

Pattern Discovery in Smartphone Pricing and Hardware Specifications

Victor Urey

Kennesaw State University

Marietta, GA

vurey1@students.kennesaw.edu

Abstract—This document aims to investigate patterns between smartphone hardware and their success on the global market. It explores pricing, hardware specification, brand strategy, and consumer buying trends to paint a broader picture of the smartphone market over the last thirty years. Two data sets are used to: (1) discover how specific hardware features relate to price tiers across leading brands, and (2) which hardware components appear among the most popular phones and how do companies iterate on them in the future. Patterns are found using a combination of clustering and association rules to reveal specification hardware combinations that have catapulted a few brands to international success in the smartphone market.

Index Terms—hardware, clustering, association rules, pricing correlation, consumer trends

I. INTRODUCTION

Within the last three decades the world of smartphone technology has skyrocketed beyond any prediction from the late 1990s. Devices became smaller, more powerful, and catered to consumers' every need. They have applications for work, gaming, socializing, shopping, and now provide quick access to cutting-edge artificial intelligence. This demand for increasingly stronger devices has strained dated hardware systems and has forced companies to innovate in order to develop the next best thing.

Because of this, the smartphone market is now notable for the speed at which new hardware is developed, implemented, and pushed out to consumers. These features are all too common on the market today; bigger cameras, longer battery life, foldable screens, quicker operating systems, fast charging, and brighter displays all advertised as the next big thing in smartphone technology. These differences are what helped set brands apart over the past thirty years, lending to dedicated customers who seek out the next flagship device.

This project aims to discover the patterns between smartphone hardware and consumer buying trends to reveal how hardware improvements have changed the smartphone game of today. How does battery life, screen size, camera specs, or even charging cable change how successful smartphones are? How have these hardware components changed over time, and has the every-day consumer changed their priorities in smartphone hardware over the years? These questions are crucial talking points in this project and by using advanced clustering and association rule mining I hope to uncover meaningful data which brings hardware, previously hidden behind the smartphone screen, to light.

II. DATASET DESCRIPTION

A. Kaggle - Smartphone Specifications Dataset

<https://www.kaggle.com/datasets/devgondaliya007/smartphone-specifications-dataset>

Size: 986 rows, 27 columns

This data set provides a comprehensive view of smartphone models and their specifications on the market today. It includes nearly 1000 different smartphone models and is useful for studying patterns between hardware specifications, pricing, and consumer buying trends. This data was collected across multiple mobile comparison and e-commerce websites to provide a wide array of phone models and their pricing.

The key feature of this data set are the information it provides on phone models, price at release, operating system, NFC compatibility, ram quantity, screen size, resolution / display type, camera resolution, and refresh rate. These are all important groups that may later be used to cluster similar phone models together. There are a few data quality issues with this data set; namely, some rows are missing data values in the refresh rate category, and model naming is not entirely consistent across different brands.

B. Kaggle - World's Best-Selling Phone's Sales

<https://www.kaggle.com/datasets/muhammadroshaanriaz/global-best-selling-phone-sales>

Size: 120 rows, 7 columns

This data set provides the top 120 hottest-selling mobile phones over the past 30 years. It includes data on phone manufacturer, specific model, form factor, release year, and the units sold to date: providing a comprehensive view of phone sales and specification changes across the early 2000s. This dataset includes flip phones, bar phones, and touchscreen models to demonstrate the progression of smartphone technology over the years.

The key features in this data set are the phone manufacturer, model, form factor, release year, units sold, and global sales ranking. These features add on to the previous data set by detailing other phone specifications outside of the phone hardware. One glaring issue, however, is the lack of citation for where the data was recovered. I will need to look into this

in the future and make sure that I can find the original citation before implementing my discovery questions.

III. DISCOVERY QUESTIONS

A. Question 1: How do hardware features correlate with price tiers across popular phone brands?

Hardware is one of the most important categories when choosing which phone to buy. Finding patterns with smartphone hardware can reveal if brands have similar pricing for similar hardware, if brands tends to prioritize one hardware feature over others, and if customers tend to stick with brands that suit their hardware needs. It may even show how hardware improvements have increased the prices of certain phones while others remain largely unaffected.

B. Question 2: What hardware patterns exist in the most popular phone models sold over time across major phone brands?

Not only are hardware features important to the consumer, but they are also important to brands who constantly have to source the resources to produce it. This question may uncover how the year that products released had a large affect on the amount of success that product had in the long-run. It can show if consumer priorities have changed over the years or if brand loyalty keeps consumers coming back despite an objectively inferior product.

IV. FIGURES AND TABLES

# model	# price	# os	# NFC	# ram_gb	# storage_gb	# screen_size_in	# resolution	# display_type	# n
apple iphone 14 pro max	1699999	1205 x16	True	6	6	6.7	1296x2796	Punch Hole	49
huawei mate x	1699999		True	8	8	8.0	2280x2488	Punch Hole	50
samsung galaxy z fold 4	1699999		True	12	12	7.6	1812x2176	Punch Hole	50
huawei mate xs 2	1629999		True	8	8	7.8	2280x2488	Punch Hole	50
samsung Galaxy z fold 4	1549998		True	12	12	7.6	1812x2176	Punch Hole	50

Fig. 1. Sample data table from "Smartphone Specifications Dataset"

# Rank	# Manufactur...	# Model	# Form Factor	# Smartphone...	# Year	# Units Sold ...
1	Nokia	1100	Bar	FALSE	2003	250
2	Nokia	1110	Bar	FALSE	2005	247.5
3	Apple	iPhone 6 & 6 Plus	Touchscreen	TRUE	2014	222.4
4	Nokia	185 Series	Bar	FALSE	2013	200
5	Apple	iPhone 6S & 6S Plus	Touchscreen	TRUE	2015	174.1

Fig. 2. Sample data table from "World's Best-Selling Phone's Sales"

V. PLANNED TECHNIQUES

A. Technique 1: Association Rules (Apriori, FP-Growth)

a) : Relation to Discovery Question 1: Useful for discovering different combinations of specifications that occur within price tiers. It can be used to define how individual specifications and groups of specifications contribute to phone pricing. This provides a means to group phone models into tiers by their contributing specifications and the overall pricing.

b) : Relation to Discovery Question 2: Using association rules is useful for discovering patterns between phone models and the years they were released. It can categorize the highest selling model over time as well as defining which brands consistently win in this category. This question is focused on how the advancement of hardware has contributed to the overall units sold for popular brands like Apple, Samsung, or Vivo which are expected to be consistent top sellers.

B. Technique 2: Clustering (K-Means, Hierarchical, DBSCAN)

a) : Relation to Discovery Question 1: Clustering is useful for grouping phone models into categories that are hardware-specific. These groups could include phone models with the longest battery life, models with the best cameras, cheapest phones, and even the phone models and brands leading the industry. It provides a basis for separating phones by price tiers and how these brands prioritize certain features over others within each of these tiers.

b) : Relation to Discovery Question 2: Since this question is focused on how the hardware of specific phone models has contributed to it being a top seller, clustering by common hardware among top sellers could be very useful. It could show patterns on model lifespan and how each brand chooses to iterate on their previous models. Clustering can also provide some insight into consumer buying trends and how their priorities may have shifted between hardware components.

VI. PRELIMINARY TIMELINE

M2 - Clean data values, join both (and possibly third) data set, explore common models across both, apply basic clustering and association rules mining techniques.

M3 - Complete advanced pattern mining, convert clustering and association rules results into visuals, combine visual aids into one graph.

M4 - Compile all findings, finalize visual reports, report issues throughout the process, contextualize and document usefulness of data, presentation preparation.

A. Anticipated Challenges

With any form of pattern mining there are bound to be challenges. Firstly, combining two data sets, each with their own naming conventions, is easier said than done. Model names and some data values are not documented the same way across data sets so standardizing all brand and model formats before conducting any pattern mining is crucial. Secondly, using cluster methods may yield phone models grouped by brand, not very useful for comparing similar phones across shared hardware or pricing. Thirdly, grouping values like battery, pricing, and megapixels into discernable categories like small, medium, and large could benefit association rules mining. This way there are no arbitrary groups being formed and the resulting data can be better understood. Lastly, tweaking the discovery algorithms to discovery shared hardware

specifications across different brands seems to be the most difficult part. Lots of trial and error may be used and I may find a more suited algorithm or pattern discovery tool down the line.

VII. GITHUB REPOSITORY

A GitHub Repository has been set up and linked below:

<https://github.com/victorurey/cs4412-smartphone-data-project>

This repository includes a README document that outlines the purpose for the project, links to the datasets, and developer information. There are folders for data as well as the cleaned data sets and source code when those become available.

REFERENCES

- [1] CodewithDev, “Smartphone Specifications Dataset,” Kaggle.com, 2026. <https://www.kaggle.com/datasets/devgondaliya007/smartphone-specifications-dataset> (accessed Feb. 06, 2026).
- [2] M. J. Zaki and W. Meira, Jr., Data Mining and Machine Learning: Fundamental Concepts and Algorithms. Cambridge, U.K.: Cambridge University Press, 2020.
- [3] Muhammad Roshan Riaz, “World’s Best-Selling Phone’s Sales,” Kaggle.com, 2018. <https://www.kaggle.com/datasets/muhammadroshaanriaz/global-best-selling-phone-sales>