

Importing necessary modules

In [1]:

```
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
```

Importing data

In [2]:

```
1 game=pd.read_csv(r"K:\Desktop\NIIT\tables\gamer.csv")
2 game
```

Out[2]:

	Game	Year	Genre	Publisher	North_America_Sales_M\$	Europe_Sales_M\$	Japan_Sales_M\$	Other_Sales_M\$	Global_Sales_M\$
0	Grand Theft Auto V	2016	Action	Rockstar Games	6.06	9.71	0.60	3.02	19.39
1	Call of Duty: Black Ops 3	2017	Shooter	Activision	6.18	6.05	0.41	2.44	15.09
2	Red Dead Redemption 2	2020	Action-Adventure	Rockstar Games	5.26	6.21	0.21	2.26	13.94
3	Call of Duty: WWII	2019	Shooter	Activision	4.67	6.21	0.40	2.12	13.40
4	FIFA 18	2019	Sports	EA Sports	1.27	8.64	0.15	1.73	11.80
...
704	Hyper Light Drifter	2019	Role-Playing	Active Gaming Media	0.00	0.00	0.01	0.00	0.01
705	Teslagrad	2016	Platform	Rain Games	0.00	0.01	0.00	0.00	0.01
706	Darius Burst: Chronicle Saviours	2019	Action	Kadokawa Games	0.00	0.00	0.01	0.00	0.01
707	The Escapists: The Walking Dead	2018	Action	Team17 Software	0.00	0.00	0.00	0.00	0.01
708	Super Meat Boy	2018	Platform	Team Meat	0.00	0.00	0.00	0.00	0.01

709 rows × 9 columns

Task1 : To find probability of selecting a female who is working part time

In [3]:

```
1 print("Probabilit of selecting woman = 0.46")
2 print("Probability of selecting part time working woman = ",(0.46*0.25)/0.17)
3
```

Probabilit of selecting woman = 0.46
Probability of selecting part time working woman = 0.676470588235294

Task2: To determine which office changes would increase productivity

In [4]:

```
1 print("Reducing noise = 0.7")
2 print("Increasing storage = 0.67")
3 print("Both = 0.57")
4 print("probability of finding more storage guy is  = ",0.57/0.7)
```

Reducing noise = 0.7
Increasing storage = 0.67
Both = 0.57
probability of finding more storage guy is = 0.8142857142857143

Task3: Game data

Subtasks 3.1.1 Random sampling without replacement

In [5]:

```
1 rand_samp=game.sample(30,random_state=13,ignore_index=True)
2 rand_samp
```

Out[5]:

	Game	Year	Genre	Publisher	North_America_Sales_M\$	Europe_Sales_M\$	Japan_Sales_M\$	Other_Sales_M\$	Global_Sales_M\$
0	Hidden Agenda	2019	Adventure	Sony Interactive Entertainment	0.08	0.42	0.00	0.09	0.59
1	We Sing (2018)	2018	Music	THQ Nordic	0.02	0.00	0.00	0.01	0.03
2	Gal Gun: Double Peace	2017	Action	PQube	0.00	0.00	0.01	0.00	0.01
3	Nobunaga's Ambition: Creation	2016	Strategy	Tecmo Koei	0.04	0.00	0.00	0.01	0.05
4	LEGO The Hobbit	2016	Action-Adventure	Warner Bros. Interactive Entertainment	0.14	0.39	0.00	0.10	0.62
5	The Elder Scrolls V: Skyrim	2018	Role-Playing	Bethesda Softworks	1.24	1.50	0.08	0.54	3.36
6	Until Dawn: Rush of Blood	2018	Shooter	Sony Interactive Entertainment	0.50	0.21	0.00	0.15	0.85
7	Shin Hayarigami 2	2018	Adventure	Nippon Ichi Software	0.00	0.00	0.03	0.00	0.03
8	Golf Club 2	2019	Sports	Maximum Games	0.05	0.06	0.00	0.02	0.13
9	FIFA 18	2019	Sports	EA Sports	1.27	8.64	0.15	1.73	11.80
10	Punch Line	2017	Adventure	5pb	0.03	0.00	0.00	0.01	0.04
11	NieR Automata	2019	Role-Playing	Square Enix	0.56	0.51	0.42	0.22	1.71
12	Natsuiro High School: Seishun Hakusho	2017	Action	D3Publisher	0.00	0.00	0.02	0.00	0.02
13	Dynasty Warriors 8: Empires	2016	Misc	Tecmo Koei	0.07	0.07	0.04	0.03	0.21
14	Assassin's Creed IV: Black Flag	2015	Action-Adventure	Ubisoft	1.07	1.55	0.06	0.51	3.19
15	Far Cry 5	2020	Action	Ubisoft	1.44	1.73	0.15	0.62	3.95
16	Farming Simulator 17	2018	Simulation	Focus Home Interactive	0.19	0.52	0.01	0.13	0.85
17	The Witcher 3: Wild Hunt	2017	Role-Playing	Namco Bandai Games	1.48	2.82	0.28	0.81	5.39
18	Metal Gear Survive	2020	Action	Konami Digital Entertainment	0.08	0.06	0.05	0.03	0.23
19	ClaDun Sengoku	2019	Action	NIS America	0.02	0.00	0.00	0.00	0.03
20	Wolfenstein: The Old Blood	2017	Shooter	Bethesda Softworks	0.13	0.25	0.02	0.07	0.48
21	WWE 2K18	2019	Sports	2K Sports	0.45	0.67	0.00	0.22	1.33
22	NBA Live 18	2019	Sports	EA Sports	0.20	0.00	0.00	0.05	0.24
23	Robinson: The Journey	2018	Action-Adventure	Sony Interactive Entertainment	0.14	0.09	0.00	0.05	0.27
24	XCOM 2	2018	Strategy	2K Games	0.16	0.10	0.00	0.05	0.31
25	Injustice: Gods Among Us	2015	Fighting	Warner Bros. Interactive Entertainment	0.46	0.35	0.00	0.17	0.98
26	Werewolves Within	2018	Misc	Ubisoft	0.04	0.00	0.00	0.01	0.05
27	WWE 2K16	2017	Sports	Take-Two Interactive	0.43	0.65	0.00	0.21	1.29
28	Battlezone	2018	Action	Sony Interactive Entertainment	0.09	0.04	0.00	0.03	0.16
29	Syberia III	2019	Adventure	Microids	0.04	0.00	0.00	0.01	0.05

Subtasks 3.1.2 Random sampling with replacement

In [6]:

```
1 rand_samp_r=game.sample(30,random_state=13,replace=True,ignore_index=True)
2 rand_samp_r
```

Out[6]:

	Game	Year	Genre	Publisher	North_America_Sales_M\$	Europe_Sales_M\$	Japan_Sales_M\$	Other_Sales_M\$	Global_Sales_M\$
0	Eagle Flight	2018	Adventure	Ubisoft	0.09	0.08	0.00	0.03	0.20
1	WipEout Omega Collection	2019	Racing	Sony Interactive Entertainment	0.12	0.49	0.00	0.11	0.72
2	Need for Speed Rivals	2015	Racing	Electronic Arts	0.75	1.04	0.03	0.35	2.17
3	Exist Archive: The Other Side of the Sky	2017	Role-Playing	Aksys Games	0.02	0.00	0.03	0.00	0.06
4	LEGO Harry Potter Collection	2018	Action	Warner Bros. Interactive Entertainment	0.07	0.68	0.00	0.13	0.88
5	Farming Simulator 2017	2017	Simulation	Koch Media	0.13	0.22	0.00	0.07	0.41
6	Omega Labyrinth Z	2019	Role-Playing	D3Publisher	0.00	0.00	0.02	0.00	0.02
7	Mega Man Legacy Collection 2	2019	Action	Capcom	0.06	0.00	0.02	0.01	0.10
8	TrackMania Turbo	2018	Racing	Ubisoft	0.04	0.35	0.00	0.07	0.47
9	Aegis of Earth: Protonovus Assault	2018	Action	PQube	0.03	0.01	0.00	0.01	0.04
10	The Crew 2	2020	Racing	Ubisoft	0.22	0.42	0.05	0.12	0.82
11	Final Fantasy X/X-2 HD Remaster	2017	Role-Playing	Square Enix	0.34	0.36	0.07	0.14	0.90
12	The Legend of Heroes: Trails of Cold Steel III	2019	Role-Playing	Nihon Falcom Corporation	0.00	0.00	0.13	0.00	0.13
13	Ni-Oh	2019	Action	Tecmo Koei	0.34	0.36	0.20	0.14	1.03
14	Legend of Kay	2017	Action	Nordic Games	0.03	0.03	0.00	0.01	0.08
15	PlayStation VR Worlds	2018	Misc	Sony Interactive Entertainment	0.36	1.40	0.09	0.32	2.16
16	Call of Duty: Black Ops 3	2017	Shooter	Activision	6.18	6.05	0.41	2.44	15.09
17	Professional Farmer 2016	2017	Action	Namco Bandai Games	0.00	0.02	0.00	0.00	0.02
18	Naruto to Boruto: Shinobi Striker	2020	Action	Bandai Namco Games	0.19	0.03	0.02	0.05	0.29
19	Danganronpa 1o2 Reload	2019	Adventure	NIS America	0.08	0.00	0.02	0.02	0.12
20	Star Wars Battlefront 2017	2017	Shooter	Electronic Arts	3.31	3.19	0.23	1.30	8.03
21	Fate/Extella: The Umbral Star	2018	Action	PQube	0.12	0.00	0.12	0.03	0.27
22	Tony Hawk's Pro Skater 5	2017	Sports	Activision	0.11	0.08	0.00	0.04	0.23
23	Black Mirror	2019	Adventure	THQ Nordic	0.04	0.00	0.00	0.01	0.05
24	Ziggurat	2018	Shooter	Soedesco	0.00	0.01	0.00	0.00	0.01
25	Atelier Sophie: The Alchemist of the Mysteriou...	2017	Role-Playing	Tecmo Koei	0.04	0.03	0.07	0.01	0.15
26	Dead Rising 2	2018	Action	Capcom	0.04	0.00	0.02	0.01	0.07
27	Project Setsuna	2018	Role-Playing	Square Enix	0.00	0.00	0.07	0.00	0.07
28	The Crew: Wild Run	2017	Racing	Ubisoft	0.00	0.12	0.00	0.02	0.14
29	Metal Gear Solid: Ground Zeroes	2016	Action-Adventure	Konami Digital Entertainment	0.36	0.50	0.17	0.17	1.20

Subtasks 3.1.3 Systematic sampling

In [7]:

```
1 game.iloc[np.arange(len(game),step=24)]
```

Out[7]:

	Game	Year	Genre	Publisher	North_America_Sales_M\$	Europe_Sales_M\$	Japan_Sales_M\$	Other_Sales_M\$	Global_Sales_M\$
0	Grand Theft Auto V	2016	Action	Rockstar Games	6.06	9.71	0.60	3.02	19.39
24	Monster Hunter: World	2020	Action	Capcom	1.03	1.06	2.17	0.42	4.67
48	Madden NFL 16	2017	Sports	EA Sports	2.39	0.31	0.00	0.61	3.30
72	Need for Speed: Payback	2019	Racing	Electronic Arts	0.62	1.18	0.04	0.34	2.18
96	Persona 5	2018	Role-Playing	Deep Silver	0.62	0.34	0.48	0.20	1.64
120	Metal Gear Solid: Ground Zeroes	2016	Action-Adventure	Konami Digital Entertainment	0.36	0.50	0.17	0.17	1.20
144	Injustice: Gods Among Us	2015	Fighting	Warner Bros. Interactive Entertainment	0.46	0.35	0.00	0.17	0.98
168	Hitman (2016)	2019	Action	Square Enix	0.24	0.36	0.06	0.11	0.78
192	LEGO The Hobbit	2016	Action-Adventure	Warner Bros. Interactive Entertainment	0.14	0.39	0.00	0.10	0.62
216	One Piece: Pirate Warriors 3	2017	Action	Namco Bandai Games	0.11	0.22	0.11	0.06	0.52
240	Trials Fusion	2016	Misc	Ubisoft	0.13	0.23	0.01	0.07	0.43
264	The King of Fighters XIV	2018	Fighting	Deep Silver	0.21	0.02	0.06	0.05	0.34
288	Onechanbara Z2: Chaos	2016	Action	NIS America	0.16	0.03	0.05	0.04	0.28
312	Digimon World: Next Order	2019	Role-Playing	Namco Bandai Games	0.13	0.05	0.02	0.04	0.24
336	Dragon's Dogma	2019	Role-Playing	Capcom	0.07	0.07	0.03	0.03	0.21
360	Putty Squad	2015	Platform	System 3	0.06	0.07	0.00	0.03	0.16
384	Shovel Knight	2017	Platform	Yacht Club Games	0.08	0.04	0.00	0.02	0.14
408	Teenage Mutant Ninja Turtles: Mutants in Manha...	2018	Action	Activision	0.07	0.03	0.00	0.02	0.12
432	Mega Man Legacy Collection	2018	Platform	Capcom	0.09	0.00	0.00	0.02	0.11
456	Toy Soldiers: War Chest	2017	Action	Ubisoft	0.07	0.00	0.00	0.02	0.09
480	Resident Evil 5 HD	2018	Action	Capcom	0.05	0.00	0.01	0.01	0.07
504	Steins;Gate 0	2017	Adventure	PQube	0.00	0.00	0.07	0.00	0.07
528	Exist Archive: The Other Side of the Sky	2017	Role-Playing	Aksys Games	0.02	0.00	0.03	0.00	0.06
552	SŽbastien Loeb Rally Evo	2018	Racing	Milestone S.r.l	0.00	0.04	0.00	0.01	0.05
576	Unbox: Newbie's Adventure	2019	Action	Merge Games	0.03	0.00	0.00	0.01	0.04
600	Raiden V	2019	Action	UFO Interactive	0.02	0.00	0.00	0.01	0.03
624	Shin Hayarigami 2	2018	Adventure	Nippon Ichi Software	0.00	0.00	0.03	0.00	0.03
648	Big Buck Hunter Arcade	2018	Shooter	GameMill Entertainment	0.02	0.00	0.00	0.00	0.02
672	Space Hulk: Deathwing	2017	Shooter	Focus Home Interactive	0.00	0.01	0.00	0.00	0.02
696	Lock's Quest	2019	Strategy	THQ Nordic	0.01	0.00	0.00	0.00	0.01

Subtasks 3.1.4 Stratified sampling for year and genre

In [8]:

```
1 genres=game.Genre.unique()
2 genre_group=game.groupby("Genre")
3 year_group=game.groupby("Year")
4 years=game.Year.unique()
5 strat_genre=pd.DataFrame()
6 strat_year=pd.DataFrame()
7
8 for i in genres:
9     garb=genre_group.get_group(i).sample(30,replace=True,ignore_index=True,random_state=10)
10    strat_genre=pd.concat([strat_genre,garb])
11 for i in years:
12     garb=year_group.get_group(i).sample(30,replace=True,ignore_index=True,random_state=10)
13     strat_year=pd.concat([strat_year,garb])
14
```

In [9]: 1 strat_year

Out[9]:

	Game	Year	Genre	Publisher	North_America_Sales_M\$	Europe_Sales_M\$	Japan_Sales_M\$	Other_Sales_M\$	Global_Sales_M\$
0	Middle-Earth: Shadow of Mordor	2016	Action	Warner Bros. Interactive Entertainment	1.03	1.60	0.05	0.51	3.19
1	Madden NFL 15	2016	Sports	EA Sports	1.58	0.25	0.00	0.41	2.25
2	Akiba's Trip 2	2016	Action	Acquire	0.09	0.00	0.02	0.02	0.13
3	Final Fantasy XIV: A Realm Reborn	2016	MMO	Square Enix	0.35	0.23	0.08	0.12	0.79
4	Pro Evolution Soccer 2015	2016	Sports	Konami Digital Entertainment	0.09	0.45	0.12	0.10	0.76
...
25	Just Dance 2014	2015	Music	Ubisoft	0.21	0.18	0.00	0.08	0.47
26	Lego Marvel Super Heroes	2015	Action	Warner Bros. Interactive Entertainment	0.59	1.00	0.01	0.30	1.91
27	NBA 2K14	2015	Sports	2K Sports	0.91	0.35	0.01	0.27	1.54
28	Just Dance 2014	2015	Music	Ubisoft	0.21	0.18	0.00	0.08	0.47
29	Madden NFL 25	2015	Sports	EA Sports	0.64	0.15	0.00	0.17	0.96

180 rows × 9 columns

In [10]: 1 strat_genre

Out[10]:

	Game	Year	Genre	Publisher	North_America_Sales_M\$	Europe_Sales_M\$	Japan_Sales_M\$	Other_Sales_M\$	Global_Sales_M\$
0	Far Cry 5	2020	Action	Ubisoft	1.44	1.73	0.15	0.62	3.95
1	ZombiU	2018	Action	Ubisoft	0.00	0.04	0.01	0.01	0.07
2	Dying Light	2017	Action	Warner Bros. Interactive	1.18	0.74	0.13	0.40	2.45
3	Fate/Extella: The Umbral Star	2018	Action	PQube	0.12	0.00	0.12	0.03	0.27
4	Dead Rising	2018	Action	Capcom	0.04	0.00	0.02	0.01	0.08
...
25	Ao no Kanata no Four Rhythm	2019	Visual Novel	Sprite	0.00	0.00	0.03	0.00	0.03
26	Chaos;Child	2017	Visual Novel	PQube	0.00	0.00	0.02	0.00	0.02
27	Psycho-Pass: Mandatory Happiness	2018	Visual Novel	NIS America	0.10	0.02	0.00	0.03	0.15
28	Ao no Kanata no Four Rhythm	2019	Visual Novel	Sprite	0.00	0.00	0.03	0.00	0.03
29	The Silver Case	2019	Visual Novel	NIS America	0.05	0.00	0.01	0.01	0.07

510 rows × 9 columns

Subtasks 3.1.5 Clustered sampling based on year and genre

```
In [11]: 1 genres=game.Genre.unique()
2 genre_group=game.groupby("Genre")
3 year_group=game.groupby("Year")
4 years=game.Year.unique()
5 clust_genre=pd.DataFrame()
6 clust_year=pd.DataFrame()
7
8 for i in genres:
9     garb=genre_group.get_group(i).sample(frac=0.2,replace=True,ignore_index=True,random_state=13)
10    clust_genre=pd.concat([clust_genre,garb])
11 for i in years:
12     garb=year_group.get_group(i).sample(frac=0.2,replace=True,ignore_index=True,random_state=13)
13    clust_year=pd.concat([clust_year,garb])
14
```

In [12]: 1 clust_year

Out[12]:

	Game	Year	Genre	Publisher	North_America_Sales_M\$	Europe_Sales_M\$	Japan_Sales_M\$	Other_Sales_M\$	Global_Sales_M\$
0	The Walking Dead: Season Two	2016	Misc	Square Enix	0.13	0.09	0.03	0.05	0.30
1	Infamous: First Light	2016	Action	Sony Computer Entertainment	0.00	0.03	0.00	0.00	0.03
2	LittleBigPlanet 3	2016	Platform	Sony Computer Entertainment	0.78	0.93	0.01	0.34	2.06
3	Alien: Isolation	2016	Adventure	Sega	0.29	0.59	0.05	0.16	1.09
4	MotoGP 14	2016	Racing	Milestone S.r.l.	0.05	0.11	0.00	0.03	0.20
...
38	Hasbro Family Fun Pack Conquest Edition	2018	Misc	Ubisoft	0.07	0.00	0.00	0.02	0.08
39	Berserk and the Band of the Hawk	2018	Action	Tecmo Koei	0.07	0.02	0.05	0.02	0.16
0	FIFA Soccer 14	2015	Sports	EA Sports	0.62	2.18	0.12	0.51	3.43
1	Call of Duty: Ghosts	2015	Shooter	Activision	1.79	1.64	0.05	0.69	4.17
2	Madden NFL 25	2015	Sports	EA Sports	0.64	0.15	0.00	0.17	0.96

142 rows × 9 columns

In [13]: 1 clust_genre

Out[13]:

	Game	Year	Genre	Publisher	North_America_Sales_M\$	Europe_Sales_M\$	Japan_Sales_M\$	Other_Sales_M\$	Global_Sales_M\$
0	Resident Evil: Revelations	2019	Action	Capcom	0.06	0.06	0.04	0.02	0.18
1	Outlast Trinity	2019	Action	Warner Bros. Interactive Entertainment	0.09	0.09	0.00	0.04	0.22
2	Uncharted: The Lost Legacy	2019	Action	Sony Interactive Entertainment	0.58	1.38	0.04	0.36	2.37
3	Assassin's Creed Chronicles	2018	Action	Ubisoft	0.02	0.05	0.02	0.01	0.11
4	ClaDun Sengoku	2019	Action	NIS America	0.02	0.00	0.00	0.00	0.03
...
3	Armello	2020	Strategy	League of Geeks	0.02	0.00	0.00	0.00	0.02
0	The Talos Principle	2017	Puzzle	Nighthawk Interactive	0.00	0.04	0.00	0.01	0.04
1	Angry Birds: Star Wars	2015	Puzzle	Activision	0.11	0.09	0.00	0.04	0.23
0	The Silver Case	2019	Visual Novel	NIS America	0.05	0.00	0.01	0.01	0.07
1	Psycho-Pass: Mandatory Happiness	2018	Visual Novel	NIS America	0.10	0.02	0.00	0.03	0.15

142 rows × 9 columns

Task3: To show measures of dispersion for samples and population

```
In [14]: 1 mean=[]
2 median=[]
3 std=[]
4 for i in range(1,6):
5     if(i==5):
6         mean.append(game["Global_Sales_M$"].mean())
7         median.append(game["Global_Sales_M$"].median())
8         std.append(game["Global_Sales_M$"].std())
9         measures=pd.DataFrame([mean,median,std],columns=["Samp_10","Samp_20","Samp_30","Samp_40","Population"],index=["Mean","Median","Standard_dev"])
10    else:
11        samp=game.sample(i*10,replace=True,ignore_index=True,random_state=13)
12        mean.append(samp["Global_Sales_M$"].mean())
13        median.append(samp["Global_Sales_M$"].median())
14        std.append(samp["Global_Sales_M$"].std())
15 measures
16
```

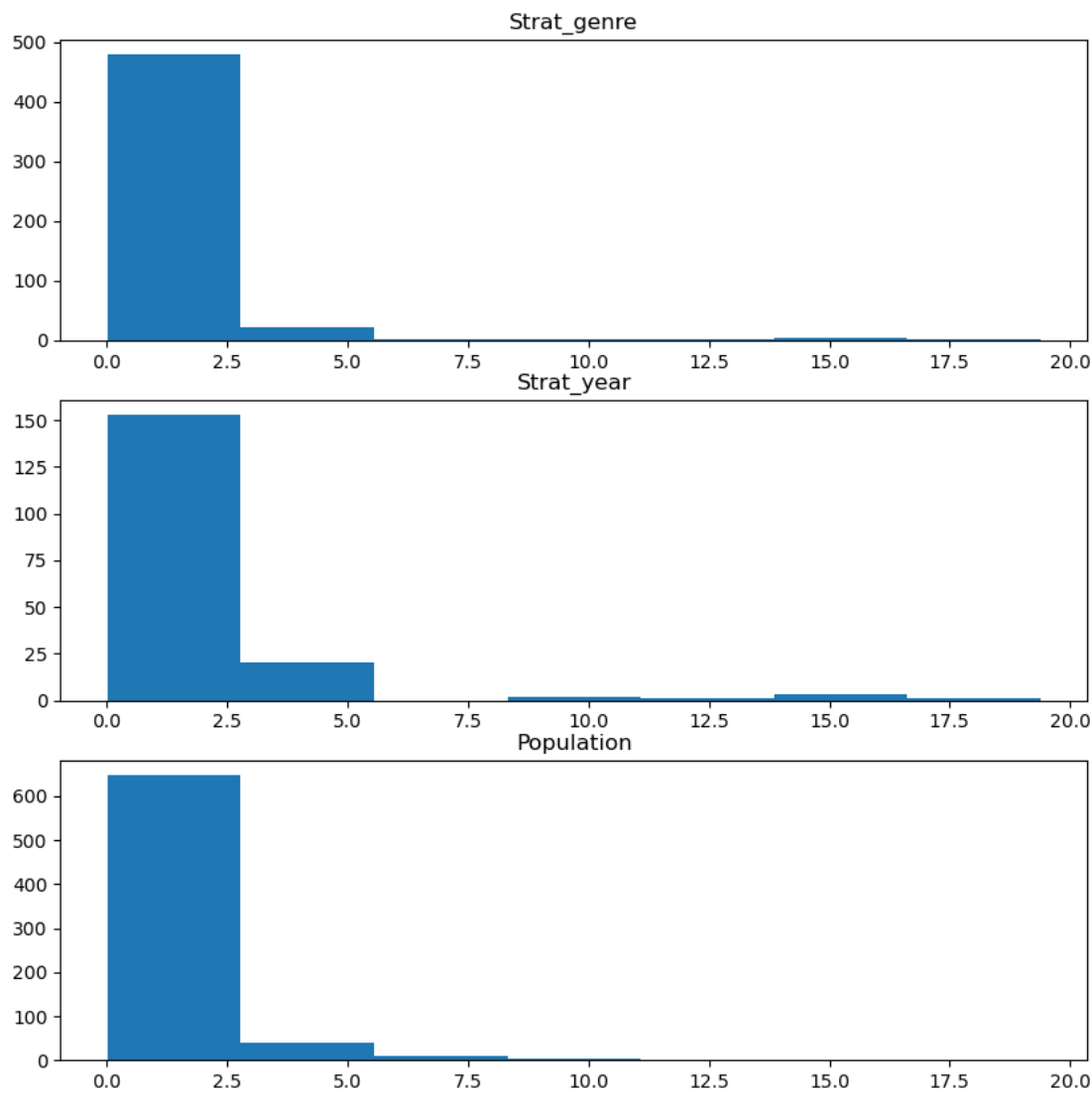
Out[14]:

	Samp_10	Samp_20	Samp_30	Samp_40	Population
Mean	0.507000	1.285500	1.197667	0.943000	0.839803
Median	0.305000	0.350000	0.215000	0.195000	0.170000
Standard_dev	0.656236	3.312141	3.020359	2.643848	1.853915

Task 3.2 To present stratified data and population data

```
In [15]: 1 fig,ax=plt.subplots(3,1,figsize=(10,10))
2 ax[0].hist(strat_genre["Global_Sales_M$"],bins=7)
3 ax[0].set_title("Strat_genre")
4 ax[1].hist(strat_year["Global_Sales_M$"],bins=7)
5 ax[1].set_title("Strat_year")
6 ax[2].hist(game["Global_Sales_M$"],bins=7)
7 ax[2].set_title("Population")
```

Out[15]: Text(0.5, 1.0, 'Population')



```
In [16]: """ From above tables and histograms most of the global sales has happened in the range 0.1 to 2.5 which means vast number of
gamers buy games between this price bracket of [0.1 ot 2.5]$ and there are very low customers ready to by beyond these prices
the standard deviation is steadiliy decreasing from sample 30 because its obvious that the more samples fall in the spread of data is
compromised between 0.0 to 2.5 the mean and median are clearly indicating that the data is a left skewed data negative skewness"""
```

Out[16]: ' From above tables and histograms most of the global sales has happened in the range 0.1 to 2.5 which means vast number of\ngamers buy games between this price bracket of [0.1 ot 2.5]\$ a
nd there are very low customers ready to by beyond these prices\nthe standard deviation is steadiliy decreasing from sample 30 because its obvious that the more samples fall in the spread
of data is\ncompromised between 0.0 to 2.5 the mean and median are clearly indicating that the data is a left skewed data negative skewness'