```
In [2]:  import pandas as pd
```

```
In [3]:  import numpy
```

```
In [4]:  from pathlib import Path
```

```
In [5]:  %matplotlib inline
```

```
In [6]:  csv_file = Path("tb_product_202410181030.csv")
```

```
In [7]:  df = pd.read_csv(csv_file,
                header=0,
                index_col= ["id"],
                low_memory = False,
                usecols=["id","id_trademark","id_measurement_unit","id_record_status","id_c
                sep = ',')
```

```
In [8]:  print (df)
```

```
        id_trademark  id_measurement_unit  id_record_status  \
id
678876        8277.0                  NaN                 1
458570        2947.0               5382.0                 1
176386         292.0                683.0                 1
458571        3405.0               5382.0                 1
589884        7336.0                  NaN                 1
...              ...                  ...               ...
604469        7341.0                  NaN                 1
579329           NaN                  NaN                 1
172391         290.0                683.0                 1
656432        8153.0                  NaN                 1
391878           NaN               2202.0                 1

        id_company_variation  id_product_category     code  \
id
678876                   893                38885    10487
458570                   572                23337     1801
176386                    97                 6082   162278
458571                   572                23394     1802
589884                   846                42386    18310
...                      ...                  ...      ...
604469                   846                36887    31592
579329                   799                    0     4474
172391                    97                 1390   158283
656432                   900                41732     2956
391878                   246                 7981      770

                                             name registration_date  \
id
678876                          CAB ALTA 121 121         2020-04-09
458570  Rodamientos Concha Biela A 010 Jeep Liberty 20...  2018-09-14
176386  Steering Tie Rod End Front Right Outer Mevotec...  2017-10-24
458571                 Guardapolvo Dodge Caliber Rh Sp     2018-09-14
589884                         STOP CORSA 3P DER           2020-01-17
...                                           ...                ...
604469                    VARILLA MED ACEITE RODEOV6        2020-01-17
579329  filtro aire dodge neon 96-99 lebaron terrano 3660  2019-12-18
172391  A/C Accumulator with Hose Assembly-w/Hose Asse...  2017-10-24
656432                            F.A/A GOLF 2006           2020-04-09
391878        Transmision Lav. Asiatica Jn-102-a 35-30     2018-06-26

                                      description        model  \
id
678876                          CAB ALTA 121 121    121, 1997
458570  RODAMIENTOS CONCHA BIELA 3.7\n\nTIPO DE CONCHA...  6-4840A-010
176386                                        NaN          NaN
458571  GUARDAPOLVO DE EXCELENTE CALIDAD\n\nDODGE CALI...  DG11053AR
589884                         STOP CORSA 3P DER        CORSA
...                                           ...          ...
604469                    VARILLA MED ACEITE RODEOV6      RODEO
579329                                        NaN          NaN
172391                                        NaN          NaN
656432                            F.A/A GOLF 2006       GOLF
391878                                        NaN            -
```

```
          manufacturer            keyword   id_subscription  \
id
678876           MAZDA                NaN             701.0
458570   DIAMOND POWER  CONCHA BIELA 3.7             165.0
176386             NaN                NaN             105.0
458571       LION TUOH     GUARDAPOLVO RH             165.0
589884       CHEVROLET                NaN             701.0
...                ...                ...              ...
604469       CHEVROLET                NaN             701.0
579329             NaN                NaN             654.0
172391             NaN                NaN             105.0
656432      VOLKSWAGEN                NaN             701.0
391878             NaN                NaN              NaN

         id_product_category_l1                 sku
id
678876                 38420.0       B3C7-18-140B**
458570                 23244.0          6-4840-010
176386                  6001.0          ME:MS30617
458571                 23244.0            DG11053AR
589884                 42385.0  000000000093232514=
...                        ...                  ...
604469                 36872.0  000000008971486972
579329                     0.0       06036048124001
172391                  1319.0            MI:YL209
656432                 41727.0          1H0819644B
391878                     NaN                 NaN

[555935 rows x 15 columns]
```

In [9]: `df.dtypes`

Out[9]:
```
id_trademark            float64
id_measurement_unit     float64
id_record_status          int64
id_company_variation      int64
id_product_category       int64
code                      int64
name                     object
registration_date        object
description              object
model                    object
manufacturer             object
keyword                  object
id_subscription         float64
id_product_category_l1  float64
sku                      object
dtype: object
```

In [10]: `df_na = df.dropna()`

In [11]: `print(df_na)`

```
        id_trademark   id_measurement_unit   id_record_status   \
id
458570         2947.0               5382.0                  1
458571         3405.0               5382.0                  1
622187         2263.0               5592.0                  1
562683         3331.0               6568.0                  1
458597         2913.0               5382.0                  1
...               ...                  ...                ...
740649          111.0                199.0                  1
488623         2261.0               5592.0                  1
740646          111.0                199.0                  1
740647          111.0                199.0                  1
478472         2383.0               5592.0                  1


        id_company_variation   id_product_category   code   \
id
458570                   572                 23337   1801
458571                   572                 23394   1802
622187                   587                 24472   1456
562683                   683                 29013    349
458597                   572                 23298   1828
...                      ...                   ...    ...
740649                    13                   447   2532
488623                   587                 24500    866
740646                    13                   446   2529
740647                    13                   451   2530
478472                   587                 24437    707


                                              name registration_date   \
id
458570  Rodamientos Concha Biela A 010 Jeep Liberty 20...        2018-09-14
458571                 Guardapolvo Dodge Caliber Rh Sp        2018-09-14
622187  Gomas Valvula Explorer 4.6 Fx4 Triton F150 F25...        2020-03-11
562683     Correa Tiempo Gates Mitsubishi Outlander 3.0L        2019-11-21
458597  Muñon Inferior Dodge Neon 00 AL 06 Chrysler Pt...        2018-09-14
...                                            ...               ...
740649                                    Bomba agua        2020-07-22
488623  Liga Freno DOT 4 LV Motorcraft Original 473 ML...        2018-11-22
740646                                       Anillos        2020-07-22
740647                             Estopera tapa cadena        2020-07-22
478472  Kit Reparaciones Bomba Freno Ford F100 F150 Br...        2018-10-19


                                       description   \
id
458570  RODAMIENTOS CONCHA BIELA 3.7\n\nTIPO DE CONCHA...
458571  GUARDAPOLVO DE EXCELENTE CALIDAD\n\nDODGE CALI...
622187  FORD GOMAS DE VALVULA \n\nAPLICA PARA: EXPLORE...
562683  SOMOS DISTRIBUIDORES AUTORIZADOS GATES\nLament...
458597  MUÑÓN INFERIOR DE EXCELENTE CALIDAD\n\nPARA LO...
...                                            ...
740649          APLICA a 2da VERSION MOTOR 1GRFE TOYOTA
488623   MOTORCRAFT  Liga Freno \n\nNumero de parte: PM20
740646                      1.5 x 1.5 x 2.5mm / 102mm
740647                                             --
478472  MISAKAMI Kit Reparaciones Bomba Freno \n\nApli...
```

```
                               model          manufacturer   \
        id
        458570           6-4840A-010          DIAMOND POWER
        458571             DG11053AR             LION TUOH
        622187   EXPLORER 4.6 FX4 5.4                  Ford
        562683                  T339                 Gates
        458597             TVA-K7147                   TVA
        ...                     ...                   ...
        740649                     -  DNJ ENGINE COMPONENTS
        488623                  FORD                  Ford
        740646                     -  DNJ ENGINE COMPONENTS
        740647                     -  DNJ ENGINE COMPONENTS
        478472       F100 F150 Bronco                  Ford

                              keyword   id_subscription  id_product_category_l1  \
        id
        458570           CONCHA BIELA 3.7          165.0                 23244.0
        458571             GUARDAPOLVO RH          165.0                 23244.0
        622187             GOMAS DE VALVULA        165.0                 24406.0
        562683   Correa De Tiempo Outlander        549.0                 29011.0
        458597           MUÑÓN INFERIOR 00-06      165.0                 23244.0
        ...                     ...               ...                   ...
        740649               Bomba agua           105.0                  444.0
        488623               Liga Freno           165.0                24499.0
        740646                  Anillos           105.0                  444.0
        740647       Estopera tapa cadena         105.0                  444.0
        478472   Reparaciones Bomba Freno          165.0                 24406.0

                           sku
        id
        458570      6-4840-010
        458571       DG11053AR
        622187   CHV3L3Z6571DA
        562683            T339
        458597       TVA-K7147
        ...              ...
        740649           WP990
        488623            PM20
        740646       PR4224.20
        740647          TC4224
        478472          MK97935

        [6449 rows x 15 columns]
```

In [12]:
```python
grouped_data = df_na.groupby('manufacturer')
```
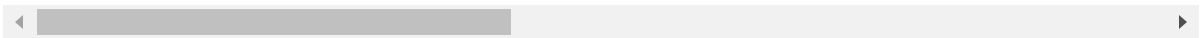
In [13]:
```python
grouped_data.describe()
```

Out[13]:   **id_tradema**

manufacturer	count	mean	std	min	25%	50%	75%	m
-	16.0	2334.125000	3573.333511	449.0	503.75	515.0	1688.5	1116
226ERS	44.0	15819.000000	0.000000	15819.0	15819.00	15819.0	15819.0	1581
A FILTERS	4.0	6407.000000	0.000000	6407.0	6407.00	6407.0	6407.0	640
A1FILTER	5.0	12409.000000	0.000000	12409.0	12409.00	12409.0	12409.0	1240
ACDELCO	4.0	2961.000000	0.000000	2961.0	2961.00	2961.0	2961.0	296
...	...	...	...	...	...	...	...	
motorcraft	6.0	2261.000000	0.000000	2261.0	2261.00	2261.0	2261.0	226
secutronik	6.0	3327.333333	3808.270771	483.0	838.25	1103.5	6297.0	849
tiendababy	7.0	1909.000000	0.000000	1909.0	1909.00	1909.0	1909.0	190
westingHouse	1.0	2296.000000	NaN	2296.0	2296.00	2296.0	2296.0	229
xiaomi	2.0	2320.000000	0.000000	2320.0	2320.00	2320.0	2320.0	232

615 rows × 64 columns

In [17]:
```python
count_data = df_na.groupby('manufacturer')['name'].nunique()
```

In [33]:
```python
count_data_mayores = count_data.sort_values(ascending=False)
```
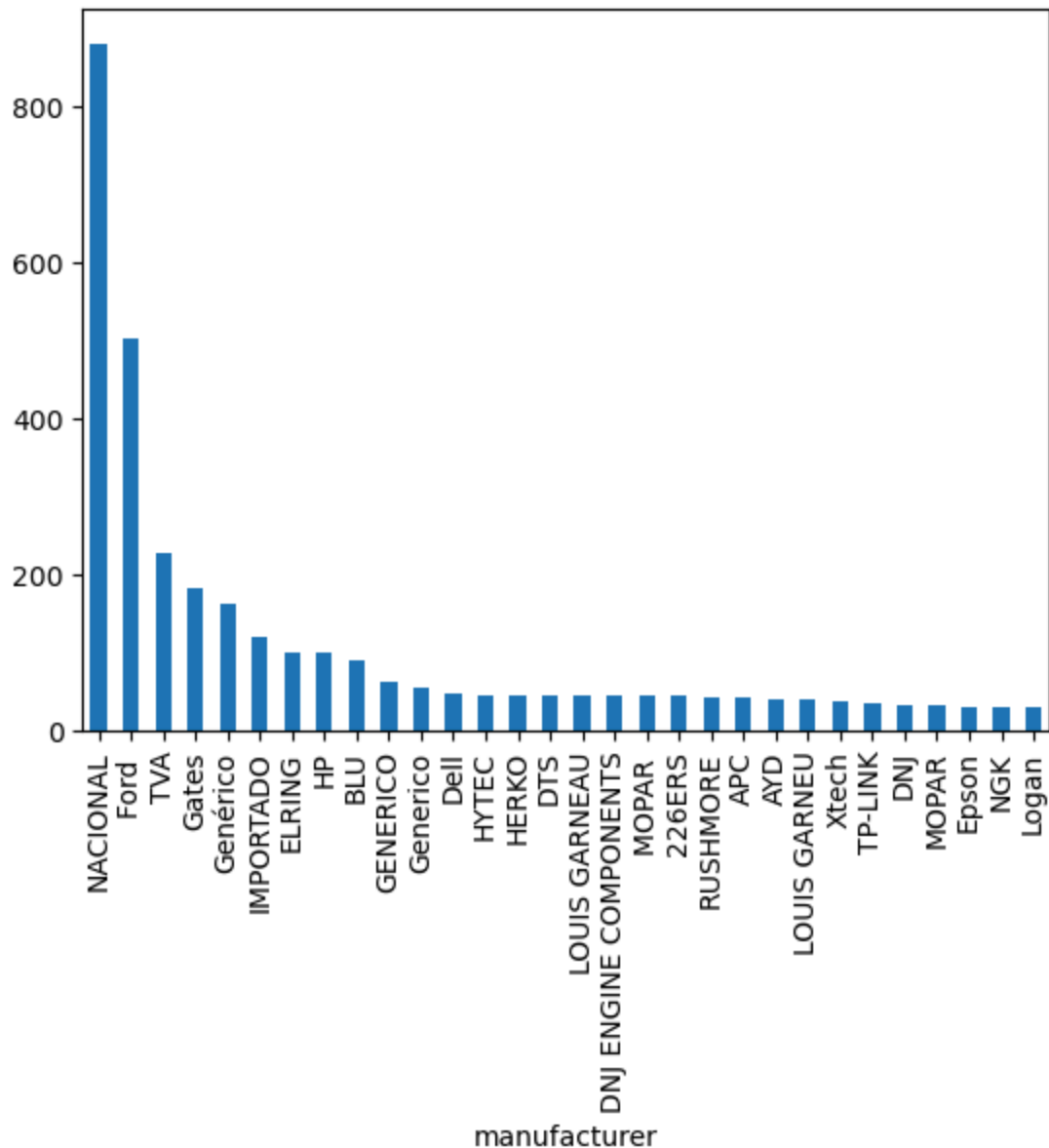
In [39]:
```python
print(count_data_mayores.head(50))
```

```
manufacturer
NACIONAL                   879
Ford                       502
TVA                        227
Gates                      183
Genérico                   161
IMPORTADO                  120
ELRING                     101
HP                         100
BLU                         91
GENERICO                    63
Generico                    54
Dell                        47
HYTEC                       46
HERKO                       46
DTS                         45
LOUIS GARNEAU               45
DNJ ENGINE COMPONENTS       44
MOPAR                       44
226ERS                      44
RUSHMORE                    43
APC                         42
AYD                         40
LOUIS GARNEU                40
Xtech                       37
TP-LINK                     35
DNJ                         32
MOPAR                       32
Epson                       31
NGK                         31
Logan                       29
FORD                        29
Samsung                     28
U Corporation               27
RGA                         26
Victor Reinz                26
TRENDER                     25
CROW                        25
China                       25
NATZUKI                     24
AMERICANO                   23
SHIMANO                     23
Lenovo                      23
WIREPLUS                    22
HANDUP                      22
MARUSSI                     21
HikVision                   21
Huawei                      21
WAYPRO                      21
Xiaomi                      20
LEVO                        20
Name: name, dtype: int64
```

In [41]: `count_data_mayores.head(30).plot(kind='bar')`

Out[41]: `<Axes: xlabel='manufacturer'>`

```
In [51]:   count_data2 = df_na.groupby('keyword')['name'].nunique()
```

```
In [55]:   count_data_mayores2 = count_data2.sort_values(ascending=False)
```

```
In [57]:   print(count_data_mayores2)
```

```
keyword
Celular                        164
UPS                             43
Laptop                          29
MAILLOT , CAMISAS , ROPA        29
MICA                            28
                              ...
Correa Unica Multicanal          1
Correa Unica Super Duty 7.3      1
Correa de Gates 207 Citroen C3   1
Correa de Tiempo                 1
Correa Unica                     1
Name: name, Length: 3102, dtype: int64
```
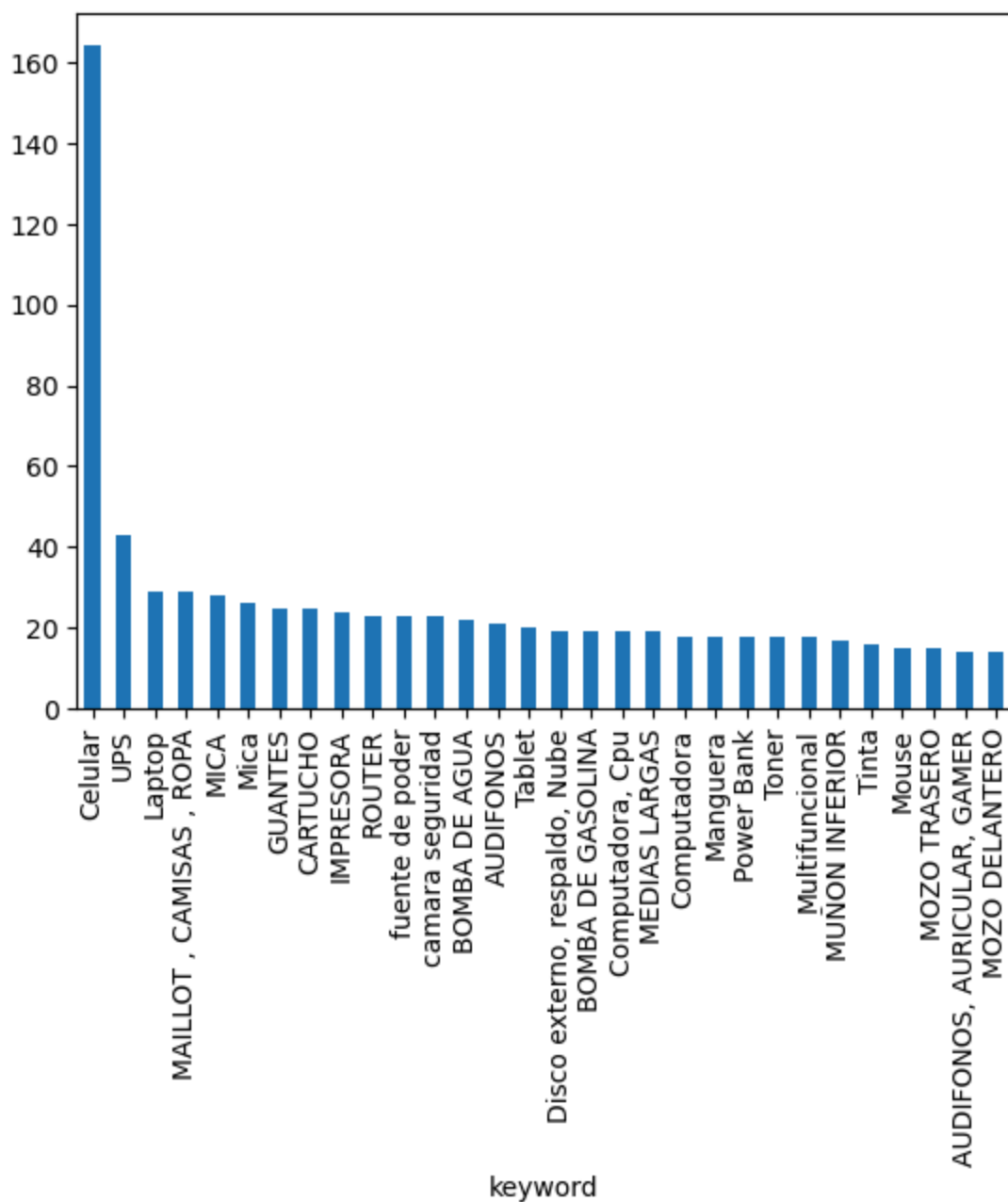
In [59]: `print(count_data_mayores2.head(50))`

```
keyword
Celular                            164
UPS                                 43
Laptop                              29
MAILLOT , CAMISAS , ROPA            29
MICA                                28
Mica                                26
GUANTES                             25
CARTUCHO                            25
IMPRESORA                           24
ROUTER                              23
fuente de poder                     23
camara seguridad                    23
BOMBA DE AGUA                       22
AUDIFONOS                           21
Tablet                              20
Disco externo, respaldo, Nube       19
BOMBA DE GASOLINA                   19
Computadora, Cpu                    19
MEDIAS LARGAS                       19
Computadora                         18
Manguera                            18
Power Bank                          18
Toner                               18
Multifuncional                      18
MUÑON INFERIOR                      17
Tinta                               16
Mouse                               15
MOZO TRASERO                        15
AUDIFONOS, AURICULAR, GAMER         14
MOZO DELANTERO                      14
ZAPATOS                             14
TONER                               13
CAUCHO , CAUCHOS                    13
CORREA UNICA                        13
FORRO                               13
Televisor                           12
PLASTIC BOTTLE                      12
PUNTA TRIPOIDE                      12
Corneta                             12
cases, gamer, gaming                11
Disco Duro                          11
PENDRIVE                            11
Multifuncional, Impresora           11
GUANTES CORTOS                      11
Monitor                             10
MUÑON SUPERIOR                      10
VIDRIO                              10
Corneta Pórtatil                    10
BOLSO                               10
ROTULA                              10
Name: name, dtype: int64
```

In [61]: `count_data_mayores2.head(30).plot(kind='bar')`

Out[61]: `<Axes: xlabel='keyword'>`

In [ ]: