

Assignment 7

March 7, 2021

0.0.1 Portfolio assignment 7

15 min: Look at the histogram of at least 2 columns with numerical data in the dataset that you chose in portfolio assignment 4. Do you recognise the distribution? Does it look like a uniform or normal distribution or something else? If it doesn't look like a uniform or normal distribution, take a quick look here to see if you can find the distribution shape: <https://www.itl.nist.gov/div898/handbook/eda/section3/eda366.htm>

```
[1]: import pandas as pd
import numpy as np
```

```
[2]: steam = pd.read_csv(r"steam_games.csv")
steam.head()
```

```
[2]:
```

	url	types	\
0	https://store.steampowered.com/app/379720/DOOM/	app	
1	https://store.steampowered.com/app/578080/PLAY...	app	
2	https://store.steampowered.com/app/637090/BATT...	app	
3	https://store.steampowered.com/app/221100/DayZ/	app	
4	https://store.steampowered.com/app/8500/EVE_On...	app	

	name	\
0	DOOM	
1	PLAYERUNKNOWN'S BATTLEGROUNDS	
2	BATTLETECH	
3	DayZ	
4	EVE Online	

	desc_snippet	\
0	Now includes all three premium DLC packs (Unto...	
1	PLAYERUNKNOWN'S BATTLEGROUNDS is a battle roya...	
2	Take command of your own mercenary outfit of '...	
3	The post-soviet country of Chernarus is struck...	
4	EVE Online is a community-driven spaceship MMO...	

	recent_reviews	\
0	Very Positive,(554),- 89% of the 554 user revi...	
1	Mixed,(6,214),- 49% of the 6,214 user reviews ...	

2 Mixed,(166),- 54% of the 166 user reviews in t...
 3 Mixed,(932),- 57% of the 932 user reviews in t...
 4 Mixed,(287),- 54% of the 287 user reviews in t...

all_reviews release_date \
 0 Very Positive,(42,550),- 92% of the 42,550 use... May 12, 2016
 1 Mixed,(836,608),- 49% of the 836,608 user revi... Dec 21, 2017
 2 Mostly Positive,(7,030),- 71% of the 7,030 use... Apr 24, 2018
 3 Mixed,(167,115),- 61% of the 167,115 user revi... Dec 13, 2018
 4 Mostly Positive,(11,481),- 74% of the 11,481 u... May 6, 2003

developer publisher \
 0 id Software Bethesda Softworks,Bethesda Softworks
 1 PUBG Corporation PUBG Corporation,PUBG Corporation
 2 Harebrained Schemes Paradox Interactive,Paradox Interactive
 3 Bohemia Interactive Bohemia Interactive,Bohemia Interactive
 4 CCP CCP,CCP

popular_tags \
 0 FPS,Gore,Action,Demons,Shooter,First-Person,Gr...
 1 Survival,Shooter,Multiplayer,Battle Royale,PvP...
 2 Mechs,Strategy,Turn-Based,Turn-Based Tactics,S...
 3 Survival,Zombies,Open World,Multiplayer,PvP,Ma...
 4 Space,Massively Multiplayer,Sci-fi,Sandbox,MMO...

game_details \
 0 Single-player,Multi-player,Co-op,Steam Achieve...
 1 Multi-player,Online Multi-Player,Stats
 2 Single-player,Multi-player,Online Multi-Player...
 3 Multi-player,Online Multi-Player,Steam Worksho...
 4 Multi-player,Online Multi-Player,MMO,Co-op,Onl...

languages achievements \
 0 English,French,Italian,German,Spanish - Spain,... 54.0
 1 English,Korean,Simplified Chinese,French,Germa... 37.0
 2 English,French,German,Russian 128.0
 3 English,French,Italian,German,Spanish - Spain,... NaN
 4 English,German,Russian,French NaN

genre \
 0 Action
 1 Action,Adventure,Massively Multiplayer
 2 Action,Adventure,Strategy
 3 Action,Adventure,Massively Multiplayer
 4 Action,Free to Play,Massively Multiplayer,RPG,...

game_description \
 0

```

0 About This Game Developed by id software, the...
1 About This Game PLAYERUNKNOWN'S BATTLEGROUN...
2 About This Game From original BATTLETECH/Mec...
3 About This Game The post-soviet country of Ch...
4 About This Game

mature_content \
0 NaN
1 Mature Content Description The developers de...
2 NaN
3 NaN
4 NaN

minimum_requirements \
0 Minimum:,OS:,Windows 7/8.1/10 (64-bit versions...
1 Minimum:,Requires a 64-bit processor and opera...
2 Minimum:,Requires a 64-bit processor and opera...
3 Minimum:,OS:,Windows 7/8.1 64-bit,Processor:,I...
4 Minimum:,OS:,Windows 7,Processor:,Intel Dual C...

recommended_requirements original_price \
0 Recommended:,OS:,Windows 7/8.1/10 (64-bit vers... $19.99
1 Recommended:,Requires a 64-bit processor and o... $29.99
2 Recommended:,Requires a 64-bit processor and o... $39.99
3 Recommended:,OS:,Windows 10 64-bit,Processor:,... $44.99
4 Recommended:,OS:,Windows 10,Processor:,Intel i... Free

discount_price
0 $14.99
1 NaN
2 NaN
3 NaN
4 NaN

```

0.1 Test 1 - Distribution in achievements

```
[3]: steam[['name', 'achievements']].sort_values(by='achievements', ascending=False)
```

```
[3]:
```

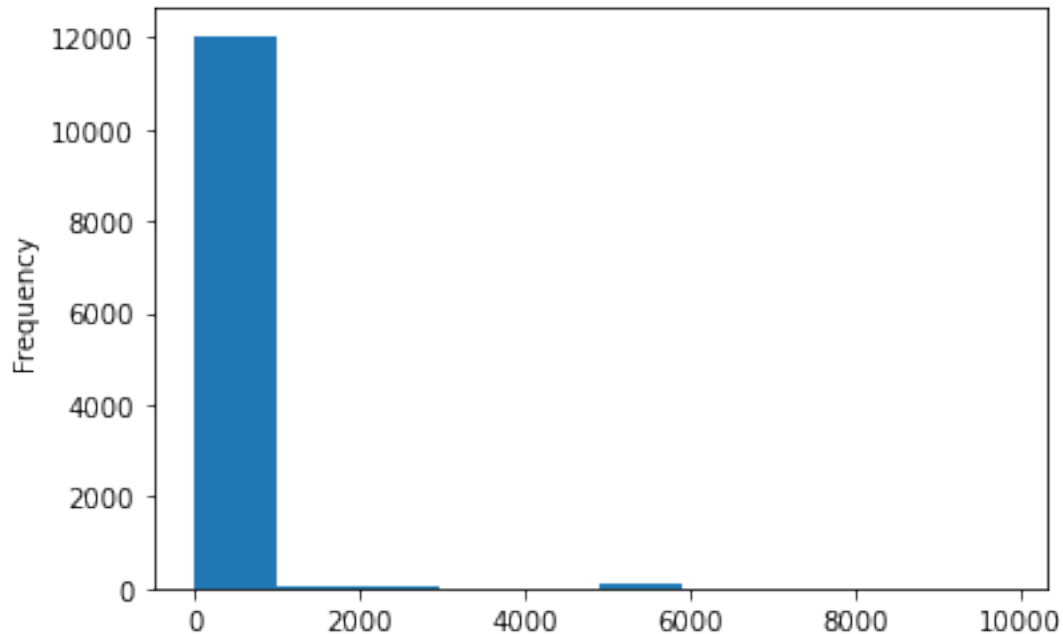
	name	achievements
20589	LOGistical	9821.0
12720	Trivia Vault: Movie Trivia	5000.0
24971	Panda Run	5000.0
29201	The Dropping of The Dead	5000.0
5520	USA 2020	5000.0
...
40828	Rocksmith® 2014 Edition - Remastered - Sabaton...	NaN
40829	Rocksmith® 2014 Edition - Remastered - Stone T...	NaN

40830	Fantasy Grounds - Quests of Doom 4: A Midnight...	NaN
40831	Mega Man X5 Sound Collection	NaN
40832	Stories In Stone	NaN

[40833 rows x 2 columns]

```
[4]: pd.Series(steam.achievements).plot(kind='hist', bins=10)
```

```
[4]: <AxesSubplot:ylabel='Frequency'>
```



It's clear there's no distribution in this graph, since the majority had 0 achievements

0.2 Test 2 - Distribution between price and discount

```
[5]: steam[['name', 'original_price']].sort_values(by='original_price')
```

```
[5]:
```

	name \
2940	2D Paintball
36679	HORIZON VANGUARD
6514	Super Hipster Lumberjack
39119	On Board 4 PC
26588	Space Wars
...	...
40816	VICCP
40822	Bare Boob Brawlerz: Novel 01 (Visual Novel)
40823	Gamers Club

```

40824          Gravia
40826          Alive

original_price
2940  "Better than a demo!" 2D Paintball - Online | ...
36679          "Location Test" Demo
6514          $0.50
39119          $0.90
26588          $0.90
...
40816          NaN
40822          NaN
40823          NaN
40824          NaN
40826          NaN

```

```
[40833 rows x 2 columns]
```

As you can see, some of these values are also in other currencies, but for this test it will be fine.

```
[6]: steam["original_price"] = steam["original_price"].str.replace('$','') # remove
      ↪dollar sign
      steam["original_price"] = pd.to_numeric(steam["original_price"],
      ↪downcast="float", errors='coerce') # convert to float

```

```
[7]: steam[['name','original_price']].sort_values(by='original_price')

```

```

[7]:
      name  original_price
6514  Super Hipster Lumberjack      0.5
39119          On Board 4 PC      0.9
26588          Space Wars      0.9
25371          Hex Phase      0.9
38041          On Board      0.9
...
40816          VICCP      NaN
40822  Bare Boob Brawlerz: Novel 01 (Visual Novel)      NaN
40823          Gamers Club      NaN
40824          Gravia      NaN
40826          Alive      NaN

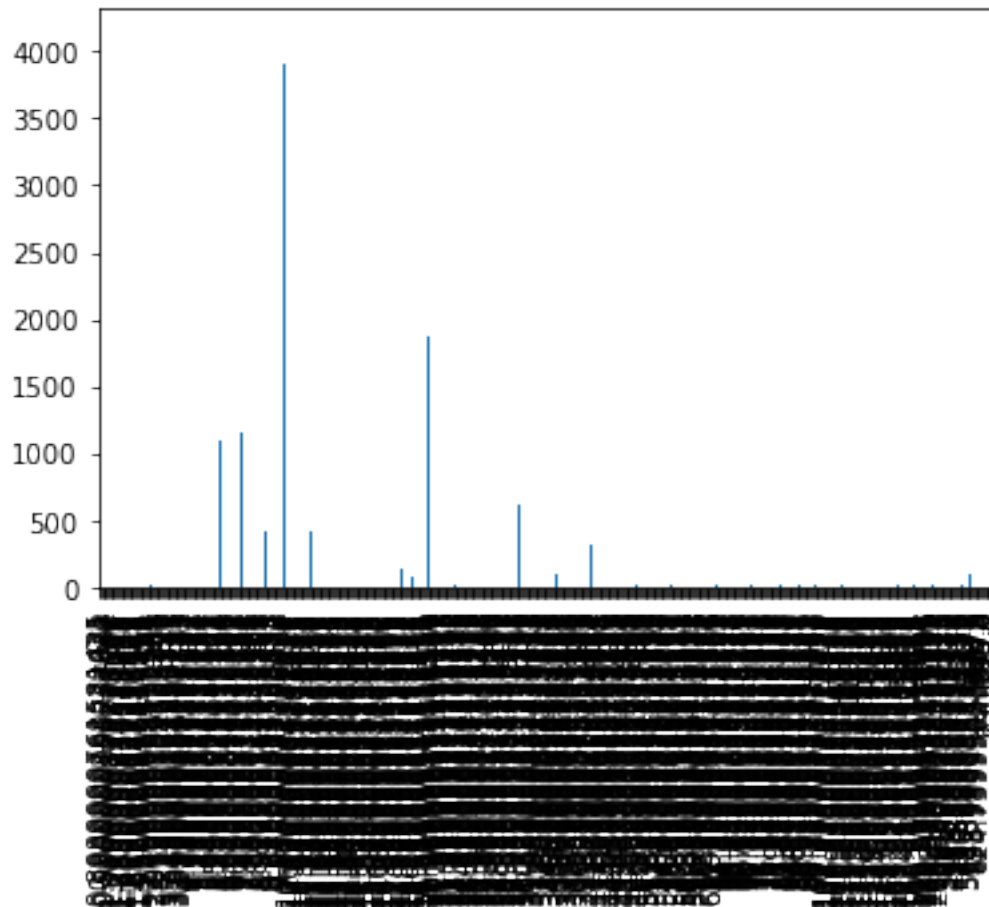
```

```
[40833 rows x 2 columns]
```

```
[8]: priceDist = pd.Series(steam.original_price).dropna()
```

```
[9]: priceDist.value_counts().sort_index().plot(kind='bar')
```

```
[9]: <AxesSubplot:>
```



The amount of games make it difficult to see, but you can definitely see a shape in this graph