Product Analysis Sample

October 3, 2021

[ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
plt.rcParams["figure.figsize"] = (20,10)
```

1 Data analysis

The Customer Service team at Profeco (Mexican Consumer Protection Agency) wants to analyze the monitored products in Mexico.

[ ]:
```
chunksize = 10 ** 6
#filename="data/profeco/all_data.csv"
filename= "data/sample.csv"
```

[ ]:
```
data = pd.read_csv(filename, chunksize=chunksize, iterator=True, header=0,␣
 ↪low_memory=False)
df = pd.concat(data, ignore_index=True)
```

1.1 Exploratory Analysis

[ ]:
```
# Head of data to review attributes and types.
df.head()
```

[ ]:
```
                producto                                  presentacion  \
0  CUADERNO FORMA ITALIANA          96 HOJAS PASTA DURA. CUADRICULA CHICA
1                CRAYONES              CAJA 12 CERAS. JUMBO. C.B. 201423
2                CRAYONES      CAJA 12 CERAS. TAMANO REGULAR C.B. 201034
3       COLORES DE MADERA  CAJA 12 PIEZAS LARGO. TRIANGULAR. C.B. 640646
4             COLOR LARGO        CAJA 36 PIEZAS. CON SACAPUNTAS. 68-4036

      marca         categoria          catalogo  precio  \
0  ESTRELLA  MATERIAL ESCOLAR  UTILES ESCOLARES    25.9
1   CRAYOLA  MATERIAL ESCOLAR  UTILES ESCOLARES    27.5
2   CRAYOLA  MATERIAL ESCOLAR  UTILES ESCOLARES    13.9
3  PINCELIN  MATERIAL ESCOLAR  UTILES ESCOLARES    46.9
4   CRAYOLA  MATERIAL ESCOLAR  UTILES ESCOLARES   115.0

             fechaRegistro     cadenaComercial        giro  \
```

```
0  2011-05-18 00:00:00.000  ABASTECEDORA LUMEN  PAPELERIAS
1  2011-05-18 00:00:00.000  ABASTECEDORA LUMEN  PAPELERIAS
2  2011-05-18 00:00:00.000  ABASTECEDORA LUMEN  PAPELERIAS
3  2011-05-18 00:00:00.000  ABASTECEDORA LUMEN  PAPELERIAS
4  2011-05-18 00:00:00.000  ABASTECEDORA LUMEN  PAPELERIAS

                          nombreComercial  \
0  ABASTECEDORA LUMEN SUCURSAL VILLA COAPA
1  ABASTECEDORA LUMEN SUCURSAL VILLA COAPA
2  ABASTECEDORA LUMEN SUCURSAL VILLA COAPA
3  ABASTECEDORA LUMEN SUCURSAL VILLA COAPA
4  ABASTECEDORA LUMEN SUCURSAL VILLA COAPA

                               direccion            estado  \
0  CANNES No. 6 ESQ. CANAL DE MIRAMONTES  DISTRITO FEDERAL
1  CANNES No. 6 ESQ. CANAL DE MIRAMONTES  DISTRITO FEDERAL
2  CANNES No. 6 ESQ. CANAL DE MIRAMONTES  DISTRITO FEDERAL
3  CANNES No. 6 ESQ. CANAL DE MIRAMONTES  DISTRITO FEDERAL
4  CANNES No. 6 ESQ. CANAL DE MIRAMONTES  DISTRITO FEDERAL

                                  municipio   latitud   longitud
0  TLALPAN                                  19.29699 -99.125417
1  TLALPAN                                  19.29699 -99.125417
2  TLALPAN                                  19.29699 -99.125417
3  TLALPAN                                  19.29699 -99.125417
4  TLALPAN                                  19.29699 -99.125417
```

[ ]:
```
# Tail of data to review attributes and types.
df.tail()
```

[ ]:
```
                              producto                 presentacion  \
1994                          REFRESCO  BOTELLA 1 LT. NO RETORNABLE
1995                 GELATINA EN POLVO      CAJA 84 GR. AGUA. LIMON
1996    POLVO P/BEBIDAS SABOR CHOCOLATE                 LATA 400 GR.
1997  POLVO PARA BEBIDAS SABOR CHOCOLATE                BOLSA 400 GR.
1998    POLVO P/BEBIDAS SABOR CHOCOLATE                 LATA 400 GR.

                   marca              categoria  catalogo  precio  \
1994           COCA COLA     REFRESCOS ENVASADOS   BASICOS    11.0
1995              PRONTO  CHOCOLATES Y GOLOSINAS   BASICOS     6.5
1996             NESQUIK  CHOCOLATES Y GOLOSINAS  MERCADOS    29.0
1997  MARCA INSTITUCIONAL  CHOCOLATES Y GOLOSINAS   BASICOS    20.0
1998           CHOCO MILK  CHOCOLATES Y GOLOSINAS  MERCADOS    32.0

                fechaRegistro cadenaComercial                    giro  \
1994  2011-01-10 00:00:00.000        WAL-MART  TIENDA DE AUTOSERVICIO
1995  2011-01-10 00:00:00.000        WAL-MART  TIENDA DE AUTOSERVICIO
```

```
1996  2011-01-10 00:00:00.000        WAL-MART  TIENDA DE AUTOSERVICIO
1997  2011-01-10 00:00:00.000        WAL-MART  TIENDA DE AUTOSERVICIO
1998  2011-01-10 00:00:00.000        WAL-MART  TIENDA DE AUTOSERVICIO

                 nombreComercial  \
1994  WAL MART SUCURSAL ECHEGARAY
1995  WAL MART SUCURSAL ECHEGARAY
1996  WAL MART SUCURSAL ECHEGARAY
1997  WAL MART SUCURSAL ECHEGARAY
1998  WAL MART SUCURSAL ECHEGARAY

                                          direccion  estado  \
1994  EMILIANO ZAPATA NO. 7     ENTRE REFORMA Y CUAU…  MÉXICO
1995  EMILIANO ZAPATA NO. 7     ENTRE REFORMA Y CUAU…  MÉXICO
1996  EMILIANO ZAPATA NO. 7     ENTRE REFORMA Y CUAU…  MÉXICO
1997  EMILIANO ZAPATA NO. 7     ENTRE REFORMA Y CUAU…  MÉXICO
1998  EMILIANO ZAPATA NO. 7     ENTRE REFORMA Y CUAU…  MÉXICO

                                    municipio    latitud  longitud
1994  NAUCALPAN                                19.483909 -99.23318
1995  NAUCALPAN                                19.483909 -99.23318
1996  NAUCALPAN                                19.483909 -99.23318
1997  NAUCALPAN                                19.483909 -99.23318
1998  NAUCALPAN                                19.483909 -99.23318
```

[ ]:
```
# Count of not null rows by attribute
df.count()
```

[ ]:
```
producto           1999
presentacion       1999
marca              1999
categoria          1999
catalogo           1999
precio             1999
fechaRegistro      1999
cadenaComercial    1999
giro               1999
nombreComercial    1999
direccion          1999
estado             1999
municipio          1999
latitud            1229
longitud           1229
dtype: int64
```

[ ]:
```
#Basic information about data. In small datasets it shows count, mean, std,␣
 ↪min, etc of numeric data.
```

```
df.describe()
```

[ ]:

	precio	latitud	longitud
count	1999.000000	1229.000000	1229.000000
mean	293.945953	19.399105	-99.185158
std	1270.865893	0.048795	0.032065
min	1.200000	19.296990	-99.233180
25%	13.500000	19.366903	-99.181344
50%	26.100000	19.366903	-99.181344
75%	56.100000	19.400879	-99.181344
max	13497.000000	19.483909	-99.125417

1.2 Questions

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

[ ]:
```
df['cadenaComercial'].nunique()
```

[ ]:
```
4
```

In total we have 4 different commercial chains that are monitored by the system.

2.What are the top 10 monitored products by State? Due to the fact that this question is a bit ambiguous this question will be answered in 3 different ways.

2.1 Top 10 of product count by state

[ ]:
```
df_group= df.groupby(by=["estado","producto"]).agg({"producto":"count"}).
↪rename(columns={"producto":"count_producto"})
df_group
```

[ ]:

		count_producto
estado	producto	
DISTRITO FEDERAL	ACEITE	11
	ACEITE DE OLIVA	6
	ACEITUNA	3
	ACEITUNA. GORDAL	1
	ACEITUNA. RELLENA CON PIMIENTO	1
…	…	…
MÉXICO	ZESTRIL	1
	ZOVIRAX	2
	ZUCARITAS	2
	ZWAN	8
	ZWAN. PREMIUM	4

[546 rows x 1 columns]

[ ]:

```
df_count=df_group.assign(rn_count= df_group.
 ↪sort_values(by=["count_producto"],ascending=[False]).groupby(by=["estado"]).
 ↪cumcount() + 1)
df_count
```

[ ]:

		count_producto	rn_count
estado	producto		
DISTRITO FEDERAL	ACEITE	11	19
	ACEITE DE OLIVA	6	45
	ACEITUNA	3	104
	ACEITUNA. GORDAL	1	183
	ACEITUNA. RELLENA CON PIMIENTO	1	182
…		…	…
MÉXICO	ZESTRIL	1	222
	ZOVIRAX	2	146
	ZUCARITAS	2	143
	ZWAN	8	27
	ZWAN. PREMIUM	4	95

[546 rows x 2 columns]

This is the list of top 10 productos count by state

[ ]:
```
df_top_count= df_count.sort_values(['estado','rn_count'],␣
 ↪ascending=[True,True]).where(df_count["rn_count"]<=10).dropna()
df_top_count
```

[ ]:

		count_producto	rn_count
estado	producto		
DISTRITO FEDERAL	DETERGENTE P/ROPA	25.0	1.0
	LECHE ULTRAPASTEURIZADA	23.0	2.0
	REFRESCO	22.0	3.0
	CHILES EN LATA	21.0	4.0
	MAYONESA	20.0	5.0
	SHAMPOO	18.0	6.0
	TELEVISORES	17.0	7.0
	JABON DE TOCADOR	16.0	8.0
	GALLETAS DULCES	16.0	9.0
	PAN DE CAJA	14.0	10.0
MÉXICO	REFRESCO	26.0	1.0
	DETERGENTE P/ROPA	24.0	2.0
	SHAMPOO	19.0	3.0
	CHILES EN LATA	18.0	4.0
	MAYONESA	17.0	5.0
	TOALLA FEMENINA	17.0	6.0
	LECHE ULTRAPASTEURIZADA	17.0	7.0
	TELEVISORES	15.0	8.0
	DESODORANTE	15.0	9.0

```
                    YOGHURT                                 15.0       10.0
```

[ ]:
```
#Testing the rows count - 14
df.where((df["estado"]=="DISTRITO FEDERAL") & (df["producto"]=="PAN DE CAJA")).
 ↪dropna()
```

[ ]:
```
         producto                                        presentacion  \
779   PAN DE CAJA         PAQUETE MEDIANO 480 GR. INTEGRAL. REBANADO
780   PAN DE CAJA          PAQUETE GRANDE 680 GR. INTEGRAL. REBANADO
781   PAN DE CAJA            PAQUETE MEDIANO 460 GR. REBANADO BLANCO
955   PAN DE CAJA         PAQUETE MEDIANO 475 GR. INTEGRAL. REBANADO
956   PAN DE CAJA  PAQUETE GRANDE 680 GR. REBANADO BLANCO. CON FIBRA
1279  PAN DE CAJA          PAQUETE GRANDE 680 GR. INTEGRAL. REBANADO
1280  PAN DE CAJA                   PAQUETE 210 GR. TOSTADO. CLÁSICO
1281  PAN DE CAJA             PAQUETE GRANDE 680 GR. REBANADO BLANCO
1517  PAN DE CAJA         PAQUETE MEDIANO 480 GR. INTEGRAL. REBANADO
1518  PAN DE CAJA            PAQUETE MEDIANO 460 GR. REBANADO BLANCO
1556  PAN DE CAJA  PAQUETE GRANDE 680 GR. REBANADO BLANCO. CON FIBRA
1567  PAN DE CAJA         PAQUETE MEDIANO 475 GR. INTEGRAL. REBANADO
1655  PAN DE CAJA                   PAQUETE 210 GR. TOSTADO. CLÁSICO
1656  PAN DE CAJA             PAQUETE GRANDE 680 GR. REBANADO BLANCO

              marca categoria catalogo  precio            fechaRegistro  \
779           BIMBO       PAN  BASICOS   22.00  2011-01-10 00:00:00.000
780   WONDER. 100%       PAN  BASICOS   27.00  2011-01-10 00:00:00.000
781           BIMBO       PAN  BASICOS   20.00  2011-01-10 00:00:00.000
955   WONDER. 100%       PAN  BASICOS   22.50  2011-01-10 00:00:00.000
956         WONDER       PAN  BASICOS   24.00  2011-01-10 00:00:00.000
1279          BIMBO       PAN  BASICOS   26.00  2011-01-10 00:00:00.000
1280          BIMBO       PAN  BASICOS   14.00  2011-01-10 00:00:00.000
1281          BIMBO       PAN  BASICOS   24.00  2011-01-10 00:00:00.000
1517          BIMBO       PAN  BASICOS   16.00  2011-01-10 00:00:00.000
1518          BIMBO       PAN  BASICOS   15.55  2011-01-10 00:00:00.000
1556        WONDER       PAN  BASICOS   21.95  2011-01-10 00:00:00.000
1567  WONDER. 100%       PAN  BASICOS   18.65  2011-01-10 00:00:00.000
1655          BIMBO       PAN  BASICOS   12.15  2011-01-10 00:00:00.000
1656          BIMBO       PAN  BASICOS   19.30  2011-01-10 00:00:00.000

         cadenaComercial                    giro  \
779   COMERCIAL MEXICANA  TIENDA DE AUTOSERVICIO
780   COMERCIAL MEXICANA  TIENDA DE AUTOSERVICIO
781   COMERCIAL MEXICANA  TIENDA DE AUTOSERVICIO
955   COMERCIAL MEXICANA  TIENDA DE AUTOSERVICIO
956   COMERCIAL MEXICANA  TIENDA DE AUTOSERVICIO
1279  COMERCIAL MEXICANA  TIENDA DE AUTOSERVICIO
1280  COMERCIAL MEXICANA  TIENDA DE AUTOSERVICIO
1281  COMERCIAL MEXICANA  TIENDA DE AUTOSERVICIO
```

```
1517        I.S.S.S.T.E.  TIENDA DE AUTOSERVICIO
1518        I.S.S.S.T.E.  TIENDA DE AUTOSERVICIO
1556        I.S.S.S.T.E.  TIENDA DE AUTOSERVICIO
1567        I.S.S.S.T.E.  TIENDA DE AUTOSERVICIO
1655        I.S.S.S.T.E.  TIENDA DE AUTOSERVICIO
1656        I.S.S.S.T.E.  TIENDA DE AUTOSERVICIO

                              nombreComercial  \
779   COMERCIAL MEXICANA SUCURSAL INSURGENTES
780   COMERCIAL MEXICANA SUCURSAL INSURGENTES
781   COMERCIAL MEXICANA SUCURSAL INSURGENTES
955   COMERCIAL MEXICANA SUCURSAL INSURGENTES
956   COMERCIAL MEXICANA SUCURSAL INSURGENTES
1279  COMERCIAL MEXICANA SUCURSAL INSURGENTES
1280  COMERCIAL MEXICANA SUCURSAL INSURGENTES
1281  COMERCIAL MEXICANA SUCURSAL INSURGENTES
1517                 I.S.S.S.T.E. SUCURSAL 7
1518                 I.S.S.S.T.E. SUCURSAL 7
1556                 I.S.S.S.T.E. SUCURSAL 7
1567                 I.S.S.S.T.E. SUCURSAL 7
1655                 I.S.S.S.T.E. SUCURSAL 7
1656                 I.S.S.S.T.E. SUCURSAL 7

                                  direccion            estado  \
779   PERPETUA NO. 35 ESQ. INSURGENTES SUR  DISTRITO FEDERAL
780   PERPETUA NO. 35 ESQ. INSURGENTES SUR  DISTRITO FEDERAL
781   PERPETUA NO. 35 ESQ. INSURGENTES SUR  DISTRITO FEDERAL
955   PERPETUA NO. 35 ESQ. INSURGENTES SUR  DISTRITO FEDERAL
956   PERPETUA NO. 35 ESQ. INSURGENTES SUR  DISTRITO FEDERAL
1279  PERPETUA NO. 35 ESQ. INSURGENTES SUR  DISTRITO FEDERAL
1280  PERPETUA NO. 35 ESQ. INSURGENTES SUR  DISTRITO FEDERAL
1281  PERPETUA NO. 35 ESQ. INSURGENTES SUR  DISTRITO FEDERAL
1517              CORUÑA S/N ESQ. SUR 65-A  DISTRITO FEDERAL
1518              CORUÑA S/N ESQ. SUR 65-A  DISTRITO FEDERAL
1556              CORUÑA S/N ESQ. SUR 65-A  DISTRITO FEDERAL
1567              CORUÑA S/N ESQ. SUR 65-A  DISTRITO FEDERAL
1655              CORUÑA S/N ESQ. SUR 65-A  DISTRITO FEDERAL
1656              CORUÑA S/N ESQ. SUR 65-A  DISTRITO FEDERAL

                                     municipio    latitud   longitud
779   BENITO JUAREZ                            19.366903 -99.181344
780   BENITO JUAREZ                            19.366903 -99.181344
781   BENITO JUAREZ                            19.366903 -99.181344
955   BENITO JUAREZ                            19.366903 -99.181344
956   BENITO JUAREZ                            19.366903 -99.181344
1279  BENITO JUAREZ                            19.366903 -99.181344
1280  BENITO JUAREZ                            19.366903 -99.181344
```

```
1281  BENITO JUAREZ                          19.366903 -99.181344
1517  IZTACALCO                              19.400879 -99.131672
1518  IZTACALCO                              19.400879 -99.131672
1556  IZTACALCO                              19.400879 -99.131672
1567  IZTACALCO                              19.400879 -99.131672
1655  IZTACALCO                              19.400879 -99.131672
1656  IZTACALCO                              19.400879 -99.131672
```

2.2 Top 10 of products present in more states

[ ]:
```
df_group_state= df.groupby(by=["producto"]).agg({"estado":"nunique"}).
↪rename(columns={"estado":"count_estado"})
df_group_state
```

[ ]:
```
                      count_estado
producto
ACEITE                           2
ACEITE DE OLIVA                  2
ACEITUNA                         2
ACEITUNA. GORDAL                 2
ACEITUNA. MANZANILLA             1
…                              …
ZESTRIL                          1
ZOVIRAX                          1
ZUCARITAS                        2
ZWAN                             2
ZWAN. PREMIUM                    2

[330 rows x 1 columns]
```

[ ]:
```
df_rank_estado = df_group_state.sort_values(['count_estado'],␣
↪ascending=[False]).assign(rn_count_estado= df_group_state.reset_index().
↪index+1)
df_rank_estado
```

[ ]:
```
                          count_estado  rn_count_estado
producto
ACEITE                               2                1
LECHE ULTRAPASTEURIZADA              2                2
LOS VOLCANES                         2                3
LIMPIADOR LIQUIDO P/PISO             2                4
LIMON                                2                5
…                                  …                …
CHAMPIÑON                            1              326
LINCOCIN                             1              327
LIMA                                 1              328
CINTA ADHESIVA                       1              329
NORVAS                               1              330
```

[330 rows x 2 columns]

[ ]:
```
df_top_count_estado= df_rank_estado.sort_values(['producto','rn_count_estado'],␣
↪ascending=[True,True]).where(df_rank_estado["rn_count_estado"]<=10).dropna()
df_top_count_estado
```

[ ]:

producto	count_estado	rn_count_estado
ACEITE	2.0	1.0
JITOMATE	2.0	10.0
LECHE PASTEURIZADA	2.0	9.0
LECHE ULTRAPASTEURIZADA	2.0	2.0
LECHUGA	2.0	8.0
LENTEJA	2.0	7.0
LICUADORAS	2.0	6.0
LIMON	2.0	5.0
LIMPIADOR LIQUIDO P/PISO	2.0	4.0
LOS VOLCANES	2.0	3.0

[ ]:
```
#Testing the rows count
df.where(df["producto"]=="ACEITE")["estado"].dropna().unique()
```

[ ]:
```
array(['MÉXICO', 'DISTRITO FEDERAL'], dtype=object)
```

2.3 Top 10 products by avg price by state

[ ]:
```
df["precio"]= pd.to_numeric(df['precio'], errors='coerce')
```

[ ]:
```
df_avg_precio= df.groupby(by=["estado","producto"],dropna=True)["precio"].
↪mean().reset_index(drop=False).dropna()
df_avg_precio
```

[ ]:

	estado	producto	precio
0	DISTRITO FEDERAL	ACEITE	27.746364
1	DISTRITO FEDERAL	ACEITE DE OLIVA	87.418333
2	DISTRITO FEDERAL	ACEITUNA	26.223333
3	DISTRITO FEDERAL	ACEITUNA. GORDAL	67.500000
4	DISTRITO FEDERAL	ACEITUNA. RELLENA CON PIMIENTO	67.500000
..	…	…	…
541	MÉXICO	ZESTRIL	246.120000
542	MÉXICO	ZOVIRAX	172.735000
543	MÉXICO	ZUCARITAS	31.955000
544	MÉXICO	ZWAN	59.512500
545	MÉXICO	ZWAN. PREMIUM	76.480000

[546 rows x 3 columns]

```
[ ]: df_rank= df_avg_precio.assign(rn=df_avg_precio.sort_values(['precio'],␣
      ↪ascending=[False]).groupby(by=["estado"]).cumcount() + 1)
     df_top= df_rank.sort_values(['estado','rn'], ascending=[True,True]).
      ↪where(df_rank["rn"]<=10).dropna()
     df_top
```

```
[ ]:              estado                                 producto       precio  \
     222  DISTRITO FEDERAL                              TELEVISORES  7867.941176
     236  DISTRITO FEDERAL                             VIDEOCAMARAS  4664.333333
     69   DISTRITO FEDERAL                  COMPONENTES DE AUDIO     3769.428571
     217  DISTRITO FEDERAL  SISTEMAS DE TEATRO EN CASA HOME THEATER  2928.000000
     199  DISTRITO FEDERAL                       RADIOGRABADORAS     2311.333333
     38   DISTRITO FEDERAL                     CAMARAS DIGITALES     2283.714286
     120  DISTRITO FEDERAL                   HORNO DE MICROONDAS     1548.000000
     88   DISTRITO FEDERAL                           DVD / BLU RAY      698.000000
     170  DISTRITO FEDERAL                OLLA DE PRESION EXPRESS      674.200000
     145  DISTRITO FEDERAL                             LICUADORAS      596.428571
     520            MÉXICO                              TELEVISORES  8804.466667
     492            MÉXICO                          REFRIGERADORES   6295.000000
     417            MÉXICO                               LAVADORAS   5529.400000
     367            MÉXICO                                 ESTUFAS   4497.666667
     535            MÉXICO                             VIDEOCAMARAS  4321.000000
     511            MÉXICO  SISTEMAS DE TEATRO EN CASA HOME THEATER  3148.500000
     344            MÉXICO                  COMPONENTES DE AUDIO     2931.500000
     302            MÉXICO                     CAMARAS DIGITALES     2098.000000
     362            MÉXICO                         DVD / BLU RAY     1751.714286
     489            MÉXICO                       RADIOGRABADORAS     1458.000000

           rn
     222  1.0
     236  2.0
     69   3.0
     217  4.0
     199  5.0
     38   6.0
     120  7.0
     88   8.0
     170  9.0
     145 10.0
     520  1.0
     492  2.0
     417  3.0
     367  4.0
     535  5.0
     511  6.0
     344  7.0
     302  8.0
```

```
362   9.0
489  10.0
```

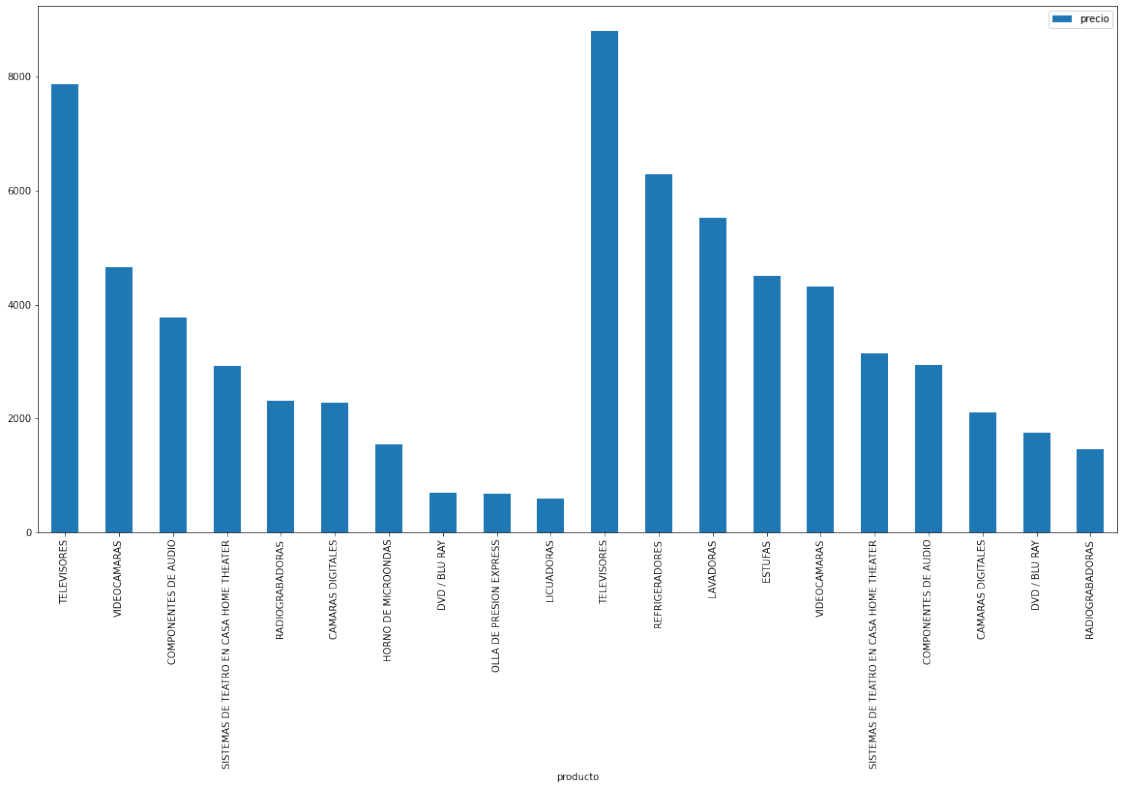
[ ]:
```
#Test some states
df_estado= df_top.where((df_top["estado"]=='DISTRITO FEDERAL')|(df_top["estado"]=='MÉXICO')).dropna()
df_estado
```

[ ]:
```
              estado                                 producto       precio  \
222  DISTRITO FEDERAL                              TELEVISORES  7867.941176
236  DISTRITO FEDERAL                             VIDEOCAMARAS  4664.333333
69   DISTRITO FEDERAL                    COMPONENTES DE AUDIO  3769.428571
217  DISTRITO FEDERAL  SISTEMAS DE TEATRO EN CASA HOME THEATER  2928.000000
199  DISTRITO FEDERAL                          RADIOGRABADORAS  2311.333333
38   DISTRITO FEDERAL                        CAMARAS DIGITALES  2283.714286
120  DISTRITO FEDERAL                     HORNO DE MICROONDAS  1548.000000
88   DISTRITO FEDERAL                             DVD / BLU RAY   698.000000
170  DISTRITO FEDERAL                OLLA DE PRESION EXPRESS   674.200000
145  DISTRITO FEDERAL                              LICUADORAS   596.428571
520           MÉXICO                              TELEVISORES  8804.466667
492           MÉXICO                            REFRIGERADORES  6295.000000
417           MÉXICO                                 LAVADORAS  5529.400000
367           MÉXICO                                   ESTUFAS  4497.666667
535           MÉXICO                             VIDEOCAMARAS  4321.000000
511           MÉXICO  SISTEMAS DE TEATRO EN CASA HOME THEATER  3148.500000
344           MÉXICO                    COMPONENTES DE AUDIO  2931.500000
302           MÉXICO                       CAMARAS DIGITALES  2098.000000
362           MÉXICO                           DVD / BLU RAY  1751.714286
489           MÉXICO                          RADIOGRABADORAS  1458.000000

       rn
222   1.0
236   2.0
69    3.0
217   4.0
199   5.0
38    6.0
120   7.0
88    8.0
170   9.0
145  10.0
520   1.0
492   2.0
417   3.0
367   4.0
535   5.0
511   6.0
```

```
344   7.0
302   8.0
362   9.0
489  10.0
```

[ ]:
```
ax = df_estado.plot.bar(x='producto', y='precio',rot=90)
```



[ ]:
```
df2 = df_estado.set_index(['estado','producto'])
df2
```

[ ]:

estado	producto	precio	rn
DISTRITO FEDERAL	TELEVISORES	7867.941176	1.0
	VIDEOCAMARAS	4664.333333	2.0
	COMPONENTES DE AUDIO	3769.428571	3.0
	SISTEMAS DE TEATRO EN CASA HOME THEATER	2928.000000	4.0
	RADIOGRABADORAS	2311.333333	5.0
	CAMARAS DIGITALES	2283.714286	6.0
	HORNO DE MICROONDAS	1548.000000	7.0
	DVD / BLU RAY	698.000000	8.0
	OLLA DE PRESION EXPRESS	674.200000	9.0
	LICUADORAS	596.428571	10.0

```
MÉXICO          TELEVISORES                              8804.466667   1.0
                REFRIGERADORES                           6295.000000   2.0
                LAVADORAS                                5529.400000   3.0
                ESTUFAS                                  4497.666667   4.0
                VIDEOCAMARAS                             4321.000000   5.0
                SISTEMAS DE TEATRO EN CASA HOME THEATER  3148.500000   6.0
                COMPONENTES DE AUDIO                     2931.500000   7.0
                CAMARAS DIGITALES                        2098.000000   8.0
                DVD / BLU RAY                            1751.714286   9.0
                RADIOGRABADORAS                          1458.000000  10.0
```

[ ]:
```
df2.index.levels
```

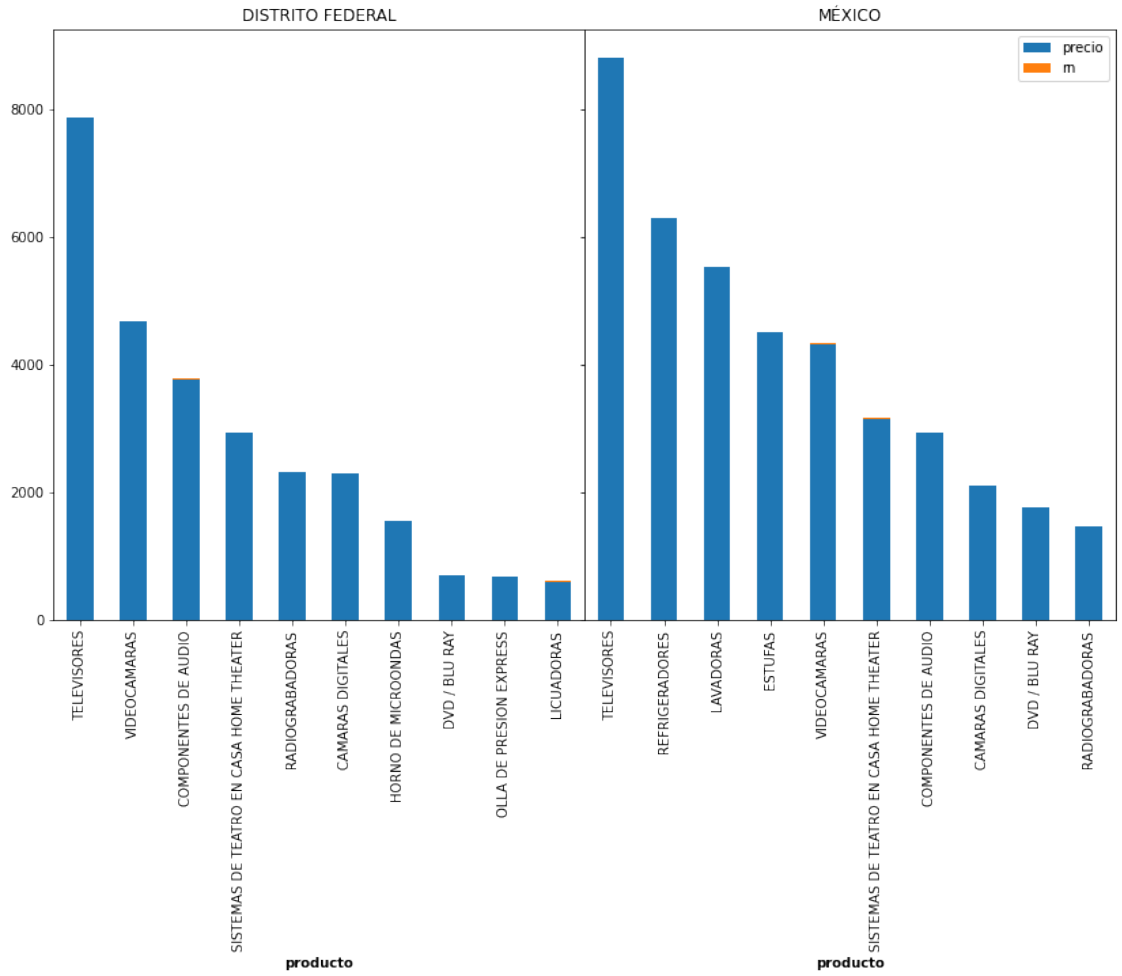
[ ]:
```
FrozenList([['DISTRITO FEDERAL', 'MÉXICO'], ['CAMARAS DIGITALES', 'COMPONENTES
DE AUDIO', 'DVD / BLU RAY', 'ESTUFAS', 'HORNO DE MICROONDAS', 'LAVADORAS',
'LICUADORAS', 'OLLA DE PRESION EXPRESS', 'RADIOGRABADORAS', 'REFRIGERADORES',
'SISTEMAS DE TEATRO EN CASA HOME THEATER', 'TELEVISORES', 'VIDEOCAMARAS']])
```

[ ]:
```
def plot_function(x, ax):
    ax= graph[x]
    ax.set_xlabel(x, weight='bold')
    ax.set_title(x)
    return df2.xs(x).plot(kind='bar', stacked='True', ax=ax, legend=False)

n_subplots = len(df2.index.levels[0])
fig, axes = plt.subplots(nrows=1, ncols=n_subplots, sharey=True, figsize=(14,␣
 ↪8))  # width, height

graph = dict(zip(df2.index.levels[0], axes))
plots = list(map(lambda x: plot_function(x, graph[x]), graph))
ax.tick_params(axis='both', which='both', length=0)
fig.subplots_adjust(wspace=0)

plt.legend()
plt.show()
```



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

[ ]:
```
df_highest_commercial_chain= df.groupby("cadenaComercial",as_index=False).
 ↪agg({"producto":"nunique"}).rename(columns={'producto':'cuenta_productos'})
df_highest_commercial_chain.sort_values(['cuenta_productos'],␣
 ↪ascending=[False]).head(1)
```

[ ]:
```
      cadenaComercial  cuenta_productos
1  COMERCIAL MEXICANA               310
```

The commercial chain with the highest number of monitored produts (unique) is COMERCIAL MEXICANA.

4. Use the data to find an interesting fact: - How many products types of products are sold nationwide (32 states) - In large dataset

[ ]:
```
df_rank_estado.where(df_rank_estado["count_estado"]==32).count()
```

```
[ ]: count_estado       0
     rn_count_estado    0
     dtype: int64
```

A total of 0 products are sold throughout the country.

5. What are the lessons learned from this exercise? There are undoubtedly many lessons learned. For my part it is always good to practice with the Pandas library, it is very good and widely used in the data world and especially in data science. Maybe not quite to my taste because I don't consider it superior to SQL but it is an excellent tool. With this process I was able to face the challenge of handling a high volume of data in a local computer, data management in memory, data cleansing, information adjustment. I also realized how important it is to define the questions to be solved and to avoid ambiguity. Another lesson is the importance of knowing the content and type of data. In short, excellent exercise.

6. Can you identify other ways to approach this problem? Explain. Being a problem that can escalate greatly with the amount of data and with possible new business needs such as opening the market to other countries. The problem can be addressed using other tools such as dash, which is a distributed system designed to compensate for this shortcoming in Pandas. It could also be a very important tool as an alternative Spark that works on memory in distributed systems, much better than pandas to handle large volumes of data and has something that for me is a super winning point. SparkSQL allows you to use this language on the data making the query process much more enjoyable.

Another interesting point that I would like to contribute is that although for ease of use I use Jupyter Notebook, this process should be done in a system that allows to be automated and divided into processes and easily ported to other environments and pipelines.