# **Exploratory Data Analysis of Activities**

[Game Type Distribution](#_1kj0cygcrn9y)

[Activity Level Distribution](#_50l4wg84o9pw)

[Level x Age Distribution](#_8x7uijrg8xmv)

[Level x Game Type Distribution](#_iw7xlaesnjyh)

[Level x Activity Type Distribution](#_26fk232otplj)

[Level x Culture Distribution](#_ep55mgx5ke5s)

[Activity\_Type x Hits x Level](#_5nniq5u5bqwh)

[Game Type x Hits x Level](#_s10o6br218ms)

[Cultures x Game Type](#_4oj3ie1nch6p)

[Cultures x Activity Type](#_updj1amoc9yb)

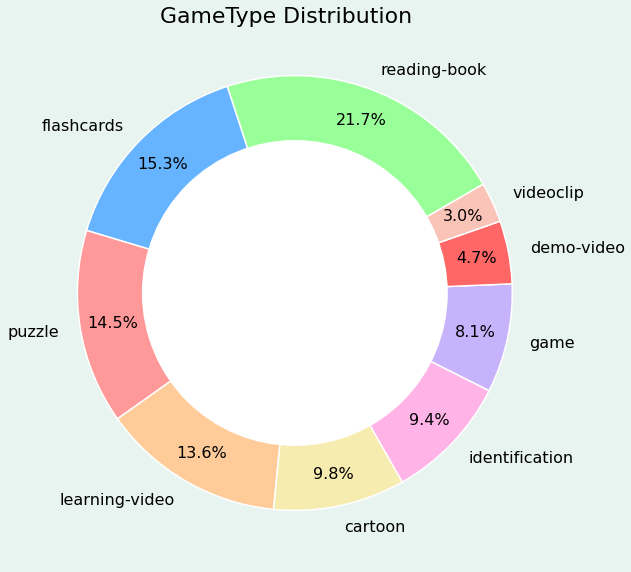
[Final Thoughts](#_btl7gn3notx7)

**An overview**:

* The EDA is based on how activities are distributed over different factors of the **Activities Dataset**.
* The majority of the inferences were drawn on the basis of the distribution of levels across the dataset.
* And, also on the number of hits across other parameters.

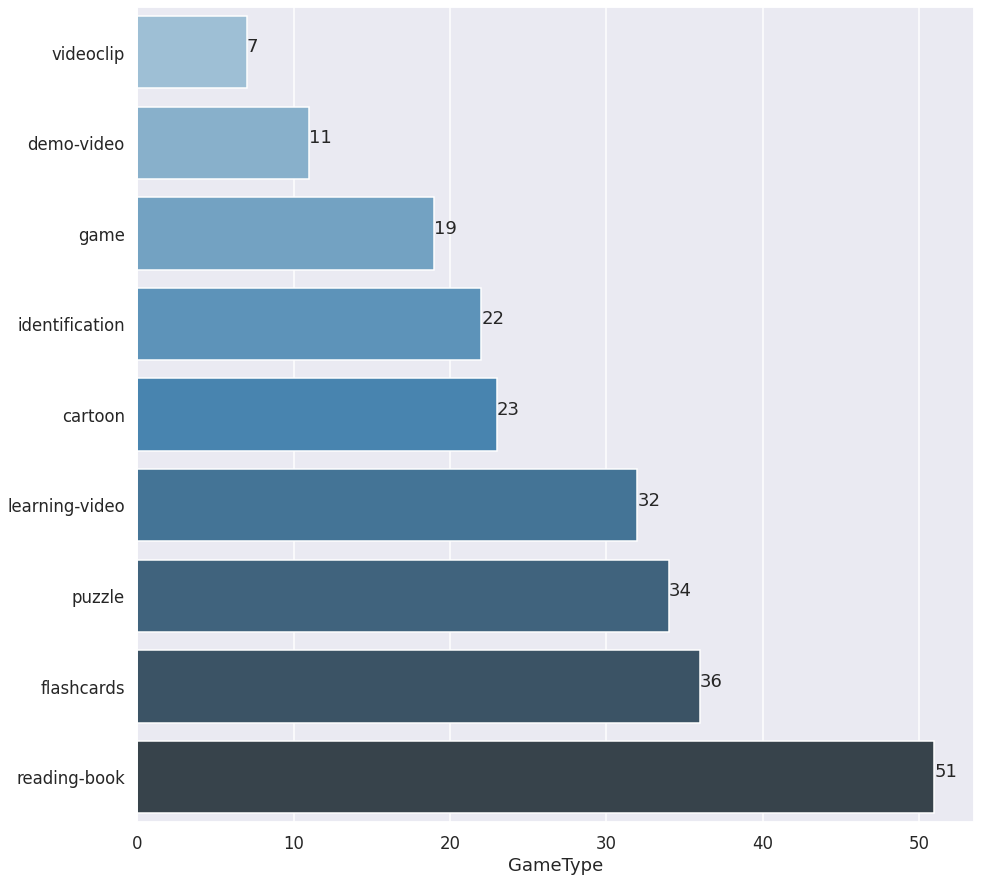
**Note**: The data and inferences are made before the weekly meet. I will be sharing the updated one in the coming week.

## Game Type Distribution



**In Numbers**

* reading-book 51
* flashcards 36
* puzzle 34
* learning-video 32
* cartoon 23
* identification 22
* game 19
* demo-video 11
* videoclip 7



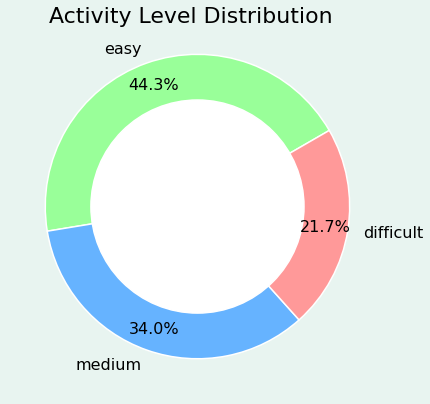
**Insights**

* The majority of the Games are of type **‘reading-book’**.
* The next two are **‘flashcards’** and **‘puzzle’** respectively.
* **‘demo-video’** and **‘videoclip’** have the least two counts of activities.

**Questions**

* If the distribution is not based on any specific theory, can the activities be distributed more fairly?
* If the demo-video is a demo of how to perform a certain activity, do we have enough of them?

## Activity Level Distribution



**In Numbers**

* easy 104
* medium 80
* difficult 51

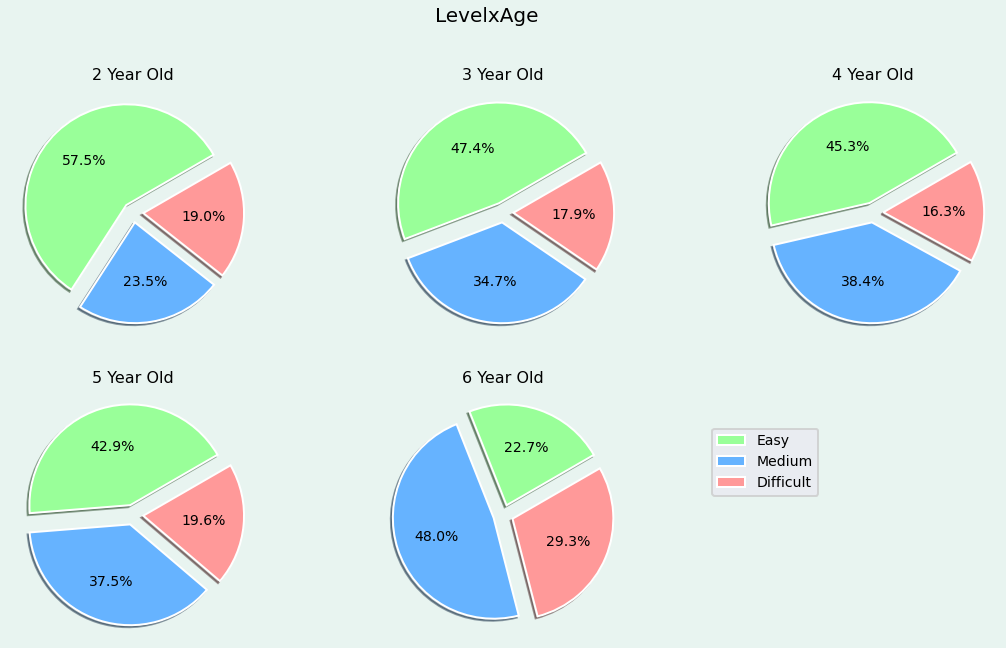
**Insights**

* Almost half-of the activities are of level ’**easy’**
* A fast learner may feel the activities are too easy

**Questions**

* Can more difficult activities be added?

## Level x Age Distribution



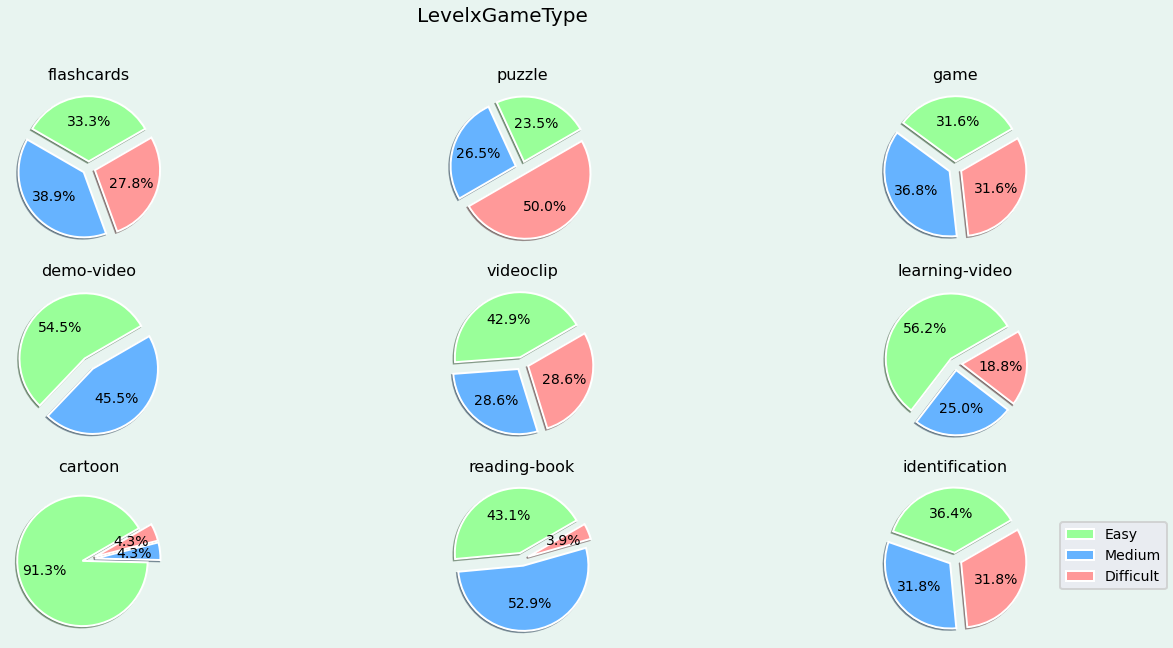
**Insights**

* The number of level **‘easy’** activities tend to decrease as the age increases
* The number of level **‘medium’** activities comprises 34% of the overall activities
* The 6-year-old activities consist of double the number of level **‘medium**’ activities of a 2-year-old.
* The number of level ‘**difficult’** activities of a 2-year-old is more than the number that of 3, 4 and 5-year-old.

**Questions**

* Can level **‘medium’** and level **‘difficult’** be more fairly distributed for a 6-year-old?

## Level x Game Type Distribution



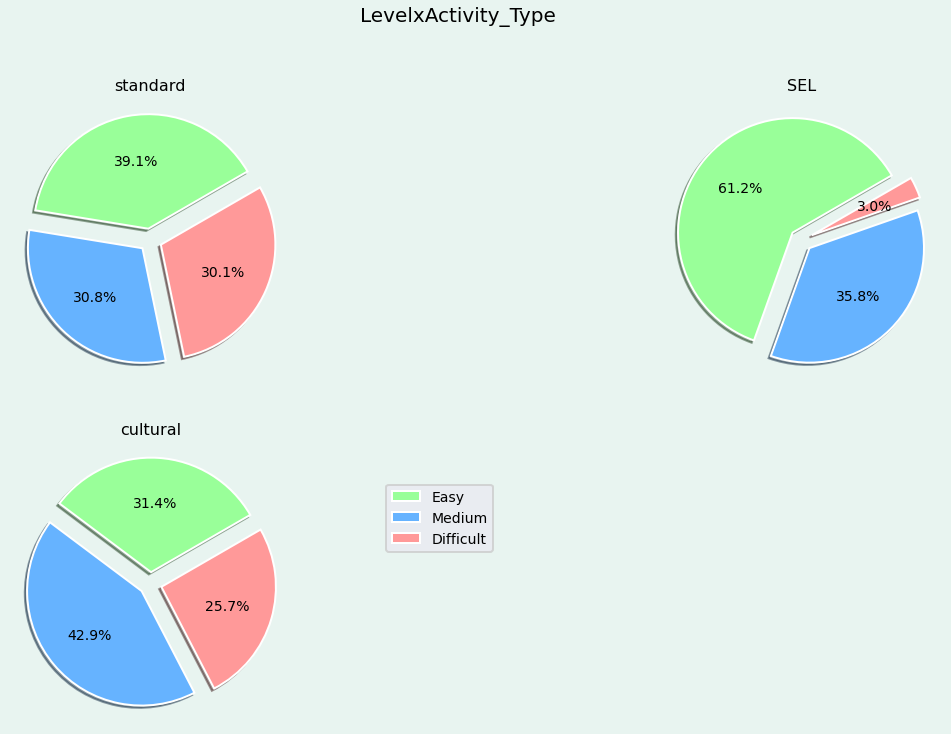
**Insights**

* Half the number of **‘puzzle**’ are of level **‘difficult**’
* No activities of level ‘difficult’ among **‘demo-videos**’
* More than half the number of ‘**learning-video’** are of level **‘easy**’
* 90% of the ‘**cartoon**’ is of level ‘easy’
* Only 4% of the **‘reading-book**’ are of level ‘difficult’

**Questions**

* Can more level **‘medium’** and level ‘**easy’** be added for Game Type **‘puzzle’**?
* Can levels be fairly distributed for Game Type ‘**reading-book’**?
* Can levels be fairly distributed for Game Type ‘**cartoon’**?
* Can levels be fairly distributed for Game Type **learning-video**?

## Level x Activity Type Distribution



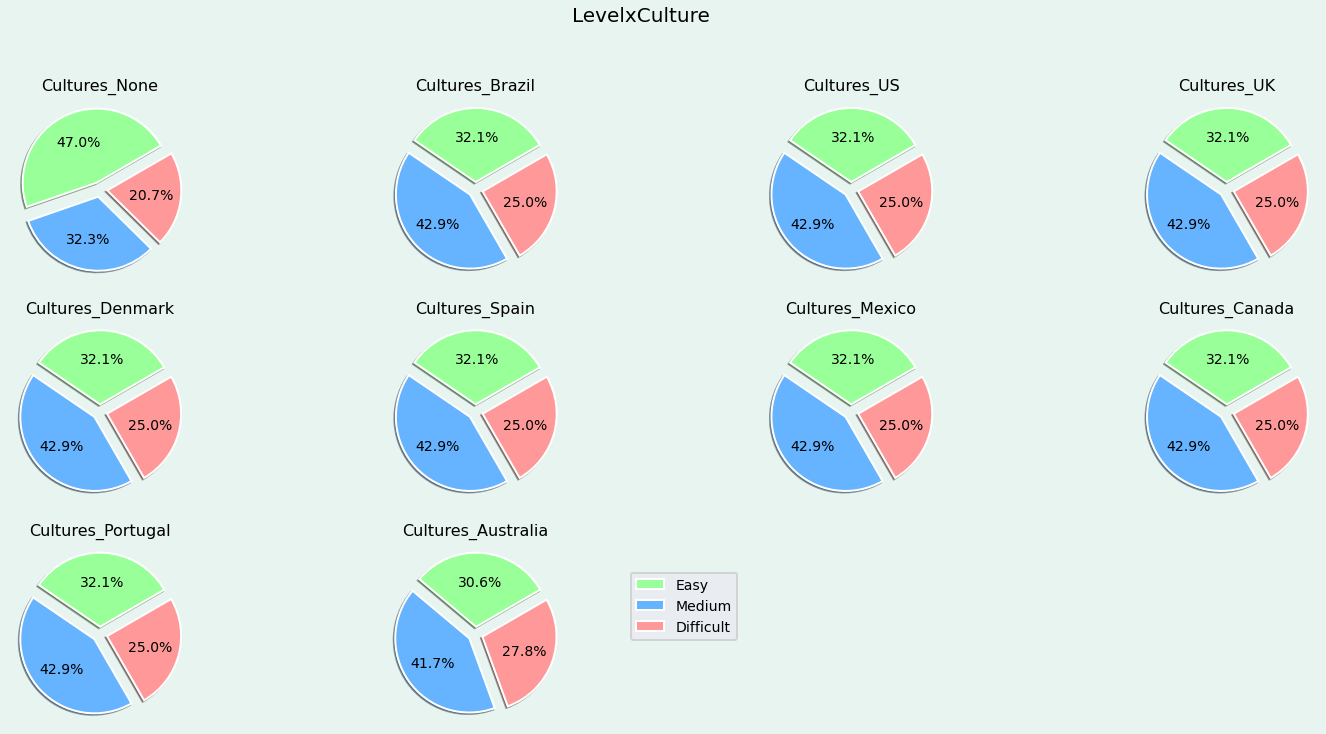
**Insights**

* Levels are fairly distributed for ‘**standard’** and **‘cultural**’ Activity Types.
* Levels are unfairly distributed for the **‘SEL**’ Activity Type.

**Questions**

* Can more level ‘**difficult**’ be added for Activity Type ‘**SEL**’?

## Level x Culture Distribution



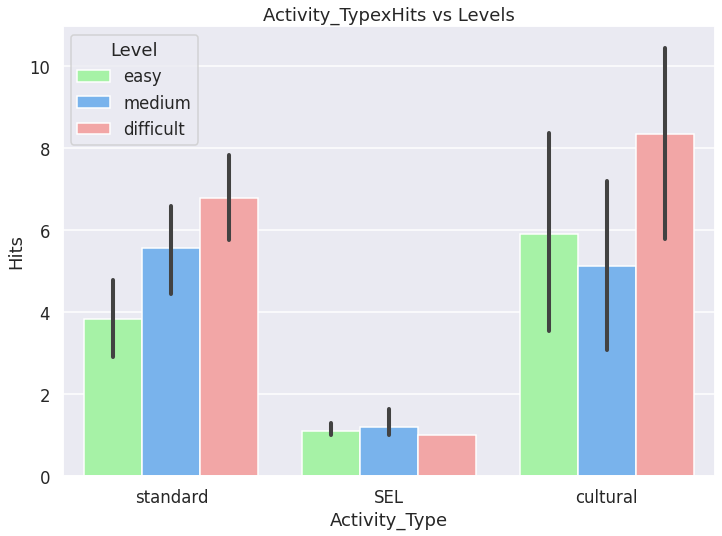
**Insights**

* Levels are fairly distributed at a percentage of easy - 32.1%, medium - 25%, and difficult - 25% for cultures such as Brazil, US, UK,   
  Denmark, Spain, Mexico, Canada, Portugal
* Cultures\_Australia has a different distribution of activity levels compared to other cultures.
* Cultures like France, Italy and Germany do not have any activities as of now.

**Questions**

* What is Cultures\_None?
* Should Culture\_Australia levels be distributed like other cultures?

## Activity\_Type x Hits x Level



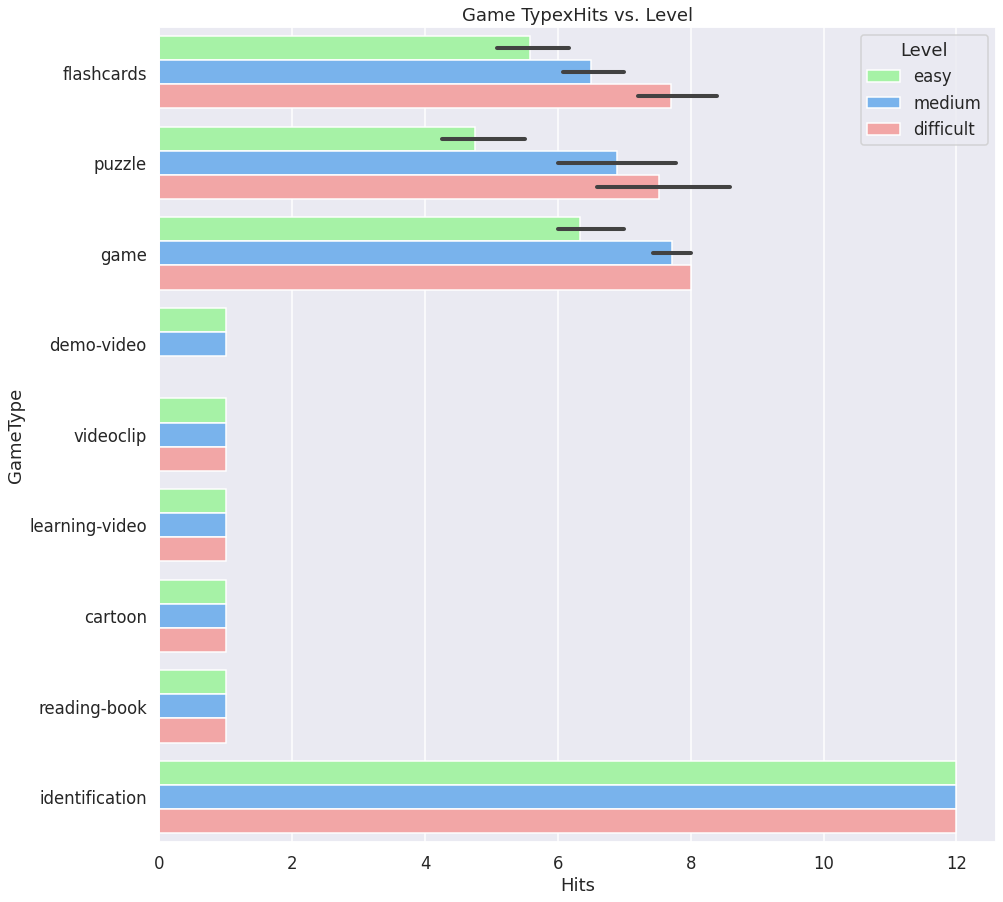
**Insights**

* The average **‘Hits’** for the **‘standard’** activity type seems to increase with the increase in levels.  
  The variance is low when compared to the **‘cultural’** activity type.
* The average **‘Hits’** for the **‘SEL’** activity type seems to increase and decrease slightly for level ‘**difficult**’.   
  No variance in level **‘difficult**’
* The average **‘Hits**’ for the **‘cultural**’ activity type seems to decrease slightly for level ‘**medium**’ and increase for level **‘difficult**’.   
  High variance among all three levels.

**Questions**

* Can more activities of higher hits of level **‘difficult**’ be added to **‘SEL**’ activity type?
* Can more activities of higher hits of level **‘medium**’ be added to ‘**cultural**’ activity type?
* Can the variance in hits be looked at for the **‘cultural**’ activity type?

## Game Type x Hits x Level



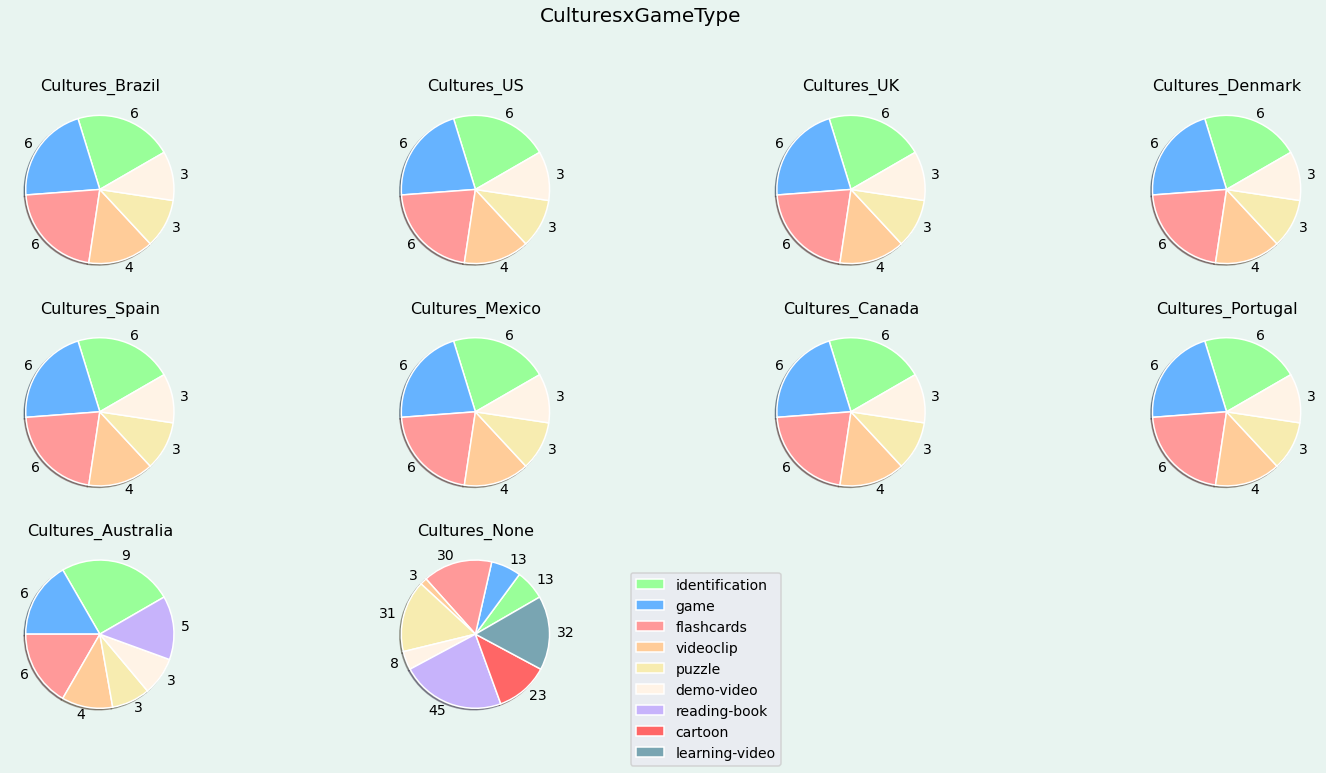
**Insights**

* Game Types like ‘**demo-video’, ‘videoclip’, ‘cartoon’, ‘reading-book’** and **‘identification’** have the same number of hits for   
  all the activities of different levels.
* Game Types like **‘flashcards’**, ‘**puzzle’**, and **‘game’** have different number of **hits** across different levels.  
  Also, we can say the variance in hits except for ‘**game’** Game Type’s ‘**difficult**’ level

**Questions**

* No ques from my end, please feel free to add in if you have any.

## Cultures x Game Type



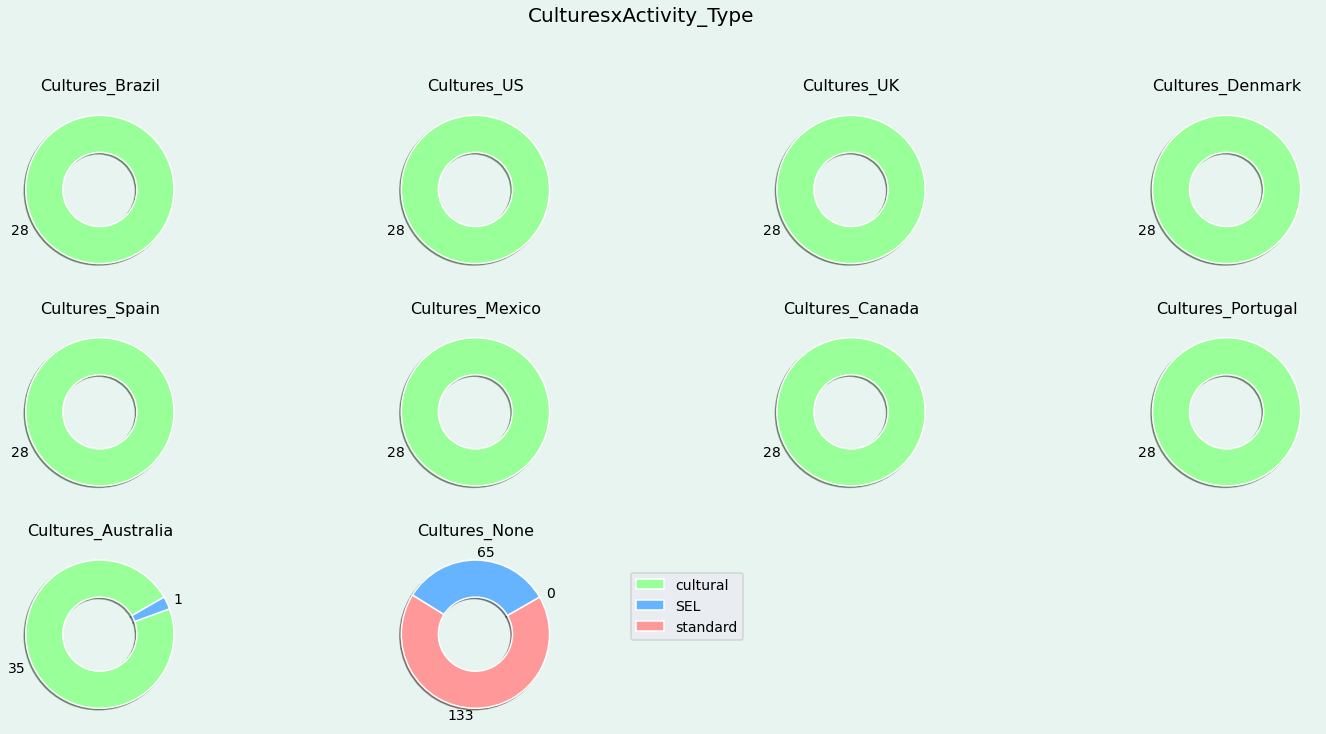
**Insights**

* Culture\_Australia has a different distribution compared to the rest of the Cultures.
* France, Germany and Italy do not have any activities.

**Questions**

* Can activities be added to missing cultures?
* Can other cultures also have all Game Type activities like in Cultures\_None?

## Cultures x Activity Type



**Insights**

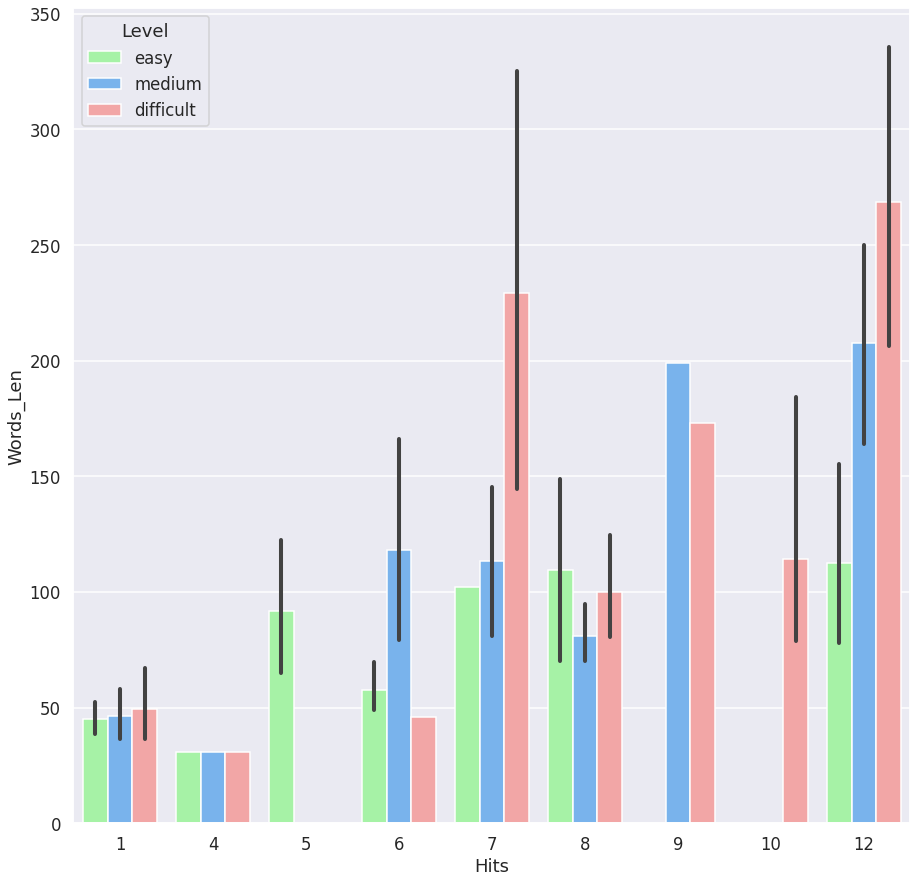
* All cultures except for Australia have only activities from the **‘cultural**’ Activity Type.
* Cultures\_None doesn’t contain any cultural activities, which makes sense.

**Questions**

* Should **SEL** activities be added to the rest of the cultures as well?

## Final Thoughts

* Please feel free to add in questions or thoughts that you want to be visualised, I will try to address them.
* The levels seem to have a fair distribution wrt to other parameters.
* Cultures like France, Italy, and Germany do not have any activities.
* I also did some analysis on how the length of ‘**words\_list’** varied with the number of hits, but I didn’t put it here as Leticia said that there is no direct relation between the two. Below is one of the graphs:



* Need to add activities of type **‘SEL**’ across all cultures.
* Activities like **‘learningvideo**’, ‘**reading-book**’ and ‘**demovideo**’ are not present in any of the cultures other than Cultures\_None.
* The distribution of **Game Type x Hits x Level** to be looked at once.
* The fairest distribution comes in the comparison of **Levels x Cultures**
* Activities of Game Type **demo-video** and **videoclip** are very few in number when compared to other activities.
* The colab notebook is not documented and is naive, I’d like to document it and share in the upcoming days.
* Also, I’d be doing EDA for Students\_Activities and Student\_Performance dataset so please do join me if you are interested.

Thank you for taking your time to go through the analysis, please do add in some suggestions! (:

**Add your questions/thoughts below:**