

Task 2:

Video games sales (see attached) dataset contains sales data for video games from all around the world, across different platforms, genres and regions. Armed with this data and analysis, future developers can better understand what types of gameplay and mechanics resonate more with players to create a new gaming experience. Undertake a comprehensive analysis to produce a report on

- Insights on what games successful (hit) across the globe
- What other things can we learn from this dataset
- Submit your code and report via Github repo

Result:

The libraries which are needed for the EDA process are imported and the dataset also has been read into a variable 'sales'. The preliminary process is conducted to understand the dataset better. Information such as the type of the variables are revealed. It helps deciding whether the variable is categorical or numerical. For, the video games sales dataset the attributes index and rank are of type integer. The attributes Game title, Platform, Genre, Publisher are of type Object. The attributes Year, North America, Europe, Japan, Rest of World, Global, Review is of type float.

Null value treatment is done as the next step after the preliminary exploration of the dataset. The attributes Year and Publisher has null values. But they make only a small amount in the dataset so those rows are dropped.

Normalization method min max scaling is performed to the attributes North America, Europe, Japan, Rest of World and Global with feature range between 1 to 10

The Univariate Analysis is performed on the dataset.

- Platform – The top 5 platforms are PS2, PS, X360, PS3 and Wii.
- Year – The sales of the video games were high in the years 2008, 2007, 2009, 2010, 2004.
- Genre – The popular genres which made higher sales than other genres are Sports, Action, Shooter, Platform and Racing.
- Publisher – The Game Publishers who made higher sales than other publishing companies are Electronic Arts, Nintendo, Sony Computer Entertainment, Activision and Libisoft.
- Game Title – Popular games which made high sales are Fifa Soccer 08, WWe SmackDown vs Raw 2008, The Simpsons Game, Pro Evolution Soccer 2008 and LEGO Indiana Jones: The Original Adventures.
- North America – The sales in North America reveals a Right-skewed histogram. So, the normalized values are transformed using Box Cox transformation to normalize the data. It results in a plot with normal distribution.
- Europe – The sales in Europe reveals a Right-skewed histogram. So, the normalized values are transformed using Box Cox transformation to normalize the data. It results in a plot with normal distribution.

- Japan – The sales in Japan reveals a Right-skewed histogram. So, the normalized values are transformed using Box Cox transformation to normalize the data. It results in a plot with normal distribution.
- Rest of World – The sales in Rest of World reveals a Right-skewed histogram. So, the normalized values are transformed using Box Cox transformation to normalize the data. It results in a plot with normal distribution.
- Global – The sales in Global reveals a Right-skewed histogram. So, the normalized values are transformed using Box Cox transformation to normalize the data. It results in a plot with normal distribution.
- Review – The sales in Review reveals a Left-skewed histogram. So, the normalized values are transformed using Box Cox transformation to normalize the data. It results in a plot with normal distribution.

The Bivariate Analysis is performed on the dataset.

- Successful game across the globe – The attributes Game Title and Global are grouped by the attribute Game Title and they are plotted which reveals the games which are sold most across the globe. They are Wii sports, Super Mario Bros, Tetris, Mario Kart Wii, Wii Sports Resort, New Super Mario Bros, Wii Play, Duck Hunt, Call of Duty: Modern Warfare 3 and Call of Duty: Black Ops.
- Game with Good Review – The attributes Game Title and Review are grouped by the attribute Game Title and they are plotted which reveals the games which are sold most across the globe. They are Fifa Soccer 08, Star Wars: The Force Unleashed, Lego Indiana Jones: The Original Adventures, Fifa Soccer 10, Pro Evolution Soccer 2008 etc.
- Global Sales by Year – The attributes year and global are grouped by the attribute year and they are plotted which reveals the games which reveals the highest game sales by year. They are 2008, 2007, 2009, 2006, 2010, 2005, 2004, 2011, 2002, 2001. These years had the highest sales.
- Popular Publisher – The publishing company which sold the greatest number of games are Nintendo, Electronic Arts, Sony Computer Entertainment, Activision, Take-Two Interactive, Ubisoft, Microsoft games studios, THQ, Sega, Capcom.
- Popular Genre – The popular genre which sold the greatest number of games are Sports, Action, Platform, Shooter, Role-Playing, Racing, Misc, Fighting, Adventure, Simulation.