# **Case Study Orlen – Jan Koppan**

**Postup řešení domácí case study - BI analýza**

Tato případová studie popisuje můj konkrétní postup při řešení analytické úlohy nad daty maloobchodní sítě. Celý proces probíhal od prvního otevření Excelového souboru až po interpretaci výsledků v interaktivním dashboardu. Začal jsem základní analýzou a úpravou dat, pokračoval jejich nahráním a transformací do Power BI a vytvořením výpočtů pomocí DAXu. Nakonec jsem připravil přehledný dashboard a doplnil jej hlavními závěry, trendy a doporučeními, které z něj plynou.

**1. Seznámení se zadáním a daty**

Nejprve jsem si otevřel Excelový soubor a seznámil se s jeho strukturou. Prohlédl jsem si jednotlivé sloupce, především FS ID, segmenty jako Paliva, Gastro, Shop a Cigarety, dále tržby, marže, časové a regionální informace. Zkontroloval jsem konzistenci a upravil správnost formátu dat a identifikoval klíčové hodnoty, se kterými budu dál pracovat. Už při prvním pohledu bylo zřejmé, že se dataset dělí do čtyř hlavních kategorií, kterým jsem pak přizpůsobil i analytickou strukturu.

**2. Čištění a příprava dat**

V Excelu jsem se nejprve snažil najít odstranil duplicity, prázdné nebo chybně zapsané hodnoty. U sloupce „Gastro koncept“ jsem sjednotil odlišné zápisy jako „–“ a „unknown“ pod jednu kategorii „unknown“. U těchto dvou poboček je nutné zjistit proč je gastro koncept neznámý. Všiml jsem si několika podezřelých, často se opakujících hodnot, které jsem ponechal, protože dataset je zřejmě cvičný. I přesto jsem data ale zvýraznil žlutě v Excel souboru. Zároveň jsem si všiml že ne každá pobočka obsahuje všechny kategorie a vytvoří to tedy chyby.

Vytvořil jsem také nové pomocné listy. Dopočítal jsem zisk pomocí vzorce tržeb a marží. Aby vzorec fungoval i při chybějících hodnotách, použil jsem funkce jako IFERROR a XLOOKUP.

**3. Úprava údajů o otevírací době**

Otevírací doba byla zapsaná různě, takže jsem ji musel sjednotit. Pomocí VBA jsem napsal jednoduché makro, které mi počet hodin otevření spočítalo z rozdílných otevíracích dob. Výsledky jsem následně ručně zkontroloval pomocí funkcí jako HODNOTA, ZLEVA, ZPRAVA nebo DÉLKA. U složitějších zápisů jsem část zpracoval ručně nebo využil pomoc od ChatGPT. Tento výpočet potvrdil správnost makra.

Z počtu hodin kdy je otevřeno jsem dopočítal zisk na hodinu. U tří segmentů (Gastro, Shop, Cigarety) jsem dopočítal i náklady pomocí rozdílu tržby a zisk. U Paliva to ale možné nebylo, protože jsme neměli uvedenou ani cenu za litr, ani tržbu v korunách, pouze v litrech. Proto jsem náklady zobrazil jen pro ty tři zmíněné segmenty.

**4. Nahrání a modelování v Power BI a DAX metriky**

Po důkladném zkontrolování jsem připravená data jsem nahrál do Power BI přes Power Query, kde jsem provedl další transformace sloupců a typů dat. Vytvořil jsem relační model, který propojuje tabullky přes FS ID. Pomocí DAXu jsem si dovytvořil další potřebné metriky jako celkový zisk na hodinu, průměrné marže.

Pomocí DAXu jsem vytvořil také samostatné tabulky pro TOP 10 nejvýdělečnějších poboček podle celkového zisku a podle zisku na hodinu. To samé jsem udělal také pro Bottom deset, tedy nejméně ziskové podle celkového zisku, podle zisku na hodinu a podle tržeb. Výpočet zisku na hodinu vycházel z dopočtu otevírací doby v Excelu. Napadlo mě, že v praxi by šlo některé kroky složitější nahradit Pythonem, ale v tomto případě jsem zůstal u VBA. Pokud bych potřeboval rychle filtrovat nebo analyzovat data, využil bych SQL. Například jednoduchý výběr TOP 10 poboček v kategorii Gastro by šel zapsat pomocí SQL dotazu:

Obsah obrázku text, snímek obrazovky, Písmo

Obsah vygenerovaný umělou inteligencí může být nesprávný.

**5. Tvorba dashboardu**

Dashboard jsem sestavil ve firemní identitě Orlenu. Použil jsem bílo-červené barvy a přidal logo firmy. Do dashboardu jsem také přidal úvodní rozklikávací navigační okno a po stranách ikonky doleva a doprava které pomáhají s navigací po stránkách.

Při vytváření dashboardu jsem se snažil dodržovat osvědčené postupy (best practices) v oblasti vizualizace dat. Důraz jsem kladl na přehlednost, funkčnost a logickou návaznost všech částí dashboardu.

Na úvodní stránce se zobrazují základní čísla jako počet stanic, regionů a manažerů. Dále jsem zobrazil srovnání tržeb, zisků a nákladů v čase pro všechny segmenty a taktéž pro celkové zisky a zisky na hodinu od ledna 2022 do září 2024. Přidal jsem možnost filtrování podle segmentu, regionu, otevřených hodin, gastro konceptu, regionálního manažera a dalších parametrů. Pro obchod jsem přidal filtr podle maržové kategorie, pro palivo filtr podle cenové kategorie. Také jsem upravil popisky a názvý všech grafů a tabulek do přehledné podoby.

Vytvořil jsem také další vizualizace jako mapu rozložení stanic podle regionů, koláčový graf, který ukazuje podíl regionálních manažerů na celkovém počtu stanic a také výběr TOP/BOTTOM poboček dle zisku, tržeb nebo zisku na hodinu. Dashboard je interaktivní a umožňuje rychlou navigaci mezi segmenty.

**6. Analytická interpretace a storytelling**

Na základě grafů a vizualizací jsem identifikoval klíčové trendy. Z těchto trendů jsem postupně udělal přehledné slidy, kde uvádím tyto trendy k jednotlivým kategoriíí, interpretuji výsledky a následně dávám doporučení. Mezi hlavní závěry mimo jiné patří: Segment Palivo táhl zvýšené zisky za sledované období. Gastro a Shop mají sezónní charakter. Cigarety stagnují, ale stále jsou stabilním zdrojem tržeb. Letní měsíce bývají pro síť nejvýnosnější. Zároveň se ukázalo, že pobočky s nejvyšším ziskem na hodinu nemusí být nutně ty největší. Výsledky jsem převedl do doporučení: Zvýšit efektivitu poboček na základě zisku na hodinu, zaměřit se na aktivní řízení segmentů Gastro a Shop, dále rozvíjet segment Palivo jako stabilní pilíř výkonu a zjednodušit reporting pomocí připravených dashboardů.

**Shrnutí kroků:**

1. Seznámení se s daty a pochopení jejich struktury
2. Čištění a úprava dat, výpočty zisku, otevírací doba
3. Nahrání dat do Power BI, jejich transformace a vytvoření relačního modelu
4. Tvorba metrik a pomocných tabulek a sloupců v DAXu
5. Design a sestavení přehledného dashboardu s funkčními prúřezy
6. Nalezení a vysvětlení hlavních trendů a sestavení doporučení