DATA ANALYST PORTFOLIO

PREPARED BY

[Your Name]

Professional Background

I am an Electrical Engineering graduate with a passion for I.T and Telecommunications. I worked with Huawei Telecommunications as a Datacom Product Manager, this inspired me to study further and I obtained a Master's degree from the Blekinge Institute of Technology, Karlskrona, Sweden also in Electrical Engineering (Telecommunications). This helped me become a lot more confident and gave me the qualification needed to work in the Telecoms/I.T Industry.

In the course of my work I have been exposed to various applications, gained experience in Internet/Network Management, Technical Presentations, Application Support, IP Network Designs, and had the opportunity to work on projects using Unix and Perl scripting, programming in nesC & Python, Oracle SQL/Sybase, XML/MXML/HTML and various network Management Systems (like SNMPc, Transmode etc).

My journey into Data Analytics started during the COVID lockdown when I had a lot more time on my hands but realized it was difficult securing a remote job that would not only keep me busy but afford me the opportunity to earn some more and also gain a sense of career fulfilment and satisfaction. The Data Analyst course with Entry Level has taught me a whole lot; I now use google sheets as well as Excel, and their functions comfortably in a more robust manner, Pivot tables and other visualization Applications like Tableau in particular to manipulate and analyze data in such a way that I can clearly see trends and make conclusions and recommendations based on the information contained in the Data set.

I am motivated by the need to be a lot more versatile and flexible; the need to be relevant and part of the cutting-edge professional workforce that is driving technology by expanding my horizons, being able to work remotely and gain better earning potential. I am quite methodical, analytical and always ready to learn which has helped develop a keen sense of acquiring and employing the expertise needed to analyze problems and implement corresponding solutions.

I look forward to being an excellent Data Analyst, which involves continuous self-development, delving deeper into technologies involved and gaining to expert level the skills which entry level has so graciously exposed me to.

Table of Contents

Udemy Project Description

This project was given as a hypothetical case where I was to assume I was working as an employee of the Online Educational Technological giant, Udemy. The task was for me as a Data Analyst to find the best way to increase revenue and profitability for Udemy in the next quarter and my manager who is the head of Curriculum at Udemy needed this information as a report which she can send to the CEO within the next three weeks.

I was provided with data on the courses that were being taken so that I can present the statistics on course revenue, understand where the opportunities to increase earnings may lie and track the performance of the various courses. My manager already suggested that the courses in Web Development should be encouraged to charge more since she believes they are the most popular courses and as such will easily and quickly increase the revenue being generated.

I need to look closely at the data to know exactly what is happening, identify the business problem and suggest my solution backed up with evidence from my analysis.

EntryLevel

The Problem

The goal of this project is to increase revenue generation for Udemy. What is the best way to go about increasing the revenue and quarterly earnings in such a way that will not only be seamless and subtle but in such a manner that will not cause a negative backlash for Udemy. The problem basically is how to increase the quarterly earnings for Udemy in an easy and seamless way.

Obviously, the paid applications are the ones contributing directly to Udemy's revenue generation, so the problem is should the cost of the paid applications be increased? in particular the web development courses as already suggested by the head of curriculum since they seem to be the most popular; or there is another way to go about generating more revenue for example could some free courses be changed to become paid courses? This will increase the number of courses being paid for and increase earnings.

The raw data given mainly consists of the subjects and course titles that were being offered as at the time, it also details the number of subscribers, the cost, the level type (i.e. expert, intermediate, Beginner or combined "all levels"), ratings and reviews of each of these courses. There was therefore a need to take a critical look at the information I had to get a clear picture of what the current earnings are, the courses that are trending and other pertinent information that can be deduced; this will then be used to project, predict and make informed decisions that will guarantee the desired result which was to increase Udemy's quarterly earnings as required.

Design

The data came in four separate blocks where each block was a representation of each of the subjects Udemy was offering. These individual blocks were combined and consolidated into one using Google Sheets to capture all the information as a single entity so it can be easy to work with and analyze. To properly track the performance of the courses a detailed scrutiny of the data was required, to gain clarity and the proper perspective needed to make sound decisions and aid wholesome management I had to also confirm the data was reliable. Ensuring good, reliable data, and improving the quality of the data set was done by cleaning the data.

I had to remove duplicate information as well as blank cells and incomplete information; also, improperly formatted values were corrected so that the data was consistent and easy to work with.

I also used Tableau public to perform further in-depth analysis and get a pictorial overview and proper representation/visualization to support theories, deductions and inferences made.

Findings

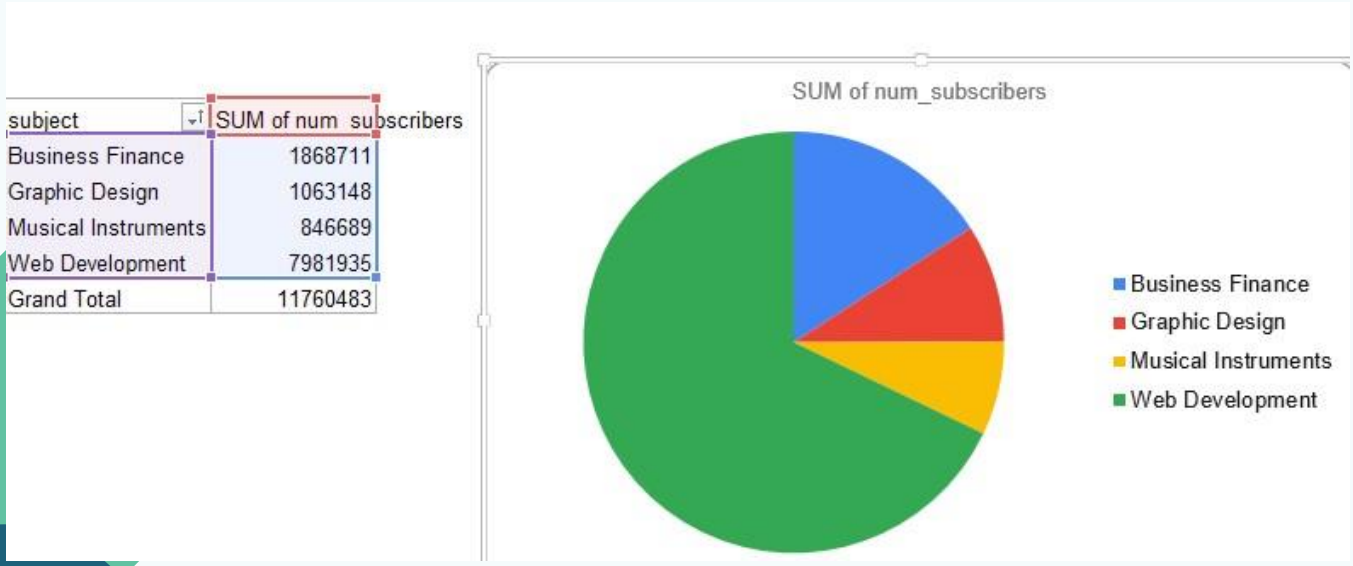
Based on the data given the earnings made amounted to 885,160,005 that is, the sum of all paid subscriptions for the courses installed (possibly in dollars or pounds not sure of the currency) but this is over a time period spanning the years between 2011 to 2017, so it's difficult to deduce accurately what the quarterly earnings have been in the past but we can precisely see what the earnings have been for each business year.

Table 1. Yearly Paid Subscriptions

Sum of Cost Row Labels	Business Finance	Graphic Design	Musical Instruments	Web Development	Grand Total
2011				11643420	11643420
2012	190740	1329850	766405	9486475	11773470
2013	7298950	3085300	7479930	51541640	69405820
2014	35870820	8364490	20899910	42027165	107162385
2015	38702015	23273795	12363235	240171350	314510395
2016	30727750	23538210	7458615	215969765	277694340
2017	10945040	17391525	4390960	60242650	92970175
Grand Total	**123735315**	**76983170**	**53359055**	**631082465**	**885160005**

We can see that the earnings for Udemy where highest in the years 2015, in 2017 earnings fell drastically which could have been caused by any number of factors. Also we can see that in the year 2011 only web development courses made any earnings.

Figure 1. Number of Subscribers per Subject



Without a doubt at first glance (as shown in Figure 1 above) it is evident that the Web Development courses have the highest number of subscribers of all the courses and it will be quite easy to say and conclude that more should be charged for Web Development courses to increase revenue. However, further analysis shows that while this holds true there is more to the data than that which is apparent at first glance.

A critical look at the data first of all showed that the top 20 courses were not necessarily free courses and that they were majorly web design courses. In fact of the top 20 courses, 17 are Web Development courses as shown below in Figure 2. Also these top 20 courses were offered majorly between the year 2012 and 2014 with a couple of outliers and 12 of these 20 courses were offered for free.

Figure 2. Top 20 Courses

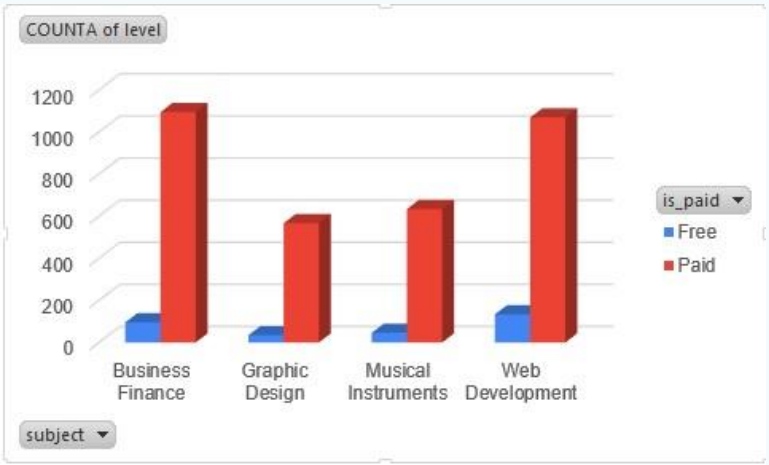
1	course_title	price	is_paid	num_subscribers	num_reviews	num_lectures	level	rating	content_duration	subject	date
2	Learn HTML5 F	0	Free	268923	8629	45	Beginner Level	0.82	10.5	Web Development	2013-02-14
3	Coding for Entr	0	Free	161029	279	27	Expert Level	0.69	3.5	Web Development	2013-06-09
4	Build Your Firs	0	Free	120291	5924	30	All Levels	0.78	3	Web Development	2014-04-08
5	Free Beginner	0	Free	101154	1042	95	All Levels	0.88	4.5	Musical Instruments	2012-06-15
6	Web Design fo	0	Free	98867	6512	20	All Levels	0.82	3	Web Development	2015-04-13
7	Practical PHP:	0	Free	83737	4598	45	Intermediate Lev	0.85	6.5	Web Development	2014-07-19
8	Beginner Photo	0	Free	73110	1716	22	All Levels	0.94	2	Web Development	2012-07-27
9	Web Developm	0	Free	72932	2575	21	All Levels	0.39	1	Web Development	2013-09-25
10	HTML and CSS	0	Free	70773	5660	50	All Levels	0.57	6	Web Development	2015-03-19
11	Bitcoin or How	0	Free	65576	936	24	All Levels	0.56	8	Business Finance	2013-04-20
12	Quickstart Ang	0	Free	64128	4047	17	Beginner Level	0.96	1.5	Web Development	2014-11-22
13	Learn Respons	0	Free	59639	2692	24	All Levels	0.55	4.5	Web Development	2013-12-09
14	Learn Javascrip	30	Paid	84897	2685	10	All Levels	0.79	2	Web Development	2013-10-10
15	Become a Wel	120	Paid	69186	2408	197	All Levels	0.61	27.5	Web Development	2011-11-19
16	JavaScript: Und	175	Paid	79612	16976	85	All Levels	0.69	11.5	Web Development	2015-03-12
17	Learn and Und	175	Paid	59361	11580	55	Beginner Level	0.87	7	Web Development	2014-09-24
18	Angular 4 (form	190	Paid	73783	19649	329	Beginner Level	0.9	22	Web Development	2016-02-11
19	The Web Deve	200	Paid	121584	27445	342	Beginner Level	0.89	43	Web Development	2015-11-02
20	The Complete '	200	Paid	114512	22412	304	Beginner Level	0.55	30.5	Web Development	2016-03-08
21	Pianoforall - Inc	200	Paid	75499	7676	362	Beginner Level	0.96	30	Musical Instruments	2014-08-07

A more comprehensive look at the entire data across all the subjects reveal that the number of subscribers for the free courses were actually much less in total to the number of subscribers of the paid courses (as seen in Table 2 and the bar chart in Figure 3), with Web Development having the highest number of subscribers in total for both its paid and free courses. However, it can also be seen that Business Finance as a subject actually has the highest numbers of paid subscribers.

Table 2. Free Versus Paid Courses Statistics

COUNTA of level	is_paid		
subject	Free	Paid	Grand Total
Business Finance	96	1095	1191
Graphic Design	35	567	602
Musical Instruments	46	634	680
Web Development	134	1069	1203
Grand Total	311	3365	3676

Figure 3. Free versus Paid Bar Chart



Analysis

The data set is showing us a lot more information and giving a lot more details than what was initially obvious. While Web development is the course with the highest numbers of subscribers, Business Finance actually has the highest number of paid courses subscribed for. As shown in Figure 4 below we can also see that the level of the courses seem to also have impacted the subscription rate. The “All Levels” was the highest subscribed level across all subjects for both free and paid courses, closely followed by the “Beginner Level” courses. We can also see that generally, there is a higher number of subscribers for the paid courses compared to the free courses.

Figure 4. Number of Subscribers per level

COUNTA of level		is_paid		
subject	level	Free	Paid	Grand Total
Business Finance	All Levels	52	**581**	**633**
	Beginner Level	26	**373**	**399**
	Expert Level	4	21	25
	Intermediate Level	14	120	134
Business Finance Total		96	**1095**	1191
Graphic Design	All Levels	26	309	335
	Beginner Level	4	180	184
	Expert Level		7	7
	Intermediate Level	5	71	76
Graphic Design Total		35	567	602
Musical Instruments	All Levels	20	304	324
	Beginner Level	19	247	266
	Expert Level	1	11	12
	Intermediate Level	6	72	78
Musical Instruments Total		46	634	680
Web Development	All Levels	71	**562**	**633**
	Beginner Level	51	**371**	**422**
	Expert Level	2	12	14
	Intermediate Level	10	124	134
Web Development Total		134	**1069**	1203
Grand Total		311	3365	3676

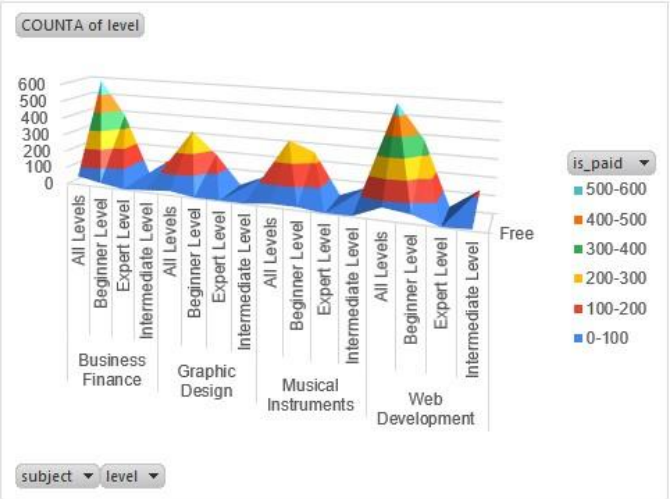
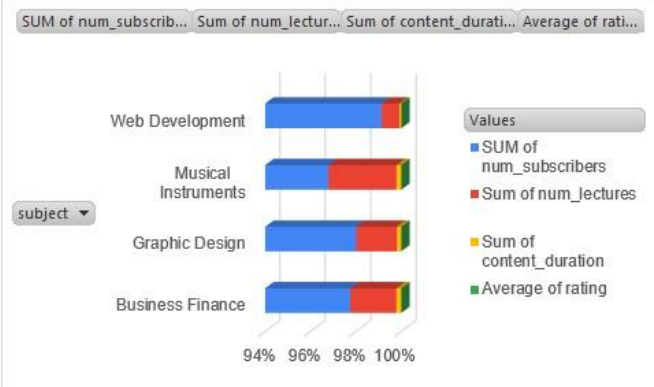


Figure 5. Content Duration and Ratings

subject	SUM of num_subscrib	Sum of content_dural	Average of rating
Business Finance	1868711	4237.216667	0.690352645
Graphic Design	1063148	2158.45	0.73038206
Musical Instruments	846689	1940.983333	0.308911765
Web Development	7981935	6720.75	0.643050707
Grand Total	11760483	15057.4	0.610867791



We can also see from Figure 5 that the subject with the lowest number of subscribers (Musical Instruments) has the lowest average rating and also its sum of content duration is the lowest too. Also, even though Business Finance had the highest Average rating, this did not translate to it having the highest number of subscribers either, while Web Development had the highest sum of number of subscribers as well as the highest sum of content duration.

Analyzing further using the 5 Whys model also reveals an interesting point.

Why is the number of subscribers for the paid courses more than that of the free courses? You would expect that lot more subscribers will register for free courses as opposed to paid courses.

Why does the highest average rating not translate to having the highest number of subscribers?

Why does the subject with the highest content duration not have the highest number of subscribers?

Why is Web Development the top course being offered even though its average cost per course is higher than that of the other subjects?

Further analysis show that even though at first glance the content duration of the Web Development courses are longer than that of other courses, it is actually shorter on the average compared to the other subjects which might mean the web development courses are probably better structured and well suited based on time period/ schedule and quite practicable so subscribers don't mind paying to get on these courses as there is value for money.

Conclusion

The data shows that the paid ‘All levels’ courses have the highest number of subscribers across board for all the subjects. Analysis also further shows that the Beginner level paid courses for the Business Finance and the Web Development subjects also have a high number of subscribers.

The data given doesn’t contain information on the status of the subscribers like if they are mainly students or if they are professionals looking to upskill or gain an additional certificate or just out to learn a new skill and so can afford to spend more, so we can’t use any of these parameters to make predictions on what a significant increase in price on one subject could mean for subscribers.

My recommendation therefore is that the increase in cost should be distributed across the subjects. This means that the cost for the ‘All level’ package should be increased for all subjects across board, especially for Business Finance and Web Development courses which have the highest paid subscribers, also an increase in the price for the Beginner levels of the Business Finance and Web development courses will also yield a significant increase in the quarterly earnings for the educational technological company Udemy.

CAPSTONE PROJECT

Capstone Project Description

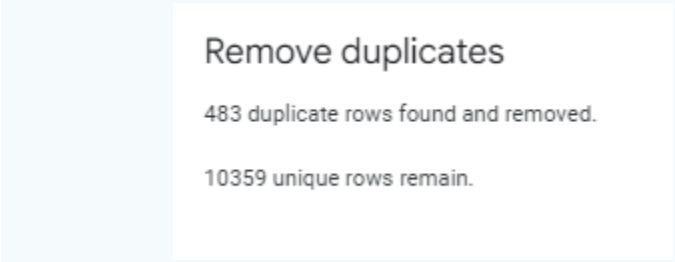
This project deals with the data collected on all the applications available on google store.

This project was chosen because I was very interested in understanding the trends of mobile applications on google store and getting a better understanding of what makes applications successful and popular amongst users. The data set was downloaded from Kaggle and was last updated in 2019 and therefore represents data as at 2019 and before. I wanted to understand what drives the success of applications amongst users and the best way to bring profitability from app development to the coder/programmer which requires that I understand the current user trends. I also realized that there were a lot of free applications compared to paid ones which was quite surprising and I wanted to know if there was a possible reason for this. There was therefore a need to take a critical look at the information to get a clear picture of what the costs are, the applications that are trending and other pertinent information that can be deduced; this will then be used to project, predict and make informed decisions in response.

EntryLevel

Data Design

The data came as a csv (comma separated values) file which was worked on (offline) using Excel and then later transferred to google sheets to do some fine tuning and comparisons, though I used Excel a lot more on this project. Tableau was also used to do further comparisons and create dashboards for additional data analysis, ease of presentation and to support inferences, deductions and conclusions made. To properly track the performance of the applications a detailed scrutiny of the data was done; in order to gain clarity and the proper perspective required to make sound judgements and aid wholesome management I made sure the data was reliable. Ensuring good, reliable data, and improving the quality of the data set was done by cleaning the data; for instance, duplicate information was removed (about 483 duplicate values were found) as well as blank cells and incomplete information (1 row was incomplete and had lots of blank cells), also improperly formatted values were corrected so that the data was consistent across board; the ratings column had over 1000 rows of NaN values, rather than discard these rows, I replaced them with the number 0 to give a better context to my analysis. I also noticed that there were a few applications that still had duplicates left in the data set even after removing duplicates i.e. the same application in the same Category and Genre but with different number of reviews yet all other features were the same, I believe this was an error in the way the data was probably collected. I removed these duplicates as well.



Findings

Based on the data set there seems to be a lot of free applications which in turn had a lot of downloads, the total earnings made by the applications that are paid amounted to **$293,149,447.79**. The data showed that the Games category had the highest number of downloads, followed by Communication and then Family category comes in at a close third as seen in Table 3 below.

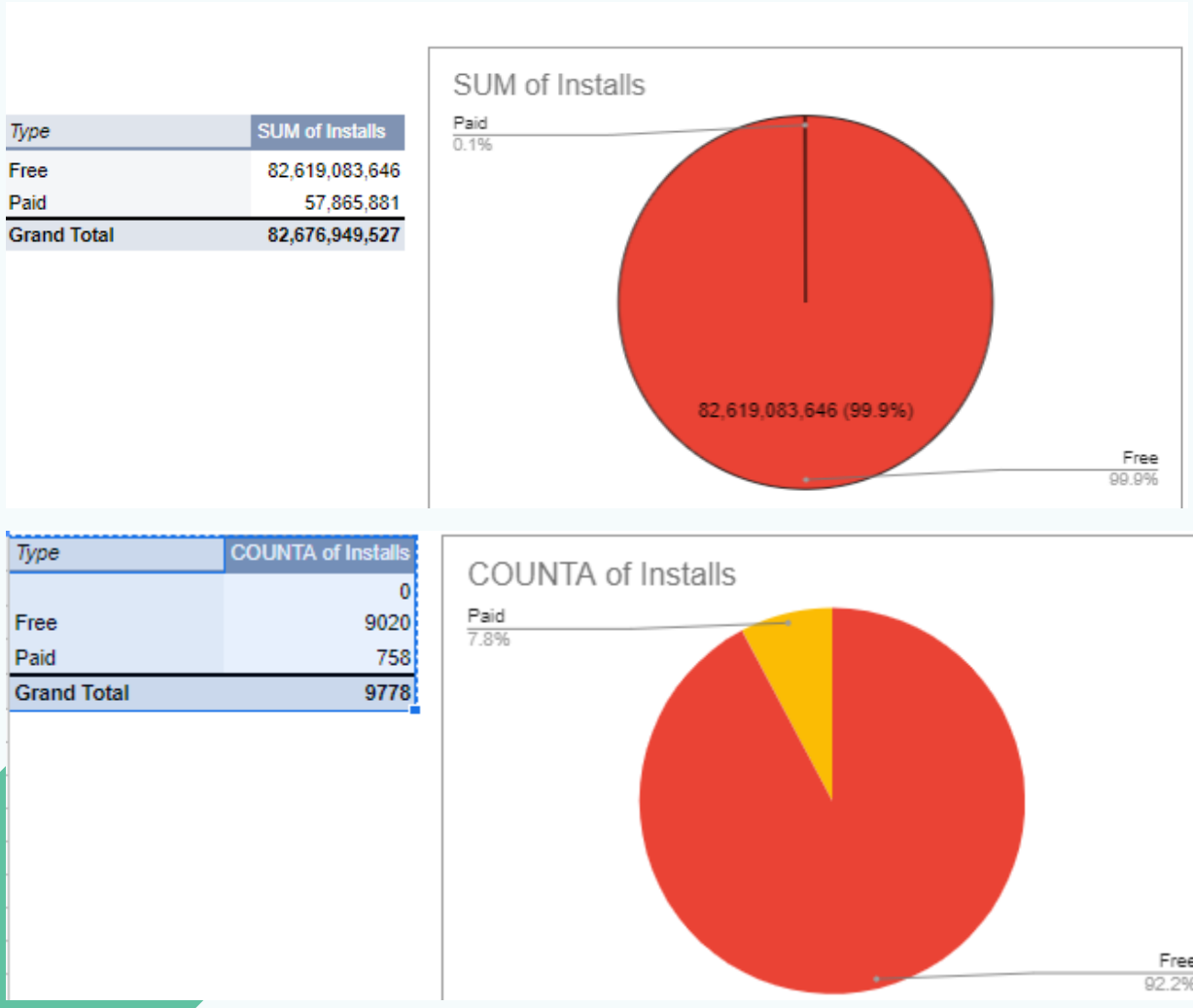
Table 3. App Downloads & Earnings per Category

Category	SUM of Installs	SUM of Earning $
ART_AND_DESIGN	114,338,100	31,840.00
AUTO_AND_VEHICLES	53,130,211	100,148.50
BEAUTY	27,197,050	0.00
BOOKS_AND_REFERENCE	1,665,969,576	90,226.74
BUSINESS	747,164,865	1,050,543.25
COMICS	44,981,150	0.00
COMMUNICATION	11,039,276,251	4,247,364.50
DATING	140,926,107	48,461.50
EDUCATION	352,952,000	2,403,980.00
ENTERTAINMENT	2,113,660,000	798,000.00
EVENTS	15,973,161	109.99
FAMILY	8,995,642,505	115,680,835.36
FINANCE	460,348,734	25,726,678.98
FOOD_AND_DRINK	212,898,751	284,400.00
GAME	13,993,924,415	40,688,636.85
HEALTH_AND_FITNESS	1,144,022,512	1,426,068.90
HOUSE_AND_HOME	97,212,461	0.00
LIBRARIES_AND_DEMO	52,995,910	99.00
LIFESTYLE	503,823,539	57,583,939.40
MAPS_AND_NAVIGATION	608,281,890	1,240,789.00
MEDICAL	40,304,177	5,957,535.67
NEWS_AND_MAGAZINES	2,879,217,760	6,445.00
PARENTING	31,521,110	249,959.00
PERSONALIZATION	1,532,494,782	6,796,947.56
PHOTOGRAPHY	5,769,147,655	8,547,767.70
PRODUCTIVITY	5,808,091,369	4,313,374.95

SHOPPING	1,671,348,785	30,149.00
SOCIAL	5,497,867,902	5,940.00
SPORTS	1,301,474,498	4,706,211.95
TOOLS	8,022,771,915	5,464,820.59
TRAVEL_AND_LOCAL	2,894,887,146	1,151,504.40
VIDEO_PLAYERS	4,432,002,720	335,290.00
WEATHER	411,100,520	4,181,380.00
Grand Total	**82,676,949,527**	**293,149,447.79**

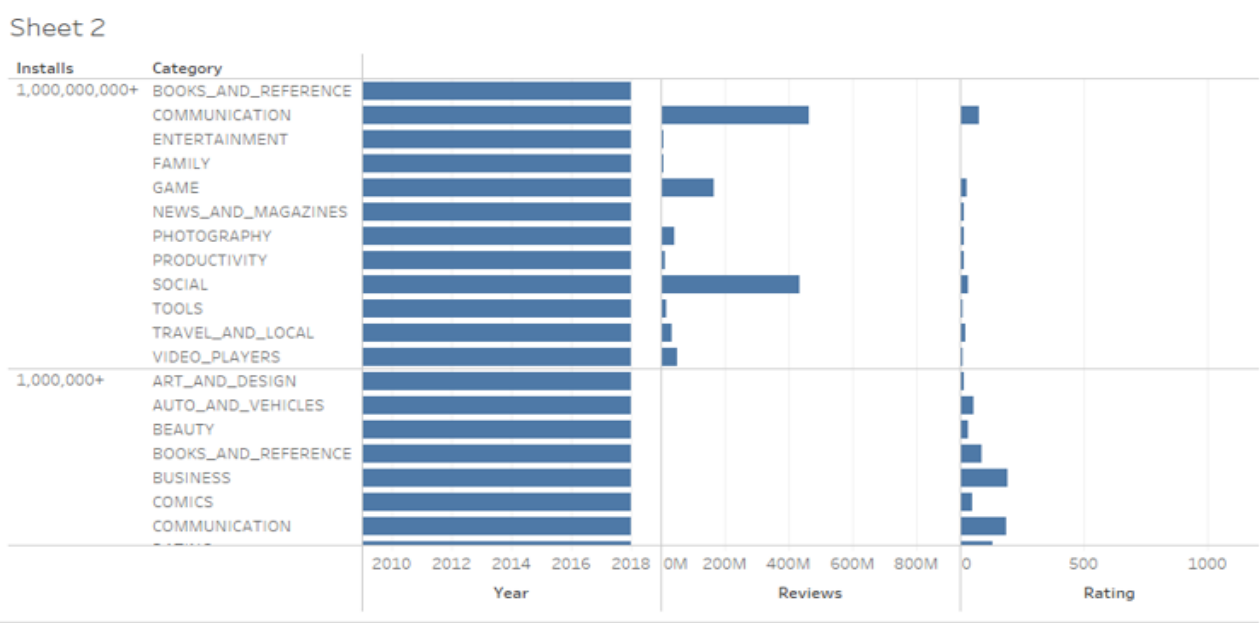
Figure 6 shows again that the number of downloads of the paid applications were less than 1% compared to the number of downloads of the free applications and there were over 9000 free applications compared to just 758 paid applications.

Figure 6. Free Versus Paid Apps based on number of Installs



The snapshot of the chart in Figure 7 below shows the Installs, Reviews and Ratings per category, the full visualization using Tableau (public) is available on my capstone dashboard (the link is given in the Appendix)

Figure 7. Installs, Reviews and Ratings per Category



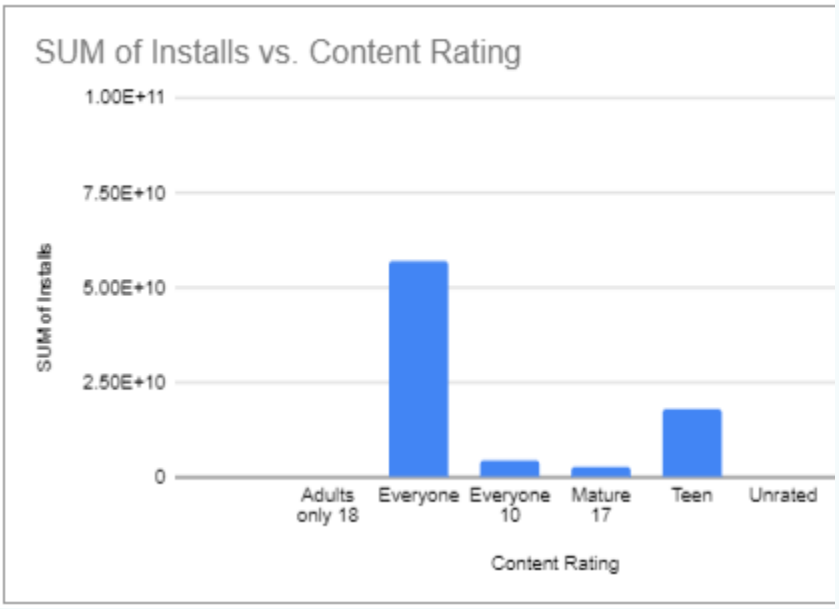
Without a doubt (as seen in Figure 7 above) it is evident that the free applications have the highest number of installs compared to the paid applications for all categories. It is also interesting to note that at first glance, the number of reviews seem to have no direct relationship to the number of installs. I would have assumed that the categories with the highest reviews will also have the highest downloads and installs but this is not exactly the case. Further analysis also shows that rating has no direct relationship to the number of application downloads, the applications that have the highest ratings do not have the highest downloads neither does it seem the size of the application nor the Android version has any relationship or correlation to the number of installs. However, it is clear that specific Categories and genres seem to have a high popularity and hence a lot more installs than others.

Some preliminary visualizations done in tableau (check CapstoneP | Tableau Public) show that the communication and the social applications had the highest number of reviews as well as high rates of installs compared to the other genres. Communication, Games, and Family had the highest number of free apps installed over and above all other genres.

We can see as well that the content rating was done by 5 main categories of people; “Everyone”, “Everyone 10”, “Teen”, “Adults only 18”, and “Mature 17”. We can see that “Everyone” and “Teen” were the highest contributors to the number of downloads as shown in Figure 8 below.

Figure 8. Content Rating per Sum of Downloads

Content Rating	SUM of Installs
Adults only 18	2,000,000
Everyone	56,951,364,961
Everyone 10	4,632,271,795
Mature 17	2,947,986,878
Teen	18,143,275,393
Unrated	50,500
Grand Total	**82,676,949,527**



We can see (in Table 4 below) that this also translates across the number of reviews as well, though there is a slight change when we consider it against the average rating as we can see that the "Adults only 18" gave the highest average ratings followed by "Everyone 10" and then "Mature 17". This means that while the rating of an App should be considered important, the data shows that the rating did not really contribute significantly to the number of downloads in this case; while, there is actually a correlation between the number of reviews and the number of installs/downloads when broken down further to the level of content rating.

Table 4a. Reviews & Ratings by Content Rating

Content Rating	SUM of Installs	SUM of Reviews	AVERAGE of Rating
Adults only 18	2,000,000	81,348.00	4.30
Everyone	56,951,364,961	1,440,995,095.00	3.50
Everyone 10	4,632,271,795	294,505,512.00	4.03
Mature 17	2,947,986,878	98,871,045.00	3.75
Teen	18,143,275,393	548,108,191.00	3.73
Unrated	50,500	1,187.00	2.05
Grand Total	**82,676,949,527**	**2,382,562,378.00**	**3.55**

Table 4b. Reviews & Ratings by Category

Category	SUM of Installs	SUM of Reviews	AVERAGE of Rating
ART_AND_DESIGN	114,338,100	1,419,203.00	4.15
AUTO_AND_VEHICLES	53,130,211	1,163,666.00	3.60
BEAUTY	27,197,050	396,240.00	3.39
BOOKS_AND_REFERENCE	1,665,969,576	16,721,314.00	3.31
BUSINESS	747,164,865	11,170,045.00	2.57
COMICS	44,981,150	2,342,071.00	4.03
COMMUNICATION	11,039,276,251	285,828,897.00	3.35
DATING	140,926,107	3,623,589.00	3.11
EDUCATION	352,952,000	13,364,148.00	4.33
ENTERTAINMENT	2,113,660,000	34,762,650.00	4.14
EVENTS	15,973,161	161,018.00	3.12
FAMILY	8,995,642,505	352,367,669.00	3.70
FINANCE	460,348,734	12,943,741.00	3.60
FOOD_AND_DRINK	212,898,751	6,335,995.00	3.51
GAME	13,993,924,415	624,217,122.00	4.04
HEALTH_AND_FITNESS	1,144,022,512	21,361,360.00	3.59
HOUSE_AND_HOME	97,212,461	1,929,847.00	3.48
LIBRARIES_AND_DEMO	52,995,910	906,831.00	3.18
LIFESTYLE	503,823,539	11,832,671.00	3.34
MAPS_AND_NAVIGATION	608,281,890	22,711,099.00	3.64
MEDICAL	40,304,177	1,308,863.00	3.07
NEWS_AND_MAGAZINES	2,879,217,760	34,963,027.00	3.32
PARENTING	31,521,110	958,331.00	3.58
PERSONALIZATION	1,532,494,782	53,543,081.00	3.43
PHOTOGRAPHY	5,769,147,655	124,115,860.00	3.90
PRODUCTIVITY	5,808,091,369	55,827,688.00	3.37
SHOPPING	1,671,348,785	57,551,886.00	3.78
SOCIAL	5,497,867,902	229,191,885.00	3.61
SPORTS	1,301,474,498	59,470,975.00	3.38
TOOLS	8,022,771,915	224,789,504.00	3.51
TRAVEL_AND_LOCAL	2,894,887,146	26,819,760.00	3.47
VIDEO_PLAYERS	4,432,002,720	74,114,771.00	3.68
WEATHER	411,100,520	14,347,571.00	3.87
Grand Total	**82,676,949,527**	**2,382,562,378.00**	**3.55**

Table 5 makes it clear that for the paid applications, Family, Game and Personalization in that order had the highest number of Installs while for the free applications it was Game, Communication and Family categories that topped the list of installs and this eventually translated into the highest number of installs overall.

Table 5. Highest Category of Installs by Type

SUM of Installs	*Type*		
Category	Free	Paid	Grand Total
ART_AND_DESIGN	114,322,100	16,000	114,338,100
AUTO_AND_VEHICLES	53,080,061	50,150	53,130,211
BEAUTY	27,197,050		27,197,050
BOOKS_AND_REFERENCE	1,665,946,260	23,316	1,665,969,576
BUSINESS	746,952,090	212,775	747,164,865
COMICS	44,981,150		44,981,150
COMMUNICATION	11,037,916,201	1,360,050	11,039,276,251
DATING	140,914,757	11,350	140,926,107
EDUCATION	352,350,000	602,000	352,952,000
ENTERTAINMENT	2,113,460,000	200,000	2,113,660,000
EVENTS	15,973,160	1	15,973,161
FAMILY	8,974,491,691	21,150,814	8,995,642,505
FINANCE	460,163,132	185,602	460,348,734
FOOD_AND_DRINK	212,838,751	60,000	212,898,751
GAME	13,972,924,450	20,999,965	13,993,924,415
HEALTH_AND_FITNESS	1,143,548,402	474,110	1,144,022,512
HOUSE_AND_HOME	97,212,461		97,212,461
LIBRARIES_AND_DEMO	52,995,810	100	52,995,910
LIFESTYLE	502,644,429	1,179,110	503,823,539
MAPS_AND_NAVIGATION	608,160,790	121,100	608,281,890
MEDICAL	39,743,344	560,833	40,304,177
NEWS_AND_MAGAZINES	2,879,212,260	5,500	2,879,217,760
PARENTING	31,471,010	50,100	31,521,110
PERSONALIZATION	1,529,235,988	3,258,794	1,532,494,782
PHOTOGRAPHY	5,767,268,915	1,878,740	5,769,147,655
PRODUCTIVITY	5,806,679,314	1,412,055	5,808,091,369
SHOPPING	1,671,338,685	10,100	1,671,348,785
SOCIAL	5,497,861,902	6,000	5,497,867,902
SPORTS	1,300,230,683	1,243,815	1,301,474,498
TOOLS	8,021,044,474	1,727,441	8,022,771,915

TRAVEL_AND_LOCAL	2,894,704,086	183,060	2,894,887,146
VIDEO_PLAYERS	4,431,931,720	71,000	4,432,002,720
WEATHER	410,288,520	812,000	411,100,520
Grand Total	**82,619,083,646**	**57,865,881**	**82,676,949,527**

Data Analysis

The data was analyzed employing the use of the 5 Whys (5Y) root cause analysis technique. This technique seeks to narrow down problems to reveal underlying factors or causes so that solutions called counter measures can be applied correctly. This method drills down to the root cause of a problem by asking the question why.

i. Why does it seem like the number of reviews have no direct impact on the number of installs as the applications with the highest reviews do not necessarily have the highest downloads? Though we see that categories with high reviews seem to have a relatively larger number of Installs than those with lower number of reviews. For instance (as seen in Table 4b. Reviews & Ratings by Category) we see that the Family Category that had a higher number of reviews still had a much lower number of installs compared to the Communication Category.

We know that actually, reviews could be positive or negative and we don't have this information given in this data set, so a high number of reviews do not necessarily translate to a good/positive reviews or feedback. This may be why we do not see it having a significant impact on the number of downloads.

ii. Why does the type of application seem to matter in the choice of installs?

The type of application (whether it is free or paid) seems to have had a direct impact on the number of downloads as the free applications across all genres had a higher number of installs compared to the paid applications. In the case of the Family and Communication category that I looked at above we see that the number of free applications installed for Communications were much higher than Family. Also, we see (in Figure 9) that the year 2018 (that is the last time the application was updated) had the highest rate of installs generally. Actually, the number of installs steadily increases as the last year of update also increased from 2010 until 2018, there were no installs for paid applications at all for updated applications in the year 2010. Also, the Beauty and Comics category had no paid applications and the Events Category had only one paid for application which in turn had only one install.

Figure 9. Installs based on Last Year App was updated

SUM of Installs	*Last Update*									
Type	2010	2011	2012	2013	2014	2015	2016	2017	2018	Grand Total
Free	100,000.00	17401000.00	13,982,000	102,434,650	266,880,960	404,261,065	1,286,956,641	3,007,770,581	77,519,296,749	82,619,083,646
Paid		6100.00	10,210	651,715	2,904,266	2,179,410	6,409,152	3,341,158	42,363,870	57,865,881
Grand Total	0	100,000.00	17407100.00	13,992,210	103,086,365	269,785,226	406,440,475	1,293,365,793	3,011,111,739	77,561,660,619	82,676,949,527

Family, Lifestyle and Game were the highest category (in that order) in earnings for the paid applications but interestingly enough Game, Communication and Family had the highest number of installs amongst the free applications (see Table 5 above). This shows that amongst the paid applications these categories were considered important enough to install despite the cost attached to it. Though Games, Communication and Family had the highest sum of installs across both types (free/paid) and categories.

Figure 10a. Installs per unit cost

Category	Rating	Reviews	Size	Installs	Downloads	Type	Price	Earning $	Content Rat	Genres
GAME	4.6	4108	13M	10,000	10,000+	Paid	$17.99	179900	Everyone 10	Arcade
GAME	3.8	252	11M	10,000	10,000+	Paid	$16.99	169900	Teen	Action
GAME	4.6	17	4.7M	100	100+	Paid	$14.00	1400	Teen	Casino
GAME	4.5	38207	15M	100,000	100,000+	Paid	$9.99	999000	Teen	Role Playing
GAME	4.4	1644	15M	10,000	10,000+	Paid	$7.99	79900	Everyone	Board
GAME	4.4	348962	26M	1,000,000	1,000,000+	Paid	$6.99	6990000	Mature 17	Action
GAME	4.2	13004	46M	100,000	100,000+	Paid	$6.99	699000	Teen	Action
GAME	4.5	514	5.9M	5,000	5,000+	Paid	$6.99	34950	Teen	Adventure
GAME	4.6	144	6.3M	1,000	1,000+	Paid	$6.99	6990	Everyone	Adventure
GAME	4.6	10795	69M	100,000	100,000+	Paid	$5.99	599000	Everyone	Simulation
GAME	4.6	87766	12M	1,000,000	1,000,000+	Paid	$4.99	4990000	Everyone	Action

Category	Rating	Reviews	Size	Installs	Downloads	Type	Price	Earning $	Content Rat	Genres
FAMILY	4	856	8.7M	10,000	10,000+	Paid	$399.99	3999900	Everyone	Entertainment
FAMILY	4.4	201	2.7M	5,000	5,000+	Paid	$399.99	1999950	Everyone	Entertainment
FAMILY	4.3	6	1.5M	100	100+	Paid	$399.99	39999	Everyone	Entertainment
FAMILY	3.6	217	4.9M	10,000	10,000+	Paid	$389.99	3899900	Everyone	Entertainment
FAMILY	0	2	8.5M	10	10+	Paid	$46.99	469.9	Everyone	Entertainment
FAMILY	0	0	14M	5	5+	Paid	$39.99	199.95	Everyone	Education
FAMILY	5	1	96M	10	10+	Paid	$29.99	299.9	Mature 17	Education
FAMILY	4.5	7812	27M	50,000	50,000+	Paid	$19.99	999500	Everyone 10	Role Playing
FAMILY	5	1	40M	10	10+	Paid	$19.99	199.9	Everyone	Education
FAMILY	4.7	54	50M	1,000	1,000+	Paid	$19.40	19400	Everyone	Education
FAMILY	4.4	3358	17M	100,000	100,000+	Paid	$14.99	1499000	Everyone	Role Playing
FAMILY	4.7	1647	16M	10,000	10,000+	Paid	$14.99	149900	Everyone	Role Playing
FAMILY	4.7	1667	17M	10,000	10,000+	Paid	$14.99	149900	Teen	Role Playing
FAMILY	4.8	4	23M	10	10+	Paid	$14.99	149.9	Teen	Casual
FAMILY	4.3	8450	1.1M	100,000	100,000+	Paid	$13.99	1399000	Everyone	Role Playing

Looking at the paid applications, the highest installs was the Game category even though Family had the highest Earnings. Drilling down further (as seen in Figure 10a above), shows that the application with the highest cost in the Game Category cost $17.99 each while the application with the highest cost in the Family Category was going for $399.99. Also, we can see that the application with the lowest cost (see Figure 10b below) in the Games Category was $0.99 while for the Family Category it was $6.99.

The application going for $17.99 in the Game Category had over ten thousand (10,000+) installs -earning $179,900 and the application with the highest installs of ten million plus (10,000,000+) in the Games category cost $0.99 earning $9,900,000 which is more than 10 times over and above what was made by the former. Furthermore, in the Family Category that had the highest earnings we see that the application with the highest cost of $399.99 also had 10,000+ installs earning $3,999,900 while the application with the highest downloads of over ten million (10,000,000+) cost $6.99 and earned $69,900,000 which is also more than 10 times what was earned by the costliest application.

Comparing both cases shows that the applications that cost the highest did not ultimately make the most earnings since they did not have a very high install rate to make the figures.

Figure 10b. Installs per unit cost

Category	Rating	Reviews	Size	Installs	Downloads	Type	Price	Earning $
FAMILY	4.5	2376564	Varies with	10,000,000	10,000,000+	Paid	$6.99	69900000
FAMILY	4.6	190086	94M	1,000,000	1,000,000+	Paid	$2.99	2990000
FAMILY	4.3	129603	23M	1,000,000	1,000,000+	Paid	$2.99	2990000
FAMILY	4.7	188740	69M	1,000,000	1,000,000+	Paid	$1.99	1990000
FAMILY	4.6	61264	43M	1,000,000	1,000,000+	Paid	$0.99	990000
FAMILY	4.3	21804	44M	1,000,000	1,000,000+	Paid	$0.99	990000
FAMILY	4.3	56444	43M	500,000	500,000+	Paid	$4.99	2495000
FAMILY	4.7	31100	24M	500,000	500,000+	Paid	$3.99	1995000
FAMILY	4.2	32812	14M	500,000	500,000+	Paid	$2.99	1495000
FAMILY	4.6	73919	40M	500,000	500,000+	Paid	$2.99	1495000
FAMILY	4.6	84114	23M	500,000	500,000+	Paid	$0.99	495000
FAMILY	4.4	3358	17M	100,000	100,000+	Paid	$14.99	1499000

Category	Rating	Reviews	Size	Installs	Downloads	Type	Price	Earning $
GAME	4.6	408292	29M	10,000,000	10,000,000+	Paid	$0.99	9900000
GAME	4.4	348962	26M	1,000,000	1,000,000+	Paid	$6.99	6990000
GAME	4.6	87766	12M	1,000,000	1,000,000+	Paid	$4.99	4990000
GAME	4.6	100805	50M	1,000,000	1,000,000+	Paid	$2.99	2990000
GAME	4.3	85468	36M	1,000,000	1,000,000+	Paid	$0.99	990000
GAME	4.4	32496	99M	1,000,000	1,000,000+	Paid	$0.99	990000
GAME	4.5	32344	29M	1,000,000	1,000,000+	Paid	$0.99	990000
GAME	4.4	38419	100M	1,000,000	1,000,000+	Paid	$0.99	990000
GAME	4.3	13604	96M	1,000,000	1,000,000+	Paid	$0.99	990000
GAME	3.7	42529	92M	500,000	500,000+	Paid	$1.99	995000
GAME	4.5	38207	15M	100,000	100,000+	Paid	$9.99	999000
GAME	4.2	13004	46M	100,000	100,000+	Paid	$6.99	699000
GAME	4.6	10795	69M	100,000	100,000+	Paid	$5.99	599000

iii. Why do some genres seem to be more popular and garner a large number of installs as opposed to other categories? Is it because a lot more people are interested in the use of these applications? Is it just because they are free applications? Or they probably meet a certain need and provide services that are considered important and therefore are seen as “must-haves”? What factors personal or external influences the installs and downloads of these applications from the google play store.

However, it is pertinent to note that while being free could have contributed somewhat to the number of downloads of these applications but it is evident that those with the highest downloads have a lot more to offer and an appeal that goes beyond and above just being a free application. There are a lot of other free applications that did not make it to the list of the highest number of installs.

I did note that generally the average ratings for the paid applications were higher than those of the free applications which might be quite logical, considering that a developer will probably pay more attention to an application that is paid-for than a free one and as such a paid application would most likely have fewer glitches, errors and bugs which might translate to a better user experience on the whole.

Figure 11. The Sum of Installs across Type & Category

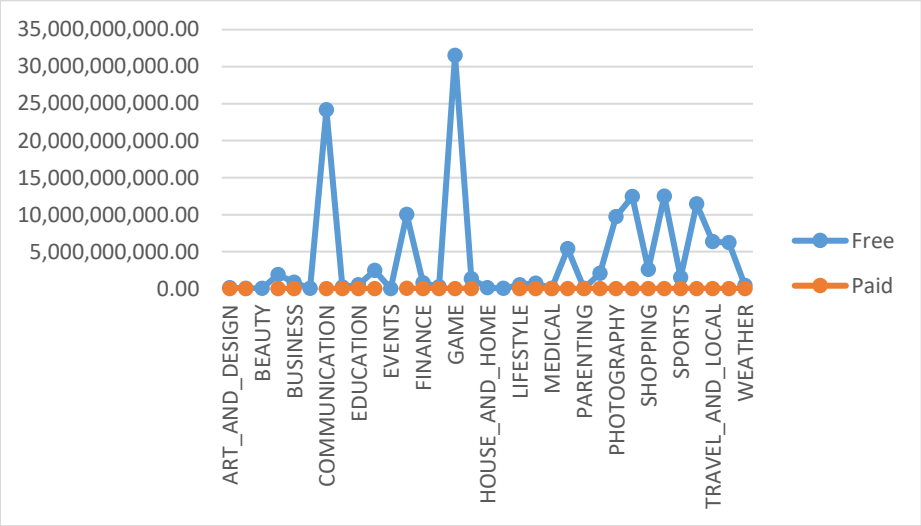
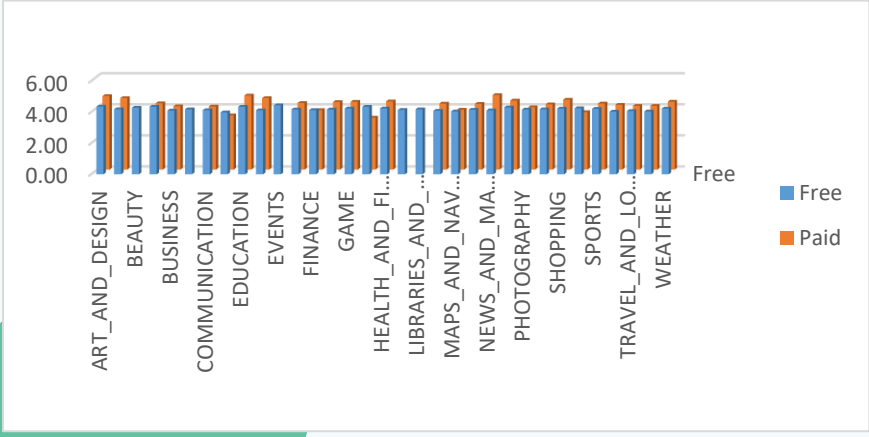


Figure 12. Application Ratings for Types



Conclusion

I set out to find trends and see what information can be gotten from the data that will be advantageous to a developer. The data shows that the genres that enjoyed the highest rate of installs were Games, Communication, and the Family categories. In the Communication category digging deeper showed that Hangouts, Google Chrome, and WhatsApp amongst many others had the highest installs while in the Family Category it's the entertainment genre that carried the day having the highest installs and downloads. Also, the data showed that the last year updated (that is the year 2018 in particular) had the highest rate of installs for all applications; this shows that installers were particular about the last time the applications were updated and therefore had possibly installed the latest versions of the applications.

The data set does not tell us how many reviews were positive and how many were negative, this could have been used to better analyze the effect the reviews had on the number of installs for the applications. Also it is evident that some applications, in particular, have a high rate of installs like Facebook, Gmail, Google Chrome, Google, Hangouts, Instagram, WhatsApp, Subway Surfers, Youtube, amongst others, showing a preference for applications that fit into a specific genre. Also, I have not been able to correctly assess whether the size of the application has any direct impact on the number of installs.

My recommendation, therefore, is that application developers who want to design applications that trend, or that will enjoy a high rate of installs have to look to design applications which fall within these trending genres and categories (Games, Communication and Family). The developer must also be ready to keep the application updated regularly as the latest updated year 2018 had the highest rate of installs across board. Also, if the application will have a cost attached to it, that is, it's going to be a "paid" application then it should not cost more than $6.99 as it is clear that the applications with costs higher than seven dollars ($6.99) did not have enough installs to earn a substantial amount in millions.

Appendix

- Google Sheets Data Set for Udemy Project

Link here: https://docs.google.com/spreadsheets/d/1k80YwQkl-VpG3Dbl_UAyEbfm0PK47EfXsUuLJvAcI3k/edit?usp=sharing

- Tableau Data Visualization for Udemy Project

Link here: EntryL_UdemyDashB | Tableau Public

- Google Sheets Data Set for Capstone Project

Link here: https://docs.google.com/spreadsheets/d/1R429__XHzWH3AWknziF6ExuIUr-LG-89ZfEG2vYmlZI/edit?usp=sharing

- Tableau Data Visualization for Google Play Store App Data (Capstone Project)

Link here: CapstoneP | Tableau Public