Cairo University

Faculty of Engineering
Cairo University

Data Science Project Phase #1

Team 18

Name	Section	BN
غياث عمر	2	8
محمد أكرم	2	12
مريم محمد	2	22
مريم ناصر	2	23

Idea/Problem description

Data Description

Our data of choice is about video games sales.
We are trying to answer general questions about video game sales that could be of importance for developers, publishers, customers or some 3rd party promoting service.

Meta info:

1. Link
2. Date of collection: 2016
3. Number of rows: 16719
4. Number of columns: 16

Columns:

1. **Name**
   a. 11563 unique values
   b. Zero nulls

2. **Platform**
   a. 31 unique values
   b. Zero nulls
   c. Describes the platform this game published on
   d. "2600" "3DO" "3DS" "DC" "DS" "GB" "GBA" "GC" "GEN" "GG" "N64" "NES" "NG" "PC" "PCFX" "PS" "PS2" "PS3" "PS4" "PSP" "PSV" "SAT" "SCD" "SNES" "TG16" "Wii" "WiiU" "WS" "X360" "XB" "XOne"

3. **Year_of_Release**
   a. 1980 <= x <= 2020
   b. 269 N/A

4. **Genre**
   a. 31 unique values
   b. Zero nulls
   c.

Action	Adventure	Fighting	Misc	Platform	Puzzle
Racing	Role-Playing	Shooter	Simulation	Sports	Strategy

5. **Publisher**
    a. 582 unique values
    b. 54 “N/A” values

6. **NA_Sales**
    a. Numeric value
    b. 0 <= x <= 41.36
    c. 4511 Zeros

7. **EU_Sales**
    a. Numeric value
    b. 0 <= x <= 28.96
    c. 5874 Zeros

8. **JP_Sales**
    a. Numeric value
    b. 0 <= x <= 10.22
    c. 10515 Zeros

9. **Other_Sales**
    a. Numeric value
    b. 0 <= x <= 10.57
    c. 6604 Zeros

10. **Global_Sales**
    a. Numeric value
    b. 0.01 <= x <= 82.53
    c. No Zeros!

11. **Critic_Score**
    a. Aggregate score compiled by Metacritic staff
    b. Numeric value out of 100
    c. 8582 nulls

12. **Critic_Count**
    a. The number of critics used in coming up with the Critic_score
    b. Numeric value 3 <= x <= 113
    c. 8582 nulls

13. User_Score

a. Score by Metacritic's subscribers
b. Numeric value 0 <= x <= 10
c. 9129 nulls

14. User_Count

a. Number of users who gave the user_score
b. Numeric value 4 <= x <= 10k
c. 9129 nulls

15. Developer

a. Party responsible for creating the game
b. 1697 unique values
c. 6623 nulls

16. Rating

a. Rating - The ESRB ratings
b. 6769 nulls
c. “ ESRB ratings provide information about what’s in a game or app so parents and consumers can make informed choices about which games are right for their family.”
d.

Value	AO	E	E10+	EC	K-A	M	RP	T
Count	1	3991	1420	8	3	1563	3	2961

Exploratory Questions

#	Question	Work plan
1	***For some specific game of multiple versions, Does rating get better or worse over time?***	● Extract same game name; it might have different versions, example: GTA 1, GTA 2 ● Group games by name ● Plot increasing or decreasing curve over time
2	***Does the name of the publisher have a role in increasing the sales?***	● Group games that were published by the same publishers. ● Compute the global -overall- sales for all groups. ● Compare the number of games for each group and their equivalent sales.
3	***Compare platforms based on how long they stay competitive in market***	● Get overall sales for each platform grouped by year ● Extract effective lifespan of each platform ● Get average of all platforms and compare each one with the average
4	***Do users or critics rate a specific platform or genre higher than others?***	● For each platform/genre get average ratings ● Get the average of these averages ● Compare each with that average
5	***Which regions prefer which genres ?***	● Compare summation of global sales over all games, grouped by region and genre
6	***Is there a certain publisher whose sales are most coming from a certain region?***	● Compare summation of global sales over all games, grouped by region and publisher.

Descriptive Questions

1	***For games rated more than X, what is the most popular genre ?***	● Filter games based on rating ● Group by genre and do summation over sales
2	***Is there a certain genre that has the highest sales?***	● Group the game names by genre and select the genre with the max global sales

Predictive Questions

1	***Predict sales of a game given release date, platform, publisher and developer***	● Establish predictive model for sales like Time-series model,Regression models, Exponential smoothing models:
2	***For a specific genre , will its sales increase/decrease over the upcoming years?***	● A predictive model will be implemented as Time-series models, Regression models, Exponential smoothing models, ARIMA models or maybe Neural network models.

Mechanistic Questions

1	***How does the choice of platform affect sales for a specific genre ?***	● Sum sales for games of that genre grouped by platform ● Explain the relation if it exists ● Reason could be joysticks and controllers for each platform example: Joysticks make playing sport games easier but shooting games harder