

```
In [ ]: # Content based manuplations
# Change in Data types
# DF.astype()
#DF.columnName = DF.columnName.astype(str)
#DF.columnName = DF.columnName.astype(float)
#DF.columnName = DF.columnName.astype(int)
#DF.columnName = DF.columnName.astype(bool)
```

```
In [13]: #pd.to_numeric()
```

```
import pandas
import pandas as pd
```

```
In [9]: n1=pd.Series([1,33,777,0,0.0001,1000])
n1.astype(bool)
```

```
Out[9]: 0      True
1      True
2      True
3     False
4      True
5      True
dtype: bool
```

```
In [11]: income=pd.Series(["1","33","777","0","0.0001","1000"])
income.astype(float)
```

```
Out[11]: 0      1.0000
1     33.0000
2    777.0000
3      0.0000
4      0.0001
5   1000.0000
dtype: float64
```

```
In [12]: income2=pd.Series(["1","33","777","0","0.0001","1000","NA"])
pd.to_numeric(income2,errors="coerce")
# it has argument called errors = "ignore" it not work!
# we set it to "coerce" meaning force:
# we force the unconvertables to missing values
#missing/blank in python is denoted by NaN
```

```
Out[12]: 0      1.0000
1     33.0000
2    777.0000
3      0.0000
4      0.0001
5   1000.0000
6         NaN
dtype: float64
```

```
In [14]: # Python cannot understand dates
DOJ=pd.Series(["12-Mar-2016","31-Dec-2011","19-Jan-2021","27-Feb-2001"])
```

```
In [15]: DOJ
```

```
Out[15]: 0    12-Mar-2016
1    31-Dec-2011
2    19-Jan-2021
3    27-Feb-2001
dtype: object
```

```
In [16]: pd.to_datetime(DOJ,format = "%d-%b-%Y")
```

```
Out[16]: 0    2016-03-12
1    2011-12-31
2    2021-01-19
3    2001-02-27
dtype: datetime64[ns]
```

```
In [ ]: #strptime - string parsing of time:converts str to time
#day
#    day no                %d                formate=""
#    weekday Abbr         %a                13-Mar-2014  -> %d-%b-%Y
#    weekday name         %A                15/12/1990  -> %d/%m%Y
#Month                    %m                01 26 1990  -> %m%d%Y
#    month no             %m
#    abbr                 %b
#    name                  %B
#Year
#    with century(1989)   %Y
#    w/o century (89)     %y
#hours
#    12 hours             %I
#    Am/Pm                 %P
```

```
# 24 hour %H
#mins %M
#secs %S
```

```
In [18]: # now we have proper dates
# Every datetime64 series gets a new attribute :dt
#dt has sub-attributes all date time utility function
```

```
In [ ]: DOJ1.dt.day_name()
DOJ1.dt.month_name()
DOJ1.dt.quarter()
```

```
In [ ]: DOJ1.dt.strftime("%A, %d %B %Y")
#1.datetime64 is always displayed as YYYY-MM-DD
#2. If we try formatting it , the ser/column will no longer remain datetime64
# we are converting it into strings : with a desired formate
```

Content Based Changes -filtering -applying conditions to data -sorting -removing redundancy/duplicates -imputing missing values(NaN) -capping outliers
-binning -encoding + dummy variables -group by -merging -get insights -apply stats models -solve a business problem -report and visualize it

```
In [13]: ## Structural manuplations
import pandas as pd
amazon = pd.read_csv(r"C:\Users\sival\Documents\datasets\amazon-products.csv",sep=",")
amazon
```

Out[13]:	timestamp	title	seller_name	brand	description	initial_price	final_pric
0	2023-08-08 00:00:00.000	Saucony Men's Kinvara 13 Running Shoe	Orv[REDACTED]tor[REDACTED]	Saucony	When it comes to lightweight speed, nothing cr...	NaN	"57.79"
1	2023-08-09 00:00:00.000	Kishigo Premium Black Series Heavy Duty Unisex...	Ama[REDACTED].co[REDACTED]	Kishigo	The Kishigo Premium Black Series Heavy Duty Ve...	NaN	"28.99"
2	2024-02-04 00:00:00.000	TWINSLUXES Solar Post Cap Lights Outdoor - Wat...	Tw[REDACTED]uxe[REDACTED]	TWINSLUXES	Solar Post Cap Lights Waterproof LED Fence Pos...	"49.99"	"33.99"
3	2024-06-09 00:00:00.000	Accutire MS-4021B Digital Tire Pressure Gauge ...	Cit[REDACTED]ran[REDACTED]Dir[REDACTED]	Accutire	About this item Heavy duty construction and ru...	1.7950000000000000e+01	1.7950000000000000e+C
4	2024-01-16 00:00:00.000	SAURA LIFE SCIENCE Adivasi Ayurvedic Neelgiri ...	PRA[REDACTED]EN[REDACTED]PRI[REDACTED]	SAURA LIFE SCIENCE	This extraordinary fusion is designed to nour...	"1299"	"799"
...
995	2023-08-19 00:00:00.000	Madison Park Amherst Faux Silk Comforter Set-C...	Ama[REDACTED].co[REDACTED]	Madison Park	The Madison Park Amherst 7 Piece Comforter Set...	"154.99"	"111.99"
996	2023-08-15 00:00:00.000	BIC Gel-Ocity Gel Pens, Medium Point Retractable...	Ama[REDACTED].co[REDACTED]	BIC	Gel-ocity is a smooth writing retractable gel ...	"10.99"	"9.99"
997	2024-06-10 00:00:00.000	Elmer's E842L Carpenter's Interior Wood Filler...	Cer[REDACTED]ied[REDACTED]ole[REDACTED]ly	Elmer's	Product Description Elmer's E842L Carpenter's ...	1.3980000000000000e+01	1.3980000000000000e+C
998	2024-05-28 00:00:00.000	Office Chair Mat for Carpet, 30" x 48" Transpa...	Hom[REDACTED]Lif[REDACTED]	HOMEK	HOMEK Office Chair Mat for Carpet HOMEK floo...	3.9950000000000000e+01	3.9950000000000000e+C
999	2024-03-03 00:00:00.000	Bavilk Kids Bike Helmet,Boys Girls Helmet Mult...	Bav[REDACTED]Dir[REDACTED]	Bavilk	TODDLER KIDS BIKE HELMETS,Boys Girls Helmet Mu...	"32.99"	"16.49"

1000 rows × 55 columns

```
In [ ]: # 1. filtering
#
```

```
In [18]: # select * from amazon where availability is instock
amazonstock = amazon.loc[(amazon.availability == "In Stock")]
amazonstock.reset_index(drop=True)
```

Out[18]:

	timestamp	title	seller_name	brand	description	initial_price	final_price
0	2023-08-08 00:00:00.000	Saucony Men's Kinvara 13 Running Shoe	Orv[REDACTED]tor[REDACTED]	Saucony	When it comes to lightweight speed, nothing cr...	NaN	"57.79"
1	2023-08-09 00:00:00.000	Kishigo Premium Black Series Heavy Duty Unisex...	Ama[REDACTED].co[REDACTED]	Kishigo	The Kishigo Premium Black Series Heavy Duty Ve...	NaN	"28.5"
2	2024-02-04 00:00:00.000	TWINSLUXES Solar Post Cap Lights Outdoor - Wat...	Tw[REDACTED]uxe[REDACTED]	TWINSLUXES	Solar Post Cap Lights Waterproof LED Fence Pos...	"49.99"	"33.99"
3	2024-06-09 00:00:00.000	Accutire MS-4021B Digital Tire Pressure Gauge ...	Cit[REDACTED]ran[REDACTED]Dir[REDACTED]	Accutire	About this item Heavy duty construction and ru...	1.7950000000000000e+01	1.7950000000000000e+01
4	2023-09-10 00:00:00.000	KASOTT Replacement Airpod Pro Ear Tip Premium ...	KAS[REDACTED] of [REDACTED]e	KASOTT	About this item [Exclusive Customization] Ma...	"17.88"	"17.88"
...
586	2024-04-30 00:00:00.000	Zak Designs Water Bottle for Travel and At Hom...	Ama[REDACTED].co[REDACTED]	Zak Designs	Make hydration easy and fun with a reusable wa...	2.3990000000000000e+01	"17.32"
587	2023-05-16 00:00:00.000	VELCRO Brand - 91302 Thin Clear Dots with Adhe...	NaN	VELCRO Brand	VELCRO Brand Thin Clear fasteners blend into g...	"11.39"	"12.03"
588	2023-08-15 00:00:00.000	BIC Gel-Ocity Gel Pens, Medium Point Retractable...	Ama[REDACTED].co[REDACTED]	BIC	Gel-ocity is a smooth writing retractable gel ...	"10.99"	"9.99"
589	2024-06-10 00:00:00.000	Elmer's E842L Carpenter's Interior Wood Filler...	Cer[REDACTED]ied[REDACTED]ole[REDACTED]ly	Elmer's	Product Description Elmer's E842L Carpenter's ...	1.3980000000000000e+01	1.3980000000000000e+01
590	2024-05-28 00:00:00.000	Office Chair Mat for Carpet, 30" x 48" Transpa...	Hom[REDACTED]e Lif[REDACTED]	HOMEK	HOMEK Office Chair Mat for Carpet HOMEK floo...	3.9950000000000000e+01	3.9950000000000000e+01

591 rows × 55 columns

In [26]:

```
#select title,sellername,brand where reviews_count is less than 100
amazon.loc[(amazon.reviews_count < 100),["title","seller_name","brand"]]
```

Out[26]:

	title	seller_name	brand
4	SAURA LIFE SCIENCE Adivasi Ayurvedic Neelgiri ...	PRA[REDACTED]EN[REDACTED]PRI[REDACTED]	SAURA LIFE SCIENCE
5	KASOTT Replacement Airpod Pro Ear Tip Premium ...	KAS[REDACTED] of [REDACTED]e	KASOTT
8	Ridgid 62990 T-201 5" Straight Auger	Emp[REDACTED] Ridgid[REDACTED]	RIDGID
14	Stahlwille 96410704 Double Ended Ring Wrench S...	Ama[REDACTED].co[REDACTED]	Stahlwille
16	General Tools 3-Pack Retractable Utility Knife...	Ama[REDACTED].co[REDACTED]	General Tools
...
972	KNIPEX Electronics Oblique Cut	Ama[REDACTED].co[REDACTED]	KNIPEX
975	Euro Tool Dual Eurocrimper, 5 3/4 Inches PLR...	Jew[REDACTED]yTo[REDACTED]	EURO TOOL
976	FBLFOBELI EVA Hard Carrying Case Compatible Wi...	FBL[REDACTED]bel[REDACTED]	FBLFOBELI
978	Puroflo Tankless RO System Under sink 600GPD H...	NaN	PUROFLO
979	10X Magnifying Glass with Light and Clamp, KIR...	KIR[REDACTED]	KIRKAS

221 rows × 3 columns

In [28]:

```
# 2. sorting
# DF.sort_values()
```

```
# sort amazon based on brand asc order
amazon.sort_values(["brand"]).reset_index(drop=True)
```

Out[28]:

	timestamp	title	seller_name	brand	description	initial_price	final_price	currenc
0	2022-09-17 00:00:00.000	3D MAXpider Front Row Custom Fit All-Weather F...	NaN	3D MAXpider	3D MAXpider Custom Fit All Weather Floor Liner...	NaN	"92.89"	USI
1	2023-05-20 00:00:00.000	3D MAXpider Custom Fit Kagu Floor Mat (Black) ...	NaN	3D MAXpider	Product Description .aplus-v2 { display: block...	NaN	"92.89"	USI
2	2023-08-08 00:00:00.000	Sperry Top-Sider Leeward 2 Eye Boat Shoe	Ama .co	6pm Sperry Top-Sider Footwear RP	Leeward 2-Eye Corduroy Boat Shoe Best known fo...	"110"	"69.99"	USI
3	2023-08-16 00:00:00.000	A-Premium Electric Fuel Pump Compatible with C...	Pre .mpa . Who	A-Premium	Compatible with following Chevrolet Models: 20...	"29.99"	"29.99"	USI
4	2024-01-15 00:00:00.000	Aco Jasmine Premium Scent, Long Lasting, Fresh...	S V T	ACO perfumes	About this item Long Lasting Perfume: An allur...	"320"	"310"	INI
...
995	2024-06-10 00:00:00.000	Hanging File Organizer, 3 Pocket Wall Mount Do...	Sup . pag . rec	superbpag	Great quality This wire file holder features a...	2.9990000000000000e+01	2.9990000000000000e+01	USI
996	2024-06-25 00:00:00.000	thermomate Electric Tank Water Heater, 6.5 Gal...	the . flo	thermomate	Previous page Next page 1 Space Saving 2 Insta...	2.2999000000000000e+02	2.2999000000000000e+02	USI
997	2024-06-10 00:00:00.000	Clip Mount for Eufy Baby Monitor Camera, Flexi...	Woc . US	wochel	Flexible Baby Monitor Holder - Specially Desig...	1.5990000000000000e+01	1.5990000000000000e+01	USI
998	2024-05-13 00:00:00.000	wusoda Silent Basketball Dribbling Indoor Drib...	HMO . S	wusoda	Silent Basketball Dribbling Indoor Dribble Dre...	1.8990000000000000e+01	1.7990000000000000e+01	USI
999	2022-09-06 00:00:00.000	Replenix Retinol Smooth + Tighten Body Lotion ...	NaN	NaN	Previous page Advanced Performance Medical-Gra...	"93.00"	"84.75"	USI

1000 rows × 55 columns

```
In [31]: #sort amazon based on brand in desc order
amazon.sort_values(["brand"],ascending=False).reset_index(drop=True)
```

Out[31]:

	timestamp	title	seller_name	brand	description	initial_price	final_price	currentc
0	2024-05-13 00:00:00.000	wusoda Silent Basketball Dribbling Indoor Dribble Dre...	HMO S	wusoda	Silent Basketball Dribbling Indoor Dribble Dre...	1.8990000000000000e+01	1.7990000000000000e+01	USI
1	2024-06-10 00:00:00.000	Clip Mount for Eufy Baby Monitor Camera, Flexi...	Woc US	wochel	Flexible Baby Monitor Holder - Specially Desig...	1.5990000000000000e+01	1.5990000000000000e+01	USI
2	2024-06-25 00:00:00.000	thermomate Electric Tank Water Heater, 6.5 Gal...	the flo	thermomate	Previous page Next page 1 Space Saving 2 Insta...	2.2999000000000000e+02	2.2999000000000000e+02	USI
3	2024-06-10 00:00:00.000	Hanging File Organizer, 3 Pocket Wall Mount Do...	Sup pag rec	superbpag	Great quality This wire file holder features a...	2.9990000000000000e+01	2.9990000000000000e+01	USI
4	2024-04-29 00:00:00.000	Motorcycle Covers Waterproof Outdoor Storage, ...	Sec s U	seceles	300D Heavy Duty Better Waterproof Made of 300D...	3.1990000000000000e+01	"31.99"	USI
...
995	2023-08-16 00:00:00.000	A-Premium Electric Fuel Pump Compatible with C...	Pre mpa Who	A-Premium	Compatible with following Chevrolet Models: 20...	"29.99"	"29.99"	USI
996	2023-08-08 00:00:00.000	Sperry Top-Sider Leeward 2 Eye Boat Shoe	Ama .co	6pm Sperry Top-Sider Footwear RP	Leeward 2-Eye Corduroy Boat Shoe Best known fo...	"110"	"69.99"	USI
997	2022-09-17 00:00:00.000	3D MAXpider Front Row Custom Fit All-Weather F...	NaN	3D MAXpider	3D MAXpider Custom Fit All Weather Floor Liner...	NaN	"92.89"	USI
998	2023-05-20 00:00:00.000	3D MAXpider Custom Fit Kagu Floor Mat (Black) ...	NaN	3D MAXpider	Product Description .aplus-v2 { display: block...	NaN	"92.89"	USI
999	2022-09-06 00:00:00.000	Replenix Retinol Smooth + Tighten Body Lotion ...	NaN	NaN	Previous page Advanced Performance Medical-Gra...	"93.00"	"84.75"	USI

1000 rows × 55 columns

```
In [32]: #sort amazon based on brand asc orer and title in desc order
amazon.sort_values(["brand", "title"],ascending=[True,False]).reset_index(drop=True)
```

Out[32]:

	timestamp	title	seller_name	brand	description	initial_price	final_price	currenc
0	2022-09-17 00:00:00.000	3D MAXpider Front Row Custom Fit All-Weather F...	NaN	3D MAXpider	3D MAXpider Custom Fit All Weather Floor Liner...	NaN	"92.89"	USI
1	2023-05-20 00:00:00.000	3D MAXpider Custom Fit Kagu Floor Mat (Black) ...	NaN	3D MAXpider	Product Description .aplus-v2 { display: block...	NaN	"92.89"	USI
2	2023-08-08 00:00:00.000	Sperry Top-Sider Leeward 2 Eye Boat Shoe	Ama[REDACTED].co[REDACTED]	6pm Sperry Top-Sider Footwear RP	Leeward 2-Eye Corduroy Boat Shoe Best known fo...	"110"	"69.99"	USI
3	2023-08-16 00:00:00.000	A-Premium Electric Fuel Pump Compatible with C...	Pre[REDACTED]mpa[REDACTED]Who[REDACTED]	A-Premium	Compatible with following Chevrolet Models: 20...	"29.99"	"29.99"	USI
4	2024-01-15 00:00:00.000	Aco Jasmine Premium Scent, Long Lasting, Fresh...	S V T	ACO perfumes	About this item Long Lasting Perfume: An allur...	"320"	"310"	INI
...
995	2024-06-10 00:00:00.000	Hanging File Organizer, 3 Pocket Wall Mount Do...	Sup[REDACTED]pag[REDACTED]rec[REDACTED]	superbpag	Great quality This wire file holder features a...	2.9990000000000000e+01	2.9990000000000000e+01	USI
996	2024-06-25 00:00:00.000	thermomate Electric Tank Water Heater, 6.5 Gal...	the[REDACTED]flo[REDACTED]	thermomate	Previous page Next page 1 Space Saving 2 Insta...	2.2999000000000000e+02	2.2999000000000000e+02	USI
997	2024-06-10 00:00:00.000	Clip Mount for Eufy Baby Monitor Camera, Flexi...	Woc[REDACTED]US[REDACTED]	wochel	Flexible Baby Monitor Holder - Specially Desig...	1.5990000000000000e+01	1.5990000000000000e+01	USI
998	2024-05-13 00:00:00.000	wusoda Silent Basketball Dribbling Indoor Drib...	HMO[REDACTED]S	wusoda	Silent Basketball Dribbling Indoor Dribble Dre...	1.8990000000000000e+01	1.7990000000000000e+01	USI
999	2022-09-06 00:00:00.000	Replenix Retinol Smooth + Tighten Body Lotion ...	NaN	NaN	Previous page Advanced Performance Medical-Gra...	"93.00"	"84.75"	USI

1000 rows × 55 columns

In [38]:

```
# Removing duplicates
#two ways to identify something duplicates:
#1. based on each cell in a row
amazon seller = amazon.seller_name
amazon seller.duplicated()
```

Out[38]:

```
0      False
1      False
2      False
3      False
4      False
...
995    True
996    True
997    False
998    False
999    False
Name: seller_name, Length: 1000, dtype: bool
```

In [39]:

```
amazon seller.loc[~amazon.duplicated()] # removes all the duplicate records
```

```
Out[39]: 0      Orv      tor
1      Ama      .co
2      Twi      uxe
3      Cit      ran      Dir
4      PRA      EN      PRI

...
995     Ama      .co
996     Ama      .co
997     Cer      ied      ole      ly
998     Hom      Lif
999     Bav      Dir
Name: seller_name, Length: 1000, dtype: object
```

```
In [41]: amazon seller.loc[amazon.duplicated()]

Out[41]: Series([], Name: seller_name, dtype: object)
```

```
In [44]: #Dropping duplicates
amazon seller.drop_duplicates()
#exactly similar to amazon seller.loc[-amazon.duplicated()]

Out[44]: 0      Orv      tor
1      Ama      .co
2      Twi      uxe
3      Cit      ran      Dir
4      PRA      EN      PRI

...
988     BFJ      E S
994     LA      hio      deal
997     Cer      ied      ole      ly
998     Hom      Lif
999     Bav      Dir
Name: seller_name, Length: 570, dtype: object
```

```
In [46]: # 4. Imputing missing values
# Excel/csv or SQL database
## Error #N/A-->NaN
## SQL :NULL ...>NaN
amazon.isna() #values missing
```

Out[46]:

	timestamp	title	seller_name	brand	description	initial_price	final_price	currency	availability	reviews_count	...	root_bs_category	t
0	False	False	False	False	False	True	False	False	False	False	...	True	
1	False	False	False	False	False	True	False	False	False	False	...	True	
2	False	False	False	False	False	False	False	False	False	False	...	True	
3	False	False	False	False	False	False	False	False	False	False	...	False	
4	False	False	False	False	False	False	False	False	False	False	...	True	
...	
995	False	False	False	False	False	False	False	False	False	False	...	True	
996	False	False	False	False	False	False	False	False	False	False	...	True	
997	False	False	False	False	False	False	False	False	False	False	...	False	
998	False	False	False	False	False	False	False	False	False	False	...	False	
999	False	False	False	False	False	False	False	False	True	False	...	True	

1000 rows × 55 columns

```
In [52]: amazon.availability.isna()
nmiss =sum(amazon.availability.isna())
nmiss
```

```
Out[52]: 6
```

```
In [65]: amazon1=amazon.reviews_count
amazon1.dropna()
amazon1

Out[65]: Series([], Name: reviews_count, dtype: int64)
```

```
In [72]: amazon1.fillna(amazon1.mean())

Out[72]: Series([], Name: reviews_count, dtype: int64)
```

```
In [ ]:
```