**Customer Personality Analysis**

Analysis of company's ideal customers

<https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis>

Kontext

Vyhlásenie problému

Analýza osobnosti zákazníka je podrobná analýza ideálnych zákazníkov spoločnosti. Pomáha firme lepšie pochopiť svojich zákazníkov a uľahčuje jej úpravu produktov podľa špecifických potrieb, správania a obáv rôznych typov zákazníkov.

Analýza osobnosti zákazníka pomáha firme upraviť jej produkt na základe cieľových zákazníkov z rôznych typov segmentov zákazníkov. Napríklad namiesto toho, aby spoločnosť míňala peniaze na marketing nového produktu každému zákazníkovi v databáze spoločnosti, môže analyzovať, ktorý segment zákazníkov si s najväčšou pravdepodobnosťou produkt kúpi, a potom produkt propagovať iba tomuto konkrétnemu segmentu.

<https://www.kaggle.com/code/alexandrabelskaya/project-1-customer-segmentation-for-retail>

Popis kodu;

Tento kód vykonáva **segmentáciu zákazníkov** na základe ich nákupného správania a demografických údajov. Používa **K-means clustering**, vizualizuje výsledné segmenty pomocou **PCA, radarového grafu a párových grafov**, a porovnáva ich podľa rôznych metrík ako príjem, frekvencia nákupov či míňané sumy. Na záver uloží výsledný dataset so segmentmi do CSV súboru.

ULOHA SIMKO oprava kodu

### Koľko riadkov a stĺpcov má databáza a 2–3 príklady stĺpcov a čo znamenajú

Databáza obsahuje **2240 riadkov** a **29 stĺpcov**.

Príklady stĺpcov:

* **Year\_Birth** – rok narodenia zákazníka,
* **Income** – ročný príjem zákazníka v eurách,
* **MntWines** – suma peňazí minutá zákazníkom na vína.

### Nové stĺpce vytvorené a ich význam

* **Age** – vek zákazníka, vypočítaný ako rozdiel medzi rokom 2025 a rokom narodenia (Year\_Birth).
* **Kids** – celkový počet detí v domácnosti, súčet počtu malých detí (Kidhome) a tínedžerov (Teenhome).
* **TotalPurchases** – celková suma peňazí, ktorú zákazník minul na rôzne kategórie produktov (víno, ovocie, mäso, ryby, sladkosti, zlaté produkty).

### Pozorovania z grafov

* Väčšina zákazníkov je vo veku približne medzi 30 a 60 rokmi, čo naznačuje, že cieľová skupina je dospelá populácia v produktívnom veku.
* Príjmy zákazníkov sa najčastejšie pohybujú medzi 30 000 a 80 000 eur, pričom príjmové rozdelenie má niekoľko výrazných odľahlých hodnôt (outliers).
* Najčastejší počet detí v domácnosti je 1 až 2, čo môže ovplyvniť nákupné správanie a preferencie zákazníkov.

### Uloženie upraveného datasetu

Upravený dataset som uložil do súboru s názvom **customer\_cleaned.csv**. Tento súbor obsahuje vyčistené dáta vrátane nových stĺpcov, ktoré sme vytvorili, a je pripravený na ďalšiu analýzu alebo spracovanie.