Magic Mirror: Wehkamp's AR app
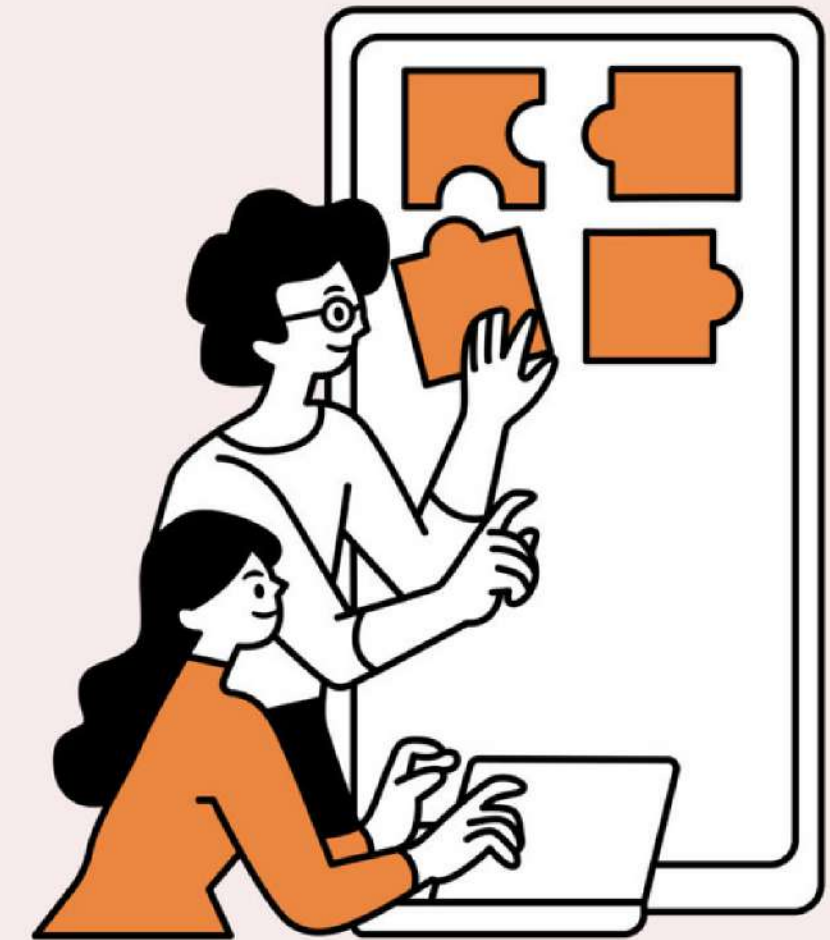
Importance of the tool

7%
of Wehkamp's revenue

14%
of Wehkamp's overall sessions and screen views

24%
of Wehkamp's users are also users of the beauty_ar feature

1.16 visits per user

Most most users visited the tool one time, with some making more than one visit.

High amount of returning customers due to

Lack of marketing efforts

Without heavy marketing, visitors may be less likely to feel like they need to "check out" other businesses in the area (new visitors only account for 0.65%) They may be more likely to return to a business they've had a positive experience with rather than trying something new.

Customer Loyalty & Satisfaction

The high percentage of returning visitors could be attributed to customer satisfaction and/or customer loyalty. When visitors are happy with a service, they are more likely to return and utilise it.

**1/4** Wehkamp's users access the beauty_ar tool.

1/10 of Wehkamp's users access the beauty_ar tool without the use of ads

Advertising

is the main tool in attracting visitors

The use of ads produces a 15% increase in beauty_ar's traffic, which is a significant amount. Therefore, Wehkamp puts effort in attracting visitors to the beauty_Ar tool through advertising.

1.88 Groningen

1.87 Almere

Zwolle **1.97**
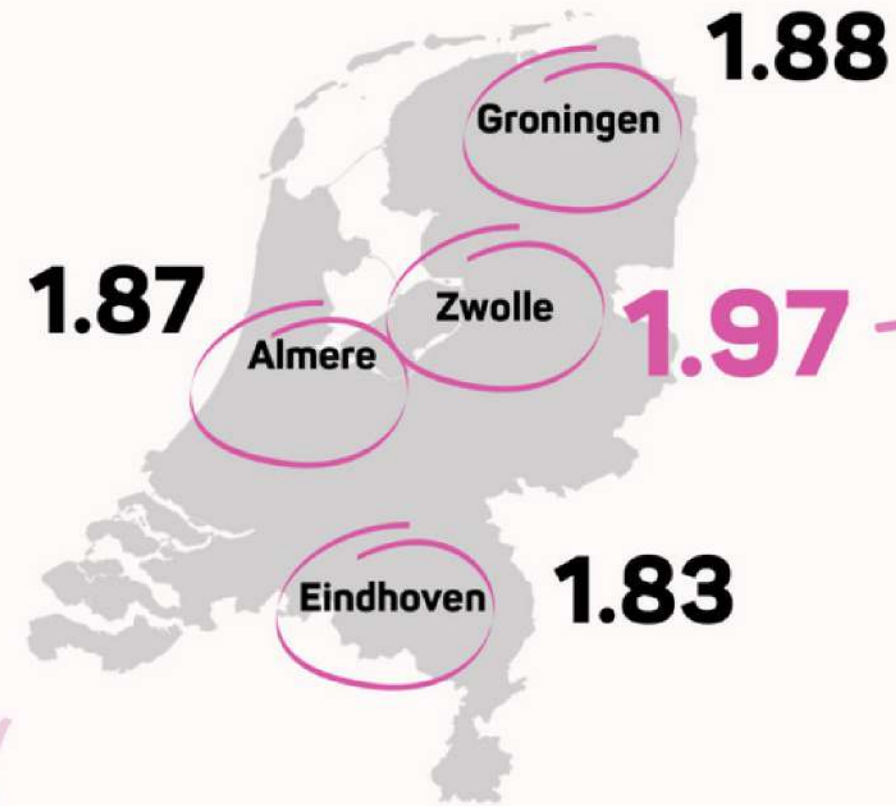
Eindhoven **1.83**

Sessions per user

The city with the highest number of sessions per user is Zwolle, with each user accounting for almost 2 sessions (1.97). Cities such as Groningen (1.88), Eindhoven (1.87) and Almere (1.83) also have a high number of sessions per user.

The rankings differ significantly regarding the cities which account for the highest and lowest number of sessions per user.

The only city with users having a similar number of sessions for both beauty_ar and for all users is Tillburg.

Zwolle stands out because...

- Users in this area account for the highest number of Wehkamp sessions
- They are responsible for the lowest number of beauty_ar sessions
- They account for the highest revenue per person when using this feature.

This could indicate that users in Zwolle are satisfied with the beauty_ar feature and are convinced to make a purchase from less sessions.

The most effective medium for the beauty_ar tool...

By computing profit per user, sessions per user and profit per session, we conclude that the medium "share" is the most effective since it produces about 25 euros of profit per user, and 22 euros of profit per session.

Desktop's Appearance

Despite the app only being designed for touch screen devices (phone or tablet), the Google Play Store can be accessed through every computer through an emulator, making the app available in Mac and PC desktop. The ability to mirror apps from a device to a computer enables a new category in the app report called "desktop", which contains results different from the mobile or tablet category.

beauty_ar VS all_users

Higher peaks in conversion rate

The higher peaks of the conversion rate when AR technology is included could be attributed to the convenience of trying on beauty products at home or on the go, the interactivity involved, offering an entertaining interface to engage with customers, realistic simulations that show each customer exactly what they will receive, and in general, a greater understanding of the customer experience.

Which one is more realistic?

AR users account for

24% of total users   **15%** of the sessions   ≫ Not the most representative group of all

The **popularity and accessibility** of this technology still needs to be questioned:

- Equipment is limited and needs improvement
- There are issues with the computing power and battery life needed to render these environments

> Implementing AR is a slow process and it still hasn't reached mass adoption (older demographics might not be as drawn to it), therefore it is not the most reliable conversion rate when looking at the overall success of the makeup category of Wehkamp.

Tablet / Mobile

Tablet	Stage	Mobile
681 100%	Total visitors	14,109 100%
591 86.78%	Product view	10,599 75.12%
262 38.47%	Item in Cart	4936 34.98%
119 17.47%	Check out	2167 15.36%
93 13.66%	Return	1683 11.93%
3.81%	DROP OFF	3.43 %

For which brand is the beauty_ar tool most important for its sales?

When discussing sales it would be most beneficial to look at the buy-to-detail rate, which is the ratio between product detail views to buying a product.

We propose that if there is a drastic difference between the BTD values of the two user types in favor of the AR users, then it is logical to deduce that the AR tool contributed to some extent to the purchase decision. The biggest discrepancy between the values of this metric is observed in the BOSS brand, where we see that Mirror App users have more than 2x the BTD rate as compared to the BTD rate of all users. (140.21% vs. 61.56%, respectively)

Strange numbers in the brand BOSS

> The first strange finding is the gap between the number of users who added a product to cart in relation to the number of cart removals. We hypothesized that this is because some users added these products to their cart in the previous period and have removed it from their cart in this period.

> Next we see that the difference between product checkouts and products added to cart is also large. We again hypothesized that this is because those users had already chosen products to buy which they've added to their cart and have just now decided to buy it.

> The add to cart is not always equal to remove from cart + checkout, because the add to cart could have been registered in a past period.

Wehkamp is piloting The Magic Mirror App.
After analyzing its performance we found it:

Generated Additional Revenue[1] & Transactions[2]

Promoted Repeat Purchases[3]

Increased Customer Value in Beauty[5]

€155,035
3393 extra items

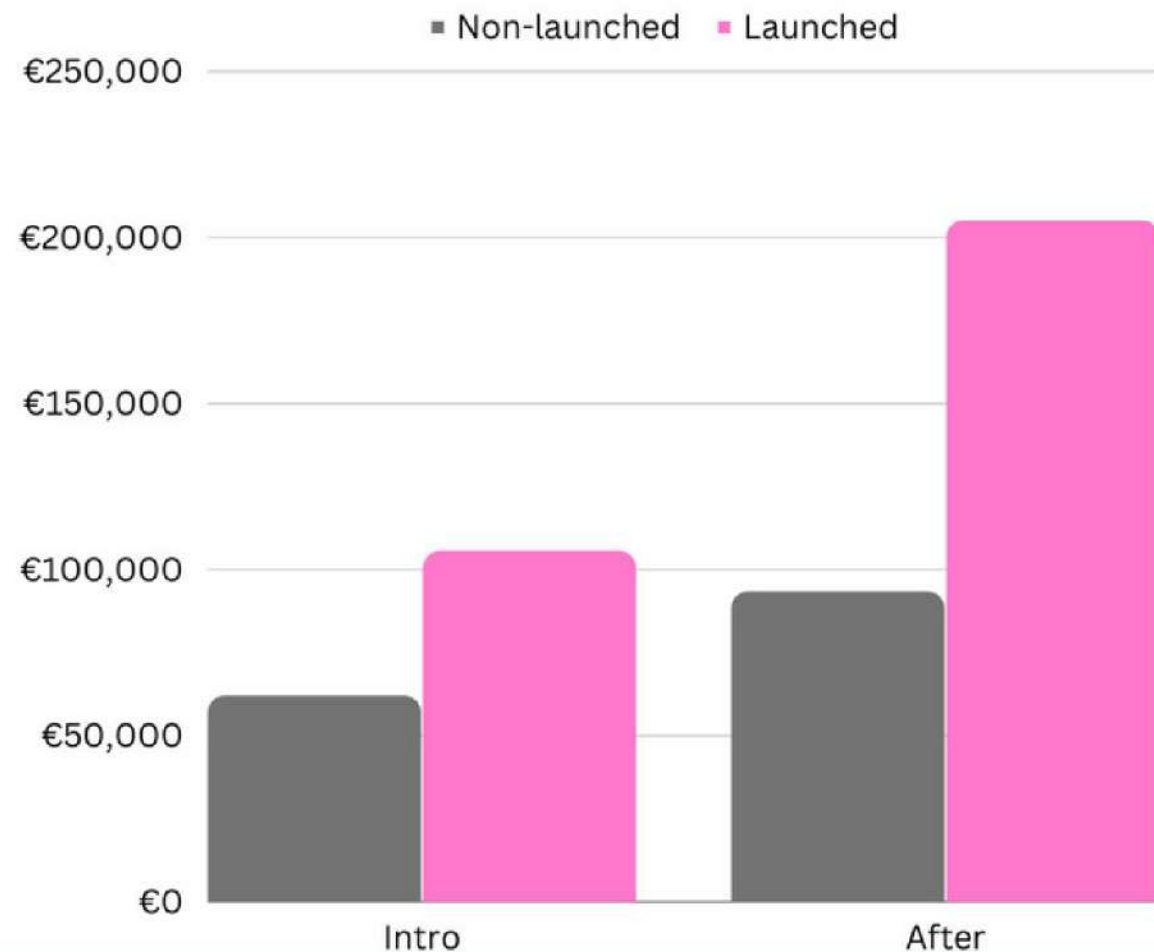**15,000 buyers** repurchasing **500 items** in launch week

Peak in beauty revenue during launch lead to a higher overall level

1.1 Extra revenue is generated in the introduction period

Method

We calculate extra revenue in the introduction period of launch. It is then compared to the non-launch period. Once this is done, the results show the latter is 45% smaller than the former. We then test the assumption of normality for the entire revenue (through a Kolmogorov-Smirnov test). The null hypothesis is rejected, therefore normality can not be assumed (p-value= 2.2e-16). Because data is not normally distributed, we perform a Wilcoxon test to determine the significance of the difference between the two groups. We conclude the mean of the two groups to be significantly different (p-value= 0.01074).

1.2 Extra revenue is generated in the after period

Method

Following the same steps taken in the introduction period, we calculate extra revenue in the period after the product has been launched. The difference is once again, 2.05x times smaller in non-launch than in launch. Because data is not normally distributed, we perform a Wilcoxon test to determine the significance of the difference between the two groups. We conclude the mean of the two groups to be significantly different (p-value= 1.916e-09).

Revenue: Intro vs. After in Launched vs. Non-Launch

Legend: ■ Non-launched ■ Launched

Conclusion

Finally, we combine the extra revenue generated in the introduction phase with the one generated in the after phase to obtain the total. The result is €155,035.10 of extra revenue, as illustrated in the graph.

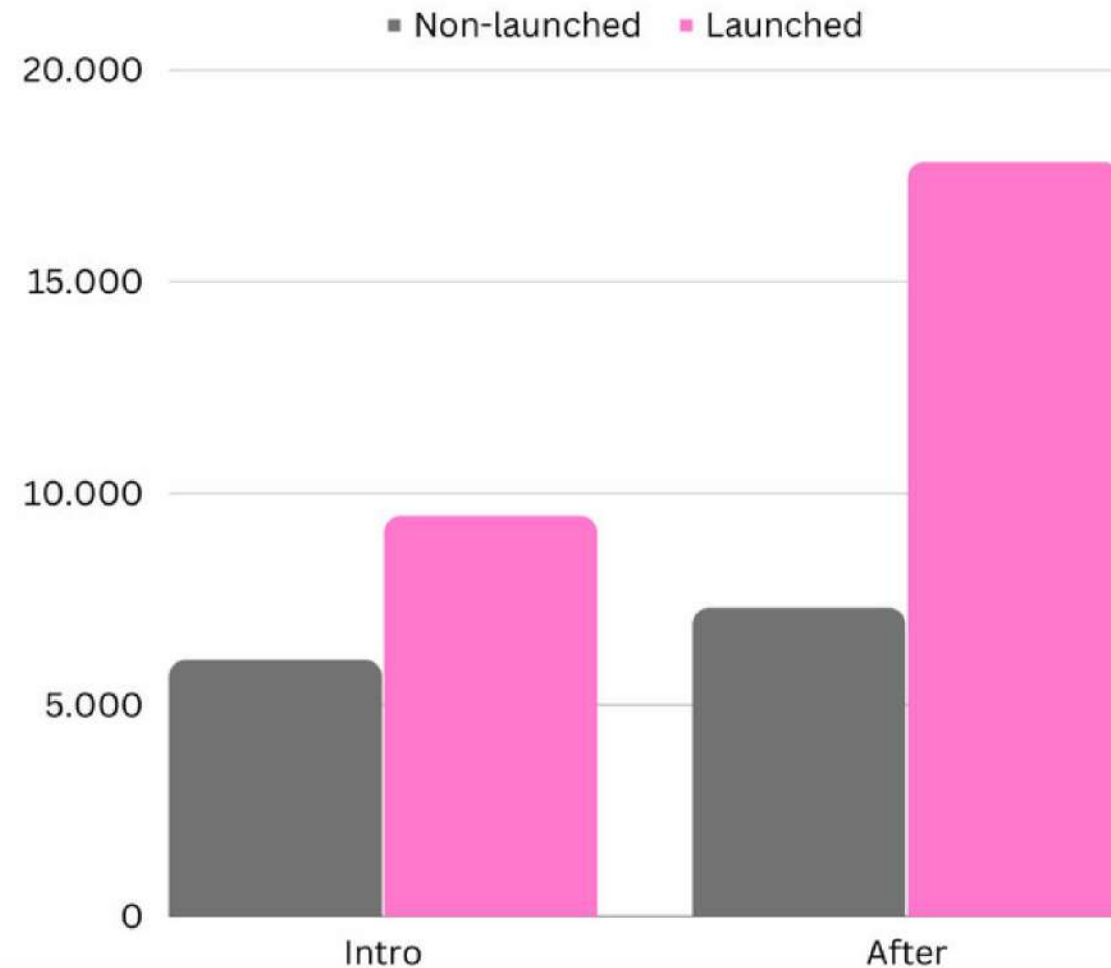| Intro	Before	Total
45% more revenue	**105%** more revenue	**€155,035** total extra revenue

2.1 Extra items are sold in the introduction period

Method

We calculate extra items sold in the introduction period of launch. It is then compared to the non-launch period. Once this is done, the results show 3393 extra items were sold in the introduction period when the product was launched vs non-launched. We then test the assumption of normality for the entire revenue (through a Kolmogorov-Smirnov test). The null hypothesis is rejected, therefore normality can not be assumed (p-value= 2.2e-16). Because data is not normally distributed, we perform a Wilcoxon test to determine the significance of the difference between the two groups. We conclude the mean of the two groups to be significantly different (p-value= 0.003052).

2.2 Extra items are sold in the after period

Method

Following the same steps taken in the introduction period, we calculate extra items in the period after the product has been launched. The difference is once again, 2.05x times smaller in non-launch than in launch. Because data is not normally distributed, we perform a Wilcoxon test to determine the significance of the difference between the two groups. We conclude the mean of the two groups to be significantly different (p-value = 1.916e-09).

Quantity: Intro vs. After in Launched vs. Non-Launch

■ Non-launched ■ Launched

Conclusion

Finally, we combine the total extra items sold in the introduction phase with the one generated in the after phase to obtain the total. The result is **13918 extra items**. The extra items are illustrated in the graph.

Intro
3393 extra items
in the first week

Before
2.05x
extra items

Total
13,918
total extra items

Strange phenomenon:
Outliers in the Revenue variable

There is an **exceptionally high revenue during the month of December**. Some outliers in the revenue are detected midway through December, which could be explained by the increase in demand that comes during the holiday season.

Seasonality

Holiday shopping causes revenue to spike, resulting in outliers in the data. These peaks are considered a strange phenomenon, when put in contrast with the revenue levels generated in other months where there isn't a particularly high spending.

Revenue Outliers in App & Web
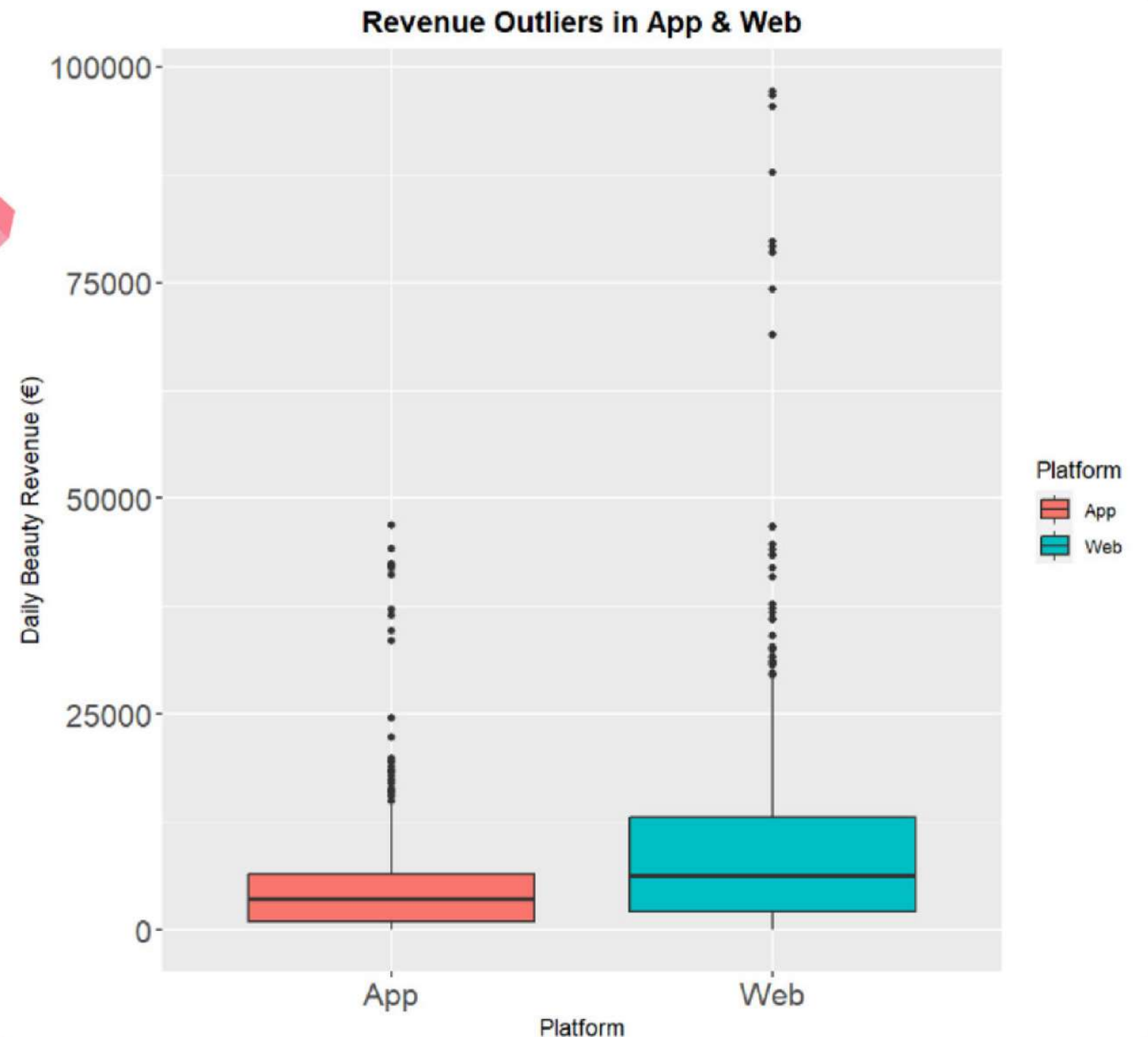
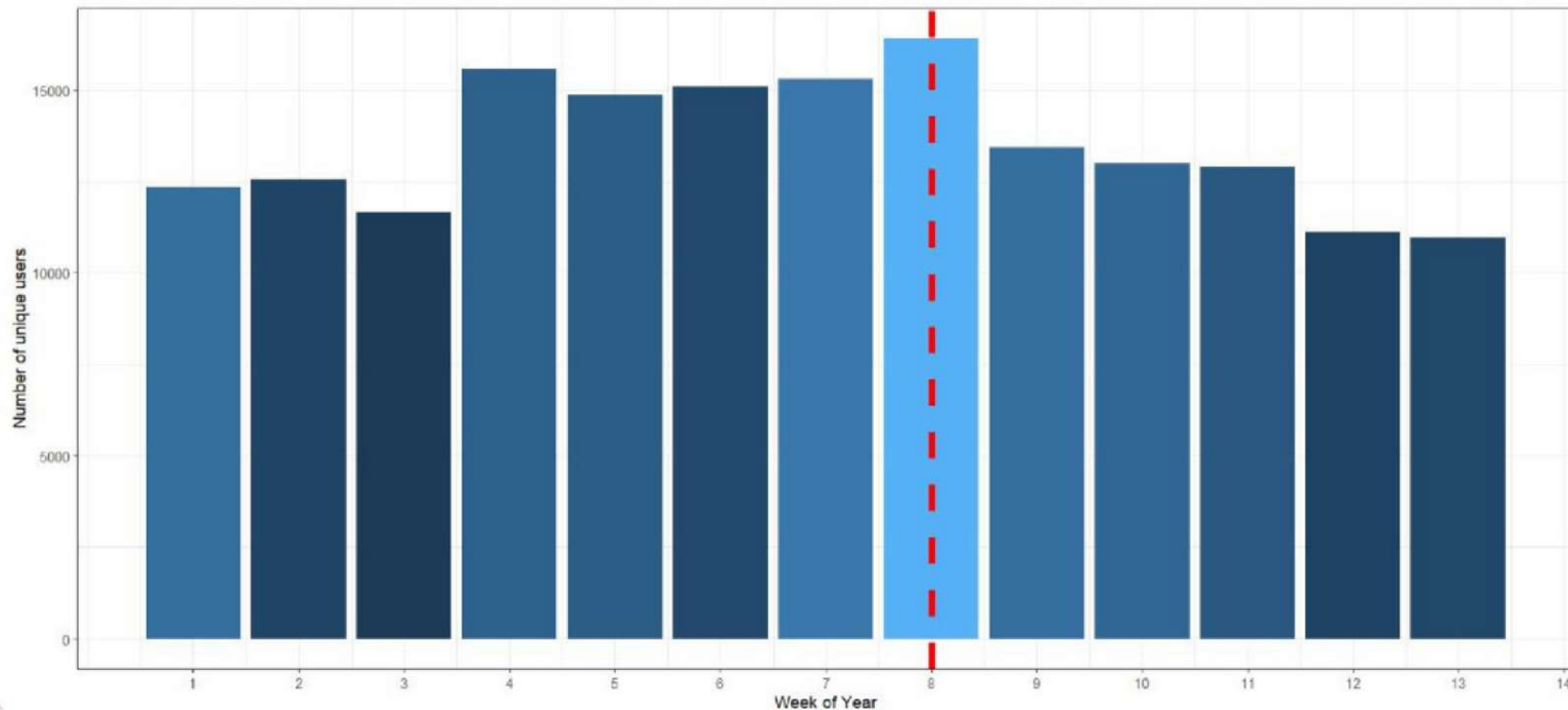# 3 Repeat Purchases

Method

The Magic Mirror's ability to encourage more recurrent sales of beauty products will be examined. We calculated the average of the weekly customers that made multiple purchases from the Magic Mirror. We also develop a bar plot to visually represent the results from the table.

Findings

" In launch week we saw **over 15000 buyers repurchasing 500 items** from the Magic Mirror "

Customers made a lot of repeat purchases from the Magic Mirror for many weeks, as we can see from the bar layout. Yet, week 8 saw the highest number of repeat purchases, with over 15000 buyers repurchasing 500 items.

Limitations

Due to **dataset characteristics**, such as the lack of Customer ID, determining repeat purchases and other customer characteristics or behavior is limited. Finally, the **period of the launch** being during holiday seasons and lack of more post-launch observations presents a seasonality issue, which affects all results.

Unique users Purchasing Beauty Items over Time

Legend

Launch of Magic Mirror

Quantity Beauty Items
500
400
300
200
100
0

Number of unique users

Week of Year

4 Frequency of App use

Weekly Sessions per User Over Time
Mobile & Tablet Combined

Legend
- Launch of Magic Mirror
- Weekly Sessions per User
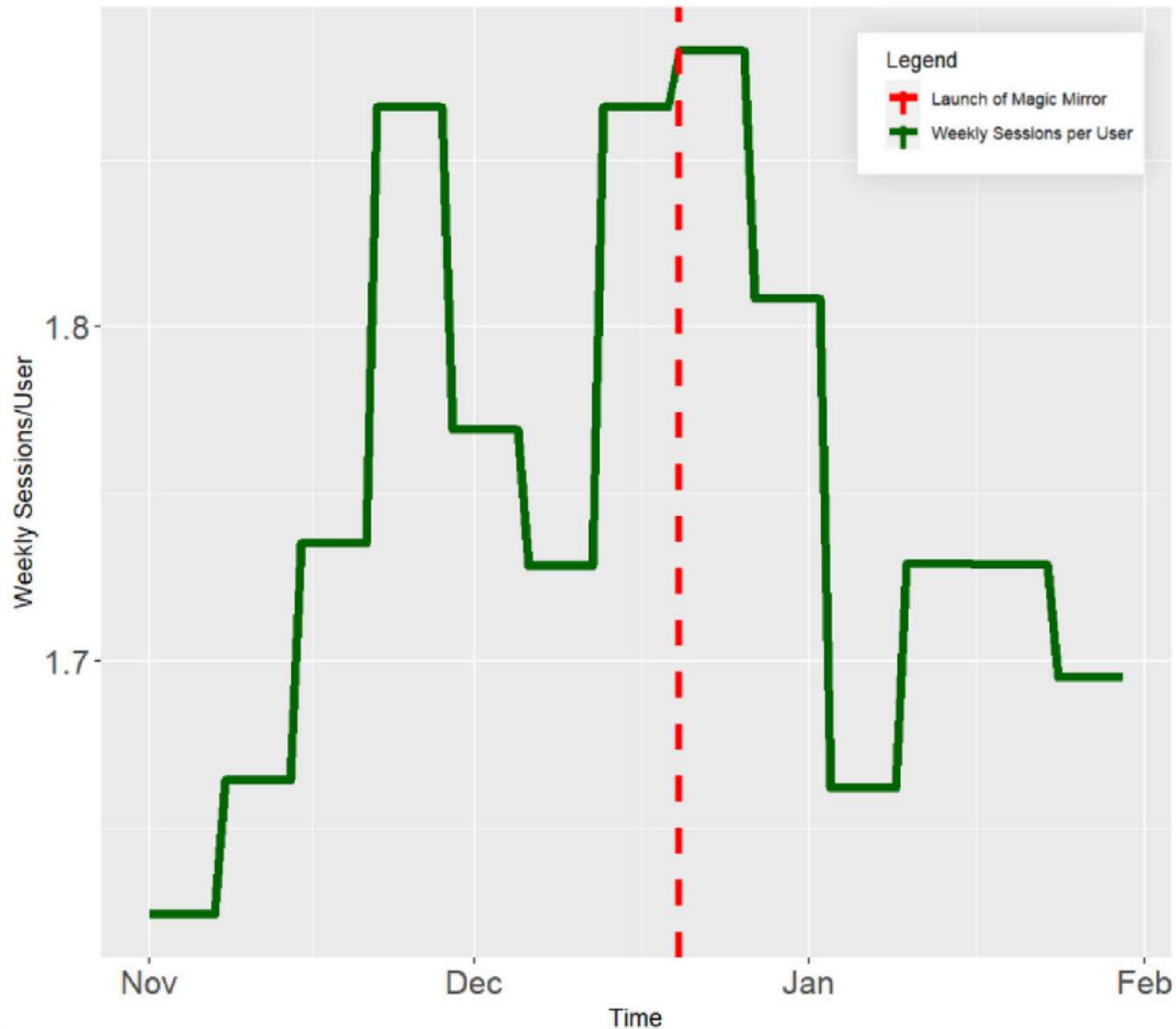
Method

In order to see if there is an increase in the frequency of the app use, we first **aggregate tablet and mobile users** (since the app is both on mobile and tablet). Next, we **plot the weekly sessions per user** over time. Finally, we perform the **Wilcoxon test** to check the statistical validity of our visual findings.

Findings

❝ We cannot clearly see if the Magic Mirror promoted a higher frequency of use. ❞

The plot shows a peak a little past the middle of December, just when the app is launched. The increase in usage is maintained for a few days, before there is a sudden decline just before January. There is another deep decrease almost immediately after, making the weekly sessions per user the lowest of the entire post-launch period. There is one last increase that starts in January and ends in February. Finally, we perform the Wilcoxon test, and the p-value is 0.6626, indicating that the distribution of the two groups is not different.

5.1 Customer Value in Beauty

Weekly Beauty Revenue per User Over Time
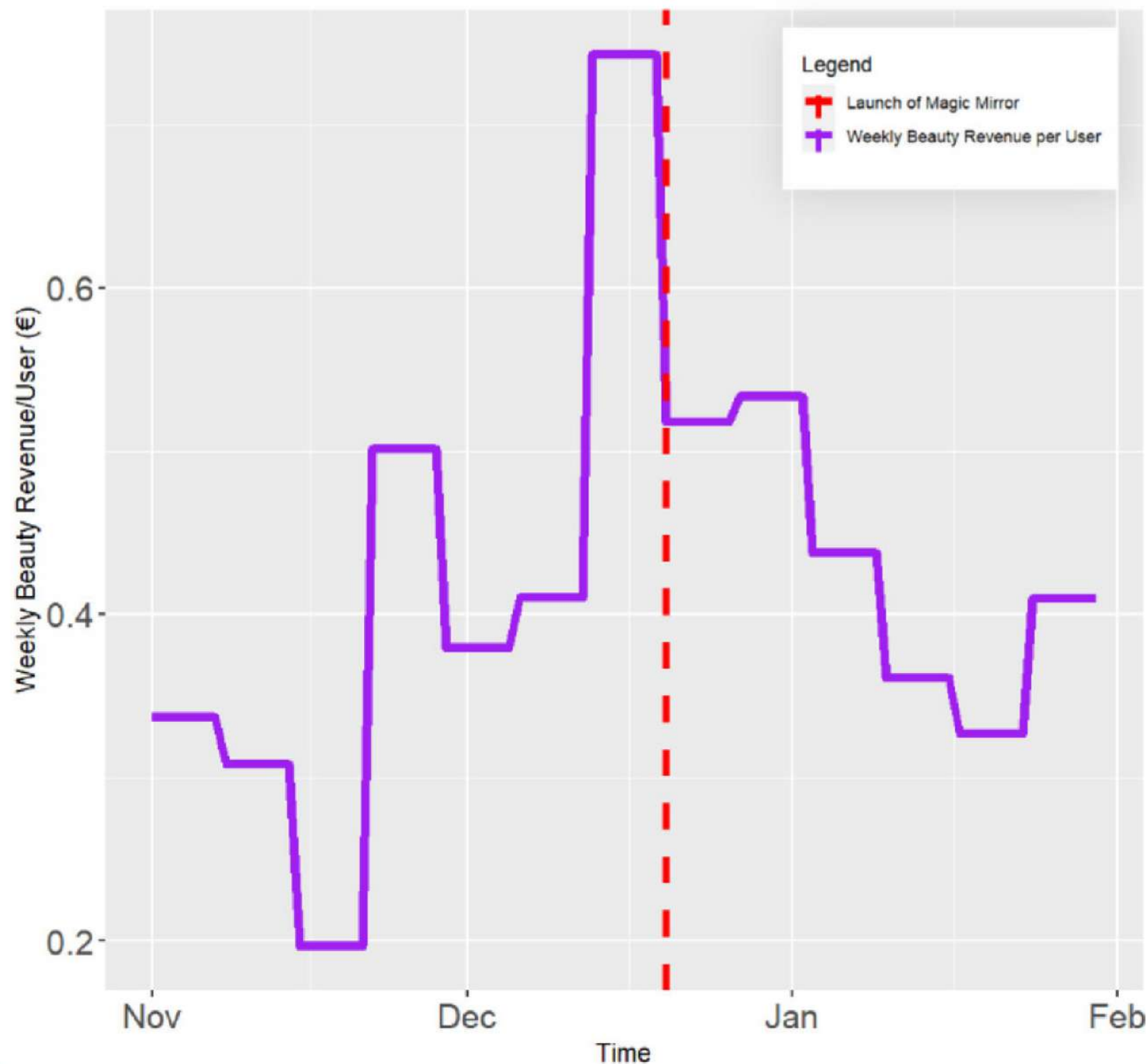**Mobile & Tablet Combined**

Method

To assess whether the Magic Mirror raised the worth of the clients. We first determine the weekly total revenue, non-beauty revenue (total - beauty), and beauty revenue alone. In particular, we make a distinction between the latter two to observe the app's effects on **beauty revenue per user (value)**, as well as the **spillover effects** demonstrated by non-beauty revenue.

Findings

66 We observed a **steep increase in Customer Value** in Beauty pre- and during launch. 99

The graph illustrates the weekly revenue per user. As we can see before the launch time, the value of the customers was rising quickly, but especially **after the Magic Mirror was introduced, we saw a rise in the value of the customers who used it**. Nonetheless, the value for the consumers was still sufficient in the end. Finally, the Wilcoxon test result's p-value of 0.9768 demonstrates that the distributions of the two groups are identical.
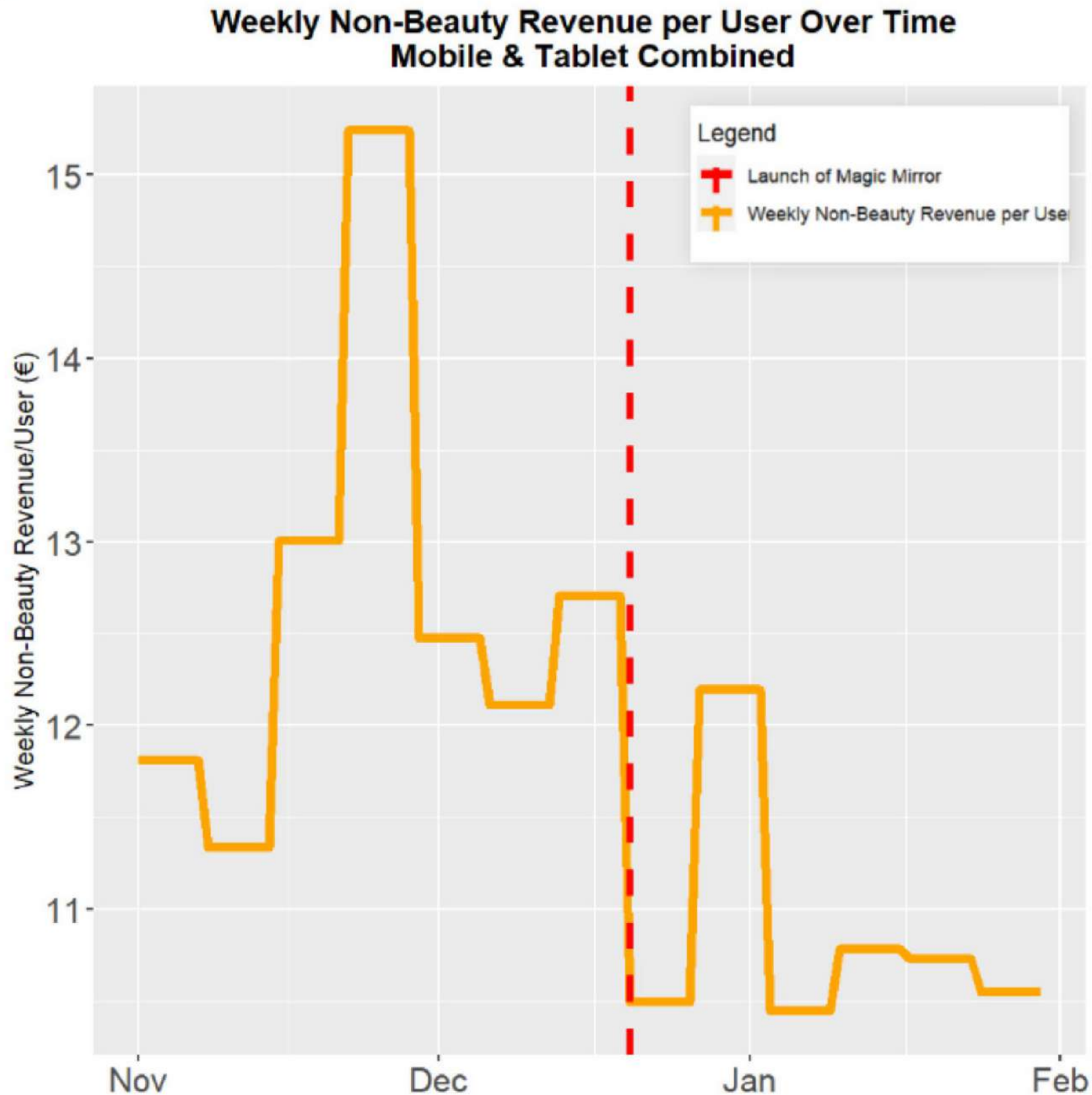
5.2 Customer Value Non-Beauty

Weekly Non-Beauty Revenue per User Over Time
Mobile & Tablet Combined

Findings

❝ We observed a **decrease in Customer Value** in non-Beauty post-launch. ❞

To investigate **spillover effects from the Magic Mirror into non-beauty categories**, the non-beauty revenue generated per user over time is shown on the graph. As we can see, before the introduction, the value of the customers was increasing quickly. However, after December, we observe a decline. Also, we observe that non-beauty revenue drastically decreased after the introduction of the magic mirror as a result of the magic mirror's increased revenue. Finally, when we perform the Wilcoxon test, the result's p-value of 1.398e-15 shows that the distributions of the two groups are different.

Suggestions & Improvements

(a) **Increase marketing effort** after launch as the use of magic mirror declined after launch

(b) From Google analytics we found that there was effort involved to download Magic Mirror, a possible improvement could be to **add a promotion of the app in Wehkamp's landing page**
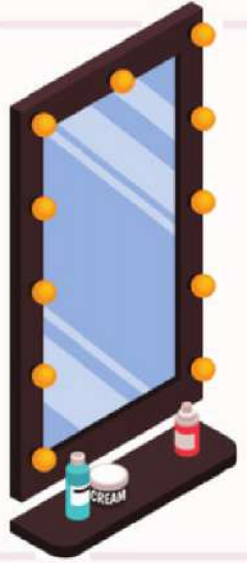

(c) **Introduce a web version** of the app giving the option to upload a picture integrating web and app

(d) **Add a QR code** in each beauty product page on Website which you can scan and access the Magic Mirror feature with this product loaded to AR try-on

Limitation & Solutions

(a) Lighting: The virtual try-on feature may not accurately reflect how the makeup will look in different lighting conditions. Therefore, users should be cautious when selecting shades or colors based solely on the virtual try-on.

To prevent a possible increase in returns due to this AR limitation, Wehkamp could include a notice of this limitation with tips for best results and recommendations based on past makeup purchases.

(b) Skin Tone: The virtual try-on may not be able to accurately match a user's skin tone. As a result, the makeup products may not appear as they would in real life. Moreover, the algorithm should be trained in a way that is inclusive for all skin tones.

An essential step is to train the algorithm on a representative sample of the target population, including people of all skin colors in order to diminish inaccuracies based on skin tone.

(c) Personal preferences: The virtual try-on feature may not account for personal preferences such as texture, smell, or feel of the product, which can only be determined by physically trying the product.

As returns data accumulates due to this, identify problematic products in terms of smell, texture and feel by analyzing return reason text. Aggregate reasoning to build customized notices warning clients when they AR try-on these products.

APPENDIX

The R script behind the insights.