Semestrálny projekt v Power BI

Cieľ projektu. Cieľom je vytvoriť ucelené riešenie v Power BI, ktoré prejde celým analytickým procesom: od výberu a pochopenia dát cez ich čistenie v Power Query, návrh dátového modelu, tvorbu DAX výpočtov až po interaktívny report a formulovanie vecných záverov. Projekt nemá byť iba súborom pekných grafov. Dôležité je preukázať, že rozumiete dátam, modelu a viete pomocou reportu odpovedať na konkrétne analytické otázky.

1. Výber témy a dát

Tému si každý zvolí sám. Odporúčané sú transakčné alebo udalostné dáta: predaje a e-commerce, doprava, šport, turizmus, reality, energetika, životné prostredie, verejné financie, filmy, videohry, anonymizované zdravotnícke štatistiky, dopravné nehody a podobne. Pri osobných alebo citlivých údajoch možno použiť iba verejné, anonymizované alebo syntetické dáta.

Oblasť	Faktová tabuľka	Možné dimenzie	Čo možno analyzovať
E-commerce	objednávky	produkt, zákazník, región, dátum	tržby, marža, trendy
Doprava	jazdy/lety	dopravca, trasa, letisko, čas	meškania, vyťaženosť
Šport	zápasy/výkony	hráč, tím, súťaž, dátum	výsledky a výkonnosť
Turizmus	návštevy/rezervácie	krajina, zariadenie, obdobie	sezónnosť, návštevnosť
Reality	predaje/prenájmy	lokalita, typ nehnuteľnosti, čas	ceny a trendy
Energetika	merania	zdroj, región, čas	výroba, spotreba
Životné prostredie	merania	stanica, lokalita, čas	kvalita ovzdušia, teplota
Verejné financie	výdavky/príjmy	organizácia, kategória, región, čas	štruktúra výdavkov
Filmy/streaming	filmy/hodnotenia	žáner, krajina, tvorca, rok	popularita, hodnotenia
Videohry	predaje/hodnotenia	hra, platforma, žáner, vydavateľ	trendy trhu
Zdravotníctvo	anonymizované udalosti	diagnóza, zariadenie, čas	počty a trendy
Kriminalita/nehody	udalosti	typ, miesto, dátum	vývoj a geografické rozdiely

Projekt musí začínať analytickým problémom. Pred tvorbou reportu si sformulujte približne 4–6 otázok, na ktoré chce pomocou dát odpovedať. Napríklad: Ktoré kategórie prinášajú najvyššie tržby? Ako sa výsledky menia v čase? Ktoré regióny alebo skupiny sa odlišujú? Kde vznikajú problémy alebo príležitosti?

2. Minimálne požiadavky na dataset

Požiadavka	Minimum / očakávanie
Štruktúra	Aspoň 3 logicky súvisiace tabuľky; odporúčané sú 4–6 tabuliek.
Faktová tabuľka	Aspoň jedna tabuľka transakcií alebo udalostí (napr. objednávky, jazdy, zápasy, merania).
Dimenzie	Aspoň 2 dimenzie, napr. dátum, produkt, zákazník, lokalita, kategória, tím alebo dopravca.
Čas	Dataset musí obsahovať dátum alebo čas a umožňovať analýzu vývoja v čase.
Číselné údaje	Aspoň 2 zmysluplné číselné veličiny, ktoré možno agregovať alebo z nich odvodiť KPI.
Kategórie	Aspoň 2 kategorické atribúty vhodné na segmentáciu a filtrovanie.
Rozsah	Odporúčané sú tisíce riadkov vo faktovej tabuľke. Veľkosť sama osebe nie je cieľom; dôležitejšia je analytická použiteľnosť.
Kvalita dát	Dáta nemusia byť dokonale čisté. Projekt má obsahovať reálne transformácie v Power Query.
Zdroj	Je potrebné uviesť zdroj dát a stručne vysvetliť ich pôvod, význam a prípadné obmedzenia.

Nevhodné dáta: jedna malá hotová tabuľka s niekoľkými desiatkami riadkov, dataset bez časovej dimenzie, alebo už kompletne pripravený „dashboard dataset“, pri ktorom nie je potrebné riešiť transformácie, vzťahy ani dátový model.

3. Povinné časti riešenia

- **Power Query:** načítanie dát, kontrola dátových typov a aspoň niekoľko zmysluplných transformačných krokov; napr. odstránenie chýb/duplicít, úprava textu, rozdelenie alebo zlúčenie stĺpcov, odvodenie údajov, merge/append podľa charakteru projektu.
- **Dátový model:** logický model s korektnými vzťahmi a kardinalitou; preferovaná je hviezdicová schéma, ak je pre dané dáta vhodná. Mali by ste vedieť vysvetliť, čo je faktová tabuľka, čo sú dimenzie a ako sa šíria filtre.
- **Kalendár:** samostatná dátumová tabuľka, ak je potrebná na časovú analýzu, a korektné prepojenie s modelom.
- **DAX:** minimálne 6 vlastných measures. Aspoň časť z nich má ísť nad rámec jednoduchého SUM/COUNT; podľa témy napr. percentuálny podiel, priemer, marža, medziročná zmena, kumulatívna hodnota alebo porovnanie s predchádzajúcim obdobím.
- **Report:** minimálne 2 obsahovo odlišné reportové stránky. Vizuály musia mať analytický význam, čitateľné názvy a primerané formátovanie.
- **Interaktivita:** slicery/filtre a zmysluplné interakcie medzi vizuálmi; podľa vhodnosti drill-down, drill-through, tooltip alebo iný interaktívny prvok.
- **Interpretácia:** aspoň 4 konkrétne zistenia vyplývajúce z analýzy. Nestačí opis typu „najvyšší stĺpec je X“; záver má vysvetľovať význam výsledku.

4. Hodnotiaca tabuľka – 100 bodov

Kritérium	Body	Čo sa hodnotí
Výber, pochopenie a zdokumentovanie dát	**10**	Vhodnosť datasetu, jasný analytický cieľ, zdroj a pochopenie významu údajov.
Power Query a kvalita dát	**20**	Správne dátové typy, čistenie, transformácie, primeraná efektívnosť a zrozumiteľnosť krokov.
Dátový model	**20**	Kvalita modelu, vzťahy, kardinality, fakt/dimenzie, kalendár, správne filtrovanie.
DAX a výpočty	**20**	Správnosť a zmysluplnosť measures, vhodné použitie DAX, overenie výsledkov.
Vizualizácie, UX a interaktivita	**15**	Výber vhodných vizuálov, čitateľnosť, konzistentnosť, filtre a interakcie.
Analytické závery	**10**	Schopnosť interpretovať výsledky a odpovedať na pôvodné analytické otázky.
Dokumentácia a odovzdanie	**5**	Úplnosť, zdroje, stručný popis riešenia a technicky funkčný odovzdaný súbor.

5. Čo sa nebude považovať za kvalitný projekt

- iba import jednej hotovej tabuľky a vytvorenie niekoľkých grafov bez dátového modelu;
- veľký počet vizuálov bez jasnej analytickej otázky;
- DAX measures vytvorené iba formálne bez pochopenia výsledku;
- nevhodné alebo zavádzajúce grafy a dekoratívne prvky na úkor čitateľnosti;
- projekt, pri ktorom nebudete vedieť vysvetliť zdroj dát, transformácie, vzťahy alebo vlastné výpočty;
- odovzdanie cudzieho hotového dashboardu alebo riešenia bez preukázateľnej vlastnej práce.

6. Odovzdanie a krátka obhajoba

Odovzdávané výstupy: Power BI projekt (štandardne PBIX), zdrojové dáta alebo funkčný odkaz na ich zdroj a krátka dokumentácia. Dokumentácia má obsahovať cieľ, zdroj dát, stručný opis modelu, zoznam hlavných measures a 4–6 najdôležitejších zistení.

Obhajoba: stručné predstavenie problému, modelu a reportu, odpovedanie na otázky k Power Query, vzťahom, DAX a interpretácii výsledkov. Cieľom je overiť, že rozumiete vlastnému riešeniu.

7. Kontrolný zoznam pred odovzdaním

☐ Viem vysvetliť, čo predstavuje jeden riadok mojej faktovej tabuľky.

☐ Poznám primárne/jedinečné kľúče dimenzií a rozumiem vzťahom medzi tabuľkami.

☐ Dátové typy a transformácie v Power Query sú správne.

☐ Model neobsahuje zbytočné alebo nejednoznačné vzťahy.

☐ Moje measures vracajú správne výsledky aj pri filtrovaní.

☐ Každý vizuál má jasný účel a vhodný typ grafu.

☐ Report odpovedá na pôvodne stanovené analytické otázky.

☐ Závery sú podložené dátami a viem ich slovne interpretovať.

☐ Zdroj dát je uvedený a projekt neobsahuje nevhodné osobné/citlivé údaje.

☐ Odovzdaný súbor sa otvorí a funguje bez chýb.

Hodnotí sa kvalita analytického riešenia, nie množstvo grafov.