

Česká zemědělská univerzita v Praze

Provozně ekonomická fakulta

Katedra informačních technologií



Bakalářská práce

Produktová analytika za použití dat z Google Analytics

Tomáš Vopička

© 2023 ČZU v Praze

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Provozně ekonomická fakulta

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Tomáš Vopička

Informatika

Název práce

Produktová analytika za použití dat z Google Analytics

Název anglicky

Product analytics base on Google Analytics data

Cíle práce

Hlavním cílem bakalářské práce je využití aktuálně nevyužívaných produktových dat pro lepší uživatelskou zkušenost a obchodní zisk firmy.

Dílním cílem teoretické části práce je:

- deskripce webové analytiky a měření na webu,
- realizace několika analýz
- interpretace výsledku,
- návrh použití výsledků.

Metodika

Práce je založena na studiu vědecké a odborné literatury. V teoretické části práce bude provedena deskripce a charakteristika základů webového měření jak z pohledu implementačního, technického tak uživatelského. Budou analyzovány jednotlivé základní metriky jako Návštěvy, Uživatelé, Míra okamžitého opuštění, Zobrazení a další.

V praktické části bude analýza uživatelské chování v rámci nákupního procesu, výkon jednotlivých produktů a hledání produktů s nevyužitým potenciálem, identifikaci zástupných produktů, využívání seznamu uživatelů.

Na základě zjištěných faktů bude syntetizován závěr práce.

Doporučený rozsah práce

40 – 50 stránek

Klíčová slova

webová analytika, produktová analytika, Google Analytics, analýza, web

Doporučené zdroje informací

- BRUNEC, Jan. Google analytics. Praha: Grada Publishing, 2017. Průvodce (Grada). ISBN 978-80-271-0338-6.
- CLIFTON, B. *Google Analytics : podrobný průvodce webovými statistikami*. Brno: Computer Press, 2009. ISBN 978-80-251-2231-0.
- Dan Croxen-John. Johann Van Tonder. Optimalizace webových stránek: Proč si 95% návštěvníků vašich stránek nic nekoupí a jak to můžete změnit. Lingea, 2022. ISBN 978-80-7508-716-4
- DAVIS, Joel. MOKALIS Alexa. Google Analytics Demystified (4th Edition). CreateSpace, 2018. ISBN 978-1545486917
- KAUSHIK, A. *Webová analytika 2.0 : kompletní průvodce analýzami návštěvnosti*. Brno: Computer Press, 2011. ISBN 978-80-251-2964-7.
- RIMAN Tarek. The Secret to Capitalizing on Analytics: A Web Analytics Approach for Beginners. Independently published (April 11, 2019). ISBN 978-1796616194
- SHARMA, Himanshu. Maths and Stats for Web Analytics and Conversion Optimization. Blurb, 2020. ISBN 978-1364849184
- TONKIN, S. – WHITMORE, C. – CUTRONI, J. *Výkonnostní marketing s Google Analytics*. Brno: Computer Press, 2011. ISBN 978-80-251-3339-2.

Předběžný termín obhajoby

2022/23 LS – PEF

Vedoucí práce

Ing. Alexandr Vasilenko, Ph.D.

Garantující pracoviště

Katedra informačních technologií

Elektronicky schváleno dne 14. 7. 2022

doc. Ing. Jiří Vaněk, Ph.D.

Vedoucí katedry

Elektronicky schváleno dne 27. 10. 2022

doc. Ing. Tomáš Šubrt, Ph.D.

Děkan

V Praze dne 11. 03. 2023

Čestné prohlášení

Prohlašuji, že svou bakalářskou práci "Produktová analytika za použití dat z Google Analytics" jsem vypracoval samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou citovány v práci a uvedeny v seznamu použitých zdrojů na konci práce. Jako autor uvedené bakalářské práce dále prohlašuji, že jsem v souvislosti s jejím vytvořením neporušil autorská práva třetích osob.

V Praze dne 15.3.2023

Poděkování

Chtěl bych touto formou poděkovat mému vedoucímu bakalářské práce Ing. Alexandru Vasilenkovi, Ph.D., který mi pomohl a poradil při psaní práce. Dále bych chtěl poděkovat zástupcům firmy Sportisimo, kteří mi poskytli data a cenné informace o firmě.

Produktová analytika za použití dat z Google Analytics

Abstrakt

Tato bakalářská práce se věnuje produktové analytice za použití dat z Google Analytics, především využití dosud nepoužívaných dat k získání dalších informací o zákaznících. Dílčím cílem práce je hledání produktů s nevyužitým potenciálem a tedy, zda zákazníci kupují, to co si přidali na seznam oblíbených výrobků, interpretace výsledků analýz a návrh na použití zjištěných výsledků. Hlavním důvodem pro výběr tohoto tématu mé bakalářské práce byl můj zájem o to, jak mohou firmy data z webových stránek získat a jakým způsobem je mohou využít. První část práce se věnuje teoretickým poznatkům o webové analytice, službě Google Analytics a analytickým metodám pro práci s daty. Praktická část se zaměřuje na analýzu uživatelského chování internetového obchodu společnosti Sportisimo. Tato část se zabývá hledáním produktů s nevyužitým potenciálem a tedy, zda zákazníci kupují to, co si přidali na seznam oblíbených výrobků, interpretaci výsledků a návrh na použití zjištěných výsledků.

Klíčová slova: webová analytika, produktová analytika, Google Analytics, analýza, metriky, konverzní poměr, míra opuštění, analýza chování, analýza návštěvnosti, analýza nákupního košíku

Product analytics base on Google Analytics data

Abstract

This bachelor's thesis is devoted to product analytics using data from Google Analytics, especially the use of previously unused data to obtain additional information about customers. The partial goal of the work is the search for products with unused potential and therefore whether customers buy what they have added to the list of favorite products, the interpretation of the analysis results and a proposal for the use of the results. The main reason for choosing this topic of my bachelor's thesis was my interest in how companies can obtain data from websites and in what way they can use it. The first part of the thesis is devoted to theoretical knowledge about web analytics, the Google Analytics service and analytical methods for working with data. The practical part focuses on the analysis of user behavior of Sportissimo's online store. This part deals with the search for products with untapped potential and therefore whether customers buy what they have added to the list of favorite products, the interpretation of the results and a proposal for the use of the found results.

Keywords: web analytics, product analytics, Google Analytics, analysis, metrics, conversion rate, bounce rate, behaviour analysis, traffic analysis, shopping cart analysis

Obsah

1 Úvod.....	10
2 Cíl práce a metodika	11
2.1 Cíl práce	11
2.2 Metodika.....	11
3 Teoretická východiska	12
3.1 Úvod do webové analytiky.....	12
3.2 Nástroj Google Analytics	16
3.2.1 Historie Google Analytics.....	16
3.2.2 Rozdíl mezi verzemi Google Analytics	17
3.2.3 Nástroj Matomo	17
3.2.4 Metodika měření	18
3.3 Dimenze a Metriky	21
3.3.1 Dimenze	21
3.3.2 Datové kostky	21
3.3.3 Metriky.....	22
3.3.4 Základní metriky	23
3.3.5 Životní cyklus použití metrik.....	25
3.4 Analytické úlohy	26
3.4.1 Segmentace	26
3.4.2 Analýza nákupního košíku.....	26
3.5.1 A/B testování.....	27
4 Vlastní práce	28
4.1 Představení firmy	28
4.2 Struktura použitých datasetů	29
4.2.1 Události produktů.....	29
4.2.2 Produktový katalog.....	30
4.2.3 Transakční data	31
4.2.4 Nákupní listy	31
4.2.5 Nákupy zákazníků.....	32
4.3 Agregace dat.....	33
4.3.1 Denní agregace.....	33
4.3.2 Agregace na úrovni produktu a měsíce.....	34
4.3.3 Produktová agregace	36
4.4 Závislost kroků v nákupním procesu.....	37
4.4.1 Základní myšlenka	37
4.4.2 Chování zákazníka v datech.....	37

4.4.3	Analýza kroků Denní S1-S2	38
4.5	Porovnání datasetů	42
4.5.1	Závislost kroků na denní agregaci – dataset „All“	42
4.5.2	Závislost kroků na denní agregaci – rozdělené datasety	43
4.6	Meziměsíční vývoj	45
4.7	Srovnání na úrovni produktů	46
4.7.1	Produkty s nevyužitým potenciálem	46
4.7.2	Hledání alternativních produktů	48
4.8	Analýza používání funkce „Na seznam“	50
4.8.1	Přidané produkty na seznam	51
4.8.2	Počet dní od přidání po nákup	52
4.8.3	Průměrný počet dní od znovu použití	53
5	Výsledky a diskuse	54
6	Závěr.....	56
7	Seznam použitých zdrojů	58
8	Seznam obrázků, tabulek, grafů a zkratk.....	60
8.1	Seznam obrázků	60
8.2	Seznam tabulek	60
8.3	Seznam grafů.....	60

1 Úvod

Využívání různých webových stránek patří k běžným činnostem naprosté většiny z nás. Už delší dobu mě zajímalo, jak webové stránky fungují z pozice firem a jak mohou firmy data ze svých webových stránek získat a jakým způsobem je mohou využít. Proto jsem si jako téma bakalářské práce vybral produktovou analytiku za použití dat z Google Analytics.

Téměř všechny firmy s daty o návštěvnicích webů již pracují, proto jsem se zaměřil na data, která firma sice má k dispozici, ale zatím je nijak nevyužívá. Cílem mé bakalářské práce tedy je využití dosud nepoužívaných dat k tomu, aby firma získala další informace o svých zákaznících a případně tak zvýšila svůj obchodní zisk.

V teoretické části je představena webová analytika, služba Google Analytics a její nástroje a analytické metody pro práci s daty. Praktická část se zaměřuje na analýzu uživatelského chování internetového obchodu společnosti Sportisimo. Tato část se zabývá hledáním produktů s nevyužitým potenciálem a tedy, zda zákazníci kupují to, co si přidali na seznam, interpretací výsledků a návrh na použití zjištěných výsledků.

2 Cíl práce a metodika

2.1 Cíl práce

Hlavním cílem bakalářské práce je využití aktuálně nevyužívaných produktových dat pro lepší uživatelskou zkušenost a obchodní zisk firmy.

Dílčí cíle:

- hledání produktů s nevyužitým potenciálem

- nakupují zákazníci to, co si přidali na seznam
- atraktivita produktu
- interpretace výsledků
- návrh použití výsledků

2.2 Metodika

Práce je založena na studiu vědecké a odborné literatury.

V teoretické části práce budou popsány (základy) charakteristika základů webového měření jak z pohledu implementačního, technického tak uživatelského. Budou analyzovány jednotlivé základní metriky jako Návštěvy, Uživatelé, Míra okamžitého opuštění, Zobrazení a další.

V praktické části bude analýza uživatelské chování v rámci nákupního procesu, výkon jednotlivých produktů a hledání produktů s nevyužitým potenciálem, identifikaci zástupných produktů, využívání seznamu uživatelů. Na základě zjištěných faktů bude syntetizován závěr práce.

3 Teoretická východiska

3.1 Úvod do webové analytiky

Webová analytika cílí především na sběr, měření, sledování a vyhodnocování internetových dat. Webový analytik Stéphane Hamel jako definici uvádí: „Webová analytika zahrnuje pokročilé využívání kvalitativních a kvantitativních dat, statistických analýz, exploračních a prediktivních modelů za účelem kontinuálního vylepšování online aktivit a lepší návratnosti investic.“ (1)

Kaushik definuje webovou analytiku jako analýzu kvalitativních a kvantitativních dat z vlastního webu i webů konkurence, a to za účelem kontinuálního vylepšování uživatelského prožitku současných i potenciálních zákazníků, což se promítá do požadovaných výsledků na webu i mimo něj. (2)

Na základě výše uvedených definic můžeme konstatovat, že webová analytika se zabývá zjišťováním informací o chování návštěvníků na webových stránkách, například kolik návštěvníků a odkud na web přichází, jak se na stránkách pohybují, jaká klíčová slova zadávají do vyhledávání. Webová analytika firmám pomáhá s pochopením současného fungování a podporou manažerského rozhodování pro optimalizaci.

Typy dat ve webové analytice

Veškeré údaje, které lze získat a zanalyzovat se dělí na kvantitativní a kvalitativní data.

Kvantitativní data jdou vyjádřit číselně a měří, co a jak se událo. Nejčastěji se měří zobrazování stránek na webu, prokliky z newsletteru, přihlášení k odběru newsletteru, transakce, míra opuštění, objednávky a další různé typy interakcí. Většinou se data sbírají přes JavaScript měřicí kód uvedený na stránce. (3)

Díky kvalitativním datům lze získat hlubší porozumění tomu, proč se určité okolnosti dějí nebo staly. Tyto informace jsou nejčastěji získány pomocí různých metod výzkumů, jako například dotazníky nebo jiné kvalitativní techniky na vzorku návštěvníků. (3)

Ve webové analytice se používá mnoho různých typů nástrojů tzv. clickstream data, která popisují pohyb návštěvníků po webu. Jinými slovy je to forma analýzy, která sleduje a analyzuje návštěvníky webových stránek. (2)

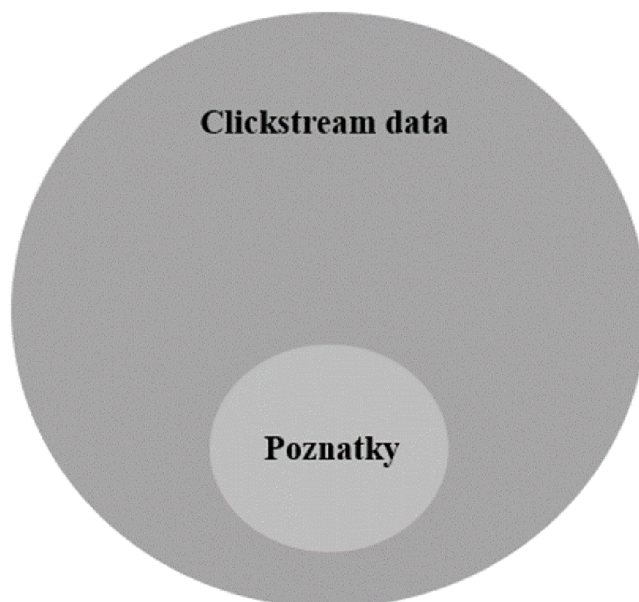
Clickstream data mohou zahrnovat následující typy informací o návštěvnících:

- Zda je návštěva webu jedinečná nebo opakovaná.
- Jakou část webu navštíví jako první.
- Čas, jaký na webové stránce stráví.
- Kdy a jak je položka přidána nebo odebrána z nákupního košíku.
- Kam uživatel odchází dál.
- Kdy použije tlačítko zpět.

(4)

Avšak i clickstream data mají svá omezení. Jedno z hlavních omezení clickstream dat je znázorněno v obrázku níže:

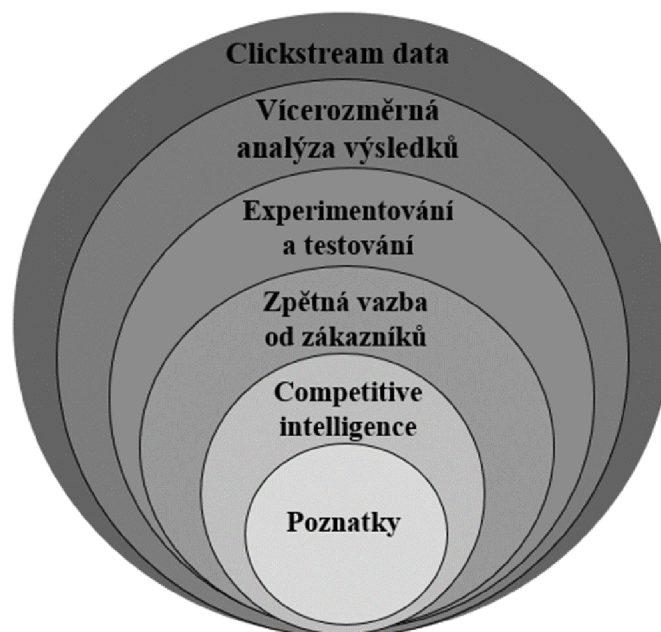
Obrázek 1 Staré paradigma webové analytiky



Obrázek vlastní úprava, zdroj: (2)

Kaushik zavádí pojem webová analytika 2.0, protože se dle něj začíná přehodnocovat význam webové analytiky a využití jejích nástrojů. Webová analytika 2.0 se zabývá jednak analýzou kvalitativních a kvantitativních dat, ale také analýzou dat z konkurenčních webových stránek. Díky tomu lze získat odpovědi na složitější otázky o uživatelích, proč se daná věc děje a pochopit tak lépe chování zákazníků. (2)

Obrázek 2 Klíčové otázky spojené s webovou analytikou 2.0



Obrázek vlastní úprava, zdroj (2)

3.2 Nástroj Google Analytics

Google Analytics je online analytické řešení od společnosti Google Inc. V současnosti je Google Analytics hlavním nástrojem pro 94 % nejvíce recenzovaných českých internetových obchodů. Existují také firmy, které využívají alternativní nástroje jako je například Adobe SiteCatalyst, Webtrends nebo Matomo. Google Analytics slouží k získávání dat o uživatelích webových stránek. Pomocí nástroje Google Analytics je možné sledovat jak aktuální, tak historická data, která se týkají chování uživatelů na webových stránkách, a to včetně informací o jejich vlastnostech, jako je doba kterou uživatel na stránce stráví, kolikrát web navštívil opakovaně nebo položky které přidal do svého nákupního košíku. Jako základní princip můžeme označit sběr analytických dat z měřeného webu, jejich a finální zobrazení vybraných dat. (5) (6) (7)

Služba Google Analytics funguje pomocí měřicího kódu. Tento kód se musí vložit na všechny stránky, které mají být do měření zahrnuty. Měřicí kód sleduje, které stránky si návštěvníci zobrazují. Některé údaje ukládá do anonymního souboru cook. (5)

3.2.1 Historie Google Analytics

Služba Google Analytics vznikla v roce 2005. Během této doby prošla několika vývojovými fázemi. V roce 1997 byl vyvinut software Urchin samostatnou firmou, kterou Google koupil v roce 2005 a přejmenoval ho na Google Analytics. Poté v roce 2007 přišla druhá vývojová fáze Classic Analytics. Služba se postupně rozšiřovala o nové funkcionality, a tak v roce 2012 vznikla další fáze Universal Analytics. Od roku 2016 začala fungovat i placená verze Google Analytics Premium, která byla později přejmenována na Google Analytics 360. Nejnovější verze, která se jmenuje Google Analytics 4, byla vydána v roce 2020. Jednou z největších novinek této verze je, že umožňuje měřit společně webové stránky i mobilní aplikace. (5) (8)

3.2.2 Rozdíl mezi verzemi Google Analytics

Služba Google Analytics je pro uživatele dostupná ve dvou modifikacích. Primární verzi lze využít zdarma. V neplacené verzi uživatelé narazí především na omezení limitů a kvót pro počet požadavků na server Google Analytics a to konkrétně na 10 milionů požadavků měsíčně. V základní verzi může nastat i prodleva mezi měřením a zpracováním dat. (9)

Základní verzi doplňuje placená verze Google Analytics 360, jejíž ceníková cena začíná na 135 000 eurech měsíčně. Placená verze zvyšuje limit na 500 000 000 požadavků měsíčně. Tato verze garantuje maximální dobu zpracování na 4 hodiny. Mezi další výhody oproti základní verzi patří navýšení limitů:

- 200 vlastních dimenzí a 200 vlastních metrik (místo dosavadních 20 + 20),
- 50 kalkulovaných metrik (místo dosavadních 5),
- export až 3 000 000 řádků (místo dosavadních 5 000),
- kardinalita tabulky až 500 000 řádků denně (místo dosavadních 75 000),
- až 2 000 požadavků na jednu návštěvu (místo standardních 500).

(9)

3.2.3 Nástroj Matomo

Nástroj Matomo, dříve známý pod názvem Piwik, se řadí mezi jednu z nejpobulárnějších alternativ k nástroji Google Analytics. Je to software s otevřeným zdrojovým kódem, který lze stáhnout zdarma. Aplikace funguje pomocí skriptu, který se vloží do zdrojového kódu stránky, pro kterou se bude analýza provádět. Velkou výhodou oproti Google Analytics je fakt, že provozovatelé webových stránek ukládají všechna data na svých vlastních serverech. Nezávislým hostováním dat mají provozovatelé webových stránek plnou kontrolu nad všemi daty. Zároveň je to výhodné z hlediska ochrany údajů, protože Matomo nevyžaduje od návštěvníků webu souhlas se sledováním. (10) (11)

3.2.4 Metodika měření

Ke sledování provozu na webových stránkách služba Google Analytics využívá měřicí kód v jazyce JavaScript. Tento kód se nachází na všech stránkách webu, sbírá údaje o návštěvníkovi webu a odesílá je na servery společnosti Google. (12)

Služba Google Analytics, i konkurenční služby jako Adobe Analytics nebo Matomo, využívají pro všechna měření měřicí kód. Stejně jako se vyvíjí celé odvětví webové analytiky, tak se vyvíjí i nástroje a měřicí kódy. U společnosti Google byl vývoj následující:

- Google Analytics rok 2005 měřicí kod ga.js
- Classic Analytics rok 2007 analytics.js
- Universal Analytics rok 2012 gtag.js
- Google Analytics Premium nebo 360 rok 2016
- Google Analytics 4 rok 2020

(8)

Originální kód ga.js je dnes již označován jako zastaralý, Google doporučuje migraci na novější způsoby měření.

Obrázek 3 Ukázka vložení dat do dataLayeru

```
<script type="text/javascript">
  var _gaq = _gaq || [];
  _gaq.push(['_setAccount', 'UA-XXXXX-X']);
  _gaq.push(['_trackPageview']);

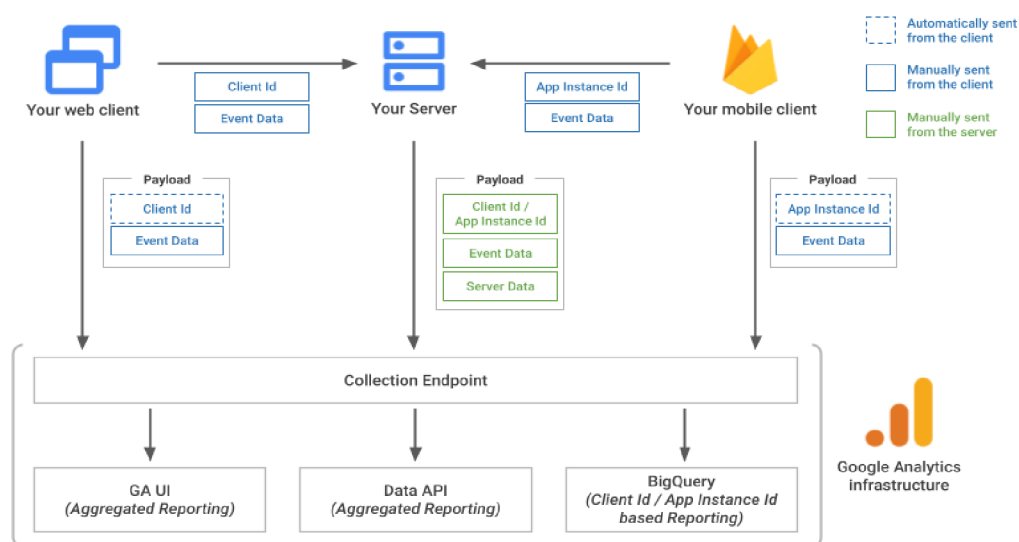
  (function() {
    var ga = document.createElement('script'); ga.type = 'text/javascript';
    ga.async = true;
    ga.src = ('https:' == document.location.protocol ? 'https://ssl' :
'http://www') + '.google-analytics.com/ga.js';
    var s = document.getElementsByTagName('script')[0];
    s.parentNode.insertBefore(ga, s);
  })();
</script>
```

Obrázek zdroj: (13)

Implementace měřicího kódu přímo do kódu webu je dnes již také historií. Samozřejmě je to možné, ale pro tyto účely se dnes již používají značkovací manažeři. Jedním takovým je Google Tag Manažer (GTM). Jedná se o prostředníka mezi nástroji a webem. (13)

Pomocí tohoto manažera lze do webu vkládat a spravovat různé kódy jako například Google Analytics nebo měřicí kódy marketingových platforem, kde firmy inzerují. Významnou změnou bylo zavedení takzvané datové vrstvy sloužící k předávání dodatečných informací, které se dají využít pro analytické nebo marketingové účely. Výhodou tohoto přístupu je to, že taková hodnota je pomocí datové vrstvy předána GTM a ten ji pak může distribuovat do více systémů. V roce 2014 byl také představen Measurement protokol, který umožňoval nový způsob vkládání dat do Google Analytics a to dokonce i informace, které se staly po konverzi návštěvníka jako například storno objednávky. Nejvýznamnější změnu však přináší nástroj Google Analytics 4, který umožňuje měřit webové prezentace, a i mobilní aplikace společně. Na následujícím obrázku je architektura Measurement Protokolu aktuální verze Google Analytics 4. (13)

Obrázek 4 Přehled architektury



Obrázek zdroj: (14)

Architektura znázorňuje možnosti měření od klasického přes server-site, až po měření mobilních aplikací. Data se pak ukládají do BigQuery, odkud se zpracovávají v UI, případně k nim lze přistupovat přes API nebo v Google Analytics UI. Další součástí měření jsou také cookies, což jsou krátké textové soubory umožňující zjistit, odkud návštěvník přichází, jaký typ operačního systému používá, jaký typ prohlížeče používá či jakou má nastavenou jazykovou mutaci. (13) (5)

Cookies se obecně dají rozdělit do následujících skupin:

- Bezpečnostní
- Analytické
- Marketingové
- A další...

(5) (15)

Různé nástroje samozřejmě mohou vytvářet vlastní cookies dle potřeb každého nástroje. Po sběru všech těchto údajů sledovací kód nastaví několik souborů cookies na počítači návštěvníka. Využitím cookies jsou do Google Analytics zasílána data například o tom, jestli návštěvník zobrazil stránku poprvé nebo ji již v minulosti navštívil, případně kolikrát se za určitou dobu vrátil, nebo jak dlouho trvalo než stránku opět navštívil. (15)

Pro službu Google Analytics je využívání souborů cookie klíčové. V případě, kdy služba Google Analytics nemůže nastavit soubory cookie na počítači návštěvníka, není možné ho sledovat. Služba Google Analytics využívá ke sledování návštěvníků webů čtyři případně pět vlastních souborů cookie. (12)

Tabulka 1 Základní sledovací soubory cookie služby Google Analytics

Název souboru cookie	Účel
_utma	Uchovává identifikátory návštěvníka. Obsahuje číselný identifikátor. Služba Google Analytics přepočítává soubor _utma, aby sledovala počet návštěvníků daného webu, známých jako Jedineční návštěvníci.
_utmb	Uchovává identifikátory relace. Služba Google Analytics používá tento soubor cookie a soubor _utmc k výpočtu časově založených metrik jako je např. Doba na stránce a Doba na webu.
_utmc	Uchovává identifikátory relace. Služba Google Analytics používá tento soubor cookie a soubor _utmb k výpočtu časově založených metrik jako je např. Doba na stránce a Doba na webu.
_utmZ	Uchovává zdroj návštěvníka. Identifikuje, odkud návštěvník této konkrétní návštěvy pochází. Sleduje např. marketingové kampaně, klíčová slova a odkazující stránky.
_utmv	Uchovává vlastní proměnné. Tento soubor cookie se používá s funkcí zvanou Vlastní proměnné pro uložení informací, které se se asociují s návštěvníky webu.

Tabulka vlastní zpracování, zdroj: (12)

3.3 Dimenze a Metriky

Dimenze a metriky tvoří základní strukturu ukládaných dat, dimenze mají charakter kategoričkých informací a říkají čeho se data týkají, zato metriky jsou počítatelné. (16)

3.3.1 Dimenze

Dimenze je vlastnost měřené hodnoty tedy většinou vlastnost návštěvníka. Příkladem dimenzí může být například geografická poloha, zdroj zařízení, varianta produktu nebo typ stránky. Dalším příkladem dimenzí může být také čas. Dimenze tvoří v Google Analytics obvykle první sloupec tabulky, při použití sekundární dimenze sloupečky dva. (5) (17)

3.3.2 Datové kostky

Jsou druhem vícedimenzionální matice, která uživatelům umožňuje zkoumat a analyzovat soubor dat z mnoha různých perspektiv, vyhledat a vytvořit mezi nimi souvislosti a vazby. Datové kostky mají obvykle tři rozměry – dimenze. S narůstajícím počtem rozměrů se kostky stávají rozptýlenějšími – mnoho buněk představujících specifické kombinace atributů je tedy prázdných, neobsahují žádné agregované údaje. (18) (19) (20)

Datové kostky agregují metriku (často označovanou jako fakt uloženou v tabulce faktů), například "display_cnt" nebo "clicked_cnt" přes dimenze. Umožňují uspořádat velké objemy dat tak, aby bylo možné získat srozumitelné výstupy, které jsou využívány pro analýzu a rozhodování. (18) (21) (20)

3.3.3 Metriky

Pod pojmem metrika se ve webové analytice rozumí statisticky měřitelná veličina nebo ukazatelé, které slouží k ohodnocení výkonů. Metriky jsou kvantitativní měření, které charakterizuje jednotlivé dimenze. Metriky se především používají v souvislosti s webovou analytikou k analýze chování návštěvníků na webu nebo k měření výsledků marketingových online kampaní. Pod pojmem klíčové indikátory výkonu se rozumí metrika, která pomáhá porozumět, jak se posunout ke stanoveným cílům a určit, zda jsou cíle dosaženy. (2) (16)

Dimenze mohou obsahovat jednu nebo více metrik. Například dimenze „zdroj návštěv“ může nabývat těchto metrik:

- Počet návštěv
- Počet nových návštěvníků
- Průměrný čas strávený na stránce
- Míra odchodu
- Počet konverzí (5)

Zde je uvedeno pouze několik příkladů, samozřejmě je také možné si vytvořit své vlastní metriky, podle potřeby a uzpůsobit je potřebné analýze.

3.3.4 Základní metriky

Celkový výčet metrik by byl velmi dlouhý, proto je zde uvedeno pouze několik základních a nejpoužívanější metrik a dále pak metriky vztahující se k práci samotné.

Návštěvy a návštěvníci

Návštěvy a návštěvníci tvoří základ prakticky každého výpočtu webové metriky.

Návštěvy (sessions)

Návštěvy z anglického sessions, někdy uváděny také jako relace, jsou definovány jako doba, po kterou je uživatel aktivní na webu nebo v aplikaci. Ve výchozím nastavení, pokud je uživatel neaktivní po dobu 30 minut nebo déle, je jakákoliv budoucí aktivita připsána nové relaci. Uživatelé, kteří opustí web a vrátí se do 30 minut, se počítají jako součást původní relace. (2)

Unikátní návštěvníci

Matematický odhad skutečné návštěvnosti založené na otisku prohlížeče uloženém v cookies. Výpočet není znám, ale zahrnuje problematiku povolených cookies i fakt, že za daným zařízením nemusí vždy sedět stejná osoba. Nelze zaměňovat s počtem návštěv, který udává, kolikrát uživatel přišel na web. (2) (22)

Čas strávený na stránce a Čas strávený na webu

Další významnou metrikou, hned po návštěvách a návštěvnících, je ve webové analytice čas. Jedná se o měření času, který návštěvníci stráví na jednotlivé stránce a o čas, po který se pohybují po celém webu, a to v rámci jedné relace. (2)

Míra opuštění

Jedná se o metriku, která je standardně obsažena v mnoha nástrojích pro webovou analytiku. Dá se použít na více úrovních, zejména při identifikaci lehce opravitelných problémů. Přesnější definice technického charakteru říká, že se jedná o procento relací, které při návštěvě webu skončí prohlédnutím stránky. Za optimální hodnotu se považuje hodnota pod 50 %. Průměrná hodnota míry opuštění se pohybuje mezi 26 % a 40 %.

(2) (23) (24) (25)

Míra odchodů

Míra odchodů ukazuje, kolik uživatelů opustilo web z určité zadané stránky. Problém nastává v tom, že každý, kdo navštíví web, musí někudy také odejít. Uživatelé budou typicky procházet stránky webu a opouštět je prakticky libovolnou z internetových stránek tohoto místa. Ovšem jejich odchod ze stránek nemusí v tomto případě nutně znamenat přílišnou velikost nebo nějaký nedostatek na stránce. (2) (24) (25)

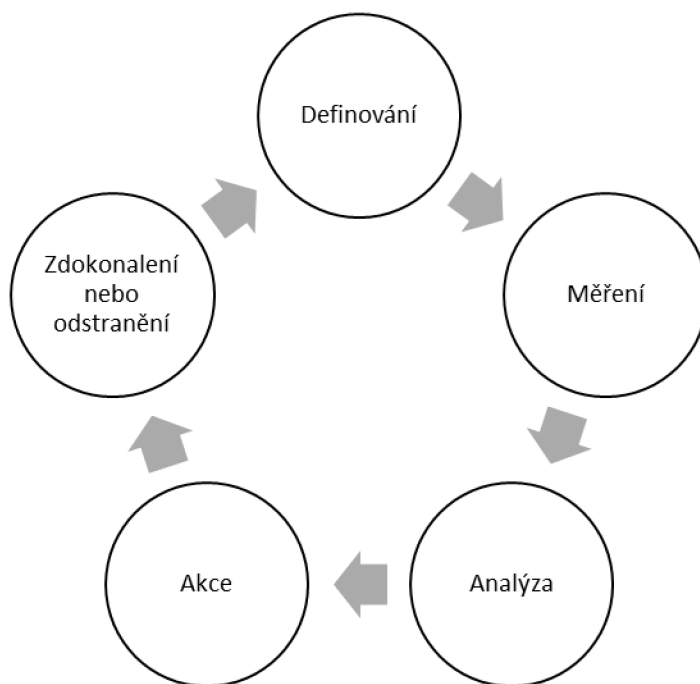
Míra konverze

Míra konverze vyjádřená v procentech se definuje jako počet konverzí dělený počtem unikátních návštěvníků (nebo návštěv). Počet konverzí je obvykle nabídka objednávek na webu zaměřeném na elektronické obchodování. Optimální konverzní poměr se pohybuje od 2 % do 5 %. (2) (26) (27)

3.3.5 Životní cyklus použití metrik

Pokud se rozhodneme pro využití metriky nebo metrik pro vyhodnocení výkonu, aktivit nebo něčeho jiného, je třeba také říct, že webová analytika je v podstatě iterativní proces, který se týká i používání samotných metrik. (2)

Obrázek 5 Životní cyklus webových metrik



Obrázek vlastní zpracování, zdroj: (2)

Nejlépe je si pro začátek zvolit více metrik, které se opakovaným použitím a rozšíří nebo eliminují na počet metrik, které budou vyhovovat danému použití. (2)

3.4 Analytické úlohy

3.4.1 Segmentace

Ve webové analytice se často používá pojem segmentace, která umožňuje rozdělit návštěvníky a zákazníky do různých skupin na základě různých kritérií jako například typ zařízení, ze kterého návštěvník přichází (počítač, tablet, mobilní telefon), prohlížeč, operační systém, poloha návštěvníka, odkud návštěvník přichází a zda se jedná o nového či vracejícího se zákazníka. Segmentace se využívá především k optimalizaci obsahu, který je pro dané skupiny návštěvníků nejdůležitější. (28)

Pomocí segmentace je možné z celkových průměrných hodnot najít odlišnosti mezi těmi typy zákazníků, u nichž je vyšší pravděpodobnost dokončení konverze. Na základě toho je potom možné upravit například cílení kampaní. (28)

3.4.2 Analýza nákupního košíku

Analýza nákupních košíků se používá k pochopení chování uživatelů, zejména k získání informací o tom, jaké produkty jsou nakoupeny společně nebo naopak jako samostatné položky. Analýza nákupních košíků může pomoci při rozhodování, které produkty se navzájem doplňují a bylo by vhodné je umístit vedle sebe, což může zlepšit prodejnost produktů. Tento typ prodeje se nazývá „Cross-selling“. Mezi další strategie pro zvyšování průměrné hodnoty objednávky je „Up-selling“. Tento typ nabízí zákazníkovi dražší verze produktu, které si již přidal do nákupního košíku. Využití analýzy nákupního košíku může pomoci jak při „Cross-sellingu“ tak při „Up-sellingu“, protože umožňuje identifikovat produkty, o které zákazník projevil zájem.

3.5.1 A/B testování

Pro optimalizaci konverzního poměru se téměř vždy používá A/B testování jako statistická metoda pro ověřování hypotéz.

Návštěvnost na vstupní stránce se rozdělí na minimálně dvě kontrolované skupiny, jimž se zobrazí odlišné verze stránek. Vyhodnocuje se, která z verzí měla vyšší konverzní poměr, a vítězná varianta se poté použije. (28)

A/B test pomůže také zjistit, které úpravy byly vhodné a které raději vrátit zpět. Poradí, který text více prodává. Konečným cílem každého procesu CRO (optimalizace zaměřená na zvýšení konverzního poměru) by mělo být zvýšení celkových tržeb. (5)

Testy A/B jsou nejvhodnější pro testování zásadních změn týkajících se rozvržení stránky a šablon u internetových stránek, popřípadě tam, kde chcete přidat či odstranit stránky a šablon u internetových stránek. Dále je také možné využít A/B testování k porovnání produktů, reklamních kampaním nebo obsahu na webové stránce (2)

Výhody A/B testování

A/B testování je zřejmě nejlevnějším způsobem, jak s testováním začít. Je to tím, že se využívají interní zdroje a není ani příliš složité, co se týče vynaloženého úsilí. Výsledky jsou jasně srozumitelné. (2)

Nevýhody A/B testování

Při A/B testování je snadné odhadnout, která varianta stránky je ta lepší, ale bývá obtížnější výběr toho, které prvky na internetové stránce přispívají k lepšímu výsledku. (2)

4 Vlastní práce

4.1 Představení firmy

V této části práce bych chtěl představit společnost Sportisimo, kterou jsem si vybral pro použití produktové analytiky pomocí Google Analytics. Sportisimo je původně česká síť prodejen, která byla založena v roce 2000. Sportisimo byl prvním významným sportovním řetězcem, který umožnil zákazníkům nákup z pohodlí domova a v roce 2008 spustil internetový obchod. Momentálně firma působí v dalších 8 zemích s více než 200 prodejny a prodává více než 400 značek.

Cílem firmy je poskytnout zákazníkům veškeré vybavení nejen ke sportu, ale ke všem zábavným aktivitám. Internetový obchod obsahuje stovky kategorií, které zobrazují tisíce produktů od více jak 400 značek. Zobrazení takového množství produktů a jejich variant neumožňuje dát stejný prostor k prezentaci všem výrobkům.

4.2 Struktura použitých datasetů

V následující části jsou popsány datasety, které byly získány exportem z Google Analytics, výjimkou je produktový katalog, který je z datového skladu. Pro přípravu i zpracování datasetů byly použity nástroje Keboola Connection. Pro potřeby bakalářské práce byly také některé sloupce „zaheshovány“ z důvodu bezpečnosti a anonymity, tyto sloupce byly také zkráceny ze třiceti dvou znaků na deset, a to z důvodu přehlednějšího zobrazení použitých tabulek.

4.2.1 Události produktů

Tabulka 2 Události produktů

product_sku	product_variant	date	page_type	displayed_cnt	clicked_cnt	detailed_cnt	added_to_cart_cnt	purchased_quantity
c5911397ee	91f9c9d53b	2021-05-23	category	738	29	0	0	0
02a3b55be5	d41d8cd98f	2021-07-06	category	1990	40	0	0	0
0624656ae7	d41d8cd98f	2021-07-13	category	88	1	0	0	0
161e496225	a7fdf55ae2	2021-06-17	product	0	0	99	9	0
b07db95f35	f10306a8c2	2021-06-17	product	0	0	48	8	0
2d9c209f7f	0974c8ad77	2021-09-28	product	0	0	147	8	0
d41d8cd98f	d41d8cd98f	2021-10-11	category brand	824	95	0	0	0
1943355b6d	0518d21ec0	2021-05-03	category brand	393	8	0	0	0
02a3b55be5	7652d23e26	2021-05-11	purchase	0	0	0	0	5
02a3b55be5	80601428de	2021-05-10	purchase	0	0	0	0	3

Tabulka zdroj: vlastní zpracování

Tabulka obsahuje několik sloupců, jako například „produkt_SKU“, primární klíč, který slouží k identifikaci produktu, zatímco „product_variant“ identifikuje varianty produktu. Datum, ke kterému se události na úrovni produktu staly, je označen jako „date“. Sloupec „Page_type“ zachycuje typ stránky, na které se události staly. Důležité jsou také sloupce, které zaznamenávají chování uživatelů, například „displayed_cnt“ udává počet zobrazení produktu v dané kategorii, počet prokliků je zaznamenán v „clicked_cnt“, počet zobrazení detailu v „detailed_cnt“ a počet přidanych produktů „added_to_cart_cnt“. V poslední sloupci „purchased_cnt“ je množství zakoupených produktů

Tabulka událostí reflektuje aktivitu chování zákazníka ve vztahu k produktu, některé události se mohou stát pouze na určitých typech stránek. Například „displayed_cnt“ je validní jen v kombinaci „page_type = category“, stejně tak zobrazení detailu produktu se měří pouze na stránce „page_type = detail“. Hodnoty v tabulce udávají počet událostí daného typu, což samozřejmě znamená, že jeden návštěvník může danou událost provést v jedné návštěvě opakovaně.

4.2.2 Produktový katalog

Produktový katalog slouží k poskytnutí informací o produktech, které firma prodává. Firemní produktový katalog obsahuje desítky sloupců, pro účel bakalářské práce bylo vyexportováno pouze několik sloupců. První sloupec v produktovém katalogu obsahuje primární klíč, který slouží k identifikaci produktu. Následující sloupce obsahují názvy jednotlivých kategorií, které popisují produkt. Sloupec „ProductSeasonalityName“ určuje, jestli je produkt v nabídce stále nebo se objevuje na základě sezóny. V posledním sloupci je zaznamenáváno, pro koho je produkt určený, zda je produkt určen pro muže, ženy či děti.

Tabulka 3 Produktový katalog

Product_SK	Name	Brand	CategoryName	CategoryL1Text
9c953d23c6	BLU/YES/RED ZABKY	ARESS	OBUV - CHLAPECKÉ - VOLNÝ ČAS - PANTOFLE	OBUV
b07b27053e	UNE	UMBRO	OBUV - UNISEX DĚTSKÉ - TURISTIKA - VYCHÁZKOVÉ - NÍZKÉ	OBUV
9be7919fa4	SS20S52-9990 MUMBIA	CROSSROAD	OBUV - DÁMSKÉ - VOLNÝ ČAS - SANDÁLE	OBUV
1d04f2e066	HIKELITE 32	OSPREY	HW - UNISEX DOSPĚLÉ - TURISTIKA - BATOHY - TREK DO 50L	HW
92d25cf850	COPA SENSE.4 FXG J	ADIDAS	OBUV - CHLAPECKÉ - FOTBAL - FG	OBUV
1d9c7784b9	GRN DETSKA KOLA12	ARCORE	HW - CHLAPECKÉ - CYKLISTIKA - KOLA - HORSKÁ (MTB)	HW
01147df176	AURIZ JR	PUMA	OBUV - CHLAPECKÉ - OSTATNÍ MÍČOVÉ SPORTY - VOLEJBAL - HALA - NÍZKÉ	OBUV
9a294c486c	SCO SHIRT JR TRAIL 10 S/SL	SCOTT	TEXTIL - UNISEX DĚTSKÉ - DRESY - KRÁTKÉ RUKÁVY - CYKLISTIKA	TEXTIL
dfd76ed53d	GRASSOR JR RAIL+FJ7 JR	FISCHER	HW - UNISEX DĚTSKÉ - SJEZD - LYŽE - ALLMOUNTAIN	HW
6d53c8b647	J4 RED	ALPINA	HW - UNISEX DĚTSKÉ - SJEZD - LYŽÁKY - ALLMOUNTAIN	HW

CategoryL2Text	CategoryL3Text	CategoryL4Text	CategoryL5Text	CategoryL6Text	ProductSeasonalityName	BusinessGender
CHLAPECKÉ	VOLNÝ ČAS		PANTOFLE		letní	DĚTSKÉ
UNISEX DĚTSKÉ	TURISTIKA	TURISTIKA	VYCHÁZKOVÉ	NÍZKÉ	stálý	DĚTSKÉ
DÁMSKÉ	VOLNÝ ČAS		SANDÁLE		letní	DÁMSKÉ
UNISEX DOSPĚLÉ	TURISTIKA	TURISTIKA	BATOHY	TREK DO 50L	stálý	PÁNSKÉ
CHLAPECKÉ	FOTBAL	FOTBAL	FG		stálý	DĚTSKÉ
CHLAPECKÉ	CYKLISTIKA	CYKLISTIKA	KOLA	HORSKÁ (MTB)	letní	DĚTSKÉ
CHLAPECKÉ	OSTATNÍ MÍČOVÉ SPORTY	VOLEJBAL	HALA	NÍZKÉ	stálý	DĚTSKÉ
UNISEX DĚTSKÉ	DRESY	KRÁTKÉ RUKÁVY	CYKLISTIKA	CYKLISTIKA	stálý	DĚTSKÉ
UNISEX DĚTSKÉ	SJEZD	SJEZD	LYŽE	ALLMOUNTAIN	zimní	DĚTSKÉ
UNISEX DĚTSKÉ	SJEZD	SJEZD	LYŽÁKY	ALLMOUNTAIN	zimní	DĚTSKÉ

Tabulka zdroj: vlastní zpracování

4.2.3 Transakční data

Transakční obsahují informace o jednotlivých aktivitách. Prodeje produktu jsou agregovány na úrovni jednotlivých objednávek. Sloupec „Product_SK“ slouží k identifikaci produktů, zatímco „ProductSKU_SK“ slouží k identifikaci varianty produktu. K identifikaci transakce je použit sloupec „Transaction_Number“, datum této transakce je zaznamenán ve sloupci „SalesDate“. Následný sloupec „SalesChannel“ neboli prodejní kanál transakce, určuje, zda byl produkt nakoupen na e-shopu nebo v kamenném odchodu. V poledním sloupci je uvedeno množství produktové varianty, které se v rámci transakce prodalo.

Tabulka 4 Transakční data

Product_SK	ProductSKU_SK	TransactionNumber	SalesDate	SalesChannel	SalesQty
ed1fa94a2f	a35661fee9	0d73e7d925	2021-09-04	Direct sale	1
de73ff1c92	c0eaa565f8	436ac27b7c	2021-09-04	Direct sale	1
cea08058e1	0a46550235	9e6c6c7a01	2021-09-04	Direct sale	1
0cf3631c28	ca72e01580	0b3b98e09e	2021-09-04	Direct sale	1
cbd63ac981	424cc465a3	35856269c1	2021-09-04	Direct sale	1
f06f93bde0	5377ff626b	7509e44c16	2021-09-04	Direct sale	1
4e8843cec7	429102ebd1	ccf1aa672d	2021-09-04	Direct sale	1
c9924bc446	b8ebc57d97	58d20fb768	2021-09-04	Direct sale	1
a0f99a7e5c	816024923	5134b2d3c8	2021-09-04	Direct sale	1
f8b537e3bb	8009189d97	806fb92f34	2021-09-04	Direct sale	1

Tabulka zdroj: vlastní zpracování

4.2.4 Nákupní listy

Jsou extraktem tabulky a dat získaných z webu pomocí funkce „Na seznam“ umožňující zákazníkům si ukládat produkty do vlastních seznamů.

Tabulka 5 Nákupní listy

shopping_list_id	name	date_list_created	is_default	customer_id	product_variant_id	product_code	quantity	date_product_added
3c9e91ee53	Oblíbené	2021-04-12 11:01:04	1	4602fc349f	0b0a0d9fa4	7ccc9296f5	1	2021-09-18 20:36:12
6168a10845	Oblíbené	2021-04-12 21:44:23	1	e4abdeaf8a	0be8d0c049	434936a347	1	2021-09-18 21:30:41
c391afe3b3	Oblíbené	2021-04-12 06:43:20	1	1ff537294b	0d3d118397	4559ce9868	1	2021-06-06 16:38:23
7e7048f0fc	Oblíbené	2021-04-12 22:22:52	1	a15016b8e5	0d5e558907	66b59d8afd	1	2021-04-13 09:12:53
5e8900fcc2	Oblíbené	2021-04-12 11:11:06	1	d865667177	0ea5430655	d44a1d4d3f	1	2021-07-30 23:13:48
5f49072249	Boty	2021-04-12 19:50:31	0	8596bc7765	0eb21063d0	07c5118e57	1	2021-04-20 10:20:50
8aa9b99cd7	Oblíbené	2021-04-12 14:29:01	1	ae989e91f5	0eec2fbac9	8c75ff4df6	1	2021-04-12 14:35:13
b9e86cf10f	Nový	2021-04-12 22:25:02	0	bce2bf7a9b	0fbab29b79	efbd57ba60	1	2021-04-12 22:25:02
c391afe3b3	Oblíbené	2021-04-12 06:43:20	1	1ff537294b	0fc399525f	681f50f661	1	2021-08-30 10:38:32
c8c39a2476	Fotbal	2021-04-12 20:04:41	0	e3e8f9f07f	1163eb416f	0815a4ac12	1	2021-04-12 20:07:53
333eca27b4	já	2021-04-12 07:55:56	0	bf29dd4122	11d005d357	f0b01ba5ef	1	2021-04-12 08:02:59

Tabulka zdroj: vlastní zpracování

Význam jednotlivých sloupců:

shopping_list_id – Primární klíč seznamu, který si uživatel vytvořil nebo použil defaultní list „Oblíbené“.

name – Název seznamu.

date_list_created – Datum vytvoření seznamu.

is_default – Určuje, zda je seznam výchozí nebo není. Pokud je seznam výchozí v tabulce je zaznamenána „1“, pokud byl vytvořen uživatelem bude zaznamenána „0“.

customer_id – Id zákaznického účtu.

product_variant_id – Identifikace produktové varianty.

product_code – Identifikace produktu, podle které se daný produkt identifikuje (Id).

quantity – Množství produktové varianty, které bylo přidáno do seznamu.

date_product_added – Datum, kdy byl produkt přidán na seznam.

4.2.5 Nákupy zákazníků

Následující tabulka „Nákupy zákazníků“ obsahuje primární klíč objednávky „order_id“, který identifikuje objednávku, kterou uživatel realizoval. V následujícím sloupci je primární klíč, který byl přidělen každému unikátnímu zákazníkovi „customer_id“. Datum, kdy byla objednávka zaplacená je uveden ve sloupci „date_purchased“. Identifikační číslo produktu, které bylo přidáno do objednávky je ve sloupci „product_variant_id“ a jeho množství ve sloupci „quantity“.

Tabulka 6 Nákupy zákazníků

order_id	customer_id	date_purchased	product_variant_id	rfid	quantity
261c675825	fd0b89cee0	2021-03-24 20:31:56	dd77913860	c8e7b1b374	2
805da6c351	d6116f9905	2021-05-18 11:37:42	5769d7e01e	cd441512ec	3
e80eea26d2	cefcecc440a	2021-03-04 10:06:38	acbe807316	fd47794bb5	2
d3a744a6a8	c45b2a69ba	2021-03-04 13:18:54	4f7cee686a	e55bbcb047	2
5458ab3454	81c5602954	2021-04-19 20:49:36	c2d37f1ce1	aa15ac83ad	2
5a9190ab16	9ad97b2eff	2021-07-12 21:53:03	04ef931e25	08efadf2db	5
e84a2d5b76	b6d5e6bdf6	2021-01-03 19:32:24	21431ed840	912935e3fb	2
d9db65ca9f	01e8318c28	2021-12-08 20:36:21	0e904985a4	73af884ef9	2
c0b6e353a6	7f64ce3e81	2021-03-14 16:44:24	9efd40a404	ff525e447c	2
3690691dc7	70e0884869	2021-01-08 09:39:16	792bf64300	a2d18db95f	2

Tabulka zdroj: vlastní zpracování

4.3 Agregace dat

Pro samotné analýzy byly zvoleny agregace dat popsané v následujících podkapitolách. Vždy je uvedena ukázka SQL kódu a deset řádků dat jako vzorek po agregaci. Práce s daty byla realizována v Keboola Connection na Snowflake backendu, toto prostředí bylo poskytnuto společností Sportisimo.

4.3.1 Denní agregace

Agregace, která počítá unikátní produkty a varianty a sumy jednotlivých událostí, funkce TRY_TO_NUMBER byla použita pro případ, že by sloupce obsahovaly nevalidní hodnoty. Data v tomto formátu byla použita pro seznámení se s celou problematikou a pro základní analýzy, které vycházejí z následujících agregovaných dat.

Obrázek 6 Ukázka kódu, denní agregace dat

```
CREATE OR REPLACE TABLE "daily-traffic-AGG" AS (  
SELECT "date"  
      ,COUNT(DISTINCT "product_sku") AS "uniqueProducts"  
      ,COUNT(DISTINCT "product_variant") AS "uniqueProductVariants"  
      ,SUM(TRY_TO_NUMBER("displayed_cnt")) AS "displayed_cnt"  
      ,SUM(TRY_TO_NUMBER("clicked_cnt")) AS "clicked_cnt"  
      ,SUM(TRY_TO_NUMBER("detailed_cnt")) AS "detailed_cnt"  
      ,SUM(TRY_TO_NUMBER("added_to_cart_cnt")) AS "added_to_cart_cnt"  
      ,SUM(TRY_TO_NUMBER("purchased_quantity")) AS "purchased_quantity"  
FROM "events-cz-sample"  
GROUP BY 1);
```

Obrázek zdroj: (29) (30), vlastní zpracování

Výsledek agregace poté vypadá následovně.

Tabulka 7 Výsledná data po agregaci na úrovni dne

date	uniqueProducts	uniqueProductVariants	displayed_cnt	clicked_cnt	detailed_cnt	added_to_cart_cnt	purchased_quantity
27.05.2021	38320	64813	5917763	95286	196673	15953	6020
17.07.2021	38237	47885	8138444	139916	271666	22904	8335
27.06.2021	37662	43606	8357514	140201	304760	33639	13896
17.08.2021	33593	44720	998240	104577	222164	18615	7180
09.06.2021	37842	45373	7122662	114016	234879	19666	8072
20.09.2021	27136	51813	1289890	169626	353462	24335	10019
23.10.2021	39940	50354	8461879	143504	276311	18940	7371
06.06.2021	37660	45368	7284742	116915	247447	21847	9014
05.08.2021	38660	41409	5276912	91076	189984	14234	5651
06.05.2021	38402	67382	6861032	116859	217436	19299	6929

Tabulka zdroj: vlastní zpracování

4.3.2 Agregace na úrovni produktu a měsíce

Tato agregace slouží pro vytvoření dat na základě, kterých je možné ověřit vývoj v čase. Zda se data nějak mění v závislosti na ročním období nebo na základě sezónnosti. Měsíc je doplněn o agregaci na úrovni produktu tak, aby byla zachována možnost práce na původní myšlence a zároveň bylo pro každý měsíc dostatek pozorování.

Obrázek 7 Ukázka kódu, měsíční a produktová agregace dat

```
CREATE OR REPLACE TABLE "product-monthly-traffic-AGG" AS (  
SELECT "product_sku"  
      ,LEFT("date",7) AS "YYYY-MM"  
      ,COUNT(DISTINCT "product_sku") AS "uniqueProducts"  
      ,COUNT(DISTINCT "product_variant") AS "uniqueProductVariants"  
      ,SUM(TRY_TO_NUMBER("displayed_cnt")) AS "displayed_cnt"  
      ,SUM(TRY_TO_NUMBER("clicked_cnt")) AS "clicked_cnt"  
      ,SUM(TRY_TO_NUMBER("detailed_cnt")) AS "detailed_cnt"  
      ,SUM(TRY_TO_NUMBER("added_to_cart_cnt")) AS "added_to_cart_cnt"  
      ,SUM(TRY_TO_NUMBER("purchased_quantity")) AS "purchased_quantity"  
FROM "events-cz-sample"  
GROUP BY 1,2);
```

Obrázek zdroj: vlastní zpracování

Výsledná data neobsahují tak velká čísla jako denní agregace, což je zcela pochopitelné. Dataset ukazuje počty variant jednotlivých produktů (sloupeček uniqueProductVariants).

Tabulka 8 Výsledná data po agregaci na úrovni měsíce a produktu

product_sku	YYYY-MM	uniqueProductVariants	displayed_cnt	clicked_cnt	detailed_cnt	added_to_cart_cnt	purchased_quantity
d6a079808d	2021-09	2	0	0	91	0	
5a448d1771	2021-05	5	11657	337	762	20	5
6d8baa4b65	2021-07	6	3951	68	156	15	9
5590a4c216	2021-08	2	1201	22	79	1	
b47ff33927	2021-06	3	1801	11	17	1	
ca2ad104aa	2021-07	6	6808	85	212	13	9
e955ea13cb	2021-08	6	7210	67	260	28	12
861f3d7940	2021-09	2	0	0	458	40	18
a4c3e88b95	2021-10	8	10471	41	244	7	4
6204c45e35	2021-09	5	0	0	154	15	5

Tabulka zdroj: vlastní zpracování

Data také obsahují zdánlivě nelogické hodnoty, jako například první řádek tabulky. Jak je možné, že produkt má návštěvy (detailed_cnt), když nemá žádná shlédnutí (displayed_cnt). Tento fakt je způsoben tím, že data shlédnutí (displayed_cnt) a kliknutí (clicked_cnt) jsou sbírána z výpisu produktů. Produkt může být zobrazován i v sekcích doporučené, která není klasickým výpisem produktů na stránce jako v případě kategorie. Produkt může být také navštěvován z newsletteru nebo na něj mohou vést externí kampaně, takováto návštěvnost vždy končí na detailní produktové stránce. Metrika návštěv stránky detailu produktu (detailed_cnt) je měřena bez ohledu na zdroj návštěvnosti nebo předchozí stránku.

4.3.3 Produktová agregace

Tato agregace vytváří data pro ověření závislosti na úrovni produktů, všechny metriky jsou opět stejné jako u předchozích agregací.

Obrázek 8 Ukázka kódu, produktová agregace dat

```
CREATE OR REPLACE TABLE "product-traffic-AGG" AS (  
SELECT "product_sku"  
      ,COUNT(DISTINCT "product_variant") AS "uniqueProductVariants"  
      ,SUM(TRY_TO_NUMBER("displayed_cnt")) AS "displayed_cnt"  
      ,SUM(TRY_TO_NUMBER("clicked_cnt")) AS "clicked_cnt"  
      ,SUM(TRY_TO_NUMBER("detailed_cnt")) AS "detailed_cnt"  
      ,SUM(TRY_TO_NUMBER("added_to_cart_cnt")) AS "added_to_cart_cnt"  
      ,SUM(TRY_TO_NUMBER("purchased_quantity")) AS "purchased_quantity"  
FROM "events-cz-sample"  
GROUP BY 1);
```

Obrázek zdroj: vlastní zpracování

Výstup agregace vypadá následovně, data byla použita pro hledání produktu s nevyužitým potenciálem.

Tabulka 9 Výsledná data po agregaci na úrovni produktu

product_sku	uniqueProductVariants	displayed_cnt	clicked_cnt	detailed_cnt	added_to_cart_cnt	purchased_quantity
28e04a2082	6	21953	642	1053	241	44
bba025fb8e	6	15873	312	1710	100	30
58d3b9b875	6	13170	161	662	26	14
68b8712712	7	6345	297	2433	64	33
5e2c2421d5	7	8927	137	1030	58	31
e684df7a3b	8	257245	3225	5680	716	364
2a9b95a2f3	7	9380	106	461	113	14
2b3cf29d704	6	15003	101	149	12	4
9b79977583	7	5400	86	404	25	4
657eafdf8e	9	36524	716	2296	299	82

Tabulka zdroj: vlastní zpracování

Zde je vidět rozdíl ve výkonu jednotlivých produktů. Tento rozdíl je na řádku 8 a 9, kde jsou produkty, které mají stejnou purchased_quantity, ale poměr zobrazení je cca 3:1. Tyto produkty mají přitom proklikovost 10:8, ale zobrazení detailu 15:40 a také přidání do košíku je ve prospěch druhého produktu, a to v poměru 12:25. Je nutno podotknout, že se jedná o vzorek náhodně vybraných 10 záznamu.

4.4 Závislost kroků v nákupním procesu

Následující sekce se zabývá hledáním závislosti mezi jednotlivými kroky nákupního procesu. K tomu byla použita regresní analýza v prostředí MS Excel a jeho doplňky.

4.4.1 Základní myšlenka

Pokud je produkt v řazení kategorií na nižších místech, má menší návštěvnost a v závislosti na to také prodejnost. „Co zákazník nevidí to nekoupí.“

Nulová hypotéza – počet nakoupených produktů není závislý na počtu zobrazení produktu.

Alternativní hypotéza – popírá platnost nulové hypotézy. Mezi proměnnými existuje závislost. Tudiž počet nakoupených produktů je závislý na počtu zobrazení daného produktu.

4.4.2 Chování zákazníka v datech

Tabulka 10 Chování zákazníků na webu

product_sku	displayed_cnt	clicked_cnt	detailed_cnt	added_to_cart_cnt	purchased_quantity
28e04a2082	21953	642	1053	241	44
bba025fb8e	15873	312	1710	100	30
58d3b9b875	13170	161	662	26	14
68b8712712	6345	297	2433	64	33
5e2c2421d5	8927	137	1030	58	31
e684df7a3b	257245	3225	5680	716	364
2a9b95a2f3	9380	106	461	113	14
2b3cf29d704	15003	101	149	12	4
9b79977583	5400	86	404	25	4
657eafdf8e	36524	716	2296	299	82

Tabulka zdroj: vlastní zpracování

Význam jednotlivých sloupců a jejich následné značení:

produkt_SKU – Primární klíč sloužící k identifikaci jednotlivých produktů.

displayed_cnt – První krok - (S1) nákupního procesu, udává počet zobrazení daného produktu.

clicked_cnt – Druhý krok - (S2) sleduje počet kliknutí na daný produkt na stránce.

detailed_cnt – Třetí krok - (S3) udává počet zobrazení daného produktu v detailu.

added_to_cart_cnt – Čtvrtý krok - (S4) znázorňuje počet přidávaných produktů do košíku

purchased_quantity – Pátý krok - (S5) nákupního procesu vyjadřuje počet zakoupených produktů.

4.4.3 Analýza kroků Denní S1-S2

Pro názornost následující označení „Denní S1 – S2“ udává, že při analýze byla použita data s denní agregací a závislost je počítána mezi krokem jedna (S1) – zobrazení v kategorii a krokem dva (S2) – kliknutí na produkt.

Výsledky analýzy závislosti kliků na zobrazený produkt ve výpisu v kategorii, pomocí regresní statistiky jsou zobrazeny v následující tabulce.

Tabulka 11 Regresní analýza – koeficienty závislosti mezi „S1-S2“

<i>Regresní statistika</i>	
Korelační koeficient	0,078446977
Koeficient determinace	0,006153928
Adjustovaný koeficient determinace R^2_{adj}	0,000693236
Chyba stř. hodnoty	20666,78476
Pozorování	184

Tabulka zdroj: vlastní zpracování

Korelační koeficient – Korelace je lineární závislost mezi dvěma proměnnými. Míru korelace vyjadřuje korelační koeficient, který nabývá hodnot od -1 do 1 . Hodnota 0 udává, že mezi proměnnými lineární závislost není. Zatímco kladné hodnoty udávají, že mezi proměnnými existuje závislost. S růstem jedné proměnné roste i proměnná druhá. Naopak záporné hodnoty znamenají, že mezi proměnnými existuje vztah „čím více – tím méně“, tedy s růstem jedné veličiny druhá veličina klesá. Čím blíže jsou hodnoty korelačního koeficientu 1 nebo -1 , tím je závislost mezi proměnnými vyšší a vztah silnější. Pokud se hodnoty pohybují kolem 0 znamená to, že mezi proměnnými není závislost. (31) (32)

Koeficient determinace – označuje míru kvality regresního modelu a jaký podíl variability závislé proměnné vysvětluje. Kvalitu modelu vyjadřuje v procentních bodech a říká nám kolik procent variability výstupu (Y) je způsobeno naším vstupem (X). Koeficient determinace nabývá hodnot od 0 do 100. Přičemž 100 % - znamená dokonalou predikci hodnot závisle proměnné. Naopak hodnota 0 (resp. 0 %) znamená, že model nepřináší pro poznání závisle proměnné žádnou informaci, je zcela neúčinný. Koeficient determinace nezohledňuje počet proměnných, přidáním další proměnné se vždy zvýší. Z tohoto důvodu je lepší použít Adjustovaný koeficient determinace, který tento problém eliminuje. (32)

Chyba střední hodnoty – Vyjadřuje tedy kolísání výběrových průměrů kolem teoretické (skutečné) střední hodnoty v celém základním souboru.

Pozorování – Udává počet prvků, které pozorujeme. (32)

Tabulka 12 Regresní analýza – výsledek regresní analýzy

	Koeficienty	Chyba stř. hodnoty	t Stat	Hodnota P	Dolní 95%	Horní 95%	Dolní 95,0%	Horní 95,0%
Hranice (Konstanta) b0	117325,1935	2919,845299	40,18198961	2,0222E-92	111564,0932	123086,2937	111564,0932	123086,2937
displayed_cnt (Směrnice) b1	0,000531857	0,000501006	1,061579063	0,289833469	-0,000456669	0,001520384	-0,000456669	0,001520384

Tabulka zdroj: vlastní zpracování

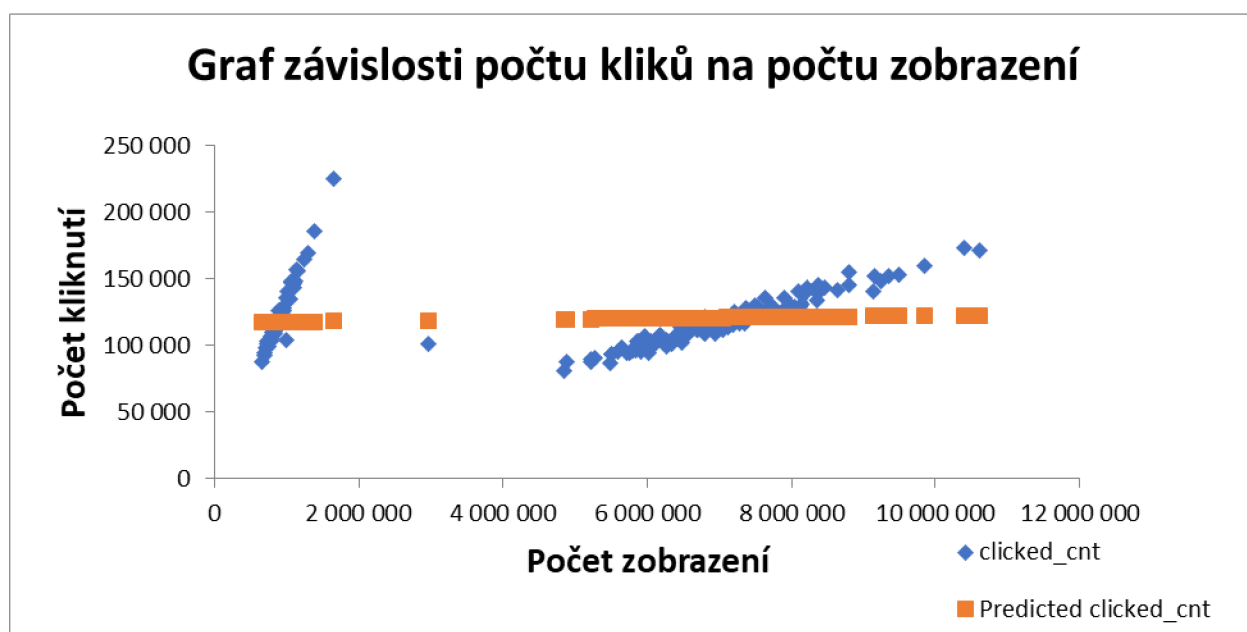
Konstanta B0 – hodnota Y pro X=0 (průsečík s osou Y)

Směrnice B1 – říká o kolik jednotek se zvýší Y, pokud se X zvýší o jednotku

Hodnota P – Tato hodnota představuje maximální možnou hladinu testu alfa, pro kterou hypotézu H0 ještě nezamítneme. Pokud tedy hladina testu alfa bude větší než uvedená hodnota P, zamítneme H0. Nebo obráceně, bude-li hodnota P menší než námi předem zvolená hladina testu, nulovou hypotézu zamítáme, v opačném případě nulovou hypotézu nelze zamítnout. (32)

Jak je vidět z tabulek výše, závislost (korelace) stejně jako determinace je velmi slabá. Což zcela odporuje základní hypotéze celé práce. Po pohledu na následující graf bylo zjištěno, že vizuální rozložení dat je nějaké podezřelé. Graf ukazuje dvě množiny bodů s odlišnou směrovou orientací a jeden bod mezi nimi uprostřed. Což je pro lineární regresi samozřejmě velmi špatně řešitelný problém.

Graf 1 Závislost počtu prokliků na počtu zobrazení produktu



Graf zdroj: vlastní zpracování

Levá množina „Red“ vykazuje daleko nižší displayed cnt, tím pádem daleko vyšší konverzní poměr. Naopak pravá množina prvků „White“ vykazuje vyšší displayed cnt a nižší konverzní poměr. Bod uprostřed znázorňuje den, kdy byla použita změna kódu a měření bylo vráceno do normálu.

Bohužel změna je skoková od 2021-08-17 a končí 2021-10-18, výrazná změna je pouze v hodnotách `displayed_cnt`, ze kterých se zmíněné dvě metriky počítají. Na základě tohoto zjištění byl dataset rozdělen na dva a data z 2021-10-18 vyloučena úplně. Tyto změny v hodnotách jsou způsobené chybou měření, jejichž příčinou mohlo být nasazení nové funkcionality webu nebo změnou měření v GTM. Tuto domněnku bylo možné konzultovat a potvrdit se společností Sportisimo.

Rozdělení datasetu je ilustrováno v posledním sloupečku „cluster“, pro potřeby práce budou datasety označovány jako „red“ (snížené hodnoty `displayed_cnt`) a „white“. **V práci bude dále použit pouze dataset „White“.**

Tabulka 13 Rozdělení dat

date	displayed_cnt	clicked_cnt	purchased_quantity	S1-S2 CR%	S1-S5 CR%	cluster
2021-05-01	7 630 905	135 950	7 216	1,78%	0,09%	white
...
2021-08-16	7 002 100	115 975	8 427	1,66%	0,12%	white
2021-08-17	998 240	104 577	7 180	10,48%	0,72%	red
2021-08-18	784 292	105 559	6 974	13,46%	0,89%	red
...
2021-10-16	894 820	126 136	7 134	14,10%	0,80%	red
2021-10-17	738 448	101 251	6 510	13,71%	0,88%	red
2021-10-18	2 961 769	101 043	6 734	3,41%	0,23%	x
2021-10-19	5 960 902	100 530	6 505	1,69%	0,11%	white
...
2021-10-31	6 547 247	108 159	7 141	1,65%	0,11%	white

Tabulka zdroj: vlastní zpracování

4.5 Porovnání datasetů

Originální data získaná o chování zákazníků od společnosti Sportisimo po denní agregaci jsou dále označovány jako „all“ data jsou za období od 2021-05-01 do 2021-10-31 a obsahují tedy 184 řádků (pozorování). Datasety „red“ a „white“ vznikly rozdělením původního datasetu „all“. Red obsahuje data od 2021-08-17 do 2021-10-17 (včetně) tedy 62 řádků (pozorování) a white od 2021-05-01 do 2021-08-16 a 2021-08-19 do 2021-10-31 (včetně) 121 řádků (pozorování).

4.5.1 Závislost kroků na denní agregaci – dataset „All“

Tabulka 14 Závislost kroků v datasetu "All" na denní agregaci

	S1-S2	S1-S3	S2-S3	S3-S4	S4-S5	S1-S5
Korelační koeficient	0,078447	0,029495	0,966728	0,791207	0,968964	0,346587
Koeficient determinace	0,006154	0,00087	0,934563	0,626009	0,938891	0,120123
Koeficient determinace (%)	1%	0%	93%	63%	94%	12%
Adjustovaný koeficient determinace R^2_{adj}	0,000693	-0,00462	0,934203	0,623954	0,938555	0,115288
Chyba stř. hodnoty	20666,78	39272,97	10050,67	2892,502	462,8739	1756,392
Pozorování	184	184	184	184	184	184
t Stat	1,061579	-0,39809	50,98326	17,45402	52,87996	4,984676
Hodnota P	0,289833	0,691031	1,1E-109	1E-40	2,1E-112	1,44E-06

Tabulka zdroj: vlastní zpracování

V první tabulce u sloupců S1-S2 a S1-S3 je vidět dopad nesouladu v datech na koeficienty korelace a determinace v původním datasetu „all“. Z toho byla následně zjištěna minimální závislost mezi kroky S1-S2 a S1-S3. Na základě hodnoty P, která byla vyšší než kritická hodnota pro hladinu alfa = 0,05 byla nezamítnutá nulová hypotéza a výsledek je považován za statisticky nevýznamný.

4.5.2 Závislost kroků na denní agregaci – rozdělené datasey

Tabulka 15 Závislost kroků v datasetu "White" na denní agregaci

	S1-S2	S1-S3	S2-S3	S3-S4	S4-S5	S1-S5
Korelační koeficient	0,977661	0,960629	0,956077	0,932153	0,967229	0,882914
Koeficient determinace	0,955821	0,922809	0,914083	0,868909	0,935531	0,779538
Koeficient determinace (%)	96%	92%	91%	87%	94%	78%
Adjustovaný koeficient determinace R ² adj	0,95545	0,92216	0,913361	0,867807	0,934989	0,777685
Chyba stř. hodnoty	3840,444	9232,826	9740,672	1845,179	511,3246	945,5588
Pozorování	121	121	121	121	121	121
t Stat	50,74047	37,71765	35,58176	28,08499	41,55539	20,5128
Hodnota P	1,83E-82	4,91E-68	2,88E-65	2,45E-54	1,08E-72	6,99E-41

Tabulka zdroj: vlastní zpracování

Na výsledků z rozdělených subsetů, byla prokázána statistická významnost na hladině alfa = 0,05, byla zamítnuta nulová hypotéza a přijata alternativní hypotéza. Tudiž byla prokázána silná závislost v chování návštěvníků mezi jednotlivými kroky nákupního procesu.

Tabulka 16 Závislost kroků v datasetu "Red" na denní agregaci

	S1-S2	S1-S3	S2-S3	S3-S4	S4-S5	S1-S5
Korelační koeficient	0,983309	0,974729	0,986038	0,861231	0,977691	0,846894
Koeficient determinace	0,966897	0,950097	0,972271	0,74172	0,95588	0,717229
Koeficient determinace (%)	96,69%	95,01%	97,23%	74,17%	95,59%	71,72%
Adjustovaný koeficient determinace R ² adj	0,966345	0,949265	0,971809	0,737415	0,955145	0,712517
Chyba stř. hodnoty	4454,191	10339,33	7707,103	1944,043	329,2762	833,6072
Pozorování	62	62	62	62	62	62
t Stat	41,86296	33,79832	45,86761	13,12652	36,05466	12,33638
Hodnota P	4,11E-46	9,24E-41	2,02E-48	2,73E-19	2,29E-42	4,2E-18

Tabulka zdroj: vlastní zpracování

Tento celý problém by bylo možné dále řešit pomocí složitější funkce než lineární regrese, například logaritmickou nebo exponenciální funkcí. Tyto funkce však nejsou vhodné pro určení závislosti z tohoto důvodu byla použita lineární regrese.

V následující tabulce jde vidět, že rozdělení původního datasetu na nové dva datasety toto rozdělení vedlo ke zlepšení hodnot koeficientů ve všech porovnání, pouze pro S2-S3 je „White dataset“ lehce horší než původní dataset. Toto zhoršení je však přijatelné vzhledem k ostatním výsledkům.

Tabulka 17 Porovnání koeficientů z datasetů

	Korelační koeficient			Koeficient determinace		
	All	Red	White	All	Red	White
S1-S2	0,078447	0,983309	0,977661	0,006154	0,966897	0,955821
S1-S3	0,029495	0,974729	0,960629	0,00087	0,950097	0,922809
S2-S3	0,966728	0,986038	0,956077	0,934563	0,972271	0,914083
S3-S4	0,791207	0,861231	0,932153	0,626009	0,74172	0,868909
S4-S5	0,968964	0,977691	0,967229	0,938891	0,95588	0,935531
S1-S5	0,346587	0,846894	0,882914	0,120123	0,717229	0,779538

Tabulka zdroj: vlastní zpracování

4.6 Meziměsíční vývoj

V rámci každého roku je několik významných nákupních období, která se opakují a mohou ovlivňovat chování zákazníku (novoroční výprodeje, prázdniny, začátek školního roku, černý pátek). Pro ověření velikosti změny na našich datech byl vybrán začátek prázdnin.

Pro toto porovnání byly použity první tři měsíce z datasetu „all“. Ve vybraných měsících je nabízený sortiment konstantní a chování zákazníku by mělo být ovlivněno pouze jejich potřebou, která se s příchodem prázdnin může promítnout v čase.

Tabulka 18 Měsíční vývoj závislosti jednotlivých kroků nákupního procesu

	5. měsíc			6. měsíc			7. měsíc		
	S1-S2	S2-S3	S1-S5	S1-S2	S2-S3	S1-S5	S1-S2	S2-S3	S1-S5
Korelační koeficient	0,98097	0,95663	0,94072	0,96982	0,97263	0,88938	0,96905	0,96463	0,8401
Koeficient detrmínace %	96%	92%	88%	94%	95%	79%	94%	93%	71%
Chyba stř. hodnoty	4028,89	11237,8	777,069	4314,79	7999,05	932,669	3359,99	6977,39	657,12
Pozorování	31	31	31	30	30	30	31	31	31
Hodnota P	3,4E-22	4,5E-17	3,8E-15	1,1E-18	2,7E-19	5,1E-11	3,6E-19	2,5E-18	3,4E-09

Tabulka zdroj: vlastní zpracování

V tabulce jsou vidět změny v krocích S1-S2 a S2-S3 v jednotkách procent, výrazná změna koeficientu determinace říká, že v sedmém měsíci je nakupování zákazníků méně závislé na zobrazení produktů.

4.7 Srovnání na úrovni produktů

V této sekci je popsána analýza provedená na originálních datech s agregací na produktu. Získaná čísla slouží jako výchozí hodnoty pro další analýzy a další kroky popsané v následujících podsekcích, do vyhodnocení byly zahrnuty všechny produkty, které se v datech vyskytovaly.

Tabulka 19 Závislost jednotlivých kroků nákupního produktu na úrovni produktů

	S1-S2	S1-S3	S2-S3	S3-S4	S4-S5	S1-S5
Korelační koeficient	0,809797	0,630378	0,834957	0,860207	0,964148	0,584428
Koeficient determinace	0,655772	0,397377	0,697154	0,739956	0,929582	0,341556
Koeficient determinace (%)	66%	40%	70%	74%	93%	34%
Adjustovaný koeficient determinace R ² adj	0,655766	0,397367	0,697149	0,739951	0,92958	0,341545
Chyba stř. hodnoty	302,578	1357,643	962,44	94,54615	21,32458	65,20738
Pozorování	59902	59902	59902	59902	59902	59902
t Stat	337,8054	198,7429	371,3357	412,8502	889,2297	176,2725
Hodnota P	0	0	0	0	0	0

Tabulka zdroj: vlastní zpracování

Jak lze vidět z tabulky výše, závislost mezi zobrazením a nakoupením produktu je slabá. Jelikož produkt může mít například 1 000 zobrazení a pouze desetkrát byl nakoupen. Druhý produkt může mít také 1 000 zobrazení, ale byl nakoupen stokrát. Na základě tohoto zjištění byla provedena analýza, které hledá produkty s nevyužitým potenciálem.

4.7.1 Produkty s nevyužitým potenciálem

Jako produkt s „nevyužitým potenciálem“ byl brán takový produkt, který má velké množství zobrazení, ale podprůměrné obchodní výsledky. Respektive je hodně zobrazovaný, ale málo prodáváný.

Pro hledání produktů s nevyužitým potenciálem byla vybrána následující data: Top 100 produktů – nejvíce zobrazované produkty, které byly seřazeny podle celkového množství zobrazení. Bot 100 produktů – nejméně zobrazované produkty, které byly také seřazeny podle celkového množství zobrazení, kde počet zobrazení byl větší než 40 000. Přibližně 200 zobrazení denně. Tento počet zobrazení byl vybrán z důvodu statistické významnosti.

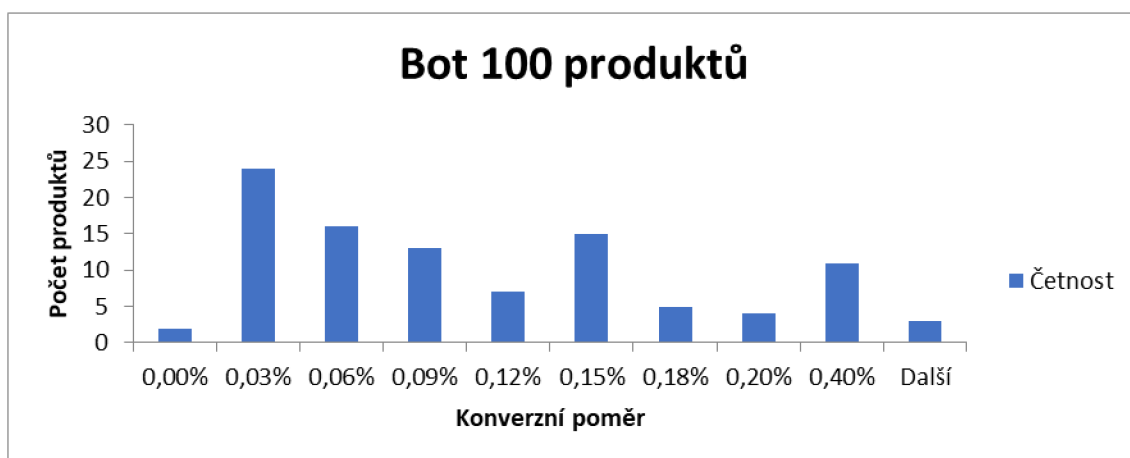
Graf 2 Konverzní poměr S1-S5 Top 100 produktů



Graf zdroj: vlastní zpracování

V následujícím histogramu je zobrazen konverzní poměr Top 100 produktů mezi kroky zobrazení produktu a nakoupením produktu.

Graf 3 Konverzní poměr S1-S5 Bot 100 produktů



Graf zdroj: vlastní zpracování

Téměř polovina produktů z datasetu „Bot 100“ vykazuje konverzní poměr od 0,15 % a výš, kdežto polovina datasetu „Top“ 100“ vykazuje konverzní poměr od 0 do 0,15 %. Z tohoto důvodu by bylo vhodné nahradit právě produkty z datasetu „Top 100“, které nevykazují vysoký konverzní poměr právě daty z „Bot“, které mají konverzní poměr vyšší. Tuto změny by bylo potřeba následně nasadit do zvýhodněných pozic ve výpisech a po čase ověřit, zda si produkty zachovají podobnou výkonnost i při vyšším počtu zobrazení.

4.7.2 Hledání alternativních produktů

Pro hledání alternativních produktů bylo použito jednoduchého kódu pracujícího s produktovým katalogem. Každý produkt v katalogu je porovnán s ostatními přes všechny dostupné parametry a jejich hodnoty. Před samotným porovnáním bylo třeba originální podobu katalogu transformovat do následujícího formátu a odstranit záznamy s hodnotami, které nenesou validní informace jako například („‘,‘-‘,‘unknown‘). (33)

Obrázek 9 Transformovaný produktový katalog

"Product_SK"	"parameter"	"value"
d068c993124185d91f92a2ebf384d1b6	Brand	LEVELWEAR
d068c993124185d91f92a2ebf384d1b6	CategoryL4Text	
d068c993124185d91f92a2ebf384d1b6	CategoryL3Text	FANS
d068c993124185d91f92a2ebf384d1b6	CategoryL2Text	UNISEX DOSPĚLÉ
d068c993124185d91f92a2ebf384d1b6	CategoryL1Text	HW
d068c993124185d91f92a2ebf384d1b6	CategoryName	HW - UNISEX DOSPĚLÉ - FANS - TRIKA - KRÁTKÉ RUKÁVY
d068c993124185d91f92a2ebf384d1b6	CategoryL5Text	TRIKA
d068c993124185d91f92a2ebf384d1b6	CategoryL6Text	KRÁTKÉ RUKÁVY
d068c993124185d91f92a2ebf384d1b6	ProductSeasonalityName	stálý
d068c993124185d91f92a2ebf384d1b6	BusinessGender	PÁNSKÉ
d06900942502063d0cf0cc6452058609	CategoryL3Text	ŠATY
d06900942502063d0cf0cc6452058609	CategoryL2Text	DÁMSKÉ

Obrázek zdroj: vlastní zpracování

Produkt, který nemá parametry vyplněny, nebo je velmi specifický, bude mít velmi málo alternativ s malým počtem shodných parametrů/atributů. Následovně pak dochází k seřazení produktů podle počtu shodných parametrů/atributů. Vzhledem k tomu, že není jak podobnost dále hodnotit, slouží sloupeček „sort_order“ spíše k orientaci, kolik existuje alternativních produktů se stejným počtem shodných parametrů/atributů.

Obrázek 10 Ukázka kódu na hledání alternativních produktů

```
CREATE OR REPLACE VIEW "shared_attributes" AS
SELECT a."Product_SK",
       b."Product_SK" AS "product_match",
       count(*) AS "shared_attributes"
FROM "sample" a
     INNER JOIN "sample" b ON b."parameter" = a."parameter" AND b."value" = a."value" AND b."Product_SK" <> a."Product_SK"
GROUP BY a."Product_SK", b."Product_SK";

CREATE OR REPLACE TABLE "product-alternatives" AS (
SELECT *, ROW_NUMBER()
     OVER (PARTITION BY "Product_SK" ORDER BY "shared_attributes" DESC) AS "sort_order"
FROM "shared_attributes"
ORDER BY 1,3 DESC
);
```

Obrázek zdroj: vlastní zpracování

Cena produktu by se také dala zahrnout do hledání, například jako vážený parametr od -1 do 1, poskytnutý katalog cenu neobsahuje z několika důvodů. Z pohledu návrhu databáze se nejedná o parametr, cena se bude měnit v čase a také dle domény, na které se zákazník nachází. Dále může být upravena aktuálně probíhajícími akcemi. Výpočet je realizován na úrovni **produktu** tak, aby se eliminovalo porovnání produktových variant, které se liší zejména velikostí a barvou.

Obrázek 11 Nalezené alternativní produkty a jejich sdílené atributy

Product_SK	product_match	shared_attributes	sort_order
594657e1f77d5bd81e49ac89af99	729cea7aed19e81783e80b698c9143c7	10	16
64b2594657e1f77d5bd81e49ac89af99	acd75eec232c9556b9bf7962dfce5d3c	10	15
64b2594657e1f77d5bd81e49ac89af99	d03694e2ec3a265c8ae52943f08677f5	10	14
64b2594657e1f77d5bd81e49ac89af99	236e834c6966eb5f868af020c6acd908	10	13
64b2594657e1f77d5bd81e49ac89af99	ebb2e64e4f17d75de741bbd605c29d0d	10	12
64b2594657e1f77d5bd81e49ac89af99	400d83e467e1eb371a2bb728ebbc81a6	10	11
64b2594657e1f77d5bd81e49ac89af99	d90774ef77b1c49220b5077542fae963	10	10
64b2594657e1f77d5bd81e49ac89af99	b0792c850f464000585babb812c769b4	10	9
64b2594657e1f77d5bd81e49ac89af99	7055c7c83936ec744c3e1ecce57f7a8a	10	8
64b2594657e1f77d5bd81e49ac89af99	25ce98d9f20cc925365158b1b60ce228	10	7
64b2594657e1f77d5bd81e49ac89af99	70aae010ceda94d8f82c8d00a3bb9ed9	10	6
64b2594657e1f77d5bd81e49ac89af99	86dd89b07b1a29ea901be9851d4cbacd	10	5
64b2594657e1f77d5bd81e49ac89af99	f8d02733451cd716311625bc1319753c	10	4
64b2594657e1f77d5bd81e49ac89af99	2f0401943a9866b57d5b4f65bf4714b8	10	3

Obrázek zdroj: vlastní zpracování

Analýza byla prováděna nad produktovým katalogem čítajícím více jak 18 tisíc produktů, pro představu je zde uveden počet alternativ jednoho produktu.

Obrázek 12 Počet nalezených variant k danému produktu

Product_SK	shared_attributes	product_alternatives
64b2594657e1f77d5bd81e49ac89af99	10	26
64b2594657e1f77d5bd81e49ac89af99	9	29
64b2594657e1f77d5bd81e49ac89af99	8	65
64b2594657e1f77d5bd81e49ac89af99	7	106
64b2594657e1f77d5bd81e49ac89af99	6	316
64b2594657e1f77d5bd81e49ac89af99	5	1100
64b2594657e1f77d5bd81e49ac89af99	4	2305
64b2594657e1f77d5bd81e49ac89af99	3	5018
64b2594657e1f77d5bd81e49ac89af99	2	7945
64b2594657e1f77d5bd81e49ac89af99	1	13371

Obrázek zdroj: vlastní zpracování

Takto napočítanou tabulku lze použít k doporučování produktů. Tabulku je třeba přepočítávat pouze při rozšiřování katalogu o nové parametry nebo při přidání nových produktů na detail produktů, při odstranění produktu z košíku, vyprodání produktu nebo dané velikosti.

4.8 Analýza používání funkce „Na seznam“

Tato část je zaměřena na používání funkcionality „Na seznam“ na webu firmy. Funkcionalitu je možné najít ve výpisu produktů vpravo nahoře u každého produktu, po najetí myši na ikonu srdíčka.

Obrázek 13 Náhled produktu a přidání tlačítka "Na seznam"



Obrázek zdroj: Sportisimo.cz

Nebo také jako tlačítko v pravé horní části detailu produktu mezi hodnocením a přidáním produktů do porovnání s jinými produkty. Funkce „Na seznam“ umožňuje přidávat produkty do seznamu „Oblíbené“, případně si vytvořit seznamy vlastní a následně je editovat. Pojetí této části je více interpretační a obchodně popsáno v následujících podsekcích.

4.8.1 Přidané produkty na seznam

V datech bylo nalezeno 49 841 unikátních zákazníků, kteří v roce 2021 funkci „Na seznam“ využili, z toho bylo zjištěno, že 39 144 uživatelů v roce 2021 nenakoupilo ani jeden produkt ze svého seznamu. Naopak 10 697 uživatelů nakoupilo alespoň jeden produkt, který si přidali na svůj seznam, z toho průměrně nakoupí 28 % produktů.

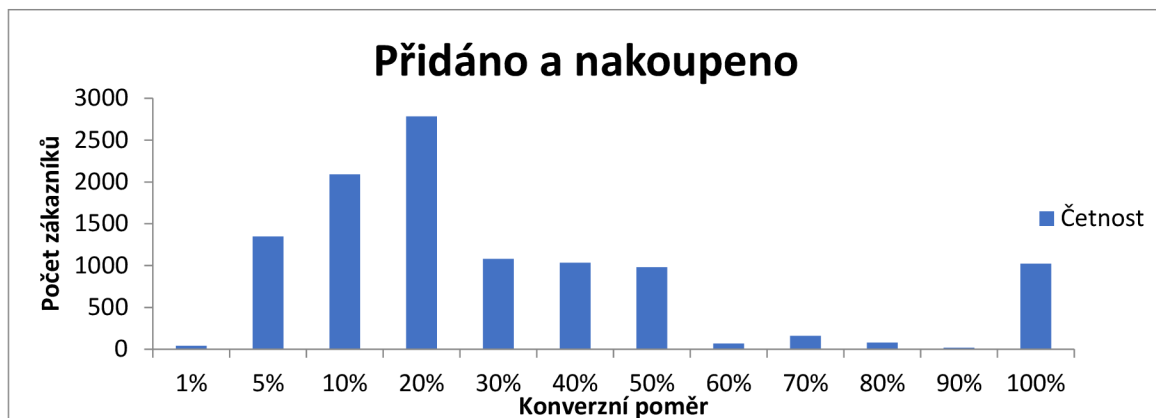
Tabulka 20 Kolik procent z toho, co zákazník přidá a následně nakoupí

Id uživatele	Přidané položky do seznamu	Nakoupené položky ze seznamu	Konverzní poměr
000a6ac17f03f393d94	32	1	3%
0012a83c1e6cda9d73e	2	1	50%
00225e015223f4a512c	3	3	100%
0022e8439e01662208;	25	1	4%
004709e32b0176d63e	9	2	22%
004e2536ac571d1143f	24	2	8%
0052511648d677908e!	1	1	100%
0055a2521ee9dd528a	83	3	4%
0056e7e1318951a02fk	5	1	20%
0057eba7235ad14f93	24	1	4%

Tabulka zdroj: vlastní zpracování

Průměrný počet položek v seznamu, které si uživatel přidá „Na seznam“ je devatenáct produktů. Pokud se na data podíváme o jednu úroveň hlouběji, tedy na množství produktů, které uživatel přidal a následně nakoupil, vidíme, že se chování zákazníků liší v poměru přidaných a nakoupených produktů. Množina 39 144 zákazníků, kteří nenakoupili žádný produkt ze seznamu, nejsou v grafu uvedeny z důvodu správného zobrazení grafu.

Graf 4 Konverzní poměr toho, co zákazník přidá na seznam a následně nakoupí

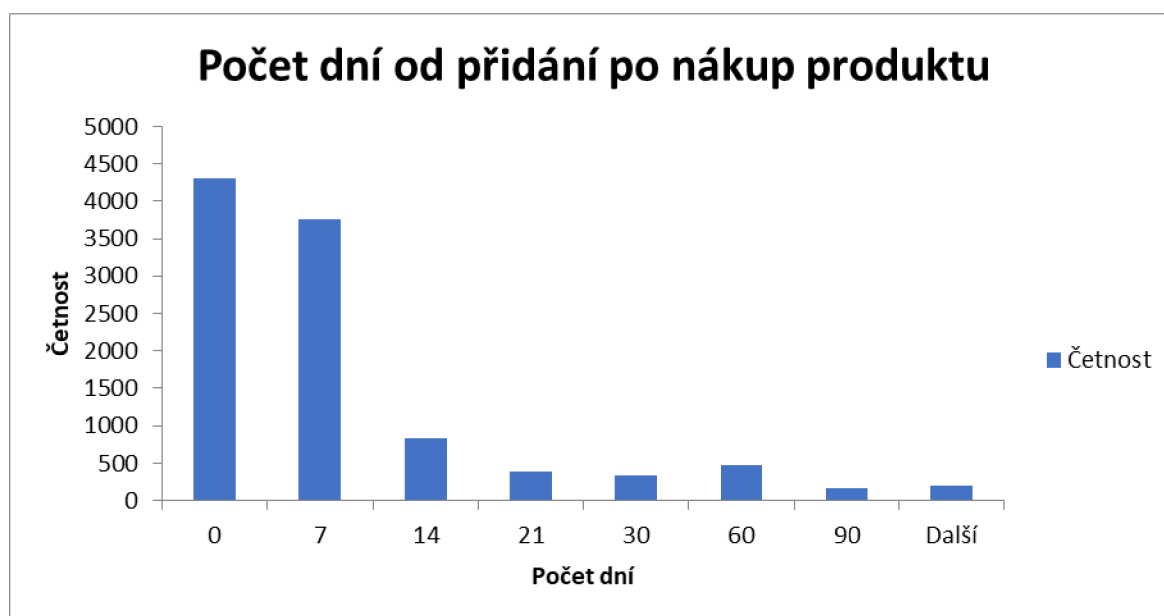


Graf zdroj: vlastní zpracování

4.8.2 Počet dní od přidání po nákup

Z celkového počtu 10 697 zákazníků si produkt přidalo a následně nakoupilo později 10 219 z nich. Z toho 4 242 zákazníků si nakoupilo produkt v ten samý den, ve který byl přidán na jejich seznam. Pouze 478 zákazníků si produkt nakoupilo a poté přidalo na svůj seznam. Následně byla zjištěna doba, která uplynula od přidání po nákup produktu. Výsledná doba byla průměrně 9,13 dní od přidání produktu po nákup daného produktu. Rozdíly v chování jednotlivých zákazníků jsou vidět v následujícím histogramu.

Graf 5 Počet dní od přidání produktu na seznam po jeho koupi



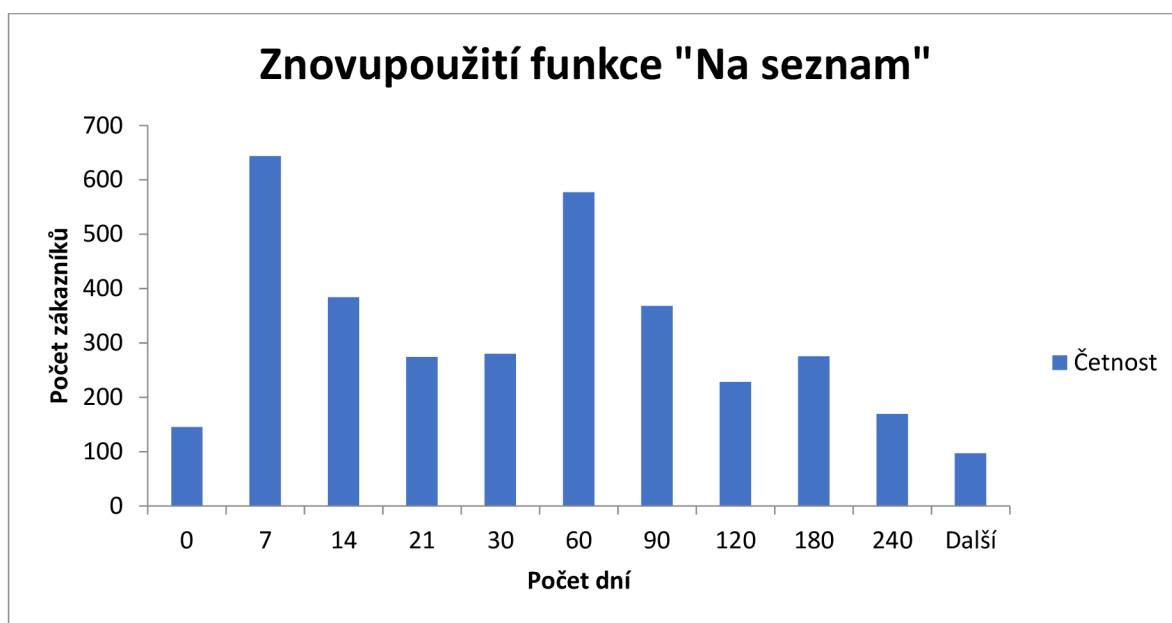
Graf zdroj: vlastní zpracování

4.8.3 Průměrný počet dní od znovu použití

V předcházející analýze bylo zjištěno, že téměř polovina zákazníků, kteří funkci použili, nakoupili ve stejný den jako byl produkt přidán na seznam. Na základě tohoto zjištění byla provedena doplňující analýza, a to zjištění po jaké době zákazníci opět použijí funkci „Na seznam“

Zde je důležité připomenout, že data byla dostupná za 12 měsíců a do analýzy bylo zahrnuto 42 295 zákazníků, kteří přidali produkt na seznam, který následně zakoupili. Z nich 3 441 funkci použilo znovu alespoň jednou. Průměrný počet dní od znovu použití funkce se pohybuje průměrně okolo dvou měsíců – 57,7 dní, což znázorňuje následující Histogram.

Graf 6 Počet dní od znovupoužití funkce "Na seznam"



Graf zdroj: vlastní zpracování

5 Výsledky a diskuse

Cílem práce bylo využití dat z webové analytiky pro lepší fungování e-shopu. Všechna zjištění byla učiněna vlastní prací na popsaných datasetech. Analýzy byly zaměřeny na chování zákazníků a produkty. Základní hypotéza, která říkala, zda existuje závislost mezi množstvím zakoupených produktů a počtem zobrazení produktu ve výpisu na webu, byla v rámci práce několikrát potvrzena, a to při různých pohledech na data. Toto vyhodnocení závislosti bylo součástí všech analýz v první části praktické části práce.

Výsledky koeficientů korelace a determinace mezi S1-S2 a S2-S3 vedli k rozhodnutí o rozdělení původního datasetu. Analýza vývoje v čase sloužila zejména pro ověření reality a realizaci, jak moc se mohou hodnoty měnit v čase. Jak již bylo zmíněno vzhledem k tomu, že S3 (množství zobrazení detailu produktu) může být ovlivněno dalšími vlivy, proto byl kladen důraz především na S1-S2 a S1-S5.

Srovnání na produktové úrovni také odhalilo několik málo desítek produktů, které měly nižší desítky zobrazení detailu a ostatní metriky nulové. To je fakticky způsobeno tím, že produkty jsou i po jejich vyprodání stále dostupné pomocí odkazu, na webu se však nezobrazují. Pokud si tedy otevřete třeba rok starý newsletter a prokliknete se z něj na web vytvoříte právě takový záznam. Tyto záznamy byly z datasetu odstraněny.

Pro analýzu produktů s nevyužitým potenciálem bylo vybráno 100 produktů s největším počtem zobrazení a 100 produktů se zobrazením 40 000 a nižší. Z datasetu TOP100 bylo vybráno 20 produktů s nejnižší výkonností, které by se daly nahradit 20 produkty s nejvyšším výkonem ze skupiny BOT100. Rozdíl výkonnosti byl v průměru dvojnásobný a větší, tyto produkty by bylo dobré nasadit do zvýhodněných pozic ve výpisech a po čase ověřit, zda si produkty zachovávají podobnou výkonnost i při vyšším počtu zobrazení. Samozřejmě stejným způsobem jde postupovat i u produktů, které se nacházejí mezi zmíněnými skupinami produktů. Do případného algoritmu na výměnu produktů by se dala přidat i podobnost produktů napočítaná v sekci „Hledání alternativních produktů“

Pro další analýzu byli vybráni návštěvníci používající funkci „Na seznam“ a to z důvodů domněnky, že tito uživatelé mohou být stejně jako uživatelé, kteří použili vyhledávání na webu skupinou návštěvníků s vyšším konverzním potenciálem. V rámci analýzy byl kladen důraz především na tři aspekty chování, a to množství produktů, čas mezi interakcemi a retenci používání funkce. Tuto funkci v roce 2021 využilo něco málo přes 40 000 návštěvníků webu, transakci provedla zhruba čtvrtina z nich. Průměrný uživatel funkce nakoupí 28 % zboží. Tato informace může být lehce zavádějící, protože někteří uživatelé přidali na seznam pouze jeden produkt, který posléze nakoupili, avšak existují také uživatelé s desítkami produktů na seznamech, proto byl do vyhodnocení přidán také histogram.

Dále bylo zjištěno, že více jak 40 % uživatelů nakoupí produkt ve stejný den jako si ho přidají do košíku, což není zanedbatelný počet uživatelů, proto by bylo zajímavé se na tyto uživatele podívat detailněji, například na úrovni dní a hodin, zda se nenajde nějaký vzorec chování jako například „v práci najdu – večer proberu s manželkou“. Průměr ostatních uživatelů je pak 9 dní mezi přidáním na seznam a následným nákupem. Velmi zajímavým poznatkem v této analýze bylo že přibližně 4 % uživatelů funkci použili až po nákupu produktu čili právě naopak, než bylo zamýšleno. Při pohledu na retenci uživatelů v histogramu jsou výrazné dva intervaly a to 7 a 60 dní od nákupu k přidání dalšího produktu na seznam.

6 Závěr

V rámci práce se prokázalo, že čistota a porozumění datům jak po technické, tak i obchodní stránce jsou nedílnou součástí a předpokladem každé analýzy. Práce se zaměřuje na využití již dostupných zdrojů, které nejsou aktivně využívány. V práci jsem se zaměřil především na využití sbíraných dat o produktech. Tato data byla analyzována za účelem získání podkladů pro návrh relevantnějšího řazení produktů v produktových výpisech na webu. Pro realizaci je pak možné použít konverzního poměru mezi zobrazeními a prokliky produktu. Tato hodnota byla označena jako koeficient atraktivity, jeho průměrná hodnota je rovna 2,95 % (přes produkty s relevantním množstvím prokliku). Alternativou je také konverzní poměr mezi zobrazeními a množstvím zakoupených produktů. Zde je průměrná hodnota 0,24 % a koeficient korelace a determinace je zde nižší. Dalším zdrojem může být také analýza alternativních produktů. Případně je možné přístupy kombinovat nebo naopak testovat vůči sobě, součástí nasazení takové funkcionality by mělo být A/B testování.

Jako produkty s nevyužitým potenciálem bereme takové produkty, které mají i přes velké množství zobrazení nízké (podprůměrné) konverzní koeficienty S1-S2 a S1-S5. V práci byla pro názornost uvedena analýza na vybraných skupinách produktů. TOP100 reprezentuje nejzobrazovanější produkty naopak BOT100 obsahuje nejméně zobrazované produkty, ale jsou to produkty, které jsou statisticky významné. Záměna mezi těmito skupinami byla navržena realizovat u 20 produktů. Bylo vybráno dvacet produktů s nejvyšším výkonem ze skupiny BOT100, které budeme preferovat ve výpisech tak, aby se dostaly do TOP100, následně je pak nutné porovnat změnu výkonnosti. Tuto záměnu není možno realizovat 1:1 a to především z důvodu, že produkty patří do jiných kategorií, které se na základě poptávky mohou výrazně lišit návštěvností.

Analýza funkce „Na seznam“ je zaměřena na uživatelské chování, které je zachyceno v datech, které generuje používání funkce na webu. U těchto uživatelů stejně jako u uživatelů, kteří použili vyhledávání na webu, byl očekáván vyšší konverzní potenciál. Konverzní poměr této funkce se pohybuje kolem 25 %. Průměrný uživatel funkce nakoupí 28 % zboží, což je zkruseno uživateli, kteří si na seznam přidali jeden produkt, který posléze nakoupili. Více jak 40 % uživatelů nakoupí uložený produkt ve stejný den. Z pohledu retence uživatelů vyčnívají dva intervaly, a to 7 a 60 dní od nákupu do přidání dalšího produktu na seznam. Následné kroky zde mohou být detailní analýza uživatelů, kteří přidany produkt nakoupí v ten samý den, zejména s ohledem na čas. Potenciál zde vidím v tom, že funkci využívá jen minimum zákazníků, a tak ji může provozovatel představit i těm co ji aktuálně nevyužívají. Na produkty v seznamech se dá navázat komunikace ve stylu „opuštěný košík“ a také ideální akční ceny, které budou personalizovány podle toho, co má uživatel na seznamech.

Práce se celkově snaží poukázat na potenciál využití aktuálně dostupných dat pro vhodnější produkty, ze kterých mohou mít prospěch jak zákazníci, tak provozovatel. Možnosti rozšíření práce jsou zřejmé, a to realizace navrhovaných experimentů dle možností, kapacit a priorit vlastníka dat čili firmy Sportisimo.

7 Seznam použitých zdrojů

1. **Hamel, Stéphane.** Definition of Web Analytics. *Blog immeria*. [Online] 8 2009. [Citace: 14. 10 2022.] <http://blog.immeria.net/2009/08/definition-of-web-analytics.html>.
2. **Kaushik, Avinash.** *Webová Analytika 2.0*. Holandská 3, 639 00 Brno : Computer Press, a.s., 2011. ISBN 978-80-251-2964-7.
3. **Pavel, Jašek.** Webová analytika v kostce. *Slideshare*. [Online] 2011. [Citace: 25. 9 2022.] <https://www.slideshare.net/dobryweb/webov-analytika-v-kostce>.
4. **Gillis, Alexander S.** clickstream data. *TechTarget*. [Online] 5. 5 2022. [Citace: 15. 10 2022.] <https://www.techtarget.com/searchcustomerexperience/definition/clickstream-analysis-clickstream-analytics?amp=1&fbclid=IwAR0PsJIF1A3GOzF9XMein8u7MJWpHEHqnTP3f969EmJSt9e0JvyCx3s3uWs>.
5. **Brunec, Jan.** *Google Analytics*. U Průchodu 22, Praha 7 : Grada Publishing, a.s., 2017. ISBN 978-80-271-0918-0.
6. **Davis, Joel a Mokalis, Alexa.** *Google Analytics Demystified (4th Edition)*. místo neznámé : CreateSpace, 2018. ISBN 978-1545486917.
7. **Riman, Tarek.** *The Secret to Capitalizing on Analytics*. místo neznámé : Independently published, 2019. ISBN 978-1796616194.
8. **Rajtmajer, František.** Co jsou Google Analytics. *Nazakladedat.cz*. [Online] 27. 5 2020. [Citace: 10. 9 2022.] <https://nazakladedat.cz/co-jsou-google-analytics/>.
9. **Tichý, Jan.** 3 důvody, proč přejít na placené Google Analytics 360. *Digichief*. [Online] 11. 2 2021. [Citace: 18. 10 2022.] https://digichief.cz/google-analytics-360?fbclid=IwAR3b86y45u70VDbvRevGkQgfbCenQjGUro4vFENzx845K7sjPxo2X8F_MrE.
10. **Widner, Christina.** Toto jsou nejlepší alternativy Google Analytics 6. *Raidboxes*. [Online] 21. 11 22. [Citace: 20. 1 2023.] <https://raidboxes.io/cs/blog/wordpress/google-analytics-alternative/>.
11. **Consentmanager.** Používejte Matomo v souladu s GDPR. *Consentmanager*. [Online] [Citace: 20. 1 2023.] <https://www.consentmanager.cz/znalost/vseobecne/matomo-v-souladu-s-gdpr/>.
12. **Tonkin, Sebastien, Caleb, Whitmore a Cutroni, Justin.** *Výkonnostní marketing s Google Analytics*. Holandská 3, 639 00 Brno : Computer Press, a.s., 2011. ISBN 978-80-251-3339-2.
13. **Rajtmajer, František.** Co je dataLayer. *Nazakladedat.cz*. [Online] 16. 6 2022. [Citace: 25. 2 2023.] <https://nazakladedat.cz/co-je-datalayer/>.
14. **GoogleAnalytics.** Measurement Protocol (Google Analytics 4). *GoogleAnalytics*. [Online] [Citace: 12. 2 2022.] <https://developers.google.com/analytics/devguides/collection/protocol/ga4>.
15. **Clifton, Brian.** *Google Analytics: podrobný průvodce webovými statistikami*. Brno : Computer Press, 2009. ISBN 9788025122310.
16. **Zeman, Milan.** Průvodce Google Analytics pro začátečníky. *Můjsvětmarketingu*. [Online] 5. 12 2018. [Citace: 19. 9 2022.] <https://mujsvetmarketingu.cz/digitalni-marketing/pruvodce-google-analytics-pro-zacatecniky/>.
17. **Rajtmajer, František.** Co jsou dimenze a metriky. *Nazakladedat.cz*. [Online] 7. 2 2020. [Citace: 10. 10 2022.] <https://nazakladedat.cz/dimenze-a-metriky/>.
18. **JavaTpoint.** What is Data Cube? *JavaTpoint*. [Online] [Citace: 11. 11 2022.] <https://www.javatpoint.com/data-warehouse-what-is-data-cube>.

19. **Navrátil, Lukáš.** Datové kostky. *Univerzita Karlova*. [Online] 15. 11 2011. [Citace: 18. 10 2022.]
https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fwww.ksi.mff.cuni.cz%2F~pokorny%2Fdj%2Fprezentace%2F3_67.pptx&wdOrigin=BROWSELINK.
20. **Russel, Kay.** Zaostřeno na datové kostky. *Science Word*. [Online] [Citace: 5. 10 2022.]
<https://www.scienceworld.cz/neziva-priroda/zaostreno-na-datove-kostky-2223/>.
21. **Metabase.** Data cubes. *Metabase*. [Online] [Citace: 24. 8 2022.]
<https://www.metabase.com/learn/databases/data-cube>.
22. **Omniconvert.** What is... Unique visitors. *Omniconvert*. [Online] 29. 1 2021. [Citace: 2. 11 2022.] <https://www.omniconvert.com/what-is/unique-visitors/>.
23. **Zantal-Wiener, Amanda.** What Is Bounce Rate? (And How Can I Fix Mine?). *HubSpot*. [Online] 3. 1 2022. [Citace: 1. 10 2022.]
<https://blog.hubspot.com/marketing/what-is-bounce-rate-fix>.
24. **Upqode.** Bounce Rate vs Exit Rate – What is the Difference? *Upqode*. [Online] 20. 9 2022. [Citace: 12. 11 2022.] <https://upqode.com/bounce-rate-vs-exit-rate/>.
25. **Masooma, Memon.** Exit Rate vs Bounce Rate – How to Analyze Both to Improve Your Content Strategy? *Databox*. [Online] 24. 1 2022. [Citace: 10. 1 2023.]
<https://databox.com/exit-rate-vs-bounce-rate>.
26. **Larry, Kim.** What's a Good Conversion Rate? (It's Higher Than You Think). *WordStream*. [Online] 12. 11 2022. [Citace: 10. 12 2022.]
<https://www.wordstream.com/blog/ws/2014/03/17/what-is-a-good-conversion-rate>.
27. **Sharma, Himanshu.** *Maths and Stats for Web Analytics and Conversion Optimization*. místo neznámé : Blurb, 2015. ISBN 978-1364849184.
28. **Kolektiv autorů.** *Online marketing*. Brno : Computer Press, a.s., 2014. ISBN 978-80-251-4155-7.
29. **Hayes, Adam.** Aggregate Function: Definition, Examples, and Uses. *Investopedia*. [Online] 30. 3 2022. [Citace: 16. 12 2022.]
<https://www.investopedia.com/terms/a/aggregate-function.asp>.
30. **Gupta, Aryan.** SQL Aggregate Functions: SUM(), COUNT(), AVG(), Functions. *Simplilearn*. [Online] 10. 1 2023. [Citace: 3. 2 2023.]
<https://www.simplilearn.com/tutorials/sql-tutorial/sql-aggregate-functions>.
31. **Fiřtová, Lenka.** Korelace. *ExcelTown*. [Online] [Citace: 14. 9 2022.]
<https://exceltown.com/navody/pokrocila-analyza-regrese-korelace/korelace-co-to-vlastne-je/>.
32. **Zikmundová, Markéta.** Lineární regrese. *Vysoká škola chemicko-technologická v Praze*. [Online] [Citace: 20. 11 2022.]
https://web.vscht.cz/~zikmundm/as_piga2016_soubory/lin-regrese/reseni_linearni_regrese_excel.pdf.
33. **Corexen-John, Dan a Van Tonder, Johann.** *Optimalizace webových stránek: Proč si 95% návštěvníků vašich stránek nic nekoupí a jak to můžete změnit*. [překl.] Petra Šmídová. Brno : Lingea, 2022. ISBN 978-80-7508-716-4.
34. **Alhlou, Feras, Asif, Shiraz a Fettman, Eric.** *Google Analytics Breakthrough*. Hoboken, New Jersey : John Wiley & Sons, Inc., 2016. ISBN 9781119231707.

8 Seznam obrázků, tabulek, grafů a zkratk

8.1 Seznam obrázků

Obrázek 1 Staré paradigma webové analytiky.....	14
Obrázek 2 Klíčové otázky spojené s webovou analytikou 2.0	15
Obrázek 3 Ukázka vložení dat do dataLayeru	18
Obrázek 4 Přehled architektury.....	19
Obrázek 5 Životní cyklus webových metrik	25
Obrázek 6 Ukázka kódu, denní agregace dat.....	33
Obrázek 7 Ukázka kódu, měsíční a produktová agregace dat	34
Obrázek 8 Ukázka kódu, produktová agregace dat.....	36
Obrázek 9 Transformovaný produktový katalog	48
Obrázek 10 Ukázka kódu na hledání alternativních produktů.....	48
Obrázek 11 Nalezené alternativní produkty a jejich sdílené atributy	49
Obrázek 12 Počet nalezených variant k danému produktu	49
Obrázek 13 Náhled produktu a přidání tlačítka "Na seznam"	50

8.2 Seznam tabulek

Tabulka 1 Základní sledovací soubory cookie služby Google Analytics	20
Tabulka 2 Události produktů.....	29
Tabulka 3 Produktový katalog	30
Tabulka 4 Transakční data	31
Tabulka 5 Nákupní listy	31
Tabulka 6 Nákupy zákazníků.....	32
Tabulka 7 Výsledná data po agregaci na úrovni dne	33
Tabulka 8 Výsledná data po agregaci na úrovni měsíce a produktu.....	34
Tabulka 9 Výsledná data po agregaci na úrovni produktu.....	36
Tabulka 10 Chování zákazníků na webu	37
Tabulka 11 Regresní analýza – koeficienty závislosti mezi „S1-S2“	38
Tabulka 12 Regresní analýza – výsledek regresní analýzy.....	39
Tabulka 13 Rozdělení dat	41
Tabulka 14 Závislost kroků v datasetu "All" na denní agregaci.....	42
Tabulka 15 Závislost kroků v datasetu "White" na denní agregaci	43
Tabulka 16 Závislost kroků v datasetu "Red" na denní agregaci	43
Tabulka 17 Porovnání koeficientů z datasetů	44
Tabulka 18 Měsíční vývoj závislosti jednotlivých kroků nákupního procesu.....	45
Tabulka 19 Závislost jednotlivých kroků nákupního produktů na úrovni produktů.....	46
Tabulka 20 Kolik procent z toho, co zákazník přidá a následně nakoupí.....	51

8.3 Seznam grafů

Graf 1 Závislost počtu prokliků na počtu zobrazení produktu.....	40
Graf 2 Konverzní poměr S1-S5 Top 100 produktů.....	47
Graf 3 Konverzní poměr S1-S5 Bot 100 produktů	47
Graf 4 Konverzní poměr toho, co zákazník přidá na seznam a následně nakoupí	51
Graf 5 Počet dní od přidání produktu na seznam po jeho koupi.....	52
Graf 6 Počet dní od znovupoužití funkce "Na seznam"	53