

Week 5 Sunday – Taibah Valley – Task 9 – Extracting useful information from a Comma Separated Values (CSV file)

S. Y. Al-Kafrawi

6th of July 2020

1 - I have downloaded CSV datasheet which encompasses data about the following topics:

- App Name
- Category
- Rating
- Reviews
- Size
- Installs
- Type
- Price
- Content Rating
- Genre
- Last Updated
- Current Ver
- Android Ver

2 - I have interest about three aspects, these aspects are rating, popularity and pricing. Ergo, I cleaned the rows which include ambiguous data like “Nan” and I cleared the unnecessary columns and kept the necessary columns only, these columns are: **App, Rating, Reviews, Installs and Type.**

3 - I used **Excel** for filtering and clearing and a program called **Number** for charts. Moreover, I have established few questions regarding the aforementioned topics with respect to a certain criterion in order to enhance the reliability of the extracted information.

4 – Rating Aspect

Q1 – How many apps are rated above 4.5?

Q2 – How many apps are rated 3.6-4.5?

Q3 – How many apps are rated 2.6-3.5?

Q4 – How many apps are rated 1.6-2.5?

Q5 – How many apps are rated 0.6-1.5?

Q6 – How many apps are rated below 0.5?

Criterion: Number of reviews must be more than 100.

A1 = 1290 Apps

A2 = 5649 Apps

A3 = 452 Apps

A4 = 49 App

A5 = 0 App

A6 = 0 App

I have attached an example in Fig 1 illustrators the filtering procedure.

Rating

Sort

A Z Ascending

Z A Descending

By color: None

Filter

By color: None

Greater Than4.5

And

Or

Choose One

Search

(Select All)

1.6

1.7

1.8

1.9

2

2.1

2.2

Clear Filter

Rating

Sort

A Z Ascending

Z A Descending

By color: None

Filter

By color: None

Less Than or Equal To4.5

And

Or

Greater Than or Equal To3.5

Search

(Select All)

1.6

1.7

1.8

1.9

2

2.1

2.2

Clear Filter

Count: 1291

Figure 1 filter settings.

Rating	
above 4.5	1290
3.6-4.5	5649
2.6-3.5	452
1.6-2.5	49
0.6-1.5	0
0.5 or below	0
Total expenses	SAR 7,440

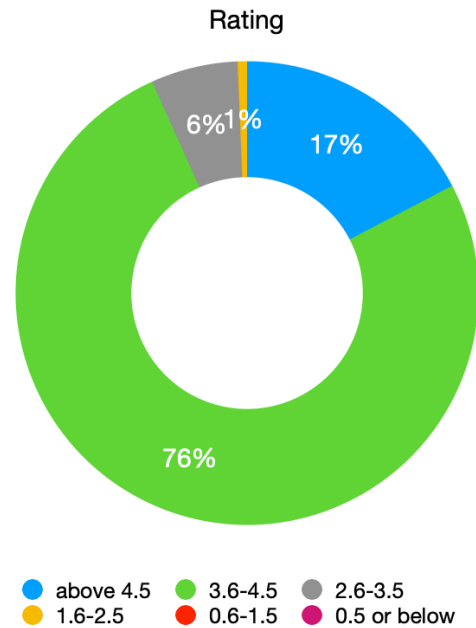


Fig.2 Visualization for rating data.

5 – Popularity aspect

Q1 – How many apps are downloaded more than 1 Million?

Q2 – How many apps are downloaded between 100K and 1 Million?

Q3 – How many apps are downloaded 100K or below?

A1= 4409

A2= 2155

A3= 3269F

Downloads	Units Number of apps
more than 1 Million	4,409
between 100K and 1 Million	2,155
100K or below	3,269

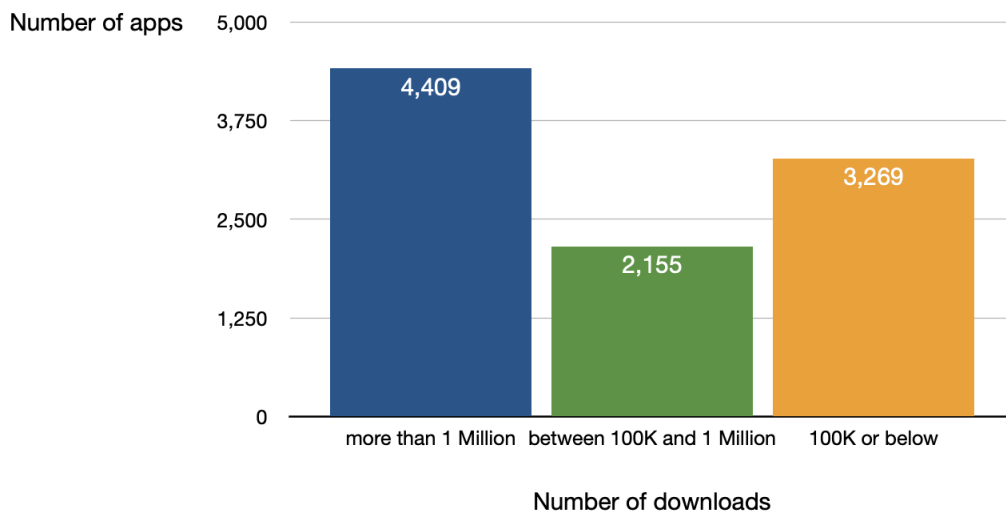


Fig.3 Visualization for popularity data.

6 – Pricing aspect

Q1 – How many apps are free?

Q2 – How many apps are paid?

A1= 8719

A2= 647

Cost	Number of apps
Free	8,719
Paid	647

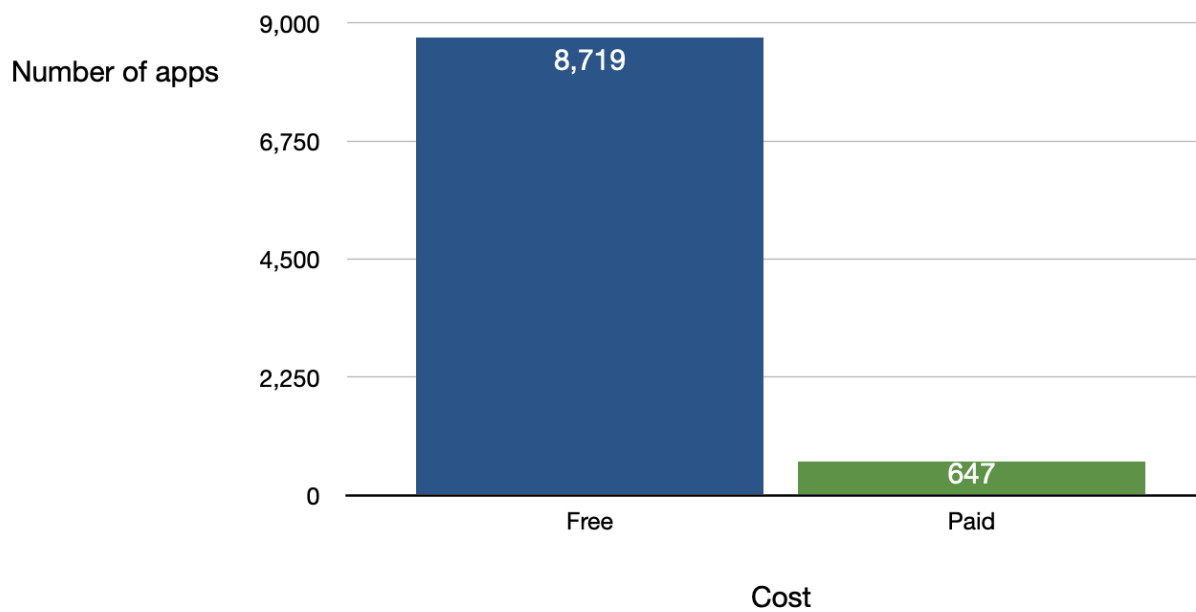


Fig.3 Visualization for cost data.

7 – Kindly, check the modified CSV file here:

https://drive.google.com/file/d/1jVlla6Uxb9UUtkp5qa_vmDjOSxZaZU0e/view?usp=sharing