**Report:**

1. The sources of data that you will extract from.
   * Used the data from Kaggle.com

First data set is a csv format: <https://www.kaggle.com/gregorut/videogamesales>

Second data set is a JSON format: <https://www.kaggle.com/rsrishav/youtube-trending-video-dataset?select=US_category_id.json>

1. The type of transformation needed for this data (cleaning, joining, filtering, aggregating, etc).
   * Cleaned the csv data to change to right data types.
   * Removed Nan values and copied to a new data frame
   * Normalized the json tables and created a data frame
   * Renamed the columns and converted the ‘Object’ to ‘Int’ datatypes
2. The type of final production database to load the data into (relational or non-relational).

These tables below were created in Postgres

<!-- -- Create the vidgames table -->

CREATE TABLE vidgames (

Rank INT,

Name VARCHAR,

Platform VARCHAR,

Year INT,

Genre VARCHAR,

Publisher VARCHAR,

NA\_Sales FLOAT,

EU\_Sales FLOAT,

JP\_Sales FLOAT,

Other\_Sales FLOAT,

Global\_Sales FLOAT

);

<!-- --Create youtube list -->

CREATE TABLE youtubelist (

ID INT,

Title VARCHAR,

ETag VARCHAR,

Flag Bool,

Kind VARCHAR,

ChannelId VARCHAR

);

1. The final tables or collections that will be used in the production database.
   * select \* from vidgames
   * select \* from youtubelist