1-Identify and explain database

Video Games Database

This database stores information about the top 11,000 or so video games that have had the most sales of all time. It captures video game sales in North America, Europe, Japan, and other regions grouped as "Other". It has games from various platforms: PC, recent consoles, and old consoles.

The sample data was obtained by a free online data source. There are about 11,000 games, 31 platforms, 577 publishers, and 4 regions in this dataset.

Table Explanations

Genre

This table has a short list of game genres, such as Strategy or Racing. It's helpful if you want to see what genre a game is, or the types of genres that are making sales.

Game

A list of all of the games in the database. It only contains a title, and a genre_id to link to the genre table. There is no concept of a series in this data (e.g. "Call of Duty" or "Need for Speed").

Publisher

A company that publishes video games.

Game_Publisher

This is a joining table that defines the games and their publishers, because in the source data, a game can have many publishers. One example is different publishers work with the game on different platforms (e.g. one on PC and one on XBOX).

Platform

A list of the different gaming platforms in the source data. For example, PC, Playstation 3, XBOX360, Nintendo.

Game_Platform

This table links the games, publishers, and the platform. Using this table, you can see the games, who published them, and on which platform (when you join to the related tables).

It also includes the release_year. Why is the release_year here and not on the game table?

This is because, in the source data, a game can be released by different publishers on a different platform in a different year. For example, SimCity 2000 was released on PC in 1992 and PS in 1996.

Region

This lists 4 different regions that sales were captured in: North America, Europe, Japan, and Others.

Region_Sales

This table shows all sales made for games in each region. It's the main table used to calculate the sales for queries.

The num_sales column is the number of sales of the game in the region, and **the number is in millions**.

2- What are the measures you have in the database?

-Top-selling video games

3- What KPIs you are going to measure

Top-selling video games, by platform

Top-selling video games, by region

Top-selling video games, by publisher

Top-selling video games, by year and region

4- ERD for Video Games Database

genre
PK	id	INT
	genre_name	VARCHAR

game
PK	id	INT
FK	genre_id	INT
	game_name	VARCHAR

game_publisher
PK	id	INT
FK	game_id	INT
FK	publisher_id	INT

publisher
PK	id	INT
	publisher_name	VARCHAR

game_platform
PK	id	INT
FK	game_publisher_id	INT
FK	platform_id	INT
	release_year	INT

platform
PK	id	INT
	platform_name	VARCHAR

region_sales
PK, FK	region_id	INT
PK, FK	game_platform_id	INT
	num_sales	INT

region
PK	id	INT
	region_name	VARCHAR