ : int [1:2] 7242018 7242019
## .. ..$ : int [1:2] 7242020 7242021
## .. ..$ : int 8480272
## .. ..$ : int 8480273
## .. ..$ : int [1:3] 6097170 6097171 6097172
## .. ..$ : int [1:4] 5067300 5067301 5067302 5067303
## .. ..$ : int [1:2] 7242022 7242023
## .. ..$ : int [1:2] 7242024 7242025
## .. ..$ : int [1:2] 7242026 7242027
## .. ..$ : int [1:2] 7242028 7242029
## .. ..$ : int [1:4] 5067304 5067305 5067306 5067307
## .. ..$ : int 8480274
## .. ..$ : int 8480275
## .. ..$ : int 8480276
## .. ..$ : int [1:7] 3017663 3017664 3017665 3017666 3017667 3017668 3017669
## .. ..$ : int [1:4] 5067308 5067309 5067310 5067311
## .. ..$ : int 8480277
## .. ..$ : int 8480278
## .. ..$ : int [1:4] 5067312 5067313 5067314 5067315
## .. ..$ : int 8480279
## .. ..$ : int [1:7] 3017670 3017671 3017672 3017673 3017674 3017675 3017676
## .. ..$ : int [1:14] 1021526 1021527 1021528 1021529 1021530 1021531 1021532 1021533 1021534 1021535
## .. ..$ : int 8480280
## .. ..$ : int [1:21] 408560 408561 408562 408563 408564 408565 408566 408567 408568 408569 ...
## .. ..$ : int 8480281
## .. ..$ : int 8480282
## .. ..$ : int [1:2] 7242030 7242031
## .. ..$ : int [1:4] 5067316 5067317 5067318 5067319
## .. ..$ : int [1:4] 5067320 5067321 5067322 5067323
## .. ..$ : int [1:6] 3561374 3561375 3561376 3561377 3561378 3561379
## .. ..$ : int [1:6] 3561380 3561381 3561382 3561383 3561384 3561385
## .. ..$ : int [1:4] 5067324 5067325 5067326 5067327
## .. ..$ : int [1:4] 5067328 5067329 5067330 5067331
## .. ..$ : int [1:10] 1850768 1850769 1850770 1850771 1850772 1850773 1850774 1850775 1850776 1850777
## .. ..$ : int [1:3] 6097173 6097174 6097175
## .. ..$ : int 8480283
## .. ..$ : int [1:3] 6097176 6097177 6097178
## .. ..$ : int [1:5] 4258195 4258196 4258197 4258198 4258199
## .. ..$ : int [1:4] 5067332 5067333 5067334 5067335
## .. ..$ : int [1:3] 6097179 6097180 6097181
## .. ..$ : int [1:2] 7242032 7242033
## .. ..$ : int [1:4] 5067336 5067337 5067338 5067339
## .. ..$ : int [1:7] 3017677 3017678 3017679 3017680 3017681 3017682 3017683
## .. ..$ : int [1:6] 3561386 3561387 3561388 3561389 3561390 3561391
## .. ..$ : int [1:8] 2546271 2546272 2546273 2546274 2546275 2546276 2546277 2546278

```

```

## .. ..$ : int [1:4] 5067340 5067341 5067342 5067343
## .. ..$ : int [1:3] 6097182 6097183 6097184
## .. ..$ : int [1:4] 5067344 5067345 5067346 5067347
## .. ..$ : int 8480284
## .. ..$ : int [1:2] 7242034 7242035
## .. ..$ : int 8480285
## .. ..$ : int [1:5] 4258200 4258201 4258202 4258203 4258204
## .. ..$ : int [1:4] 5067348 5067349 5067350 5067351
## .. ..$ : int [1:4] 5067352 5067353 5067354 5067355
## .. ..$ : int [1:5] 4258205 4258206 4258207 4258208 4258209
## .. ..$ : int [1:3] 6097185 6097186 6097187
## .. ..$ : int [1:6] 3561392 3561393 3561394 3561395 3561396 3561397
## .. ..$ : int [1:2] 7242036 7242037
## .. ..$ : int [1:5] 4258210 4258211 4258212 4258213 4258214
## .. ..$ : int [1:2] 7242038 7242039
## .. ..$ : int [1:2] 7242040 7242041
## .. ..$ : int 8480286
## .. ..$ : int [1:4] 5067356 5067357 5067358 5067359
## .. ..$ : int [1:5] 4258215 4258216 4258217 4258218 4258219
## .. ..$ : int [1:4] 5067360 5067361 5067362 5067363
## .. ..$ : int [1:2] 7242042 7242043
## .. ..$ : int 8480287
## .. ..$ : int [1:8] 2546279 2546280 2546281 2546282 2546283 2546284 2546285 2546286
## .. ..$ : int [1:4] 5067364 5067365 5067366 5067367
## .. ..$ : int [1:7] 3017684 3017685 3017686 3017687 3017688 3017689 3017690
## .. ..$ : int [1:3] 6097188 6097189 6097190
## .. ..$ : int [1:19] 525773 525774 525775 525776 525777 525778 525779 525780 525781 525782 ...
## .. ..$ : int [1:6] 3561398 3561399 3561400 3561401 3561402 3561403
## .. ..$ : int [1:2] 7242044 7242045
## .. ..$ : int 8480288
## .. ..$ : int 8480289
## .. ..$ : int 8480290
## .. ..$ : int [1:5] 4258220 4258221 4258222 4258223 4258224
## .. ..$ : int 8480291
## .. ..$ : int [1:4] 5067368 5067369 5067370 5067371
## .. ..$ : int 8480292
## .. ..$ : int 8480293
## .. ..$ : int [1:3] 6097191 6097192 6097193
## .. ..$ : int 8480294
## .. ..$ : int [1:2] 7242046 7242047
## .. ..$ : int [1:2] 7242048 7242049
## .. ..$ : int 8480295
## .. ..$ : int [1:2] 7242050 7242051
## .. ..$ : int [1:4] 5067372 5067373 5067374 5067375
## .. .. [list output truncated]
## .. ..@ ptype: int(0)
## ..- attr(*, ".drop")= logi TRUE

```

The join is ON t\_dat... so why are there NA's in the t\_dat??

```

## customer_id      t_dat      n      age      age_group
##           0      2823      0        0        2823

```