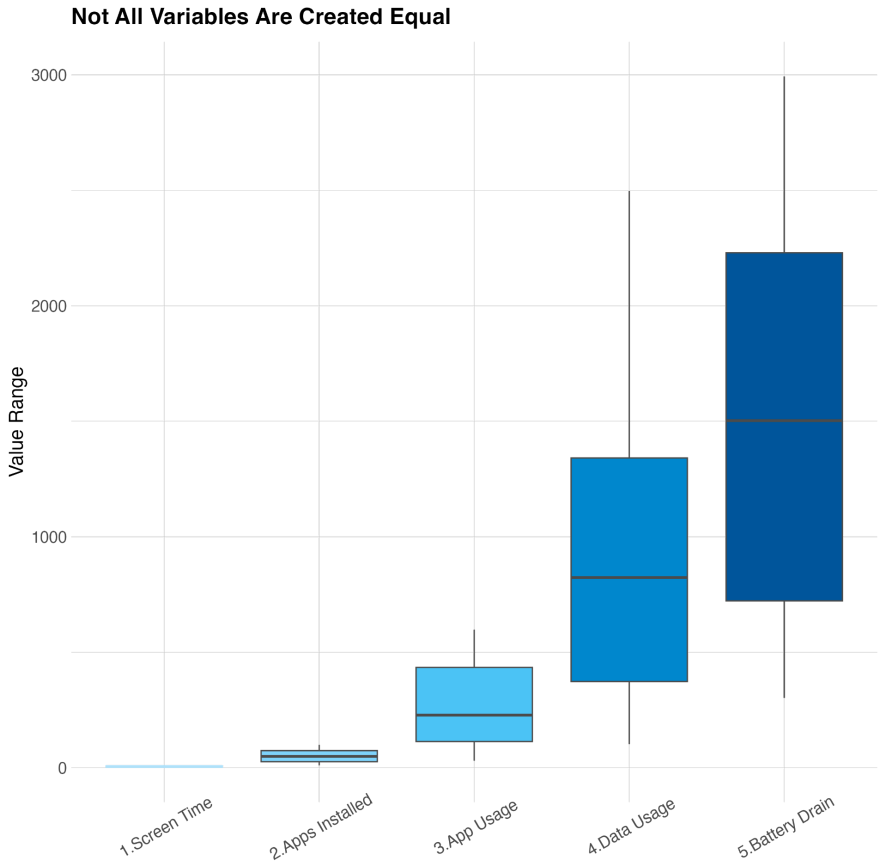
SMM635 Mid-Term Project

Konstantinos Bakogiannis
Valentina Donado Agudelo
Blanca Fariña Alcedo
Marinos Kampisioulis

Venus Inc. is a recently established startup, dealing with womens' health and well-being, based in Tokyo, Japan. The company has been interested in entering the digital goods market by developing an app, named **VenuSync**, which helps women accurately track their menstrual cycle. To minimise the expenditure (and maximise the profits) of this endeavour, the company aims to develop the app for only one operating system (Android vs. iOS). To that end, we gathered data pertaining to Mobile Device User Behaviour, to help them make this decision. Specifically, how certain metrics, including Screen Time, Data Usage and Number of Apps installed varies across users of different ages, genders and nationalities. Through our exploratory data analysis VenuSync will be able to secure the largest (and most lucrative) potential user base and succeed in this incredibly competitive market.

Plot 1 to Visualise an Intrinsic Weaknesses



This boxplot displays five key variables - Screen Time, Apps Installed, App Usage, Data Usage, and Battery Drain - chosen thanks to their primarily continuous nature and effective representation of phone user behaviour. The boxplot highlights intrinsic weaknesses in the dataset by showing the variability and spread of each variable, which helps in identifying potential data quality issues. Our selected visual form is particularly suited for this analysis, as each box represents the range of values for each variable, allowing us to quickly observe and compare differences across them.

The line in the middle of each box shows the median, the measure of the centre, or location, of the distribution which helps us identify the “typical” value for each variable. The top and bottom of the boxes represent the upper and lower quartiles. The box height or the distance between these two values represents the interquartile range, a measure of the spread of the distribution, which is the range where the central 50% of values fall. The interquartile range of

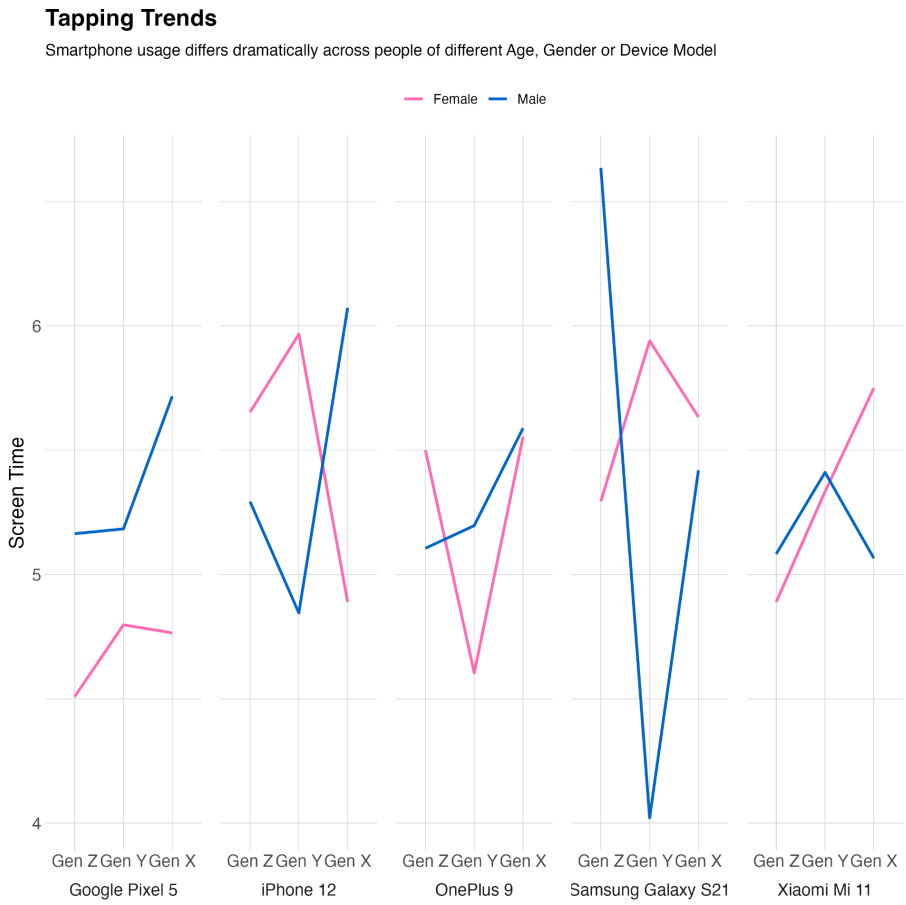
Screen Time, Apps Installed and App Usage is small, indicating that the middle data are tightly packed around the median. Data Usage, and Battery Drain have a large interquartile range, suggesting that the middle data spread out far from the median. The lines extending from the boxes encode the adjacent values and they show the approximate range of the data, typically 1.5 times the interquartile range. Any data points outside this range would be considered outliers, though none are visible in this plot.

The boxplot reveals substantial differences in variability among the five variables. The variability in each box's height shows that some variables are more consistent across users than others. Battery Drain and Data Usage have much larger interquartile ranges compared to Screen Time and Apps Installed. A larger interquartile range means more variation in that variable. High variability in some variables, suggests that user behaviour might not be consistent, and some variables may capture user actions or device performance with more detail, making it difficult to directly compare these variables. For instance, Apps Installed has a relatively compact interquartile range, meaning most users have a similar number of apps installed, while Battery Drain has a much wider interquartile range, indicating more diversity in user's battery consumption. This variability represents an intrinsic weakness in terms of stability, as it suggests that user experiences or device efficiency are not consistent across the dataset.

Furthermore, the range from the lower to the upper adjacent value also varies greatly among variables. Battery Drain and Data Usage have much larger ranges than Screen Time and Apps Installed, which means that some variables cover a much wider scale. Those Differences in range create scaling issues when these variables are analysed together without adjustments. Variables with large ranges, like Battery Drain, could dominate or skew our statistical analysis, potentially leading to biassed results.

In summary, the boxplot highlights variability and range differences as key intrinsic weaknesses in the data, revealing areas where adjustments may be needed for balanced analysis and meaningful insights. These quality issues are essential to address in order for VenusSync to ensure accurate, reliable conclusions.

Plot 2 to visualise Novelty



This plot visualises screen time across different generations, genders and mobile devices. It allows us to understand which generation and device types our client should prioritise in order to have high engagement in their app. Each line in the graph gives insights into how different groups of generation use their devices. To facilitate rapid comparison between genders, the plot presents two types of coloured lines, the blue line represents males, while the pink one represents females. In terms of usage patterns across different types of mobile devices, the x-axis is divided into five different types of models (Google Pixel 5, iPhone 12, OnePlus 9, Samsung Galaxy S21, Xiaomi Mi 11). For each of these, screen time variations across different generations and genders is displayed. On the other hand, the y-axis illustrates average daily screen time in hours.

Overall, the data shows significant variability in time usage across generations, device type and gender. By examining these, some trends have been identified. Firstly, when comparing

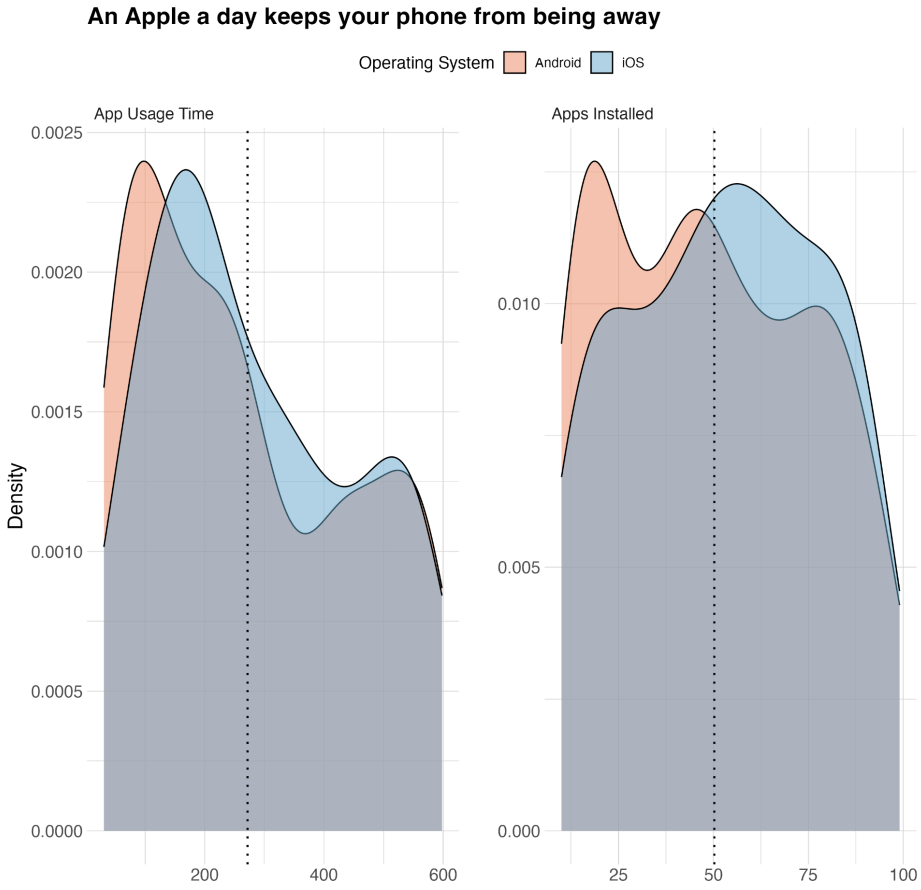
Android to iOS devices, the plot indicates that for Android, on average, older women (Gen X) have a higher screen time than the youngest generation. On the other hand, Gen Z women have higher screen time in iOS compared to older generations. This reflects different preferences or usage habits across these gender groups.

Another valuable finding that is exhibited in this plot is that regardless of gender, Samsung galaxy S21 presents the highest screen usage among Android consumers. This suggests that individuals using this model device are more engaged with their phones, perhaps due to its futures or user experience.

Further, if gender is compared, it can be concluded that women's behaviour is more predictable than men's. Women show more patterns and less variation in screen time hours across all devices and generations. Contrarily, men show a more significant variability. This could be interpreted as men's engagement being influenced by device features, or other external factors that women either do not consider or are less affected by.

To conclude, these novelty findings can strongly empower VenuSync when making their decision. If they decide to prioritise iOS, their focus would be toward higher engagement among Gen Z women, as they show a higher screen time on this device model. If they decide to go the other route, focusing on Samsung Galaxy S21 may enable them to capture another meaningful engaged segment, as this device shows highest usage across individuals that own Android. The findings further suggest that women have more consistent usage patterns with less variability, showing predictability that benefits the company's engagement strategies.

Plot 3 to Visualise the Business Analytics Implications



In order to help VenusSync decide on what OS to launch their app to better target their audience, we explored the comparison between iOS and Android with density estimation that displays the concentration and spread of app usage time and the number of apps installed for both systems filtering only female users. To identify similarities and differences in user behaviour across platforms we use the overlapping density plots which reveal where the distributions differ between the two OS types (Orange: Android, Blue: iOS) and include a dotted line to display the mean of each variable to visualise which OS system has greater presence above average.

At a first glance we notice two key results. First, Android users have shorter app usage time than iOS users, highlighting differences in user behaviour and engagement levels across OS

systems. We can imply that iOS female users engage with apps over longer sessions or have a more varied usage behaviour, whereas Android female users may engage in shorter but possibly more frequent sessions. Secondly, iOS users have a wider range of apps installed, while Android have higher concentration around certain values. For the company this is vital because having more apps installed forecasts a higher probability of users downloading their app. These results imply differences in app adoption behaviour. Android female users tend to install less apps which reveals a more selective approach to their app choices, while iOS female users have a wider range of apps installed showing curious behaviour to try out different or new apps.

These insights create potential avenues for further data exploration. Knowing VenusSync is currently developing an app, we can encourage them to **create platform specific features** based on differences in app usage and installation patterns. If our client decides to launch in the iOS Store, they can focus on prioritising content and features that promote longer engagement such as interactive elements that foster continuous browsing since these users have more apps installed and spend more time on them. Conversely, for Android, where users might be more selective with apps, they can focus on user performance and experience by implementing quick and efficient interactions in the app like notifications for quick check-ins or updates. Additionally, they can keep track of app usage metrics to increase engagement and retention. In the case where there is a consistent decrease in average usage time for users, our client can easily see the declining interest in the app prompting an analysis of content relevance or further user experience.

To further explore VenusSync's business implications of our EDA, we can recommend **personalised marketing** based on which OS system they decide to use. If they decide on launching with an iOS system, it is worth exploring campaigns with more aggressive time-intensive features such as in-app events, promotions, or push notifications since this type of users tend to have higher engagement. On the other hand, for Android users, whose usage time is shorter, the company can focus on retention marketing strategies like personalised recommendations. Loyalty programs, or exclusive content that keep users coming back in order to increase engagement. Overall, we would strongly recommend that the app gets designed for the iOS, because based on our analysis, this would be the most lucrative approach in the scenario that simultaneous development for both operating systems is deemed unaffordable.