

Report of Exploratory Analysis for the Customer Service team at Profeco

```
In [14]: # Import libraries
import pandas

In [15]: # Import dataframe
df = pandas.read_csv(r'data\all_data.csv', dtype='unicode')
```

How many commercial chains are monitored, and therefore, included in this database?

```
In [16]: print((There are {} commercial chains monitored.'.format(df['cadenaComercial'].nunique()), end= "\n\n"))

There are 705 commercial chains monitored.
```

What are the top 10 monitored products by State?

```
In [17]: for s in df['estado'].unique():
    print('The top 10 products for {} are:'.format(s))
    print(df.loc[df['estado'] == s].groupby(['producto'])['producto'].count().sort_values(ascending=False).head(10))

The top 10 products for DISTRITO FEDERAL are:
REFRESCO      287463
FUD            207569
LECHE ULTRAPASTEURIZADA 175640
DETERGENTE P/ROPA 173452
YOGHURT        136720
CERVEZA        136686
MAYONESA       131103
CHILES EN LATA 130598
JABON DE TOCADOR 129889
SHAMPOO        125603

The top 10 products for MÉXICO are:
REFRESCO      194939
FUD            149141
DETERGENTE P/ROPA 132862
LECHE ULTRAPASTEURIZADA 116522
JABON DE TOCADOR 97330
YOGHURT        94852
MAYONESA       94286
CHILES EN LATA 92539
SHAMPOO        92307
CERVEZA        91747

The top 10 products for COAHUILA DE ZARAGOZA are:
FUD            28876
REFRESCO      27164
DETERGENTE P/ROPA 26706
SHAMPOO        21401
LECHE ULTRAPASTEURIZADA 21047
JABON DE TOCADOR 20283
PANTALLAS      20216
LAVADORAS      19801
CHILES EN LATA 19759
MAYONESA       19026

The top 10 products for GUANAJUATO are:
REFRESCO      49441
DETERGENTE P/ROPA 36618
VARIOS         35278
FUD            33363
LECHE ULTRAPASTEURIZADA 32633
LAVADORAS      31037
CHILES EN LATA 27190
JABON DE TOCADOR 26788
SHAMPOO        26772
MAYONESA       25205

The top 10 products for HIDALGO are:
REFRESCO      18137
LAVADORAS      16519
DETERGENTE P/ROPA 15946
LECHE ULTRAPASTEURIZADA 13537
COMPONENTES DE AUDIO 13171
PLANCHAS       12486
REFRIGERADORES 12378
SHAMPOO        12278
JABON DE TOCADOR 11848
CHILES EN LATA 11803

The top 10 products for JALISCO are:
REFRESCO      81718
LECHE ULTRAPASTEURIZADA 81149
DETERGENTE P/ROPA 73854
FUD            61680
CHILES EN LATA 55783
JABON DE TOCADOR 55196
SHAMPOO        54009
MAYONESA       53886
DESODORANTE    46335
TOALLA FEMENINA 46019

The top 10 products for MICHOACÁN DE OCAMPO are:
DETERGENTE P/ROPA 40144
REFRESCO      38016
LECHE ULTRAPASTEURIZADA 31175
CHILES EN LATA 25054
JABON DE TOCADOR 24587
MAYONESA       23988
LAVADORAS      23886
SHAMPOO        22475
FUD            21763
COMPONENTES DE AUDIO 20297

The top 10 products for NUEVO LEÓN are:
DETERGENTE P/ROPA 50307
REFRESCO      49592
LECHE ULTRAPASTEURIZADA 43803
CHILES EN LATA 40561
SHAMPOO        39827
JABON DE TOCADOR 38442
MAYONESA       37607
FUD            37200
TOALLA FEMENINA 32895
DESODORANTE    31875

The top 10 products for OAXACA are:
LECHE ULTRAPASTEURIZADA 18078
REFRESCO      14501
DETERGENTE P/ROPA 14396
FUD            13850
LAVADORAS      12774
CHILES EN LATA 11053
COMPONENTES DE AUDIO 11026
JABON DE TOCADOR 10564
SHAMPOO        10361
PANTALLAS      10325

The top 10 products for PUEBLA are:
REFRESCO      31597
LAVADORAS      27076
LECHE ULTRAPASTEURIZADA 24864
DETERGENTE P/ROPA 24078
FUD            20397
PLANCHAS       19547
JABON DE TOCADOR 19229
REFRIGERADORES 19150
PANTALLAS      19009
COMPONENTES DE AUDIO 18907

The top 10 products for QUERÉTARO are:
LECHE ULTRAPASTEURIZADA 27868
DETERGENTE P/ROPA 26700
REFRESCO      25494
LAVADORAS      20625
FUD            19574
JABON DE TOCADOR 19544
SHAMPOO        19417
CHILES EN LATA 18556
COMPONENTES DE AUDIO 16619
DESODORANTE    16269

The top 10 products for QUINTANA ROO are:
FUD            34846
REFRESCO      34367
LAVADORAS      32347
LECHE ULTRAPASTEURIZADA 31155
DETERGENTE P/ROPA 28691
COMPONENTES DE AUDIO 24836
PANTALLAS      24718
PLANCHAS       24526
SHAMPOO        24414
REFRIGERADORES 24096

The top 10 products for SINALOA are:
REFRESCO      33115
DETERGENTE P/ROPA 27177
SHAMPOO        22435
LAVADORAS      21207
COMPONENTES DE AUDIO 20827
JABON DE TOCADOR 20122
LECHE ULTRAPASTEURIZADA 18645
PANTALLAS      17650
MAYONESA       17138
REFRIGERADORES 16475

The top 10 products for SONORA are:
REFRESCO      34731
DETERGENTE P/ROPA 31997
SHAMPOO        24758
JABON DE TOCADOR 23459
CHILES EN LATA 21761
MAYONESA       20210
LAVADORAS      19217
DESODORANTE    18399
PANTALLAS      18304
GALLETAS       18239

The top 10 products for TABASCO are:
REFRESCO      28754
DETERGENTE P/ROPA 26431
LAVADORAS      26361
LECHE ULTRAPASTEURIZADA 25331
FUD            25004
VARIOS         21577
COMPONENTES DE AUDIO 21085
CHILES EN LATA 20949
JABON DE TOCADOR 19841
PLANCHAS       19808

The top 10 products for TLAXCALA are:
REFRESCO      43904
DETERGENTE P/ROPA 41398
LECHE ULTRAPASTEURIZADA 32169
JABON DE TOCADOR 30977
SHAMPOO        30235
CHILES EN LATA 25479
MAYONESA       24884
ACEITE         23236
TOALLA FEMENINA 22565
LECHE EN POLVO 22034

The top 10 products for VERACRUZ DE IGNACIO DE LA LLAVE are:
REFRESCO      14576
LECHE ULTRAPASTEURIZADA 12051
DETERGENTE P/ROPA 10345
FUD            9482
LAVADORAS      9231
SHAMPOO        8630
JABON DE TOCADOR 8405
DESODORANTE    8213
PANTALLAS      7814
CHILES EN LATA 7658

The top 10 products for YUCATÁN are:
LECHE ULTRAPASTEURIZADA 35991
DETERGENTE P/ROPA 33390
REFRESCO      33235
FUD            31885
LAVADORAS      27013
CHILES EN LATA 24472
MAYONESA       23629
SHAMPOO        23433
PLANCHAS       22816
LECHE EN POLVO 22696

The top 10 products for AGUASCALIENTES are:
FUD            12005
DETERGENTE P/ROPA 10188
LECHE ULTRAPASTEURIZADA 9824
SHAMPOO        9654
REFRESCO      9481
DESODORANTE    8859
JABON DE TOCADOR 8517
CHILES EN LATA 7946
YOGHURT        7401
MAYONESA       7173

The top 10 products for COLIMA are:
DETERGENTE P/ROPA 17193
LECHE ULTRAPASTEURIZADA 14425
REFRESCO      14383
FUD            13757
JABON DE TOCADOR 13298
LAVADORAS      13235
COMPONENTES DE AUDIO 13154
PANTALLAS      13092
CHILES EN LATA 12819
SHAMPOO        12678

The top 10 products for ZACATECAS are:
DETERGENTE P/ROPA 20884
LECHE ULTRAPASTEURIZADA 17309
LAVADORAS      16072
MAYONESA       15927
JABON DE TOCADOR 15926
FUD            15541
SHAMPOO        15012
CHILES EN LATA 14866
COMPONENTES DE AUDIO 14799
REFRESCO      13925

The top 10 products for nan are:
Series([], )

The top 10 products for DURANGO are:
DETERGENTE P/ROPA 9696
LAVADORAS      9644
FUD            8866
REFRESCO      8579
PANTALLAS      8329
SHAMPOO        8074
COMPONENTES DE AUDIO 7876
JABON DE TOCADOR 7778
PLANCHAS       7732
CHILES EN LATA 7425

The top 10 products for BAJA CALIFORNIA are:
REFRESCO      37243
DETERGENTE P/ROPA 23395
FUD            19967
SHAMPOO        19123
JABON DE TOCADOR 18348
CHILES EN LATA 16676
GALLETAS       15873
PANTALLAS      15703
CEREALES       15398
DESODORANTE    14748

The top 10 products for BAJA CALIFORNIA SUR are:
REFRESCO      27770
FUD            17776
PANTALLAS      17383
DETERGENTE P/ROPA 15510
LAVADORAS      13242
SHAMPOO        12844
JABON DE TOCADOR 12799
CHILES EN LATA 11537
PLANCHAS       11301
COMPONENTES DE AUDIO 11036

The top 10 products for CAMPECHE are:
FUD            12960
REFRESCO      11333
PANTALLAS      10449
LAVADORAS      9729
DETERGENTE P/ROPA 8307
LECHE ULTRAPASTEURIZADA 8027
REFRIGERADORES 7572
SHAMPOO        7560
JABON DE TOCADOR 7543
PLANCHAS       7139

The top 10 products for CHIAPAS are:
REFRESCO      14452
FUD            10417
DETERGENTE P/ROPA 8831
PANTALLAS      8395
JABON DE TOCADOR 7841
CHILES EN LATA 7707
SHAMPOO        7601
LECHE ULTRAPASTEURIZADA 7278
DESODORANTE    7038
CERVEZA        6980

The top 10 products for CHIHUAHUA are:
DETERGENTE P/ROPA 15982
FUD            14348
CHILES EN LATA 14189
SHAMPOO        14179
JABON DE TOCADOR 13749
REFRESCO      13554
PANTALLAS      13064
MAYONESA       12237
DESODORANTE    11741
LAVADORAS      11479

The top 10 products for GUERRERO are:
REFRESCO      8932
FUD            8846
DETERGENTE P/ROPA 6455
SHAMPOO        6283
LECHE ULTRAPASTEURIZADA 5837
LAVADORAS      5389
CERVEZA        5289
JABON DE TOCADOR 5169
CHILES EN LATA 5020

The top 10 products for MORELOS are:
REFRESCO      9632
FUD            6909
DETERGENTE P/ROPA 5829
LAVADORAS      5485
PANTALLAS      4989
SHAMPOO        4908
REFRIGERADORES 4647
LECHE ULTRAPASTEURIZADA 4539
DESODORANTE    4521
JABON DE TOCADOR 4500

The top 10 products for NAYARIT are:
REFRESCO      8003
PANTALLAS      7083
FUD            6544
LAVADORAS      6549
DETERGENTE P/ROPA 5514
COMPONENTES DE AUDIO 4999
SHAMPOO        4975
JABON DE TOCADOR 4680
REFRIGERADORES 4578
PLANCHAS       4509

The top 10 products for SAN LUIS POTOSÍ are:
FUD            10164
PANTALLAS      8662
LAVADORAS      7934
DETERGENTE P/ROPA 7380
LECHE ULTRAPASTEURIZADA 7095
REFRESCO      6940
SHAMPOO        6738
JABON DE TOCADOR 6389
ESTUFAS        6266

The top 10 products for TAMAULIPAS are:
REFRESCO      22457
FUD            21113
DETERGENTE P/ROPA 20291
SHAMPOO        19141
JABON DE TOCADOR 17167
CHILES EN LATA 16008
DESODORANTE    15994
CEREALES       15398
PANTALLAS      15180
MAYONESA       15014

The top 10 products for estado are:
producto      20
```

Which is the commercial chain with the highest number of monitored products?

```
In [18]: cch = df.groupby(['cadenaComercial'])['producto'].nunique().sort_values(ascending=False).head(1)
print('The commercial chain with the highest number of monitored products is {} with {} brands.'.format(cch.index.values[0], cch.values[0]))

The commercial chain with the highest number of monitored products is SORIANA with 1059 products
```

Use the data to find an interesting fact.

```
In [19]: ccmb = df.groupby(['cadenaComercial'])['marca'].nunique().sort_values(ascending=False).head(1)
print('The commercial chain offering the most brands is {} with {} brands.'.format(ccmb.index.values[0], ccmb.values[0]))

bmp = df.groupby(['marca'])['producto'].nunique().sort_values(ascending=False).head(1)
print('The brand with most products is {} with {} products.'.format(bmp.index.values[0], bmp.values[0]))

print('We are monitoring {} brands.'.format(df['marca'].nunique()))
```

The commercial chain offering the most brands is WAL-MART with 1338 brands.
The brand with most products is S/M with 565 products.
We are monitoring 2079 brands.

What are the lessons learned from this exercise?

The main lesson of this project was to use the Pandas library. I had not worked with Pandas before, at first it was a little complex to understand the functions of this library, but once understood, it became easier and easier to use this library.

So I can say that I was pleasantly surprised by the knowledge of Pandas, and surely it would have made my life easier in the past processing large volumes of data.

Can you identify other ways to approach this problem? Explain.

In a less efficient way, I would have used functions, pure Python properties and some regular expressions. Specifically, commands such as "open" and "read" would be used.

Another possible approach would have been to insert all the data into a MySQL or MSSQL type database and from there perform different database queries until the expected results were obtained.

Needless to say, using Pandas is the best option by far.