

Spreadsheet Data Analytics

CONTINUOUS ASSESSMENT 1 – CASE STUDY APP USAGE ANALYSIS

ASSESSMENT

This assessment is due on Sunday 24th October 2021 and is worth 10%. It should be submitted through Moodle in the form of a single excel file which includes a number of sheets including one named Notes, three for the features specified below and one for the prediction task below. It will also include a number of sheets with data and analysis. The Marking scheme is as follows:

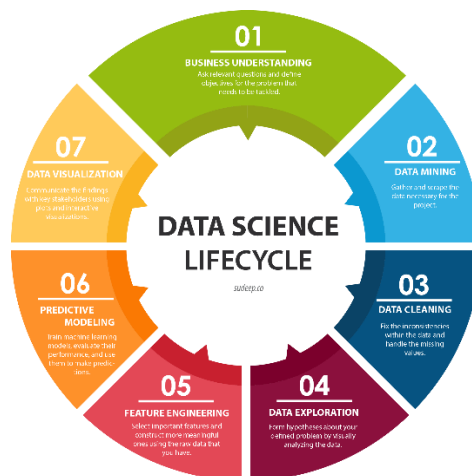
Spreadsheet structure including notes and details of checks and cleaning: 20%

Analysis of features 2,3, 5 & Prediction 2: 80% (20% each).

Background

In this exercise we will investigate each of the steps in the Data Analytics Lifecycle shown below through a real (but anonymized) data set which details the usage of a mobile app in the early months post release. A number of the tasks will form part of CA1; we will work through the remainder of the tasks in class.

Data Analytics Lifecycle



Investigation

Step 1: Business Understanding

The data file titled “AppUsage.xlsx” available on Moodle contains the registration and session dates and times for users of a mobile App as recorded on the server. It contains

- Email (anonymized)
- Gender (1-Male, 2-Female)
- Date of birth
- Location of user (1-US, 2-Europe, 3-Africa, 4-Asia, 5-South America, 6–Australasia)
- App registration – date and time.
- App session – date and time at start of session.

Note that the data is incomplete for the period after 31 March 2018. All this data should be ignored.

Note that the data is incomplete for periods post 31 March 2018 so any data after this point should be ignored.

Examine the fields in the data set. Features 1-5 and Prediction 1 below set out a number of important business problems or questions that this data set could potentially answer.

These questions may relate to the profile of users in terms of age, gender and location, the number of users over time and also whether usage by an individual is changing over time and particularly if users are discontinuing their usage of the app.

Step 2: Data Mining

In this case the data has been provided to you. In other problems you may decide there is other data from other apps which might be good for comparison purposes. For this assessment there is no need to access other data.

Step 3: Data Cleaning

It is good practice to first make a copy of your data set and perform whatever checks and cleaning on the data as you see fit on this copy. Note in a separate sheet named “Notes” the checks you have made and any changes to the data when cleaning it and also the details of the cleaning checks you have performed. You should also use this sheet to detail the *version history* of the document as well as any other important notes.

Step 4: Data Exploration

For this activity you would normally examine the distribution of each of the variables often using average, median, mode, max and min functions. You often include histograms and frequency distributions which we will look at later in the course. For this assessment we will perform some of this exploratory activity. We will look at it in more detail in later assignments and exercises.

Step 5: Feature Engineering (Analysis)

Feature 1: Gender Effect

Using appropriate analysis investigate whether there is a gender difference in registered users. Display the results visually and include an appropriate interpretation of your findings.

You should first create a summary sheet with a single row for each registered user which shows the sex, age and location of the user. This summary sheet is easier to analyse than the complete raw data and will form the basis for much of the analysis in this exercise. You can use the Remove Duplicates

[You will then need two columns in your solutions sheet, one for the genders and one to count the number of users for each gender using a COUNTIF function. A third column can be used to calculate the percentage in each location. A pie chart may be a good way to display this information.]

Feature 2: Location Effect

Investigate the location profile of registered users. Display the results visually and include a short interpretation.

Feature 3: Age Effect

Using appropriate analysis investigate the age distribution of users. Is there a gender difference in the age distribution? You should measure the average and spread of the age values and you should also produce a histogram for the age of the users. Display the results visually and include a short interpretation.

Feature 4: Times of Sessions

Investigate the times of sessions. Is there any useful information for the App's product manager in this data? In particular, maintenance and updates can take up to one hour and so when should the company schedule this work to minimize disruption? Include a short interpretation.

Feature 5: Usage Patterns 1 (Hard Problem!)

How often do users use the App? Produce metrics to analyse this and investigate if there is a difference between genders. The usage value can be found as either the Sessions per Day (divide number sessions by App days) or as the Days or hours between Sessions (divide App days by number of sessions).

Feature 6: Retention Rate (Hard Problem!)

It is very important to understand how many users are still using the app at a certain point. If retention is defined as the percentage of app users who have used the app in the previous 4 weeks, find the retention rate at the end of each month and explore whether this rate is increasing or decreasing.

For those users who are not retained at 31/03/2018 investigate the number of sessions that app users record before they stop using the app.

Note you will need to exclude all registrations that occur in the four week period previous to the time point you are examining.

Step 6: Prediction

Prediction 1: Growth in Users

The product manager for the app wants to understand the growth in the number of registered users. Investigate the trend and produce a rough prediction for the number of new registered users that there are likely to be in the three month period from 1 April through to 30 June 2018. You will need to use the regression method for the prediction. Comment on your prediction.

[In your analysis sheet you might look to find the number of new registrations in each month. To do this you might create a column with dates from 1 May 2017, 1 June 2017 etc through to 1 April 2018 and then a second column to find the number of registrations

before these dates. Using subtraction, you can then find the number of registrations in each month. You can plot this data using a scatter plot.

To predict future registrations, you can add a linear trendline to the graph and then you can either extend the trend line and visually determine an approximate value for the number of future users or you can use the trendline equation to give a more exact prediction.]

Prediction 2: Growth in Users by Gender

The company want to predict the growth in number of users in each of the US, EU and Africa regions. Develop appropriate predictions and comment on their potential accuracy.

Step 7: Data Vizualisation

It is important to explain your findings in visual form if necessary. In each of your analysis sheets you should include a clear summary of your findings in appropriate form possibly including visual, textual and numerical elements.