COMMERCIAL SUCCESS FACTORS IN THE GLOBAL HEALTH AND FITNESS APPLICATIONS MARKET

by

Hlib Usovych

A thesis submitted in partial fulfillment of the requirements for the degree of

MA in Business and Financial Economics

Kyiv School of Economics

2025

Thesis Supervisor: ________________________________ Pavlo Prokopovych

Approved by ____

Head of the KSE Defense Committee, Professor [Type surname, name]

____

____

____

Date ______________________________________

ACKNOWLEDGMENTS

The author wishes to [Click and type acknowledgments]

TABLE OF CONTENTS

LIST OF FIGURES iii
LIST OF TABLES v
LIST OF ABBREVIATIONS vi
Chapter 1. Introduction 1
Chapter 2. Industry Overview and Related Studies 4
Chapter 3. Methodology 10
Chapter 4. Data 19
Chapter 5. Results 28
    5.1. Hypothesis 1 findings 28
    5.2. Hypothesis 2 findings 29
Chapter 6. Conclusions and Recommendations 31
REFERENCES 34
APPENDIX 1

LIST OF FIGURES

Number	*Page*
Figure 1 – Geographic distribution of the annual revenue	7
Figure 2 – Compound Aggregate Growth Rate	8
Figure 3 – Google Colab script keywords	12
Figure 4 – Variance Infaltion Factor Analysis	14
Figure 5 – VIF analysis results (precise)	14
Figure 6 – Model A (OLS)	15
Figure 7 – Social integration and AI regression (source – R code)	16
Figure 8 – Model B (in R)	17
Figure 9 – Model C	18
Figure 10 – Niche category bar chart (generated in R)	20
Figure 11 – Dataset 1 Correlation matrix	21
Figure 12 – Correlation heatmap (Dataset 2 – created in Python)	27
Figure 13 – In-App Purchases Price Polynominal Fit (Dataset 2)	29
Figure 14 – In-App Purchases Price Polynominal Fit (Dataset 1)	30

LIST OF TABLES

Number *Page*

Table 1. Descriptive statistics for Dataset 1 (created in R) 20

Table 2. Table 2 – Sensor Tower descriptive statistics (part 1, calculated in Python scripts) 23

Table 3. Sensor Tower dataset descriptive statistics (Part 2) 25

LIST OF ABBREVIATIONS

CAGR Compound Annual Growth Rate

WHO World Health Organization

IAP In-app Purchases

RpD Revenue per Download

DAU/MAU Daily/Monthly active users

WW Worldwide

B2B Business to Business

B2C Business to Consumer

API Application Programming Interface

CHAPTER 1. INTRODUCTION

The global market of health and fitness applications represents one of the faster growing segments of digital health ecosystem. Based on Statista reports, it shows compound annual growth rate of of 8.58% with the yearly revenue of US$6.74bn in 2024 and is projected to grow 10.1% in 2025 (Statista, 2024)

Several factors converged to create conditions for wide market adoption of this apps. One of them was rising cost of healthcare and prevalence of chronic diseases, combined with high smatphone adoption and worldwide access to internet. COVID-19 and subsequent push for healthcare and digitalization enchanted the speed of health and fitness apps adoption.

One of the most popular niches in the market was born because there exists a growing trend of health problems connected with excessive weight, obesity and sedentiary lifestyle. By 2022, there was a two and a half billion adults who are overweight (World Health Organisation, 2022), and out of them there was around 900 million people with obesity. With growing automation of physical jobs and prevalence of unhealthy processed foods in the market, this trend, acknowledged by WHO as an obesity crisis, is unlikely to reverse.

With this in mind, there exists a great market for digital products that would provide a proven and safe information on healthcare, track important health parameters, control weight, create working sport plans for self-development and much, much more. Growing worldwide trend of health problems connected with high weight and sedentary lifestyle makes sure that this market would stay relevant and growing in the coming years.
Despite favorable market conditions, the Health and Fitness app market is characterized by high rates of failure and significant performance disparities. Most apps have problems

with user acquisition, usually demonstrating poor retention rates, and as the result – fail to achieve meaningful and stable revenue in the long term. This suggests that success on the market depends on specific, identifiable factors rather than just good timing of market entrance.

Most existing studies either research individual apps or work with too broad of context, limiting the usability of findings to the the specific market. The rapid development of app features, changing expectations of the users, and dynamic conditions of the market call for fresh empirical analysis using comprehensive datasets that capture the full spectrum of market performance indicators.

This thesis aims to understand both segment-specific and market-wide factors influencing revenue, downloads and ratings.

One of chosen hypotheses aim to find and quantify the impact of certain popular features, such as AI-driven personalization, social features and gamification on the app`s success. Other hypothethis aims to find the impact of pricing of in-app purchases on the revenue. The third hypothesis aims to find the correlation between the chosen niche and the app downloads.

Overall, the results of hypotheses testing are expected to show optimal pricing range, advantageous features and niches for upcoming app creators to enter the market.

This thesis is structured as follows: The introduction presents the research context, objectives, and hypotheses. Chapter 2 provides necessary information on the worldwide health and fitness application market, such as geographical structure of the market, growth rate and niche segmentation, as well as a review of the relevant literature. Chapter 3 outlines the methodology, focusing on the regression models used to identify the determinants of revenue, downloads and ratings. Chapter 4 describes the data used in the analysis, with information on key variables, their transformations and their sources. Chapter 5 presents the estimation results, interpreting the effects of the various factors on commercial success metrics. Finally, Chapter 6 concludes with a summary of findings.

CHAPTER 2. INDUSTRY OVERVIEW AND RELATED STUDIES

Mobile applications are already passed the stage of being perceived as groundbreaking innovate,ions and now are part of our daily routine. Therefore, there is a fair share of scientific literature and research papers on this topic. Howev,er, when working with more specific economic context and sub-niches in mobile application market, research becomes much more scarce.

The stuidies advanced gradualy during the last decade. Earliest works usually took worked with descriptive analysis – finding basic, correlations between the metrics available on the market. Lot of works were focused on behavourial, rather than financial patterns. Now the papers on behavourial and technological factors has grown quite sophisticated, especially since the smartphones, smart watches and other appliances allowing the tracking of user data are infusing the markets more and more.

Now let's go into the short overview of the works in the field. Inukollu et al. (2014) was one of the earliest works. He researched factors that influenced app quality – and consequently, the engagement of the users. So as the result, study shown that user experience and, also lifecycle management have an effect on the engagment in the long term—those ideas would later greatly influence studies of user retention.

Interestingly, the Roma, Perrone, and Valenti (2016) conducted one of the first empirical analyses in the market of mobile applications that worked precisely with the revenue.

This is one of studies that founded the notion that so called freemium models (monetization models with free basic functionality and paywalled advanced features) are more effective than the monetization models relying on paid apps or other single monetization method. Their finding also shown the importance of high user ratings as a predictor of the revenue. However, this study worked with market as the whole homogenious set which isn`t exactly true – it also missed some of the retention metrics and focused more on market as a homogenous ecosystem and did not consider specificifics of mobile Health and Fitness niche. Dataset also didn`t include retention metrics, RpD and so on, rather focusing on effects on reputation and fame – a flaw that my thesis would work with.

Then came Lin, Althoff, and Leskovec (2018) - their work engaged with the behavourial analysis at large scale. They worked with the market of activity-tracking applications and their work resulted in formulating the idea of "multiple user lifecycles". Their research shown that the complexieties of the user behaiviour didn`t stop at retention and churn rates. Rather, this finding shows that usual user at some point abandons the app and comes back later for multiple times.

One of the early works on the topic dealing with specifically mobile health application market is research of marketplace data with machine learning instruments (Gokhan Aydin, Gokhan Silahtaroglu, 2021). Their finding showed that functions dealing with social interaction, gamification and privacy disclosures had significant correlation with higher downloads and ratings. However, it dealt with very narrow topic, mainly concerning descriptions of applications and their influence on ratings. However, the topic should be researched further – exploring links to main determinants of success, such as revenue and downloads, not just ratings.

Complementary behavioral insights were provided by Yang and Koenigstorfer (2021), who empirically tested how motivational and gamification-related features influence users' physical activity intentions. Using a cross-sectional survey based on the UTAUT2 model, they demonstrated that habit formation, perceived usefulness, and education-related features are major drivers of adoption. Although valuable, their approach relied on self-reported intentions rather than real behavioral or financial data, limiting its applicability for market-level inference.

Applied Microeconomics of Freemium Pricing Strategies in Mobile App Market (Naixin Zhu, 2023) is in-depth research of the market. His dissertation aimed to help venture capital investors to find great products on early stages of development and understand how niche of the product influences its success. It introduced a niche index (a quantitive measure to understand how the positioning of app differs from competitors) and measured it effects on discoverability of app and other metrics. It also measured effects on different monetization models on consumers.

This is in-depth research on a big market, that, however, doesn`t concern itself with finding effects in smaller niches, and measuring market-specific variables. Therefore, Health and Fitness applications market is defined as a market containing apps focused on health tracking, stress managing, working out or helping with treatment of a minor conditions. Usually such apps aim to provide systematic, scheduled and interactive experiences, encouraging customers to maintain healthy lifestyles while providing them with all necessary tools for it.

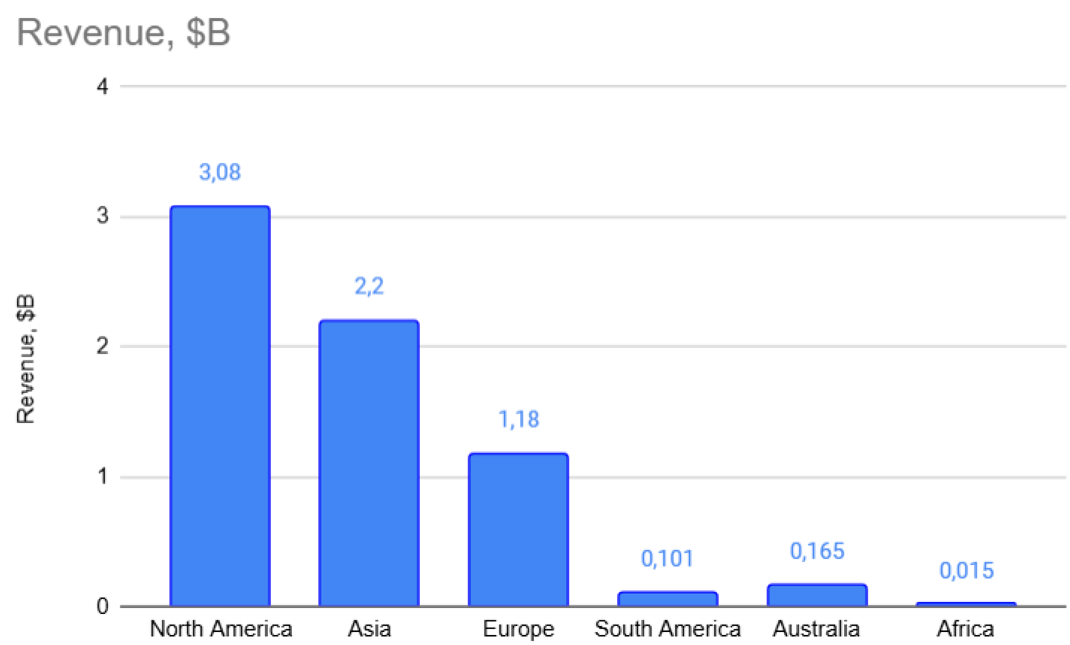
Across this literature, there several gaps for further research.

This study aims to address these gaps by conducting an empirical, data-driven investigation into the main commercial success drivers in the Health and Fitness applications market. By integrating real and up-to date marketplace data (Appmagic, Sensor Tower) with feature-level text analysis, the research tries to quantify how AI-driven personalization, social integration, affect revenue. Furthermore, it tries to find out optimal in-app purchase pricing range and explore relationships and correlations between deifferent predictors available on the market.

Based on Statista reports, it shows compound annual growth rate of of 8.58% with the yearly revenue of US$6.74bn in 2024 and is projected to grow 10.1% in 2025. The geographical structure of the health and fitness market is structured as follows. North America is the undisputable leading sector with $3.08 billions in revenue projected in 2025. Second strongest contender on the market is Asia, accounting for $2.20 billion in 2025. Europe takes the third place with $1.18 billion. Australia and Oceania fourth with 165 million. South American countries count for $101 million of the market revenue, and Africa is last with only $15 million. (Statista, 2024).

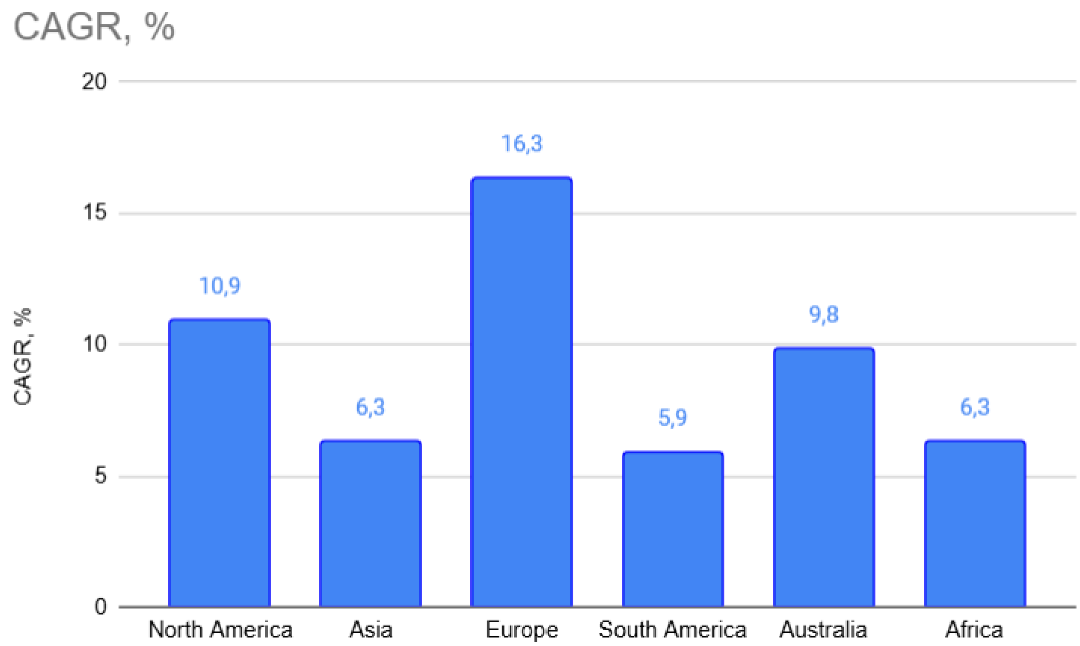
This distribution is visualized in Figure 1.

Figure 1 – Geographic distribution of the annual revenue



However, the CAGR tells us another story. North America, the highest grossing region, shows just 10,9% of CAGR in the 2024. Asia and Africa show the lowest CAGR, despite the difference in the scope of the market. Europe is the fastest-growing region, while the South America is the slowest. (Statista, 2024) This shows the market future potential for new entrants.

Figure 2 – Compound Aggregate Growth Rate



It`s also possible to categorize Health and Fitness market by main niches and geographical regions. Databases such as Appmagic and Sensor Tower use following classification:

1. Healthcare/Medicine niche – apps that are used to connect with medical professionals (booking appointments, video connection, etc), keep track of medications and health parameters, test results and so on. A great example would be epocrates+ app that is not only includes functions mentioned above, but also gives great tool to identify unknown drugs, check their interaction with each other and gives great guides on how to diagnose and treat illnesses.

2. Maternity niche – mainly "baby tracker" applications for parents to monitor sleep cycles, moving etc. of their child and help. Some apps are more focused on mother in pregnancy period instead. Good example of such app is Huckleberry.

3. Weight Control – one of the biggest niches that contains nutrition tracking apps – calorie counters, personalized food planners, tracking statistics of food and fitness and even food

scanners that analyze the products you are planned to eat. MyFitnessPal is one of the biggest applications on the market now.

4. Sport Trackers – applications to monitor workouts, creating challenging training schedules for gradual improvement, track health parameters such as heart rate, sleep, calories burned, steps taken etc. Fitbit is one of the biggest players in the market right now.

5. Yoga – applications to provide unique yoga workouts to the user's preference for all levels, usually with video guides and track progress.

6. Smart Ring – small niche of applications working with smart ring, smart watches and other appliances that can track health parameters. Usually provide customer with

7. Period Tracker – niche with applications for period tracking, managing symptoms and drawing insights from unique experience of the customer.

8. Workout – large group of apps with specialized workouts (running, core, back, etc.) that usually offer access to the thousands of workouts for the customers taste, set training goals, track progress and health parameters. There is also a subset for training equipment apps, such as Peloton.

9. Others – usually, specialized apps that can`t be described by categories above – heart rate monitors, sugar trackers, etc. For the future analysis, this category was omitted, because inclusion of applications from all types of

CHAPTER 3. METHODOLOGY

This study deals exclusively with data on the B2C market – apps that can be downloaded from major stores like Apple Store or Google Play. B2B solutions and revenue gain from out-of-app purchases and subscriptions are not included because of difficulties with tracking them.

I began my research working with existing datasets containing scraped Google Play Store data. Such datasets are extremely outdated for such a dynamic market (for example, the Kaggle dataset with more than 10 000 apps scraped data available is dated 2019). However, the structure of the data itself is useful for my research, as it sheds some light on the useful categories of information existing in the market.

1. Downloads/installs (scraped from Google Play Store/Apple App Store; can also be taken from Appmagic and Sensor Tower Datasets) which would be used as one of secondary metrics of app success (main metric for apps that don`t feature any subscriptions/in-app purchases and fully rely on ads revenue)
2. Type of the app (Paid/Free – scraped from Google Play Store/Apple App Store)
3. Minimal IAP price as a proxy for the subscription price (scraped from Google Play Store/Apple App Store) – iTunes API doesn`t clearly give the customer the list of prices, but Google Play scraper does - however
4. Rating of the app – to understand consumer sentiments.
    a. Number of the reviews
5. Revenue (downloaded from Appmagic/Sensor Tower) – main metric to judge success of the App
    a. Revenue Per Download (from Sensor Tower)
6. Niche (as binary variable) to understand if some niches are more successful than others
7. Existence or absence of a particular features (AI-driven insights, social integration features, etc.) (collected with use of text scraping)

8. Daily and Monthly active users – useful metrics to understand user retention (can be found on Sensor Tower)
9. Minimum In-App Purchase Price – proxy for the monthly subscription price, in USD
10. Retention for Day 1, 7, 30 (sensor tower) – the percent of the users that remains of the app after the user

The amount of the data is vast and difficult to scrape because of the existing instrument limitations. I focused on key categories of the data to form my dataset – Revenue, Downloads, use of social features or AI –driven personalization, app price, ratings and niche.

Download and Revenue were collected as of August 2025 from individual from the pages on Sensor Tower, which gives monthly updates on such indicators with higher accuracy than Appmagic.

Niche category was determined by niche leaderboards on Appmagic (as of August 2025), as a binary variable, where 1 shows app belonging to particular niche.

Social and AI features in the apps posed a challenge, since to understand their existence you had to read every single description and make an expert guess. The process was automated with the text-scraping script in Python (Google Colab) that used several API to scrape the text from Google Play and Apple App Store and check the existence of a long set of keywords.

Here are lists of the keywords from both categories. It doesn`t give a 100% accuracy, but is supposed to be pretty close to the true picture. To acess the code used further, you can refer to Hlib_Usovych_Thesis Scripts.ipynb (Script 1).

Figure 3 – Google Colab script keywords

```
SOCIAL_KEYWORDS = [
    "social", "socially", "friend", "friends", "invite", "invites", "invited",
    "share", "share with", "share your", "sharing", "compare", "compare results",
    "leaderboard", "leaderboards", "rank", "ranked", "competition", "compete",
    "compete with friends", "community", "communities", "follow", "followers",
    "follow friends", "groups", "team", "teams", "challenge", "challenges",
    "connect with friends", "kudos", "club", "clubs", "matchmaking"
]

AI_KEYWORDS = [
    "ai", "ai-powered", "ai powered", "artificial intelligence", "machine learning",
    "ml", "llm", "gpt", "chatbot", "neural network", "neural networks",
    "personalized", "personalised", "personalization", "personalisation",
    "algorithm", "algorithms", "recommend", "recommendation", "recommendations",
    "adaptive", "adaptive program", "smart suggestions", "smartly", "predictive",
    "predict", "automated", "automation"
]
```

Minimum IAP, Ratings and Reviews were collected with the help of the Python script using Google Scraper API and iTunes API. Minimum IAP was used as a proxy for the monthly subscription price, which is usually the lowest priced item in the apps, as seen on most of individual app pages. For the structure of the code used, refer to the Hlib_Usovych_Thesis Scripts.ipynb (Script 2).

There is also a secondary dataset based on Sensor Tower data – much more precise but limited in scope (refer to Sensor Tower data (uncleaned) page in submitted Excel file).
It contained the 13 metrics:

- Last 30 Days Downloads (WW)
- Last 30 Days Revenue (WW)
- RPD (All Time, WW)
- Last 30 Days Average DAU (WW)
- Last Month Average MAU (WW)
- Day 1 Retention (Last Quarter, WW)

- Day 7 Retention (Last Quarter, WW)
- Day 30 Retention (Last Quarter, WW)
- Age (Last Quarter, WW)
- Genders (Last Quarter, WW)

And Social AI, Min IAP, Rating and Reviews for the appropriate apps from main dataset. This dataset was mainly used for the descriptive statistics and insights, the regressions had too much problems.

To answer the overarching question of the study it is crucial to answer several sub-questions using quantitive analysis. Those questions (and relevant hypotheses for comprehensive quantitive analysis) are:

1. What is relationship between features and commercial success metrics? Which app features (social, AI-driven insights, etc.) best predict revenue?
    a. H1: Social integration and/or AI-driven insights features will generate significantly higher Revenue compared to apps without those features.
2. How do different app prices and in-app purchases impact the generation of the revenue?
    a. H2: Prices between 4 and 8 dollars would be associated with the highest revenue, with revenue going up on the 0-4 zone and down on 8+ zone.

For the testing of the Hypothethis 1, I made several steps.

Firstly, created the correlation diagram to see if those parameters correlated with the revenue.

Then I considered regression analysis with the two datasets. Appmagic one (named MainDataset in Excel – data on more than 1900 apps of seven different niches, which was, however, not super accurate, but gave the basic data for further analysis). The second one (Sensor Tower data (uncleaned) in Excel) had very precise data on 186 apps from weight control niche (described better in the Data chapter).

To prepare the first dataset for the regression running, I cleaned the data as a part of the python script. It resulted in the data of the 1117 apps remaining. To test data for multicollinearity, VIF test was performed.

You can see the results of the VIF test in Figure 4 and 5

Figure 4 - Variance Infaltion Factor Analysis

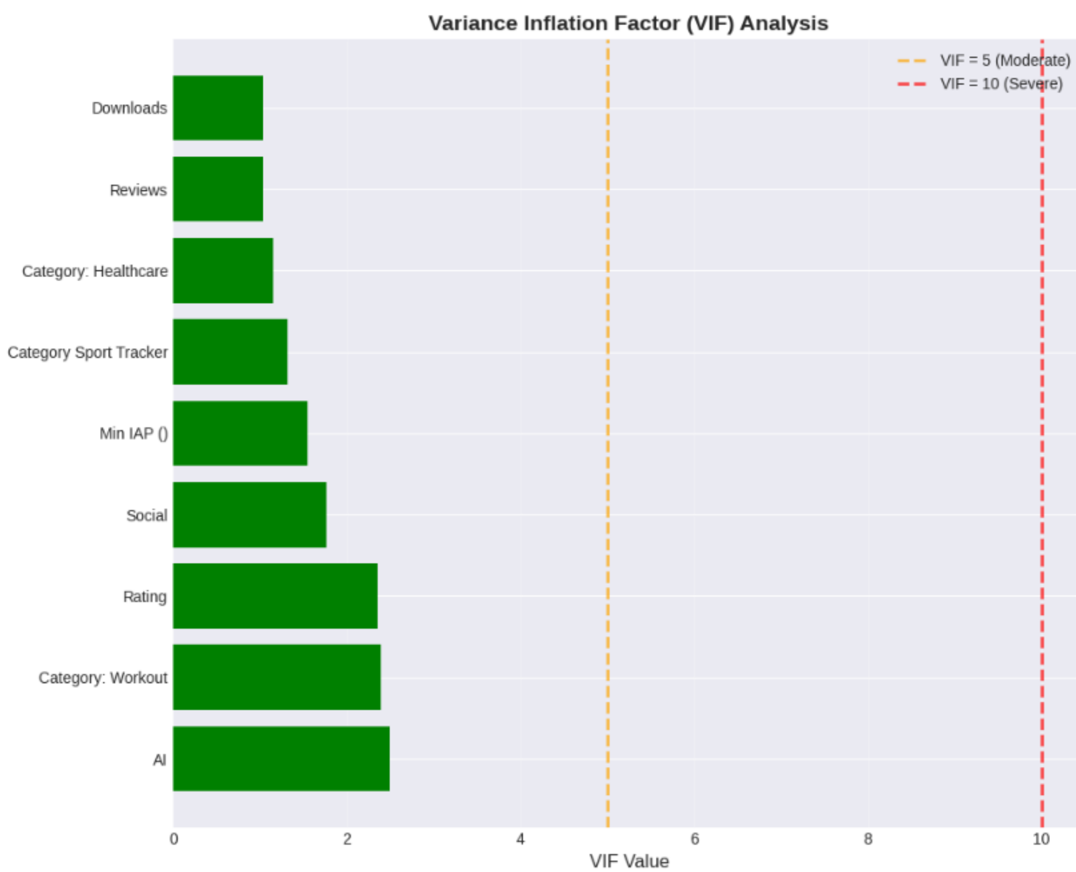


Figure 5 – VIF analysis results (precise)

```
Variance Inflation Factor (VIF) Results:
                   Feature      VIF
                        AI 2.489538
         Category: Workout 2.385397
                    Rating 2.345173
                    Social 1.755500
                Min IAP () 1.537415
  Category Sport Tracker 1.315998
      Category: Healthcare 1.144835
                   Reviews 1.037665
                 Downloads 1.034961
```

All types of the data demonstrated correlation between 1 and 5 – low correlation between the data that wouldn`t cause problems for further analysis.

The first model used to find the effects of revenue using Appmagic dataset were ordinary least squares model with most of the data. Here is select few findings, which lead me to abandon the model.

Figure 6 – Model A (OLS)

```
--------------------------------------------------------------------------------
OLS Regression Results:
--------------------------------------------------------------------------------
                            OLS Regression Results
==============================================================================
Dep. Variable:                Revenue   R-squared:                       0.104
Model:                            OLS   Adj. R-squared:                  0.096
Method:                 Least Squares   F-statistic:                     14.24
Date:                Mon, 03 Nov 2025   Prob (F-statistic):           6.09e-22
Time:                        20:05:34   Log-Likelihood:                -16548.
No. Observations:                1117   AIC:                         3.312e+04
Df Residuals:                    1107   BIC:                         3.317e+04
Df Model:                           9
Covariance Type:            nonrobust
==========================================================================================
                             coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                   3.406e+04   8.04e+04      0.424      0.672   -1.24e+05    1.92e+05
Downloads                  0.1074      0.027      3.926      0.000       0.054       0.161
Social                  1.642e+05   6.68e+04      2.458      0.014    3.31e+04    2.95e+05
AI                      1.854e+04   5.86e+04      0.316      0.752   -9.65e+04    1.34e+05
Min IAP ()              5082.1643   2876.077      1.767      0.077    -561.014    1.07e+04
Rating                  6727.5433   1.73e+04      0.388      0.698   -2.73e+04    4.08e+04
Reviews                    1.4583      0.151      9.677      0.000       1.163       1.754
Category: Workout      -6.124e+04   4.87e+04     -1.258      0.209   -1.57e+05    3.43e+04
Category: Healthcare   -1.026e+05   9.36e+04     -1.097      0.273   -2.86e+05     8.1e+04
Category Sport Tracker  -8.87e+04   8.37e+04     -1.059      0.290   -2.53e+05    7.56e+04
==============================================================================
Omnibus:                     2007.284   Durbin-Watson:                   1.414
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          2140910.446
Skew:                          12.249   Prob(JB):                         0.00
Kurtosis:                     216.072   Cond. No.                     3.90e+06
==============================================================================
```

Most of the predictors in current form turned to be statistically insignificant. The only one with P>|t| lower than 0,05 were Downloads, Social integrations, and reviews. Min IAP was close (0,077) but not quite. R^2 (0,104) of the model has shown that it explained very little of revenue parameter.

Residual diagnostics all shown that the distributions are not normal. Breusch-Pagan test (see in Appendix 1) has shown that data is heteroscedasticic, which means that estimates may be unreliable.

I decided to move forvard with Hypothesis 1 and create the model focused on Social integrations data and AI data as a predictors.

The formula for the next model were as follows:

lm(formula = log_Revenue ~ Social + AI)

The result was as follows:

Figure 7 – Social integration and AI regression (source – R code)

```
lm(formula = log_Revenue ~ Social + AI, data = data_model)

Residuals:
    Min      1Q  Median      3Q     Max 
-9.4681 -1.0202 -0.9507  0.4355  6.5805 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  9.53757    0.05859 162.784  < 2e-16 ***
Social       0.51739    0.15594   3.318 0.000939 ***
AI          -0.06949    0.12611  -0.551 0.581760    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.56 on 1027 degrees of freedom
  (87 observations deleted due to missingness)
Multiple R-squared:  0.0128,	Adjusted R-squared:  0.01088 
F-statistic: 6.658 on 2 and 1027 DF,  p-value: 0.001339
```

This model explains very little of the variance in the revenue (R^2=0,0128), but it is statistically significant. Existance of social effect has shown to up the revenue by 51%

compared to the apps without one. Effect of AI in the applications is statisticaly insignificant.

Use of non-log revenue performed far worse, explaining only 0,06 of the model (see the Dataset 1 Analysis).

Than the fuller model were used:

Figure 8 – Model B (in R)

```
Call:
lm(formula = log_Revenue ~ Social + AI + log_Downloads + Rating +
    log_Reviews + `Category: Workout` + `Category: Healthcare`,
    data = data_model)

Residuals:
     Min       1Q   Median       3Q      Max
-11.0759  -0.9020  -0.4171   0.7979   5.1180

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)              5.93840    0.33367  17.797  < 2e-16 ***
Social                   0.41834    0.14457   2.894 0.003888 **
AI                      -0.08636    0.12440  -0.694 0.487719
log_Downloads            0.35264    0.03120  11.302  < 2e-16 ***
Rating                  -0.04704    0.04520  -1.041 0.298241
log_Reviews              0.06255    0.01740   3.594 0.000341 ***
`Category: Workout`     -0.12338    0.10045  -1.228 0.219597
`Category: Healthcare`  -0.11157    0.20330  -0.549 0.583269
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.442 on 1022 degrees of freedom
  (87 observations deleted due to missingness)
Multiple R-squared:  0.1612,	Adjusted R-squared:  0.1554
F-statistic: 28.06 on 7 and 1022 DF,  p-value: < 2.2e-16
```

Most of predictors has shown to be statisticaly insignificant, however, the social integrations existence shown to up the revenue by 41%. AI still statistically insignificant and the model explains just 0,1612 of the variance in the log_revenue.

There were several other models, perfoming worse, as you can see in the codes, so they weren`t mentioned here.

The only exception is this model from the python script.

Figure 9 – Model C

```
--------------------------------------------------------------------------------
Log-Transformed Regression: Log(Revenue) ~ Log(Downloads) + Others
--------------------------------------------------------------------------------
                            OLS Regression Results
==============================================================================
Dep. Variable:            Log_Revenue   R-squared:                       0.231
Model:                            OLS   Adj. R-squared:                  0.225
Method:                 Least Squares   F-statistic:                     36.97
Date:                Mon, 03 Nov 2025   Prob (F-statistic):           1.42e-57
Time:                        20:05:42   Log-Likelihood:                -1944.2
No. Observations:                1117   AIC:                             3908.
Df Residuals:                    1107   BIC:                             3959.
Df Model:                           9
Covariance Type:            nonrobust
==========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                     5.3860      0.310     17.368      0.000       4.778       5.994
Log_Downloads             0.3977      0.028     14.184      0.000       0.343       0.453
Social                    0.5052      0.140      3.605      0.000       0.230       0.780
AI                        0.0392      0.123      0.319      0.749      -0.202       0.280
Min IAP ()                0.0485      0.006      8.026      0.000       0.037       0.060
Rating                    0.0288      0.037      0.786      0.432      -0.043       0.101
Reviews                 1.39e-06   3.17e-07      4.383      0.000    7.68e-07    2.01e-06
Category: Workout        -0.3538      0.102     -3.460      0.001      -0.554      -0.153
Category: Healthcare     -0.4603      0.196     -2.344      0.019      -0.846      -0.075
Category Sport Tracker   -0.7766      0.176     -4.419      0.000      -1.121      -0.432
```

It too, shows the social factors as statisticaly significant and upping revenue by 50%.

It also shows that minimal In-App Purchase price is statistically significant predictor for the log of revenue – upping minimal price a dollar results in revenue growing 4,8% at average.

CHAPTER 4. DATA

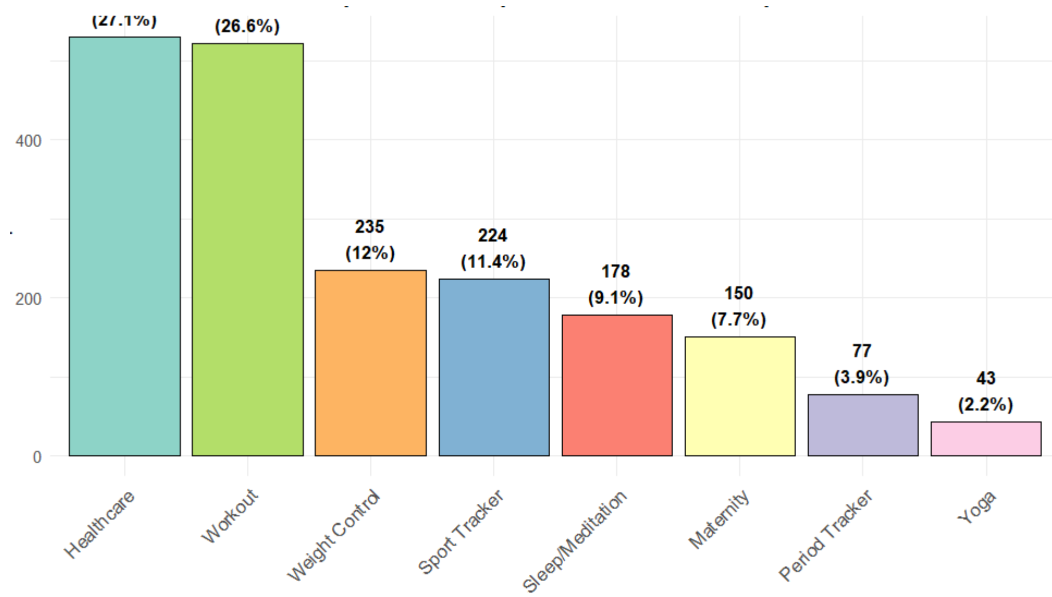
Most types of the data outlined in the term paper require a web-scraping off the app stores, and some of the data from Appmagic and Sensor Tower requires premium access to databases.

The first step was collection of the dataset – it was scraped from the Appmagic monthly charts of health and fitness apps. To keep the data of revenue relevant to each other I decided to collect the data for all metrics for August 2025 – the last month available before the beginning of data collection. Monthly data was chosen to find out accurate information of app performance in the current market. Other option available on Appmagic, lifetime data for either of those metrics, would result in the correlation being skewed heavily in favor of long-standing apps and wouldn`t showcase the status of current market.

The total volume of information collected before sorting of the data was 1836 apps, classified by seven niches. The composition of those apps were as follows:

- 519 apps in Workout category
- 530 in Healthcare category
- 235 in Weight Control category
- 224 in Sport Tracker category
- 150 apps in Maternity category
- 77 apps in Period Tracker category
- 43 apps in Yoga category

Figure 10 – Niche category bar chart (generated in R)



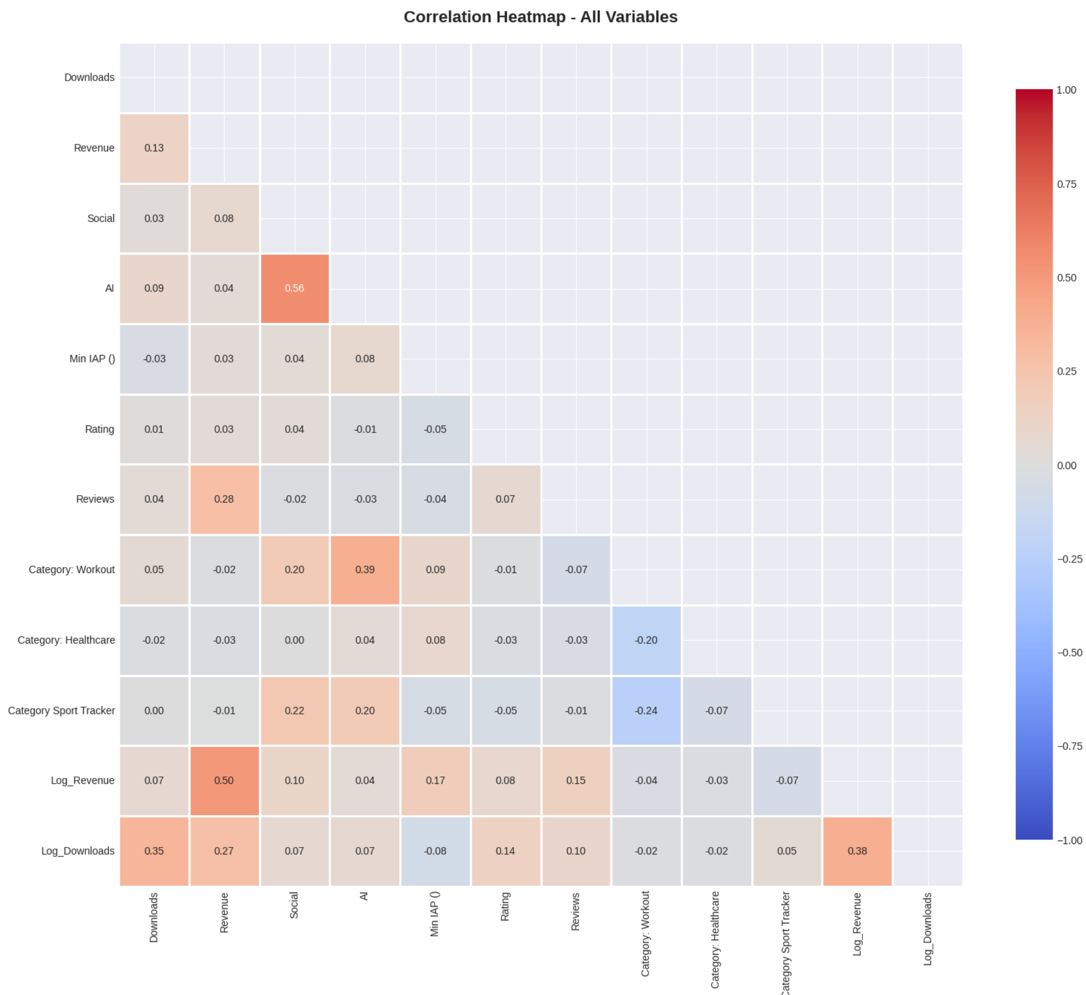
Most consequential data is the data on key performance metrics itself – downloads and revenue.

Several other metrics were collected as posiible predictors with google-play scraping – those are:

- Social
- AI
- Min IAP ()
- Rating
- Reviews

You can see the correlation matrix in the Figure 11.

Figure 11 – Dataset 1 Correlation matrix



There are also a second dataset – much more precise data on the 186 Apps in Weight Control niche, collected with help of the Sensor Tower. Peculiarities of the system that has maximum limits of the daily downloads, has the problems with mass downloading the information about the apps (having to add each app by hand, some bug with downloading data that wouldn't show anything except 2-3 additional metrics) have greatly limited the amount of apps I could work with – the main problem with the following work. Since data sample size is small, it probably couldn`t be used for the efficient enough predictive model, but they would show the key insights into the market that could be useful for further researchers.

Use of descriptive statistics on the data reveals the following picture. (descriptive statistics for both were used at the Appmagic-collected dataset. =

Table 1 – Revenue descriptive statistics for Dataset 1 (created in R)

Metric	Downloads	Revenue	Min IAP ()	Rating
Observations	1959	1092	910	1945
mean	71013.27	115956.96	5.69	3.89
Standard deviation	216233.97	620637.35	7.86	1.40
median	20000.00	5000.00	2.99	4.44
trimmed	30681.96	21258.58	4.17	4.24
min	5e+03	5e+03	5e-02	0e+00
max	5.000e+06	1.000e+07	1.199e+02	5.000e+00
range	4995000.00	9995000.00	119.94	5.00
skew	10.37	12.42	5.65	-1.97
kurtosis	168.10	181.97	58.19	2.74
Standard Error	4885.47	18781.34	0.26	0.03

The mean revenue (115956.96) is much higher than the median (5000) and mode (5000), which shows the presence of abnormally high values and very uneven distribution. Such a low median and mode are probably byproduct of the system of Appmagic labelling all apps with sub-5000 dollars revenue as 5000 – so the presence of the outliers. I expected low median and mode because the overwhelming majority of the apps hardly even produce enough money to cover the costs of development.

The standard error (18781.34) and is quite large. That could indicate substantial uncertainty and fluctuation in the mean. That's why I used trimmed mean to showcase that without outliers, most of the data still stays quite close to the median.

Standard deviaton and very high variance indicate high spread and variability within the revenue data.

The number of entries shows that revenue data for around a half of the apps wasn`t available, likely because of purely ad-driven revenue model that Appmagic can`t reliably guess the info about.

The sampled data from the Sensor Tower is much more precise, but limited in scope – its problematic to correlate only a small size with whole dataset, it was analyzed separately. After cleaning incomplete data from the dataset, it tells us a following story.

Table 2 – Sensor Tower descriptive statistics (part 1, calculated in Python scripts)

Variable	Downloads	Revenue	RPD	DAU	MAU
Mean	49845,38	239267,46	1,51	61846,38	418556,96
Median	8777,00	26223,00	0,82	42,00	199,31
St.d	93418,79	1076495,96	2,47	347715,83	1439989,76
CV%	187,42	449,91	164,21	562,23	344,04
Skewness	3,24	8,77	4,69	6,37	5,54
Kurtosis	12,05	82,27	28,19	43,61	36,82
count	111,00	111,00	111,00	111,00	111,00
min	4,00	0,00	0,00	1,00	1,19
0,25	1832,50	1038,50	0,16	11,94	42,11
0,50	8777,00	26223,00	0,82	42,00	199,31
0,75	57256,50	123555,50	2,04	157,42	722,58
max	579322,00	10799984,00	19,48	2917474,00	11894932,00

There is total of 111 full observations from the Sensor Tower. Other one don`t have either parts or full data, but uncleaned version shows partial information of 186 apps.

Over most categories of the data, data shows really high standard deviation and variance coefficient. This is especially evident with the first part of the data – Downloads, Revenue, average users and so on – those categories of data have both extremely high and extremely low outliers. It looks that there are also outdated or nonfunctional applications in the dataset – those no longer used showcase 0 in last month revenue and downloads, which also influences mean and the median.

Strong skewness and high curtosis with heavy tails show that this is not normal distribution, but rather one with the power law distribution. Such occurrence is not that rare in digital markets where very much relies on discoverability.

This finding supports general predictions from already available market research – the niche belongs to the few top leaders, while most of the apps earn just modestly or even can`t sustain themselves.

Then its time to move on to the specific metrics statistics. Downloads are skewed to the right, and have a long tail of app with good performance. Most popular apps in the niche, the global leaders, result in the mean that is x6 times higher than the median. Just a few apps have mass reach. Applications with the 57 thousand monthly downloads are in the top quartile, while being just above the mean – that means mean is not really reliable as a future benchmark, being overinflated by global leaders. Relative ranks would be better for lokking at app potential.

Interestingly, revenue distribution shows one of highest inequalities. Variance coefficient for this metric is 260%. Situation shows that market state is close to "winner takes all" situation. Lowest percentiles can`t even bring back the money spent of them. Only top percentile (or probably even smaller group) can allow stable marketing and post-release development. At median level, app is basically non-viable economically. Confidence interval shows true mean between 73 thousand and 211 thousand.

The Revenue per Download was chosen as the metric by me because it allows to control for apps popularity. Interestingly, different strategies and apps extract different value from user – and while the variation and confidence intervals are still pretty high, large RpD doesn`t necessarily mean high Revenue – rather, it depends to a monetization mode choice. RpD below half a dollar usually mean free app monetized by adds (minimal IAP is higher than that in all but edge cases) or a freemium model that not very effective in making the user pay for subscription. 1-3 dollars typically refers to better freemiums. High RpD (5+) usually means paid app or subscription based service.

Table 3 – Sensor Tower dataset descriptive statistics (Part 2)

Variable	Retention_Day1	Retention_Day7	Retention_Day30	Age	Male_Percent
Mean	0,34	0,13	0,06	33,33	47,43
Median	0,31	0,13	0,05	34,00	49,00
Std	0,13	0,07	0,05	2,91	10,26
CV%	39,26	52,10	73,31	8,73	21,64
Skewness	1,45	0,61	0,92	-0,11	-0,38
Kurtosis	5,00	0,21	0,05	-0,08	0,82
count	111,00	111,00	111,00	111,00	111,00
min	0,05	0,01	0,00	26,00	16,00
0,25	0,26	0,07	0,03	31,00	43,00
0,50	0,31	0,13	0,05	34,00	49,00
0,75	0,42	0,18	0,09	35,00	52,00
max	0,92	0,36	0,20	41,00	73,00

When I compare the retention metrics with all metrics examined before, I see drastical change. Skewness, kurtosis, standard deviation and coefficient of variance are now much lower, showing us that most of the applicainons are in the same playing ground. This is much closer to normal distribution than to Pareto one.

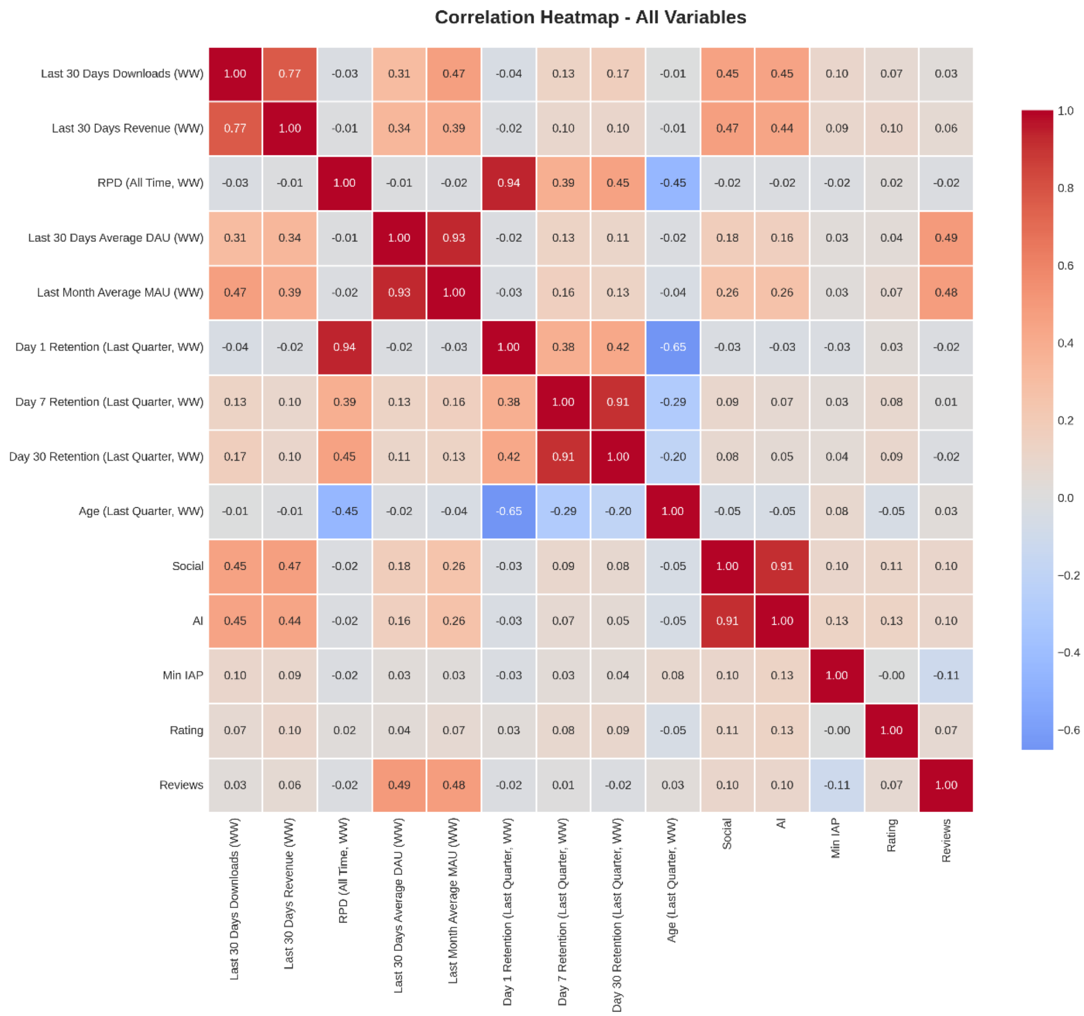
Day 1 Retention metric essentially shows us the success of initial impression, if the user liked application at all. At average, only one of three users returns back after the first day. This leads us to benchmarks that potential app should aim to and what to think as a lowest point - retention below 25% (so, in the lowest quartile) shows that the application turns off most of the users and something is due for the change. A good estimate to aim for the new app that wants to have a part of the market is 50+%, which will put it in a higher part of top quartile and allow it to cover further development and bring profit for the developing company. There are super strong performers with up to 92% of Day-1 retention.

Day 7 retention is the metric that measures how app captured attention of the user by the app features – most of them drop the application before the end of the week (mean of 0.34 and of 0.13 respectively). The apps that get dropped, from my position, either lack sufficient depth of content (only a limited amount of exercises, no special features, poor design, no long-term tracking functions etc.) or just have some unevident problems that user unearths only after a few visits. The data is very symmetrical and has lower variance. Even the stronger performers are hit proportionally with the user drop-off (almost 2/3 of user leave the app for good) and can`t pierce the 36%.

Its probably a good idea to put login streaks, bonuses and reminders to get into the top quartile. 20-25% of retention is, as I think, the good target to aim to.

Day 30 retention is the metric that shows the user activity cycles – there are not a lot of users that would stay faithfully for the one month, and if the person stays, its probably means that most of the retained group would stay with the app further – the app at that point probably endured one or two periods of user inactivity and going back. The coefficient of variance is highest out of the retention metrics and shows that the depending on strategies to form the habit of app use, perfomace will highly differ. The good range to aim for is top quartile 10-15% users that stay with you after the month of using your app.

Figure 12 – Correlation heatmap (Dataset 2 – created in Python)



Correlation heatmap reveals to me main possible predictors of the monthly revenue. Those are, in descending order:

- Monthly downloads (0.77) – most predictable one.
- Social features (0.47)
- Personalisation and AI-based features (0.44)
- Monthly and daily active users (0.39 and 0.34 respectively)

Some other interesting finds – RPD is highly correlated with Day 1 Retention and a bit worse with retention metric at other time intervals. Reviews have high correlation with the daily and monthly active users.

CHAPTER 5. RESULTS

My work on the thesis resulted in a lot of findings, some of which did directly stem from the hypotheses testing, and others were analytical insights gained from working with the different metrics and sources.

5.1. Hypothesis 1 findings

For the models and their results, look at the methodology section – here are just the findings.

The test of the first hypothesis on the 1st dataset partially proved the hypothesis, showing that the Social integrations explain revenue variance in statistically significant ways.

Most other predictors has shown to be statisticaly insignificant, however, the social integrations existence shown to up the revenue by 41%. AI still statistically insignificant and the model explains just 0,1612 of the variance in the log_revenue.

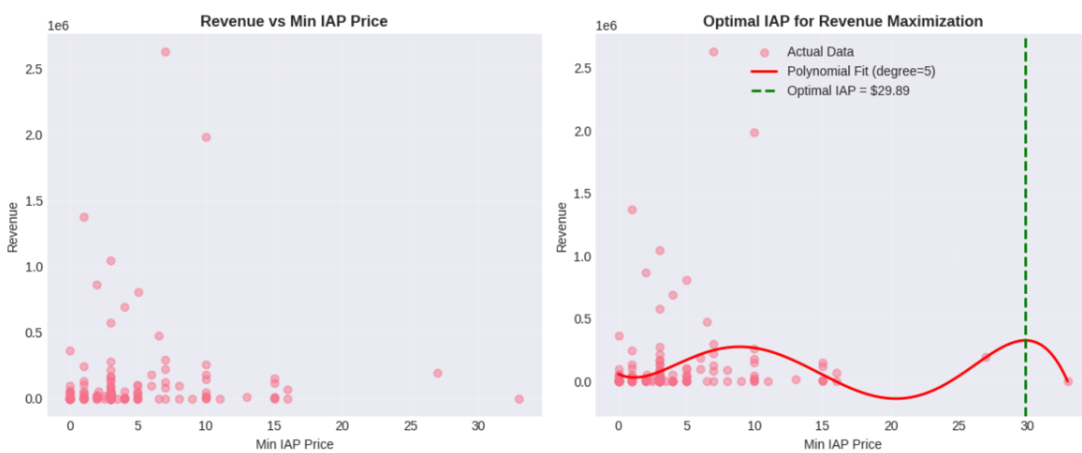
Even if the 2nd dataset (sourced form the Sensor Tower) shows that the both Ai and Social factors are heavily correlated with the revenue, this is statisticaly insignificant and can`t be used for further analysis because of multi ple problems with the data.

My finding show, that applications incorporating social features and/or AI integration show 44-47% higher average revenue. While these coefficients sometimes do not achieve statistical significance in multivariate analysis (likely due to small samples and confounding), the substantial bivariate correlations suggest these features may serve as differentiating capabilities.

5.2. Hypothesis 2 findings

Testing the Hypothesis 2 using the polynominal model to find the optimal monthly price price for subscription-based services (that I used minimal IAP as a proxy for).

Figure 13. In-App Purchases Price Polynominal Fit

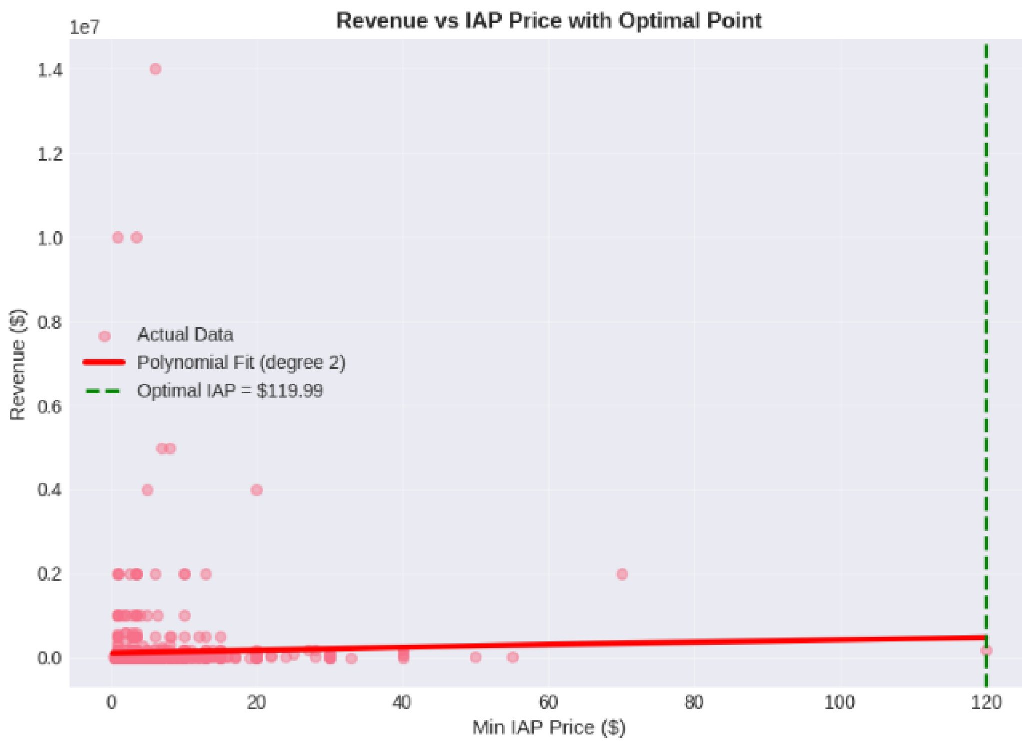


Polynominal fit predictions with the fifth degree of the fit determined two peaks of the data – the global peak at the $29,9, that is, however, to dependent on several outliers to be a reasonable insight, and the local peak around $9, with top region between the $7 and $11. This disproves my initial hypothesis – while there is indeed the value of In-App where revenue reaches the peak and then declines, I missed the target a lot. Peak IAP, which was usually the proxy for the subscription price, in Weight Control Market, is $7-11. Since such a subscription price is pretty high for the usual customer in a lot of countries, it is advised for the any developer to aim for appropriate markets (Such as

United States and, secondarily, Europe) – otherwise customers could find the price to high and not buy anything, which would tank total app revenue.

Test with the 1st dataset (Appmagic) has shown this:

Figure 14. In-App Purchases Price Polynominal Fit (Dataset 1)



This essentially tells us nothing. With the inprecise data of the Appmagic dataset on the revenue, price-revenue relationsips is weak, and the influence from the outliers influences the whole model way too much.

CHAPTER 6. CONCLUSIONS AND RECOMMENDATIONS

This thesis investigated commercial success factors in the global health and fitness applications market, with the empirical analysis of 1,117 apps, and detailed analysis of 186 weight control applications. My findings reveal several insights for any app developers who want to enter or compete in this dynamic Health and Fintess mobile applications market.

The descriptive statistics showed me the picture of market concentration. Revenue distribution is very unequal, the mean exceeds the median around x9 – that pattern is much more usual in context of the power-law distribution, not normal one. The first finding is only top top percentile of apps achieve something around basic economic viability. The median app in the weight control niche generates insufficient revenue to sustain further development, with only the top quartile (apps with $123 thousand dollars of revenue and higher) achieving significant market presence. The situation with downloads is almost the same. Applications with the 57 thousand monthly downloads are in the top quartile, while being just above the mean – that means mean is not really reliable as a future benchmark, and we better concentrate on relative ranks.

Retention metrics are mostly free from market overconcentration on top performers. Its shown by much lower coefficients of variance. My findungs show, that, althrough the best apps can retain up to 92% of the users on the first day, all apps are hit with the same user loss pattern in the long term – bleeding around 2/3 of the users in the first week, and around the half of whats left during the first month. The lowering of the churn rate on a long-term periods is consistent with Lin, Althoff, and Leskovec (2018) theory of "multiple user lifecycles", which means part of the users go back for after the period of abandonment.

Hypothesis 1 on the significance of AI and social features in the variance of the revenue yielded partial results. With the definition that relied on correct keywords for the collected data, AI-based features impact was shown to be statistically insignificant across multiple model specifications. On the contrary, regression analysis across multiple model specifications consistently demonstrated that apps incorporating social features generate 41-50,5% higher revenue compared to apps without such features. This effect remained consistent across the broad Appmagic dataset (1,117 apps). Correlation analysis of the Sensor Tower dataset also have shown significant (0.47) correlation with revenue. However, due to model overfitting, multicollinearity and low number of the observations, results of the actual model predicted the performace worse that the mean, so any findings from them are insignificant. While AI-driven personalization features had strong bivariate correlation (0.44) with revenue in the weight control niche, this relationship did not achieve statistical significance in any models. That means, as the result of the hypothesis testing tell me, social integration represents a differentiating capability that developers should prioritize. AI-features were either not defined well enough, or they truly don`t have any significant effect on app revenue.

Interestingly, minimal In-App Purchase price analysis revealed an optimal range countering the initial hypothesis. Polynomial fit modeling identified local peak of the revenue revenue at minimum in-app purchase prices of $7-11, and the local maximum around $9, significantly higher than the range I predicted. This result applies specifically to subscription-based monetization models in the weight control (trying this on Appmagic dataset failed to reveal any useful insights) and could be useful for pricing the product accordingly to have higher total revenue (it would be important for app to perform on the appropriate markets)

For the further researchers, I first and foremost suggest increasing data-gathering capabilities. Insights from the publically available information are not precise and limit the actual usefulness of the found insights on the market. My research suffered from the limited nature of the second dataset being confined to only 186 observations. Larger samples would result in more precise results and less problems with overfitting.

Some areas to look at further:

Separating the social integration effects into more niche groups for the more precise analysis of its effect on the app revenue. Defining the scripts for the popular nowadays gamification metric to measure its impact too.

Correlating the geographical structure of the userbase with the certain features to find out what appeals the most to the certain user demographic.

Trying inclusion of more features, binary variables of the niches, geographical structure, demographics of the user base into the models to better explain variations of the revenue.

REFERENCES

Amagai, Saki, Sarah Pila, Aaron J. Kaat, Cindy J. Nowinski, and Richard C. Gershon. 2022. Challenges in Participant Engagement and Retention Using Mobile Health Apps: Literature Review. *Journal of Medical Internet Research* 24 (April): e35120. doi: 10.2196/35120

Appmagic. Genre Classification. https://appmagic.rocks/tool-descriptions/genre-classification.

Aydin G, Silahtaroglu G (2021) Insights into mobile health application market via a content analysis of marketplace data with machine learning. *PLoS ONE* 16(1): e0244302. https://doi.org/10.1371/journal.pone.0244302

Gupta, L. 2019. Google Play Store Apps. Kaggle. https://www.kaggle.com/lava18/google-play-store-apps.

Inukollu, Venkata & Keshamoni, Divya & Kang, Taeghyun & Inukollu, Manikanta. (2014). Factors Influncing Quality of Mobile Apps: Role of Mobile App Development Life Cycle. *International Journal of Software Engineering & Applications*. https://doi.org/10.5121/ijsea.2014.5502 .

Lin, Zhiyuan, Tim Althoff, and Jure Leskovec. 2018. I'll Be Back: On the Multiple Lives of Users of a Mobile Activity Tracking Application. *Proceedings of the 2018 World Wide Web Conference*. https://doi.org/10.1145/3178876.3186062

Numminen, Emil, Henrik Sällberg, and Shujun Wang. 2022. The Impact of App Revenue Model Choices for App Revenues: A Study of Apps Since Their Initial App Store Launch. *Economic Analysis and Policy* 76: 325–336. https://doi.org/10.1016/j.eap.2022.08.010.

Roma, P., Valenti, F., Perrone, G. (2013). An empirical analysis of revenue drivers in the mobile app market. *In 24th Annual POM Conference. Denver, Colorado, USA, May 3 – 6, 2013 (pp.1-10). https://hdl.handle.net/10447/99369*

Statista. Health and Fitness. https://www.statista.com/outlook/amo/app/health-fitness/worldwide.

World Health Organisation. Obesity and Overweight Factsheet. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.

Yang, Yaobin, and Joerg Koenigstorfer. 2021. Determinants of Fitness App Usage and Moderating Impacts of Education-, Motivation-, and Gamification-Related App Features on Physical Activity Intentions: Cross-sectional Survey Study. *Journal of Medical Internet Research* 23 (July): e26063. https://doi.org/10.2196/26063.

Zhu, Naixin. 2023. Applied Microeconomics of Freemium Pricing Strategies in the Mobile App Market [Doctoral dissertation]. University of California, Berkeley. https://doi.org/10.48550/arXiv.2305.09479.

APPENDIX