

Analiza danych z Pearson e-learning

Daria Zadranowicz

1. Cel

Celem badania jest jak najdokładniejsza eksploracja otrzymanych danych poprzez analizę parametrów oraz sprawdzenie, czy istnieją między nimi jakiekolwiek zależności.

2. Realizacja badania

- Populacja: uczestnicy kursu Pearson e-learning.
- Jednostka: konkretne podejście uczestnika do rozwiązania danego unitu.
- Wielkość próby: 81432 podejść
- Liczba parametrów: 7

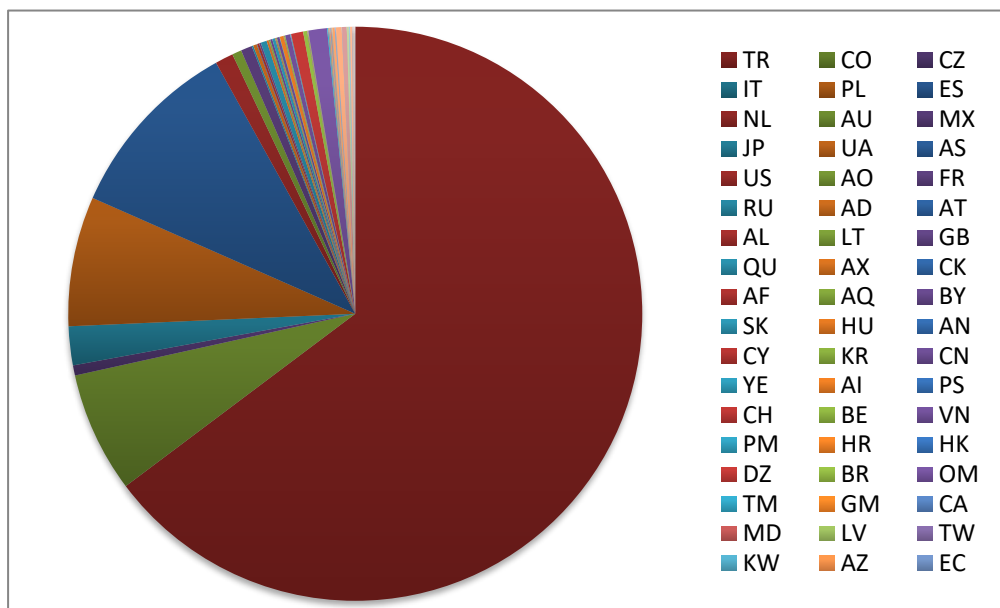
Cechy jakościowe	typ
learner_id	przypisany numer
country	1 spośród 87 krajów
in_course	t - nauka z nauczycielem, f- bez
unit	1-12/REVIEW 1-4/VIDEPODCASTS
Cechy ilościowe	zakres
avg_score	0-3
completion	0,03-1
inv_rate	0-1

Tabela 1. Opis badanych cech

3. Analiza i interpretacja

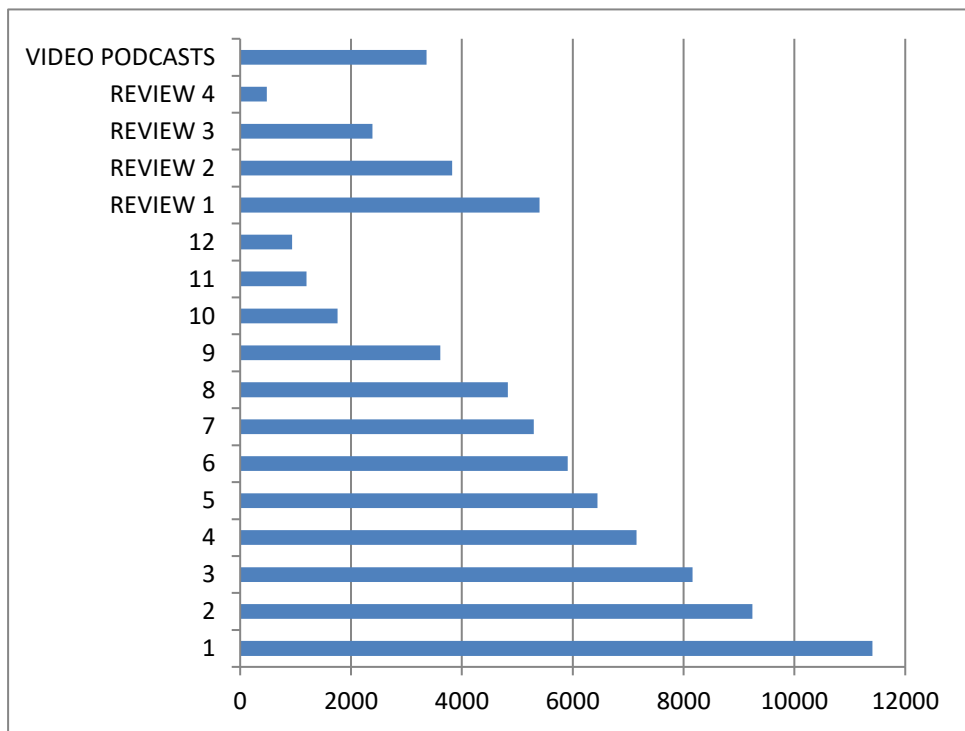
3.1. Analiza ogólna

- Jaki jest udział poszczególnych państw w kursie?



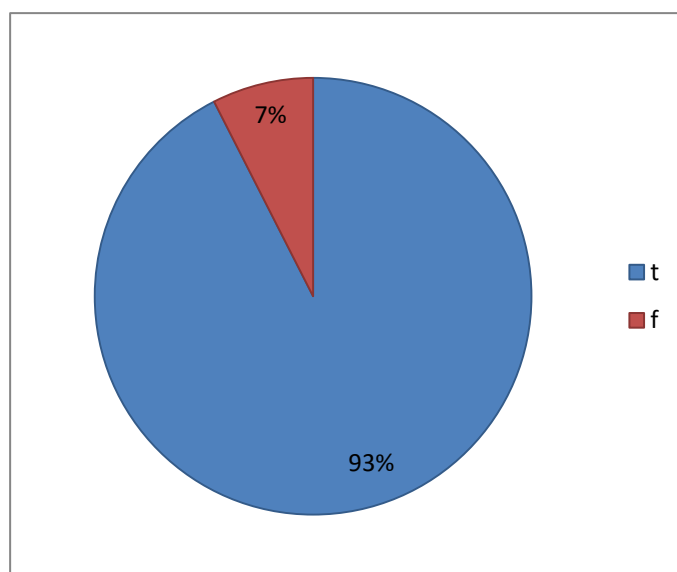
Wykres 1. Udział poszczególnych państw w kursie

- Jak często dany unit był rozwiązywany przez uczestników?



Wykres 2. Częstotliwość rozwiązywania danego rozdziału.

- Który sposób nauki jest preferowany przez uczestników kursu?

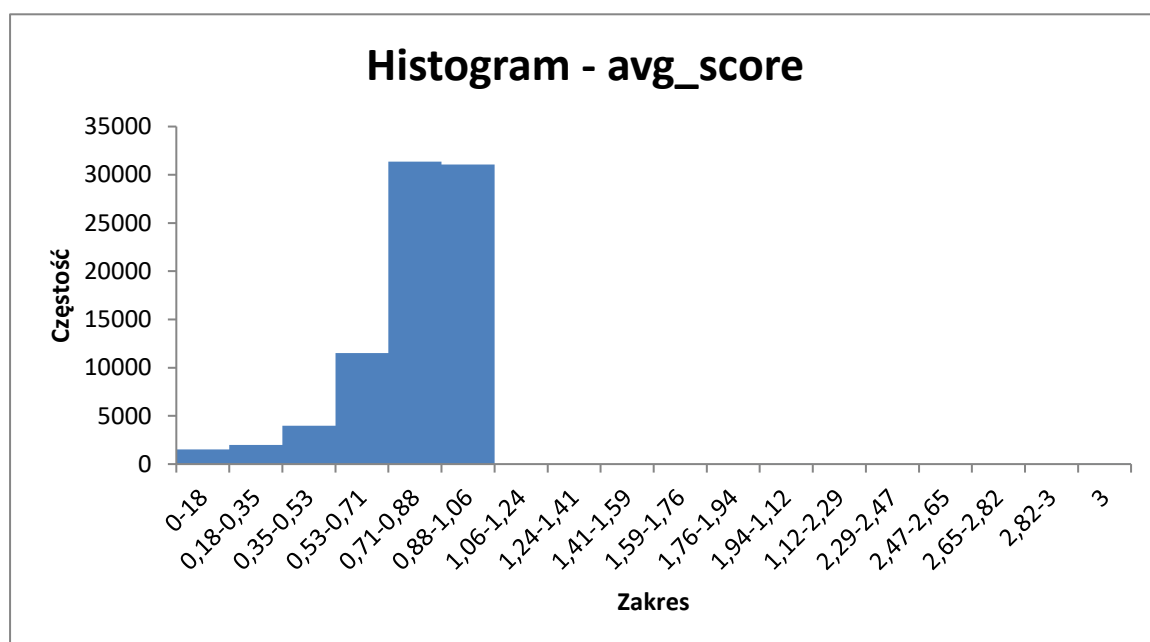


Wykres 3. Preferowany sposób nauki.

3.2. Analiza danych ilościowych:

- Metoda: budowa szeregu rozdzielczego przedziałowego
- Cel: przedstawienie histogramu prezentującego częstotliwość występowania danej cechy

a) Cecha: avg_score



Wykres 4. Histogram prezentujący rozkład cechy avg_score

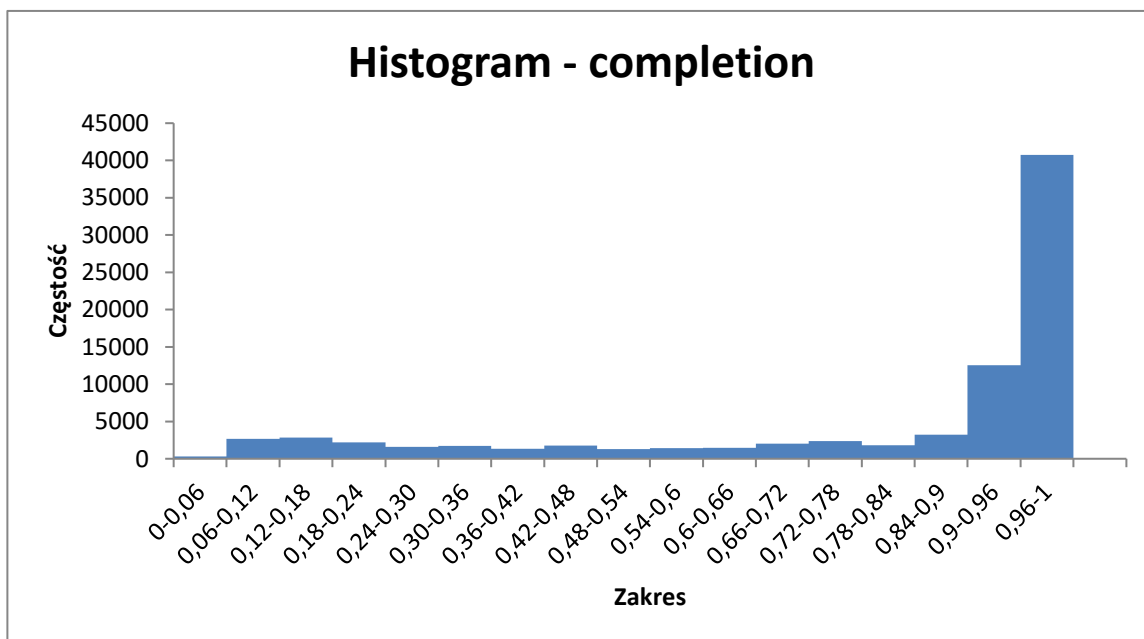
Na podstawie danych posegregowanych za pomocą histogramu można dokonać analizy miar klasycznych, dyspersji, asymetrii, koncentracji (kurtozy).

Miary klasyczne	
średnia	0,79
Miary dyspersji	
wariancja	0,04
odchylenie standardowe	0,2
współczynnik zmienności	0,25
Miary asymetrii	
współczynnik asymetrii	-1,43
Miary koncentracji	
kurtoza	5,26

Tabela 2. Miary – avg_score

Przeciętny poziom badanej cechy to 0,79. Cecha odchyła się od średniej przeciętnie o 0,2. Zróżnicowanie jest umiarkowane (0,25), asymetria (-1,43) lewostronna. Rozkład jest leptokurtyczny, co oznacza wysokie skoncentrowanie wartości wokół średniej.

b) Cecha: completion



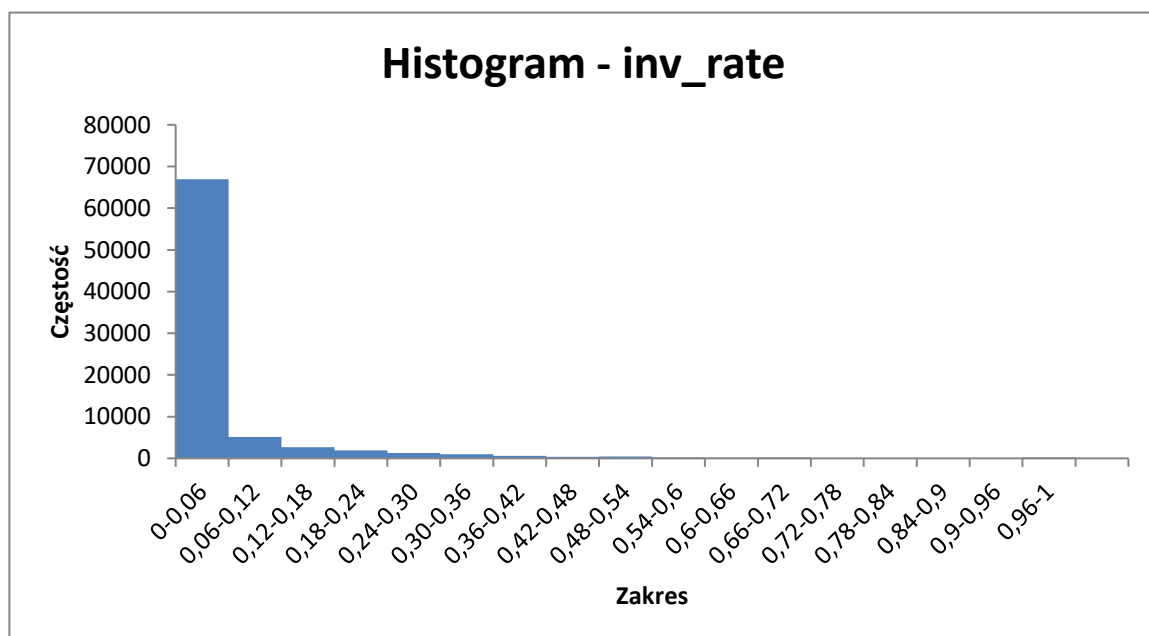
Wykres 5. Histogram prezentujący rozkład cechy completion

Miary klasyczne	
średnia	0,77
Miary dyspersji	
wariancja	0,08
odchylenie standardowe	0,27
współczynnik zmienności	0,36
Miary asymetrii	
współczynnik asymetrii	-1,40
Miary koncentracji	
kurtoza	3,48

Tabela 3. Miary – completion

Przeciętny poziom badanej cechy to 0,77. Cecha odchyła się od średniej przeciętnie o 0,27. Zróżnicowanie jest umiarkowane (0,36), asymetria (-1,40) lewostronna. Rozkład jest leptokurtyczny, co oznacza wysokie skoncentrowanie wartości wokół średniej.

c) Cecha: inv_rate



Wykres 6. Histogram prezentujący rozkład cechy inv_rate

Miary klasyczne	
średnia	0,06
Miary dyspersji	
wariancja	0,01
odchylenie standardowe	0,1
współczynnik zmienności	1,63
Miary asymetrii	
współczynnik asymetrii	4,59
Miary koncentracji	
kurtoza	28,8

Tabela 4. Miary – inv_rate

Cecha inv_rate wykazuje się rozkładem znacznie oddalonym od rozkładu normalnego, przez co miary zaprezentowane w tabeli są trudne do zinterpretowania.

3.3. Analiza danych jakościowych

Jednym ze sposobów przeanalizowania danych jakościowych jest test niezależności χ^2 . Został on przeprowadzony dla dwóch zmiennych: country oraz in_course.

Wynik testu dla 87 zmiennych:	4608,42
Wartość statystyki przy poziomie istotności równym 0,05:	108,648

Tabela 5. Wynik testu niezależności

Otrzymany wynik wskazuje na to, że nie ma podstaw do odrzucenia hipotezy o niezależności zmiennych. Innymi słowy, rodzaj nauczania wybierany przez uczestników kursu nie ma związku z krajem, z którego pochodzą.

4. Podsumowanie i wnioski

Na zaprezentowanych danych można dokonać wiele różnych analiz. Kilka z nich, przedstawionych powyżej, nasuwa następujące wnioski:

- Aż 65% użytkowników to obywatele Turcji.

- Pierwsze rozdziały zostały rozwiązane zdecydowanie więcej razy. Widać tutaj tendencję malejącą, jednak bez zmiennych czasowych dla każdego użytkownika wnioskowanie jest utrudnione.
- Zdecydowana większość uczestników (93%) preferuje rozwiązywanie kursu pod okiem nauczyciela. Być może warto ulepszyć metody pozwalające na samodzielną naukę.
- Jak widać na **Wykresie 4.**, średni wynik punktowy dla wszystkich podejść to 0,77. Może to oznaczać bardzo dobre przygotowanie użytkowników do rozwiązywania zadań.
- Cecha completion wskazuje na wysokie zaangażowanie uczestników w dokończenie przydzielonych zadań.
- Na **Wykresie 6.** Możemy zaobserwować, że odchylenia od wytycznych czynione przez uczestników kursu prawie nie występują, a jeśli tak, to w bardzo niewielkich ilościach.