PES UNIVERSITY

IDS Assignment

Analysis Of Video Game Sales

Project by:
Adithya Kiran
PES1201700231
Chirag P Tubakad
PES1201700896
Nandakrishna
PES1201701090

Dataset : Video Games Sales Across Different Genre and Different Parts Of The World

```
data = pd.read_csv("Video_Games_final.csv")
data
```

Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales	Critic_Score	Critic_Count	User_Score	User_Count
Wii Sports		2006.0	Sports	Nintendo	41.36	28.96	3.77	8.45	82.53	76.0	51.0	8	322.0
Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24	NaN	NaN	NaN	NaN
rio Kart Wii	Wii	2008.0	Racing	Nintendo	15.68	12.76	3.79	3.29	35.52	82.0	73.0	8.3	709.0
orts Resort	Wii	2009.0	Sports	Nintendo	15.61	10.93	3.28	2.95	32.77	80.0	73.0	8	192.0
Pokemon d/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37	NaN	NaN	NaN	NaN
Tetris	GB	1989.0	Puzzle	Nintendo	23.20	2.26	4.22	0.58	30.26	NaN	NaN	NaN	NaN
uper Mario Bros.	DS	2006.0	Platform	Nintendo	11.28	9.14	6.50	2.88	29.80	89.0	65.0	8.5	431.0
Wii Play	Wii	2006.0	Misc	Nintendo	13.96	9.18	2.93	2.84	28.92	58.0	41.0	6.6	129.0
uper Mario	Wii	2009.0	Platform	Nintendo	14.44	6.94	4.70	2.24	28.32	87.0	80.0	8.4	594.0

Dataset contains 15 rows, 5 of which contain categorical data, namely Genre, Platform, Developer, Name and Publisher. The other 10 columns have numeric data, sales in 5 regions, user score, critic score, user count, critic count, and year of release.
It contains 16 thousand rows, however there are a number on missing values and values which cause bias which has been cleaned before beginning analysis.

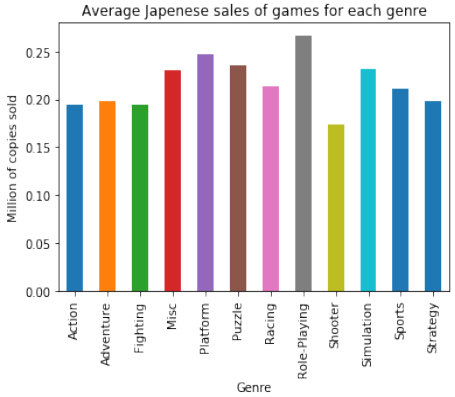
Data Cleaning

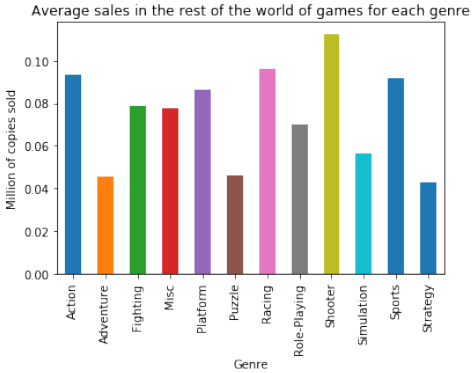
- We have cleaned the data quite thoroughly before beginning our analysis.
- We employed the cleaning methods : mean substitution and last observation carried forward.
- We also employed an additional method of dropping rows where 5 columns had missing data/ no sales as comparison would lead to bias when no sales occur in that region.

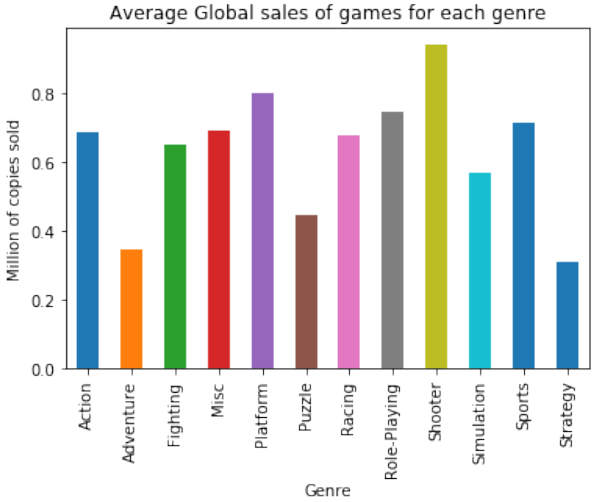
```
In [5]: data1
```

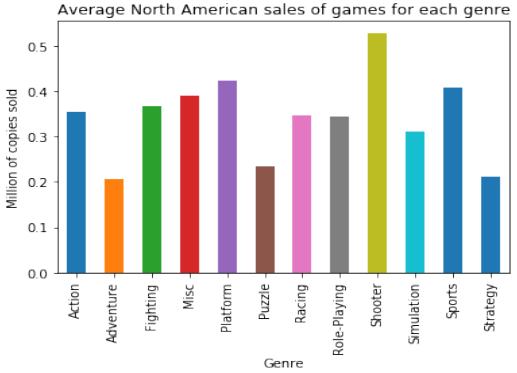
Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales	Critic_Score	Critic_Count	User_Score	User_C
Wii Sports	Wii	2006.0	Sports	Nintendo	41.360000	28.96	3.770000	8.4500	82.53	76.0000	51.00000	8.000000	322.0(
Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.680000	12.76	3.790000	3.2900	35.52	82.0000	73.00000	8.300000	709.0(
Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.610000	10.93	3.280000	2.9500	32.77	80.0000	73.00000	8.000000	192.0(
New Super Mario Bros.	DS	2006.0	Platform	Nintendo	11.280000	9.14	6.500000	2.8800	29.80	89.0000	65.00000	8.500000	431.0(
Wii Play	Wii	2006.0	Misc	Nintendo	13.960000	9.18	2.930000	2.8400	28.92	58.0000	41.00000	6.600000	129.0(
New Super Mario Bros. Wii	Wii	2009.0	Platform	Nintendo	14.440000	6.94	4.700000	2.2400	28.32	87.0000	80.00000	8.400000	594.0(
Mario Kart DS	DS	2005.0	Racing	Nintendo	9.710000	7.47	4.130000	1.9000	23.21	91.0000	64.00000	8.600000	464.0(
Wii Fit	Wii	2007.0	Sports	Nintendo	8.920000	8.03	3.600000	2.1500	22.70	80.0000	63.00000	7.700000	146.0(

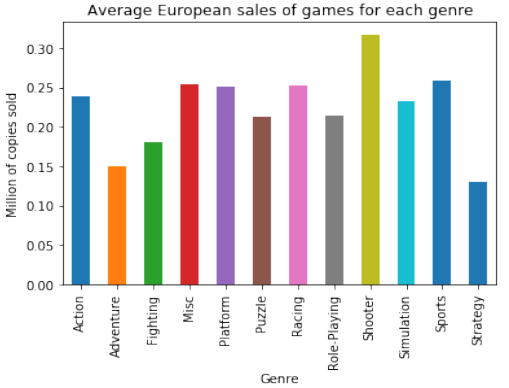
Comparison of sales per genre in different parts of the world



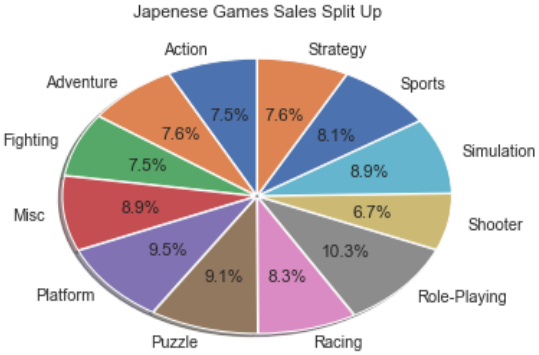


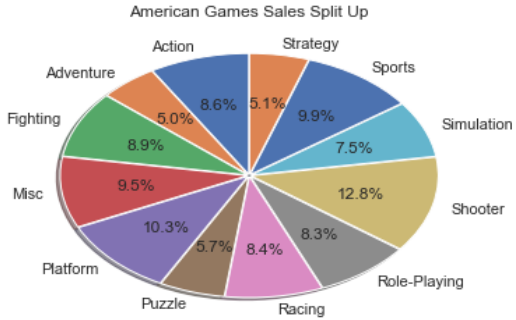




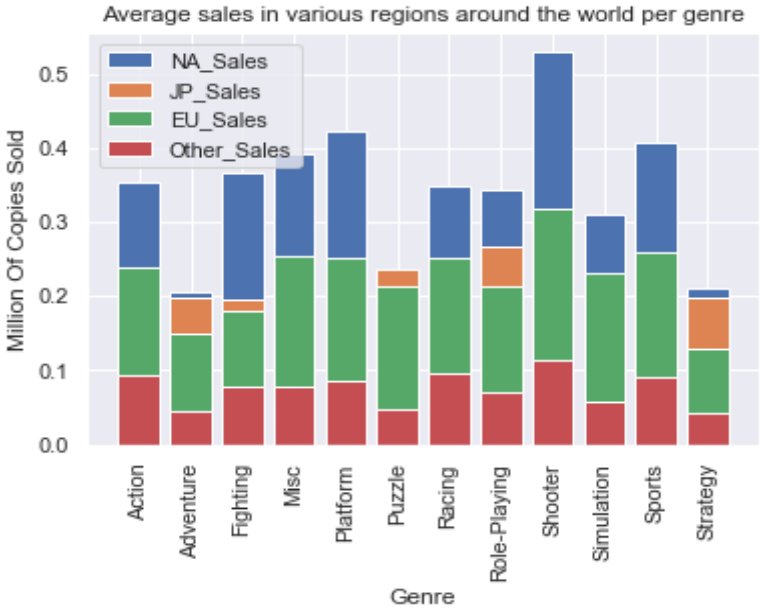


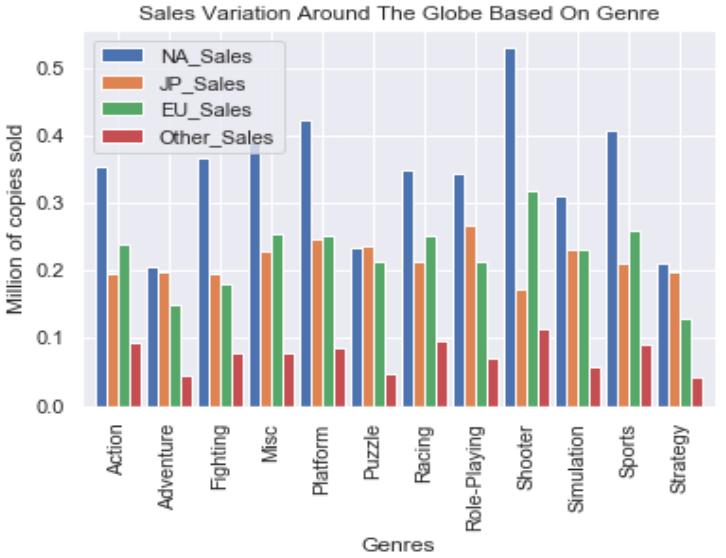
Percentage of sales per genre in different parts of the world

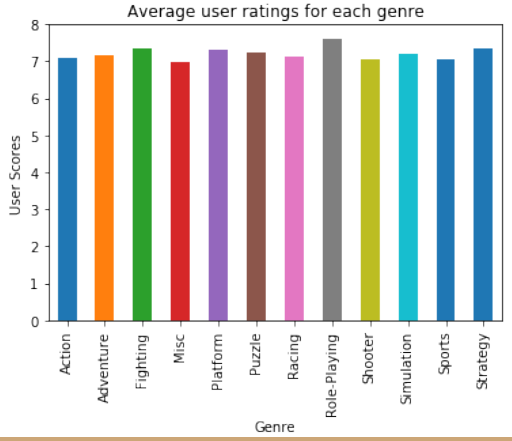


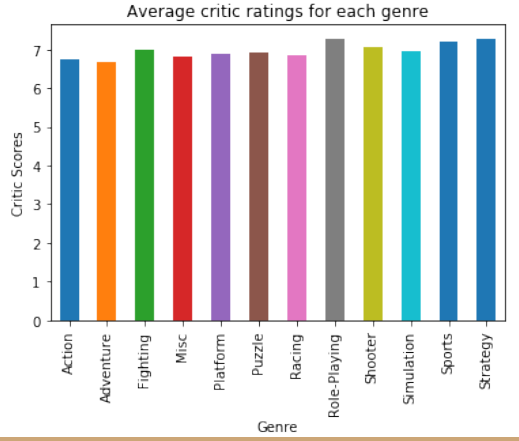


Some more comparisons

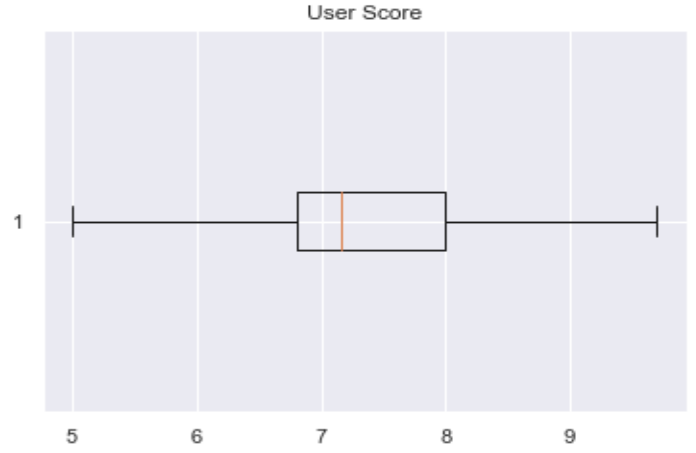


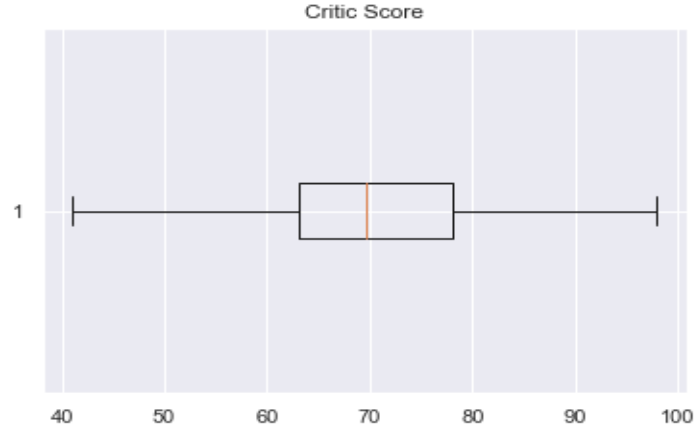




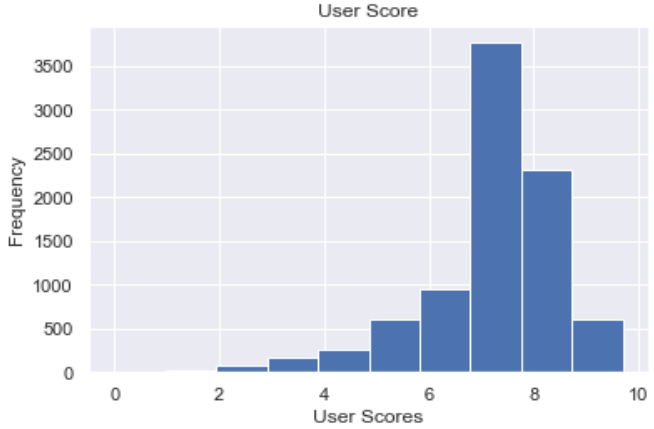


Boxplots:





Histogram



Hypothesis Testing

We made a hypothesis that the user scores on PS4 for strategy are more than 7.328255605278789 (the average score).

Thus our null hypothesis was H0: μ <= 7.328255605278789.

The result after testing was that we “failed to reject null hypothesis”.

This implies that :

```
H0 : μ <= 7.328255605278789
H1 : μ > 7.328255605278789
alpha value is : 0.05

actual z value : 1.6448536269514729
hypothesis z value : 1.3008894356055924

Failed to reject NULL hypothesis
```

PES UNIVERSITY

Project by:
Adithya Kiran
PES1201700231
Chirag P Tubakad
PES1201700896
Nandakrishna
PES1201701090

Under guidance of :
Dr. Mamatha H.R
Professor
Computer science
PES University

THANK YOU

Dataset source: www.kaggle.com/datasets