

# Univerzita Jana Evangelisty Purkyně v Ústí nad Labem

---

Přírodovědecká Fakulta

Katedra informatiky

Aplikovaná informatika

Zápočtový projekt z předmětu KI/ODM

## **Analýza prodejních dat jízdních kol pomocí OLAP modelu**

Student: David Král

Akademický rok: 2024/2025

Datum odevzdání: 5. 5. 2025

## 1. Úvod

Projekt řeší zpracování a analýzu rozsáhlé datové sady o prodeji jízdních kol. Cílem bylo vytvořit datový sklad, provést transformaci dat, sestavit hvězdicovou strukturu (fact-dimension model), realizovat analytické dotazy a nahradit data mining metodiku pokročilými SQL dotazy.

## 2. Použité nástroje

- Microsoft SQL Server Management Studio (SSMS)
- Microsoft SQL Server 2022
- CSV dataset: bike\_sales.csv (přes 450 000 záznamů)
- Python pro import a předzpracování dat

## 3. Postup řešení

### a) Import dat

CSV soubor byl naimportován do SQL Serveru přes Python (pandas + pyodbc). Data byla vyčištěna, konvertována a tabulka bike\_sales byla doplněna o primární klíč (ID INT IDENTITY).

### b) Transformace typů

Pomocí ALTER COLUMN a UPDATE byly sloupce převedeny na odpovídající typy (DATE, INT, FLOAT, BIT).

### c) Vytvoření dimenzí a faktové tabulky

Z tabulky bike\_sales\_with\_id byla vytvořena:

- DimDate (datum, den, měsíc, rok)
- DimCustomer (věk, pohlaví, stát, skupina)
- DimProduct (produkt, kategorie, značka, velikost)
- FactSales (tržby, náklady, množství, odkaz na dimenze)

Faktová tabulka byla naplněna pomocí JOIN mezi zdrojovou tabulkou a dimenzemi.

## 4. OLAP dotazy (bod 4b) [DavidKral9/OLAP](#)

Dotaz 1: Tržby podle roku

Dotaz 2: TOP 5 produktů podle zisku

Dotaz 3: Podíl ekologických produktů

Dotaz 4: Průměrný zisk podle věkových skupin

## 5. Pokročilý dotaz (náhrada data miningu – bod 4c)

Segmentace zákazníků podle průměrné hodnoty objednávky (High/Mid/Low):

Pomocí CTE dotazu byly zákazníci rozděleni do segmentů:

- High-Value (> 1000 Kč)
- Mid-Value (500–1000 Kč)
- Low-Value (< 500 Kč)

Dotaz využívá AVG, COUNT, CASE a GROUP BY.

## 6. Závěr

Projekt ukázal praktické použití relačního datového skladu a OLAP přístupů při analýze rozsáhlého datasetu. Byly vytvořeny dimenze, faktová tabulka, výkonné dotazy a pokročilé segmentace, které plně pokrývají požadavky zadání.

## 7. Zdroje

- Microsoft Learn – <https://learn.microsoft.com/sql/t-sql/>
- Dodaný dataset [Global Bike Sales Dataset \(2013-2023\)](#)
- Interní nápověda SQL Server Management Studio