**INTRODUCTION TO DATA MANAGEMENT PROJECT REPORT**

(Project Semester August-December 2020)

***VEDIO GAME SALES ANALYSIS***

Submitted by

BAMMIDI SAI CHANDRA SEKHAR

Registration No:11802374

Programme and Section: KMO76

Course Code INT-217

Under the Guidance of

**Vasudha with 23636 and asst. Professor**

**Discipline of CSE/IT**

**Lovely School of Computer Science**

**Lovely Professional University, Phagwara**



**CERTIFICATE**

This is to certify that BAMMIDI SAI CHANDRA SEKHAR bearing Registration no. 11802374 has completed INT217 project titled, **“VEDIO GAMES SALES ANALYSIS”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

**Signature and Name of the Supervisor**

**Designation of the Supervisor**

**School of COMPUTER SCIENCE AND ENGINEERING**

Lovely Professional University

Phagwara, Punjab.

Date:

**DECLARATION**

I, BAMMIDI SAI CHANDRA SEKHAR student of INT-217 (Program name) under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: Signature

Registration No. 11802374 Name of the student

**TABLE OF CONTENTS:**

[**Introduction**](#intro)

[**Scope of Analysis**](#ScopeofAnalysis)

[**Sources of Data Set**](#Source)

[**ETL Process**](#ETLPROCESS)

[**Analysis on Dataset**](#Analysisondataset)

[**Objective-1**](#obj1)

[**Objective-2**](#obj2)

[**Objective-3**](#obj3)

[**Objective-4**](#obj4)

[**Objective-5**](#obj5)

[**References**](#References)

[**Bibliography**](#Bibliography)

**INTRODUCTION:**

A video game is an electronic game that involves interaction with a UI or input device such as joystick, controller, keyboard to generate visual feedback for a player. This is then shown on a either two- or three-dimensional video display device such as a TV set, Monitor, Mobile phone, VR headset etc..,. Video games are defined based on their platform, which includes arcade games, console games and Pc games. More recently the industry has been expanded onto mobile gaming through smartphones and tablet computers, virtual and augmented reality systems, and through cloud gaming. Video games are classified into a wide range of genres based on their type of gameplay and purpose.

Today, video game development requires numerous skills to bring a game to market, including developers, publishers, distributors, retailers, console and other third-party manufacturers and other roles. Since the 2000’s the commercial importance of the video game industry has been increasing. After analyzing many recent stacks on the video-gaming analysis 2018 was a record-breaking year for our industry, with total video games sales exceeding $43.4 billion. Over 164 million adults in the US play video games.

And like China and South Korea, India is experiencing strong growth in online gaming. With between 35-50 million users, the country is attracting interest from the online video game industry, which is difficult to hack.

With a turnover of $890 million in 2018, the video game sector is still underdeveloped compared to other Asian countries, such as China and South Korea etc..,. However, many video game companies are beginning to invest, and India could become an important market for this sector. One of the factors driving the growth of the video game market is its large number of cyber cafes with more than 100,000 in 2006, 40% of which are used to play online.

This report presents an overview of gaming consoles and a comparison of how the consoles, games and the distributions of various versions over the years. we compare how the various competitors compete over years to give the best they can for the gaming lovers out there. The report concludes with an overview of other gaming platforms, and a summary of key points and trends to monitor. This report might be useful in discovering trends, perception, and other aspects of gaming industry. This dataset contains a list of video games with sales greater than 100,000 copies.

Fields include:

Rank - Ranking of overall sales

Name - The games name

Platform - Platform of the games release (i.e. PC, PS4, etc.)

Year - Year of the game's release

Genre - Genre of the game

Publisher - Publisher of the game

NA\_Sales - Sales in North America (in millions)

EU\_Sales - Sales in Europe (in millions)

JP\_Sales - Sales in Japan (in millions)

Other\_Sales - Sales in the rest of the world (in millions)

Global\_Sales - Total worldwide sales.

**Scope of Analysis:**

This report might be useful in discovering trends, perception and other aspects of gaming industry. This dataset contains a list of video games with sales greater than 100,000 copies.

**Objectives of this project:**

* To good hand on excel.
* To learn the ETL process.
* How to link one sheet to another.
* How to use pivot table.
* Learn to make dashboard.
* To make different type of graphs.
* To learn how to fetch data from other source to excel.

**Source of data set:**

<https://www.kaggle.com/gregorut/videogamesales>

**ETL PROCESS:**

**ETL** is a process that extracts the data from different source systems, then transforms the data (like applying calculations, concatenations, etc.) and finally loads the data into the Data Warehouse system. Full form of ETL is Extract, Transform and Load.

At its most basic, the ETL process encompasses data extraction, transformation, and loading. While the abbreviation implies a neat, three-step process – extract, transform, load – this simple definition doesn’t capture:

* The transportation of data
* The overlap between each of these stages
* How new technologies are changing this flow

**Why do we need ETL?**

There are many reasons for adopting ETL in the organization:

* It helps companies to analyse their business data for taking critical business decisions.
* Transactional databases cannot answer complex business questions that can be answered by ETL.
* A Data Warehouse provides a common data repository
* ETL provides a method of moving the data from various sources into a data warehouse.
* As data sources change, the Data Warehouse will automatically update.

**Steps of ETL:**

1.Extract.

2.Transformation

3.Loading

**Extraction:**

In this step, data is extracted from the source system into the staging area. Transformations if any are done in staging area so that performance of source system in not degraded. Also, if corrupted data is copied directly from the source into Data Warehouse

database, rollback will be a challenge. Staging area gives an opportunity to validate extracted data before it moves into the Data Warehouse.

Data Warehouse needs to integrate systems that have different DBMS, Hardware, Operating Systems and Communication Protocols. Sources could include legacy applications like Mainframes, customized applications, point of contact devices like ATM, Call switches, text files, spreadsheets, ERP, data from vendors, partners amongst others

Hence one needs a logical data map before data is extracted and loaded physically. This data map describes the relationship between sources and target data.

**Three Data Extraction methods**

1. Full Extraction
2. Partial Extraction – without update notification.
3. Partial Extraction – with update notification

Irrespective of the method used, extraction should not affect performance and response

time of the source systems. These source systems are live production databases. Any slow down or locking could effect company’s bottom line.

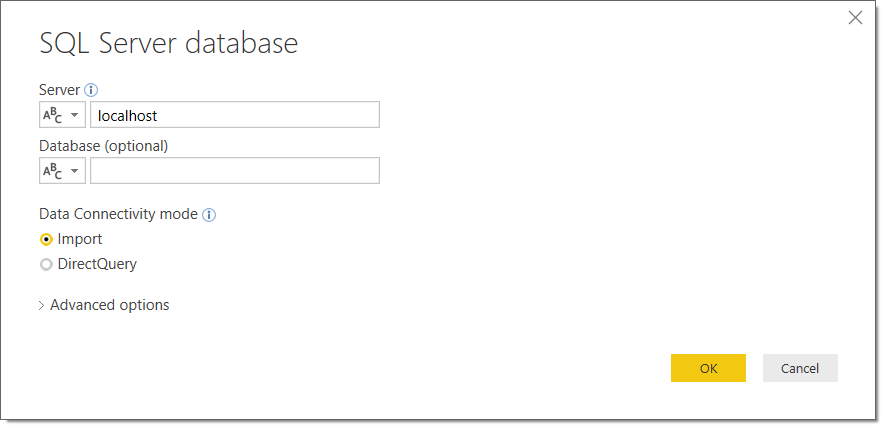
**Some validations are done during Extraction**

* Reconcile records with the source data
* Make sure that no spam/unwanted data loaded
* Data type check
* Remove all types of duplicate/fragmented data
* Check whether all the keys are in place or not

**Extraction of My data from the server:**

The data for the sales analysis I have extracted the data from the sql server database of edx site belongs to the united states of America.

To extract the data from the sql server you need to go to data in excel and the click on get data and then choose get data from the data base and then press from sql server



In the server you need to enter: msedxeus.database.windows.net In the data base enter : DAT206x01

Then choose multiple option there will be sales data table , manufacturer data table , Locations data table, products data table.

The next step is to transform the data.

**Transform:**

Data extracted from source server is raw and not usable in its original form. Therefore, it needs to be cleansed, mapped and transformed. In fact, this is the key step where ETL process adds value and changes data such that insightful BI reports can be generated.

In this step, you apply a set of functions on extracted data. Data that does not require any transformation is called as **Direct move** or **Pass through data.**

In transformation step, you can perform customized operations on data. For instance, if the user wants sum-of-sales revenue which is

not in the database. Or if the first name and the last name in a table

is in different columns. It is possible to concatenate them before loading

**Validations are done during this stage**

Filtering – Select only certain columns to load

Using rules and lookup tables for Data standardization

Character set conversion and encoding handling

Conversion of units of measurements like Date Time Conversion, Currency conversions, Numerical Conversions, etc.

Data threshold validation check. For example, age cannot be more than two digits. Data flow validation from the staging area to the intermediate tables.

Required fields should not be left blank.

Cleaning (for example, mapping NULL to 0 or Gender Male to “M” and Female to “F” etc.)

Split a column into multiples and merging multiple columns into a single column. Transposing rows and columns

Use lookups to merge data

Using any complex data validation (e.g., if the first two columns in a row are empty then it automatically rejects the row from processing)

**Loading:**

Loading data into the target data warehouse database is the last step of the ETL process. In a typical Data warehouse, huge volume of data needs to be loaded in a relatively short period. Hence, load process should be optimized for performance.

In case of load failure, recover mechanisms should be configured to restart from the point of failure without data integrity loss. Data Warehouse admins need to monitor, resume, cancel loads as per prevailing server performance.

**Types of Loading**

**Initial Load** – Populating all the Data Warehouse tables

**Incremental Load** – Applying ongoing changes as when needed periodically.

**Full Refresh** – Erasing the contents of one or more tables and reloading with fresh data.

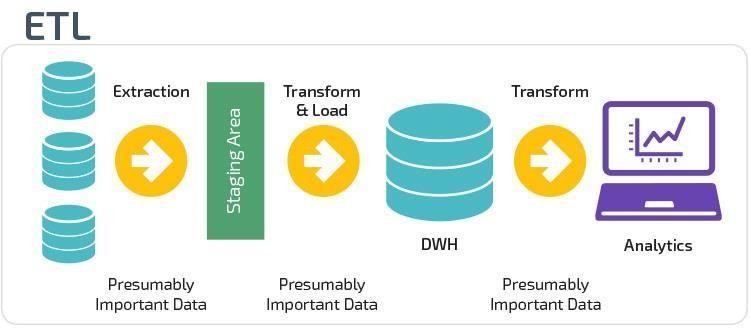
**Load Verification**

Ensure that the key field data is neither missing nor null.

1.Test modelling views based on the target tables.

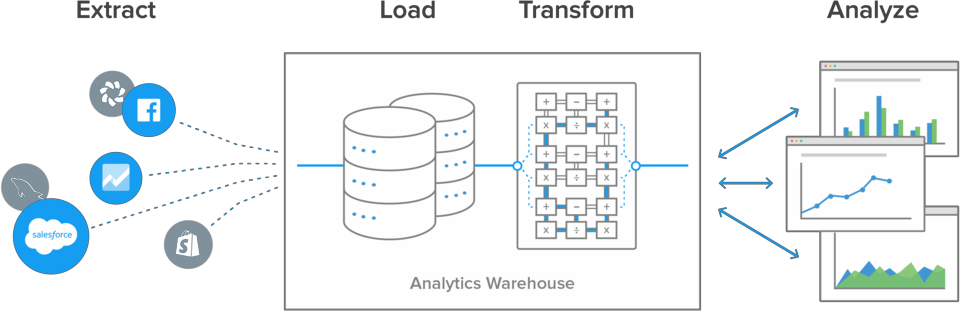
2.Check that combined values and calculated measures.

3.Data checks in dimension table as well as history table.

 4.Check the BI reports on the loaded fact and dimension table.

**Modern ETL Process**

Modern technology has changed most organizations approach to ETL, for several reasons. The biggest is the advent of powerful analytics warehouse like Amazon Redshift and Google Big Query. These newer cloud-based analytics databases have the horsepower to perform transformations in place rather than requiring a special staging area.

Also, data today is frequently analysed in raw form rather than from preloaded OLAP summaries. This has led to the development of lightweight, flexible, and transparent ETL systems with processes that look something like this:

The biggest advantage to this setup is the transformations and data modelling happen in the analytics database. This gives the BI team, data scientist, and analysts greater control over how they work with it, in a common language they all understand.

**ETL Process on my data:**

After extraction of data the cleaning process is performed all null values were changed into standardized value. The identifiers were made according to the need and representation. The attribute was changed accordingly as they were not proper and jumbled. I have also removed some of the unwanted columns such as 0’s, NO’s in order to evaluate the data easily. It helps in understanding the data quite easily. Finally, at last I have converted the csv file excel format. Graphical user interface, application, table

Description automatically generated

**ANALYSIS ON DATASET:**

Analysis on data set is done on the sales of the video games among the countries.

**Objectives:**

1. Country sales comparison.
2. Sales analysis with respect to publisher.
3. Top genre hits.
4. Number of releases per year.
5. Global sales comparison.

**Objective 1:**

Country sales comparison:

As per the records most gaming addicts are from the North America ,Japan, and Europe so comparing the sales of games provided with respective platform or by most point breaking year between these countries makes a lot of sense. In total user is provided with sale comparison between these major gaming countries.

**Analysis**:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Genre |  | Puzzle |  |  |
|  |  |  |  |  |
| **Row Labels** |  | **Sum of NA\_Sales** | **Sum of EU\_Sales** | **Sum of JP\_Sales** |
| 2600 |  | 12.72 | 0.8 | 0 |
| SAT |  | 0 | 0 | 1 |
| 3DO |  | 0 | 0 | 0.02 |
| 3DS |  | 1.29 | 2.12 | 1.8 |
| DS |  | 35.02 | 26.21 | 15.89 |
| GB |  | 29.35 | 5.03 | 12.08 |
| GBA |  | 6.65 | 2.35 | 1.44 |
| GC |  | 2.75 | 0.84 | 0.16 |
| N64 |  | 2.21 | 0.57 | 0.54 |
| NES |  | 9.14 | 1.98 | 9.44 |
| PC |  | 0.09 | 0.68 | 0 |
| PS |  | 5.29 | 2.37 | 3.8 |
| PS2 |  | 2.84 | 1.93 | 0.45 |
| PS3 |  | 0.26 | 0.05 | 0.1 |
| PS4 |  | 0 | 0.02 | 0 |
| PSP |  | 2.84 | 1.26 | 0.53 |
| PSV |  | 0.08 | 0.05 | 0.05 |
| SNES |  | 0 | 0 | 6.37 |
| Wii |  | 9.27 | 3.6 | 1.5 |
| WiiU |  | 0.66 | 0.36 | 0.22 |
| X360 |  | 0.49 | 0.04 | 0.15 |
| XB |  | 0.32 | 0.08 | 0 |
| **Grand Total** |  | **121.27** | **50.34** | **55.54** |

**Visualization**:

**Objective 2:**

Sales analysis with respect to publisher:

While collecting data on the video games it is most important to know who the creator behind its enormous success is so this objective is there to show the statistics of those great creators.

**Analysis:**

|  |  |  |  |
| --- | --- | --- | --- |
| Year | 2015 |  |  |
| Genre | Simulation |  |  |
|  |  |  |  |
| **Row Labels** | **Sum of NA\_Sales** | **Sum of EU\_Sales** | **Sum of JP\_Sales** |
| Electronic Arts | 0.05 | 0.18 | 0 |
| FuRyu | 0 | 0 | 0.03 |
| Happinet | 0 | 0 | 0.01 |
| Kalypso Media | 0.05 | 0.16 | 0.05 |
| Koch Media | 0.52 | 0.71 | 0 |
| Marvelous Entertainment | 0.04 | 0 | 0.1 |
| Nintendo | 0.47 | 1 | 1.39 |
| Paradox Interactive | 0 | 0.13 | 0 |
| Sega | 0 | 0.29 | 0 |
| Unknown | 0 | 0.04 | 0 |
| **Grand Total** | **1.13** | **2.51** | **1.58** |

**Visualization**:

**Objective 3:**

Top genre hits:

When it came to gaming each individual gamer has their own fav genre. So hands down its the comparison between different genres which is banging the leading gaming industry

**Analysis:**

|  |  |
| --- | --- |
| Year | (All) |
|  |  |
| **Row Labels** | **Sum of Global\_Sales** |
| Action | 1716.63 |
| Adventure | 229.09 |
| Fighting | 444.04 |
| Misc | 793.67 |
| Platform | 827.26 |
| Puzzle | 240.1 |
| Racing | 726.38 |
| Role-Playing | 922.76 |
| Shooter | 1020.88 |
| Simulation | 388.06 |
| Sports | 1308.84 |
| Strategy | 172.31 |
| **Grand Total** | **8790.02** |

**Visualization:**

**Objective 4:**

Number of releases per year:

to be the major competitors in the leading gaming industries one has also need to focus not only on the gaming ideas but also its releases per year

**Analysis:**

|  |  |
| --- | --- |
| Publisher | (All) |
|  |  |
| **Row Labels** | **Count of Name** |
| 1980 | 9 |
| 1981 | 46 |
| 1982 | 36 |
| 1983 | 17 |
| 1984 | 14 |
| 1985 | 14 |
| 1986 | 21 |
| 1987 | 16 |
| 1988 | 15 |
| 1989 | 17 |
| 1990 | 16 |
| 1991 | 41 |
| 1992 | 43 |
| 1993 | 60 |
| 1994 | 120 |
| 1995 | 219 |
| 1996 | 263 |
| 1997 | 289 |
| 1998 | 376 |
| 1999 | 338 |
| 2000 | 347 |
| 2001 | 478 |
| 2002 | 823 |
| 2003 | 768 |
| 2004 | 760 |
| 2005 | 927 |
| 2006 | 1003 |
| 2007 | 1186 |
| 2008 | 1414 |
| 2009 | 1412 |
| 2010 | 1246 |
| 2011 | 1124 |
| 2012 | 651 |
| 2013 | 542 |
| 2014 | 577 |
| 2015 | 605 |
| 2016 | 343 |
| 2017 | 3 |
| 2020 | 1 |
| **Grand Total** | **16180** |

**Visualization:**

**Objective 5:**

Global sales comparison:

Who does not want to know what is the top trending game across the globe? Or user also prefer to check if he/she is playing the trending game available in the market or developers want to know where their effort in the global market. This objective clears all the doubts above

**Analysis:**

|  |  |  |  |
| --- | --- | --- | --- |
| Year | 2015 |  |  |
| Genre | Action |  |  |
|  |  |  |  |
| **Row Labels** | **Sum of NA\_Sales** | **Sum of EU\_Sales** | **Sum of Global\_Sales** |
| 3DS | 2.4 | 1.63 | 12.1 |
| PC | 0.73 | 1.24 | 2.12 |
| PS3 | 1.14 | 1.78 | 5.71 |
| PS4 | 10.74 | 13.46 | 31.76 |
| PSP | 0 | 0 | 0.12 |
| PSV | 0.22 | 0.85 | 3.44 |
| Wii | 0.05 | 0.28 | 0.35 |
| WiiU | 1.27 | 0.85 | 2.34 |
| X360 | 1.48 | 1.4 | 3.14 |
| XOne | 4.8 | 3.15 | 8.74 |
| **Grand Total** | **22.83** | **24.64** | **69.82** |

**Visualization:**

**References:**

1. <https://brandessenceresearch.biz/ICT-and-Media/PC-and-Online-Gaming-Industry-in-India/Summary2>
2. <https://www.statista.com/outlook/203/119/video-games/india>
3. <https://en.wikipedia.org/wiki/Video_games_in_India>

**Bibliography:**

**Source:**

<https://www.history.com/topics/inventions/history-of-video-games>