

Assignment 3_EDA

March 15, 2024

```
[289]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[290]: import warnings
warnings.filterwarnings("ignore")

#To ignore warnings which are not critical which have appeared in the notebook
↳when I made some modifications to the data/cleaning the data
```

```
[291]: deals_data = pd.read_csv('C:\PS5 Game Deals\historic_deals.csv')
deals_data
```

```
[291]:
```

	Title \
0	Suicide Squad: Kill the Justice League - Digit...
1	Horizon Forbidden West : Burning Shores (Engli...
2	Horizon Forbidden West Complete Edition (Simp...
3	UNCHARTED: Legacy of Thieves Collection (Simpl...
4	Marvel's Spider-Man: Miles Morales Ultimate Ed...
...	...
5237	Work from Home (Simplified Chinese, English, J...
5238	YUKI (English)
5239	ZOMB (English Ver.)
5240	Spelunker HD DELUXE (Simplified Chinese, Engli...
5241	Them's Fightin' Herds: Deluxe Edition PS4 & PS...

	Publisher \
0	Warner Bros. Interactive
1	Sony Interactive Entertainment
2	Sony Interactive Entertainment
3	Sony Interactive Entertainment
4	Sony Interactive Entertainment
...	...
5237	4D CREATIVITY & TECHNOLOGY
5238	Arvore Immersive Experiences, LLC
5239	Sabec Limited
5240	Tozai, Inc.
5241	RIPPLES ASIA VENTURE PTE LTD

	Link	DiscountPCT	\
0	https://store.playstation.com/en-hk/product/UP...	40%	
1	https://store.playstation.com/en-hk/product/EP...	35%	
2	https://store.playstation.com/en-hk/product/EP...	34%	
3	https://store.playstation.com/en-hk/product/HP...	60%	
4	https://store.playstation.com/en-hk/product/HP...	43%	
...	
5237	https://store.playstation.com/en-hk/product/HP...	30%	
5238	https://store.playstation.com/en-hk/product/UP...	30%	
5239	https://store.playstation.com/en-hk/product/UP...	80%	
5240	https://store.playstation.com/en-hk/product/JP...	50%	
5241	https://store.playstation.com/en-hk/product/UP...	NaN	

	OriginalPrice	DiscountPrice	Discount_Endtime	Rating	Rating_count	\
0	HK\$780.00	HK\$468.00	13/3/2024 02:59 PM UTC	4.31	8.6k	
1	HK\$148.00	HK\$96.20	20/3/2024 03:59 PM UTC	4.55	29	
2	HK\$468.00	HK\$308.88	20/3/2024 03:59 PM UTC	4.64	32k	
3	HK\$398.00	HK\$159.20	20/3/2024 03:59 PM UTC	4.74	179k	
4	HK\$568.00	HK\$323.76	20/3/2024 03:59 PM UTC	4.58	102k	
...	
5237	HK\$148.00	HK\$103.60	14/2/2024 03:59 PM UTC	3.67	3	
5238	HK\$148.00	HK\$103.60	14/2/2024 03:59 PM UTC	4.00	20	
5239	HK\$78.00	HK\$15.60	14/2/2024 03:59 PM UTC	3.24	33	
5240	HK\$198.00	HK\$99.00	14/2/2024 03:59 PM UTC	4.19	57	
5241	NaN	HK\$398.00	NaN	4.44	152	

	Genre	ReleaseDate
0	Shooter	29/1/2024
1	Action	18/4/2023
2	Action	5/10/2023
3	Action, Adventure	27/1/2022
4	Action	11/11/2020
...
5237	Action, Casual, Party	28/9/2022
5238	Adventure, Arcade, Action	16/11/2021
5239	Arcade, Arcade, Shooter	9/3/2022
5240	Action	14/7/2021
5241	Action	17/10/2022

[5242 rows x 11 columns]

```
[292]: deals_data.dtypes
```

```
[292]: Title          object
Publisher         object
Link              object
```

```
DiscountPCT      object
OriginalPrice    object
DiscountPrice    object
Discount_Endtime object
Rating           float64
Rating_count     object
Genre            object
ReleaseDate      object
dtype: object
```

```
[293]: #checking null values
sum_null = deals_data.isnull().sum().sort_values(ascending = False)
print(sum_null)
percentage_null = ((deals_data.isnull().sum())/(deals_data.isnull().
↳count())*100).sort_values(ascending = False)
print(percentage_null)
```

```
OriginalPrice      170
Discount_Endtime   169
DiscountPCT        121
Publisher           5
Title              4
DiscountPrice      4
Rating             4
Rating_count       4
ReleaseDate        3
Genre              1
Link               0
dtype: int64
OriginalPrice      3.243037
Discount_Endtime   3.223960
DiscountPCT        2.308279
Publisher           0.095383
Title              0.076307
DiscountPrice      0.076307
Rating             0.076307
Rating_count       0.076307
ReleaseDate        0.057230
Genre              0.019077
Link               0.000000
dtype: float64
```

```
[294]: #dropping null values and checking for them again
null_drop = deals_data.dropna(axis = 0)
sum_null_drop = null_drop.isnull().sum().sort_values(ascending = False)
sum_null_drop
```

```
[294]: Title          0
      Publisher      0
      Link          0
      DiscountPCT   0
      OriginalPrice 0
      DiscountPrice 0
      Discount_Endtime 0
      Rating        0
      Rating_count  0
      Genre         0
      ReleaseDate   0
      dtype: int64
```

```
[295]: #checking the number of data points/rows dropped
      count_values_before = deals_data.count()
      print(count_values_before)
      count_values_after = null_drop.count()
      print(count_values_after)
```

```
Title          5238
Publisher      5237
Link          5242
DiscountPCT   5121
OriginalPrice 5072
DiscountPrice 5238
Discount_Endtime 5073
Rating        5238
Rating_count  5238
Genre         5241
ReleaseDate   5239
dtype: int64
Title          5066
Publisher      5066
Link          5066
DiscountPCT   5066
OriginalPrice 5066
DiscountPrice 5066
Discount_Endtime 5066
Rating        5066
Rating_count  5066
Genre         5066
ReleaseDate   5066
dtype: int64
```

```
[296]: clean_df = null_drop
```

```
[297]: clean_df['OriginalPrice'] = clean_df['OriginalPrice'].str.replace('HK$', '')
clean_df['OriginalPrice'] = clean_df['OriginalPrice'].str.replace(',', '')
clean_df['OriginalPrice'] = clean_df['OriginalPrice'].astype(float)
print(clean_df['OriginalPrice'])
```

```
0      780.0
1      148.0
2      468.0
3      398.0
4      568.0
...
5236   238.0
5237   148.0
5238   148.0
5239    78.0
5240   198.0
Name: OriginalPrice, Length: 5066, dtype: float64
```

```
[298]: mean_op = clean_df['OriginalPrice'].mean()
print(mean_op)
median_op = clean_df['OriginalPrice'].median()
print(median_op)
```

```
203.74472956968023
148.0
```

```
[299]: clean_df['DiscountPrice'] = clean_df['DiscountPrice'].str.replace('HK$', '')
clean_df['DiscountPrice'] = clean_df['DiscountPrice'].str.replace(',', '')
clean_df['DiscountPrice'] = clean_df['DiscountPrice'].astype(float)
print(clean_df['DiscountPrice'])
```

```
0      468.00
1       96.20
2      308.88
3      159.20
4      323.76
...
5236    95.20
5237   103.60
5238   103.60
5239    15.60
5240    99.00
Name: DiscountPrice, Length: 5066, dtype: float64
```

```
[300]: mean_dp = clean_df['DiscountPrice'].mean()
print(mean_dp)
median_dp = clean_df['DiscountPrice'].median()
print(median_dp)
```

88.61213580734308

59.4

```
[301]: clean_df['Discount_Value'] = clean_df['OriginalPrice'] -  
        ↪clean_df['DiscountPrice']  
print(clean_df['Discount_Value'])
```

```
0      312.00  
1       51.80  
2      159.12  
3      238.80  
4      244.24
```

```
...  
5236   142.80  
5237    44.40  
5238    44.40  
5239    62.40  
5240    99.00
```

Name: Discount_Value, Length: 5066, dtype: float64

```
[302]: mean_dv = clean_df['Discount_Value'].mean()  
print(mean_dv)  
median_dv = clean_df['Discount_Value'].median()  
print(median_dv)  
  
#In all the three above column cases, median is probably the most reliable  
↪measure of central tendency/typical value of the above columns.
```

115.13259376233714

74.0

```
[303]: clean_df[['Date', 'Time', 'AM/PM', 'Time Zone']] = clean_df['Discount_Endtime'].  
        ↪str.split(' ', expand = True)  
print(clean_df['Date'])  
print(clean_df['Time'])  
print(clean_df['AM/PM'])  
print(clean_df['Time Zone'])
```

```
0      13/3/2024  
1      20/3/2024  
2      20/3/2024  
3      20/3/2024  
4      20/3/2024
```

```
...  
5236   14/2/2024  
5237   14/2/2024  
5238   14/2/2024  
5239   14/2/2024  
5240   14/2/2024
```

```
Name: Date, Length: 5066, dtype: object
```

```
0      02:59
1      03:59
2      03:59
3      03:59
4      03:59
```

```
...
```

```
5236   03:59
5237   03:59
5238   03:59
5239   03:59
5240   03:59
```

```
Name: Time, Length: 5066, dtype: object
```

```
0      PM
1      PM
2      PM
3      PM
4      PM
```

```
..
```

```
5236   PM
5237   PM
5238   PM
5239   PM
5240   PM
```

```
Name: AM/PM, Length: 5066, dtype: object
```

```
0      UTC
1      UTC
2      UTC
3      UTC
4      UTC
```

```
...
```

```
5236   UTC
5237   UTC
5238   UTC
5239   UTC
5240   UTC
```

```
Name: Time Zone, Length: 5066, dtype: object
```

```
[304]: clean_df['AM/PM'].unique()
```

```
[304]: array(['PM'], dtype=object)
```

```
[305]: clean_df = clean_df.drop(['AM/PM'],axis = 1)
#Dropping this column cause it's all PM and therefore not useful.
```

```
[306]: clean_df['Time Zone'].unique()
clean_df = clean_df.drop(['Time Zone'], axis = 1 )
```

```
#Dropping this column as it's all UTC and therefore not useful.
```

```
[307]: clean_df['Time'].unique()
```

```
[307]: array(['02:59', '03:59'], dtype=object)
```

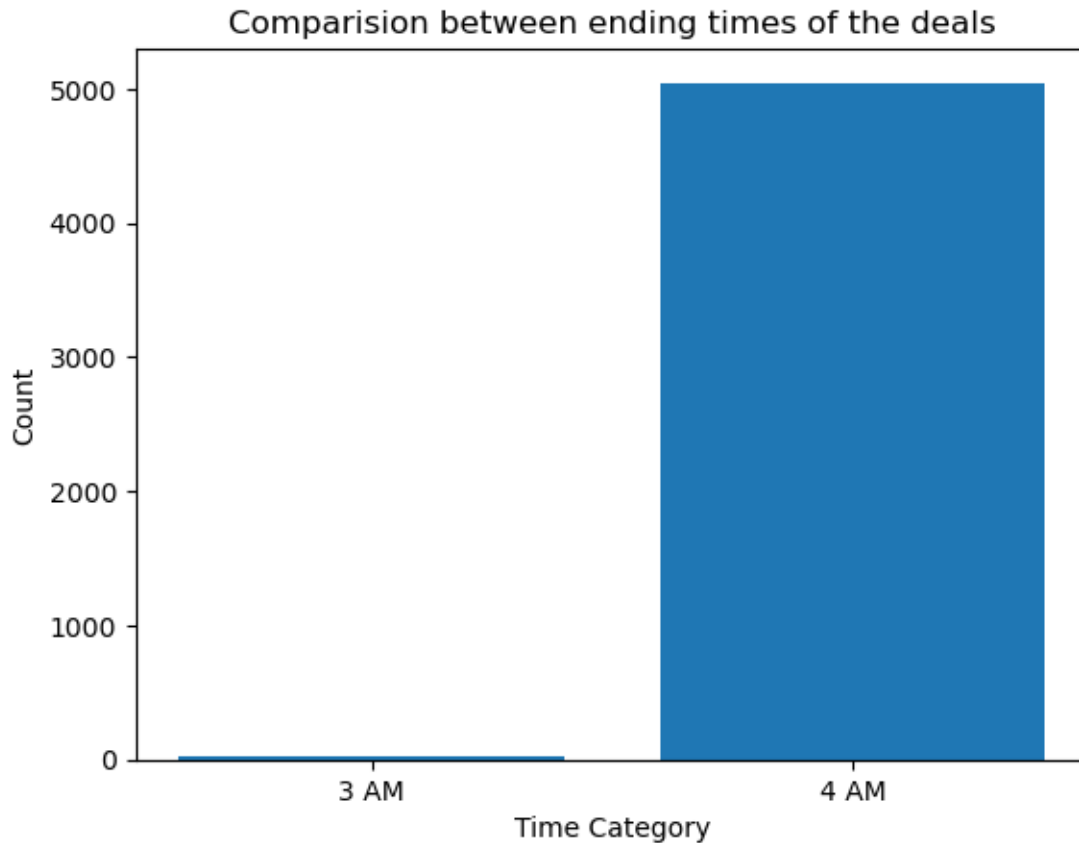
```
[308]: count_time = clean_df['Time'].value_counts()  
count_time
```

```
#It tells us that almost every deal lasts until 4 PM and only 20 deals last  
↳ untill 3 PM.
```

```
[308]: Time  
03:59    5046  
02:59     20  
Name: count, dtype: int64
```

```
[309]: time_cat = ['3 AM', '4 AM']  
count_3and4 = [20,5046]
```

```
pt.bar(time_cat, count_3and4)  
pt.xlabel('Time Category')  
pt.ylabel('Count')  
pt.title('Comparision between ending times of the deals')  
pt.show()
```

```
[310]: clean_df['Genre'].unique()
```

```
[310]: array(['Shooter', 'Action', 'Action, Adventure',
             'Action, Fighting, Adventure', 'Action, Role Playing Games',
             'Action, Adventure, Shooter',
             'Action, Adventure, Role Playing Games', 'Action, Casual, Family',
             'Family', 'Sport', 'Role Playing Games, Action',
             'Action, Simulation, Shooter', 'Action, Shooter, Unique',
             'Action, Adventure, Strategy', 'Adventure, Horror', 'Horror',
             'Role Playing Games',
             'Adventure, Role Playing Games, Role Playing Games',
             'Simulation, Simulation', 'Simulation', 'Simulation, Simulator',
             'Shooter, Action', 'Fighting, Action', 'Adventure',
             'Role Playing Games, Action, Role Playing Games',
             'Action, Fighting', 'Action, Casual, Fighting', 'Driving/Racing',
             'Driving/Racing, Simulation, Sport',
             'Role Playing Games, Adventure, Role Playing Games',
             'Simulation, Strategy', 'Action, Shooter', 'Fighting', 'Strategy',
             'Adventure, Action, Horror', 'Adventure, Horror, Action', 'Unique',
             'Action, Strategy', 'Role Playing Games, Strategy',
```

'Role Playing Games, Role Playing Games',
 'Driving/Racing, Simulation', 'Strategy, Strategy',
 'Action, Horror', 'Action, Party, Unique', 'Strategy, Shooter',
 'Action, Action', 'Action, Action, Role Playing Games',
 'Horror, Adventure', 'Arcade, Driving/Racing',
 'Driving/Racing, Simulation, Unique', 'Action, Adventure, Horror',
 'Adventure, Arcade, Family', 'Action, Puzzle',
 'Adventure, Adventure, Puzzle', 'Action, Adventure, Arcade',
 'Action, Adventure, Driving/Racing', 'Action, Horror, Unique',
 'Shooter, Action, Unique', 'Action, Unique',
 'Action, Adventure, Fighting',
 'Role Playing Games, Strategy, Simulation', 'Party',
 'Adventure, Family, Puzzle', 'Action, Party',
 'Action, Horror, Shooter', 'Action, Horror, Action',
 'Adventure, Puzzle, Action', 'Adventure, Casual, Simulation',
 'Action, Arcade, Fighting', 'Arcade, Driving/Racing, Simulation',
 'Shooter, Action, Role Playing Games', 'Action, Adventure, Unique',
 'Fighting, Action, Party', 'Shooter, Shooter',
 'Adventure, Role Playing Games', 'Puzzle', 'Sport, Sport',
 'Arcade', 'Arcade, Arcade, Fighting',
 'Role Playing Games, Simulation', 'Adventure, Puzzle',
 'Horror, Action, Horror', 'Music/Rhythm',
 'Horror, Adventure, Horror', 'Adventure, Unique, Adventure',
 'Shooter, Action, Shooter', 'Simulator', 'Action, Horror, Horror',
 'Arcade, Music/Rhythm', 'Adventure, Shooter, Shooter',
 'Action, Adventure, Simulator', 'Adventure, Action',
 'Action, Arcade', 'Action, Role Playing Games, Role Playing Games',
 'Action, Adventure, Family', 'Simulation, Adventure',
 'Action, Action, Shooter', 'Adventure, Simulation',
 'Action, Adventure, Simulation', 'Arcade, Simulation',
 'Action, Arcade, Shooter', 'Role Playing Games, Shooter',
 'Action, Music/Rhythm', 'Action, Shooter, Action',
 'Adventure, Horror, Adventure', 'Arcade, Driving/Racing, Sport',
 'Adventure, Puzzle, Unique', 'Adventure, Unique',
 'Unique, Shooter', 'Adventure, Casual', 'Action, Shooter, Shooter',
 'Adventure, Adventure, Adventure',
 'Adventure, Role Playing Games, Unique',
 'Action, Role Playing Games, Shooter',
 'Role Playing Games, Strategy, Adventure', 'Arcade, Action',
 'Driving/Racing, Simulation, Simulator',
 'Simulation, Simulation, Simulation', 'Action, Shooter, Strategy',
 'Action, Simulation', 'Unique, Adventure', 'Casual, Sport',
 'Action, Action, Action', 'Casual', 'Arcade, Family, Party',
 'Brain Training, Educational', 'Action, Family, Party',
 'Adventure, Casual, Puzzle',
 'Action, Role Playing Games, Simulation', 'Casual, Simulator',
 'Adventure, Adventure', 'Action, Sport',

'Arcade, Shooter, Shooter', 'Adventure, Music/Rhythm',
 'Arcade, Sport', 'Arcade, Horror', 'Action, Action, Fighting',
 'Unique, Adventure, Puzzle', 'Action, Family, Role Playing Games',
 'Adventure, Simulation, Simulator', 'Action, Adventure, Puzzle',
 'Puzzle, Adventure', 'Music/Rhythm, Arcade',
 'Role Playing Games, Puzzle', 'Puzzle, Sport', 'Casual, Family',
 'Casual, Family, Puzzle', 'Puzzle, Brain Training',
 'Simulation, Role Playing Games, Strategy',
 'Adventure, Adventure, Unique', 'Casual, Puzzle',
 'Action, Shooter, Simulation', 'Simulation, Casual, Strategy',
 'Action, Action, Arcade', 'Role Playing Games, Action, Adventure',
 'Educational, Family, Puzzle', 'Simulation, Adventure, Casual',
 'Arcade, Shooter', 'Puzzle, Strategy', 'Adventure, Action, Puzzle',
 'Adventure, Shooter', 'Brain Training, Casual, Puzzle',
 'Action, Casual, Party', 'Driving/Racing, Party, Unique',
 'Casual, Strategy, Strategy', 'Action, Horror, Adventure',
 'Brain Training, Puzzle, Strategy', 'Puzzle, Educational',
 'Simulation, Action, Puzzle', 'Action, Casual, Adventure',
 'Simulation, Educational, Family',
 'Shooter, Arcade, Role Playing Games', 'Horror, Puzzle',
 'Arcade, Unique', 'Brain Training, Family, Puzzle',
 'Puzzle, Unique', 'Simulation, Casual, Family', 'Unique, Arcade',
 'Action, Role Playing Games, Adventure',
 'Family, Simulation, Sport', 'Puzzle, Arcade',
 'Music/Rhythm, Arcade, Brain Training',
 'Action, Shooter, Adventure', 'Brain Training, Puzzle',
 'Simulation, Sport, Casual', 'Action, Adventure, Adventure',
 'Adventure, Brain Training, Puzzle', 'Horror, Shooter',
 'Casual, Action, Adventure', 'Arcade, Unique, Party',
 'Action, Arcade, Role Playing Games', 'Adventure, Simulator',
 'Fighting, Sport, Fighting', 'Horror, Action, Adventure',
 'Action, Adventure, Casual', 'Action, Adventure, Action',
 'Fighting, Fighting', 'Role Playing Games, Adventure, Strategy',
 'Role Playing Games, Role Playing Games, Simulation',
 'Action, Driving/Racing, Family', 'Casual, Action, Driving/Racing',
 'Role Playing Games, Adventure', 'Brain Training, Puzzle, Adult',
 'Driving/Racing, Action, Strategy',
 'Action, Casual, Driving/Racing', 'Action, Simulator',
 'Simulation, Action', 'Horror, Adventure, Puzzle',
 'Role Playing Games, Simulator', 'Casual, Puzzle, Family',
 'Arcade, Arcade, Shooter', 'Action, Arcade, Action',
 'Action, Fighting, Action',
 'Role Playing Games, Strategy, Strategy',
 'Action, Action, Simulation', 'Family, Simulation',
 'Action, Role Playing Games, Unique', 'Horror, Shooter, Action',
 'Driving/Racing, Sport', 'Adventure, Adventure, Shooter',
 'Strategy, Role Playing Games, Unique', 'Fighting, Unique',

'Adventure, Action, Unique',
 'Adventure, Role Playing Games, Shooter', 'Action, Family, Puzzle',
 'Adventure, Adventure, Action',
 'Adventure, Role Playing Games, Simulation',
 'Action, Action, Strategy',
 'Role Playing Games, Role Playing Games, Shooter',
 'Shooter, Simulation', 'Adventure, Strategy',
 'Driving/Racing, Simulation, Strategy', 'Simulation, Sport',
 'Adventure, Action, Adventure', 'Adventure, Action, Shooter',
 'Horror, Horror', 'Role Playing Games, Action, Unique',
 'Horror, Shooter, Music/Rhythm',
 'Fighting, Role Playing Games, Action', 'Puzzle, Puzzle, Shooter',
 'Arcade, Action, Arcade', 'Shooter, Strategy',
 'Driving/Racing, Simulation, Action', 'Adventure, Puzzle, Puzzle',
 'Action, Party, Strategy', 'Simulation, Casual',
 'Sport, Simulator', 'Action, Arcade, Sport',
 'Casual, Family, Party', 'Arcade, Puzzle, Action',
 'Shooter, Action, Arcade', 'Casual, Family, Driving/Racing',
 'Brain Training, Music/Rhythm', 'Adventure, Simulation, Strategy',
 'Puzzle, Simulation, Strategy', 'Strategy, Strategy, Strategy',
 'Casual, Simulation, Family', 'Puzzle, Family',
 'Simulation, Action, Driving/Racing',
 'Adventure, Adventure, Strategy',
 'Adventure, Brain Training, Casual',
 'Strategy, Arcade, Brain Training', 'Adventure, Unique, Adult',
 'Adventure, Casual, Driving/Racing', 'Strategy, Sport',
 'Horror, Puzzle, Simulator', 'Unique, Simulation',
 'Family, Family', 'Strategy, Adventure, Action',
 'Simulation, Sport, Strategy', 'Simulation, Action, Unique',
 'Simulation, Simulation, Strategy',
 'Adventure, Strategy, Simulation', 'Adventure, Family, Casual',
 'Family, Sport', 'Simulation, Casual, Sport', 'Action, Family',
 'Strategy, Simulation', 'Shooter, Arcade, Driving/Racing',
 'Strategy, Strategy, Unique', 'Shooter, Unique, Action',
 'Unique, Action, Shooter', 'Family, Puzzle',
 'Educational, Simulation, Driving/Racing', 'Puzzle, Puzzle',
 'Puzzle, Simulation, Arcade', 'Adventure, Educational',
 'Fighting, Adventure, Action', 'Casual, Family, Simulation',
 'Adventure, Adventure, Role Playing Games',
 'Family, Sport, Casual', 'Arcade, Arcade, Adult',
 'Fighting, Strategy, Action', 'Adventure, Simulation, Simulation',
 'Family, Party, Puzzle', 'Casual, Puzzle, Strategy',
 'Simulation, Strategy, Simulation',
 'Casual, Role Playing Games, Simulation',
 'Action, Adventure, Music/Rhythm', 'Adventure, Casual, Arcade',
 'Arcade, Puzzle', 'Action, Arcade, Adventure',
 'Adventure, Action, Arcade', 'Adventure, Horror, Casual',

'Action, Action, Adventure', 'Puzzle, Casual',
 'Fighting, Action, Shooter', 'Arcade, Adventure, Unique',
 'Arcade, Arcade', 'Action, Arcade, Puzzle', 'Quiz',
 'Casual, Simulation', 'Simulation, Casual, Music/Rhythm',
 'Arcade, Role Playing Games, Puzzle', 'Action, Horror, Strategy',
 'Action, Puzzle, Adventure', 'Puzzle, Simulation',
 'Action, Strategy, Action', 'Adventure, Puzzle, Adventure',
 'Strategy, Simulation, Unique', 'Action, Strategy, Strategy',
 'Adventure, Family, Role Playing Games',
 'Role Playing Games, Puzzle, Simulation',
 'Adventure, Unique, Puzzle',
 'Adventure, Action, Role Playing Games', 'Family, Casual, Sport',
 'Action, Simulation, Unique',
 'Adventure, Strategy, Role Playing Games',
 'Adventure, Action, Simulation', 'Action, Puzzle, Simulation',
 'Simulation, Educational, Sport', 'Arcade, Action, Adventure',
 'Adventure, Shooter, Casual', 'Strategy, Adventure',
 'Strategy, Role Playing Games', 'Adventure, Unique, Simulation',
 'Casual, Simulation, Adult',
 'Sport, Role Playing Games, Simulation',
 'Casual, Puzzle, Simulation', 'Simulation, Strategy, Unique',
 'Arcade, Casual, Driving/Racing', 'Casual, Simulation, Sport',
 'Adventure, Simulation, Party', 'Family, Party, Party',
 'Casual, Music/Rhythm, Unique', 'Simulation, Casual, Educational',
 'Strategy, Strategy, Simulation', 'Action, Adventure, Party',
 'Simulation, Adventure, Party', 'Simulation, Family, Strategy',
 'Arcade, Strategy', 'Arcade, Family, Sport', 'Simulation, Arcade',
 'Music/Rhythm, Action, Music/Rhythm',
 'Simulation, Strategy, Simulator',
 'Role Playing Games, Strategy, Role Playing Games',
 'Casual, Action', 'Adventure, Simulation, Casual',
 'Shooter, Driving/Racing, Driving/Racing',
 'Action, Driving/Racing, Shooter',
 'Adventure, Role Playing Games, Action',
 'Family, Adventure, Driving/Racing', 'Adventure, Casual, Action',
 'Unique, Adventure, Casual', 'Puzzle, Educational, Strategy',
 'Action, Casual', 'Casual, Simulation, Simulator',
 'Educational, Simulation, Simulator', 'Adventure, Casual, Casual',
 'Adventure, Arcade, Sport', 'Educational',
 'Arcade, Adventure, Action', 'Strategy, Adventure, Puzzle',
 'Family, Driving/Racing, Sport',
 'Simulation, Adventure, Educational',
 'Puzzle, Adventure, Brain Training', 'Shooter, Simulator',
 'Action, Casual, Simulation', 'Strategy, Adventure, Strategy',
 'Strategy, Simulator', 'Brain Training, Quiz',
 'Adventure, Adventure, Family', 'Simulation, Unique, Simulator',
 'Party, Driving/Racing', 'Adventure, Action, Strategy',

```

'Action, Driving/Racing, Driving/Racing', 'Puzzle, Adult',
'Horror, Music/Rhythm, Puzzle', 'Fitness, Music/Rhythm',
'Party, Simulator', 'Family, Casual, Puzzle',
'Horror, Horror, Puzzle', 'Action, Arcade, Driving/Racing',
'Puzzle, Role Playing Games', 'Driving/Racing, Action, Arcade',
'Adventure, Sport', 'Action, Arcade, Party',
'Action, Driving/Racing', 'Adventure, Family',
'Puzzle, Adventure, Unique', 'Arcade, Party, Driving/Racing',
'Puzzle, Brain Training, Puzzle', 'Unique, Puzzle',
'Role Playing Games, Unique', 'Board Game',
'Action, Fighting, Role Playing Games',
'Unique, Adventure, Unique', 'Fighting, Action, Fighting',
'Adventure, Unique, Action', 'Puzzle, Action',
'Action, Action, Puzzle', 'Action, Fighting, Unique',
'Action, Simulation, Strategy', 'Strategy, Simulation, Simulator',
'Fighting, Fighting, Fighting', 'Arcade, Fighting, Action',
'Action, Family, Action', 'Role Playing Games, Shooter, Unique',
'Music/Rhythm, Party', 'Party, Puzzle',
'Family, Educational, Puzzle', 'Casual, Music/Rhythm, Puzzle',
'Action, Arcade, Casual', 'Arcade, Action, Shooter',
'Action, Puzzle, Adult', 'Strategy, Strategy, Simulator',
'Adventure, Horror, Puzzle', 'Horror, Adventure, Adventure',
'Action, Strategy, Role Playing Games',
'Adventure, Casual, Unique', 'Role Playing Games, Horror, Puzzle',
'Adventure, Role Playing Games, Strategy', 'Simulation, Unique',
'Action, Family, Adventure', 'Simulation, Casual, Driving/Racing',
'Simulation, Adventure, Strategy', 'Arcade, Adventure',
'Music/Rhythm, Sport', 'Family, Puzzle, Unique',
'Adventure, Arcade, Simulator', 'Fighting, Action, Arcade',
'Casual, Family, Music/Rhythm', 'Casual, Arcade, Adult',
'Unique, Action', 'Arcade, Fighting', 'Casual, Arcade',
'Adventure, Simulation, Role Playing Games',
'Arcade, Role Playing Games', 'Adventure, Arcade, Action'],
dtype=object)

```

```

[311]: clean_df['DiscountPCT'] = clean_df['DiscountPCT'].str.replace('%','')
clean_df['DiscountPCT'] = pd.to_numeric(clean_df['DiscountPCT'],errors =_
↳'coerce')

```

'''We encountered non numeric/integer values in DiscountPCT column and what we_
↳did here is removing % part of the values
and then when trying to explicitly convert object to integer values using_
↳astype(int), we encounter an error where there
are still values in this column which are not something that can be converted_
↳into numeric values. Which is why we use
to_numeric to convert the values which can be converted into numeric and the_
↳other values which be converted into NaN values

```
through errors parameter set to 'coerce' which can be removed using dropna_
↳function later'''
```

[311]: "We encountered non numeric/integer values in DiscountPCT column and what we did here is removing % part of the values\nand then when trying to explicitly convert object to integer values using astype(int), we encounter an error where there\nare still values in this column which are not something that can be converted into numeric values. Which is why we use\nto_numeric to convert the values which can be converted into numeric and the other values which be converted into NaN values\nthrough errors parameter set to 'coerce' which can be removed using dropna function later"

```
[312]: df = clean_df.dropna(axis = 0)
print(df.count())

'''dropping the rows with NaN values (which are present due to the earlier_
↳modification) using dropna(). We see that
only 6 values were non-numerical values in the DiscountPCT column. And all the_
↳rows with those values were dropped.'''
```

```
Title          5060
Publisher       5060
Link            5060
DiscountPCT     5060
OriginalPrice   5060
DiscountPrice   5060
Discount_Endtime 5060
Rating          5060
Rating_count    5060
Genre           5060
ReleaseDate     5060
Discount_Value  5060
Date            5060
Time            5060
dtype: int64
```

[312]: 'dropping the rows with NaN values (which are present due to the earlier modification) using dropna(). We see that\nonly 6 values were non-numerical values in the DiscountPCT column. And all the rows with those values were dropped.'

```
[313]: #Checking if low rated games have more discount value
pt.figure(figsize= (20,20))
pt.scatter(df['Rating'],df['Discount_Value'])
pt.xlabel('Rating')
pt.ylabel('Discount Value in HK$')
pt.show()
```

'''It would seem that for a lot of the games which are in the nearly 4 to 5
↳range have a minimum 200 discount
and possibly around 300 as the maximum discount. However the games which are
↳have a rating of 1 seem to have low
discount values with a couple of outliers with discount values at 200 or more
↳HK\$ (Hong Kong dollar).

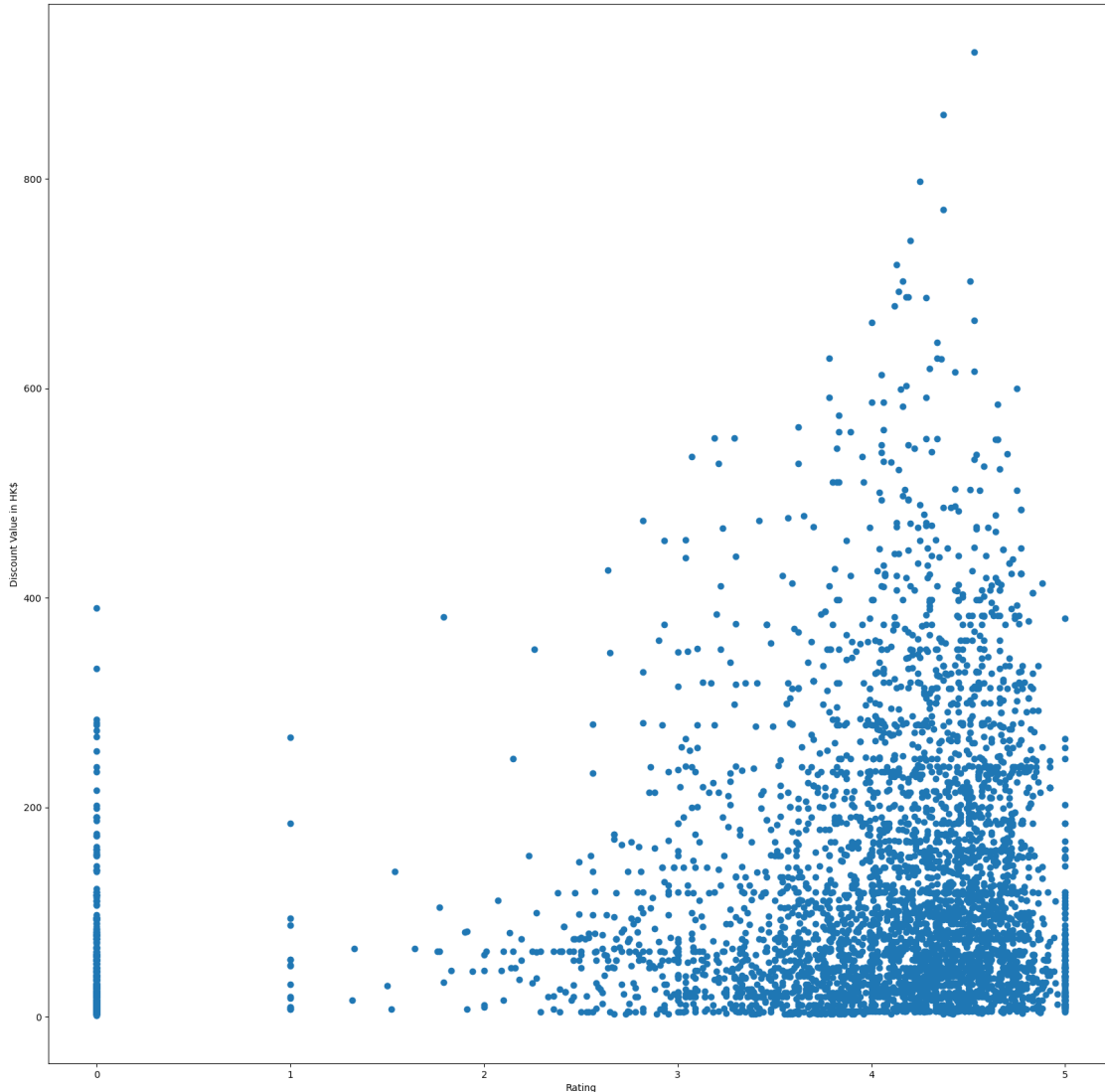
A majority of 0 rating games have discount values between 0 to 200 HK\$ with the
↳minority between 200-400 HK\$.

Games which are in the rating between 1 to 2 range have very low discounts as
↳well.

We can hypothesize that games
which are not well received and have low ratings do not get great discount
↳values, that is, they do not fall lower than their
original pricing as much as other games seem to do.

There are some outliers which get extremely great discount values and are also
↳well received (4 to 5 rating range)

'''

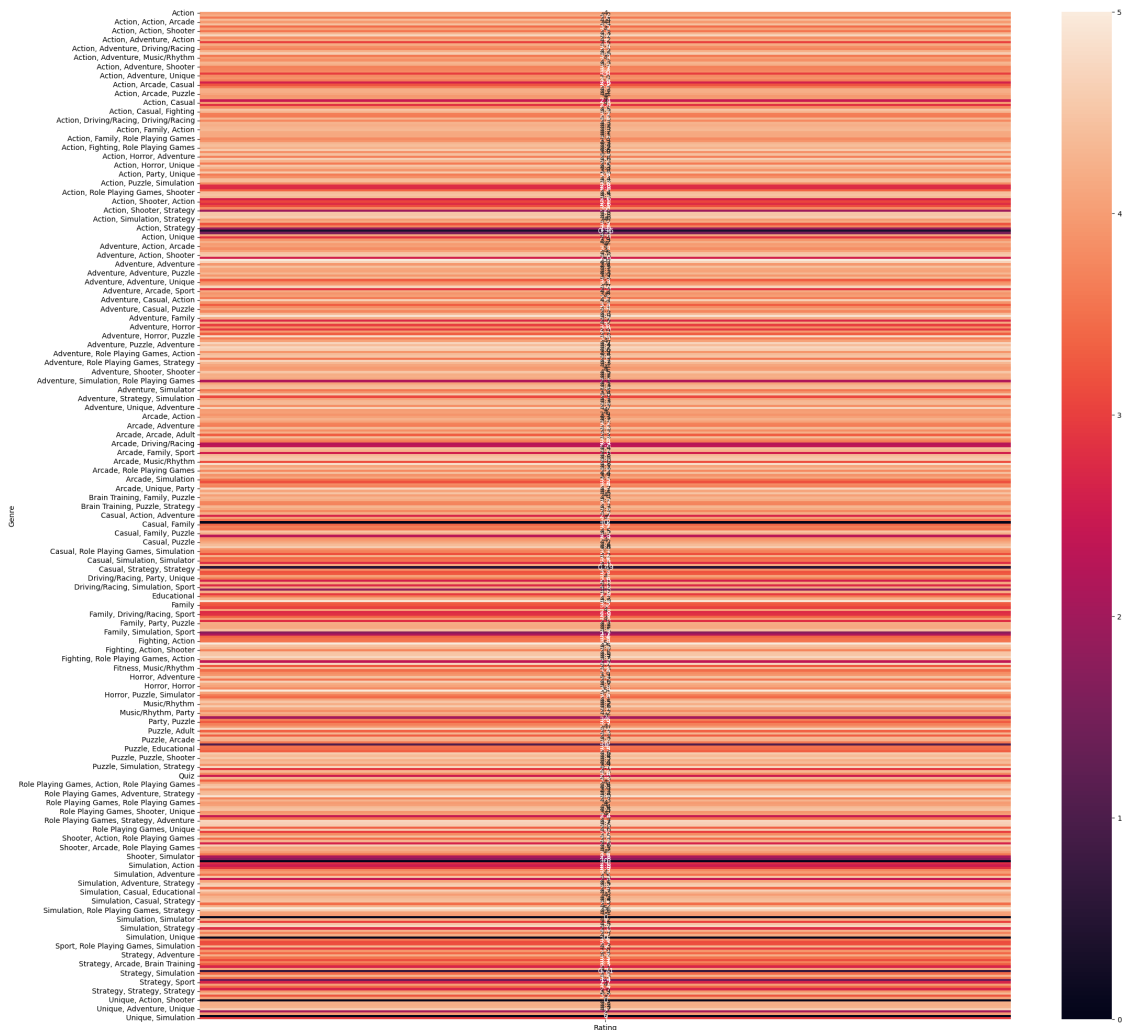


[313]: 'It would seem that for a lot of the games which are in the nearly 4 to 5 range have a minimum 200 discount \nand possibly around 300 as the maximum discount. However the games which are have a rating of 1 seem to have low \ndiscount values with a couple of outliers with discount values at 200 or more HK\$ (Hong Kong dollar). \n\nA majority of 0 rating games have discount values between 0 to 200 HK\$ with the minority between 200-400 HK\$.\n\nGames which are in the rating between 1 to 2 range have very low discounts as well. \n\nWe can hypothesize that games\nwhich are not well received and have low ratings do not get great discount values, that is, they do not fall lower than their\noriginal pricing as much as other games seem to do.\n\nThere are some outliers which get extremely great discount values and are also well received (4 to 5 rating range)\n\n'

```
[314]: df_ratings_genre = df[['Genre', 'Rating']]
```

```
[315]: pt.figure(figsize = (25,25))
sns.heatmap(df_ratings_genre.pivot_table(index = 'Genre'), annot = True)
pt.show()
```

'''There do seem to be a combination of genres which are very unpopular as seen by the dark cell lines in the heatmap. However, it's hard to tell which combinations exactly are rated low, even with zooming in further'''



[315]: "There do seem to be a combination of genres which are very unpopular as seen by the dark cell lines in the heatmap. \nHowever, it's hard to tell which combinations exactly are rated low, even with zooming in further"