**Q1**

You have been asked to take this dataset and compile a one-page report for the entire studio’s benefit.

a) *At a summary level, please describe the form and content of this report.*

**Purchase**

* Raw sales
  + Graph showing Weekly Installs by Date to understand sales and tie-back sales to marketing / advertising initiatives over time
* Customer engagement
  + New vs old users: New Unique Users as a proportion of Daily Unique Users plotted against Date. This shows how much of your playing population is being retained as a longer playing gamer
  + Customer evolution: An area graph combining 1, 3, 7 and 30 day returning users with time. With time this will hopefully show the conversion of users from 30 day towards the 1 day returning users indicating that players are becoming more engaged with the game with time. Area graphs are often frowned upon as the totals’ edges can be misunderstood. This graph is clearer when done as a percentage as no total can be inferred - see below.

**Game Play**

* Most compelling game features
  + Look at Pearsons CC between Mean Session Length and each of: Hero Win Rate, Creatures Killed, Gold Spent, Gold Earned, Chests Opened. This shows what elements of the game are most compelling to players for future product design

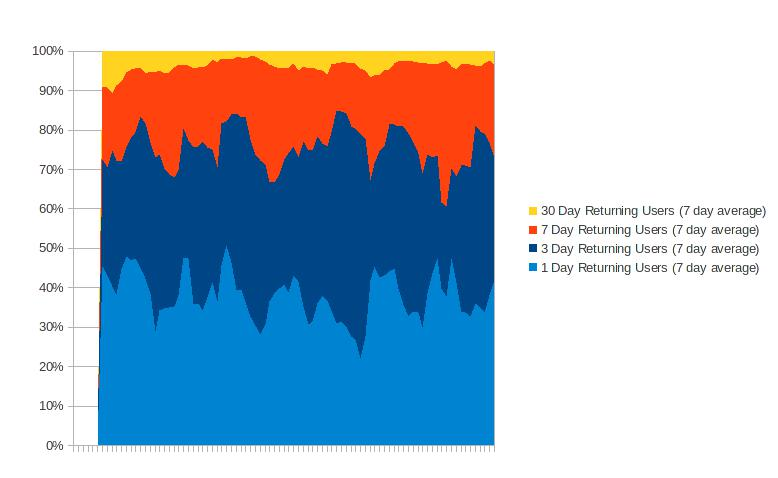
b) *Please describe how you would go about creating this report. Valuable detail would include tools used, techniques used, intended functionality and how the information would be conveyed.*

This is a report which can easily be constructed using Excel.

* Raw sales is a line graph of Weekly Installs against Date.
* New vs old users is made with an area graph of Date on the X axis vs these on the Y:
  + New Unique Users
  + (Daily Unique Users - New Unique Users) to indicate “Old” users
* Customer evolution is demonstrated below and implemented in dataset1a.ods
* Use the PEARSON(arr1, arr2) function to compare Mean Session Length and Hero Win Rate. Then do the same with the other listed game features. Present this as a set of bars on a bar chart with a scale from +1 to -1 on the Y axis.

c) *Please create a demonstration version of one element within this report. Please also provide any materials used to generate that element.*

The graph below shows the percentage of players that are 1, 3, 7 and 30 day users. This was created using the file dataset1a.ods (Open Office file format).



d) *Assuming a long-term commitment to a report of the form you specified above, are there any changes would you make over time?*

* Compare Weekly Installs with previous years’ data
* Contrast different games and the fall-off of interest from one game to another
* Causality between Mean Session Length and Hero Win Rate, Creatures Killed, etc is weak when game playing sessions are aggregated. It would be better to get each player’s playing time and then looking for correlation by user, not summations

**Q2**

a) *What recommendations would you give to the Business Manager as regards player engagement monitoring?*

* Record what game achievements / failures each player makes individually. Do longer players use more multiplayer features, open more chests per time period, etc.
  + What was the activity that happened immediately before a user turned off
* Record the start and stop time for each session played. What time of day (or day of the week) fits with longest play - consider making the game harder / easier accordingly
* Acquire information about the demographics that make up longer and shorter players for marketing purposes
  + Identify one or a set of perfect players with the aim of making more players like them by modifying the game
* Record baseline data to compare: collect or find data from other games to contrast: consider if it is more useful to use the same or different genre
* Anonymise data appropriately so customer information and company image is safeguarded.

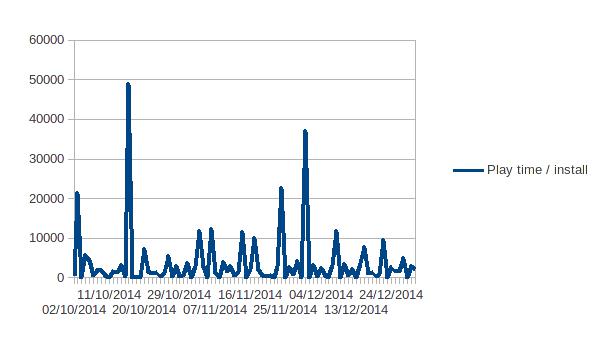
b) *What information does Dataset 1 contain which may be useful in regards to tracking long-term player engagement?*

Daily Installs and LTD Play Time (Minutes)

c) *Using Dataset 1, create an initial assessment of long-term player engagement challenges, issues and key data points. Please also provide a brief description of the techniques used.*

In the available data, Daily Installs is the field most closely resembling actual sales. (But more complicated because a player may install multiple times on the same / multiple devices, depending on recording of the data / licensing restrictions).

Plotting time played per day against the number of players, outputs the graph below:



A more useful, interactive version is here: [*https://plot.ly/46/~noelevans*](https://plot.ly/46/~noelevans). Note: This plot was created using the file dataset1b.ods.

Overall, the analysis shows an upwards trend in use per day. The data has lots of spikes which fall around public holidays. While the upwards trend can be seen through the whole period, it is hard to draw a trend line through the data as the spikes would distort the line significantly.

The data we have is aggregated so it does not give an idea of how - on a given day - new and older players are behaving on that day. For example if the date is 25 December, there may be a very significant rise in game-minutes played but a significant proportion is likely to be “first day binge” rather than a more representative day where short and long term players contribute more equally.

d) *At a high level, assess the utility of your initial assessment.*

This approach is a quick first assessment which does not allow specific users to be identified. It would be better to identify specific users based on game playing traits to analyse those identified as long-term players.

**Q3**

a) *Generate an initial 18-month revenue prediction for this title. Please provide a description of your methodology and any relevant materials.*

First the file mod\_dataset2.csv was made which contains only the relevant predictors:

* title Life To Date (months)
* LTD Revenue
* platforms
* genres

In the file, the platform and genre information was normalised to be a set of booleans for each platform and genre i.e. is\_pc, is\_moba, etc. As this is an initial prediction a linear model was used to predict LTD Revenue - month by month - using the platform, genre and revenue information for the other games using the R code in file q3.html. The model for LTD Revenue for a FPS as described is:

LTD Revenue = 28134954 \* month + 153849773

So at 18 months, the predicted revenue is: £660,200,000.

b) *Would you make any comments or observations regarding your revenue prediction?*

This is a built using linear regression and so in no way takes in to account the initial first day sales that an anticipated title may receive followed by quadratic-like decrease in sales for a given month.