School of Information Technology

King Mongkut's University of Technology Thonburi

Project Report

"***Talking Datamobile User Demographics***"

1. Yanisa	Phrompeng	56130500208
2. Panrapee	Satienpisarn	56130500220
3. Jutarut	Kongrith	56130500262
4. Nijjaree	Chuencharoensuk	56130500268

This report has been prepared as an assessment requirement
for Data Mining(CSC531)
and submitted to Asst. Prof. Dr. Chakarida Nukoolkit.

PART I: Preparing Data

Before we start this project, we have to download file name:

- Mobile_Apps_Original_CategoryLabels.csv file
- Mobile_Apps_Usage.csv file

Which Mobile_Apps_Original_CategoryLabels file is contain the data of application types and Mobile_Apps_Usage file composed of application id. Then, we analyze and separate data into 15 categories which are:

1. Games
2. Education
3. Sport
4. Books
5. Entertainment
6. Music
7. Schedule
8. Information Technology
9. Fashion
10. Transportation
11. Family
12. Restaurant
13. Shopping
14. Financial
15. Other

After, we divided the data of application types into 15 categories. We merged Mobile_Apps_Original_CategoryLabels file and Mobile_Apps_Usage file by using label_id for matching attribute and Vlookup Formular for merging

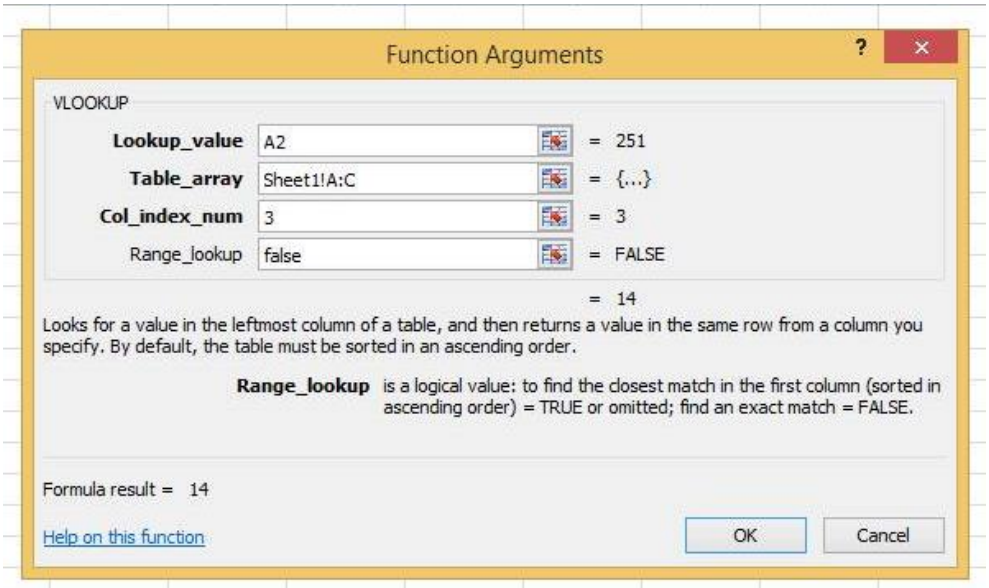

Figure 1: Vlookup Formular Tool

After we merged these 2 files together, we will have a new field in Mobile_Apps_Usage file that contained 15 categories that match to the application id.

PART II: Analyze Data

In this part, we use WEKA application to analyze data, by assign a new Mobile_Apps_Usage file into WEKA to generate a two-dimension of histogram:

X-asis represent a class (Category)

Y-asis represent a Frequency of data

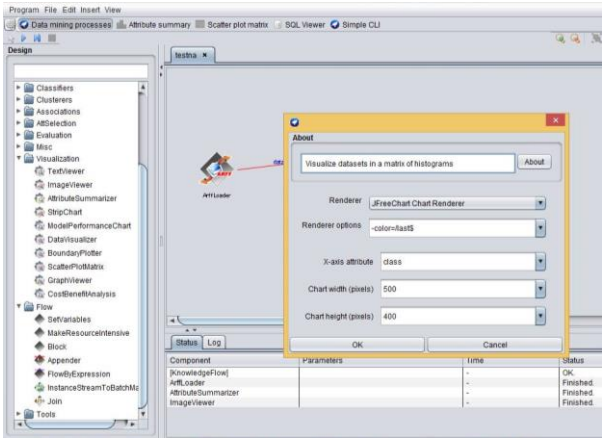

Figure 2.1: Import JFreeChart Chart Renderer Plugin.

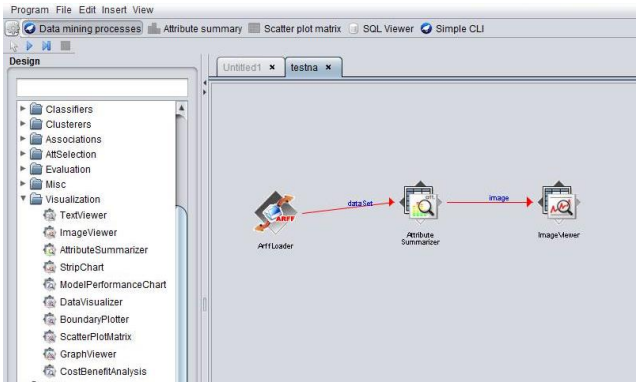

Figure 2.2: Create Histograms.

To generate a result of data set in form of histogram, we downloaded JFreeChart chart to render histogram graph, by choosing JFreeChart on renderer option as Figure 2.1.

As you can see in the Figure 2.2, that represent a complete path for generating a histogram which are:

1 Arffloader: preparing data set.

2 Attribute Summarize: converting data set into histogram (using JFreeChart).

3 Image Viewer: generating histogram into an image.

The result will be printed out as Figure 2.3, which present the frequency of 15 categories of application type, that was downloaded.

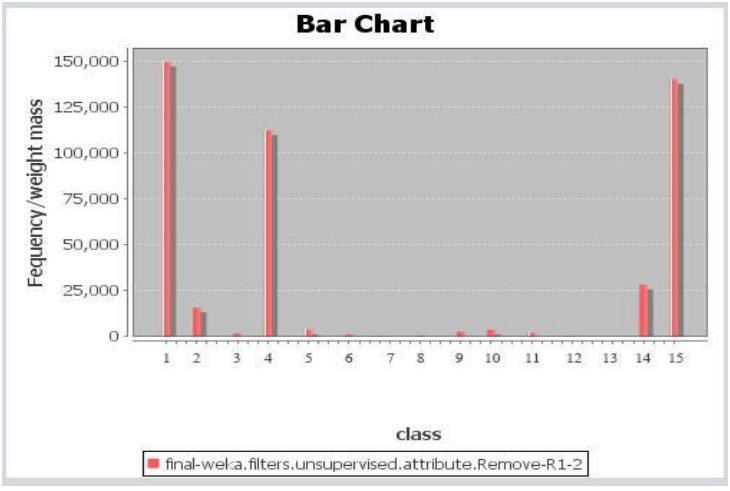

Figure 2.3: Histograms of Mobile_Apps_Usage.

In conclusion, from the histogram graph, we can conclude that there are a large amount of application downloaded in category 1 (Game) and 15 (Other).