

# DSL1\_C5\_S3\_Challenge

## Task1

In [85]:

```
female = 0.46
work_part_time = 0.25
probability = (female*work_part_time)/0.17
print("probability of selecting part time working female :", probability)
```

probability of selecting part time working female : 0.676470588235294

## Task 2

In [86]:

```
Reducing_noise = 0.7
increasing_noise = 0.67
both = 0.57
print("probability of finding more storage guy is :", both/Reducing_noise)
```

probability of finding more storage guy is : 0.8142857142857143

## Task 3

In [2]:

```
import pandas as pd
import numpy as np
import statistics as st
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
```

In [3]:

```
game = pd.read_csv('E:\Aishwarya official\Aishwarya Data Scince\course 5\DS1_C5_S3_GamesSal
game
```

Out[3]:

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

709 rows × 9 columns

### 3.1.1 : random sampling without replacement

In [6]:

```
a = game.sample(20,ignore_index=True)
a
```

Out[6]:

	Game	Year	Genre	Publisher	North_America_Sales_M\$	Europe_Sales_M\$
0	Vikings: Wolves of Midgard	2019	Action	Kalypso Media	0.03	0.01
1	Game of Thrones (Telltale)	2017	Adventure	Telltale Games	0.05	0.00
2	Darksiders III	2020	Action	THQ Nordic	0.13	0.07
3	DmC: Devil May Cry	2017	Action	Capcom	0.12	0.13
4	Spider-Man (PS4)	2020	Action-Adventure	Sony Interactive Entertainment	3.64	3.39
5	Gundam Breaker 3	2018	Action	Namco Bandai Games	0.00	0.00
6	The Pillars of the Earth	2019	Adventure	Kalypso Media	0.02	0.00
7	Destiny 2	2019	Shooter	Activision	1.92	1.44
8	The Crew 2	2020	Racing	Ubisoft	0.22	0.42
9	FIFA Soccer 14	2015	Sports	EA Sports	0.62	2.18
10	Micro Machines: World Series	2019	Racing	Codemasters	0.03	0.16
11	Bravo Team VR	2020	Shooter	Sony Interactive Entertainment	0.12	0.09
12	MLB 15: The Show	2017	Sports	Sony Computer Entertainment	0.75	0.08
13	Raiden V	2019	Action	UFO Interactive	0.02	0.00
14	Lego Batman 3: Beyond Gotham	2016	Action	Warner Bros. Interactive Entertainment	0.37	0.42
15	Thief 4	2016	Action	Square Enix	0.28	0.41
16	FIFA 16	2017	Sports	EA Sports	1.15	5.77
17	Sine Mora	2019	Action	THQ Nordic	0.03	0.01
18	Valkyria Chronicles	2018	Role-Playing	Sega	0.14	0.08

	Game	Year	Genre	Publisher	North_America_Sales_M\$	Europe_Sales_M\$	
19	Raven's Cry	2017	Role-Playing	TopWare Interactive	0.00	0.01	

### 3.1.2: random sample with replacement

In [7]:

```
random_sam = game.sample(20, replace = True ,random_state =1)
random_sam
```

Out[7]:

	Game	Year	Genre	Publisher	North_America_Sales_M\$	Europe_Sales_M\$
37	Far Cry 5	2020	Action	Ubisoft	1.44	1.73
235	Tales of Zestiria	2017	Role-Playing	Namco Bandai Games	0.18	0.16
72	Need for Speed: Payback	2019	Racing	Electronic Arts	0.62	1.18
645	God Eater Off Shot: Tachibana Sakuya-hen Twin ...	2018	Action	Namco Bandai Games	0.00	0.00
144	Injustice: Gods Among Us	2015	Fighting	Warner Bros. Interactive Entertainment	0.46	0.35
129	LEGO Worlds	2019	Misc	Warner Bros. Interactive Entertainment	0.21	0.72
583	Dungeons 3	2019	Strategy	Kalypso Media	0.03	0.00
508	Dead or Alive Xtreme 3: Fortune	2018	Sports	Tecmo Koei	0.00	0.00
390	MXGP 2	2018	Racing	Milestone S.r.l.	0.05	0.06
281	Yooka-Laylee	2019	Platform	Sold Out	0.09	0.16
178	The LEGO Movie Videogame	2016	Action-Adventure	Warner Bros. Interactive Entertainment	0.26	0.32
276	Saint Seiya: Soldiers' Soul	2017	Fighting	Namco Bandai Games	0.00	0.21
254	Vampyr	2020	Role-Playing	Focus Home Interactive	0.12	0.19
357	MXGP 3: The Official Motocross Videogame	2019	Racing	Milestone	0.06	0.08
468	Earth Defense Force 4.1: The Shadow of New Des...	2017	Shooter	PQube	0.05	0.02

	Game	Year	Genre	Publisher	North_America_Sales_M\$	Europe_Sales_M\$
252	Resident Evil Zero	2018	Action-Adventure	Capcom	0.07	0.18
490	The Walking Dead: A Telltale Series Collection	2019	Adventure	Warner Bros. Interactive Entertainment	0.06	0.00
668	FlatOut 4: Total Insanity	2019	Racing	Bigben Interactive	0.01	0.00
398	Valentino Rossi: The Game	2018	Racing	Namco Bandai Games	0.05	0.05
562	That's You	2019	Party	Sony Interactive Entertainment	0.04	0.00

## Task 3.1.3:-Systematic Sampling

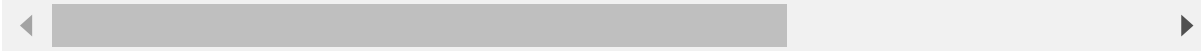
In [80]:

```
sys = np.arange(0, len(game), step=34)
game.iloc[sys]
```

Out[80]:

	Game	Year	Genre	Publisher	North_America_Sales_M\$	Europe_Sales_M\$	Ja
0	Grand Theft Auto V	2016	Action	Rockstar Games	6.06	9.71	
34	Assassin's Creed Origins	2019	Action	Ubisoft	1.22	2.11	
68	Uncharted: The Lost Legacy	2019	Action	Sony Interactive Entertainment	0.58	1.38	
102	Tomb Raider (2013)	2016	Action-Adventure	Square Enix	0.47	0.74	
136	The Evil Within II	2019	Action	Bethesda Softworks	0.31	0.49	
170	LEGO City Undercover	2019	Platform	Warner Bros. Interactive Entertainment	0.17	0.46	
204	NHL 18	2019	Sports	EA Sports	0.31	0.16	
238	Skylanders SWAP Force	2015	Misc	Activision	0.23	0.14	
272	J Stars Victory Vs.	2017	Fighting	Namco Bandai Games	0.11	0.14	
306	Just Dance 2017	2018	Misc	Ubisoft	0.08	0.12	
340	MotoGP 14E	2016	Racing	Milestone S.r.l.	0.05	0.11	
374	NASCAR Heat Evolution	2018	Racing	Dusenberry Martin Racing	0.12	0.00	
408	Teenage Mutant Ninja Turtles: Mutants in Manha...	2018	Action	Activision	0.07	0.03	
442	Natural Doctrine	2016	Role-Playing	NIS America	0.05	0.02	
476	Assetto Corsa	2018	Racing	505 Games	0.00	0.06	
510	Rugby 18	2019	Sports	Koch Media	0.05	0.00	
544	Saints Row: Gat out of Hell	2017	Action	Deep Silver	0.00	0.04	
578	Under Night In-Birth Exe:Latest	2019	Fighting	Arc System Works	0.01	0.00	

	Game	Year	Genre	Publisher	North_America_Sales_M\$	Europe_Sales_M\$	Ja
612	Romance of the Three Kingdoms 13 with Power-Up...	2019	Misc	Tecmo Koei	0.00	0.00	
646	PaRappa the Rapper Remastered	2019	Music	Sony Interactive Entertainment	0.00	0.00	
680	Hakoniwa Company Works	2019	Strategy	Nippon Ichi Software	0.00	0.00	



## 3.1.4 : Stratified Sampling

In [25]:

```
grouped_year = game.groupby('Year')
dept = game.Year.unique()
stratified_samp = pd.DataFrame()
```



In [26]:

```
for d in dept:
    d_data = grouped_year.get_group(d)
    sample_data = d_data.sample(n=5, replace=True, random_state=1)
    stratified_samp = stratified_samp.append(sample_data, ignore_index=True)
stratified_samp
```

Out[26]:

	Publisher	North_America_Sales_M\$	Europe_Sales_M\$	Japan_Sales_M\$	Other_Sales_M\$	C
re	Warner Bros. Interactive Entertainment	0.14	0.39	0.00	0.10	
ts	2K Sports	1.49	0.55	0.01	0.44	
g	Ubisoft	0.00	0.01	0.04	0.00	
on	Warner Bros. Interactive Entertainment	1.03	1.60	0.05	0.51	
g	Compile Heart	0.00	0.00	0.02	0.00	
ts	Konami Digital Entertainment	0.14	0.43	0.14	0.10	
on	Tecmo Koei	0.06	0.07	0.03	0.03	
on	Namco Bandai Games	0.00	0.01	0.00	0.00	
al el	PQube	0.00	0.00	0.02	0.00	
re	505 Games	0.04	0.09	0.00	0.02	
g	Deep Silver	0.18	0.28	0.00	0.09	
ts	Sony Interactive Entertainment	0.86	0.00	0.00	0.20	
ts	EA Sports	0.40	0.41	0.01	0.16	
g	Ubisoft	0.22	0.42	0.05	0.12	
re	Sony Interactive Entertainment	0.55	0.67	0.12	0.24	
on	Square Enix	0.24	0.36	0.06	0.11	
re	Compile Heart	0.00	0.00	0.01	0.00	

re	Publisher	North_America_Sales_M\$	Europe_Sales_M\$	Japan_Sales_M\$	Other_Sales_M\$	C
ig	Bigben Interactive	0.06	0.00	0.00	0.01	
ig	Capcom	0.19	0.00	0.02	0.04	
re	Warner Bros. Interactive Entertainment	0.06	0.00	0.00	0.01	
on	Ubisoft	0.14	0.52	0.02	0.12	
re	Bigben Interactive	0.05	0.00	0.00	0.01	
n-re	Sony Interactive Entertainment	0.14	0.09	0.00	0.05	
on	Ubisoft	0.00	0.04	0.01	0.01	
er	NIS America	0.06	0.00	0.00	0.01	
ig	Electronic Arts	0.75	1.04	0.03	0.35	
ic	Ubisoft	0.21	0.18	0.00	0.08	
sc	Activision	0.23	0.14	0.00	0.08	
ts	2K Sports	0.91	0.35	0.01	0.27	
m	System 3	0.06	0.07	0.00	0.03	



In [81]:

```
grouped_Genre = game.groupby('Genre')
Genre = game.Genre.unique()
stratified_samp1 = pd.DataFrame()
```

In [82]:

```

for d in Genre:
    d_data1 = grouped_Genre.get_group(d)
    sample_data1 = d_data1.sample(n=5, replace=True, random_state=1)
    stratified_samp1 = stratified_samp1.append(sample_data, ignore_index=True)
stratified_samp1

```

Out[82]:

	Game	Year	Genre	Publisher	North_America_Sales_M\$	Europe_Sales_M\$	Ja
0	Sébastien Loeb Rally Evo	2018	Racing	Milestone S.r.l	0.00	0.04	
1	FlatOut 4: Total Insanity	2019	Racing	Bigben Interactive	0.01	0.00	
2	Project CARS 2	2019	Racing	Namco Bandai Games	0.10	0.35	
3	F1 2018 (Codemasters)	2018	Racing	Codemasters	0.07	0.56	
4	WipEout Omega Collection	2019	Racing	Sony Interactive Entertainment	0.12	0.49	
...	...	...	...	...	...	...	
148	WipEout Omega Collection	2019	Racing	Sony Interactive Entertainment	0.12	0.49	
149	F1 2019	2019	Racing	Koch Media	0.04	0.47	
150	The Crew	2016	Racing	Ubisoft	0.44	1.04	
151	TrackMania Turbo	2018	Racing	Ubisoft	0.04	0.35	
152	Gran Turismo Sport	2019	Racing	Sony Interactive Entertainment	0.63	2.35	

153 rows × 9 columns



## Task 3.1.5 : Cluster Sampling

In [83]:

```

grouped_dep = game.groupby('Year')
dept = [2016, 2017]
Cluster_Sampling = pd.DataFrame()

```

In [84]:

```
for d in dept:
    d_data = grouped_dep.get_group(d)
    sample_data = d_data.sample(frac=0.2, replace=True, random_state=1)
    Cluster_Sampling = Cluster_Sampling.append(sample_data, ignore_index=True)
Cluster_Sampling
```

	The Peanuts							
	Movie:							
38	Snoopy's	2017	Platform	Activision	0.00	0.03	0.0	
	Grand							
	Adventure							
39	Payday 2	2017	Shooter	505 Games	0.12	0.14	0.0	
40	NBA 2K16	2017	Sports	2K Sports	2.56	0.66	0.0	
41	Rugby World	2017	Sports	Big Ben	0.00	0.05	0.0	
	Cup 2017			Interactive				
42	Sengoku							
	Basara 4:	2017	Action	Capcom	0.00	0.00	0.0	
	Sumoregi							

In [65]:

```
grouped_dep = game.groupby('Genre')
dept = ['Action','Racing']
Cluster_Sampling = pd.DataFrame()
```

In [66]:

```

for d in dept:
    d_data = grouped_dep.get_group(d)
    sample_data = d_data.sample(frac=0.2, replace=True, random_state=1)
    Cluster_Sampling = Cluster_Sampling.append(sample_data, ignore_index=True)
Cluster_Sampling

```

38	WipEout Omega Collection	2019	Racing	Sony Interactive Entertainment	0.12	0.49	0.00
39	F1 2019	2019	Racing	Koch Media	0.04	0.47	0.02
40	The Crew	2016	Racing	Ubisoft	0.44	1.04	0.03
41	TrackMania Turbo	2018	Racing	Ubisoft	0.04	0.35	0.00
42	Gran Turismo Sport	2019	Racing	Sony Interactive Entertainment	0.63	2.35	0.24

## Task 3

### 3.2.1 : mean

In [76]:

```

GS = game['Global_Sales_M$']
samp1 = GS.sample(n=10, replace=True, random_state=1)
samp2 = GS.sample(n=20, replace=True, random_state=1)
samp3 = GS.sample(n=30, replace=True, random_state=1)
samp4 = GS.sample(n=40, replace=True, random_state=1)

```

In [77]:

```
tables = [GS,samp1,samp2,samp3,samp4]
mean=[]
median=[]
std=[]
for i in tables:
    mean.append(i.mean())
    median.append(i.median())
    std.append(i.std())
pd.DataFrame([mean, median , std], columns=['GS_pop', 'samp_10', 'samp_20', 'samp_30', 'samp_40'])
```

Out[77]:

	GS_pop	samp_10	samp_20	samp_30	samp_40
mean	0.839803	0.925000	0.575500	0.810000	0.650250
std	0.170000	0.370000	0.230000	0.200000	0.155000
median	1.853915	1.263613	0.952274	1.721006	1.512594

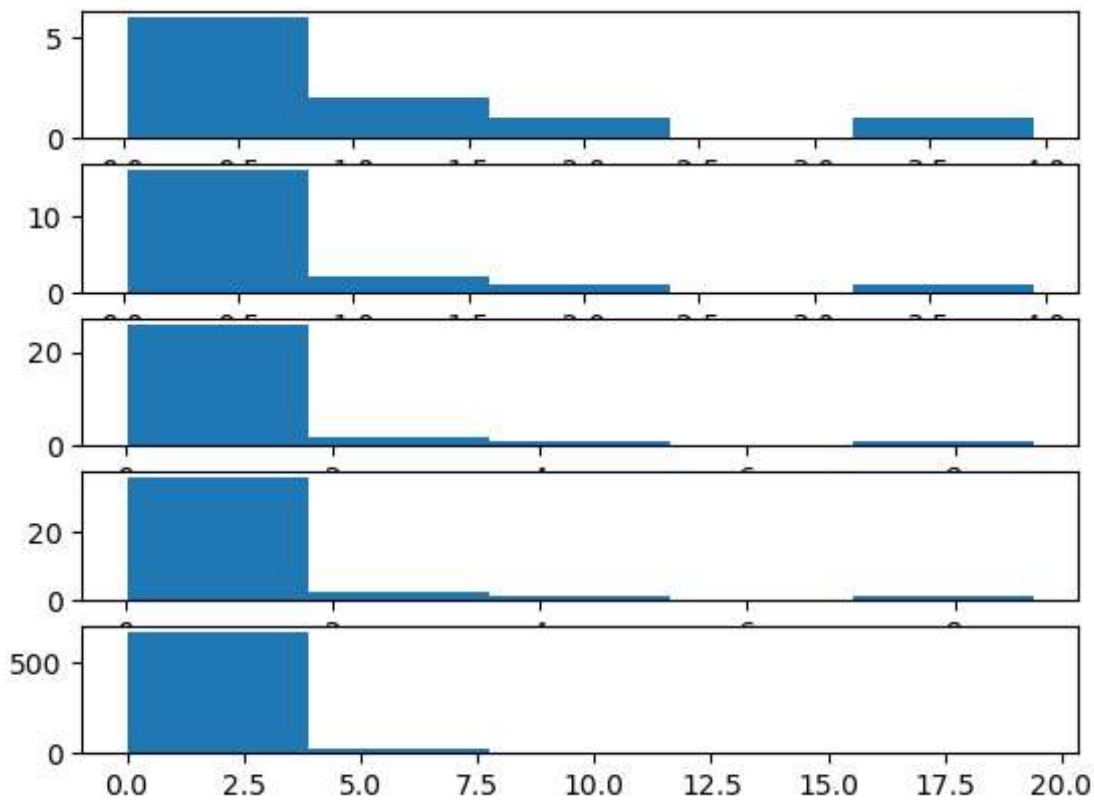
## Task 3.2

In [78]:

```
fig,ax=plt.subplots(5,1)
ax[0].hist(samp1,bins=5)
ax[1].hist(samp2,bins=5)
ax[2].hist(samp3,bins=5)
ax[3].hist(samp4,bins=5)
ax[4].hist(GS,bins=5)
```

Out[78]:

```
(array([670., 27., 7., 4., 1.]),
 array([1.0000e-02, 3.8860e+00, 7.7620e+00, 1.1638e+01, 1.5514e+01,
        1.9390e+01]),
 <BarContainer object of 5 artists>)
```



In [ ]: