Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Master thesis

Understanding Customer Behaviour to Optimize Product Sorting for E-Commerce Websites

by

Amina Voloder

Faculty of Mathematics, Informatics and Natural Sciences
Department of Informatics
Working group: Scientific Computing

Degree programme:	Informatics
Student number:	6869312
Supervising examiner:	Prof. Dr. Thomas Ludwig
Second examiner:	Dr. Julian Kunkel
First supervisor:	Dr. Julian Kunkel
Second supervisor:	Antonia Wrobel
Submission:	Hamburg, 2018-08-07

Abstract

This thesis investigates product-specific and user-specific characteristics that influence the sales in order to develop a novel sorting algorithm for application in the field of e-commerce, through the analysis of customer preferences and the nature of a given store's products, to improve the personalisation of online shopping systems. The algorithm optimises the order of displayed products by their purchase probability, determining which products users are most likely to purchase. This is determined by investigating the correlation between product sales and: product seasons, the time of day, and the devices users own. It was found that products could be classified as being sold well in particular months or regardless of the month. Similarly, products could be sold well at particular times of the day or regardless of the time. It was also determined that users of Apple devices should have more expensive products promoted to them, as they typically purchase greater quantities of expensive products. The algorithm is evaluated by means of visual and quantitative comparison against the standard sorting algorithm used within an e-commerce system by novomind AG. Test results indicate a discernible variation in the sorting order of products, as well as an increase in the variety of the eight highest sorted products. The full contribution of the algorithm to the sorting optimisation is verifiable through real-world A/B testing.

Contents

1 Introduction

This chapter explains the motivation for this thesis (Section 1.1), describes its goals (Section 1.2) and gives an overview of the thesis content (Section 1.3).

1.1 Motivation

In 2001, it was predicted that personalisation will be "*the competitive advantage that will determine the winners in the market of online shopping*" [GP01] as it significantly increases customer satisfaction, customer acquisition rates, sales and builds the long-term relationship with the customer. Today big companies like Amazon apply this technique successfully.

An important method to achieve personalisation are recommendation systems. They help users find what they are looking for and what they will be interested in. From the perspective of the shop owner, they turn shoppers into clients. The biggest drawback of most recommendation systems is that they rely on user ratings or static user profiles and neglect a large set of relationships between objects. This is still the case with the earliest form of recommendation systems which use item-based and user-based collaborative filtering. These two methods ignore a product's nature and customer preferences.

In other words "*a traditional recommendation algorithm cannot perform [the predictive task well]*" [JQ15], the main reason being that predicting the customer's impression of the product is very different from predicting the purchase probability of the product by that customer. Another drawback of traditional and some newer solutions is that customer demographics and family backgound information are not available in e-commerce.

This thesis aims to overcome the mentioned drawbacks by focusing on the purchase probability. The novelty of the approach is the use of specific product and user features which to our knowledge have not been exploited before. By analysing and applying the mentioned features, the attempt to include both first and second stage of the purchasing decision in e-commerce will be made (see Figure 1.1). The existing approaches mainly focus on Stage 2. As seen in Figure 1.1, Stage 1 refers to the consumers' motivations to pursue products while Stage 2 represents the choice between different products.

When users browse through a category or search results, the sorting itself is specific type of recommendation system application because users usually only view the first few product pages (mostly the first page). The best case for sorting would be to show what the user really wants on the first product page. The product page can be at the bottom of a category hierarchy (e.g. electronic tootbrushes) where the articles are very similar and it is not easy to differentiate between them. In this context there are aspects that need to be considered when designing a sorting algorithm. For example, new products should get a chance which means that the sorting in the end will be a mixture of user interests, association between products and the “freshness” of products.

This paper focuses on returning customers and harnesses the history of purchases in order to generate improved sortings. Since there are product features such as the position on the hype curve, there is space for developing the model further.

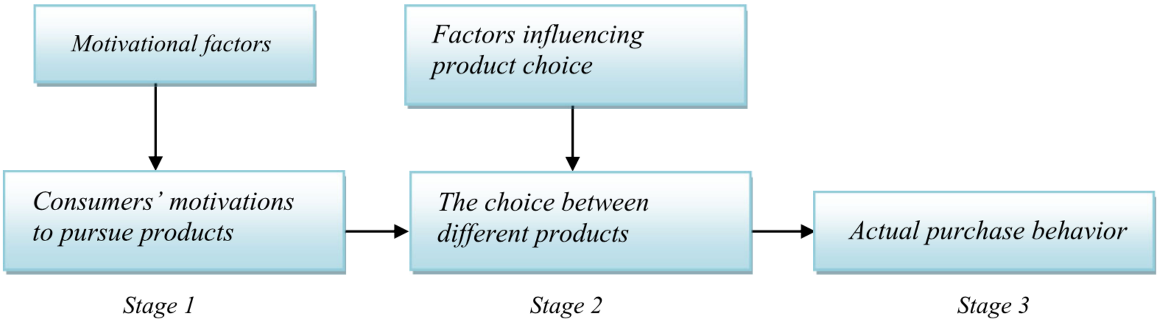

Motivational factors
Factors influencing product choice
Consumers' motivations to pursue products
The choice between different products
Actual purchase behavior
Stage 1
Stage 2
Stage 3

Figure 1.1: Consumer behaviour and the purchasing decision in the e-commerce context [JQ15]

Disclaimer:

The data used in the thesis belong to the customers of novomind AG. Due to the privacy policy, they are partially anonymised. This partial anonymisation includes:

- concrete numbers in tables and figures are either removed or mutliplied with arbitrarily chosen factors
- values of some X- and Y- axes are removed
- Chapters 4 and 5 are left out
- some parameteres used for analysis are mentioned as “certain” numbers
- IBM SPSS program screenshots are shaded

1.2 Goals

The main goal of this thesis is the exploration of the customer and product data and using this knowledge to optimise the sorting within different categories on e-commerce websites. The optimisation is done via the following methodology:

- The specific hypotheses about customers and products are chosen.
- Data are explored with help of data mining and statistics in order to support the chosen hypotheses.

- Based on the affirmative hypotheses' results, the sorting algorithm is developed.
- The approaches for sorting new products in the shop are considered.

Overall, the resulting algorithm should be a solution for sorting which works online, within every category and does not require a vast amount of data or a long history of product/customer behavior.

1.3 Outline

Apart from the introduction, this thesis contains six more chapters. In Chapter 2, the key terms necessary for understanding the thesis are presented together with the researches related to the thesis' topic. The most important chapter, Chapter 3, explains the research methodology and each of the hypotheses which will be used for the sorting algorithm. Chapter 4 presents the current sorting algorithm and designs the new sorting algorithm. It also discusses the impact of the hypotheses on the sorting order. The implementation of the new algorithm is then briefly described in Chapter 5. The results of tests obtained after the implementation are listed in Chapter 6. Finally, the conclusion is found in Chapter 7.

2 Background and related work

This chapter contains the short overview of technical terms needed to understand this thesis and the overview of papers published so far about the similar topics. Kullback-Leibler divergence (KLD), used for the quantitative analyis, is described in Section 2.1. Two tests used for the statistical hyptohesis testing can be found in Section 2.2 and Section 2.3. For differentianing device types, user agent strings are introduced in Section 2.4. Finally, the summary of related work is presented in Section 2.5.

2.1 Kullback-Leibler divergence (KLD)

KLD "*is a measure of how one probability distribution diverges from a second, expected probability distribution*". [Kul]

The advantage of the KLD is visible in comparison with Euclidean distance. "*Using Euclidean distance ... ignores the variance information of the underlying [forecast] distributions. In contrast, under normal assumption, the Kullback-Leibler distance ... [considers] both the mean and variance information[.]*" [TL] Thus, Kullback-Leibler divergence is more appropriate for comparing the distributions.

An example illustrating the difference between the Euclidean distance and KLD is shown in Figure 2.1. The two distributions are similar if the Euclidean distance is used because the mean difference is 0. If KLD is used, the value of 1.78 means that the two distributions are different.

If the KLD is 0, it means that the two distributions which are compared will behave similarly if not the same and that the information loss when replacing the first distribution with the second one is minimised.

In this thesis, the KLD will be used when comparing different distributions of product sales within a year and a day.

2.2 Mann-Whitney test

The Mann-Whitney test is used for statistical hypothesis testing when comparing differences between two independent groups, e.g., *The Mann-Whitney test can be used to test the durability of the two brands of hair dye.*

The hypotheses have the following form:

Null-hypothesis: There is no statistically significant difference between two populations.

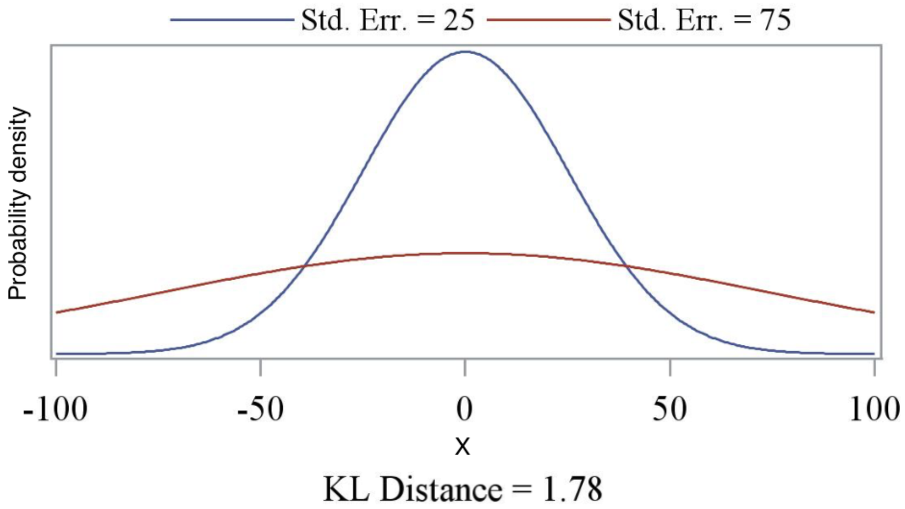


Figure 2.1: Two probability distributions with the same mean value [TL]

Alternative-hypothesis: There is a statistically significant difference between two populations.

The precondition for using this test is that the **dependant variable** needs to be either **ordinal or continuous** and that **groups are independent**. It is often used as an equivalent to the *t-test* when the distribution is not normal since it does not require normality.

The procedure of the test consists of combining the observations from both samples into one sample and ranking them from lowest to highest (from 1 to n1+n2), while keeping track of the origin sample of the observation.

The result of the test is statistic U. The U-value is the number of times observations in one sample are ranked higher than observations in the other sample. The U-value has an associated p-value which is used to indicate whether the difference between the groups (if it exists) is significant. More precisely:

If $\mathbf{p} \leq \alpha$: The difference between the groups is statistically significant and null-hypothesis is rejected.

If $\mathbf{p} > \alpha$: The difference between the groups is not statistically significant and null-hypothesis cannot be rejected,

where α is a significance level.

2.3 Chi-squared test for independence

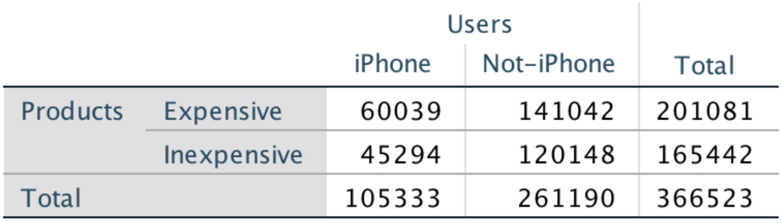
The chi-squared test for independence is useful for estimating if there is a significant relationship between two **categorical** values.

E.g. *The Chi-squared test for independence can be used to test if there is a relationship between depression and gender (male/female).*

The hypotheses have the following form:
Null-hypothesis: There is no significant relationship between the two variables.
Alternative-hypothesis: There is a significant relationship.

In order to use the Chi-square test, **each of the two variables** needs to have **at least two categories**, both **observations** and **variables** should be **independent** and the **sample size** should be relatively **large**.

The *contingency/crosstab/two-way* table is used to analyse the data. "*...Each row represents a category for one variable and each column represents a category for the other variable.*" [Jam] An example of such table in the IBM SPSS Program is shown in Figure 2.2. It shows that 60039 products are expensive and purchased by iPhone users.



		Users		
		iPhone	Not-iPhone	Total
Products	Expensive	60039	141042	201081
	Inexpensive	45294	120148	165442
Total		105333	261190	366523

Figure 2.2: An example of a two-way contingency table

The test statistic, a Chi-square value (χ^2) is the squared difference between the observed and the expected frequencies of the variables. Just as in the case of Mann-Whitney U Test, the Chi-square value has an associated p-value and:

If $\mathbf{p} \leq \alpha$: The difference between the groups is statistically significant and null-hypothesis is rejected.

If $\mathbf{p} > \alpha$: The difference between the groups is not statistically significant and null-hypothesis cannot be rejected,

where α is a significance level.

2.4 User agent strings

A user agent string serves as a browser's identification that is sent to a web server with each request. It contains the information about the browser version, the operating system, the type of device a user uses and other information. In this thesis, only a few user agent categories are of interest. They are recognised by specific keywords in the *user agent string* as well as the *user agent category* saved in the company's database. A *user agent category* is a short summary of user agent string and it contains the information about the device type e.g. smartphone, tablet etc. in a clearer form than user agent string. The overview of all observed user agent categories is listed in Table 2.1. The words in bold font are used as category keywords. Different tools for analysing user agent strings are available online.

Device	**User agent category examples**
iPhone	**Smartphone**:MOBILE_SAFARI:**Apple** Inc.:Mobile Browser:8,3:iOS 8 **Smartphone**:CHROME_MOBILE:Google Inc.:Mobile Browser:50,0:**iOS**
Apple	Tablet:MOBILE_SAFARI:**Apple** Inc.:Mobile Browser:,:iOS 8
Android smartphone	**Smartphone**:CHROME_MOBILE:Google Inc.:Mobile Browser:62,0:**Android** 4.2 Jelly Bean
Desktop	**Personal computer**:ICEWEASEL:Software in the Public Interest, Inc.:Browser:35,0:Linux
Other	Tablet:CHROME_MOBILE:Google Inc.:Mobile Browser:52,0:Android 4.1.x Jelly Bean

Table 2.1: Overview of observed device categories

2.5 Related work

A recommendation system has a goal of narrowing down selections for users and in this thesis it will be used to optimise product sorting. There are different possible classifications of recommendation systems but the most general one is:

- Collaborative filtering
- Content - based filtering
- Hybrid approaches

2.5.1 Collaborative filtering

Collaborative filtering is based on the premise that users who agreed in the past will agree in the future. In other words, the recommendations are made based on the similarity amongst users by analysing users' behaviour, activities or preferences.

Depending on what data are analysed, collaborative filtering can be:

- **Item-based:**

  In the item-based approach, the recommendation is based on the user's raking of the items. When the similarity is computed, the algorithm only knows the users' history of ratings. A rating can be a purchase so that the more two items are purchased together, the more similar they are.

  The disadvantage of this approach are:

  - item cold-start problem - the ratings for new items are not known before similar users rate them
  - scalability - computations become slow for millions of ratings
  - sparsity - for the large sets of items, only some items are purchased together

- **User-based:**

  Instead of calculating the similarity between items, the similarity between users is calculated based on their ratings i.e. opinion about items. Recommendations are then items liked by the closest users.

  The disadvantage of this approach are:

  - cold start problem - it is difficult to make recommendations for new users since it is difficult to find the similarity with other users immediately
  - item cold-start problem - the ratings for new items are not known before similar users rate them
  - sparsity - for the large sets of items, users usually rate only some of the items
  - scalability - computations become slow for millions of ratings

In both approaches of the collaborative filtering, the system does not know why the items are related. It only has an information about items being purchased together or being liked by the users with similar preferences.

2.5.2 Content - based filtering

As opposed to collaborative filtering, the content - based filtering uses the content of both item and user. These information are used to create *user* and *item* profiles. *Item profiles* are created first and they consist of a set of features for each item. Then, user profiles are build on features of items that are purchased by the single users. Finally, the similarity scores can be calculated between user and item profiles in order to recommend the items with highest scores.

The disadvantages of this approach are:

- cold start problem - a large amount of existing user data is needed which is not available for new users
- sparsity - for the large sets of items, users usually rate only some of the items
- scalability - large amount of computation power necessary for millions of products and users
- lack of diversity - recommended items are similar to the items already purchased by the user
- incorrectly/inconsistenly applied features - the quality of content-based filterring depends on the quality of item tags. With enormous number if items, it is challenging to have all features applied consistently or acurately

2.5.3 Hybrid approaches

Since both collaborative filtering and content-based filtering have major limitations, especially not performing well with large number of items and users, both approaches and multiple techniques are in most cases combined into a hybrid solutions to overcome the limitations of individual approaches.

Besides combining the approaches, it is common to use Machine learning algorithms such as cluster analyis, decision trees or aritficial neural networks to estimate the probability of user liking the item.

2.5.4 Current status of recommendation systems

Recommendation systems were first mentioned in 1990 but there is room for improvement even today. One of the major goals towards which the modern systems are developing, is to find a solution which would not be extremely personalised which would feel too intrusive, but also not too generic so that it takes users's specific taste into consideration and supports serendipity.

Another issue that needs to be taken care of is the cold start problem. A good solution should be able to offer recommendations even when little or no explicit user information is available.

This thesis aims to develop an algorithm which will be highly useful, even without much customer data being necessarily available, by using the features which have not been used before, as a preparation for the eventual Machine learning algorithm. There are many algorithms dealing with similar issues developed so far and some of them are presented in the following papers.

2.5.5 Scientific papers

Since the weaknesses of both collaborative filtering nor content-based filtering cannot be overlooked, the majority of the recommendation systems today is hybrid. For this reason, the papers mentioned in this subsection present the algorithms which use hybrid approach to make product recommendations.

Sorting with one product profile and three customer profiles using personal information [YJP17]

In this paper, clicks, basket insertions, purchases and interest fields are used for bulding a product profile and three different customer profiles based on: individual purchasing information, individual behaviour information, and individual and group behaviour information. The third customer profile model based on the individual and group behaviour information showed the best results. Its strongest point it that it uses both customer (individual and group) and product data and combines them. Its drawback is that age, gender, and occupation are used for group profiles creation and it takes time until there is enough data to create the profiles.

Predictive model based on products popularity, needs and preferences of the customers [Qiu14]

As mentioned in the tile of the paper, the predictive model called COREL (CustOmer purchase pREdiction modeL) is developed for the purposes of product sorting. COREL uses "*the needs of customers, the popularity of products and the preference of customers*" [Qiu14] to make predictions. Additionally, it uses purchase data and ratings of products to improve the model. The algorithms used are Support vector machines and Bayesian discrete choice model. COREL outperforms the baseline models and overcomes the weaknesses of collaborative filtering. The minor drawback is that it also uses another model called the Heat model to calculate the popularity of the products relying on the crowdsourcing approach. In the real world, the data would need to be labelled as *popular* or *not-popular*.

Hybrid model as a combination of data mining and collaborative filtering [GP02]

Prassas et. al combine data mining and collaborative filtering to make a hybrid model. The data mining is used for selecting a product category to be recommended by rules extracting. Then, collaborative filtering is used for selecting specific products from the chosen category. The model's strongest feature is that it can always suggest a recommendation. One of the important weaknesses is the rule processing which may cause a problem when it is done online. This model is more appropriate for grocery retail shops than clothing, furniture or similar types of stores.

Recommendation sytsem based on purchase patterns [Lu14]

The algorithm proposed in this paper uses the purchase history of users to recognise their purchasing pattern by using consecutive subsequences and then using this pattern to predict the category of next purchase. Since it can only predict the catgory but not products, the algorithm cannot be used within a single category which is the goal of this thesis.

ANN Based Recommendation Algorithm for E-commerce [AP17]

One way of sorting products is based on the customer comments and reviews on the product. The algorithm in this paper uses artificial neural networks to analyse the comments in order to determine the ranking. The system proposed consists of two modules:

- ANN based automation system: single comments about each product are processed and percentage of how much the product is buyable is given at the output
- Second module: calculates the average of all previously calculated percentage values

Once all the average vaules are calculated, the products can be ranked.

The big advantage of the suggested solution is that the user comments are automatically processed and the results can be updated thanks to ANN's ability to learn and be trained. However, not every shop has customer reviews or comments or it has only limited number of reviews. For that reson, it would be necessary to use ANN with inputs other than user comments.

Using neural networks and social networking for recommendation system [MR17]

This paper suggests using the knowledge from social networks to deal with the cold start problem when recommending products. The authors suggest that the new trend of social networks users conducting e-commerce acitivities on social networks e.g. pressing buy button from a Facebook post to purchase a product, might be an opportunity for better recommendation systems since the traditional solutions mostly use historical transaction records. The algorithm presented has three main steps:

- using recurrent neural networks (RNN) to learn user's and item's feature representation from e-commerce websites data
- deploying modified decision tree technique to transform user's social media features and add them to the existing user's features
- using feature-based matrix factorisation for a cold start product recommendation

Even though this paper investigates an important issue, cold start problem, it has a limited application field since many e-commerce webistes do not have an access to user social networks accounts.

2.5.6 iPhone vs Other users

There is a popular belief that iPhone owners spend more money than owners of other devices. The possible reason for this are the higher prices of iPhones, which also means that better developed countries have more iPhones and at the same time more money, and the greater amount of time that iPhone owners spend with their phones.

The following three studies investigating the difference between iPhone and Android users were conducted by Adobe [Inc16], Wolfgang Digital [Col17] and Pew Center for Internet and American Life [Aza12].

Adobe surveyed the Black Friday online sales. According to that sales report, "*\$368 million sales came from [iPhones], \$180 million from Android phones, \$302 million from iPads, and \$50 million from Android tablets.*" [Inc16] Wolfgang Digital analysed annual sales in e-commerce over 143 million sessions totalling €447 million in online revenue. The results showed that the average value iPhone users spend per transaction (AOV) is €27.66, which is 185% more than Android users whose AOV is only €9.69. Pew Center for Internet and American Life found out that "*iPhone users spent an average of 1 hour 15 minutes on their devices in total, whereas Android users only logged in an average of 49 minutes usage*". [Aza12]

All three studies were conducted on the American market and the question is if this holds for the German and European markets.

Summary:
As seen in the related work, there are numerous ways to realise recommendation systems and all of them have their strong and weak points. This thesis aims to develop a relatively simple algortihm which does not require large amounts of data to be captured over time before sorting and can be used to sort products within a single category.

3 Data exploration

This chapter contains the information about the data (Section 3.1), data preparation (Section 3.2) and the research methodology which will use those data (Section 3.3). Further, it explains and investigates different hypotheses which later serve as the basis for the sorting (Sections 3.4 - 3.11).

3.1 Dataset

novomind AG[1] is in charge of many e-commerce shops with focus areas in online retail and electronic customer communication. Different e-commerce systems are used for developing the shops and one of them is chosen for this thesis. Two of the shops built in the chosen system and their data are chosen for the purposes of this thesis. Data analysed are product, customer and sales data. The evaluated datasets used contained more than 200.000 product items and more than 1.000.000 customer records. Both shops have multiple category levels and the resulting algorithm can be applied on all of them or only on the specific ones.

3.2 Data cleaning

As in most shops, there are test users and test products that regularly purchase goods to make sure that the shop is running smoothly. Thus, the stored data needed to be pre-processed as follows:

1. **Test user removal:** The most challenging step, since there was initially no clear indication in the database who are test users. After further research, it was established that they had a special name, surname or email address in the private customer's database.

2. **Test product removal:** The test products were recognisable by name.

3. **Free product removal:** The shop catalogue is free and the records containing it needed to be removed because it is not a product that can be purchased but an online document.

4. **Choosing a limit:** Depending on the purpose of each research, it was necessary to decide on lower limits for the observed group, e.g., only customers who bought at least two products or only products sold 100 times are included.

[1] https://www.novomind.com

5. **Outlier removal:** For outlier removal, the interquartile range with the following formula is used:

$$IQR = Q3 - Q1 \tag{3.1}$$

Q3 is the 75th percentille and Q1 is the 25th percentille.

A data point x is an **outlier** if:

$$(x < Q1 - 1.5 \cdot IQR) \text{ or } (x > Q3 + 1.5 \cdot IQR) \tag{3.2}$$

Steps one, two and three are done only once while step four and five had to be repeated for each experiment with a experiment-specific setting.

3.3 Research methodology

The goal of this thesis is to explore and use previously unexplored features of customers (users), and products for the purposes of product sorting.

The methodology for building the sorting system will be the following:

Step 1 : Data exploration

1. Data cleaning (as described in Section 3.2)
2. Verifying the hypotheses on example data: before doing the detailed hypothesis analysis, it is useful to look for examples of data which satisfy the hypothesis
3. Verifying the hypotheses relevance: percentage/sales of users/products: after examples confirming the hypothesis are found, it is necessary to measure the percentage of users/products which are affected by the hypothesis

Step 2 : If there are enough hypotheses which are relevant for the majority of users, their impact on the sorting algorithm needs to be estimated. In other words, the hypotheses need to be translated into practical steps for the sorting algorithm

Step 3 : If there are enough hypotheses which are relevant for the majority of users, implement the sorting algorithm using Hadoop in the company's platform

Step 4 : Testing on the live e-commerce websites: the old and new sorting will be compared

Step 5 : Developing methods to infer features for new users/products from existing users/products

The fifth step is an optional step and it will be discussed at a high-level.

The biggest challenge of the sorting is that users only check the first 1-3 pages which means that the sorting has to be good enough that products most interesting to user will be on the first page.

3.3.1 Hypothesis types

Three types of hypotheses were considered before choosing the hypotheses for Step 2:

1. **Customer fixed hypotheses (CF hypotheses)** depend only on the customer features.
   e.g., iPhone users are more likely to buy expensive products.

2. **Product fixed hypotheses (PF hypotheses)** depend only on the product features.
   e.g., the majority of sales of this product takes place at certain times of the day.

3. **Both customer and product related hypotheses (CPR hypotheses)** depend on both customer and product features.
   e.g., light feather jackets are typically sold in autumn, but there are customers who don't care about seasonality.

3.3.2 Chosen hypotheses

As previously mentioned, a very important point for chosing the hypotheses was that they can be used for sorting as soon as possible without having to wait a long time for customers to purchase products or for products to be sold. For this reason, after the initial data exploration, the following hypotheses are chosen:

1. Hypothesis: A product's sales depend on the month of the purchase. Some products are sold well in all months (non-seasonal products).

2. Hypothesis: The product sales depend on the time of the day. Some products are never sold at specific hours.

3. Hypothesis:
   a) iPhone users spend more money than other users
   b) Apple users spend more money than other users
   c) iPhone and Android users spend more money than desktop and other users
   d) iPhone users purchase greater quantities of expensive products than other users
   e) Apple users purchase greater quantities of expensive products than other users

Unfortunately, it is necessary to have sales of the previous year to perform sorting based on the first hypothesis. For the second hypotheses, only a few days are enough to be able to make sorting based on the product sales. Finally, based on the third hypothesis, the sorting is done instantaneously.

The listed hypotheses will be further investigated in the next eight sections. For each hypothesis, the following analysis steps will be performed:

- **Implications of the hypothesi**s - a way in which the hypothesis affects the sorting
- **Examples confirming the hypothesis** - real world examples of customers or products from the shop which confirm the hypothesis
- **Verification of the hypothesis relevance** - obeserving the amount of customers or products for which the hypothesis is relevant

3.4 Seasonality with respect to products

Hypothesis: The product's sales depend on the month of the purchase. Some products are sold well in all months.

3.4.1 Implications of the hypothesis

If the hypothesis is correct, the following product sorting would be preferable:

1. In-season products
2. Nonseasonal products
3. Out-of-season products

Depending on the season and weather, customers have different needs. If user visits the "Shoes" category in January, winter boots or sneakers which sell well in Winter will be shown on the first page, while sandals and flip-flops will be pushed to the last pages.

One of the thesis' goals is to help customers find products they want to buy as fast as possible. At the moment of purchase, the customer is most likely to be looking for either in-season or nonseasonal products. For this reason, out-of-season products are pushed to the end. In-season products will be placed before nonseasonal ones because nonseasonal products are sold through the whole year and are not characteristic of any month.

3.4.2 Examples confirming the hypothesis

On the following four pages, in Figures 3.1, 3.2, 3.3 and 3.4, two seasonal and two nonseasonal products are shown.

Each pair of diagrams contains histogram and distibution plot for a single product sales over the two years. In histogram, number of products sold on each day of the year is shown, while the distirbution plot shows how are those sales distributed over the whole year.

Item1 and Item2 have almost identical behaviour. Approximately 20% of their purchases were made within the first three months of the year, while the remainder (and majority) of purchases were made in the last five months of the year. This makes them seasonal products because they are sold predominantly in **autumn** and **winter**.

Item3 in Figure 3.3 also shows the identical behaviour in two successive years. Its number of sales grows and falls linearly but there is only one month - June, when it is not sold at all. Therefore, Item4 a nonseasonal product.

An even better example of a product which is sold during the whole year is a Item4 which is sold nearly equally in each month, except February and October, when more sets were sold. This strongly indicates that it is a nonseasonal product.

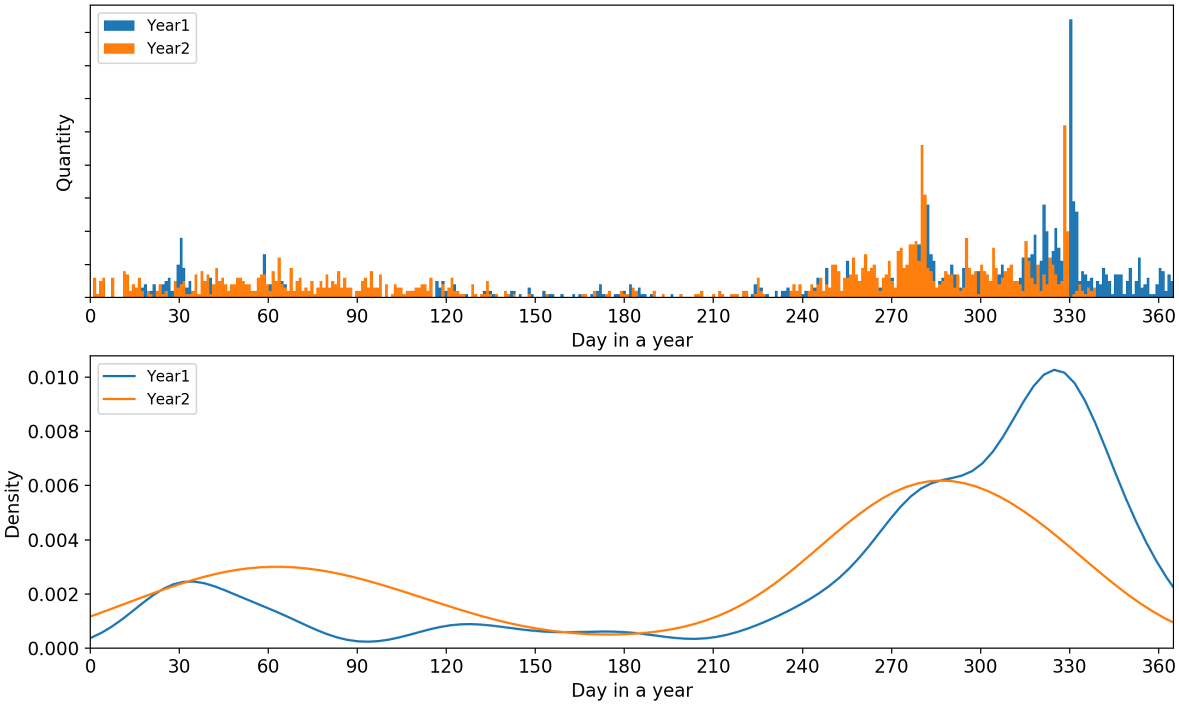


Figure 3.1: Item1: Histogram and distribution plot of sales in Year1 and Year2

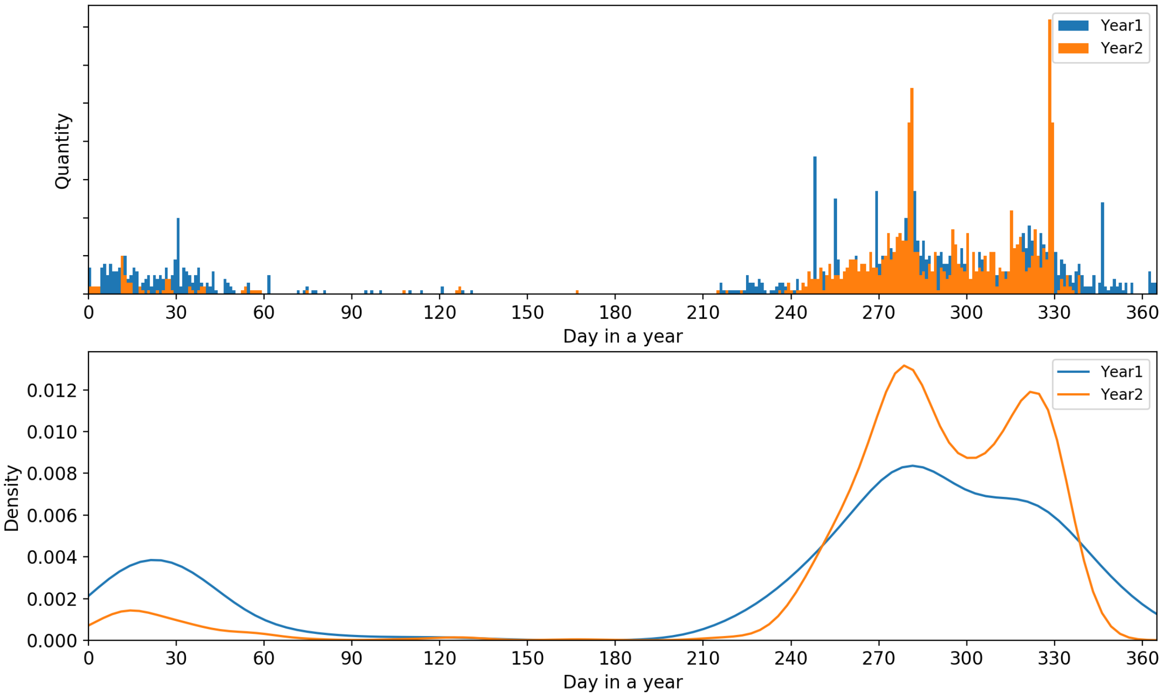


Figure 3.2: Item2: Histogram and distribution plot of sales in Year1 and Year2

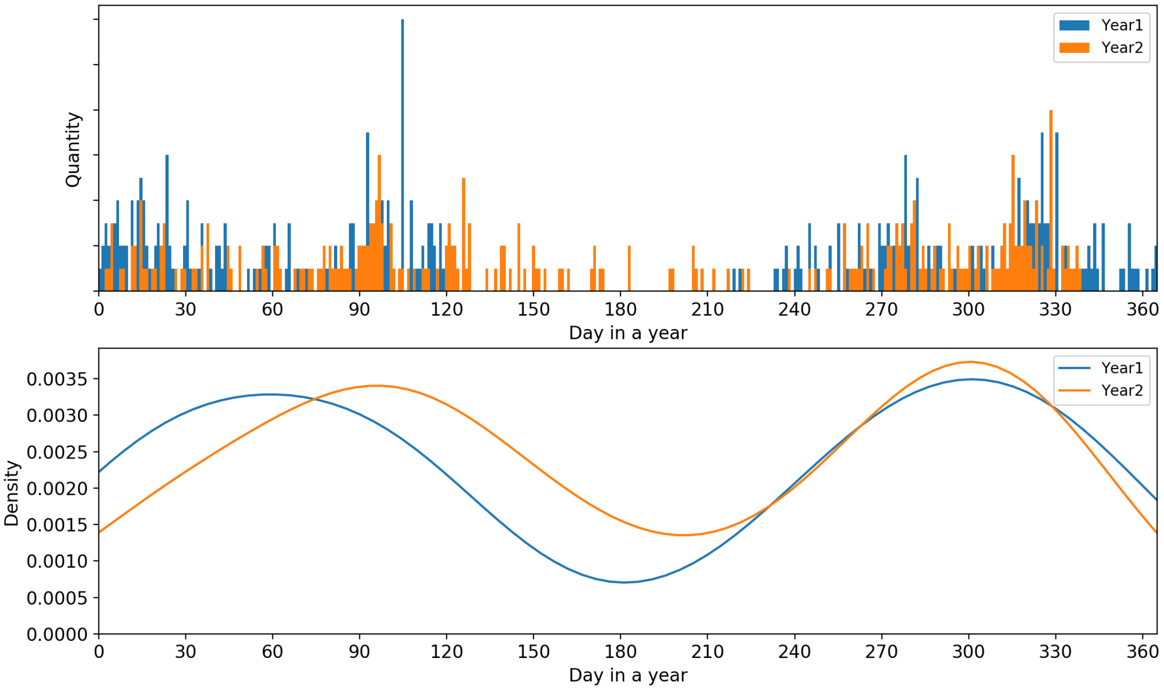


Figure 3.3: Item3: Histogram and distribution plot of sales in Year1 and Year2

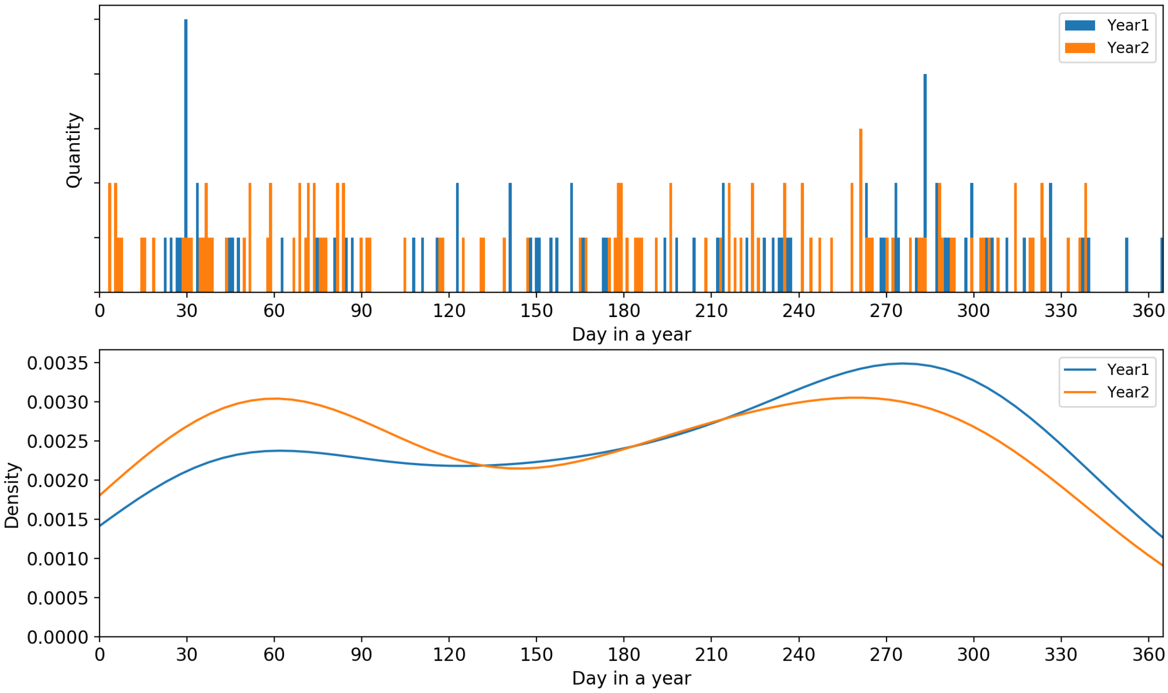


Figure 3.4: Item4: Histogram and distribution plot of sales in Year1 and Year2

3.4.3 Verification of the hypothesis' relevance

Before checking the relevance of the hypothesis, data need to be prepared for analysis. To make the results more accurate, only products which have been sold a *certain number of times* are observed.

For each product, a vector with 12 values is comuputed. Each value represents the portion of all purchases which was made in that month in two consecutive years. The portions are expressed using the unit interval (0-1). When calculating portions, two equations can be used (Equations 3.3 and 3.4). The first formula caluculates the portion for each product individually without taking into account the global sales in that month and the sales of all products. The second formula takes these factors into account. Therefore, the proof will be presented in two ways.

Once the data are computed, they can be used for evaluating the hypothesis' relevance. In order to verify the relevance, the distributions of all products need to be compared. The appropriate measure to compare distributions is the KLD, described in Section 2.1.

The proof for the hypothesis' relevance is conducted in two ways:

1. Product analysis without total sales impact
2. Product analysis including total sales impact

Product analysis without total sales impact

The analysis has four steps:

1. calculating monthly portions of sales for each product as:

$$\text{P(product,month)} = \frac{\text{sold(product,month)}}{\sum_{y}^{months}(\text{sold(product,y)})} \tag{3.3}$$

   and forming a twelve element vector where each month is an element

2. computing KLD values for all products by comparing them with the uniform distribution $(\frac{1}{12},..,\frac{1}{12})$, which represents a nonseasonal product sold throughout the year
3. separating them into two groups using the KLD value
4. observing the results for validity

The issue arrises in the second step when the threshold value of the KLD needs to be chosen. Setting the threshold to 0 would be too strict because almost no product is sold equally in each month. Setting it too high would make too many products nonseasonal. Therefore, calculations are done before choosing the threshold.

After calculating the KLD for each product, the histogram on the left side of Figure 3.5 is produced. The cut-off value for the probability is 0.17 (see Figure 3.5) which coresponds to the initial test value. 30% of the products fall below the threshold of 0.17 which means they are nonseasonal.

According to values of KLD, out of the four products in figures 3.1, 3.2, 3.3 and 3.4 two are correctly classified as nonseasonal, one was correctly classified as seasonal, and one was incorrectly classified as nonseasonal.

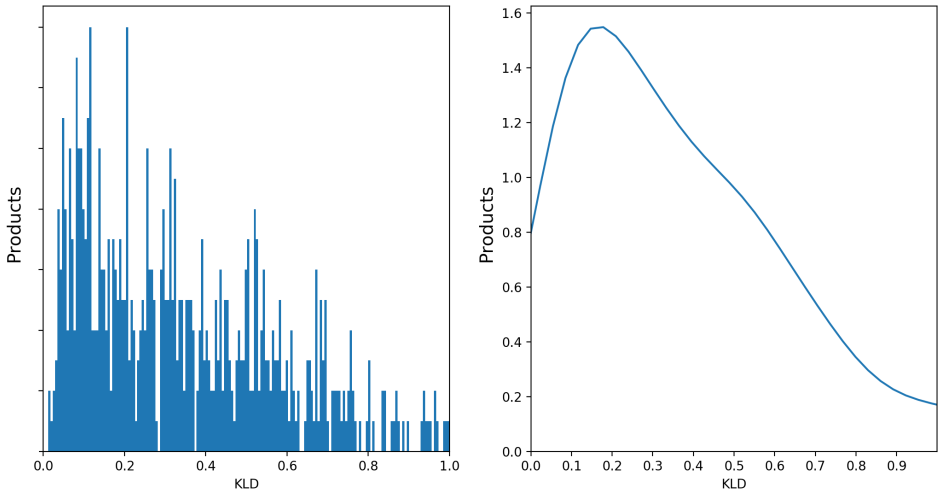
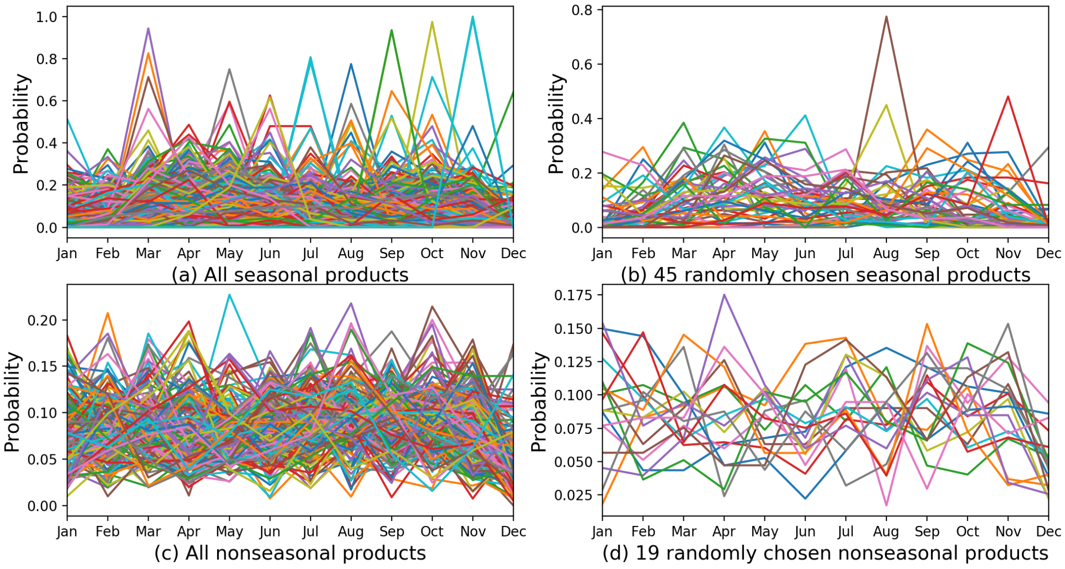


Figure 3.5: KLD histogram and distribution plot

Plotting distributions manually one-by-one to check if they are correctly classified would be an exhaustive and time-consuming task. Instead, both groups of products have been plotted in one graph each and the results are shown in Figure 3.6.

Additionally, a boxplot graph for nonseasonal products is shown in Figure 3.7.



(a) All seasonal products

(b) 45 randomly chosen seasonal products

(c) All nonseasonal products

(d) 19 randomly chosen nonseasonal products

Figure 3.6: Seasonal and nonseasonal product sales probability plots after classification

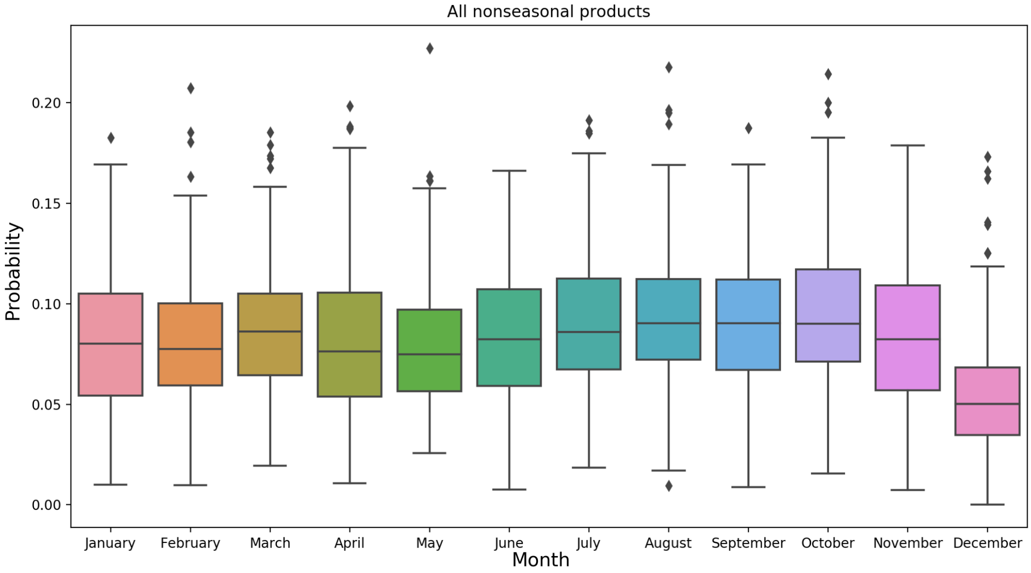


Figure 3.7: Nonseasonal products sales probability boxplot

From the results shown in the figures, it is visible that the majority of products are correctly classified. Nonseasonal products all have distribution values greater than zero and they oscillate around the middle value, which is 0.1 and amounts to 10%. On the other hand, seasonal products are much closer to zero.

This is what we expected and even though the value of the threshold might be different for each shop, the presented results are sufficient to conclude that the division of products by seasonality is possible.

Product analysis including total sales impact

The analysis including total sales impact differs from the previous analysis only in the first step:

1. calculating monthly portions of sales for each product as:

   $$\text{Probability(product,month)} = \frac{\text{sold(product,month)}}{\sum_{y}^{months}(\text{sold(product,y)})} \cdot \frac{\sum_{x,y}^{products,months}\text{sold(x,y)}}{\sum_{x}^{products}(\text{sold(x,month)})} \quad (3.4)$$

   normalising them, and forming a twelve element vector where each month is an element

2. computing KLD values for all products by comparing them with the uniform distribution ($\frac{1}{12}$,...,$\frac{1}{12}$) which represents a nonseasonal product sold throughout the year

3. separating them in two groups using the KLD value

4. observing the results for validity

As shown in Equation 3.4, the old proportion is multiplied with a ratio consisiting of *the total number of items sold* and *total number of items sold in that month.* The idea behind it is to observe the product relative to all sales made in that month. This means that if some product is sold only a few times during one month, but sales in that month were worse than usual, then the portion should be higher. In the same way, if it is a month with many promotions or events such as Black Friday, then the portion of that product should be smaller since all products sold better than usual.

In step two, new KLD values are calculated. The values for the KLD are almost the same as in the first proof up to the third or fourth decimal place. As seen in Figure 3.8, the threshold value for seasonal and nonseasonal products has not changed. It remains 0.17. The classification of products in the first and second proof differs (a product was seasonal before and is nonseasonal now and vice versa) in the case of 8% of the total number of products.

After dividing products into seasonal and nonseasonal groups again, each group is plotted separately. The result is shown in Figure 3.9.

Additionally, another boxplot for nonseasonal products is presented in Figure 3.10.

In the boxplot diagram, some “boxes”, i.e., months have a wider interquartile range than other months. This is a result of the multiplication with a factor from Equation 3.4. The biggest difference can be observed in December.

The factor for each month is shown in Figure 3.11. By using multiplication factors, a probability correction is made. As an example, if the least number of items is sold in December, then dividing the number of all items sold with the ones sold in December gives the biggest factor. This means that if an observed item X has sold less in December than in any other month, it still has a chance to be a nonseasonal product after multiplication because December had had the lowest sales in the observerd set.

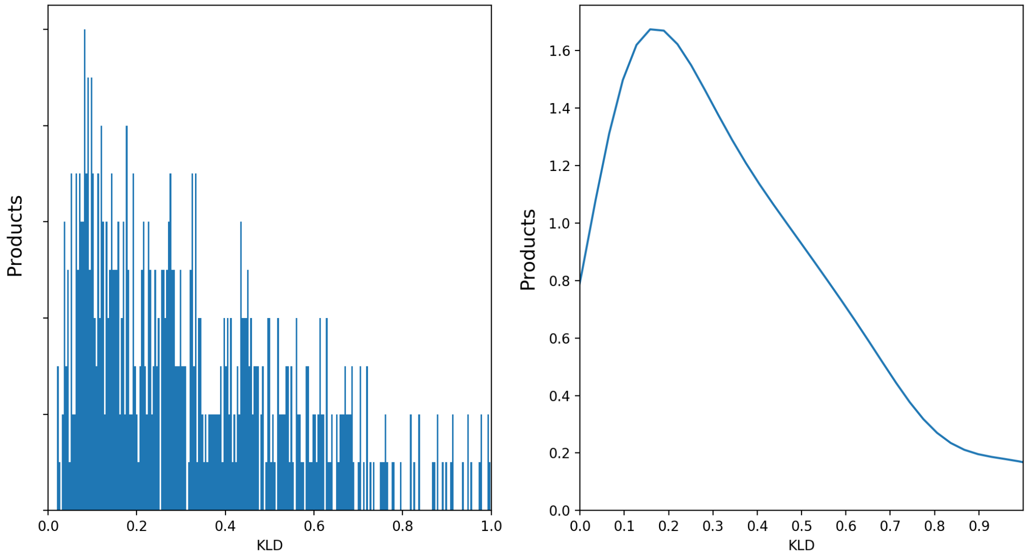
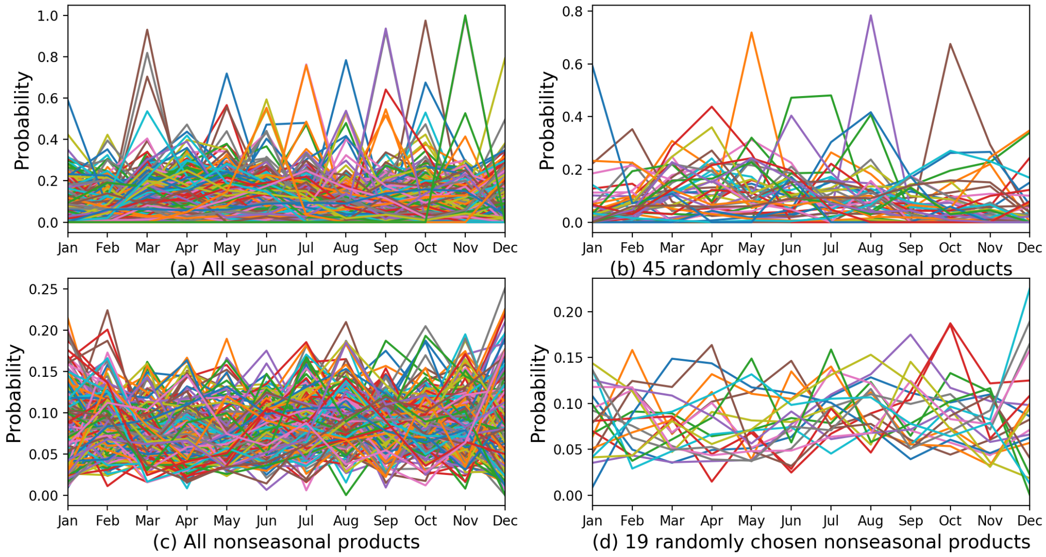


Figure 3.8: KLD histogram and distribution plot



(a) All seasonal products

(b) 45 randomly chosen seasonal products

(c) All nonseasonal products

(d) 19 randomly chosen nonseasonal products

Figure 3.9: Seasonal and nonseasonal product sales probability plots after classification

To summarise, KLD has proved as a good indicator when observing the seasonality of a product. Two proofs have been performed to show that there are indeed seasonal and nonseasonal products. In the first proof, products have been analysed independent of the shop sales in general. In the second proof, products' distributions over months has been improved by multiplying with a correctional factor. Since only products that have been sold a *certain number of times* are considered, only about approximately 1% of all items in both shops are included. This is, however, not an issue since the sorting needs to be optimised for the few first pages.

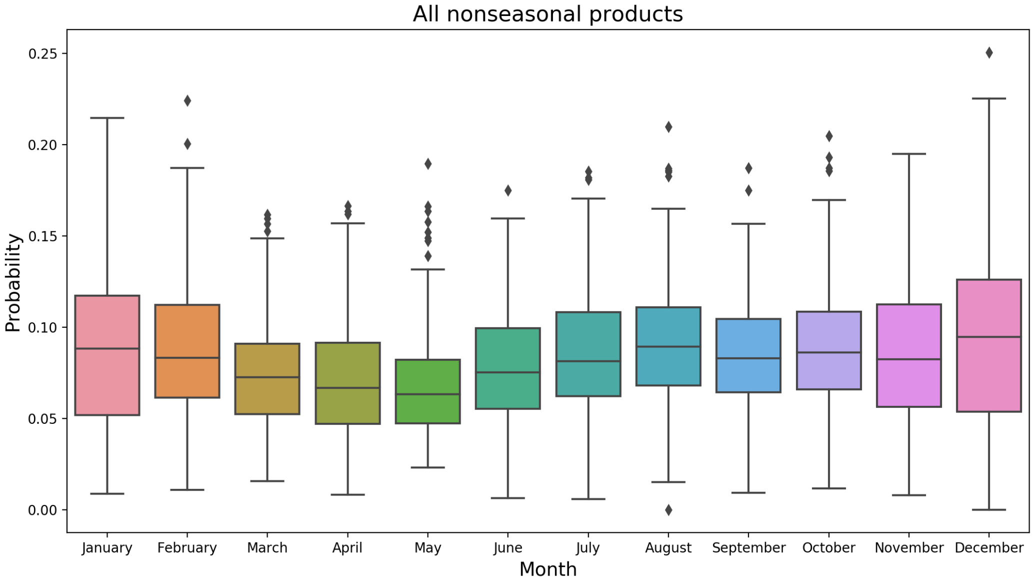


Figure 3.10: Nonseasonal products sales probability boxplot

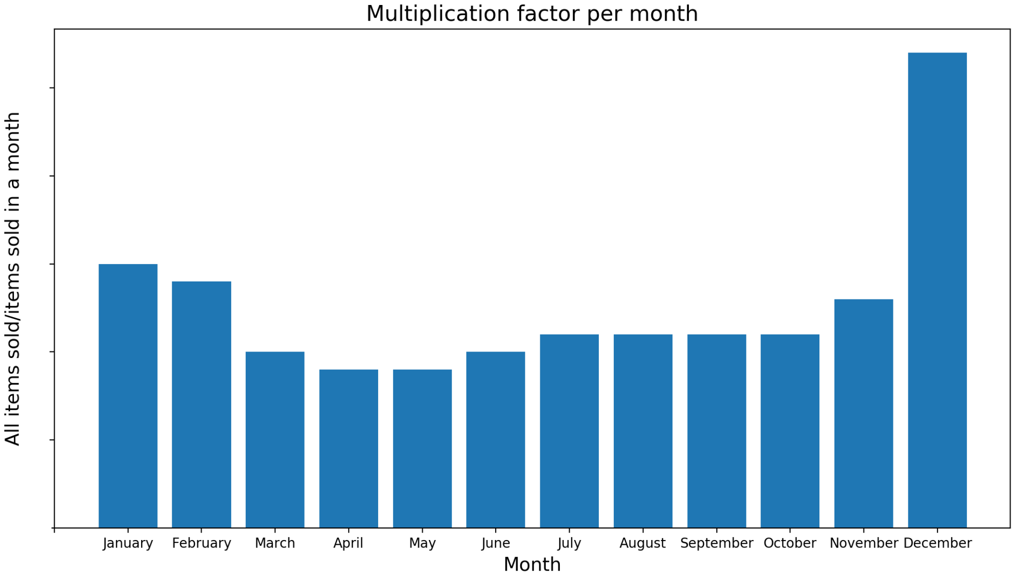


Figure 3.11: Product sales mutliplication factor for each month

3.5 Time of the day with respect to products

Hypothesis: The product's sales depend on the time of the day. Some products are rarely sold at specific hours.

3.5.1 Implications of the hypothesis

If the hypothesis is correct, depending on the time of the day when a customer is visiting the website, products that sell the best at current time of the day will be shown at first pages. **Additionally, if products have never been sold at that time, they will be pushed to the end of the list.**

It is usually the type of the product that affects the time at which the product is best sold. If a customer visits the 'Pants' category page at 2 pm and working pants are best sold at that time, they will be shown on the first few pages.

3.5.2 Examples confirming the hypothesis

Three producs are analysed on:

- *hourly basis sales*
- *hour range basis sales*

Hour ranges are defined for each period of the day. Their values are chosen with help of Figure 3.12.

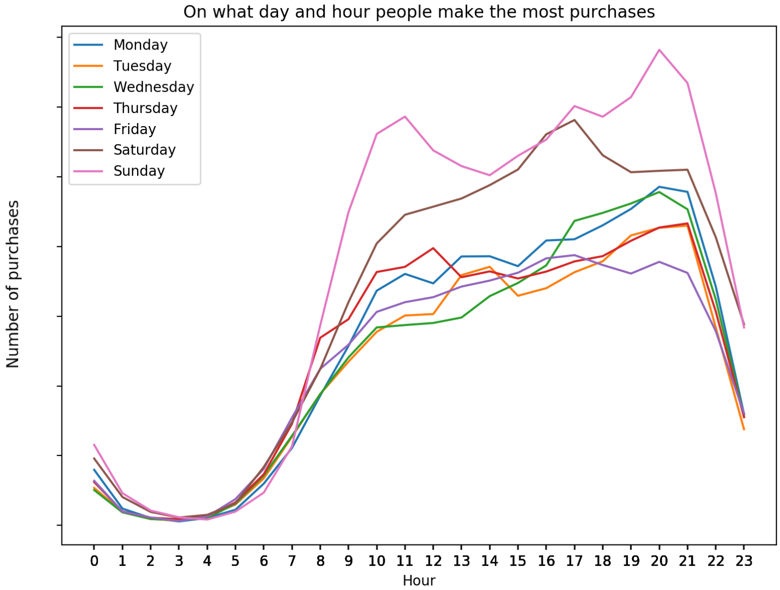


Figure 3.12: Overview of product purchases per hour across each day of the week in one year

As expected, people buy less during the late night. Around 7:00, they wake up and start shopping. Purchases keep increasing until 12:00, which is lunchtime for most working people. From 12:00 until 20:00, the number of purchases oscillates but does not rise or fall significantly. Then, after 20:00 in the evening we can see a constant fall that continues until 1:00 when customers seldom purchase.

Therefore, following ranges are used:

Time of the day	**Night**	**Morning**	**Day**	**Evening**
Hour range	[1:00-06:59]	[07:00-11:59]	[12:00-19:59]	[20:00-00:59]

Table 3.1: Hour ranges for each period of the day

The hourly based analysis

The hourly based analysis uses histograms and distribution diagrams (see Figures 3.13, 3.14 and 3.15). The histograms show the number of items bought *at each hour*, while the distribution diagrams show the portions of items bought *at each hour*. The sales portion is calculated as:

$$\frac{\text{sold(product,hour)}}{\sum_{y}^{hours}\text{sold(product,y)}} \tag{3.5}$$

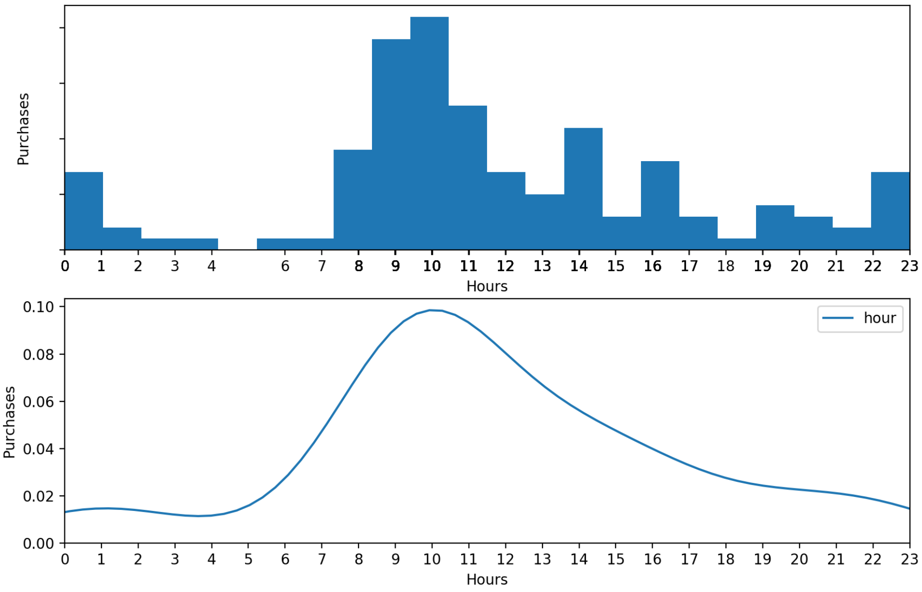


Figure 3.13: Item4: Histogram and distribution plot of sales over the course of a day

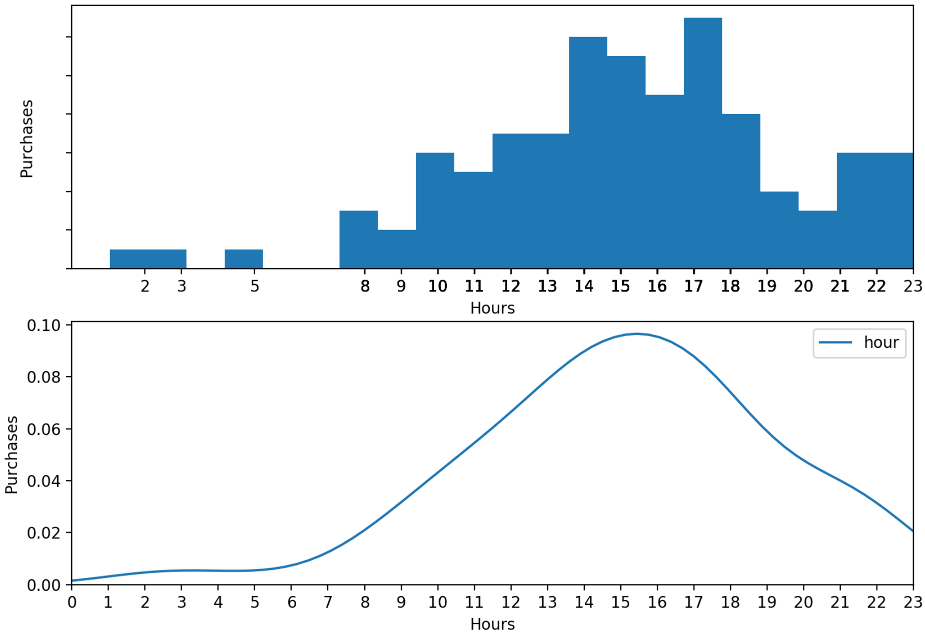


Figure 3.14: Item5: Histogram and distribution plot of sales over the course of a day

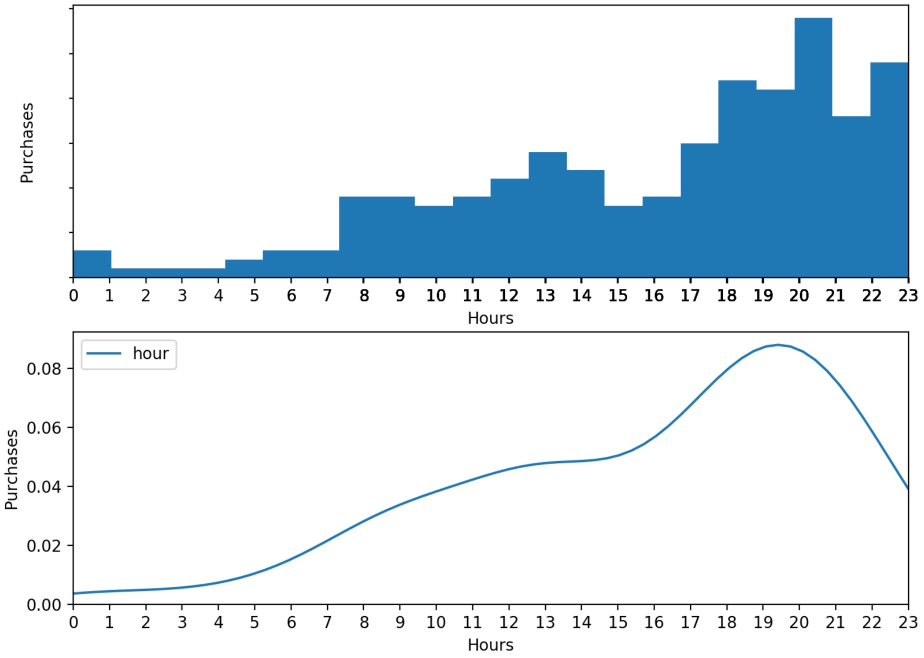


Figure 3.15: Item6: Histogram and distribution plot of sales over the course of a day

By observing distribution diagrams with *hourly portions*, the first product can be classified as a **morning** product, the second one as a **day/evening** product and the third one as an **evening** product. Some products are never sold at specific hours which confirms the second part of the hypothesis. The disadvantage of the analysis using the given portion is that a low number of purchases during certain hours might be the result of overall low numbers of purchases at that hour. This will be corrected in the *hour range* based analysis.

The hour range based analysis

Since there are only four hour ranges, the hour range based analysis uses bar diagrams. Bar diagrams are plotted using two different equations for sales proportions:

- sales portion without normalisation

$$\frac{\text{sold(product,hourRange)}}{\sum_{y}^{hourRanges}\text{sold(product,y)}} \tag{3.6}$$

- sales portion with normalisation

$$\frac{\text{sold(product,hourRange)}}{\sum_{y}^{hourRanges}\text{sold(product,y)}} \cdot \frac{\sum_{x,y}^{products,hourRanges}\textbf{sold(x,y)}}{\sum_{x}^{products}\textbf{sold(x,hourRange)}} \tag{3.7}$$

Normalisation is an attempt to calculate the scores relative to all sales made at that hour range. This means that if a product is sold *a certain number* of times during the day but 10 times during the night, it still might be a night product since there are less products sold at night anyways.

Results are presented in the Table 3.2. For each product and hour range, sales portions and normalised sales portions (bold font) are shown. Sales portion represent the portion of the product sold during that hour range up till now.

Product	**Night**	**Morning**	**Day**	**Evening**
Item4	0.06 **(1.07)**	0.46 **(1.9)**	0.33 **(1.75)**	0.14 **(0.14)**
Item5	0.04 **(1.38)**	0.42 **(1.35)**	0.37 **(0.82)**	0.17 **(0.82)**
Item6	0.02 **(0.73)**	0.17 **(0.55)**	0.5 **(1.05)**	0.32 **(1.6)**

Table 3.2: Product list with fractions of sales for times of the day

It is very important to notice how each pair of bar diagrams (see Figures 3.16, 3.17 and 3.18) shows different portions, which means that a product's class most likely changes from after multiplication of the initial portion with a factor. E.g. by reading the first diagram in Figure 3.18, one could conclude that this is a product sold mostly during the **Day**. However, after multiplication, i.e., normalisation this product is classified as a product sold mostly during the **Evening**. The overview of class changes can be seen in the Table 3.3. For each product, there are three different labels depending on the time range is used of labelling. For the first column, three diagrams in Figures 3.13, 3.14 and 3.15 are observed while the labels for the second and third column are read from the bar diagrams in Figures 3.16, 3.17 and 3.18.

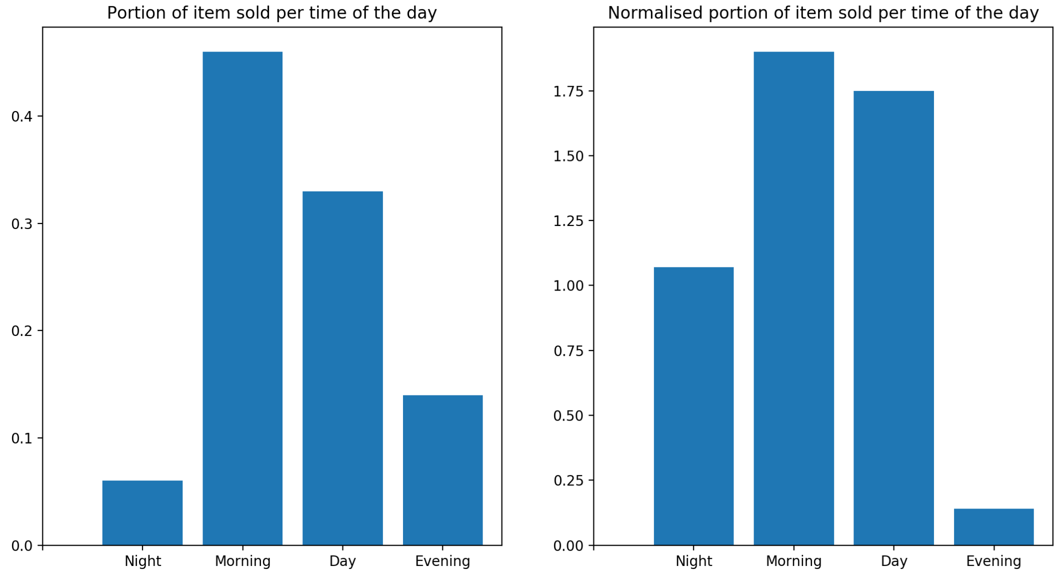


Figure 3.16: Item4: Portions and normalised portions of sales across the hour ranges

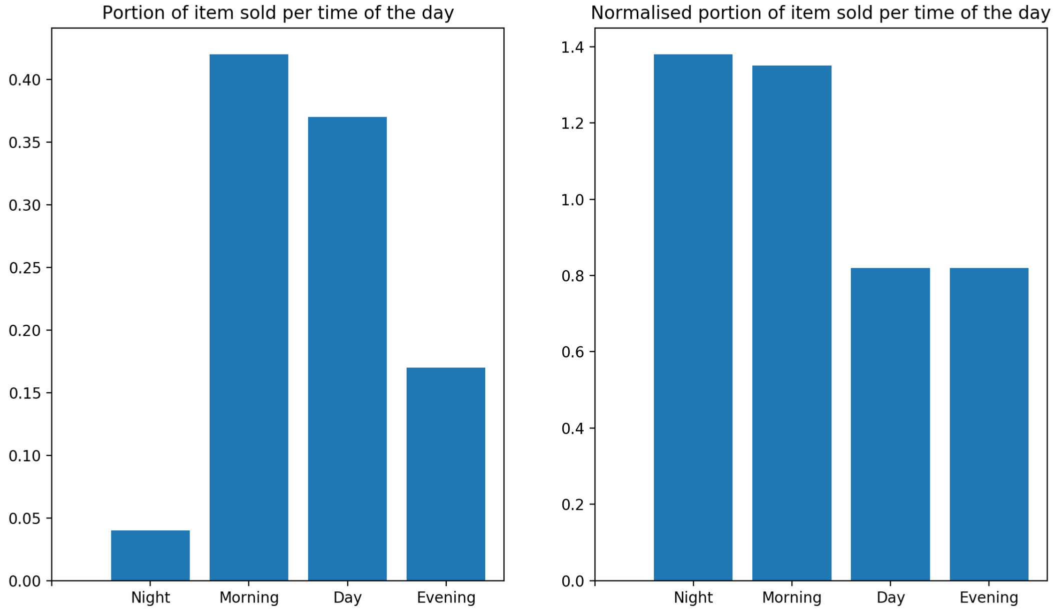


Figure 3.17: Item5: Portions and normalised portions of sales across the hour ranges

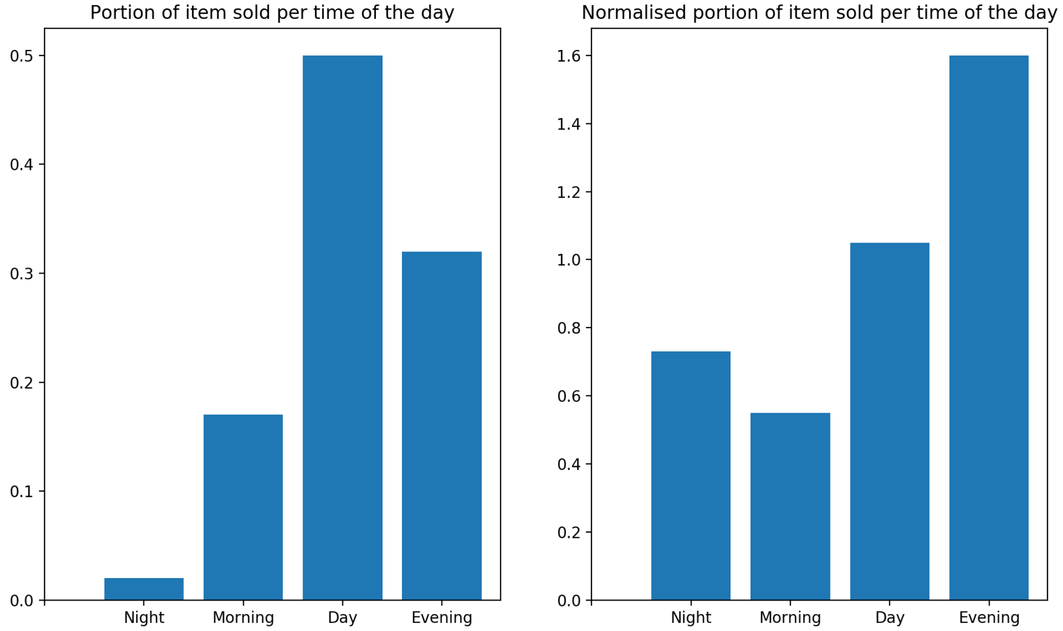


Figure 3.18: Item6: Portions and normalised portions of sales across the hour ranges

Product	**Class(Hours)**	**Class(Hour ranges)**	**Class(Hour ranges-normalised)**
Item4	Morning	Morning/Day	Morning/Day
Item5	Day/Evening	Morning/Day	Night/Morning
Item6	Evening	Morning/Day	Night

Table 3.3: Product list with classifications for the times of the day with the highest sales for the product

This section demonstrated *hour based* and *hour range* based analysis to show how product sales vary throughout the day. Since hour based analysis is more precise it will be further used, but it will be normalised. The hour ranges based analysis was useful to show why normalisation of sales portions matter and its influence on product clasess.

In the next section, the hypothesis' relevance for all products will be tested using KLD.

3.5.3 Verification of the hypothesis' relevance

Once again, to make the results more accurate, only products which have been sold more than a *certain number of times* are observed. Since it is already proven that normalisation does not change KLD values (see Section 3.4.3), all sales portions will be normalised. For precise results, portions will be calculated *per hour*. Products will be be classified as **non-all-day** or **all-day** products, depending on whether they are sold mostly during some hours or equally well during all hours respectively.

Analogous to the seasonality hypothesis' verification, the following steps are performed:

1. calculating monthly portions of sales for each product as:

$$\frac{\text{sold(product,hour)}}{\sum_{y}^{hours}\text{sold(product,y)}} \cdot \frac{\sum_{x,y}^{products,hours}\textbf{sold(x,y)}}{\sum_{x}^{products}\textbf{sold(x,hour)}} \tag{3.8}$$

   normalising them, and forming a twelve element vector where each hour is an element

2. computing KLD values for all products by comparing them with the uniform distribution $(\frac{1}{24},\ldots,\frac{1}{24})$, which represents a product sold equally throughout the day

3. separating them in two groups using the KLD value

4. observing the results for validity

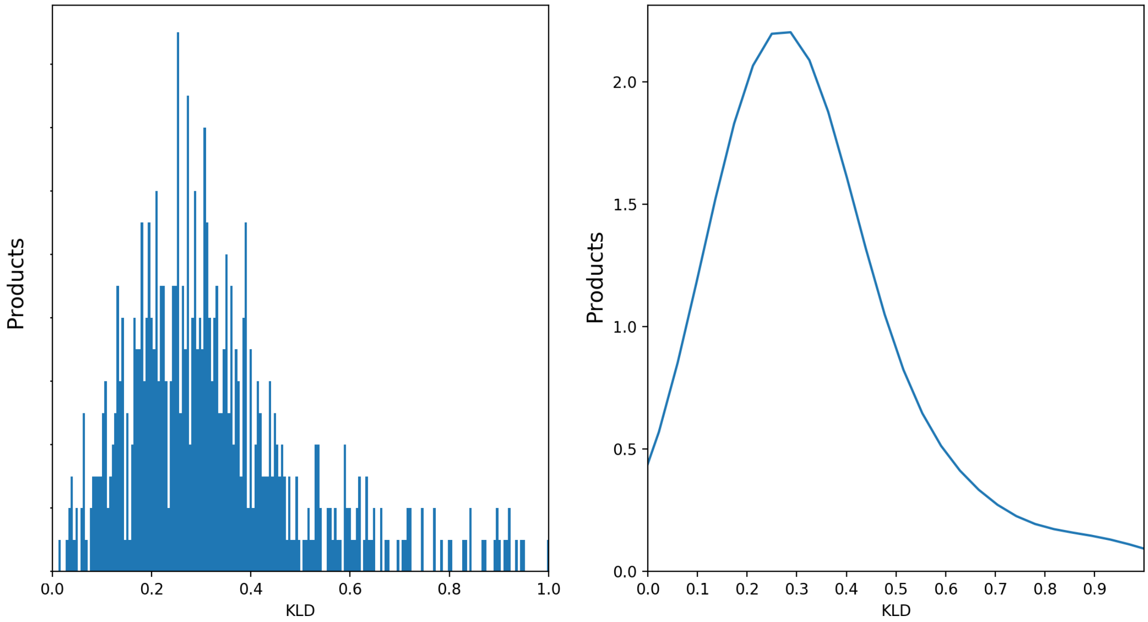


Figure 3.19: KLD histogram and distribution plot

After all the KLDs are calculated, a threshold for product classification needs to be found by plotting KLD values. The plot can be seen in Figure 3.19. The threshold chosen is 0.27 because the peak which clearly separates the products into two groups appears at that value.

Every product which has a KLD <= 0.27 is classified as an **all-day product**, or as a **non-all-day product** otherwise.

For the purpose of results verification, all products are plotted with their portions in Figure 3.20.

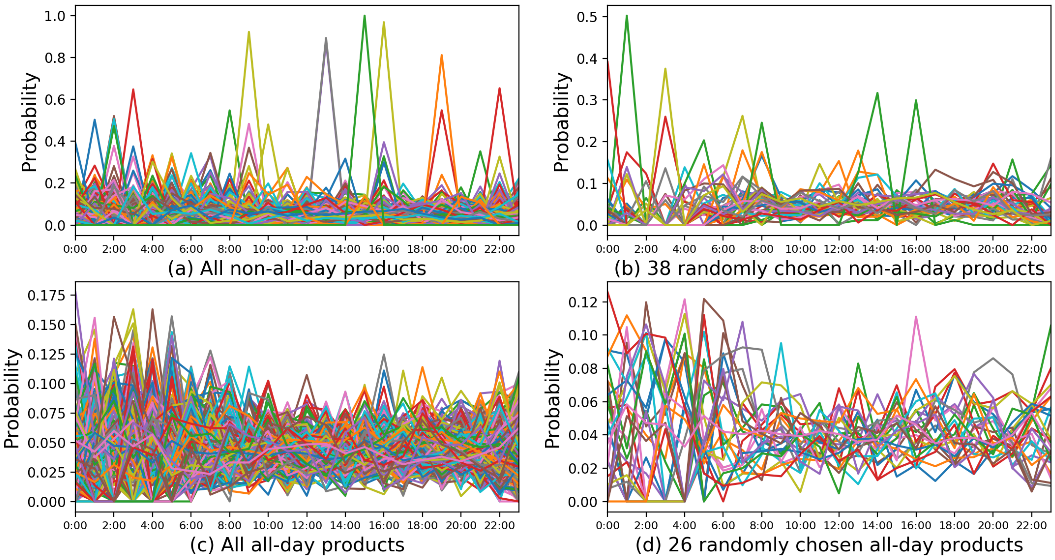


Figure 3.20: Non-all-day and all-day sales probability plots after classification

The difference between the product plot for months (Figure 3.9) and product plot for hours is that there are bigger oscillations when plotting hours (Figure 3.20), and all-day products (Figure 3.20 (c) and (d)) do not appear as an equally wide strip, which was the case with nonseasonal products. The reason lies in the fact that the number of sales per hour is very different for each hour, unlike the number of sales for months. The best example are night hours when customers sleep and seldom purchase. On the other hand, in most shops there is no month when very few customers are purchasing.

This night effect can be better seen in the boxplot for all-day products in Figure 3.21, where the height of boxes, i.e., hours significantly differ. They are the biggest for the night hours from 1:00 until 4:00. In the same way portions for night hours will be adjusted equivalently. What this means is that some products which sell poorly during the day might be labelled as selling well at night, even with the same amount of items sold, because sales at night are smaller.

Another point worth noticing is the median of all hours, which is now between 0.02 and 0.04. In the months boxplot it was closer to $0.1 = 0.05 \cdot 2$. This is correct and expected due to the reference KLD vector which is $(\frac{1}{12}, ..., \frac{1}{12})$ for months and $(\frac{1}{24}, ..., \frac{1}{24})$ for hours.

In this section, it was proven that KLD can be used for differentiating between products sold throughout the day and products sold very well only during certain hours,

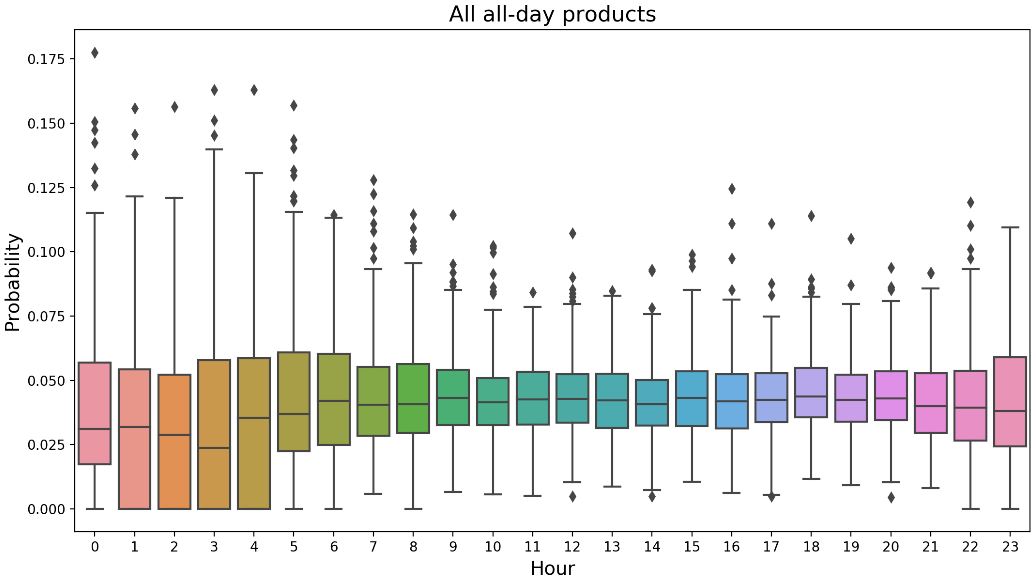


Figure 3.21: All-day products sales probability boxplot

and that products can be classified in this way. It is also shown that using a correctional factor is important for the quality of results. These findings will be used to explain the impact on the sorting algortihm and to construct the sorting algorithm.

3.6 iPhone/Apple users vs others

As observed in studies from Section 2.5.6, there is a difference in buying habits of iPhone and Android users. This thesis explores the hypothesis about iPhone and other users on two different shops. The hypothesis is formulated and explored in four versions:

- Ia: iPhone users spend more money than other users
- Ib: Apple users spend more money than other users
- IIa: iPhone users purchase greater quantities of expensive products than other users
- IIb: Apple users purchase greater quantities of expensive products than other users

The latter two hypotheses are important because a user who spends a great deal of money might be spending it on many cheap products. In such cases, the hypothesis could not be used directly to optimise the sorting.

3.7 Hypothesis Ia: iPhone users spend more money than other users

3.7.1 Implications of the hypothesis

If the hypothesis is true, it is more likely that iPhone users are willing to spend money for higher quality products and therefore more expensive products should be ranked higher than cheaper ones. This will be supported or rejected depending on the hypotheses IIa and IIb.

3.7.2 Examples confirming the hypothesis

All users are divided in two groups: iPhone users vs other users. A user is considered to be an iPhone user if at least one order was placed on an iPhone (identified by the user agent).

The first value which could contribute to the hypothesis is the Average Order Value (AOV). AOV is a the result of dividing revenue by the number of orders:

$$AOV = \frac{\sum_i^{orders}(totalprice(i))}{\text{number of orders}} \tag{3.9}$$

The AOV is calculated for each customer and then for each of the two groups: iPhone users and others. A group AOV is the average of all belonging customers' AOVs. The average values and medians of the group AOV are presented in Table 3.4.

From the results in Table 3.4 one could quickly conclude that in Shop1 iPhone users spend more money and that in Shop2 iPhone users spend less money than others.

AOV	**Shop1**			**Shop2**		
	Average	**Median**	**Users**	**Average**	**Median**	**Users**
iPhone users	569€	518 €	> 20.000	476€	417€	> 100.000
Others	525 €	458 €	> 200.000	492€	438€	> 400.000

Table 3.4: Average Order Value: iPhone users vs. others

Nevertheless, since this excerpt of data is not a valid proof, this conclusion might not be correct.

E.g. If 50% of iPhone users spend vastly more money and the other 50% spend very little, the average might be relatively high but it is not true that majority of iPhone users spend lof of money.

In other words, it is necessary to investigate the differences within both groups before making a conclusion about differences between groups. This will be done in the next section with the help of statistics.

3.7.3 Verification of the hypothesis' relevance

The goal of this subsection is to verify the findings from the previous subsection, i.e., that iPhone users spend more than others in Shop1 and less than others in Shop2 and if this difference between two groups is **significant**. Before chosing the method for testing the hypothesis, the following factors need to be considered:

- **Data:** There are two groups of data: AOV for iPhone users and AOV for others. Since AOV is an average, the type of data is a *Ratio*.
- **Data distribution**: Data are *not normally distributed* which can be seen in Figure 3.22. In Shop1, there are more iPhone users at every AOV value after approximately point B, which indictes that they spend more. In Shop2, it is the other way round where Other users spend more after approximately point A. Apart from the two trehsold values (A and B), the shapes of the graphs are very similar. This does not come as a surprise since majority of customers buy either cheap or medium expensive products.
- **Samples:** *Two samples* are compared to each other: iPhone users vs others
- **Purpose:** The purpose of testing is to *compare two statistics*, i.e., groups

Based on the listed factors, the appropriate test for the hypothesis is a test for the difference between two means not requiring the normal distibution. A test satisfying these conditions is the **Mann-Whitney U test** described in Section 2.2. A precondition for this test is that data is capable of being ranked and since averages are being analysed, this precondition is satisfied.

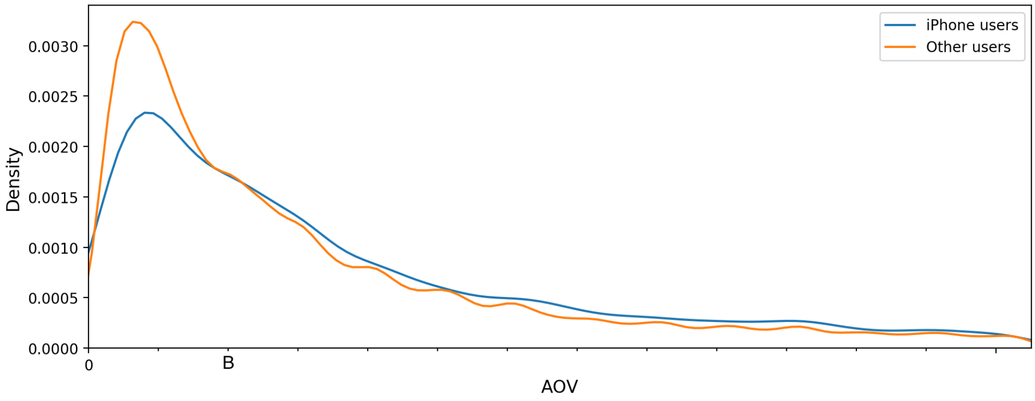


Figure 3.22: Shop 1 - AOV Distribution plot: iPhone users vs Others

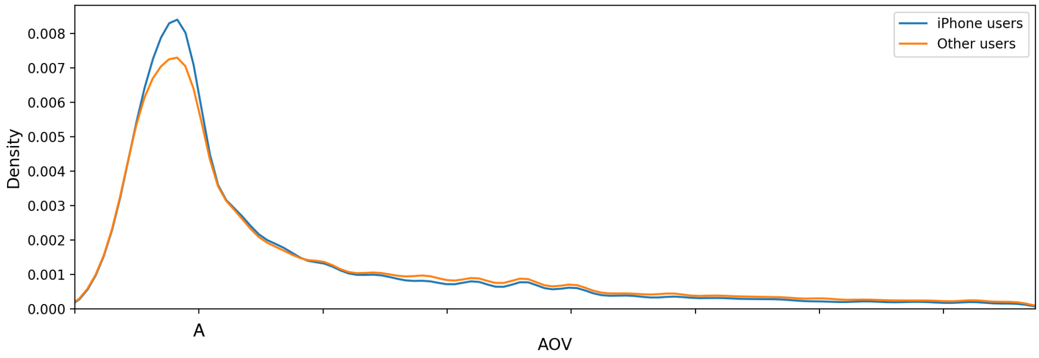


Figure 3.23: Shop 2 - AOV Distribution plot: iPhone users vs Others

Mann-Whitney U test

"*Mann-Whitney U test ... is a nonparametric test of the null hypothesis that it is equally likely that a randomly selected value from one sample will be less than or greater than a randomly selected value from a second sample.*" [Man]

The test will be conducted in the following six steps:

Step 1: State the null and alternative hypothesis

Step 2: Choose the level of significance

Step 3: Determine the critical value

Step 4: Compute the test statistic

Step 5: Analyse results

Step 6: Conclusion

Step 1: State the null and alternative hypothesis

Null-Hypothesis (H_0): iPhone users and other users spend equal amounts of money on average.

Alternative-Hypothesis (H_1): iPhone users spend significantly more or less on average than other users.

This is a two-tailed test because "*the null hypothesis will be rejected if the difference between sample means is too big or if it is too small*". [Sta] This type of test is chosen over a one-tailed test because the iPhone users of one shop spend more and the iPhone users of other shop spend less accroding to the AOV average.

Step 2: Choose the level of significance α

The level of significance α is calculated as:

$$\alpha = \frac{1 \ - \text{confidence level}}{2} \tag{3.10}$$

for two-tailed tests.

The acceptable level of confidence in e-commerce is 90% . Thus:

$$\alpha = \frac{1 \ - 0.9}{2} = 0.05 \tag{3.11}$$

Step 3: Determine the critical value

The test statistic for the Mann-Whitney test is a U value. Every U value has an associated p value which indicates whether the difference between two groups is statistically significant. The meaning of the p value can be found in Section 2.2.

Step 4: Compute the test statistic

The function used for testing is a Python function:

scipy.stats.mannwhitneyu(x, y, use_continuity=True, alternative=None)

where x and y are arrays of samples, *user_continuity = True* refers to the continuity correction and *alternative* decides if the p-value for will be calculated for the one-sided hypothesis or the two-sided hypothesis.

According to the null- and alternative-hypotheses, the following parameters are chosen:

- x = iPhone users AOV samples
- y = Other users AOV samples

- alternative = *two-sided*

After running tests with help of the *mannwhitneyu* function, the following results are observed:

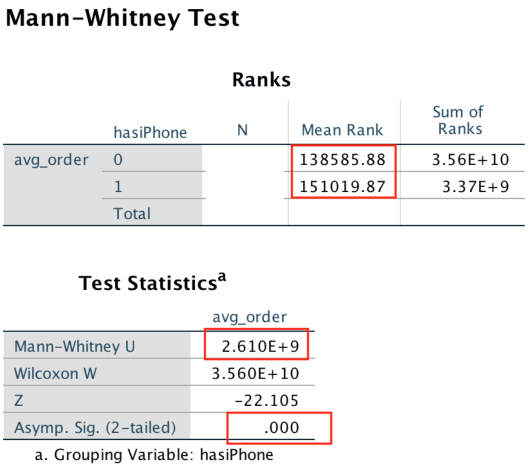
Shop	Statistic U	P-value
Shop1	3.1e9	2.8e-108
Shop2	2.7e10	1.1e-128

Table 3.5: Results of the Mann Whitney U test: iPhone users vs Others

Step 5: Analyse results

Since $p \leq \alpha$ in both shops, the difference between the iPhone users' and other users' medians is statistically significant.

In order to double-check the results and find out which distribution is significanlty greater, the data were also tested with IBM SPSS Program for Statistics and the results are shown in Figure 3.26.



Mann-Whitney Test

Ranks

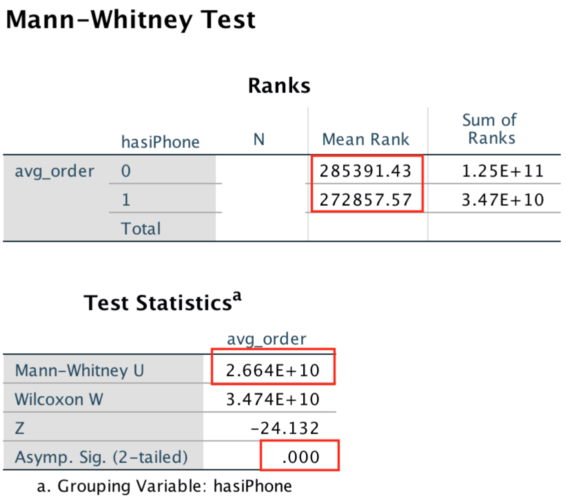
	hasiPhone	N	Mean Rank	Sum of Ranks
avg_order	0		138585.88	3.56E+10
	1		151019.87	3.37E+9
	Total			

Test Statistics[a]

	avg_order
Mann-Whitney U	2.610E+9
Wilcoxon W	3.560E+10
Z	-22.105
Asymp. Sig. (2-tailed)	.000

a. Grouping Variable: hasiPhone

Figure 3.24: Shop1



Mann-Whitney Test

Ranks

	hasiPhone	N	Mean Rank	Sum of Ranks
avg_order	0		285391.43	1.25E+11
	1		272857.57	3.47E+10
	Total			

Test Statistics[a]

	avg_order
Mann-Whitney U	2.664E+10
Wilcoxon W	3.474E+10
Z	-24.132
Asymp. Sig. (2-tailed)	.000

a. Grouping Variable: hasiPhone

Figure 3.25: Shop2

Figure 3.26: IBM SPSS Mann-Whitney Test Results: iPhone users vs Others

Once again, $p \leq \alpha$. The p values - denoted by Asymp. Sig (2-tailed) in Figure 3.26 - are the same as in Python with fewer decimal places and the *Mann-Whitney U* values differ from the Python values because there is no continuity correction in SPSS. Users who have iPhones are denoted by hasiPhone = 1. The mean rank for iPhone users is bigger than the mean rank for Others only in the first shop while it is vice verse in the second shop. The mean rank indicates which group has a greater average and this will be used in the conclusion.

Step 6: Conclusion

Shop1: A Mann-Whitney U test indicated that the Average order value was significantly greater for iPhone users (Mdn = 518€) than others (Mdn = 458€) (Mann–Whitney U = 3120116701.5, m > 20.000, n > 200.000, P ≪ 0.05 two-tailed).

Shop2: A Mann-Whitney U test indicated that the Average order value was significantly smaller for iPhone users (Mdn = 417€) than others (Mdn = 438€) (Mann–Whitney U = 26635989282.5, m > 100.000, n > 400.000, P ≪ 0.05 two-tailed).

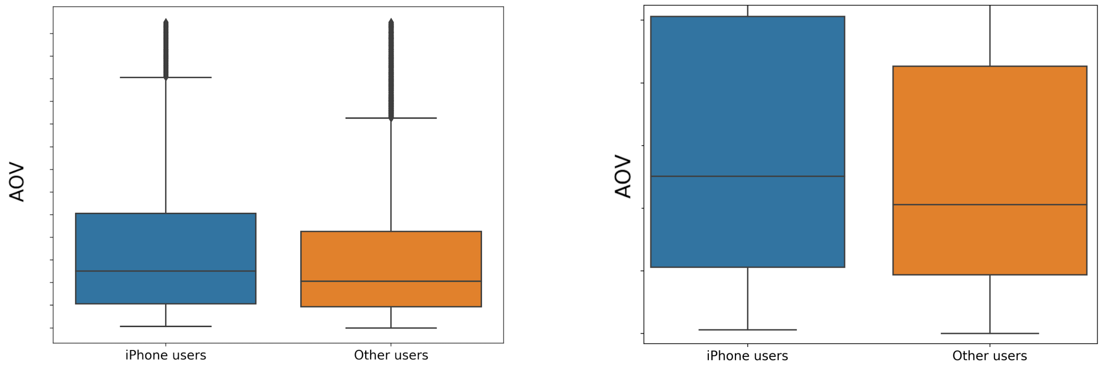


Figure 3.27: Shop1 - Boxplot of AOV of iPhone users vs Others

To understand the results better, two boxplot diagrams for Shop1 and Shop2 are shown in Figures 3.27 and 3.28. As expected, for the Shop1, the values in the iPhone box are higher than in the Others box and for the Shop2 vice versa.

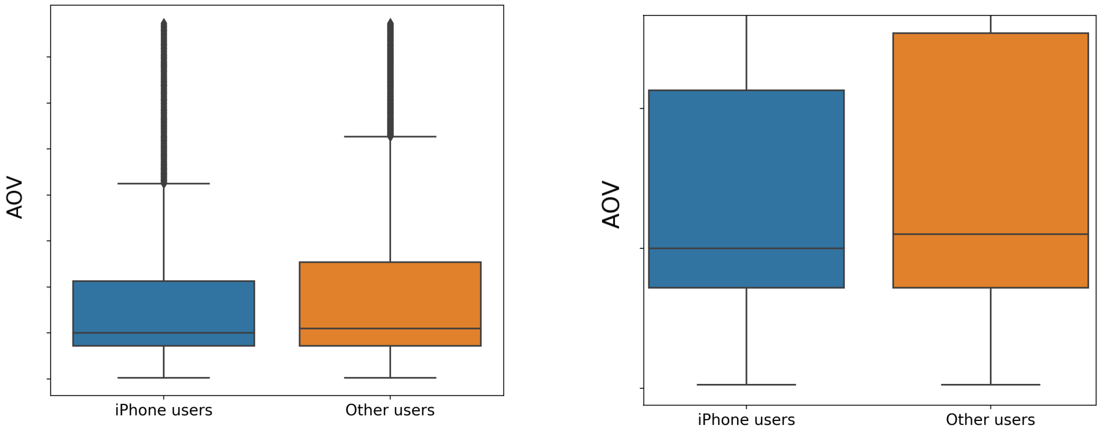


Figure 3.28: Shop2 - Boxplot of AOV of iPhone users vs Others

In this subsection, the statistical Mann-Whitney test was used to calculate the significance of the differences between iPhone and other users' AOV. The Hypothesis Ia:

iPhone users spend more money than other users was indicated as true only in the one shop.

In the next section, the AOV of Apple and other users will be investigated in order to check if Apple users spend more on average in both shops since Apple users spend more only in one shop.

3.8 Hypothesis Ib: Apple users spend more money than other users

This section does an analogous examination to Section 3.7 with the difference of iPhone users being replaced by Apple users.

3.8.1 Implications of the hypothesis

If the hypothesis is true, it is more likely that Apple users are willing to spend more money and therefore more expensive products should be ranked higher for them than cheaper ones. This will be supported or rejected depending on the hypotheses IIa and IIb.

3.8.2 Examples confirming the hypothesis

All users are divided in two groups: Apple users vs other users. A user is considered to be an Apple user if it has placed at least one order from an Apple device.

The first value which could contribute to the hypothesis is the Average Order Value (AOV). AOV is the result of dividing revenue by the number of orders:

$$AOV = \frac{\sum_{i}^{orders}(totalprice(i))}{\text{number of orders}} \tag{3.12}$$

The AOV is calculated for each customer and then for each of the two groups: Apple users and others. A group AOV is an average of all belonging customers' AOVs. The results are presented in Table 3.6.

AOV	**Shop1**			**Shop2**		
	Average	**Median**	**Users**	**Average**	**Median**	**Users**
Apple users	513€	447€	> 40.000	515€	461€	> 200.000
Others	506€	441€	> 200.000	513€	458€	> 300.000

Table 3.6: AOV: Apple users vs Others across Shop1 and Shop2

As seen in Table 3.6, the differences between the two groups are very small. In both shops, Apple users have a slightly greater average than others.

Once again, it is necessary to investigate the differences within both groups before making a conclusion about the differences between groups. This will be done in the next section with help of statistics.

3.8.3 Verification of the hypothesis' relevance

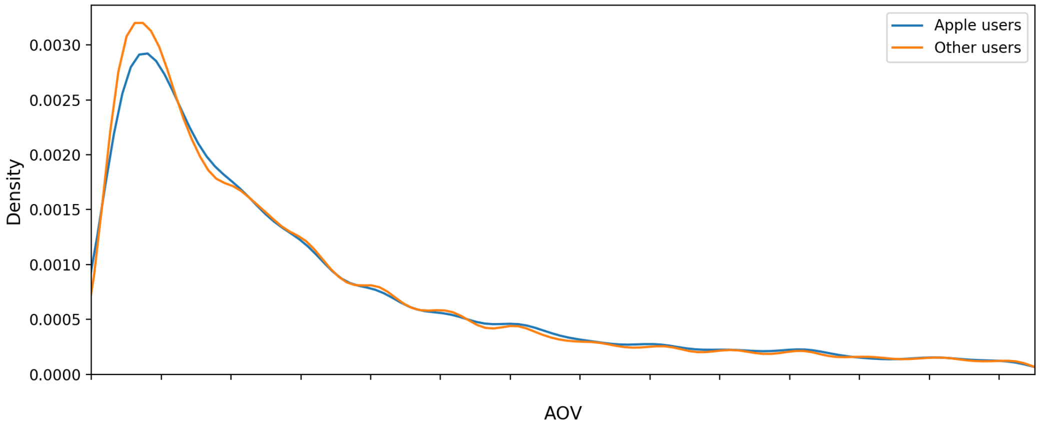


Figure 3.29: Shop1 - AOV Distribution plot: Apple users vs Others

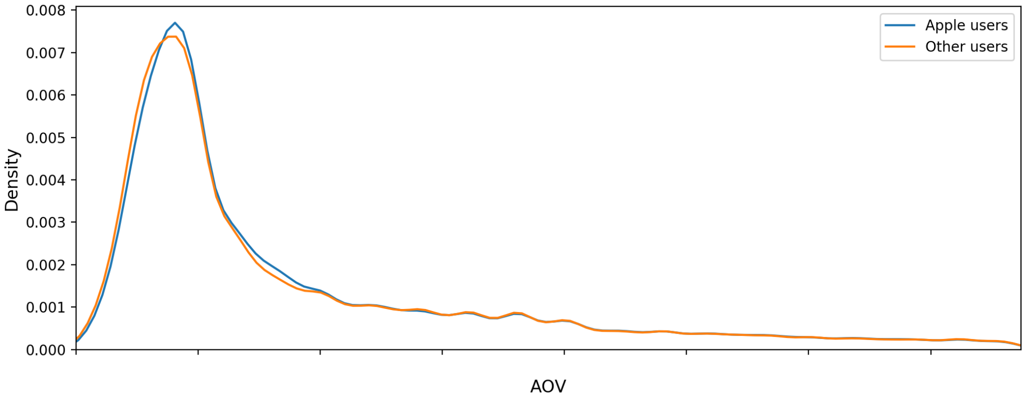


Figure 3.30: Shop2 - AOV Distribution plot: Apple users vs Others

In order to check the significance of the difference between two groups, the same statistical test as in the Subsection 3.7.3 will be used - Mann - Whitney U test. This can be done due to the nature of the data which is the same as in the case of iPhone and Others groups. The distribution plots are shown in Figures 3.29 and 3.30. They both show that the distibution is not normal and that there is very little difference between Apple users and others, especially in Shop2.

Mann-Whitney U test

The test will be conducted in six steps:

- **Step 1:** State the null and alternative hypothesis

- **Step 2:** Choose the level of significance
- **Step 3:** Determine the critical value
- **Step 4:** Compute the test statistic
- **Step 5:** Analyse results
- **Step 6:** Conclusion

Step 1: State the null and alternative hypothesis

Null-Hypothesis (H_0): Apple users and other users spend equal amount of money on average.

Alternative-Hypothesis (H_1): Apple users spend significantly more or less on average than other users.

This is a two-tailed test because "*the null hypothesis will be rejected if the difference between sample means is too big or if it is too small*". [Sta] This type of test is chosen over a one-tailed test because the Apple users of Shop1 spend more and the Apple users of Shop2 spend less according to the AOV average.

Step 2: Choose the level of significance α

The level of significance α is 0.05 as calculated in Subsection 3.7.3.

Step 3: Determine the critical value

The test statistic for the Mann-Whitney test is a U value. Every U value has an associated p value which indicates whether the difference between two groups is statistically significant. The meaning of the p value can be found in Section 2.2.

Step 4: Compute the test statistic

The function used for testing is a Python function:

scipy.stats.mannwhitneyu(x, y, use_continuity=True, alternative=None)

where x and y are arrays of samples, *user_continuity = True* refers to the continuity correction and *alternative* decides if the p-value for will be calculated for the one-sided hypothesis or the two-sided hypothesis.

According to the null- and alternative-hypotheses, the following parameters are chosen:

- x = Apple users AOV samples
- y = others AOV samples

- alternative = *two-sided*

After running tests with help of the *mannwhitneyu* function, the following results are observed:

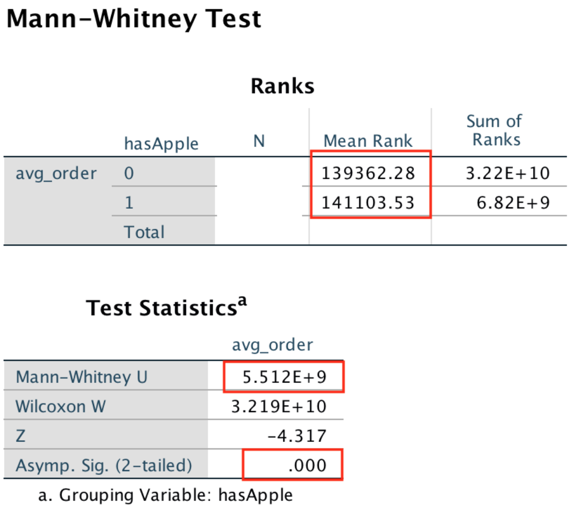
Shop	Statistic U	P-value
Shop1	5.6e9	0.00
Shop2	4e10	2.6e-40

Table 3.7: Results of the Mann Whitney U test: Apple users vs Others

Step 5: Analyse results

Since $p \leq \alpha$ in both shops, the difference between the Apple users' and other users' medians is statistically significant.

In order to double-check the results and find out which distribution is significanlty greater, the data were also tested with IBM SPSS Program for Statistics and the results are shown in Figure 3.33.



Mann-Whitney Test

Ranks

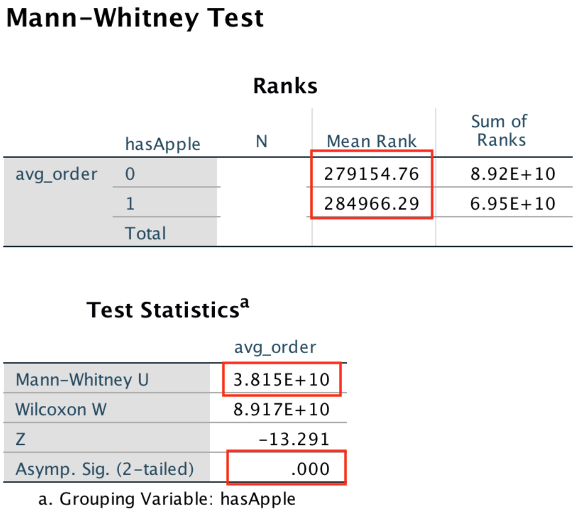
	hasApple	N	Mean Rank	Sum of Ranks
avg_order	0		139362.28	3.22E+10
	1		141103.53	6.82E+9
	Total			

Test Statistics[a]

	avg_order
Mann-Whitney U	5.512E+9
Wilcoxon W	3.219E+10
Z	-4.317
Asymp. Sig. (2-tailed)	.000

a. Grouping Variable: hasApple

Figure 3.31: Shop1



Mann-Whitney Test

Ranks

	hasApple	N	Mean Rank	Sum of Ranks
avg_order	0		279154.76	8.92E+10
	1		284966.29	6.95E+10
	Total			

Test Statistics[a]

	avg_order
Mann-Whitney U	3.815E+10
Wilcoxon W	8.917E+10
Z	-13.291
Asymp. Sig. (2-tailed)	.000

a. Grouping Variable: hasApple

Figure 3.32: Shop2

Figure 3.33: IBM SPSS Mann-Whitney Test Results: Apple users vs Others in Shop1 and Shop2

Once again, $p \leq \alpha$. The p values - denoted by Asymp. Sig (2-tailed) in Figure 3.33 - are the same as the Python values with fewer decimal places and *Mann-Whitney U* values differ slightly because there is no continuity correction in SPSS. Users who have Apple devices are denoted by hasApple = 1. The mean rank for Apple users is bigger than the mean rank for Others in both shops, which means that Apple users spend significantly more on average. The mean rank indicates which group has a greater average and this will be used in the conclusion.

Step 6: Conclusion

Shop1: A Mann-Whitney U test indicated that the Average order value was significantly greater for Apple users (Mdn = 447€) than others (Mdn = 441€) (Mann–Whitney U = 5650779943.5, m > 40.000 , n > 200.000, $P < 0.05$ two-tailed).

Shop2: A Mann-Whitney U test indicated that the Average order value was significantly greater for Apple users (Mdn = 461€) than others (Mdn = 458€) (Mann–Whitney U = 39760044709.5, m > 200.000, n > 300.000, $P \ll 0.05$ two-tailed).

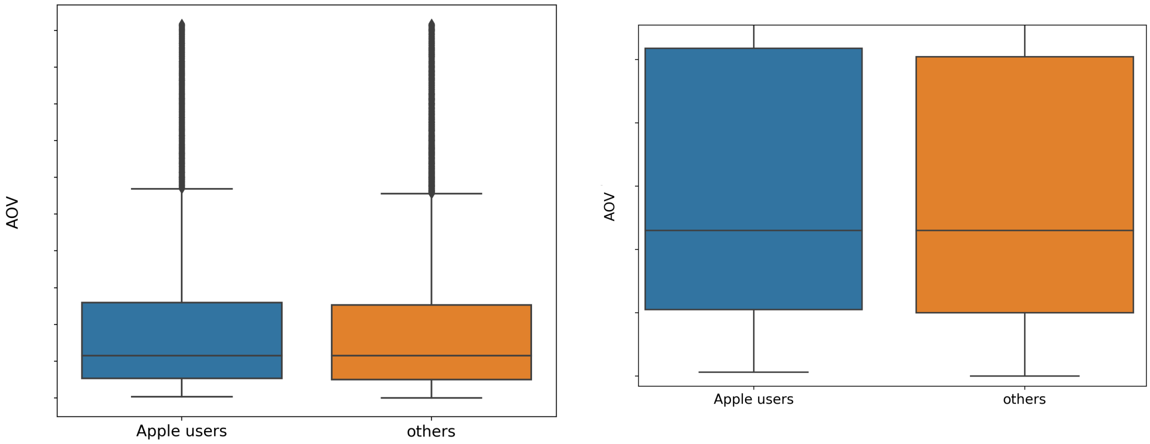


Figure 3.34: Shop1 - Boxplot of AOV of Apple users vs Others

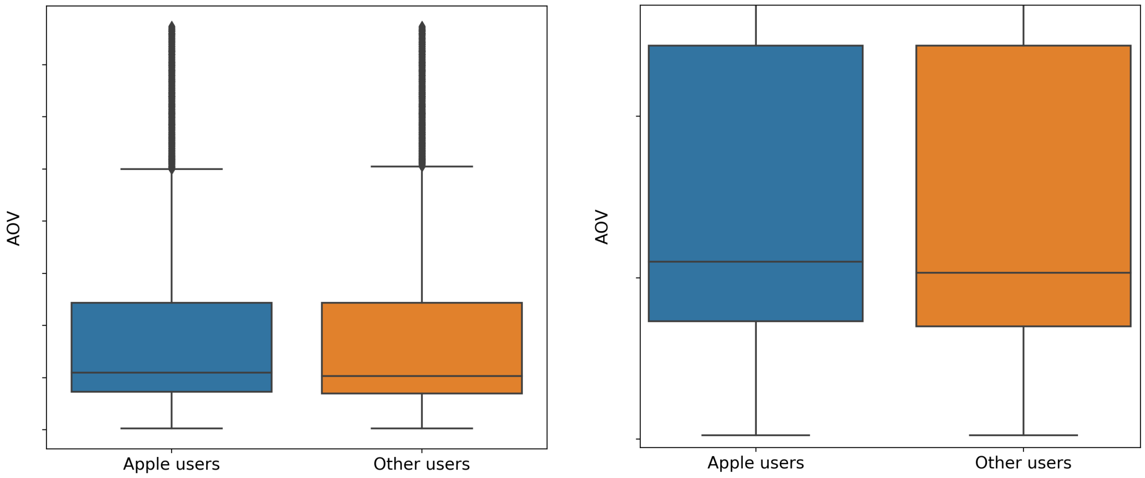


Figure 3.35: Shop2 - Boxplot of AOV of Apple users vs Others

To understand the results better, two boxplot diagrams for Shop1 and Shop2 are shown in Figures 3.34 and 3.35. Only a very small difference between Apple and Others box can be seen in Figures 3.34 and 3.35.

According to the results so far, iPhone users spend more only in the Shop1, while Apple users spend more in both shops. However, the difference between iPhone users and Others is much bigger than the difference between Apple users and others (see Figures 3.27, 3.28, 3.34 and 3.35).

In the hope of getting more accurate results, the difference between iPhone, Android, desktop and other users will be observed in the next section.

3.9 Hypothesis Ic: iPhone and Android users spend more money than desktop and other users

3.9.1 Implications of the hypothesis

If the hypothesis is true, it is more likely that iPhone and Android users are willing to spend more money and therefore more expensive products should be ranked higher for them than cheaper ones. This will be supported or rejected depending on the hypotheses IIa and IIb.

3.9.2 Examples confirming the hypothesis

All users are divided in four groups: iPhone, Android, Desktop and Other users. A user is considered to belong to a certain device group if it has placed at least one order from that device.

The first value which could contribute to the hypothesis is the Average Order Value (AOV). AOV is the result of dividing revenue by the number of orders:

$$AOV = \frac{\sum_{i}^{orders}(totalprice(i))}{\text{number of orders}} \tag{3.13}$$

The AOV is calculated for each customer and then for each of the four groups. A group AOV is an average of all belonging customers' AOVs. The results are presented in Table 3.8. Additionally, the averages and medians of Not-iPhone users from the previous section are shown for comparison purposes.

AOV	Shop1			Shop2		
	Average	Median	Users	Average	Median	Users
iPhone users	569€	518€	> 20.000	306€	247€	> 100.000
Android users	565€	527€	> 70.000	299€	238€	> 80.000
Desktop users	507€	428€	> 100.000	326€	276€	> 300.000
Others	510€	434€	> 20.000	393€	286€	> 100.000
Not-iPhone	525€	458€	> 200.000	322€	268€	> 400.000

Table 3.8: AOV: iPhone vs Android vs Desktop vs Others across Shop1 and Shop2

As seen in Table 3.8, the differences between the groups are bigger now. E.g. the difference between the iPhone and the Not-iPhone users was 44€, and it is 59€ now. The difference between the Android and the Desktop users is 66€. The Android users have an average and median AOV similar to iPhone users, while desktop and other users have an average and median like Non-iPhone users (previously Others).

The idea behind this hypothesis was to show that there is a bigger difference between (most) mobile and desktop (and other) users than between iPhone and other users.

For the sake of the statistical significance of the result, it is necessary to investigate the differences within both groups (iPhone/Android users and Desktop/other users) before making a conclusion about differences between groups. This will be done in the next section with help of statistics.

3.9.3 Verification of the hypothesis' relevance

In order to check the significance of the difference between two groups, the same statistical test as in Subsection 3.7.3 will be used - Mann - Whitney U test. This can be done due to the nature of the data which is the same as in the case of the iPhone and Others groups, and Apple and Others groups. The distributions plots for both shops are shown in the Figures 3.36 and 3.37.

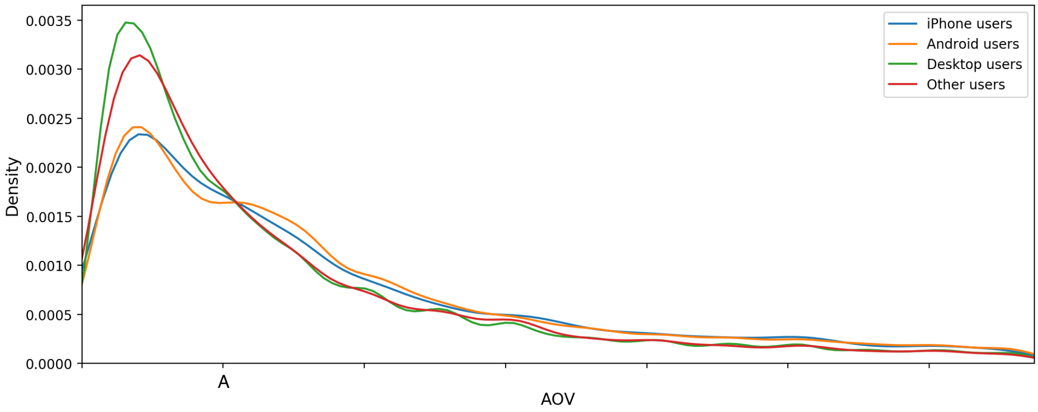


Figure 3.36: Shop 1 - AOV Distribution plot: iPhone vs Android vs Desktop vs Others

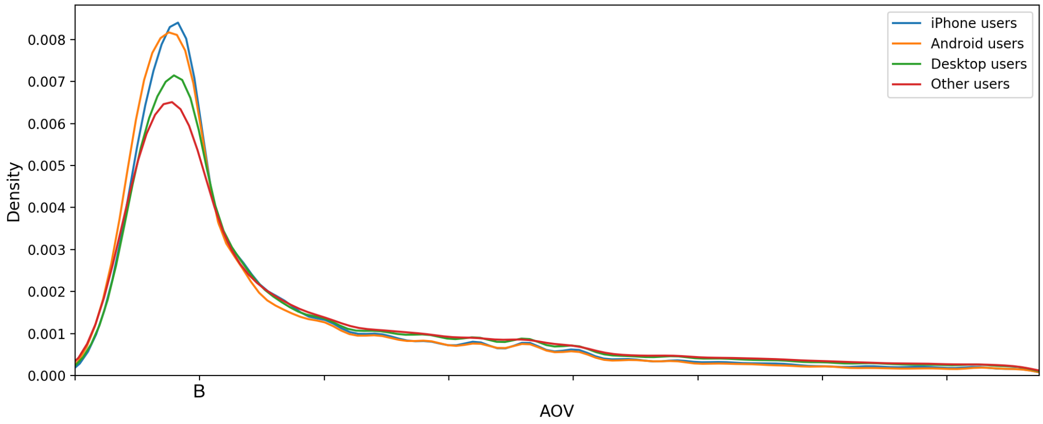


Figure 3.37: Shop 2 - AOV Distribution plot: iPhone vs Android vs Desktop vs Others

The distribution plots (Figures 3.36 and 3.37) show that the values are not normally distributed and that there is a difference between iPhone/Android users and Desktop/Other users. In Shop1, the distribution curves for the two mobile groups are always bigger than for the other two groups after point A, which confirms that iPhone/Android users spend more on average. In Shop2, the distribution curves for first two mobile groups are always bigger than that of the other two groups after point B, which indicates iPhone/Android users spend less on average.

Mann-Whitney U test

Since there are four groups observed and the Mann-Whitney U test works with only two groups, only the two following tests will be made: Android vs iPhone users and Android vs Desktop users.

Test1: Android vs iPhone users

Three different types of hypotheses are tested. The results are shown in Table 3.9. As mentioned before, the Python-based Mann-Whitney test has a parameter called *alternative* which decides whether the p-value is related to the one-sided or two-sided hypothesis. From the results, in Shop1 Android users spend significantly more on average while in Shop2 iPhone users spend significantly more. This can be concluded based on the *p-values* which are smaller than α (see Section 3.8.3) for the coresponding one-sided hypotheses.

Shop	**Statistic U**	**P-value**	**Hypothesis**
Shop1	7.7e9	0.0012	two-sided
	7.4e9	0.0006	one-sided (iPhone < Android)
	7.4e8	**0.9994**	one-sided (iPhone > Android)
Shop2	5.8e9	5.95e-64	two-sided
	5.3e9	**1.0**	one-sided (iPhone < Android)
	5.3e9	9.17e-61	one-sided (iPhone > Android)

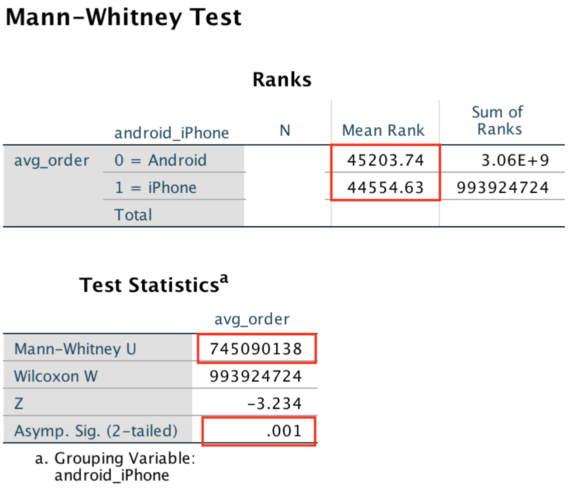
Table 3.9: Results of the Mann Whitney U test: Android users vs iPhone users

The results are equivalent to the ones obtained in IBM SPSS Statistics (Figure 3.40).

Test2: Android vs Desktop users

Three different types of hypotheses are tested. The results are shown in Table 3.10. From the results, in Shop1 Android users spend significantly more on average while in Shop2 Desktop users spend significantly more. This can be concluded based on the *p-values* which are smaller than α (see Section 3.8.3) for the coresponding one-sided hypotheses.

The results are equivalent to the ones obtained in IBM SPSS Statistics (Figure 3.43).



Mann-Whitney Test

Ranks

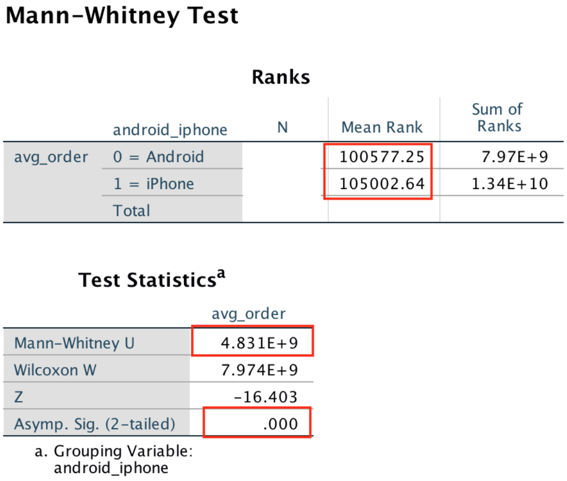
	android_iPhone	N	Mean Rank	Sum of Ranks
avg_order	0 = Android		45203.74	3.06E+9
	1 = iPhone		44554.63	993924724
	Total			

Test Statistics[a]

	avg_order
Mann-Whitney U	745090138
Wilcoxon W	993924724
Z	-3.234
Asymp. Sig. (2-tailed)	.001

a. Grouping Variable: android_iPhone

Figure 3.38: Shop1



Mann-Whitney Test

Ranks

	android_iphone	N	Mean Rank	Sum of Ranks
avg_order	0 = Android		100577.25	7.97E+9
	1 = iPhone		105002.64	1.34E+10
	Total			

Test Statistics[a]

	avg_order
Mann-Whitney U	4.831E+9
Wilcoxon W	7.974E+9
Z	-16.403
Asymp. Sig. (2-tailed)	.000

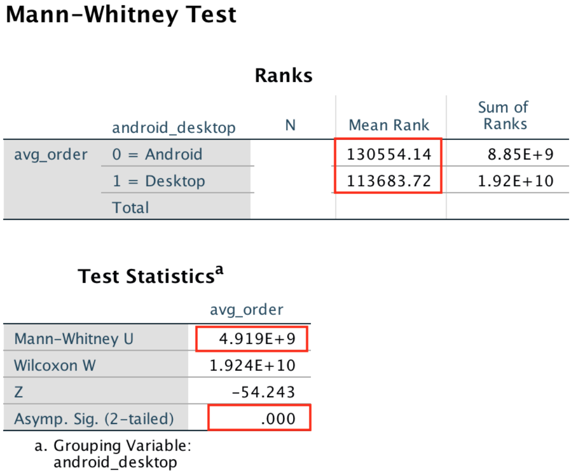
a. Grouping Variable: android_iphone

Figure 3.39: Shop2

Figure 3.40: IBM SPSS Mann-Whitney Test Results: Android vs iPhone users

Shop	**Statistic U**	**P-value**	**Hypothesis**
Shop1	9.6e9	0.0	two-sided
	6.5e9	**1.0**	one-sided (Android < Desktop)
	6.5e9	0.0	one-sided (Android > Desktop)
Shop2	9.6e9	0.0	two-sided
	9.6e9	0.0	one-sided (Android < Desktop)
	9.6e9	**1.0**	one-sided (Android > Desktop)

Table 3.10: Results of the Mann Whitney U test: Android vs Desktop users



Mann-Whitney Test

Ranks

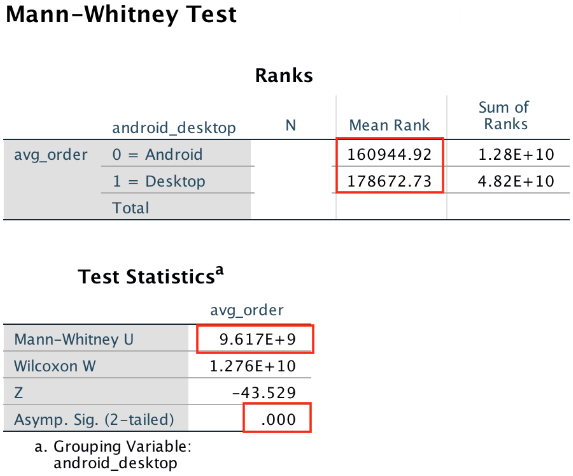
	android_desktop	N	Mean Rank	Sum of Ranks
avg_order	0 = Android		130554.14	8.85E+9
	1 = Desktop		113683.72	1.92E+10
	Total			

Test Statistics[a]

	avg_order
Mann-Whitney U	4.919E+9
Wilcoxon W	1.924E+10
Z	-54.243
Asymp. Sig. (2-tailed)	.000

a. Grouping Variable: android_desktop

Figure 3.41: Shop1



Mann-Whitney Test

Ranks

	android_desktop	N	Mean Rank	Sum of Ranks
avg_order	0 = Android		160944.92	1.28E+10
	1 = Desktop		178672.73	4.82E+10
	Total			

Test Statistics[a]

	avg_order
Mann-Whitney U	9.617E+9
Wilcoxon W	1.276E+10
Z	-43.529
Asymp. Sig. (2-tailed)	.000

a. Grouping Variable: android_desktop

Figure 3.42: Shop2

Figure 3.43: IBM SPSS Mann-Whitney Test Results: Android vs Desktop users

Conclusion

The boxplots for the previous two tests and both shops are shown in Figures 3.44, 3.45, 3.44 and 3.47. From the boxplots, (most) mobile users, i.e., iPhone and Android users spend more on average than other users only in Shop1. In Shop2, it is vice versa. These boxplots confirm the test results.

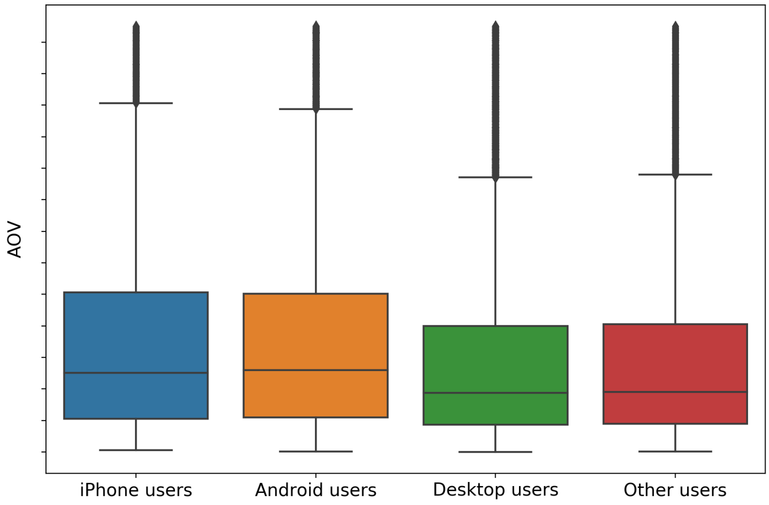


Figure 3.44: Shop 1 - AOV Boxplot: iPhone vs Android vs Desktop vs Others

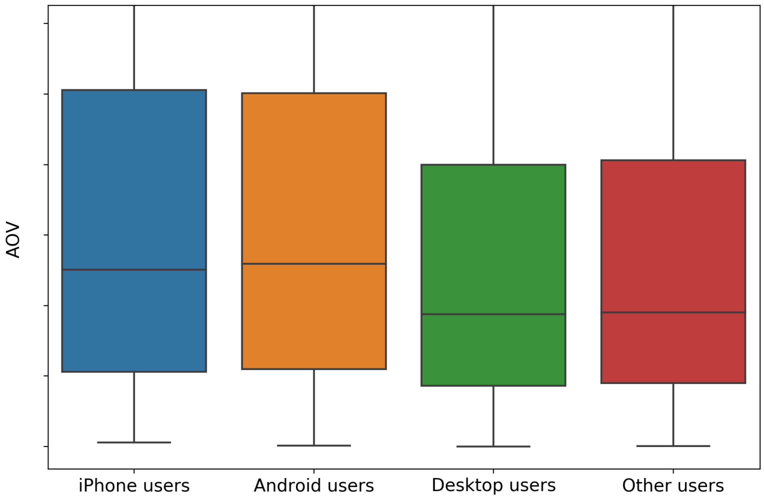


Figure 3.45: Shop 1 - Zoomed AOV Boxplot: iPhone vs Android vs Desktop vs Others

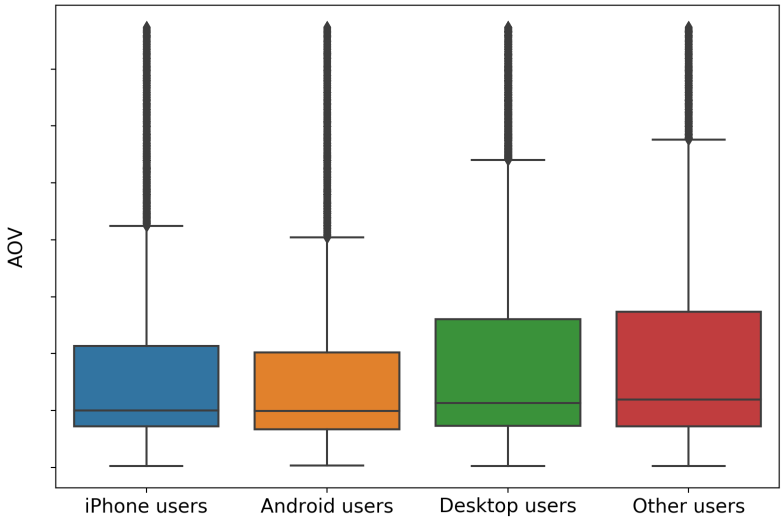


Figure 3.46: Shop 2 - AOV Boxplot: iPhone vs Android vs Desktop vs Others

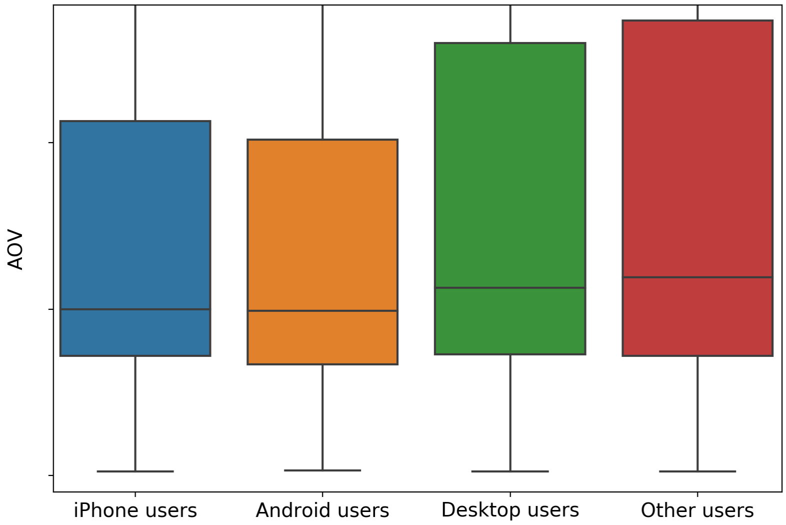


Figure 3.47: Shop 2 - Zoomed AOV Boxplot: iPhone vs Android vs Desktop vs Others

To summarise, the results have shown that iPhone users spend more on average only in the first shop (Section 3.7) and that Android/iPhone users also spend more on average only in the first shop (this section). In order to get more precise and results which can be applied to both shops, two more versions of iPhone/Apple-related hypotheses will be explored and tested.

3.10 Hypothesis IIa: iPhone users purchase greater quantities of expensive products than other users

3.10.1 Implications of the hypothesis

If the hypothesis is true, the expensive product should be ranked higher for iPhone users.

3.10.2 Examples confirming the hypothesis

All products are labelled as *expensive* or *inexpensive*. For each **shop category**, the Top 10% most expensive products are marked as "expensive" [2]. All the other products are marked as "inexpensive".

The overview of expensive products share is shown in Table 3.11. There are more expensive products in Shop2 because number of categories and categories' sizes in both shops differ.

	Shop1	**Shop2**
Expensive products	20.877	58.240
Total number of products	51.564	79.680
Total percentage of expensive products	18%	35%

Table 3.11: Overview of expensive products in Shop1 and Shop2

All users are divided in two groups: iPhone users and others. A user is considered to be an iPhone user if it has placed at least one order from an iPhone. Only users with **a minimum of two purchases** are considered for this experiment.

For each user, the percentage of the expensive products purchased is calculated. Based on these values, the average percentages for both groups are obtained. The results are shown in Table 3.12.

	Shop1	**Shop2**
iPhone Users	56%	57%
Others	48%	54%

Table 3.12: Percentage of expensive products: iPhone users vs others

As it can be seen from Table 3.12, there is a relatively small difference between iPhone users and Others in both shops. iPhone users buy 8% more expensive products on average than the others in Shop1, while they buy 3% more expensive products on average than the others in Shop2.

In the next section, these differences will be verified with the help of statistics.

[2] this value is chosen arbitrarily

3.10.3 Verification of the hypothesis' relevance

The goal of this subsection is to verify the findings from the previous subsection, i.e., that iPhone users purchase greater quantities of expensive products compared to the other users and to verify if there is a relationship between owning an iPhone and buying expensive products. Before choosing the method for testing the hypothesis, the following factors need to be considered:

- **Data:** For each purchase, there is a number of expensive and inexpensive products the customer has purchased and the device from which the purchase was made. All purchased products will be treated as a single data instance and categorised as *expensive* (amongst top 10% most expensive in its shop category) or *inexpensive*. The devices from which the product was purchased will be labeled as an *iPhone* or *not-iPhone* device. The type of data is therefore **categorical**. The data can be summarised in the contigency table, explained in Section 2.3.
- **Samples:** As shown in Figure 2.2, there is one sample with two categorical variables: iPhone ownership and the expensiveness of the product.
- **Purpose:** The purpose of testing is to **verify if there is a relationship** between customers having an iPhone and purchasing expensive products.

Based on the listed factors, the appropriate test for the hypothesis is **a Chi-square test of independence** described in Section 2.3.

Chi-square test of independence

"*The Chi-Square test of independence is used to determine if there is a significant relationship between two nominal (categorical) variables. ... The data can be displayed in a* ***contingency table*** *where each row represents a category for one variable and each column represents a category for the other variable.*" [Jam] The variables in the following test are the iPhone ownership and the product expensiveness. Each purchase represents one or more records for which it is checked whether the user has an iPhone and if the belonging products are expensive or not.

The test will be conducted in six steps:

- **Step 1:** State the null and alternative hypothesis
- **Step 2:** Choose the level of significance
- **Step 3:** Determine the critical value
- **Step 4:** Compute the test statistic
- **Step 5:** Analyse results
- **Step 6:** Conclusion

Step 1: State the null and alternative hypothesis

Null-Hypothesis (H_0): There is no relationship between users owning an iPhone and buying expensive products.

Alternative-Hypothesis (H_1): There is a relationship between users owning an iPhone and buying expensive products.

The test is a two-tailed test because it checks whether there is a relationship or not.

Step 2: Choose the level of significance α

The level of significance α is 0.05 as caluclated in Subsection 3.8.3.

Step 3: Determine the critical value

The test statistic for the Chi-square test of independence is a *chi-square* value. Every *chi-square* value has an associated p value which indicates whether the association between two variables is statistically significant. The meaning of the p value can be found in Section 2.3.

Step 4: Compute the test statistic

The function used for testing is a Python function:

scipy.stats.chi2_contingency(observed, correction=True, lambda_=None)

where *observed* is the contingency table, *correction* is the Yates' correction for continuity and *lambda* enables usage of another statistic from the Cressie-Read power divergence family.

For the purposes of this proof, the only parameter necessary is *observed*, i.e., the contingency table. The contigency tables for Shop1 and Shop2, respectively, are shown in Tables 3.13 and 3.14. The last row is not part of the contingeny table. It is only shown for the sake of clarity.

		User		
		iPhone	**Other**	**Total**
Products	**Expensive**	6.950	63.469	63.475
	Inexpensive	10.703	164.195	174.898
Total		17.654	227.664	245.318
Expensive products percentage		39.37%	27.88%	28.7%

Table 3.13: Shop1 - iPhone users vs Others: Contingency table

		User		
		iPhone	Other	Total
Products	Expensive	46.188	260.108	306.296
	Inexpensive	35.127	235.165	270.292
Total		81.316	495.274	576.590
Expensive products percentage		56.8%	52.52%	53.12%

Table 3.14: Shop2 - iPhone users vs Others: Contingency table

After running tests with the help of the *chi2_contingency* function, the following results are observed:

Shop	Statistic chi-square	P-value
Shop1	2643.12	0.0
Shop2	1800.83	0.0

Table 3.15: Results of the Chi-square test of the independence of user device and the purchase of expensive products

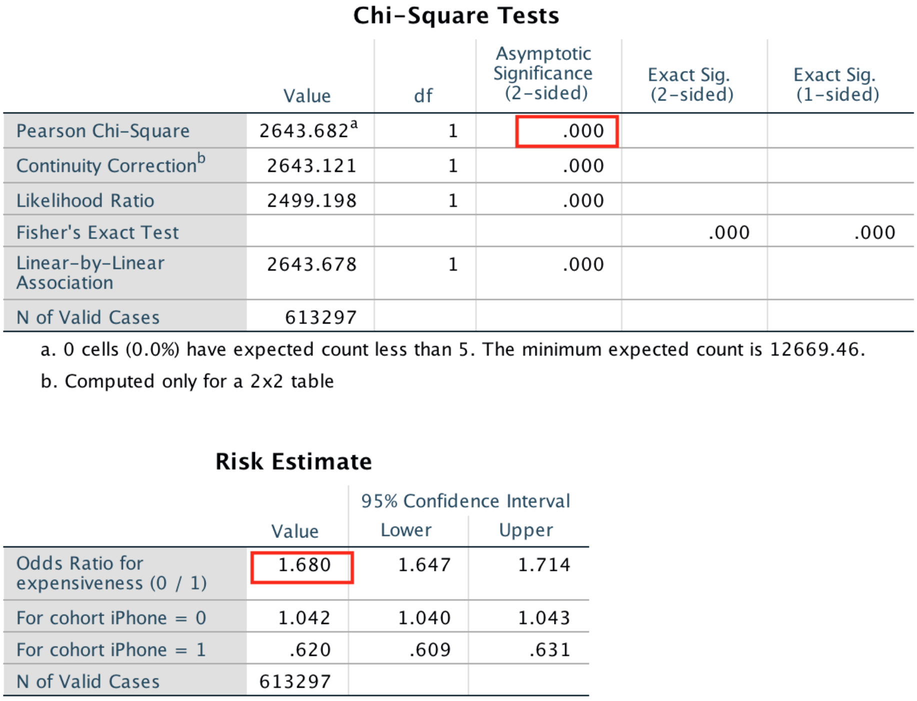
Step 5: Analyse results

Since $p \leq \alpha$ in both shops, there is a statistically significant relationship between having an iPhone and buying the expensive products.

In order to double-check the results and find out more about the odds, the data were also tested with IBM SPSS Program for Statistics and the results are shown in Figures 3.48 and 3.49 .

Once again, $p \leq \alpha$. The p values - denoted by Asymptotic Significance (two-sided) - are the same as in Python. The Chi-square statistic slightly differs. The users who have an iPhone are denoted by iPhone = 1. The expensive orders are denoted by expensiveness = 1.

Given these results, the odds of not having an iPhone are 1.1680 (1.189 in Shop2) times greater for the customers who bought an inexpensive products compared to the customers who bought an expensive product.



Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2643.682[a]	1	.000		
Continuity Correction[b]	2643.121	1	.000		
Likelihood Ratio	2499.198	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	2643.678	1	.000		
N of Valid Cases	613297				

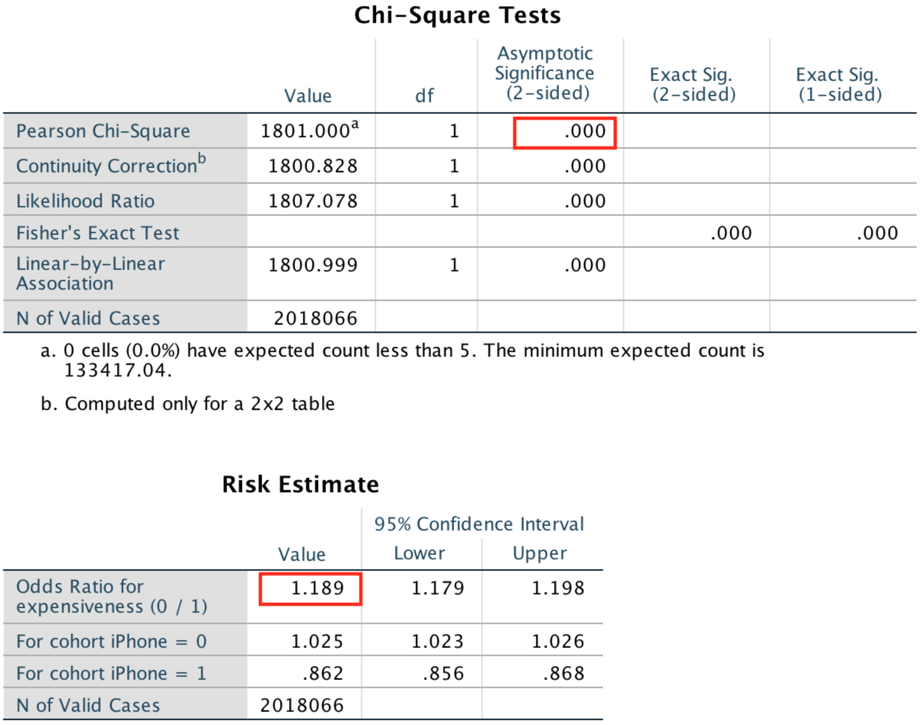
a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 12669.46.

b. Computed only for a 2x2 table

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for expensiveness (0 / 1)	1.680	1.647	1.714
For cohort iPhone = 0	1.042	1.040	1.043
For cohort iPhone = 1	.620	.609	.631
N of Valid Cases	613297		

Figure 3.48: Shop1: IBM SPSS Chi-Square Test Results



Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1801.000[a]	1	.000		
Continuity Correction[b]	1800.828	1	.000		
Likelihood Ratio	1807.078	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	1800.999	1	.000		
N of Valid Cases	2018066				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 133417.04.

b. Computed only for a 2x2 table

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for expensiveness (0 / 1)	1.189	1.179	1.198
For cohort iPhone = 0	1.025	1.023	1.026
For cohort iPhone = 1	.862	.856	.868
N of Valid Cases	2018066		

Figure 3.49: Shop2: IBM SPSS Chi-Square Test Results

Step 6: Conclusion

Shop1: A chi-square test was used to determine whether there was a significant difference between iPhone users and others in the purchasing of in-/expensive products. Only 27.88% of the products purchased by non-iPhone users were expensive, whereas 39.37% of the products purchased by iPhone users were expensive. This difference was statistically significant, $\chi^2(1) = 2643.12$, p <.000.

Shop2: A chi-square test was used to determine whether there was a significant difference between iPhone users and others in the purchasing of in-/expensive products. 52.52% of the products purchased by non-iPhone users were expensive, whereas 56.8% of the products purchased by iPhone users were expensive. This difference was statistically significant, $\chi^2(1) = 1800.83$, p <.000.

To understand the results better, two boxplot diagrams for Shop1 and Shop2 are shown in Figure 3.50. The boxplots are based on the proportion of the expensive products the single users purchased. The boxplots confirm the results of the statistical test. As seen in the figures, the median of iPhone users in both shops is much higher than the median of other users. Also, the box starts at 67% for iPhone users in both shops, and at 40% and 33% in Shop 1 and Shop2 respectively. This shows once again that iPhone users purchase greater quantities of expensive products than others.

Based on all the results presented in this section, it can be concluded that the percentage of the expensive products purchased by the iPhone users is significantly bigger than the percentage of the expensive products purchased by the non-iPhone users and that there is a significant relationship between having an iPhone and purchasing expensive products.

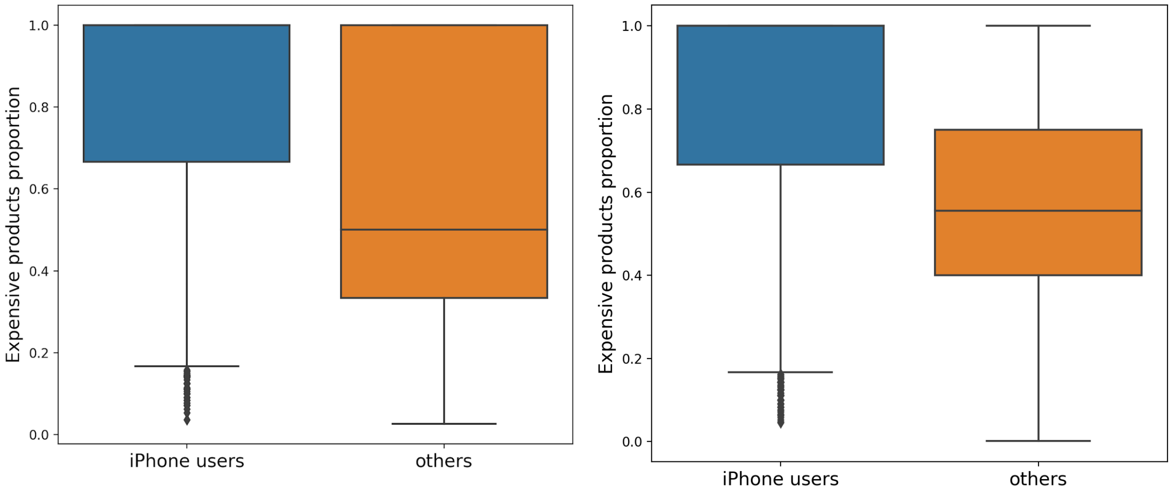


Figure 3.50: Shop1 vs Shop2: Expensive products sales proportion boxplot

3.11 Hypothesis IIb: Apple users purchase greater quantities of expensive products than other users

This section does an analogous examination to Section 3.10 with the difference of iPhone users being replaced by Apple users.

3.11.1 Implications of the hypothesis

If the hypothesis is true, the expensive products should be ranked higher for Apple users.

3.11.2 Examples confirming the hypothesis

All users are divided in two groups: Apple users and Others. A user is considered to be an Apple device user if it has placed at least one order from an Apple device. Only users with **a minimum of two purchases** are observed.

For each user, the percentage of the expensive products purchased is calculated. Based on these values, the average percentages for both groups are obtained. The results are shown in Table 3.16.

	Shop1	**Shop2**
Apple Users	53%	56%
Others	48%	54%

Table 3.16: Percentage of expensive products: Apple users vs others

As it can be seen from Table 3.16, there is a relatively small difference between Apple users and Others in both shops. Apple users buy 5% more expensive products on average than the others in Shop1, while they buy 2% more expensive products on average than the others in Shop2.

In the next section, these differences will be verified with the help of statistics.

3.11.3 Verification of the hypothesis' relevance

The goal of this subsection is to verify the findings from the previous subsection, i.e., that Apple users purchase greater quantities of expensive products compared to the other users and to verify if there is a relationship between owning an Apple product and buying expensive prouducts. Before chosing the method for testing the hypothesis, the following factors need to be considered:

- **Data:** For each purchase, there is a number of expensive and inexpensive products the customer has purchased and the device from which the purchase was made. All purchased products will be treated as a single data instance and categorised as *expensive* or *inexpensive*. The devices from which the product was purchased will be labelled as an *Apple* or *not-Apple* device. The type of data is therefore **categorical**. The data can be summarised in the contigency table.

- **Samples:** As shown in Figure 2.2, there is one sample with two categorical variables: Apple ownership and the expensiveness of the product.
- **Purpose:** The purpose of testing is to **verify if there is a relationship** between customers having an Apple and purchasing expensive products.

Based on the listed factors, the appropriate test for the hypothesis is **a Chi-square test of independence** described in Section 2.3.

Chi-square test of independence

The test will be conducted in six steps:

- **Step 1:** State the null and alternative hypothesis
- **Step 2:** Choose the level of significance
- **Step 3:** Determine the critical value
- **Step 4:** Compute the test statistic
- **Step 5:** Analyse results
- **Step 6:** Conclusion

Step 1: State the null and alternative hypothesis

Null-Hypothesis (H_0): There is no relationship between users owning an Apple device and buying expensive products.

Alternative-Hypothesis (H_1): There is a relationship between users owning an Apple device and buying expensive products.

The test is a two-tailed test because it checks whether there is a relationship or not.

Step 2: Choose the level of significance α

The level of significance α is 0.05 as caluclated in Subsection 3.8.3.

Step 3: Determine the critical value

The test statistic for the Chi-square test of independence is a *chi-square* value. Every *chi-square* value has an associated p value which indicates whether the association between two variables is statistically significant. The meaning of the p value can be found in Section 2.3.

Step 4: Compute the test statistic

Once again, the function used for testing is a Python function:

scipy.stats.chi2_contingency(observed, correction=True, lambda_=None)

For the purposes of this proof, the only parameter necessary is *observed*, i.e., the contingency table. The contigency tables for Shop1 and Shop2, respectively, are shown in Tables 3.17 and 3.18. The last row is not part of the contingeny table. It is only shown for the sake of clarity.

		User		
		iPhone	**Other**	**Total**
Products	**Expensive**	12.592	57.828	70.420
	Inexpensive	21.796	153.102	174.898
Total		34.388	210.930	245.318
Expensive products percentage		36.62%	27.41%	28.7%

Table 3.17: Shop1 - Apple users vs Others: Contingency table

		User		
		iPhone	**Other**	**Total**
Products	**Expensive**	99.794	206.503	306.297
	Inexpensive	82.529	187.763	273.292
Total		182.323	394.267	576.590
Expensive products percentage		54.73%	52.38%	53.12%

Table 3.18: Shop2 - Apple users vs Others: Contingency table

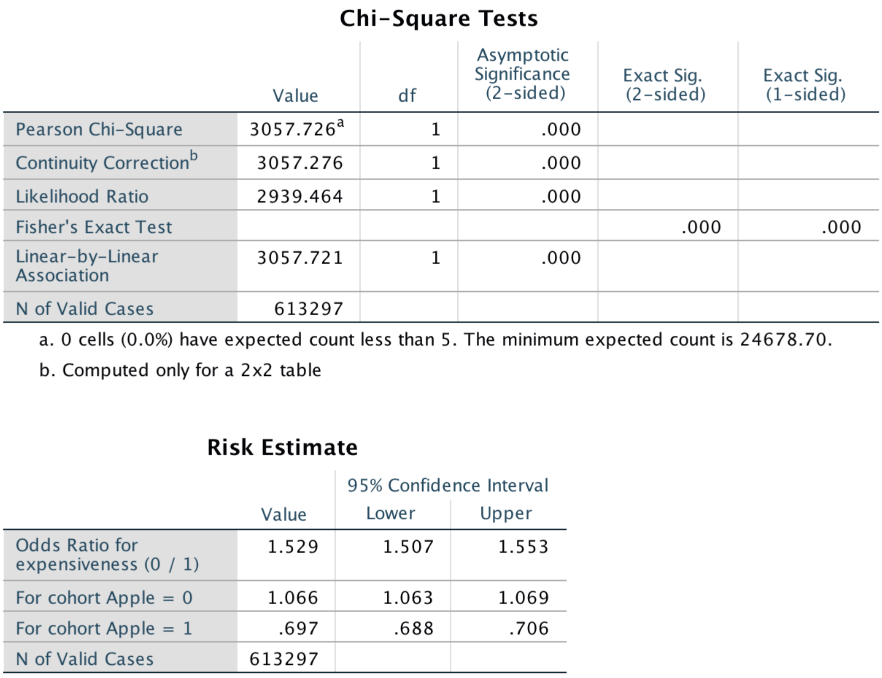
After running tests with the help of the *chi2_contingency* function, the following results are observed:

Shop	**Statistic chi-square**	**P-value**
Shop1	3057.2763724164179	0.0
Shop2	974.30422422185893	6.9131188151618246e-214

Table 3.19: Results of the Chi-square test of independence

Step 5: Analyse results

Since $p \leq \alpha$ in both shops, there is a statistically significant relationship between having an Apple device and buying the expensive products.



Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	3057.726[a]	1	.000		
Continuity Correction[b]	3057.276	1	.000		
Likelihood Ratio	2939.464	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	3057.721	1	.000		
N of Valid Cases	613297				

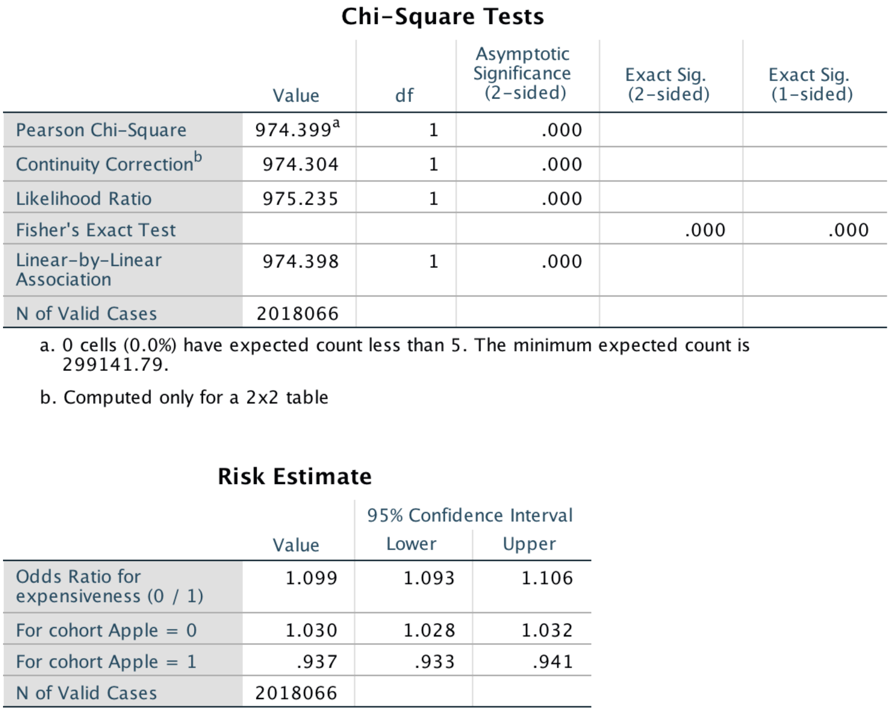
a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 24678.70.

b. Computed only for a 2x2 table

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for expensiveness (0 / 1)	1.529	1.507	1.553
For cohort Apple = 0	1.066	1.063	1.069
For cohort Apple = 1	.697	.688	.706
N of Valid Cases	613297		

Figure 3.51: Shop1: IBM SPSS Chi-Square Test Results



Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	974.399[a]	1	.000		
Continuity Correction[b]	974.304	1	.000		
Likelihood Ratio	975.235	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	974.398	1	.000		
N of Valid Cases	2018066				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 299141.79.

b. Computed only for a 2x2 table

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for expensiveness (0 / 1)	1.099	1.093	1.106
For cohort Apple = 0	1.030	1.028	1.032
For cohort Apple = 1	.937	.933	.941
N of Valid Cases	2018066		

Figure 3.52: Shop2: IBM SPSS Chi-Square Test Results

In order to double-check the results and find out more about the odds, the data were also tested with IBM SPSS Program for Statistics and the results are shown in Figures

3.51 and 3.52.

Once again, $p \leq \alpha$. The p values - denoted by Asymptotic Significance (two-sided) - are the same as in Python. The Chi-square slightly differs. The users who have an Apple device are denoted by Apple = 1. Expensive orders are denoted by expensiveness = 1.

Given these results, the odds of not having an iPhone are 1.529 (1.099 in Shop2) times greater for the customers who bought inexpensive products compared to the customers who bought an expensive product.

Step 6: Conclusion

Shop1: A chi-square test was used to determine whether there was a significant difference between Apple users and others in the purchasing of in-/expensive products. Only 27.41% of the products purchased by non-Apple users were expensive, whereas 36.62% of the products purchased by Apple users were expensive. This difference was statistically significant, $\chi^2(1) = 3057.28$, p <.000.

Shop2: A chi-square test was used to determine whether there was a significant difference between Apple users and others in the purchasing of in-/expensive products. 52.38% of the products purchased by non-Apple users were expensive, whereas 54.73% of the products purchased by Apple users were expensive. This difference was statistically significant, $\chi^2(1) = 974.30$, p <.000.

To understand the results better, two boxplot diagrams for Shop1 and Shop2 are shown in Figure 3.53. The boxplots are based on the proportion of the expensive products the single users purchased. The boxplots confirm the results of the statistical test. As seen in the figures, the median of Apple users in both shops is much higher than the median of other users. Also, the box starts at 66% for Apple users in both shops, and at 33% and 40% in Shop 1 and Shop2 respectively. This shows once again that Apple users purchase greater quantities of expensive products than others.

Based on all the results presented in this section, it can be concluded that the percentage of the expensive products purchased by the Apple users is significantly bigger than the percentage of the expensive products purchased by the non-Apple users and that there is a significant relationship between having an Apple device and purchasing expensive products.

Conclusion

This chapter investigated buying habits of iPhone and Apple user in comparison with other users. Five different hypotheses were tested on two different shops:

- Ia: iPhone users spend more money than other users
- Ib: Apple users spend more money than other users
- Ic: iPhone and Android users spend more money than desktop and other users

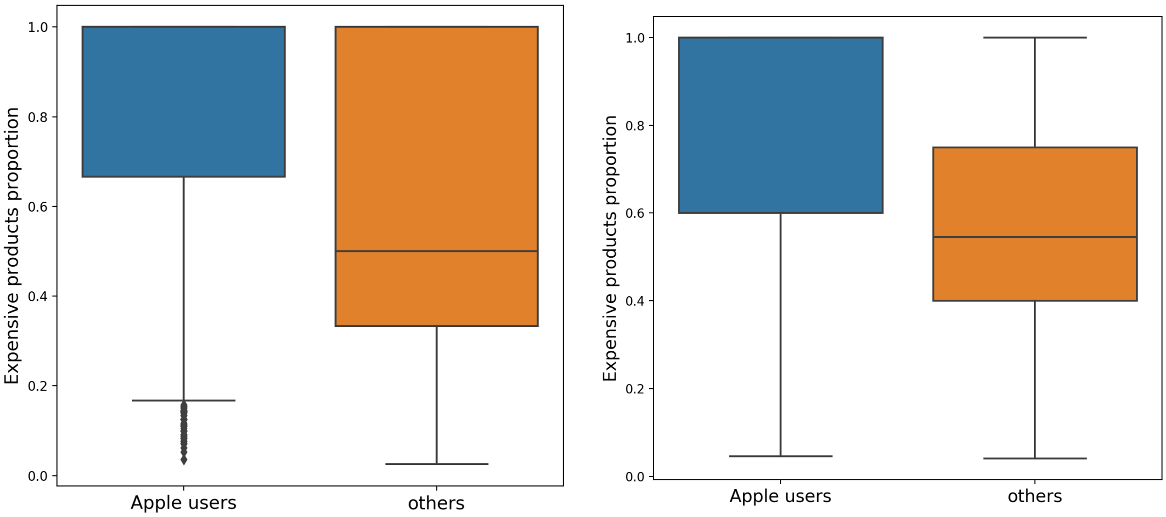


Figure 3.53: Shop1 vs Shop2: Expensive products sales proportion boxplot

- IIa: iPhone users purchase greater quantities of expensive products than other users
- IIb: Apple users purchase greater quantities of expensive products than other users

The results of the first three hypotheses tests cannot be used since they indicate that both iPhone users and iPhone&Android users spend more only in Shop1 and the difference between Apple users and others in both shops is not big.

On the other hand, all the results related to hypotheses IIa and IIb show that iPhone/Apple users purchase greater quantities of expensive products than other. Thus, these hypotheses implications can be used for the sorting.

Summary: *After a brief explanation of the data exploration and research methodology, each hypothesis has been examined separately. The examination consisted of hypothesis implications, examples confirming the hypothesis and finally, the verification of the hypothesis relevance. This was necesseary in order to check whether the hypotheses can be used for the sorting algorithm. In the next chapter, the impact of the hypothesis on the sorting algorithm will be observed as a part of high-level design for the sorting module.*

4 Design

This chapter is left out because of the company's privacy. The goal of this chapter is to explain the high-level design, investigate each hypothesis' impact on the sorting algorithm and explain the differences between the design and the real-world implementation.

5 Implementation

This chapter is left out because of the company's privacy. This chapter explains how the scoring part of the code works, how the data exploration was made and what changes were made for the new algorithms. Only basic details of the implementation are revealed.

6 Evaluation

This chapter evaluates the new sorting algorithm by comparing the product orders of old and new sorting at different hours. The evaluation settings are described in Section 6.1 and the results of the evaluation can be found in Section 6.2.

6.1 Evaluation settings

For the purposes of the evaluation, two categories from Shop1 are chosen: *Category1* with 101 products and *Category2* with 35 products.

Only two hypotheses were used for testing. The time of the day hypothesis sorts products depending on the hour of sorting, while the season hypothesis uses periods to determine the sorting.

The sorting used so far in one of the shops is called Standard sorting. The evalution is done by comparing Standard sorting with seven new sortings:

A : Sorting at 10 am with *time of the day hypothesis* only

B : Sorting at 2 pm with *time of the day hypothesis* only

C : Sorting at 4 pm with *time of the day hypothesis* only

D : Sorting at 8 pm with *time of the day hypothesis* only

E : Sorting at 2 am with *time of the day hypothesis* only

F : Sorting at 10 am in July with *season hypothesis* only

G : Sorting at 10 am in July with both *time of the day and season hypothesis*

The comparison is done in two ways:

- visually: by showing the top eight products which are the ones shown in the viewport when user views the website in a Desktop browser
- quantitatively: by checking for how many positions each product in the category has shifted either upwards or downwards and verifying the new scores i.e. new product order

The goal of the evaluation is to show that there is a change in product order after applying the new sorting algorithm and that the product order is as expected from the new algorithm.

6.2 Results

6.2.1 Visual comparsion

After visually comparing the sortings A-G with the standard sorting, the biggest difference in products for the *Category1* category was found in F, A and G with seven, six and four new products respectively. This can be seen by comparing Figure A.1 with Figures A.2, A.3 and A.4 in Appendix A.

In addition to the novelty, the advantage of new product sorting groups is that they also contain products in multiple styles, unlike the standard sorting which contains only products in one style where six out of eight products are of the same type.

When it comes to the *Category2* category, sortings with the largest number of new products are D, A and E with four new products in each sorting. The differences can be seen by comparing Figure A.5 with Figures A.6, A.7 and A.8 in Appendix A.2. It is interesting to note that Standard sorting shows four of the same type of products, while new sortings have more variety.

The visual comparison is good for describing the advantages of new sorting for users who open the category before they decide to scroll down, and for seeing the visual characteristics of products. However, to understand the changes in sorting in the whole category, it is necessary to take all the products into consideration. This is done in the next subsection.

6.2.2 Quantitative comparsion

The main goal of this section is to show to which extent products change their positions in the new sortings in comparison to the standard sorting, and that products are sorted correctly.

Position changes

For all sortings A-G and for all products, $\Delta position$, the absolute difference between the old and new product position, has been calculated (Equation 6.1).

$$\Delta position(\text{p}) = |\text{newPosition(p) - oldPosition(p)}| \tag{6.1}$$

where p is the current product.

The average $\Delta position$ is then calculated for each sortings, which can be found in Table 6.1 together with the total average per category.

The total average $\Delta position$ for Category1 is 25.67, which means that each product moved on average 25.67 positions out of 101 existing positions. For the *Category2* category, each product moved 7.69 positons on average out of 35 positions. In other words, products move approximately 75% of possible positions for both categories.

Sorting	**Average** $\Delta position$**:** **Category1 (101 products)**	**Average** $\Delta position$**:** **Category2 High (35 products)**
A	27.86	7.71
B	27.36	8.23
C	26.34	7.43
D	26.1	7.94
E	29.44	8.51
F	21.44	7.08
G	21.14	6.91
Total average $\Delta position$	25.67	7.69

Table 6.1: Average $\Delta position$ for different categories and sortings

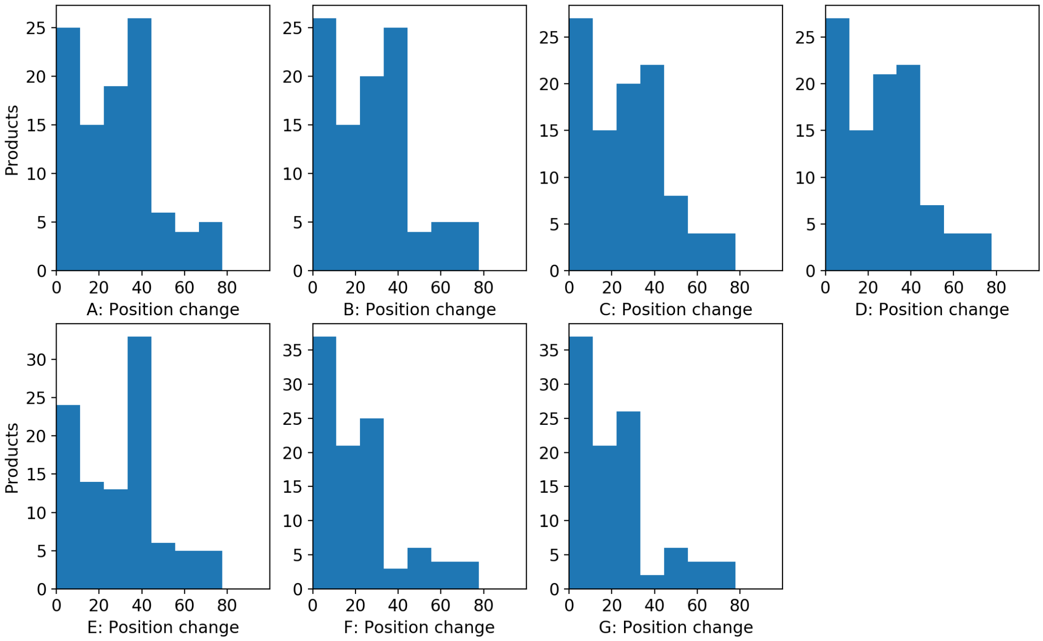


Figure 6.1: *Category1*: Histograms for products and position changes ($\Delta position$, bin range of 10)

To see the change in the positions better, histograms for the different sorting are generated.

Histograms show how many products have common $\Delta position$. The further and the higher the peaks are, the more changes in the product order occurred.

In the *Category1* category, histograms A and B, C and D, and F and G are similar. The reason for this is that sortings A-E are based on hourly changes and the changes between 10 am and 2 pm are smaller than 10 am and 8 pm due to the time difference. Histograms F and G are similar because they both have a seasonality factor, unlike the rest of the diagrams.

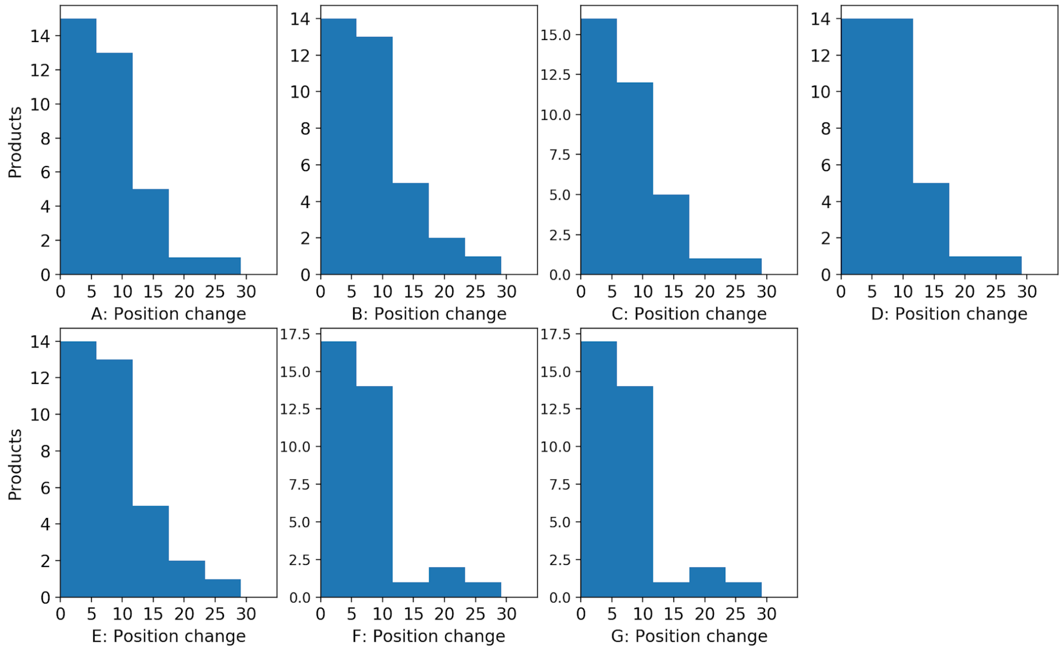


Figure 6.2: *Category2*: Histograms for products and position changes ($\Delta position$, bin range of 5)

The histograms A-E in the *Category2* category are also similar, while histograms F and G are identical. Once again, it is because they both have a seasonality factor.

Nonetheless, it is important to note that the bins in the histograms are merged which means that even though the histograms are similar, they are not exactly the same.

Score verification

As a part of the score verification, scores for the three products in sortings A (at 10 am) and F (in July), and the category *Category2* will be verified. This is enough to show that scores are calculated correctly, since the score calculation happens automatically. The scores for sortings A and F are shown in Tables 6.2 and 6.3 respectively.

Product	Amount sold at 10 am	Amount sold in total	Score
P1	50	986	0.507
P2	84	1.718	0.489
P35	4	158	0.025

Table 6.2: *Category2* A: Scores and quantities used for score calculation

Scores in the Tables 6.2 and 6.3 are taken from the website and in order to verify them *Amount sold at 10 am* is divided with *the Amount sold in total* for A sorting, and *Amount*

Product	Amount sold in July	Amount sold in total	Score
PS1	30	106	0.283
PS2	180	2.622	0.069
PS35	0	64	0

Table 6.3: *Category2* F: Scores and quantities used for score calculation

sold in July with *the Amount sold in total* for F sorting . The results of the division are equal to the scores which means that scores are calculated correctly. If the product was not sold at all in the month of testing, the score for that month is automatically zero. The order of products is also correct since the scores appear in the descending order.

To illustrate the calculations better, Figures 6.3 and 6.4 show how the scores change over hours and months for the three given products.

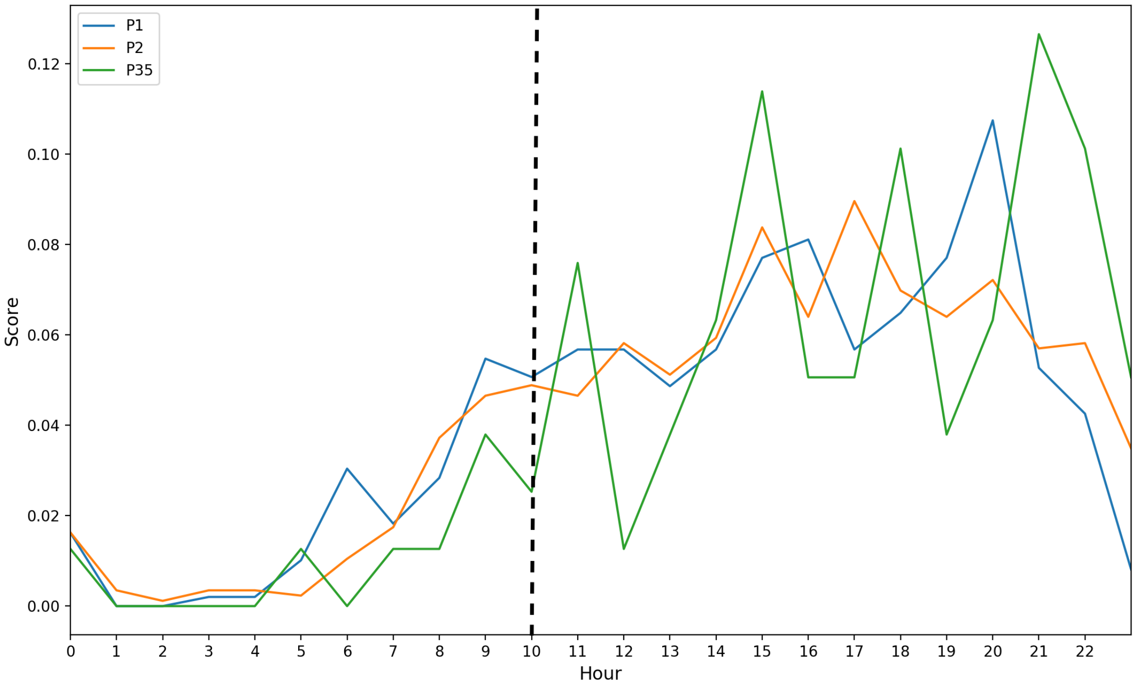


Figure 6.3: *Category2* A: *Hourly* score changes for P1, P2 and P35

Product P35 is sorted as the last one because it does not sell as well as the other two products at 10 am. However, at e.g. 3 pm and 10 pm it was sorted above them.

As for monthly changes, instead of showing all 365 scores, only one score per month is shown in Figure 6.4. Product PS1 is an *in-season* product, while products PS2 and PS35 are *out-of-season* products, and therefore have smaller scores than P1. This coresponds to the sorting order from the website.

Based on the results of both comparsions, it can be concluded that the new algorithms

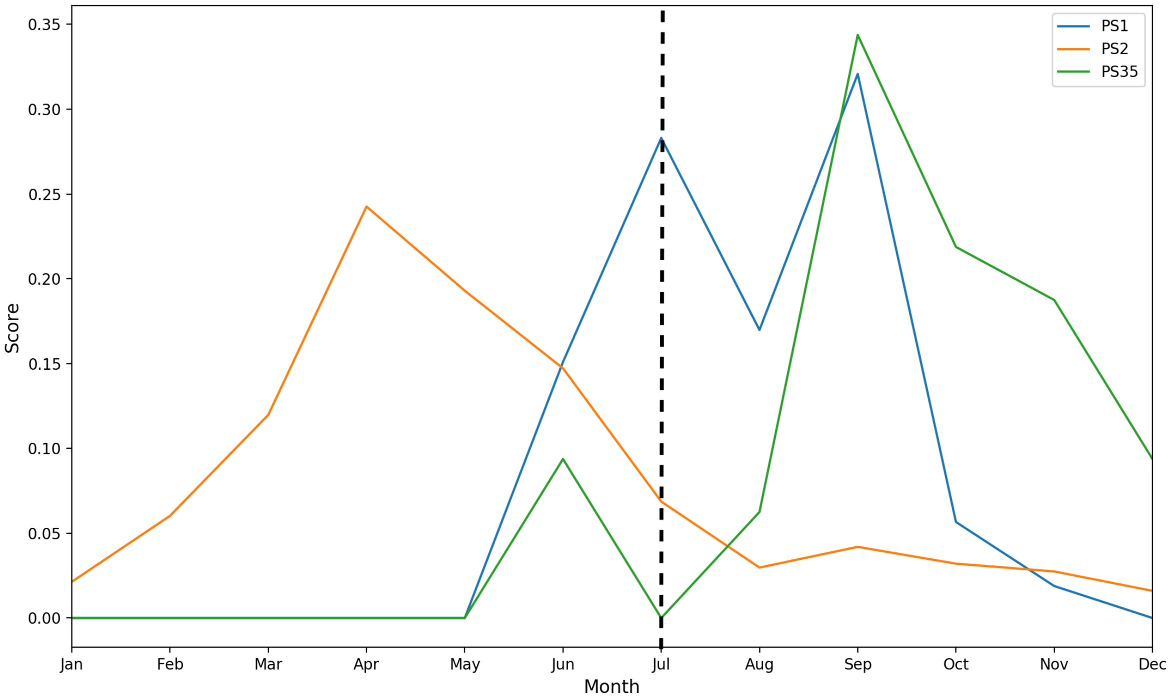


Figure 6.4: *Category2* F: *Monthly* score changes for PS1, PS2 and PS35

make a noticable difference in sorting and increase the variety of products shown at the top of the category page.

Summary:
This chapter evaluted the new sorting algorithms by means of visual and quantitative comparisons. The results have shown that there is a difference and more variety is present after using the new algorithms for sorting. The new product scores are also calculated as expected. To get better insight and evalution, it would be necessary to perform A/B tests on the Live websites to see if customers would find the products they like quicker and consequently, purchase more products.

7 Conclusion

Recommendation systems were first mentioned in 1990. and since then they continue to develop and spread. In spite of the advancements of the technology, there is a lot of space for improvement.

This thesis contributes to the existing solutions, by offering a new algorithm for product sorting optimisation by analysing both customer and product behaviour. The analysis was done using a specific e-commerce system within novomind AG. Some of the challenges faced were the limited amount of user information, no explicit users feedback e.g. reviews, need for universal solution and, as in case of all recommendation system, cold start problem - that sorting should be possible for new customers and products.

The final solution is based on the analysis of three hypotheses listed below with the summary for each hypothesis.

The product's sales depend on the month of the purchase. Some products are sold well in all months

A well-known fact in e-commerce is that there are products which sell well in specific seasons. However, many shop managers usually have to do this task manually, and apart from it being time-consuming, they might not be able to spot the products which do not look like they belong to the specific season e.g. sandals on sale which sell well in winter. The algorithm presented in this thesis automatically sorts products by season every day i.e. *in-season* products are followed by products sold throught the year, while *out-of-season* products are pushed to the bottom. The season consists of certain number of days and it since it always considers the specific number of recent days, it is able to take fast changes in e-commerce into account.

The product's sales depend on the time of the day. Some products are rarely sold at specific hours

Similar to the previous hypothesis, products can be classified as *all-day* products, selling well throughout the day, and *non-all day* products, selling well at specific hours. This has been shown in the hypothesis analyis and used for the sorting algorithm. The sorting works in such a way that products sold best at the current hour are pushed to the top and are followed by *non-all day* products, while other products which sell rarely or not at that hour are pushed to the bottom. What makes this algorithm most useful is the fact that it almost solves the cold-start problem. It is enough for a product to be online for just a couple of days before it can be sorted based on this hypoteses.

iPhone/Apple/iPhone&Android users spend more money/purchase greater quantities of expensive products than other users

This hypothesis had five different versions depending on the groups observed (iPhone/Apple/iPhone&Android users vs Others) and depending on what was compared (amount spent or percentage of expensive products). Since both iPhone users and iPhone&Android users spend more only in one shop, these hypotheses could not have been used. However, iPhone/Apple users had a larger percentage of expensive products in both shops and this can be used for sorting. This means that once an iPhone/Apple user comes to the website, products could be sorted by price (descending). It is important to mention that this hypothesis is useful only for Shop1 and Shop2. For usage in other shops, it would be necessary to investigate if their iPhone/Apple users spend more.

All the aforementioned algorithms can be used together, to produce one common ranking, or separately. They have been evaluated both together and separately by means of visual and quantitative comparison. The results show differences in sorting and more variety in top products.

7.1 Future work

For more precise evaluation of the alogorithms devloped, it would be necessary to do live A/B tests in order to see whether the new sorting increases customer satisfaction i.e. whether the customers make more purchases by finding the products they like quicker. Another aspect that needs more attention are new products and the scores they initally receive. Some of the options would be to give them the median of their belonging category, as implemented here, but they could also get the same score as the other products of the same type, brand etc.

In the long-term, those hypotheses can also be used in a bigger and more comprehensive machine learning solution e.g. as input to artificial neural networks.

Recommendation systems have a long way to go and the greatest problem to be solved is likely to be finding an optimal trade-off between showing a personalised solution which is not extremely personalised or too intrusive while allowing users to discover new things.

Bibliography

[AP17] Eshan Gupta Aditya Parashar. ANN Based Recommendation Algorithm for the Product of E-commerce. *International Journals of Advanced Research in Computer Science and Software Engineering*, 7(6), 2017.

[Aza12] Muhammad Azam. iphone vs. android: iphone users spend more time on their handsets. `http://infinigeek.com/iphone-vs-android-iphone-users-spend-more-time-on-their-handsets/`, 2012. [Online; Accessed: 2018-06-05].

[Col17] Alan Coleman. E-commerce kpi benchmarks 2017. `https://www.wolfgangdigital.com/uploads/case-studies/Wolfgang_Digital_2017_E-commerce_KPI_Benchmarks_Study.pdf`, 2017. [Online; Accessed: 2018-06-05].

[FRE13] Henrik Sattler Franz-Rudolf Esch, Andreas Herrmann. *MARKETING: Eine managementorinetierte Einführung.* Vahlen, 2013.

[GP01] Olga Papaemmanouil George Prassas, Katherine C. Pramataris. Dynamic Recommendations in Internet Retailing. 2001.

[GP02] Olga Papaemmanouil Georgios J. Doukidis George Prassas, Katherine C. Pramataris. A recommender system for online shopping based on past customer behaviour. 2002.

[Inc16] Adobe Systems Incorporated. Adobe reports mobile sales records on thanksgiving day, black friday. `https://www.businesswire.com/news/home/20131129005408/en/Adobe-Reports-Mobile-Sales-Records-Thanksgiving-Day`, 2016. [Online; Accessed: 2018-06-05].

[Jam] James Lani. Chi-square test of independence. `http://www.statisticssolutions.com/non-parametric-analysis-chi-square/`. [Online; Accessed: 2018-06-05].

[JQ15] Yinghong Li Jiangtao Qiu, Zhangxi Lin. Predicting customer purchase behavior in the e-commerce context. *Springer Science+Business Media*, 2015.

[Kul] Kullback–Leibler divergence. Kullback–leibler divergence — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/wiki/Kullback%E2%80%93Leibler_divergence`. [Online; Accessed: 2018-06-05].

[Lu14] Haiyun Lu. Recommendations Based on Purchase Patterns. *International Journal of Machine Learning and Computing*, 4(6), 2014.

[Man] Mann–Whitney U test. Mann–whitney u test — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/wiki/Mann%E2%80%93Whitney_U_test`. [Online; Accessed: 2018-06-05].

[MR17] Sachin Godse Manish Raka. Implementing Product Recommendation System using Neural Network by Connection Social Networking to E-Commerce. 3(8), 2017.

[Qiu14] Jiangtao Qiu. A predictive model for customer purchase behaviour in e-commerce context. *AIS Electronic Library*, 2014.

[Sta] Stat Trek. Hypothesis test: Difference between means. `http://stattrek.com/hypothesis-test/difference-in-means.aspx`. [Online; Accessed: 2018-06-05].

[TL] Xiangxiang Meng David Duling Taiyeong Lee, Yongqiao Xiao. Clustering Time Series Based on Forecast Distributions Using Kullback-Leibler Divergence.

[YJP17] Kun-Nyeong Chang You-Jin Park. Individual and group behavior-based customer profile model for personalized product recommendation. *Expert Systems with Applications*, 2017.

List of Figures

List of Tables

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit im Studiengang Informatik selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel – insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen – benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht. Ich versichere weiterhin, dass ich die Arbeit vorher nicht in einem anderen Prüfungsverfahren eingereicht habe und die eingereichte schriftliche Fassung der auf dem elektronischen Speichermedium entspricht.

Ort, Datum Unterschrift

Veröffentlichung

Ich bin damit einverstanden, dass meine Arbeit in den Bestand der Bibliothek des Fachbereichs Informatik eingestellt wird.

Ort, Datum Unterschrift