

cramstack_eda_wreet_sarker

October 26, 2021

```
[1]: # Let's import the necessary libraries
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
[2]: # First let's check the game sales data by converting it into a dataframe
```

```
game_sales_df = pd.read_csv("game_sales_data.csv")
game_sales_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 19600 entries, 0 to 19599
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Rank                  19600 non-null  int64
1   Name                  19600 non-null  object
2   Platform              19600 non-null  object
3   Publisher             19600 non-null  object
4   Developer             19598 non-null  object
5   Critic_Score          9969 non-null   float64
6   User_Score            2223 non-null   float64
7   Total_Shipped         19600 non-null   float64
8   Year                  19600 non-null   int64
dtypes: float64(3), int64(2), object(4)
memory usage: 1.3+ MB
```

```
[3]: # Checking the first 5 values of the dataset
```

```
game_sales_df.head()
```

```
[3]:
```

	Rank	Name	Platform	Publisher	\
0	1	Wii Sports	Wii	Nintendo	
1	2	Super Mario Bros.	NES	Nintendo	
2	3	Counter-Strike: Global Offensive	PC	Valve	
3	4	Mario Kart Wii	Wii	Nintendo	
4	5	PLAYERUNKNOWN'S BATTLEGROUNDS	PC	PUBG Corporation	

	Developer	Critic_Score	User_Score	Total_Shipped	Year
0	Nintendo EAD	7.7	8.0	82.90	2006
1	Nintendo EAD	10.0	8.2	40.24	1985
2	Valve Corporation	8.0	7.5	40.00	2012
3	Nintendo EAD	8.2	9.1	37.32	2008
4	PUBG Corporation	8.6	4.7	36.60	2017

```
[4]: # Checking the bottom 5 values of dataset
```

```
game_sales_df.tail()
```

```
[4]:
```

	Rank	Name	Platform	\
19595	19594	FirePower for Microsoft Combat Flight Simulator 3	PC	
19596	19595	Tom Clancy's Splinter Cell	PC	
19597	19596	Ashita no Joe 2: The Anime Super Remix	PS2	
19598	19597	Tokyo Yamanote Boys for V: Main Disc	PSV	
19599	19598	NadePro!! Kisama no Seiyuu Yatte Miro!	PS2	

	Publisher	Developer	Critic_Score	User_Score	\
19595	GMX Media	Shockwave Productions	NaN	NaN	
19596	Ubisoft	Ubisoft	9.4	NaN	
19597	Capcom	Capcom	NaN	NaN	
19598	Rejet	Rejet	NaN	NaN	
19599	GungHo	GungHo Works	NaN	NaN	

	Total_Shipped	Year
19595	0.01	2004
19596	0.01	2003
19597	0.01	2002
19598	0.01	2017
19599	0.01	2009

```
[5]: # Checking how many unique Platforms are in the data set and their count
```

```
game_sales_df["Platform"].value_counts()
```

```
[5]:
```

DS	2251
PS2	2214
PC	1913
PS3	1368
Wii	1352
PSP	1313
X360	1302
PS	1205
PS4	953
GBA	857

```

XB      836
PSV     643
3DS     619
GC       561
XOne    539
N64     320
NS       299
SNES    241
SAT     175
WiiU    161
2600    133
NES     102
GB       94
DC       52
GEN      32
NG       12
PSN      10
XBL       9
GBC       9
WS        7
SCD       6
3DO       4
VC        3
PCE       2
Mob       1
Amig      1
PCFX      1
Name: Platform, dtype: int64

```

```

[6]: # Creating a new modified dataframe that contains only "PC", "PS4", "PS3",
      ↪ "XOne", "X360"

game_sales_modified = game_sales_df[game_sales_df["Platform"].isin(["PC",
      ↪ "PS4", "PS3", "XOne", "X360"])]

```

```

[7]: # Checking the modified dataframe

game_sales_modified

```

```

[7]:      Rank                                     Name Platform \
2         3      Counter-Strike: Global Offensive          PC
4         5      PLAYERUNKNOWN'S BATTLEGROUNDS             PC
5         6              Minecraft                         PC
11        12      Call of Duty: Modern Warfare             PS4
15        16      Kinect Adventures!                     X360
...        ...
19585    19584                                BC Kings          PC

```

19586	19585	Grand Theft Auto: Liberty City Stories / Vice ...	PS3
19588	19587	This Is The Police	XOne
19595	19594	FirePower for Microsoft Combat Flight Simulator 3	PC
19596	19595	Tom Clancy's Splinter Cell	PC

		Publisher	Developer	Critic_Score	\
2		Valve	Valve Corporation	8.0	
4		PUBG Corporation	PUBG Corporation	8.6	
5		Mojang	Mojang AB	10.0	
11		Activision	Infinity Ward	8.0	
15		Microsoft Game Studios	Good Science Studio	6.7	
...		
19585		Strategy First	Mascot Entertainment	NaN	
19586		Moss	Unknown	NaN	
19588		THQ Nordic	Weappy Studio	NaN	
19595		GMX Media	Shockwave Productions	NaN	
19596		Ubisoft	Ubisoft	9.4	

	User_Score	Total_Shipped	Year
2	7.5	40.00	2012
4	4.7	36.60	2017
5	7.8	33.15	2010
11	3.3	30.13	2019
15	NaN	24.00	2010
...
19585	NaN	0.01	2008
19586	NaN	0.01	2014
19588	NaN	0.01	2017
19595	NaN	0.01	2004
19596	NaN	0.01	2003

[6075 rows x 9 columns]

```
[8]: game_sales_modified.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 6075 entries, 2 to 19596
Data columns (total 9 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Rank            6075 non-null   int64
1   Name            6075 non-null   object
2   Platform        6075 non-null   object
3   Publisher       6075 non-null   object
4   Developer       6075 non-null   object
5   Critic_Score    3580 non-null   float64
6   User_Score      1541 non-null   float64
7   Total_Shipped   6075 non-null   float64
```

```

      8   Year          6075 non-null   int64
dtypes: float64(3), int64(2), object(4)
memory usage: 474.6+ KB

```

```
[9]: game_sales_modified["Platform"].value_counts()
```

```

[9]: PC          1913
     PS3          1368
     X360         1302
     PS4           953
     XOne          539
     Name: Platform, dtype: int64

```

1 1. Counting the number of distinct games

```

[10]: # Checking the unique names in the Name column of the modified dataframe

len(game_sales_modified["Name"].unique())

```

```
[10]: 3899
```

```
[11]: game_sales_modified.drop_duplicates(subset="Name")
```

```

[11]:
      Rank                                     Name Platform \
2         3      Counter-Strike: Global Offensive      PC
4         5      PLAYERUNKNOWN'S BATTLEGROUNDS      PC
5         6              Minecraft      PC
11        12      Call of Duty: Modern Warfare      PS4
15        16      Kinect Adventures!      X360
...      ...
19578    19577      RPG Maker MV      PS4
19585    19584      BC Kings      PC
19586    19585  Grand Theft Auto: Liberty City Stories / Vice ...  PS3
19595    19594  FirePower for Microsoft Combat Flight Simulator 3      PC
19596    19595      Tom Clancy's Splinter Cell      PC

      Publisher      Developer  Critic_Score \
2          Valve  Valve Corporation      8.0
4  PUBG Corporation  PUBG Corporation      8.6
5          Mojang    Mojang AB      10.0
11      Activision    Infinity Ward      8.0
15  Microsoft Game Studios  Good Science Studio      6.7
...      ...
19578    NIS America    Kadokawa Games      NaN
19585    Strategy First  Mascot Entertainment      NaN
19586          Moss      Unknown      NaN
19595      GMX Media  Shockwave Productions      NaN

```

19596		Ubisoft		Ubisoft	9.4
	User_Score	Total_Shipped	Year		
2	7.5	40.00	2012		
4	4.7	36.60	2017		
5	7.8	33.15	2010		
11	3.3	30.13	2019		
15	NaN	24.00	2010		
...		
19578	NaN	0.01	2019		
19585	NaN	0.01	2008		
19586	NaN	0.01	2014		
19595	NaN	0.01	2004		
19596	NaN	0.01	2003		

[3899 rows x 9 columns]

1.0.1 So there are 3899 distinct games in the modified game sales data

2. Games with unknow publishers

```
[12]: game_sales_modified[game_sales_modified["Publisher"] == "Unknown"]
```

```
[12]:
```

	Rank		Name	Platform	\
52	53		Terraria	PC	
602	601		Super Meat Boy	PC	
660	659		Infestation: Survivor Stories	PC	
953	952		How to Survive 2	PC	
1042	1041		Stick Fight: The Game	PC	
...	
18459	18458		Dark Parables: The Exiled Prince	PC	
18835	18834	The Treasures of Mystery Island 3 Pack - Save ...		PC	
19218	19217		PDC World Championship Darts 2008	PC	
19223	19222		The Tiny Bang Story	PC	
19449	19448		Euro Fishing	PC	

	Publisher	Developer	Critic_Score	User_Score	\
52	Unknown	Re-Logic	8.8	8.6	
602	Unknown	Team Meat	9.2	NaN	
660	Unknown	Hammerpoint Interactive	3.0	NaN	
953	Unknown	EKO Software	NaN	NaN	
1042	Unknown	Landfall West	NaN	NaN	
...	
18459	Unknown	Blue Tea Games	NaN	NaN	
18835	Unknown	Viva Media	NaN	NaN	
19218	Unknown	Mere Mortals	NaN	NaN	
19223	Unknown	Colibri Games	6.3	7.2	

19449	Unknown	Dovetail Games	NaN	NaN
-------	---------	----------------	-----	-----

	Total_Shipped	Year
52	14.00	2011
602	2.97	2010
660	2.80	2012
953	2.15	2016
1042	2.01	2017
...
18459	0.01	2011
18835	0.01	2011
19218	0.01	2008
19223	0.01	2011
19449	0.01	2015

[88 rows x 9 columns]

```
[13]: # Checking the length of dataframe that has an unknown publisher
len(game_sales_modified[game_sales_modified["Publisher"] == "Unknown"])
```

[13]: 88

2.0.1 There are 88 games with unknown publishers

3. Visualization of average critic score per developer

```
[14]: game_sales_modified
```

```
[14]:
```

	Rank	Name	Platform	\
2	3	Counter-Strike: Global Offensive	PC	
4	5	PLAYERUNKNOWN'S BATTLEGROUNDS	PC	
5	6	Minecraft	PC	
11	12	Call of Duty: Modern Warfare	PS4	
15	16	Kinect Adventures!	X360	
...	
19585	19584	BC Kings	PC	
19586	19585	Grand Theft Auto: Liberty City Stories / Vice ...	PS3	
19588	19587	This Is The Police	XOne	
19595	19594	FirePower for Microsoft Combat Flight Simulator 3	PC	
19596	19595	Tom Clancy's Splinter Cell	PC	

	Publisher	Developer	Critic_Score	\
2	Valve	Valve Corporation	8.0	
4	PUBG Corporation	PUBG Corporation	8.6	
5	Mojang	Mojang AB	10.0	
11	Activision	Infinity Ward	8.0	

15	Microsoft Game Studios	Good Science Studio	6.7
...
19585	Strategy First	Mascot Entertainment	NaN
19586	Moss	Unknown	NaN
19588	THQ Nordic	Weappy Studio	NaN
19595	GMX Media	Shockwave Productions	NaN
19596	Ubisoft	Ubisoft	9.4

	User_Score	Total_Shipped	Year
2	7.5	40.00	2012
4	4.7	36.60	2017
5	7.8	33.15	2010
11	3.3	30.13	2019
15	NaN	24.00	2010
...
19585	NaN	0.01	2008
19586	NaN	0.01	2014
19588	NaN	0.01	2017
19595	NaN	0.01	2004
19596	NaN	0.01	2003

[6075 rows x 9 columns]

```
[15]: # Grouping the data by the developers
```

```
game_sales_modified.groupby("Developer").mean()["Critic_Score"]
```

```
[15]: Developer
11 bit studios      8.150000
17-Bit              NaN
1C Company          7.200000
1C Maddox Games     8.300000
1C: Maddox Games    6.550000
...
n-Space             2.300000
stillalive studios  4.600000
team BitClub        NaN
thumbfood Ltd.      NaN
tri-Ace             7.333333
Name: Critic_Score, Length: 1452, dtype: float64
```

```
[16]: grouped_by_developer = game_sales_modified.groupby("Developer").mean()
```

```
[17]: grouped_by_developer.reset_index()["Developer"]
```

```
[17]: 0      11 bit studios
      1      17-Bit
```



```

2          1C Company
3      1C Maddox Games
4      1C: Maddox Games
...
1447          n-Space
1448  stillalive studios
1449          team BitClub
1450      thumbfood Ltd.
1451          tri-Ace
Name: Developer, Length: 1452, dtype: object

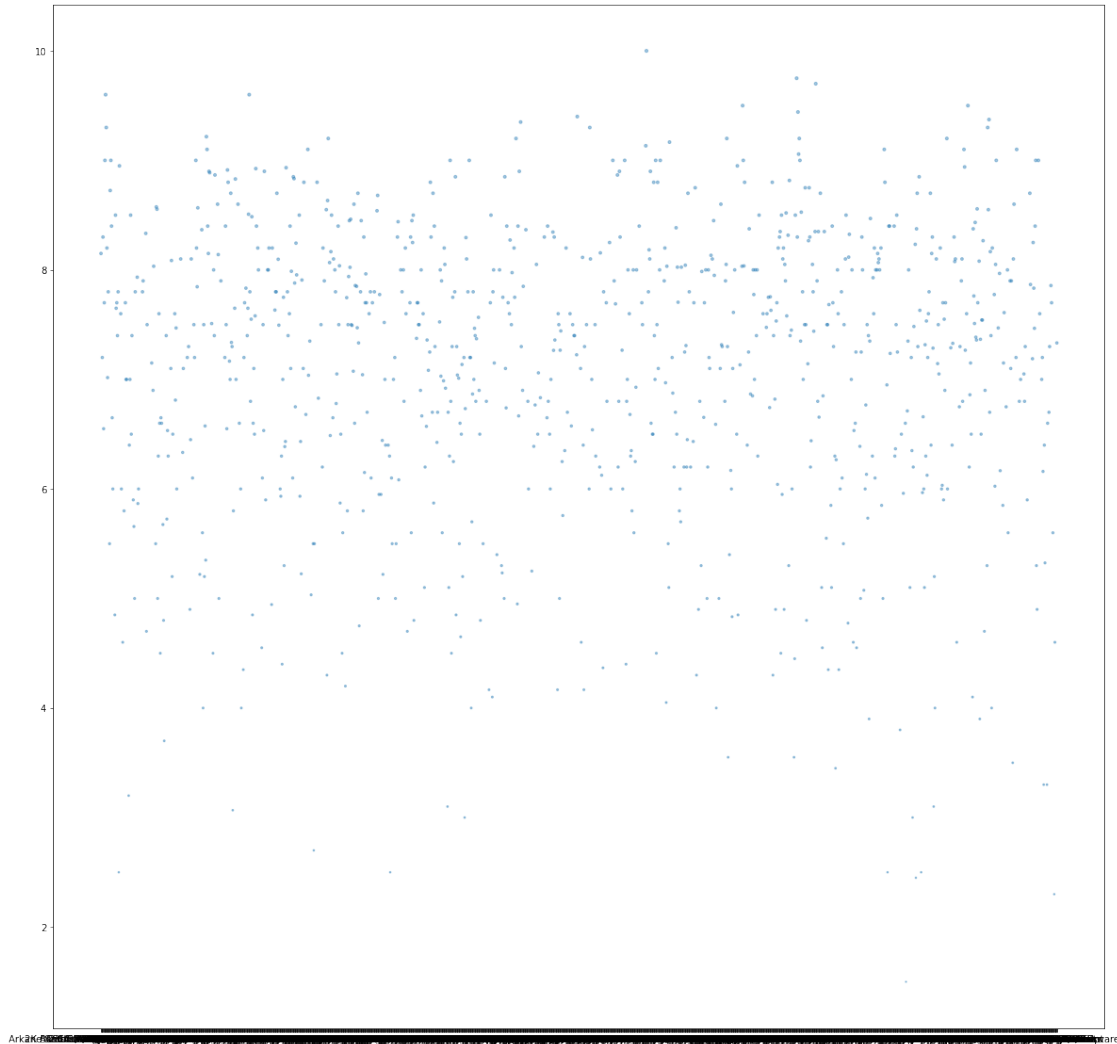
```

```
[18]: developers = grouped_by_developer.reset_index()["Developer"]
      critic_vals = game_sales_modified.groupby("Developer").mean()["Critic_Score"]
```

```
[19]: # Scatter plot to show Average critic score per developer

      plt.figure(figsize=(20, 20))
      plt.scatter(developers, critic_vals,
                  alpha=0.4, s = critic_vals)
```

```
[19]: <matplotlib.collections.PathCollection at 0x28898e77fd0>
```



4. Top 5 games with biggest differences in critic vs user score

```
[20]: # Creating a new column that contains the difference between critic score and
      ↪ user score

game_sales_modified["Critic_User_Difference"] =
      ↪ game_sales_modified["Critic_Score"] - game_sales_modified["User_Score"]
```

<ipython-input-20-5a241daa27db>:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
game_sales_modified["Critic_User_Difference"] =
game_sales_modified["Critic_Score"]- game_sales_modified["User_Score"]
```

```
[21]: game_sales_modified
```

```
[21]:
```

	Rank	Name	Platform	\
2	3	Counter-Strike: Global Offensive	PC	
4	5	PLAYERUNKNOWN'S BATTLEGROUNDS	PC	
5	6	Minecraft	PC	
11	12	Call of Duty: Modern Warfare	PS4	
15	16	Kinect Adventures!	X360	
...	
19585	19584	BC Kings	PC	
19586	19585	Grand Theft Auto: Liberty City Stories / Vice ...	PS3	
19588	19587	This Is The Police	XOne	
19595	19594	FirePower for Microsoft Combat Flight Simulator 3	PC	
19596	19595	Tom Clancy's Splinter Cell	PC	

	Publisher	Developer	Critic_Score	\
2	Valve	Valve Corporation	8.0	
4	PUBG Corporation	PUBG Corporation	8.6	
5	Mojang	Mojang AB	10.0	
11	Activision	Infinity Ward	8.0	
15	Microsoft Game Studios	Good Science Studio	6.7	
...	
19585	Strategy First	Mascot Entertainment	NaN	
19586	Moss	Unknown	NaN	
19588	THQ Nordic	Weappy Studio	NaN	
19595	GMX Media	Shockwave Productions	NaN	
19596	Ubisoft	Ubisoft	9.4	

	User_Score	Total_Shipped	Year	Critic_User_Difference
2	7.5	40.00	2012	0.5
4	4.7	36.60	2017	3.9
5	7.8	33.15	2010	2.2
11	3.3	30.13	2019	4.7
15	NaN	24.00	2010	NaN
...
19585	NaN	0.01	2008	NaN
19586	NaN	0.01	2014	NaN
19588	NaN	0.01	2017	NaN
19595	NaN	0.01	2004	NaN
19596	NaN	0.01	2003	NaN

```
[6075 rows x 10 columns]
```

```
[22]: # Sorting the dataframe by the critic and user score difference column and
      ↪ showing the top 5 values

game_sales_modified.sort_values(by="Critic_User_Difference", ascending=False).
      ↪ head()
```

```
[22]:
```

	Rank	Name	Platform	Publisher	Developer	\
157	158	FIFA 20	PS4	Electronic Arts	EA Sports	
3506	3505	Call of Duty: Ghosts	PC	Activision	Infinity Ward	
116	117	FIFA 19	PS4	Electronic Arts	EA Sports	
135	136	NBA 2K20	PS4	2K Sports	Visual Concepts	
506	505	NBA 2K18	PS4	2K Sports	Visual Concepts	

	Critic_Score	User_Score	Total_Shipped	Year	Critic_User_Difference
157	7.9	1.1	7.15	2019	6.8
3506	8.8	2.1	0.67	2013	6.7
116	8.3	1.7	9.15	2018	6.6
135	7.8	1.2	8.00	2019	6.6
506	8.1	1.7	3.34	2017	6.4