# 

# 

# 

# 

IST-718 Final Project

In-Game Assessment Scoring Engine

# 

# 

# 

# 

# 

# 

**Prepared By**

Jagannathan Govindan

Jason Min-Liang Kang

Ram Krishnan

Steve DeVito

# 

# 

[**Abstract**](#_k2z93k8gtazf) **3**

[**Problem Specification**](#_d3ejw4r2actf) **3**

[**Data Observation (Descriptive Analysis)**](#_f2mdvs4o2u8k) **4**

[About The Data](#_ns1tradiiv52) 4

[Visuals](#_ujqmpvmymp6w) 6

[**Data Modeling (Predictive Analysis)**](#_nbxaqmgrqiu5) **12**

[1. Generate a CatBoost Model](#_f2mdvs4o2u8k) 12

[2. Generate a XGBoost Classifier](#_f2mdvs4o2u8k) 13

[3. Decision Tree Model](#_f2mdvs4o2u8k) 15

[**Recommendation**](#_uqu2fo20nns2) **17**

[Our Recommendation](#_atypdqa3xuc8) 17

[PBS KIDS – Odd Squad](#_m0wh4826zhvn) 17

[WestEd Study Findings](#_c1x7brridie) 18

[**References**](#_g9mj08m8ni5a) **19**

# 

## 

# 

# Abstract

Based on our findings, the type of events that the user engages in has a correlation to the ultimate assessment score. Further analysis is needed to determine what are those events. Once those events are identified, game developers should devise a gameplay that allows users to be more engaging in those events.

# Problem Specification

This problem comes from Kaggle 2019 Data Science Bowl competition. It is the result of a collaboration between Kaggle (online data scientist community) and PBS KIDS (children programming aired by PBS). The goal is to gain insights into how gaming can help children learn important skills for success in school and life. The idea is to create a machine learning model that will predict scores on in-game assessments that will lead to better-designed games and improved learning outcomes. The model will aid in discovering important relationships between engagement with high-quality educational media and learning processes.

## 

# **Data Observation (Descriptive Analysis)**

## About the Data

The data is collected by an app called PBS KIDS Measure Up. It is an app that helps children ages 3 to 5 to learn early math concepts while navigating through a map and completing levels/assessments. The data records the number of attempts a child will take to pass a given assessment. **The training dataset is very large and has more than 11 million records.**

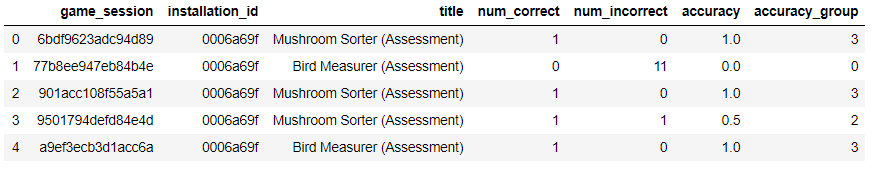
In the dataset, an assessment is a combination of multiple rows, and each row corresponds to an in-game event, identified by an event\_id. A group of events that constitute an assessment have the same event\_id. A complete list of event\_ids are given in the specs file. In the training dataset, the complete events of an assessment are given. In the testing dataset, events are truncated after the start event of an assessment.

The main data files which contain the gameplay events are train.csv and test.csv. The fields/columns in each dataset are as follows:

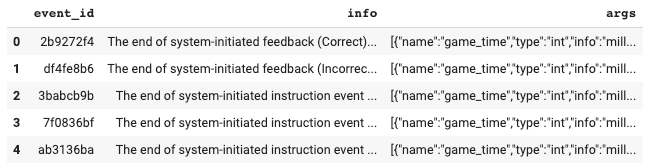
|  |  |
| --- | --- |
| event\_id | Randomly generated unique identifier for the event type. Maps to event\_id column in specs table. |
| game\_session | Randomly generated unique identifier grouping events within a single game or video play session. |
| timestamp | Client-generated datetime |
| event\_data | Semi-structured JSON formatted string containing the events parameters. Default fields are: event\_count, event\_code, and game\_time; otherwise fields are determined by the event type. |
| installation\_id | Randomly generated unique identifier grouping game sessions within a single installed application instance. |
| event\_count | Incremental counter of events within a game session (offset at 1). Extracted from event\_data. |
| event\_code | Identifier of the event 'class'. Unique per game, but may be duplicated across games. E.g. event code '2000' always identifies the 'Start Game' event for all games. Extracted from event\_data. |
| game\_time | Time in milliseconds since the start of the game session. Extracted from event\_data. |
| title | Title of the game or video. |
| type | Media type of the game or video. Possible values are: 'Game', 'Assessment', 'Activity', 'Clip'. |
| world | The section of the application the game or video belongs to. Helpful to identify the educational curriculum goals of the media. Possible values are: 'NONE' (at the app's start screen), TREETOPCITY' (Length/Height), 'MAGMAPEAK' (Capacity/Displacement), 'CRYSTALCAVES' (Weight). |

### 

**Table 1.1 Input Features**



**Table 1.2 Labels**

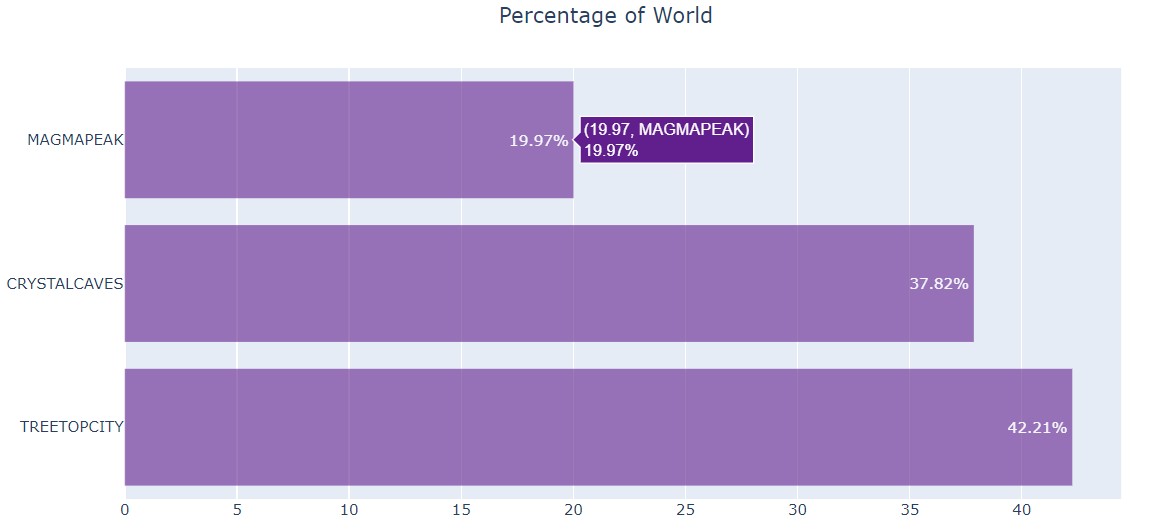
**Table 1.3 Specs**

### 

## Visuals

The game has three worlds and their percentage of users are shown in **Figure 1.1**

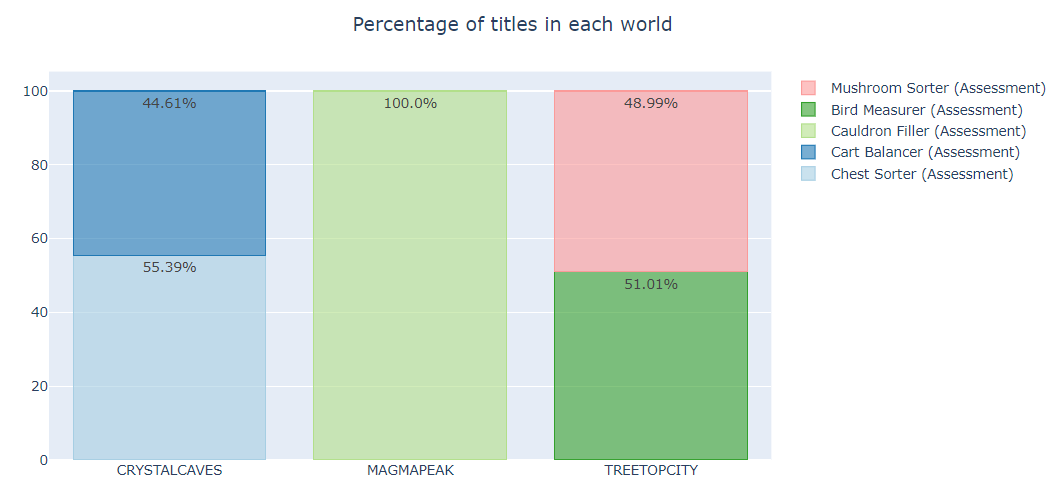
1. MAGMAPEAK (Capacity)
2. CRYSTALCAVES (Weight)
3. TREETOPCITY (Height & Length)



**Figure 1.1**

TREETOPCITY experienced the highest usage CRYSTALCAVES being a close second and MAGMAPEAK being third by a large percentage. This may be due to the progression of the game where you have to complete TREETOPCITY before moving on to CRYSTALCAVE and then finally MAGMAPEAK.

The titles in each world and their percentage of gameplay is shown in **Figure 1.2**

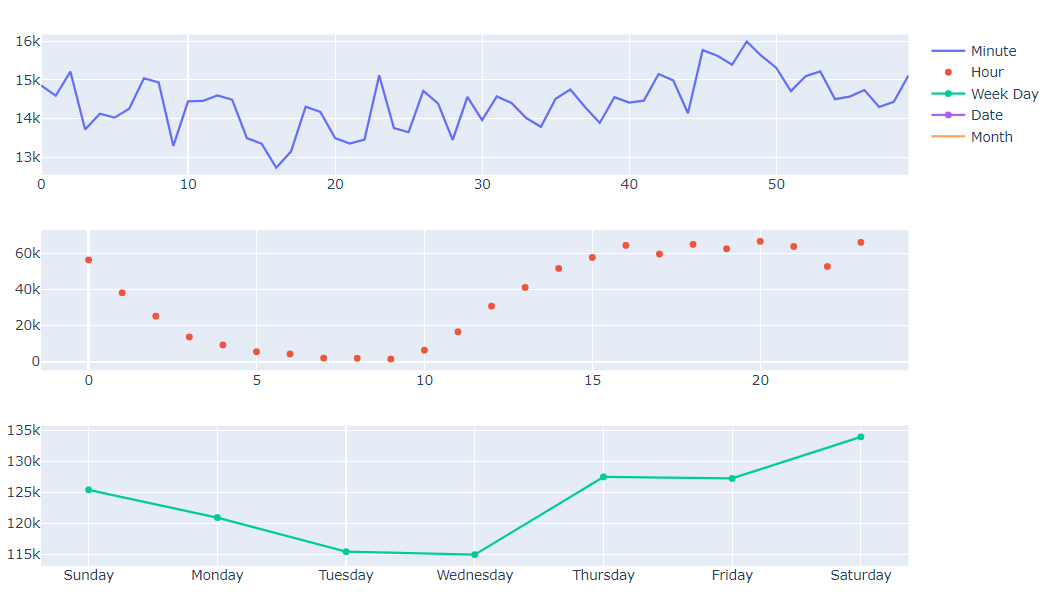


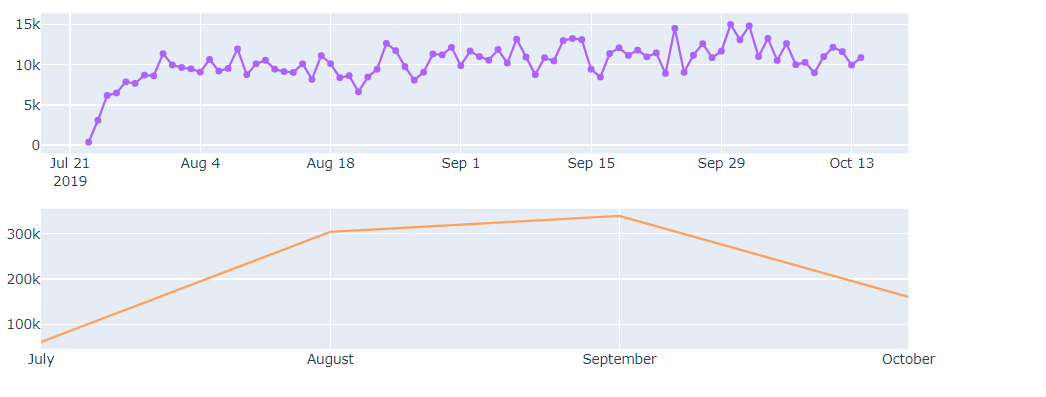
**Figure 1.2**

Each world included one or more assessments that needed to be completed, not necessarily successfully, before moving on to the next world.

From the **Figure 1.3,** the below observations can be made:

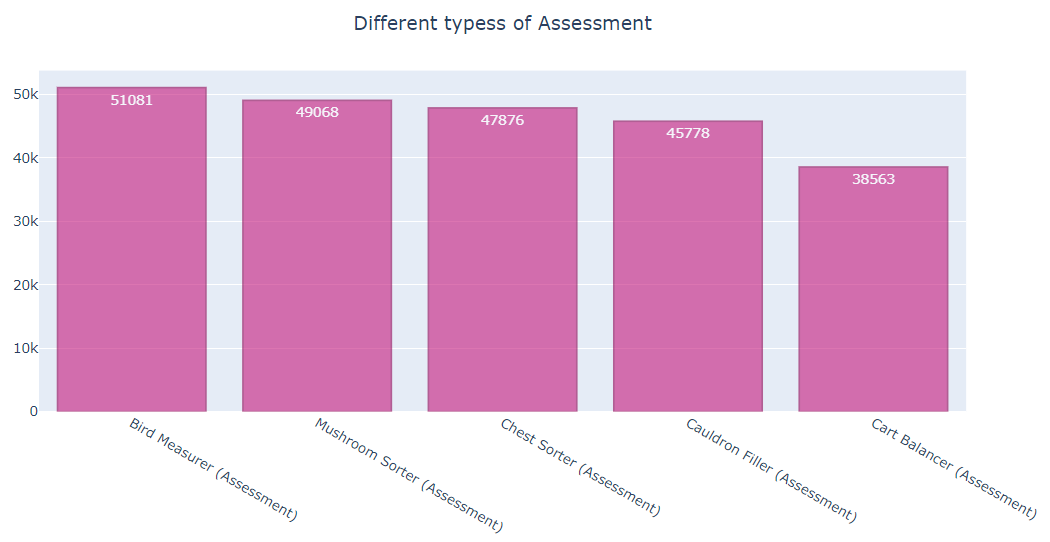
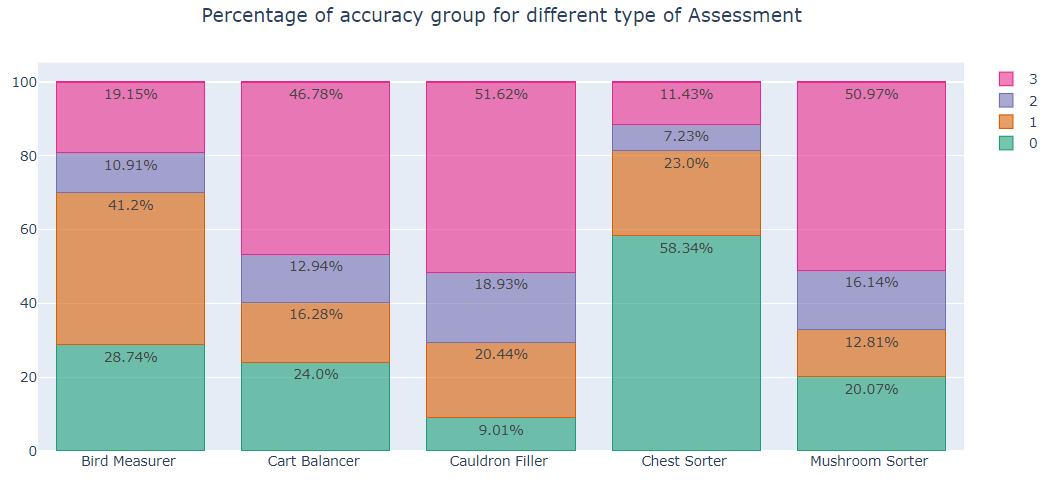
* Kids are more active from 10 AM till midnight.
* Kids are more active on Friday (Weekend starts)
* September has more traffic





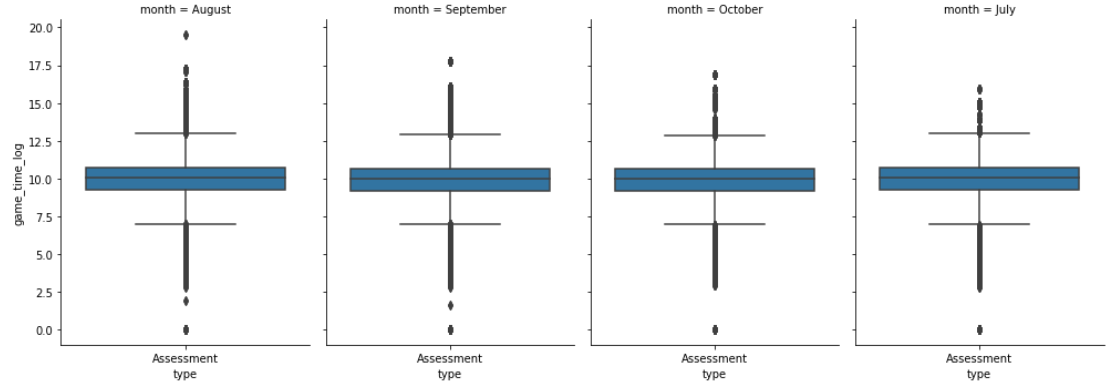
**Figure 1.3**

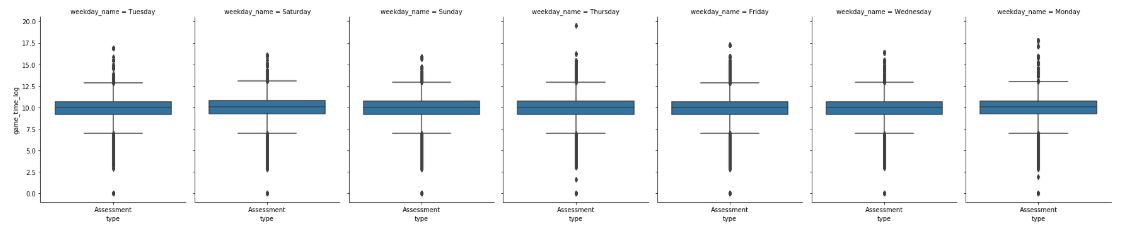
**Figure 1.4 and 1.5** shows that assessment in Cart Balancer are more easy to solve whereas the Chest Sorter had the hardest assessment



**Figure 1.4** **Figure 1.5**

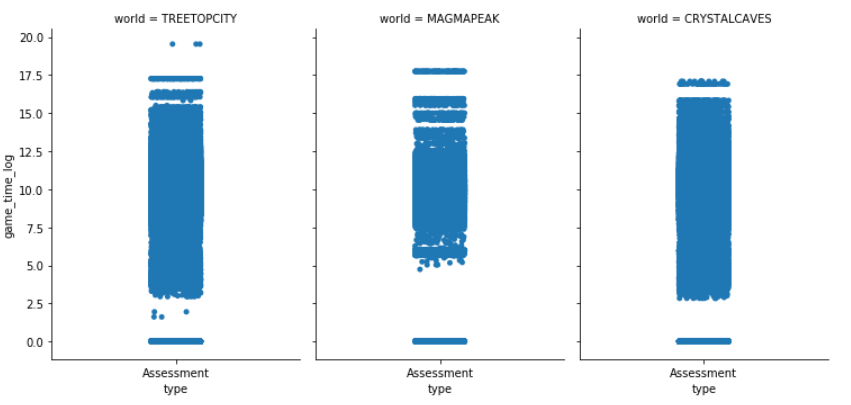
The distribution of assessments across month and week are shown in Figure 1.6 and these boxplots exhibit almost no difference in the pattern across month and weeks





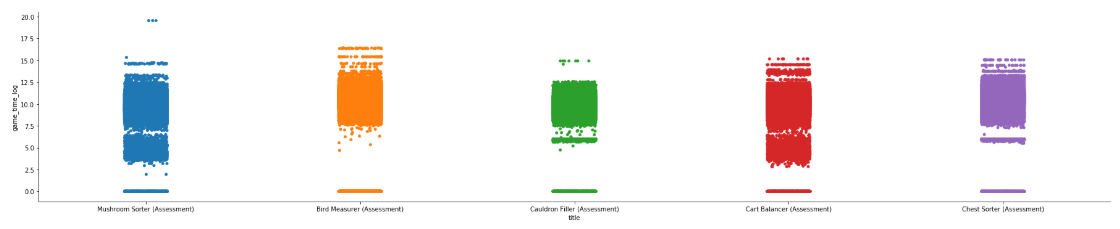
**Figure 1.6**

In **Figure 1.7**, it appears that MagmaPeak is where kids spent more time when compared to the other two worlds but less games are played in MagmaPeak when compared to other worlds. This is consistent with the way the game is structured where the user must complete one world before moving to the next with MagmaPeak being the last one.



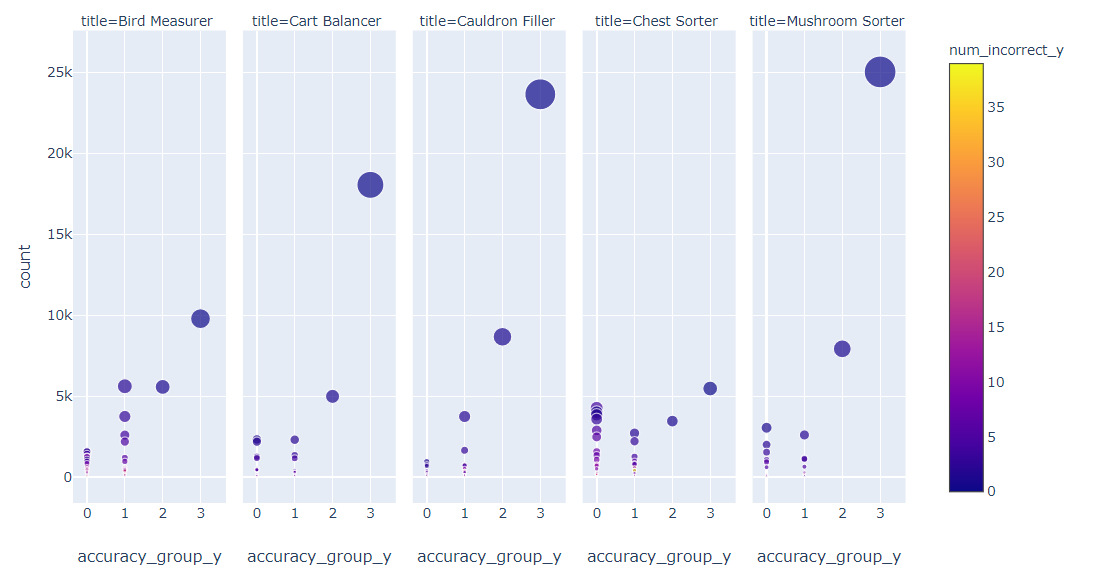
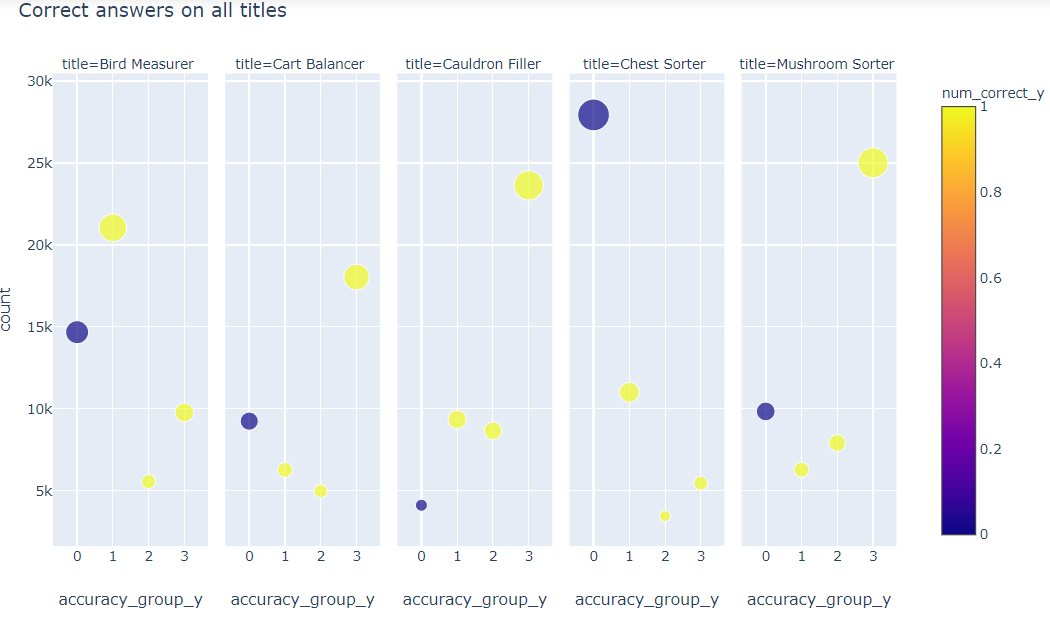
**Figure 1.7**

**Figure 1.8** shows the game time spent across all titles and the variations between them



**Figure 1.8**

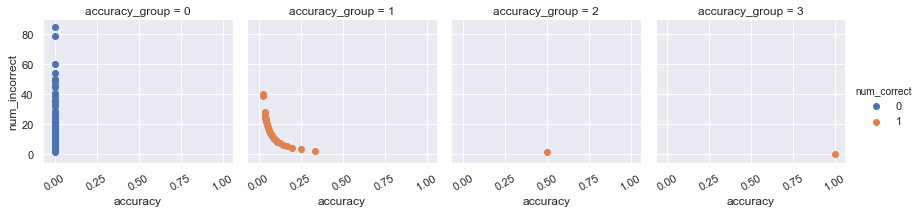
**Figure 1.8 and 1.9** shows the number of incorrect and correct responses selected by kids across different title which confirms the findings shown in **Figure 1.4 and 1.5**

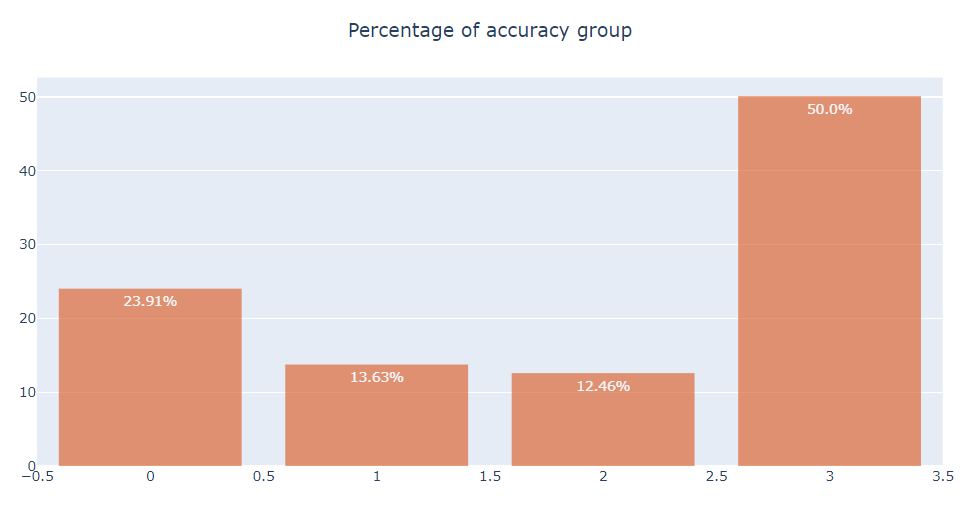
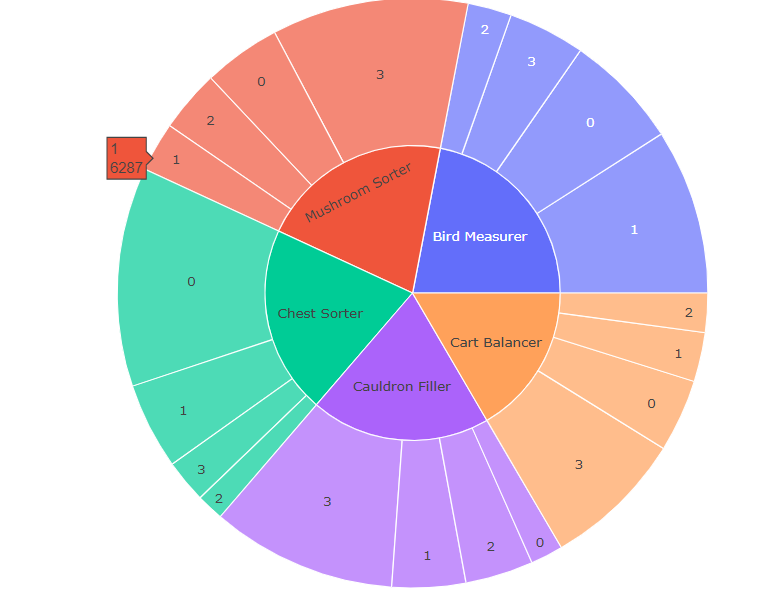
**Figure 1.9** **Figure 1.10**

Based on the below chart in **Figure 1.11,1.12 and 1.13**, the inferences are as follows

1. 0% accuracy belong to group 0
2. less than 50% accuracy belongs to group 1
3. 50% accuracy belongs to group 2
4. 100% accuracy belongs to group 3



**Figure 1.11**

**Figure 1.12 Figure 1.13**

### 

# Data Modeling (Predictive Analysis)

## 1. Generate a CatBoost Model

CatBoost is a recently open-source machine learning algorithm from Yandex. It can easily integrate with deep learning frameworks like Google’s TensorFlow and Apple’s Core ML. It can work with diverse data types to help solve a wide range of problems that businesses face today. To top it up, it provides best-in-class accuracy.

It is especially powerful in two ways:

1. It yields state-of-the-art results without extensive data training typically required by other machine learning methods, and
2. Provides powerful out-of-the-box support for the more descriptive data formats that accompany many business problems.

The “CatBoost” name comes from two words, “Category” and “Boosting”.

As discussed, the library works well with multiple Categories of data, such as audio, text, image including historical data.

“Boost” comes from gradient boosting machine learning algorithm as this library is based on gradient boosting library. Gradient boosting is a powerful machine learning algorithm that is widely applied to multiple types of business challenges like fraud detection, recommendation items, forecasting and it performs well also. It can also return very good result with relatively less data, unlike DL models that need to learn from a massive amount of data.

Hyperparameters:

CatBoostClassifier(loss\_function='MultiClass',task\_type="CPU",learning\_rate=0.05,iterations=3000,od\_type="Iter",early\_stopping\_rounds=500,random\_seed=21)

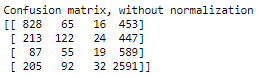
**Model Outcome:**

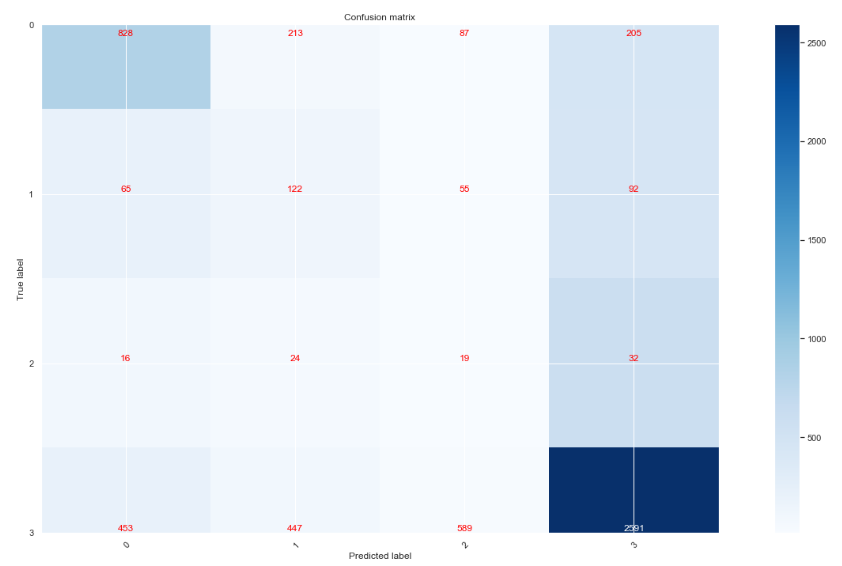
Accuracy: 0.61

precision\_score: 0.54

recall\_score: 0.61

f1\_score: 0.55





# 

## 2. Generate a XGBoost Classifier

XGBoost is an ensemble learning method. Sometimes, it may not be sufficient to rely upon the results of just one machine learning model. Ensemble learning offers a systematic solution to combine the predictive power of multiple learners. The resultant is a single model which gives the aggregated output from several models.The models that form the ensemble, also known as base learners, could be either from the same learning algorithm or different learning algorithms. Bagging and boosting are two widely used ensemble learners. Though these two techniques can be used with several statistical models, the most predominant usage has been with decision trees.

Some features of XGBoost that make it so interesting.

**Regularization**: XGBoost has an option to penalize complex models through both L1 and L2 regularization. Regularization helps in preventing overfitting

**Handling sparse data**: Missing values or data processing steps like one-hot encoding make data sparse. XGBoost incorporates a sparsity-aware split finding algorithm to handle different types of sparsity patterns in the data

**Weighted quantile sketch**: Most existing tree based algorithms can find the split points when the data points are of equal weights (using quantile sketch algorithm). However, they are not equipped to handle weighted data. XGBoost has a distributed weighted quantile sketch algorithm to effectively handle weighted data

**Block structure for parallel learning**: For faster computing, XGBoost can make use of multiple cores on the CPU. This is possible because of a block structure in its system design. Data is sorted and stored in in-memory units called blocks. Unlike other algorithms, this enables the data layout to be reused by subsequent iterations, instead of computing it again. This feature also serves useful for steps like split finding and column sub-sampling

**Cache awareness**: In XGBoost, non-continuous memory access is required to get the gradient statistics by row index. Hence, XGBoost has been designed to make optimal use of hardware. This is done by allocating internal buffers in each thread, where the gradient statistics can be stored

**Out-of-core computing**: This feature optimizes the available disk space and maximizes its usage when handling huge datasets that do not fit into memory

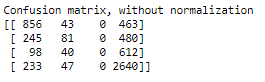
**Model Outcome:**

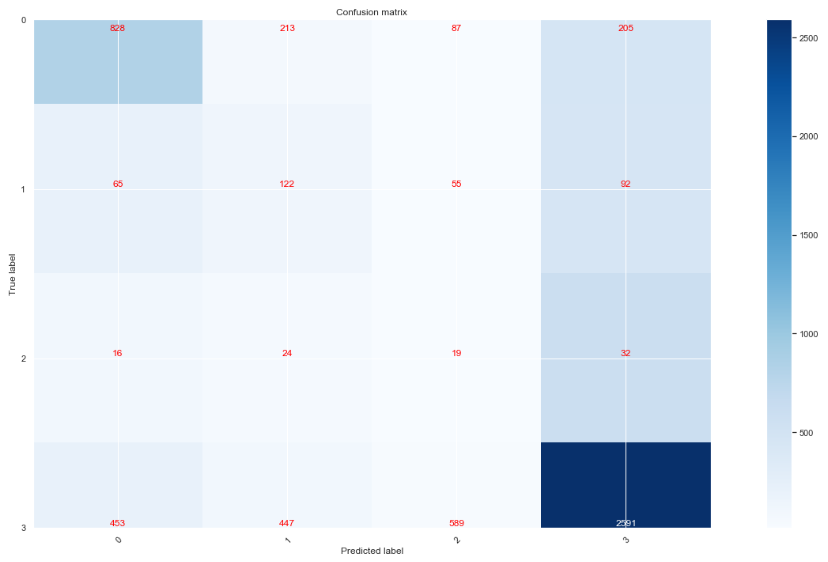
Accuracy: 0.61

precision\_score: 0.51

recall\_score: 0.61

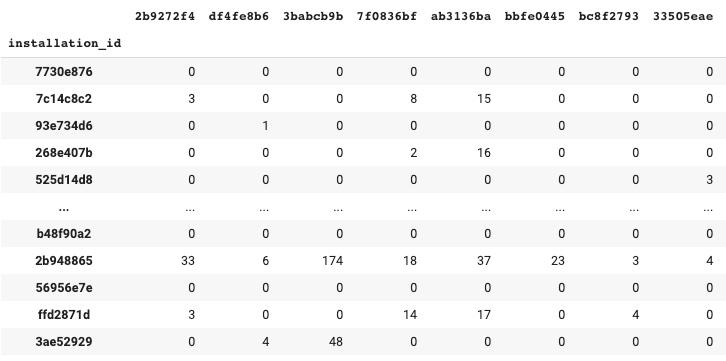
f1\_score: 0.54





## 3. Decision Tree Model

The previous approaches are based on analyzing the content inside event\_data. The resulting accuracy scores are not desirable. In this approach, analysis is performed on event\_id and installation\_id. The idea is to see whether event\_id (the type of events the user engage in) alone can draw a conclusion on the result. To perform the analysis, event\_ids are accumulated for each installation\_id (see below figure).



The resulting model yields 83% accuracy.

# **Recommendation**

## Our Recommendation

Based on our findings, the type of events that the user engages in has a correlation to the ultimate assessment score. Further analysis is needed to determine what are those events. Once those events are identified, game developers should devise a gameplay that allows users to be more engaging in those events.

## PBS KIDS – Odd Squad

PBS KIDS designed and developed the odd squad television series and transmedia suite, which included educational media, television episodes, online games and hands on resources

Table 1 and 2 shown here shows the learning objective and activities of this program

# 

## WestEd Study Findings

A study conducted by WestEd on Odd Squad (Learning Math with PBS KIDS Games and App) showed improvement in skills acquired by kids

Table 3 and 4 shown here compares the metrics accessed from KIDS pre and post Odd Squad programme

# 

# **References**

<https://www.kaggle.com/c/data-science-bowl-2019>

<https://www.analyticsvidhya.com/blog/2017/08/catboost-automated-categorical-data/>

<https://www.analyticsvidhya.com/blog/2018/09/an-end-to-end-guide-to-understand-the-math-behind-xgboost/>