* Extracting the category id and title from US\_category\_id.json file, and create an dictionary “category\_dict” to hold the key-value pairs.
* Create a dataframe for category id and title.
* Export the dataframe to a .csv file named “category\_id.csv”.
* Drop duplicates and keep the last occurrence so will have maximum amount of data for “views”, “likes”, “dislikes”, “comment\_count” columns.
* Create a dictionary “categories”, category id as the keys, category name as the values.
* Create a duplicate column of category\_id as a new column named “category\_type”. Using lambda function to apply the name of the category according to the original category\_id.
* Drop the “thumbnail\_link” column.
* Convert the datatype of “trending\_date” column to datetime format.
* Convert the datatype of “publish\_time” to datetime format, and cast it with .astype() method to datetime64[ns] datatype.
* Extracting the publish date and time from “publish\_time” column to “publish\_date” and “publish\_time” columns respectively.
* After separated, the datatype of “publish\_date” and “publish\_time” became object, so use pandas to\_datetime method to convert the “publish\_date” to datetime format.
* Extracting the time for a video to become trending from “publish\_date” and “trending\_date” and create a column name “diff\_days” to hold the values.
* Reorder the columns of the dataframe