

## Analiza vremenskih serija nad podacima prodavnice u Velikoj Britaniji

- **Opis problema**

Nad skupom podataka gde je evidentirana svaka transakcija britanske prodavnice koja se dogodila između 1.12.2010. i 9.12.2011. godine vršiće se analiza vremenskih serija. Ova britanska prodavnica posluje isključivo preko interneta i bavi se prodajom jedinstvenih poklona. Prodavnica takođe posluje sa „wholesalers“ kupcima, odnosno ljudima koji obavljaju kupovine na veliko. Cilj projekta je razumevanje obrasca ponašanja potrošača i uočavanje potencijalnih sezonskih trendova. Nakon što se ova analiza odradi, mogu se osmisлити optimalniji načini za nabavku proizvoda i generalno poslovanje kako bi se uvećao profit prodavnice.

- **Specifikacija rešenja**

- *Podaci<sup>1</sup>:*

Svaka evidentirana transakcija sadrži:

- Broj fakture
    - Kod proizvoda
    - Opis proizvoda
    - Količina kupljenog proizvoda
    - Datum i vreme fakture
    - Jedinična cena
    - Šifra kupca
    - Ime zemlje u kojoj kupac boravi

- *Elementi rešenja:*

- Pretprocesiranje podataka: rešavanje problema potencijalnih nedostajućih vrednosti, grupisanje podataka po kategoriji proizvoda i podela transakcija na dnevnom, nedeljnom ili mesečnom nivou
    - Eksplorativna analiza podataka: vizualizacija podataka, analiza po kategorijama podataka, identifikacija trendova
    - Analiza ponašanja kupaca: pronalažanje kupaca koji obavljaju kupovine na veliko i analiza koliko često obavljaju takve kupovine, pronalaženje redovnih kupaca i onih koji su kupovali u prodavnici samo jednom
    - Predviđanje budućnosti: implementacija modela vremenskih serija za predviđanje budućnosti i njihovo poređenje
    - Analiza uticaja spoljašnjih faktora

- *Tehnologije:*

- Python, biblioteke za mašinsko učenje i vizualizaciju podataka (numpy, pandas, scikit-learn, statsmodels, matplotlib, seaborn)

<sup>1</sup> Link do podataka koji će se koristiti prilikom izrade projekta:

<https://www.kaggle.com/datasets/mashlyn/online-retail-ii-uci/data>

Na gorenavedenom linku se mogu naći i drugi projekti rađeni nad istim skupom podataka kada se otvori zaglavlje „Code“

- *Model:*
  - AR, MA, ARMA, ARIMA

- **Primer gotovog rešenja**

Najsličnije rešenje pronađeno na internetu obuhvata analizu istih podataka i odgovara na pitanja:

- Postoji li trend?
- Koji su najprometniji dani u nedelji što se tiče prodaje proizvoda?
- Koji su popularni proizvodi?
- Kakav je udeo prihoda u različitim zemljama?

Link do ovog rešenja: <https://www.kaggle.com/code/smnezre/online-retail-exploratory-analysis>

<sup>1</sup> Link do podataka koji će se koristiti prilikom izrade projekta:

<https://www.kaggle.com/datasets/mashlyn/online-retail-ii-uci/data>

Na gorenavedenom linku se mogu naći i drugi projekti rađeni nad istim skupom podataka kada se otvori zaglavlje „Code“