

ZZN Projekt 2024/25

Téma 8. Databáze hodnocení aerolinek cestujícími

Úloha 2. Řešení dolovacích úloh

**Tým:**

Adam Hos (vedoucí) – xhosad00

Dominik Pop – xpopdo00

# Popis datové sady

Datová sada obsahuje údaje z průzkumu spokojenosti cestujících letecké společnosti. Shromažďuje demografické informace o cestujících, dále pak detaily o jejich letu a hodnocení různých jednotlivých vlivů před/během letu. Dataset se především zaměřuje na hodnocení služeb jako je pohodlí sedadel, kvality jídla/pití, kvality palubní obsluhy nebo třeba spolehlivosti online rezervace. Dataset se dá použít k analýze, která může letecké společnosti pomoci vylepšit své služby a porozumět lépe preferencím zákazníků.

* Id – Id záznamu
* Gender – pohlaví cestujícího (string – muž / žena)
* Customer Type – loajálnost (string – loajální / neloajální)
* Age – věk (int)
* Type of Travel – záměr cesty (string – byznys /osobní)
* Class – cestovní třída (string – Business / Eco / Eco Plus)
* Flight Distance – délka letu (int)
* Atributy míry uspokojenosti (int – [0-5])
  + Inflight wifi service – wifi připojení
  + Departure/Arrival time convenient – vhodná doba odletu / příletu
  + Ease of Online booking – kvalita online bookingu
  + Gate location – umístění brány
  + Food and drink – spokojenost s jídlem a pitím
  + Online boarding – online boarding (zařízení palubní vstupenky online)
  + Seat comfort – pohodlí sedadel
  + Inflight entertainme – zábava během letu
  + On-board service – obsluha při nalodění
  + Leg room service – prostor pro nohy
  + Baggage handling – zacházení se zavazadly
  + Checkin service – odbavení zavazadel
  + Inflight service – obsluha za letu
  + Cleanliness – čistota letadla
* Departure Delay in Minutes – zpoždění odletu v minutách (int)
* Arrival Delay in Minutes – zpoždění příletu v minutách (int)
* Satisfaction – celková spokojenost s aerolinkou (string – satisfied / neutral or dissatisfied)

# Dolovací úlohy

## Asociace mezi jednotlivými službami a spokojeností

**Popis**: Úloha zaměřená na zjištění, které poskytované služby mají největší vliv na celkovou spokojenost cestujících. V rámci úlohy budeme hledat souvislost mezi kvalitami služeb a celkovou spokojeností.

**Použití**: Letecká společnost může využít výsledky této analýzy ke zlepšení nebo prioritizaci daných důležitých služeb, za účelem zvýšení průměrné spokojenosti cestujících.

**Metody**: Asociační pravidla, Rozhodovací stromy

## Analýza dopadu zpoždění na spokojenost cestujících v závislosti na službách

**Popis**: Cílem úlohy je zjistit, zda některé služby jsou schopny zmírnit negativní vliv zpoždění na celkovou spokojenost cestujícího. V rámci úlohy budeme sledovat, jak různé služby ovlivňují spokojenost u letů se zpožděním.

**Použití**: Letecká společnost může využít výsledky analýzy k optimalizaci služeb u letů, s pravidelným, ale i nečekaným zpožděním.

**Metody**: Regresní analýza, která zkoumá, jak jednotlivé služby ovlivňují vztah mezi zpožděním a spokojeností

## Identifikace klíčových služeb v závislosti na délku letu

**Popis**: Cílem úlohy je zjistit, které služby jsou pro cestující důležitější pro krátké lety, a které naopak pro dlouhé lety.

**Použití**: Letecká společnost může využít výsledky identifikace k přizpůsobení svých služeb k délce letu a dosáhnout tak lepší spokojenosti napříč různými délkami letu.

**Metody**: K-means, Regresní stromy

## Analýza třídy letu v závislosti na věku a pohlaví cestujícího

**Popis**: Tato úloha se zaměřuje na analýzu, jak věk a pohlaví cestujícího ovlivňuje volbu cestovní třídy. Cílem je zjistit, zda existují nějaké demografické vzory v preferencích třídy.

**Použití**: Letecká společnost může využít výsledky analýzy, k vytváření cílených marketingových kampaních.

**Metody**: Klasifikace

Řešení úloh

# Charakteristiky datové sady

V datové sadě je celkově 103436 záznamů. Rozdělení můžu a žen je blízké polovičnímu, kde muži lehce převyšují (50.75%). TODO nevím, jestli to má cenu psát, jestli to vůbec chceme nějak popisovat. Každopádně by to byl výstup ze statistik týkajících se celého datasetu. Byla by to docela jen omáčka, ale mohl by na to přihlížet, že reálně víme o čem kecáme

# Asociace mezi jednotlivými službami a spokojeností

## Popis

Základem této úlohy je prozkoumání korelace jednotlivých služeb s výslednou spokojeností za celý let. Služby jsou ohodnoceny v rozmezí 1-5. Celková spokojenost je binární údaj. Pasažér je buď spokojen, nebo neutrální až nespokojený. Vyhodnocením této úlohy bude náhled, které služby jsou nejdůležitější.

## Předpřipravení dat

Všechny data se nacházeli v jednom souboru .csv, který obsahoval hlavičky atributů. Tedy samotný import do Rapid Mineru byl jednoduchý. Obsahuje ale zbytečnou informaci o číslu řádku (pro identifikování záznamu byl použit ID), které je v každém další úloze také odstraněno.

Dalším krokem je selekce pouze potřebných atributů, což jsou jednotlivé služby, ID záznamu a údaj o celkové spokojenosti. Neexistuje ani jeden záznam, které by měl chybějící hodnoty atributů, tudíž není potřeba žádný doplňovat. A tím že jsou hodnoty v rozmezí 1-5, není zapotřebí ani normalizovat. Pouze bylo převedeno celkové uspokojení z:

* satisfied => 1
* neutral or dissatisfied => 0

## Řešení

Pro výpočet korelace služeb a spokojenosti byla využita korelační matice. Důležitý je sloupec s hodnotou spokojenost. Po filtraci a seřazení vypadá následovně:

|  |  |
| --- | --- |
| **Atributy** | satisfaction |
| **Online boarding** | 0.504 |
| **Inflight entertainment** | 0.398 |
| **Seat comfort** | 0.349 |
| **On-board service** | 0.322 |
| **Leg room service** | 0.313 |
| **Cleanliness** | 0.305 |
| **Inflight wifi service** | 0.284 |
| **Baggage handling** | 0.248 |
| **Inflight service** | 0.245 |
| **Checkin service** | 0.236 |
| **Food and drink** | 0.210 |
| **Ease of Online booking** | 0.172 |
| **Gate location** | 0.001 |
| **Departure/Arrival time convenient** | -0.052 |

## Závěr

# Asociace mezi jednotlivými službami a spokojeností

## Popis

## Řešení

## Závěr

# Asociace mezi jednotlivými službami a spokojeností

## Popis

## Řešení

## Závěr