|  |
| --- |
| Battery Power smartphones\_  Individual Report Joris Cornel |

Joris Cornel

18 december 2023

Table of contents

[1 Introduction & objective 5](#_Toc153815295)

[1.1 Introduction 5](#_Toc153815296)

[1.2 Objective 5](#_Toc153815297)

[2 Data exploration 6](#_Toc153815298)

[2.1 Dataset introduction 6](#_Toc153815299)

[2.2 Missing Data 7](#_Toc153815300)

[2.3 Data Quality & Impossible values 7](#_Toc153815301)

[2.4 Descriptive Statistics 8](#_Toc153815302)

[2.4.1 output 8](#_Toc153815303)

[2.4.2 First impressions 8](#_Toc153815304)

[3 Modelling 10](#_Toc153815305)

[3.1 Correlation Analysis 10](#_Toc153815306)

[3.2 Target Variable Analysis 11](#_Toc153815307)

[4 Observations 12](#_Toc153815308)

[4.1.1 Variable observations 12](#_Toc153815309)

[4.1.2 Modelling observations 13](#_Toc153815310)

[5 Conclusion and recommendations 14](#_Toc153815311)

[5.1 Conclusion 14](#_Toc153815312)

[5.2 Recommendations 16](#_Toc153815313)

[5.2.1 For Smartphone Manufacturers: 16](#_Toc153815314)

[5.2.2 For Users: 17](#_Toc153815315)

[5.2.3 Modelling recommendations: 17](#_Toc153815316)

Project introduction

Welcome to my data science portfolio project, the end project of my Minor in Data Driven Decision Making. This project reflects my journey in applying data science skills to a real-world scenario.

**Project Overview:**

This portfolio aims to showcase my data science skills by presenting a project that follows the CRISP-DM model. Beyond algorithms and code, data science is about obtaining insights for informed decision-making. This portfolio is an example of data-driven problem-solving.

**Guiding Questions:**

As I delve into this project, I aim to answer fundamental questions: What problem am I solving? Who benefits from this analysis? Is there room for innovation in existing modelling techniques?

**The CRISP-DM Framework:**

Following the CRISP-DM framework, my approach involves the following steps: Business understanding, data understanding, data preparation, modelling, evaluation, and recommendations for employment.



Preface

In our fast-paced world, smartphones have become indispensable tools for communication, work, and entertainment. A critical aspect of the user experience is the endurance of a phone's battery life. This report delves into the complexities of smartphones to uncover the key factors influencing battery power. The insights are intended not only for manufacturers but also for users trying to optimize their experience by gaining an insight into extending the battery life.

Utilizing statistical analysis and modeling techniques, the report aims to identify the most impactful variables affecting smartphone battery life. The diverse set of variables examined includes battery capacity, processor attributes, camera specifications, memory features, screen dimensions, connectivity options, and physical characteristics.

The report's objectives are threefold:

1. **Variable Importance Assessment:** Employ statistical methods and modeling to find the significance of each variable concerning battery life.
2. **Impact Quantification:** Quantify and rank the influence of individual variables, emphasizing their contribution to the overall battery life of smartphones.
3. **Insights for Optimization:** Provide recommendations based on findings to assist both smartphone manufacturers and users in optimizing relevant features for enhanced battery life.

The ultimate goal is to empower stakeholders with the knowledge required to make informed decisions, while finding improvements in the overall efficiency of mobile devices.

# Introduction & objective

## Introduction

In our evolving society, smartphones have evolved into essential instruments for communication, work, and entertainment. At the core of the user experience lies the endurance of a phone's battery life, an important factor that influences the convenience and overall functionality of the device.

This report delves into the functionalities of smartphones to find the key factors influencing battery life. This knowledge is crucial not only for phone manufacturers but also for users seeking devices that align with their needs.

## Objective

This report employs advanced statistical analysis and modeling techniques to identify the most impactful variables affecting smartphone battery life. For this, I used a diverse set of variables, ranging from battery capacity (battery\_power) to processor attributes (clock\_speed and n\_cores), camera specifications (fc and pc), memory features (ram and int\_memory), screen dimensions (sc\_h and sc\_w), connectivity options (blue, dual\_sim, four\_g, three\_g, touch\_screen, wifi), and physical characteristics (m\_dep and mobile\_wt).

As named in the preface, we can set the following objectives:

1. Variable Importance Assessment: Utilize statistical methods and modeling to assess the significance of each variable concerning battery life.
2. Impact Quantification: Quantify and rank the influence of individual variables, emphasizing their contribution to the overall battery life of smartphones.
3. Insights for Optimization: Provide insights and recommendations based on findings to assist both smartphone manufacturers and users in optimizing relevant features for enhanced battery life.

By achieving these objectives, this report tries to provide a better understanding of the factors impacting the smartphone battery power.

# Data exploration

## Dataset introduction

The dataset under used is a comprehensive collection of mobile device specifications obtained from Kaggle, a platform for data science and machine learning datasets. Each row in the dataset represents a real-life mobile device, and the accompanying variables provide detailed insights into various aspects of these devices. The dataset encompasses a diverse range of features, which are essential attributes that contribute to the functionality and performance of mobile phones.

Below is an overview of the variables included in the dataset:

* battery\_power: Total energy a battery can store in one time, measured in mAh.
* blue: Presence or absence of Bluetooth connectivity (binary, 1 for true, 0 for false).
* clock\_speed: Speed at which the microprocessor executes instructions.
* dual\_sim: Availability of dual SIM support (binary, 1 for true, 0 for false).
* fc: Front camera resolution in mega pixels.
* four\_g: Presence or absence of 4G connectivity (binary, 1 for true, 0 for false).
* int\_memory: Internal memory capacity in gigabytes.
* m\_dep: Mobile depth in centimeters.
* mobile\_wt: Weight of the mobile phone.
* n\_cores: Number of cores in the processor.
* pc: Primary camera resolution in mega pixels.
* px\_height: Pixel resolution height.
* px\_width: Pixel resolution width.
* ram: Random access memory capacity in megabytes.
* sc\_h: Screen height of the mobile in centimeters.
* sc\_w: Screen width of the mobile in centimeters.
* talk\_time: Longest time that a single battery charge will last when in use.
* three\_g: Presence or absence of 3G connectivity (binary, 1 for true, 0 for false).
* touch\_screen: Presence or absence of a touch screen (binary, 1 for true, 0 for false).
* wifi: Presence or absence of Wi-Fi connectivity (binary, 1 for true, 0 for false).
* price\_range: The target variable indicating the price range of the mobile phone.

The dataset is publicly available on Kaggle via the following URL: <https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification/data>

For the purpose of our analysis, the dependent variable of interest is battery\_power, representing the total energy storage capacity of the mobile device's battery. This variable will serve as an important datapoint for exploring the impact of other device specifications on battery life.

## Missing Data

Upon initial examination of the dataset, it is observed that there are no missing values in any of the variables the dataset is complete.

The absence of missing data simplifies the data exploration process, eliminating the need for imputation techniques. This completeness enhances the reliability of the subsequent analyses, ensuring that the insights derived are based on an actual set of observations.

## Data Quality & Impossible values

In our evaluation of the dataset's integrity, we specifically addressed the presence of impossible values. These values are data points that deviate from the plausible range for each variable, undermining the reliability of the dataset. Plausible ranges were established based on the inherent characteristics of the features, ensuring that values remained within realistic boundaries.

After an examination, we are pleased to report that no impossible values were identified within the dataset. Consequently, no data points required removal due to impossible values.

## Descriptive Statistics

### output

|  |
| --- |
|  |
| Variable | **Count** | **Mean** | **Std** | **Min** | **25%** | **50%** | **75%** | **Max** |
| battery\_power | 2000 | 1238.5 | 439.4 | 501.0 | 851.8 | 1226.0 | 1615.2 | 1998.0 |
| blue | 2000 | 0.5 | 0.5 | 0.0 | 0.0 | 0.0 | 1.0 | 1.0 |
| clock\_speed | 2000 | 1.5 | 0.8 | 0.5 | 0.7 | 1.5 | 2.2 | 3.0 |
| dual\_sim | 2000 | 0.5 | 0.5 | 0.0 | 0.0 | 1.0 | 1.0 | 1.0 |
| fc | 2000 | 4.3 | 4.3 | 0.0 | 1.0 | 3.0 | 7.0 | 19.0 |
| four\_g | 2000 | 0.5 | 0.5 | 0.0 | 0.0 | 1.0 | 1.0 | 1.0 |
| int\_memory | 2000 | 32.0 | 18.1 | 2.0 | 16.0 | 32.0 | 48.0 | 64.0 |
| m\_dep | 2000 | 0.5 | 0.3 | 0.1 | 0.2 | 0.5 | 0.8 | 1.0 |
| mobile\_wt | 2000 | 140.2 | 35.4 | 80.0 | 109.0 | 141.0 | 170.0 | 200.0 |
| n\_cores | 2000 | 4.5 | 2.3 | 1.0 | 3.0 | 4.0 | 7.0 | 8.0 |
| pc | 2000 | 9.9 | 6.1 | 0.0 | 5.0 | 10.0 | 15.0 | 20.0 |
| px\_height | 2000 | 645.1 | 443.8 | 0.0 | 282.8 | 564.0 | 947.2 | 1960.0 |
| px\_width | 2000 | 1251.5 | 432.2 | 500.0 | 874.8 | 1247.0 | 1633.0 | 1998.0 |
| ram | 2000 | 2124.2 | 1084.7 | 256.0 | 1207.5 | 2146.5 | 3064.5 | 3998.0 |
| sc\_h | 2000 | 12.3 | 4.2 | 5.0 | 9.0 | 12.0 | 16.0 | 19.0 |
| sc\_w | 2000 | 5.8 | 4.4 | 0.0 | 2.0 | 5.0 | 9.0 | 18.0 |
| talk\_time | 2000 | 11.0 | 5.5 | 2.0 | 6.0 | 11.0 | 16.0 | 20.0 |
| three\_g | 2000 | 0.8 | 0.4 | 0.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| touch\_screen | 2000 | 0.5 | 0.5 | 0.0 | 0.0 | 1.0 | 1.0 | 1.0 |
| wifi | 2000 | 0.5 | 0.5 | 0.0 | 0.0 | 1.0 | 1.0 | 1.0 |
| price\_range | 2000 | 1.5 | 1.1 | 0.0 | 0.8 | 1.5 | 2.2 | 3.0 |

### First impressions

**Battery Power:**

* The average battery power of the mobile phones in the dataset is 1238.5 mAh, with a standard deviation of 439.4 mAh.
* The battery power ranges from a minimum of 501.0 mAh to a maximum of 1998.0 mAh.
* The battery power appears to be quite distributed, which could mean that the variable is heavily impacted by other variables.

**Connectivity:**

* Around 50% of the phones support features like Bluetooth (blue), dual SIM (dual\_sim), 4G (four\_g), and 3G (three\_g).
* WiFi (wifi) and touch screen capabilities (touch\_screen) are evenly distributed among the devices.

**Processor and Memory:**

* The average number of processor cores (n\_cores) is approximately 4.5, with a standard deviation of 2.3.
* The internal memory (int\_memory) ranges from 2.0 GB to 64.0 GB, with an average of 32.0 GB.

**Display and Camera:**

* The pixel height (px\_height) and pixel width (px\_width) suggest variations in display size, ranging from 0 to 1960 pixels in height and 500 to 1998 pixels in width.
* The front and rear camera resolutions (fc and pc) vary, with an average of 4.3 mega pixels for both.

**Design and Build:**

* The mobile weight (mobile\_wt) ranges from 80.0 grams to 200.0 grams, with an average of 140.2 grams.
* The thickness of the phone (m\_dep) varies, with an average of 0.5 cm.

**Performance:**

* RAM (ram) shows significant variability, ranging from 256.0 MB to 3998.0 MB, with an average of 2124.2 MB.
* The talk time (talk\_time) ranges from 2.0 to 20.0 hours.

The analysis shows a dataset with a diverse range of mobile devices, each with variations in components. This diversity sets the stage for a comprehensive analysis, offering an opportunity to explore and understand the influence of these specifications on the battery power of the devices.

# Modelling

## Correlation Analysis

In this section, I explored possible linear relationships between battery\_power and other variables within the dataset using a correlation matrix. The correlation matrix provides insights into the strength of these relationships.

A blue and red chart with white text

Description automatically generated

This correlation matrix was computed to explore the linear relationships between battery\_power and the other variables in the dataset. The results indicate low correlation coefficients, with most values being around 0.0. Notably, only the variable talk\_time exhibits a small correlation of 0.1, which is still too low to conclude a lineair relationship.

This pattern suggests that individual variables may not have a substantial linear impact on battery\_power. However, it is important to note that these findings do not rule out the possibility of a collective or nonlinear influence from combinations of variables. Multiple variables could collectively contribute to the battery\_power.

Further analysis, such as regression modeling, may show patterns that are not visible with the use of a correlation matrix.

## Target Variable Analysis

A screenshot of a computer screen

Description automatically generatedTo understand the relationship between battery\_power and the various smartphone specifications, a multiple linear regression model was made. The results are placed below:

**Limited Predictive Power:**

The regression model had an R-squared value of 0.012, indicating that the selected variables collectively explain only a very small part of the variation in battery\_power.

**Variable Coefficients:**

None of the individual variables showed a statistically significant impact on battery\_power.

**Conclusion:**

The analysis shows that in the scope of using this model, the selected variables do not influence battery\_power much.

# Observations

### Variable observations

**Battery Power Distribution**

One observation is the distribution of battery power among the mobile phones in the dataset. With an average capacity of 1238.5 mAh and a standard deviation of 439.4 mAh, the range from 501.0 mAh to 1998.0 mAh highlights significant variability. This suggests that battery power is likely influenced by a combination of factors.

**Connectivity Trends**

Connectivity features such as Bluetooth, dual SIM support, 4G, and 3G exhibit a balanced distribution among the devices. Approximately 50% of the phones support these features. Additionally, Wi-Fi and touch screen capabilities are evenly distributed, indicating widespread integration of these functionalities in smartphones.

**Processor and Memory Variability**

The processor-related variables, including the average number of cores (n\_cores), show an average of 4.5 cores with a standard deviation of 2.3. The internal memory (int\_memory) differs from 2.0 GB to 64.0 GB, emphasizing the diverse memory capacities across the dataset.

**Display and Camera Characteristics**

Pixel height (px\_height) and pixel width (px\_width) exhibit considerable variations, suggesting a diverse range of display sizes among the mobile devices. Front and rear camera resolutions (fc and pc) also show variability, with an average resolution of 4.3 mega pixels for both cameras.

**Design and Build Insights**

The physical characteristics of mobile phones, including weight (mobile\_wt) and thickness (m\_dep), differ a lot. Mobile weight ranges from 80.0 grams to 200.0 grams, while thickness varies with an average of 0.5 cm.

**Performance Metrics**

RAM (ram) shows variability, ranging from 256.0 MB to 3998.0 MB, indicating diverse memory capacities. Talk time (talk\_time) spans from 2.0 to 20.0 hours, showing variability in battery endurance across devices, which could also explain the variability in the battery power among the mobile devices.

### Modelling observations

**Correlation Analysis Findings**

The correlation matrix showed generally low correlation coefficients between battery\_power and other variables. Notably, only the variable talk\_time exhibited a small correlation of 0.1. This suggests that individual variables may not have a substantial linear impact on battery\_power, pointing towards the need for further investigation into possible relations.

**Regression Model Insights**

The multiple linear regression model, while providing insights into variable coefficients, demonstrated small predictive power with an R-squared value of 0.012. None of the individual variables showed a statistically significant impact on battery\_power. This suggests that individually, the variables do not explain the variable battery\_power.

# Conclusion and recommendations

## Conclusion

In the dynamic world of smartphones, the endurance of battery life remains a critical aspect of user satisfaction. This comprehensive analysis aimed to find the factors influencing smartphone battery power, utilizing advanced statistical methods and modeling techniques. I found the following observations in this report. Its important to mention, that while individual variables may not show a strong linear relationship, their collective impact could still be significant.

* **Variable Significance:**

The analysis revealed a diverse set of variables, containing variables concerning battery capacity, processor attributes, camera specifications, memory features, screen dimensions, connectivity options, and physical characteristics.

Despite the initial expectation of linear relationships, the correlation matrix indicated low coefficients, suggesting that individual variables might not have a substantial impact on battery power.

* **Regression Model Insights:**

The multiple linear regression model demonstrated limited predictive power. The selected variables collectively explained only a very small portion of the variation in battery power.

None of the individual variables exhibited a statistically significant impact on battery power, indicating that, within the scope of the model, the chosen variables did not strongly influence battery capacity.

* **Observations and Trends:**

The dataset showcased a distribution of battery power among mobile devices, emphasizing the impact of various factors on battery capacity.

Connectivity features, processor attributes, memory capacities, display and camera specifications, as well as physical characteristics, exhibited significant variability across smartphones.

The analysis suggests that battery power is likely influenced by a combination of factors, with individual variables contributing collectively to the overall endurance of a device.

* **Implications for Stakeholders:**
  + Smartphone Manufacturers: Insights from this report emphasize the need for a different approach to battery optimization, considering a combination of factors rather than individual variables. Manufacturers may benefit from exploring innovative solutions that enhance overall device efficiency. There are also opportunities for improving certain functionalities, without the need for optimizing the battery power.
  + Users: Understanding the nature of battery power allows users to make informed decisions when selecting smartphones. Optimization strategies can include managing connectivity features, choosing devices with suitable processor and memory capacities, and considering physical characteristics.
* Future Directions:

Further exploration may involve other techniques to show intricate relationships that may not be apparent in this analysis.

Expanding the current data set, along with expanding the number of variables used in this analysis, might change the conclusion we can take from the techniques used.

## Recommendations

Based on the findings of this analysis, the following recommendations are provided for both smartphone manufacturers and users:

### For Smartphone Manufacturers:

**Integrated Approach to Optimization:**

Manufacturers should adopt an integrated approach to battery optimization, considering a combination of factors rather than focusing solely on individual variables. This may involve collaborative efforts between hardware and software teams to enhance overall device efficiency.

**Innovation in Battery Technology:**

Invest in research and development of innovative battery technologies to decrease energy consumption and improve overall battery performance. Improvements in battery technology can significantly impact the endurance of mobile devices.

**Efficiency in Processor and Memory Management:**

Explore strategies to enhance the efficiency of processors and memory management systems. Optimizing the coordination between these components may lead to improved energy utilization and prolonged battery life.

**User-Centric Features:**

Consider developing features that allow users to customize and optimize battery usage based on their preferences. Providing users with more control over connectivity options, background processes, and display settings can contribute to a more personalized and efficient experience.

### For Smartphone Users:

**Optimized Connectivity Usage:**

Users can optimize battery usage by managing connectivity features such as Bluetooth, 4G, and Wi-Fi. Turning off unnecessary connectivity options when not in use can contribute to extending battery life.

**Consider Processor and Memory Needs:**

When selecting a smartphone, users should consider their specific needs for processor speed and memory capacity. Choosing a device with specifications that align with individual usage patterns can lead to better energy efficiency.

**Regular Software Updates:**

Keep devices up to date with the latest software updates provided by manufacturers. Software updates often include optimizations and bug fixes that can contribute to better overall system efficiency.

### Modelling recommendations:

**Advanced Data Analysis Techniques:**

Explore more advanced data analysis techniques, such as machine learning algorithms, to uncover intricate relationships that may not be shown through traditional statistical methods. These techniques can provide a more nuanced understanding of the factors influencing battery power.

**Expanded Dataset and Variables:**

Consider expanding the dataset with additional variables and a larger sample size. This can lead to a more comprehensive analysis, potentially revealing new insights and strengthening the conclusions drawn from the current study. For example, the actual battery life might be a better variable to predict than the battery power capacity, since that variable is directly influenced by the other variables.

