# Závěry a postup řešení analýzy

**Vypracoval: Jan Koppan**

Zadání:

V přiloženém Excelu máte dvě tabulky. Jedna představuje transakce (objednávky), které proběhly v E-shopu. Druhá představuje databázi produktů v E-shopu. Ke všem otázkám níže prosím popisujte i postup, jak jste k danému závěru došli. Pro zodpovězení otázek použijte libovolný nástroj. V ideálním případě pošlete přímo i kód, excel apod., kde bude váš postup viditelný. Odpovědi musí být v češtině.

## Postup a řešení:

#### a) Na jaké kategorii produktů máme největší obrat?

Postup: Spojil jsem sešity Transactions a Products pomocí funkce XLOOKUP. Díky tomu jsem si vytvořil nové sloupce Cena a Celková tržba v sešitu Transakce pomocí příkazu: =XLOOKUP(C2;Products!A:A;Products!C:C). Také jsem si vypočítal pomocný sloupec Turnover.

Poté jsem si vytvořil první kontingenční tabulku pro zobrazení celkových tržeb podle kategorií produktů. Do řádků jsem přidal kategorie produktů a do hodnot součet celkových tržeb.

V poslední fázi jsem seřadil všechny kategorie podle celkových tržeb.

Výsledek:

|  |  |
| --- | --- |
| **Kontingenční tabulka pro otázku č. 1 a)** | |
| Celkový obrat podle kategorií produktů | |
| **Kategorie zboží** | **Celkem tržby** |
| Televize | 1205680 |
| Tablety | 623660 |
| Mobilní telefony | 614460 |
| Audio | 416370 |
| **Celkový součet** | **2860170** |

Podle první kontingenční tabulky vidíme, že kategorie s největším obratem je kategorie Televize. Tato kategorie měla celkový objem tržeb v hodnotě 1 205 680,- Kč.

#### b) Zajímalo by mě, jestli se obrat v jednotlivých měsících mění.

**Postup:** Vytvořil jsem druhou kontingenční tabulku pro zobrazení celkového obratu podle jednotlivých měsíců, do které jsem do oblasti řádky dal Měsíce a Kategorie, a do oblasti hodnot Celkem kusů a Celkem tržeb. V poslední řadě jsem seřadil kategorie od nejvyšší podle celkových tržeb.

**Výsledek:**

|  |  |  |
| --- | --- | --- |
| **Kontingenční tabulka pro otázku č. 1 b)** | |  |
| Celkové tržby a celkem kusů podle času | |  |
| **Měsíce** | **Celkem kusů** | **Celkem tržby** |
| **II** |  |  |
| Televize | 8 | 290920 |
| Mobilní telefony | 11 | 208690 |
| Tablety | 11 | 188890 |
| Audio | 19 | 130310 |
| **III** |  |  |
| Televize | 10 | 380900 |
| Tablety | 9 | 164910 |
| Mobilní telefony | 9 | 157710 |
| Audio | 18 | 107820 |
| **IV** |  |  |
| Televize | 7 | 266930 |
| Mobilní telefony | 9 | 177510 |
| Tablety | 9 | 171910 |
| Audio | 16 | 107840 |
| **V** |  |  |
| Televize | 7 | 266930 |
| Tablety | 5 | 97950 |
| Mobilní telefony | 5 | 70550 |
| Audio | 10 | 70400 |
| **Celkový součet** | **163** | **2860170** |

Podle druhé kontingenční tabulky má kategorie **Televize** největší obrat ve všech čtyřech měsících, přičemž v měsíci březen dosáhla vrcholu s obratem 380 900. Pro ostatní kategorie se obrat v jednotlivých měsících liší, ale žádná z nich nepřekonává kategorii Televize v žádném z měsíců.

#### Který den v týdnu je nejsilnější na počet objednávek?

**Postup:** Vytvořil jsem si dva pomocné sloupce (Pomocný sloupec: Den v týdnu a Pomocný sloupec: Název dne v týdnu). První sloupec pomocí funkce: =DENTÝDNE(B2;2), vypíše v číslech, o jaký den v týdnu se jedná. Druhý sloupec potom pro lepší přehlednost převede tento číselný údaj na slovní vyjádření dne v týdnu pomocí funkce: ZVOLIT(H2; "Pondělí"; "Úterý"; "Středa"; "Čtvrtek"; "Pátek"; "Sobota"; "Neděle").

Následně jsem vytvořil kontingenční tabulku, do oblasti řádky jsem dal dny v týdnu a do hodnot počet objednávek. Kontingenční tabulku jsem poté seřadil podle nejvyššího počtu objednávek. Pokud by se výsledek nevyhledat pomocí seřazení, mohli bychom také použít funkci: =XLOOKUP(MAX(M36:M42);M36:M42;L36:L42), která nám potvrdí správný výsledek.

Výsledek:

|  |  |
| --- | --- |
| **Kontingenční tabulka pro otázku č. 2** | |
| **Popisky řádků** | **Počet objednávek** |
| Sobota | 29 |
| Pondělí | 25 |
| Středa | 22 |
| Pátek | 20 |
| Úterý | 16 |
| Čtvrtek | 15 |
| Neděle | 14 |
| **Celkový součet** | **141** |

Z kontingenční tabulky vytvořené v excelu vyplývá, že nejsilnější den na počet objednávek je **sobota**, protože došlo k 29 objednávkám.

#### Která kategorie se prodává nejčastěji spolu s produkty z kategorie Televize (resp. jsou spolu v jedné objednávce)?

**Postup:** Vytvořil jsem si pomocný sloupec: Kategorie zakoupené společně s televizí, který ukazuje každou kategorii, která byla zakoupena společně v jedné objednávce s kategorií Televize, a zároveň se nejednalo o další televizi. Tento sloupec byl vytvořen pomocí funkce: =KDYŽ(E2="Televize"; TEXTJOIN(", "; PRAVDA; KDYŽ(B:B=B2; KDYŽ(E:E<>"Televize"; E:E; ""); "")); "").  
Tento příkaz nám ukáže další kategorie zakoupené s televizí. Všechny tyto kategorie nám poté přehledněji ukazuje další sloupec: Pomocný sloupec pro otázku č. 3 pomocí příkazu: =UNIQUE(FILTER(J2:J142; (J2:J142<>"") \* (J2:J142<>0)))

Na výsledcích je ovšem vidět duplicita výsledných kategorií, protože se v jedné objednávce může unikátní kategorie koupená s televizi počítat pouze jednou. Nakonec jsem si vytvořil pomocnou závěrečnou tabulku, kde jsem si dopočetl pomocí funkce: =COUNTIF(K2#;"\*Audio\*") kolikrát se vyskytly kategorie Audio, Mobilní telefony a Tablety.

Výsledek:

|  |
| --- |
| **Výsledky pro otázku č.3** |
| Kolikrát se vyskytla kategorie Audio: |
| **10** |
| Kolikrát se vyskytla kategorie Mobilní telefony: |
| **8** |
| Kolikrát se vyskytla kategorie Tablety: |
| **8** |

Nejčastěji se spolu s produkty z kategorie Televize prodává kategorie **Audio**, protože se tato kategorie vyskytla v jedné objednávce s kategorií Televize celkem 10x.

#### Od 18. 3. 2022 jsem klukům z marketingu navýšil budget na online marketingových platformách (Google Ads, Sklik, Facebook). Dokážeš mi říct, jestli to vedlo k nějaké změně v prodeji?

Postup: Rozdělil jsem si celkové tržby na dvě období, kde je milníkem datum 18.3.2022 (Před zvýšením budgetu a po zvýšení budgetu na marketing).

Poté jsem z obou období udělal průměrný obrat pomocí funkce: =PRŮMĚR($G$2:$G$65).

Ten jsem poté porovnal a převedl na procentní vyjádření, abychom poznali změnu pomocí funkce: =((L50-L48)/L48)\*100.

Výsledek:

|  |
| --- |
| **Pomocná tabulka pro otázku č.4:** |
| Průměrný obrat v období před 18.3.2022 |
| **19858,75** |
| Průměrný obrat v období po 18.3.2022 |
| **20639,09091** |
| Percentuální změna |
| **3,929456331** |

Dalo by se tedy říct, že navýšení budgetu na reklamu mohlo opravdu zvýšit průměrné tržby pro daná období **o 3,929 %**, pokud nebereme v potaz žádné další proměnné. V analýze dat je totiž důležité brát v úvahu, že i když je zde korelace mezi navýšením rozpočtu a zvýšením průměrných tržeb, nemusí to nutně znamenat příčinnou souvislost. Existují vždy další proměnné, které by mohly ovlivnit výsledek, jako například sezónní trendy, konkurence, změny v produktu nebo chování zákazníků.

#### Na seznamu produktů v E-shopu máme produkty nějak řazeny. Vždy stejně a nemění se to. Myslím, že to je řazeno podle toho, jak je přidáváme postupně do databáze. Pak si je samozřejmě může zákazník seřadit sám podle ceny. Myslíš, že hraje roli, jaké mají ty produkty na produktových listech pořadí, tedy jak jsou seřazeny defaultně? Je mi asi jasné, že ze současných dat to nevyčteš. Máš nějaký návrh, co bychom tedy potřebovali začít měřit, abychom se v tomto tématu posunuli dále?

**Vliv defaultního pořadí produktů na chování zákazníků v e-shopu**

Určitě si myslím, že defaultní pořadí produktů na produktových listech může významně ovlivnit, jak jsou tyto produkty prezentovány zákazníkům, a zároveň jaké produkty si vybírají k prohlédnutí nebo nákupu. Ačkoliv si tento faktor mnozí nemusí uvědomovat, je důležité jeho vliv důkladně prozkoumat. Navrhl bych proto různé testovací metody, které by nám umožnily objektivně posoudit význam tohoto aspektu na chování zákazníků.

Jednou z metod, kterou bychom mohli využít, je A/B testování. Tato technika by umožnila nastavit několik verzí produktové stránky, kde by byly produkty seřazeny různými způsoby, například podle popularity nebo novinek. Měřením těchto variant bychom zjistili, zda různá uspořádání produktů ovlivňují chování zákazníků.

Dalšími užitečnými nástroji pro sledování chování zákazníků by mohly být analytické nástroje, jako je Google Analytics, a sledování heatmaps pomocí softwaru jako je Hotjar. Tyto nástroje by nám umožnily sledovat cestu zákazníka od příchodu na stránku až po dokončení nákupu, což by přispělo k hlubšímu porozumění tomu, zda defaultní řazení produktů usnadňuje, či naopak brzdí proces prodeje. Heatmapy nám poskytnou cenné informace o tom, na která místa na stránce se zákazníci nejvíce soustředí.

Kromě těchto technik bychom také mohli implementovat zákaznické dotazníky a průzkumy, které by nám mohly poskytnout důležitou zpětnou vazbu přímo od zákazníků. Tyto informace by nám pomohly lépe pochopit preference a očekávání našich uživatelů.

Další metodou jsou experimenty. Sledováním změn, například v pořadí produktů, můžeme získat cenné informace o tom, jak různé faktory ovlivňují rozhodovací proces zákazníků.

Tyto metody, spolu s mnoha dalšími metrikami, nám umožní lépe porozumět tomu, jak různé pořadí produktů ovlivňuje rozhodovací proces zákazníků a jaké změny by mohly vést k lepšímu výkonu našich produktů v e-shopu.

#### As a data analyst working with online marketing data and Google Analytics 4 (GA4), you are aware of the challenges posed by GDPR regulations and cookie consent requirements on data collection. Using a Large Language Model (like ChatGPT), outline the key impacts of GDPR and cookie consent on web analytics data collection. Then, generate a plan to mitigate these data limitations while ensuring compliance with GDPR. Provide the specific prompts you used with the LLM and summarize the responses you received.

**Key Impacts of GDPR and Cookie Consent on Web Analytics Data Collection**

The General Data Protection Regulation (GDPR) and cookie consent requirements have dramatically influenced how businesses collect and process web analytics data. The primary impacts are outlined below:

1. **Consent Requirements:** GDPR mandates that businesses must obtain explicit consent from users before collecting any personal data, including data collected via cookies. This consent requirement often leads to reduced data availability, as users may choose to reject cookies. This directly limits the amount of information businesses can collect on user behavior.
2. **Personal Data Restrictions:** GDPR imposes strict limitations on the collection and processing of personal data, including personally identifiable information (PII) such as IP addresses, user identifiers, and any other data that could potentially identify an individual. As a result, analytics platforms like Google Analytics 4 (GA4) have had to adjust their data collection practices to comply with these rules, which may reduce the granularity of the data collected.
3. **Data Anonymization and Pseudonymization:** GDPR encourages the anonymization and pseudonymization of data to protect user identities. This can impact the level of detail in web analytics, as anonymized data makes it difficult to track users across multiple sessions or devices, limiting the ability to analyze user journeys in detail.
4. **Data Retention Limitations:** Under GDPR, businesses are required to minimize the duration for which they retain personal data. This impacts long-term data analysis and trends, as organizations need to delete or anonymize data after a specific period, often leading to challenges in maintaining historical records for extended periods.
5. **User Rights to Data:** GDPR grants users rights such as the right to access, rectify, and erase their personal data. This creates operational challenges for businesses, as they need to establish systems that allow for the retrieval and deletion of personal data upon request. For web analytics, this means that businesses must implement measures to handle user data in ways that are flexible to these rights while maintaining compliance.

**Plan to Mitigate Data Limitations While Ensuring GDPR Compliance**

To address the limitations imposed by GDPR and cookie consent regulations on web analytics, businesses can adopt several strategies to ensure compliance while still collecting valuable insights. The following plan outlines key mitigation strategies:

1. **Leverage First-Party Data**: One effective approach is to prioritize the collection of first-party data, which is data collected directly by the website (e.g., user behavior on the site) rather than relying on third-party cookies. First-party data is less affected by GDPR, and businesses can still gather useful insights without the need for extensive consent from users.
2. **Implement IP Anonymization**: Businesses can mitigate the limitations imposed by personal data restrictions by implementing IP anonymization. Many analytics platforms, such as GA4, offer built-in IP anonymization tools that allow businesses to collect approximate location data without violating GDPR. This ensures that insights into geographical trends are still available while protecting users' identities.
3. **Use Aggregated Data**: Instead of focusing on granular, user-specific data, businesses can shift towards using aggregated data to track broader trends and performance metrics. Aggregated data allows for useful insights without collecting personally identifiable information, thus reducing compliance risks.
4. **Optimize Consent Mechanisms**:Improve the design and transparency of cookie consent banners to maximize user opt-in rates for tracking. Clear communication about how data will be used, combined with offering granular consent options (e.g., allowing users to select specific cookie categories), can increase the likelihood of users providing consent.
5. **Adopt Privacy-First Analytics Tools**: Consider switching to privacy-focused analytics tools, such as Matomo, which are designed with GDPR compliance in mind and do not rely on cookies for tracking. While these platforms may offer fewer data points than traditional analytics tools, they ensure that businesses can still track essential metrics without violating data protection laws.
6. **Employ Data Anonymization and Pseudonymization Techniques**: By applying advanced anonymization techniques, businesses can still perform meaningful analyses without relying on identifiable personal data. Pseudonymization, for example, allows businesses to use pseudonyms or tokens to represent users, maintaining a degree of tracking without exposing their identities.
7. **Continuous Monitoring and Auditing**: Businesses should establish ongoing processes to regularly monitor their data collection practices for GDPR compliance. This includes conducting frequent data audits, ensuring that personal data is only retained for the necessary period, and reviewing user consent records to verify compliance.
8. **Educate Users and Build Trust**: Increase user awareness of their data privacy rights and how their data is used. Clear communication about the benefits of data collection, as well as transparency in privacy policies, can help build user trust and increase the number of users willing to provide consent for tracking.

**Summary**

1. **Key Impacts of GDPR and Cookie Consent**:
   * **Consent Requirements**: GDPR mandates obtaining explicit user consent for data collection, often leading to reduced data availability as users may refuse cookies.
   * **Restrictions on Personal Data**: Strict limitations on the collection of personally identifiable information (PII) hinder detailed tracking and analysis of user behavior.
   * **Anonymization and Pseudonymization**: Businesses are encouraged to anonymize data to protect user identities, impacting the granularity of analytics.
   * **Data Retention Limitations**: GDPR requires minimizing data retention periods, complicating long-term data analysis and historical trend tracking.
   * **User Rights**: Users have rights under GDPR that affect how businesses collect, manage, and retain their data, necessitating flexible data handling practices.
2. **Mitigation Strategies**:
   * **Leverage First-Party Data**: Focus on collecting first-party data to bypass some consent requirements while still gathering useful insights.
   * **Implement IP Anonymization**: Use tools that anonymize IP addresses to protect user identities while still capturing location-based insights.
   * **Utilize Aggregated Data**: Shift towards analyzing aggregated data instead of individual user data to maintain compliance without sacrificing insights.
   * **Optimize Consent Mechanisms**: Improve cookie consent banners for clarity and transparency, enhancing user willingness to opt-in for data tracking.
   * **Adopt Privacy-First Analytics Tools**: Consider using analytics platforms designed for GDPR compliance, which do not rely heavily on cookies.
   * **Employ Data Anonymization Techniques**: Apply anonymization and pseudonymization techniques to protect user identities while collecting relevant data.
   * **Educate Users on Data Privacy**: Increase user awareness of their data privacy rights and how data is utilized to build trust and encourage consent.
   * **Establish Continuous Compliance Monitoring**: Set up processes for regular monitoring and auditing of data practices to ensure ongoing GDPR compliance.

**Prompts Used with the Chat-GPT.**

1. **Understanding GDPR and Cookie Consent:**
   * "What are the key impacts of GDPR and cookie consent regulations on web analytics data collection?"
2. **Exploring Mitigation Strategies:**
   * "What strategies can businesses implement to mitigate the limitations imposed by GDPR and cookie consent while ensuring compliance with data protection regulations?"
3. **User Rights and Data Practices:**
   * "How do user rights under GDPR affect data collection and retention practices in web analytics?"
4. **Compliance Challenges:**
   * "What challenges do businesses face in complying with GDPR and cookie consent requirements when collecting web analytics data?"
5. **Best Practices for Consent Management:**
   * "What are the best practices for managing user consent for cookie usage and data collection on websites?"
6. **Data Anonymization Techniques:**
   * "What data anonymization techniques can be used to comply with GDPR while still enabling effective web analytics?"
7. **User Education on Data Privacy:**
   * "How can businesses educate users about their data privacy rights and the use of cookies in web analytics?"
8. **Continuous Compliance Monitoring:**
   * "What processes should businesses establish for continuous monitoring of GDPR compliance in their data collection practices?"