Task 1 **Statistics**

a. Narrower

c. 0.166

e.

Task 2 **Data Analysis**

**Introduction**

This report is based on the exploratory data analysis of 5,000 players interactions with kwalee game app over 30 days.

**Exploratory Data Analysis of Players Information**

The players.csv was loaded in the jupyter notebook as a pandas dataframes, the dataset is completely cleaned with no missing values, 6 columns and 5000 rows.

Table 1 players information

|  |  |  |
| --- | --- | --- |
| Column | Non-Null Count | Data type |
| install\_datetime | 5000 non-null | object |
| player\_id | 5000 non-null | object |
| platform | 5000 non-null | object |
| country | 5000 non-null | object |
| screen\_size | 5000 non-null | Float64 |
| system\_memory | 5000 non-null | Int64 |

**Statistical Summary of Screen size and System memory**

Table 2 statistical analysis of Screen size and system memory

|  |  |  |
| --- | --- | --- |
| **Statistics** | **Screen\_size** | **System\_memory** |
| count | 5000.000000 | 5000.000000 |
| mean | 5.998411 | 3131.348400 |
| std | 1.551791 | 1615.828459 |
| min | 3.887301 | 856.000000 |
| 25% | 4.694410 | 2001.000000 |
| 50% | 5.766932 | 2845.000000 |
| 75% | 5.766932 | 3741.000000 |
| max | 13.768170 | 11562.000000 |

The exploration of players information was computed using the unique() function in python to view all unique entries per column. The players installed the game app mostly on two platforms ios and android from four countries i.e., US, AU, GB, CA.

Table 3 max values information per column

|  |  |
| --- | --- |
| **Column** | **Max** |
| install\_datetime | 2020-12-08 00:55:49.178000 |
| player\_id | 00b3e42f0aa8d3e4d40f6d2817bd6615 |
| platform | ios |
| country | US |
| screen\_size | 13.76817 |
| system\_memory | 11562 |

Table 4 min values per column

|  |  |
| --- | --- |
| **Column** | **Min** |
| install\_datetime | 2020-11-14 16:23:07.922000 |
| player\_id | 00001bc5552cd394a4b25b2ceb4ae7cd |
| platform | android |
| country | AU |
| screen\_size | 3.887301 |
| system\_memory | 856 |

Table 5 Mode values per column

|  |  |
| --- | --- |
| Column | Mode |
| player\_id | 009827dc8a88dfab41d9646d9593d23a |
| platform | ios |
| country | US |
| screen\_size | 4.69441 |
| system\_memory | 2845 |

Table 6 count of platform

|  |  |  |
| --- | --- | --- |
| Platform | counts | Percentage (%) |
| ios | 3285 | 65.7 |
| android | 1715 | 34.3 |

Table 6 showed the percentage of ios users as 65.7% while android users as 34.3%, more ios related ads can be integrated into the game app to increase ads clicks and profit margins.

Table 7 Counts of country

|  |  |  |
| --- | --- | --- |
| Country | counts | Percentage(%) |
| US | 3852 | 77.04 |
| GB | 582 | 11.64 |
| CA | 358 | 7.16 |
| AU | 208 | 4.16 |

Table 7 showed that most players are in the United States, more United States related ads should be integrated into the game app to increase clicks and conversion rate.

**Exploratory Data Analysis of Level progress information**

The level\_progress.csv was loaded in the jupyter notebook as a pandas dataframes, the dataset is completely cleaned with no missing values, 5 columns and 55826 rows, level\_number as the integer while other columns are object type.

Statistical Summary of level\_number

Table 8 statistical summary of level number

|  |  |
| --- | --- |
| Statistics | Value |
| count | 55826.000000 |
| mean | 5.991778 |
| std | 5.988455 |
| min | 0.000000 |
| 25% | 2.000000 |
| 50% | 4.000000 |
| 75% | 8.000000 |
| max | 49.000000 |

The unique() function in python was used in exploring the columns. There are three main statuses in the game; complete, start and fail, and level number 0-49.

Table 9 Max and Min Values for columns

|  |  |  |
| --- | --- | --- |
| Column | Max | Min |
| event\_id | 2020-12-16 01:17:22.240000 | 2020-11-06 19:44:32.930000 |
| player\_id | 00b3e42f0aa8d3e4d40f6d2817bd6615 | 00001bc5552cd394a4b25b2ceb4ae7cd |
| level\_number | 49 | 0 |
| status | start | complete |
| session\_id | fff568045a2a08080aaea4df7afc18b0 | 000e7a2c724b6e9dba649f341bcc3736 |

Table 10 Mode values per column

|  |  |
| --- | --- |
| Column | Mode |
| player\_id | 006c4c66f9976d96fb0bf9852b214b30 |
| level\_number | 1 |
| status | start |
| session\_id | 7c354840caa7e3790cfcd7a4d8e97281 |

Failure increases the risk of churn because the value count of level\_number showed that count of players playing session decreases as the level progresses i.e., from 7161 in level 1, to 1 in level 49. It means that the proportion of game players decreases as the difficulty level increases from 1 to 49. Meaning that some players quit playing as the difficulty level increases in the game due to failure. This could decrease clicks and ads conversion due to high churn rate of the product i.e., the game app, creating more easy levels in the game development increases players interactions with the game, reducing churn rate and therefore results in increased retention rate leading to more clicks and ads conversion.

Table 11 status count

|  |  |  |
| --- | --- | --- |
| status | count | Percentage % |
| start | 31109 | 55.72 |
| complete | 20571 | 36.85 |
| fail | 4146 | 7.43 |

Table 11 showed 55.72% starts sessions, 36.85 completed/successful sessions and 7.43% failed sessions

The two datasets players.csv and level\_progress were paired at the player\_id column using pd.merge() function. New dataframe was produced having 70816 rows and 10 columns. There were no missing values in the dataframe.

|  |  |  |
| --- | --- | --- |
| Column | Non-Null count | Data type |
| install\_datetime | 70816 | object |
| player\_id | 70816 | object |
| platform | 70816 | object |
| country | 70816 | object |
| screen\_size | 70816 | Float64 |
| system\_memory | 70816 | Int64 |
| event\_datetime | 70816 | object |
| level\_number | 70816 | Int64 |
| status | 70816 | object |
| Session\_id | 70816 | object |

**Visualization**

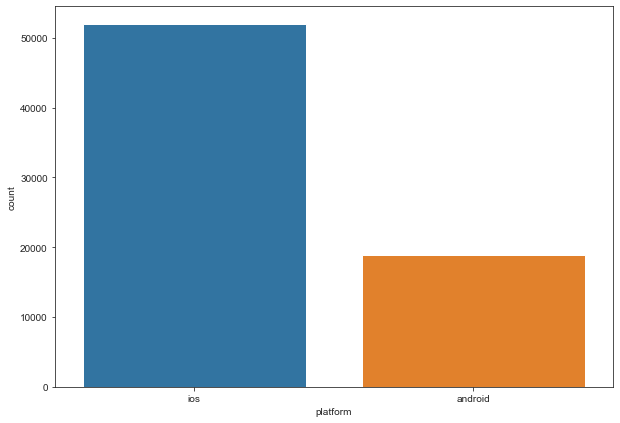
Figure countplot of status

**Chart, bar chart

Description automatically generated**

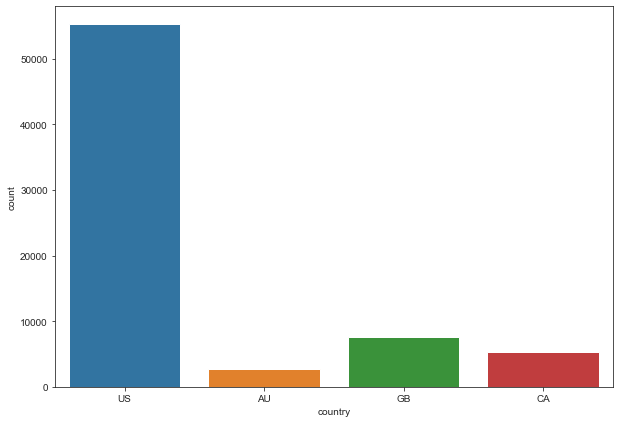
**Fig 1** showed the count of players starting sessions, count of players completed sessions and count of players failed sessions.

Figure Countplot of platform

****

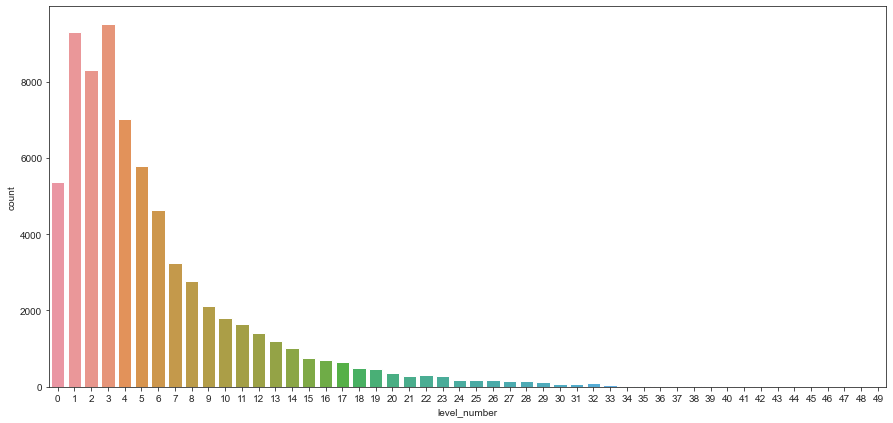
**Fig 2** showed that more game players are on ios than android, hence more ios related ads should published to increase clicks and conversion rate, also the game features should be optimized to increase usefulness, user experience of more android userssuch as game speed on android, user interface**.**

Figure country countpot

****

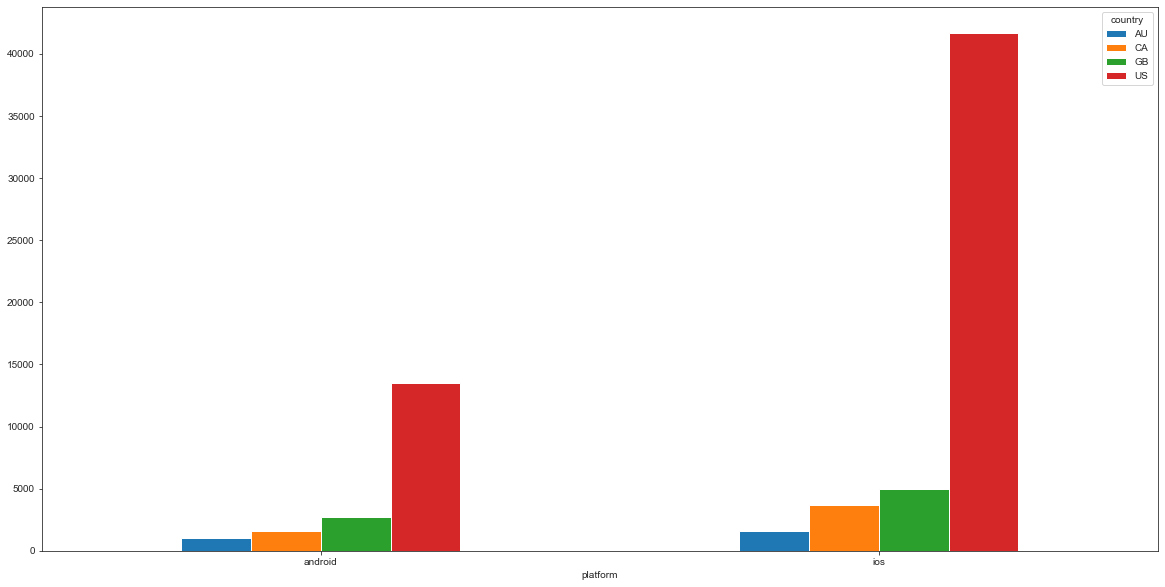
**Fig 3** showed that more game users are from the United States hence more United States related ads should be published to retain more players and game content should integrate familiar interface and scenario with AU, GB, and CA to increase game players in the countries with less players interactions.

Figure count of level number

****

**Fig 4** showed that count of level number decreases as the level increases, which means that there is decline in players interactions as the game becomes more difficult, this leads to high churn rate, more easier game level would increase more user interactions and more playing time which increases ads clicks and conversion rate**.**

Figure grouped bar plot of platform and country

****

The grouped bar plot showing the interactions on ios and android platform amongst the four countries, United States recorded the highest user interactions on the ios platform than android platform. AU recorded the lowest user interaction on both platforms. The most game players from the United States are on the ios platform.

Figure grouped bar plot of platform and status

Chart, bar chart

Description automatically generated

Fig 6 showed that most game sessions were started from the ios platform, highest fail was recorded on the ios platform and highest complete was recorded on the ios platform. Android platform recorded he lowest user game status.

Figure plot of platform and level status

Chart, histogram

Description automatically generated

Fig 7 showed more user interactions and level number on ios platform than android platform.

Figure group bar plot and status

Chart, waterfall chart

Description automatically generated

Fig 8 showed that United States recorded most game start sessions, completed session, and failed session. Meaning most players from the United States recorded highest user status. AU recorded lowest user sessions.

Figure grouped bar plot level number and country

Chart, histogram

Description automatically generated

Fig 9 showed that the highest games interactions across all levels was achieved by players in the United States.

Figure group bar plot of level number and status

Shape

Description automatically generated

Fig 10 showed decline in complete, start and fail status as level number increases, which translate to high churn rate as game becomes difficult.

**Time series Analysis**

Install\_datetime and event\_datetime were converted to datetime64 series.

Figure plot of event\_datetime and level\_number

Chart, histogram

Description automatically generated

Fig 11 showing the level number trends achieved by players across datetime, there was a minimum level number 0 in 2020-11-08 and the highest-level number on 2020-12-01.

Figure screen size and event datetime

Chart, histogram

Description automatically generated

Fig 12 showed trend based on screen size and event datetime.

Figure plot of install datetime

Chart

Description automatically generated

Fig 13 showing trend of install\_datetime, all installation took place in 2020.

Figure plot of the new dataframe

Chart

Description automatically generated

Plot showed trend between event datetime and system memory.

The trend from the Time series analysis showed that ads should be scheduled at peak event time of user interactions to increase clicks and conversion rate.

I would perform Forecasting and predictive modelling to determine player status based on other features like screen size, platform and country etc.