

Eksploracja danych

Członkowie zespołu:

Piotr Kica
Maciej Zieliński
Michał Żelasko

Temat projektu:

Analiza zbioru danych Uber Pickups.

Opis projektu:

Celem projektu będzie przeanalizowanie danych odnośnie usług wykonywanych przez firmę Uber na obszarze Nowego Jorku w latach 2014-15. W ramach projektu planowane jest znalezienie najbardziej popularnych miejsc i terminów przejazdów (wpływ pory dnia, dnia tygodnia, ewentualnie określonych dni miesiąca, świąt, pór roku etc. na częstotliwość wykonywania usług), porównanie ich z analogicznymi danymi dla innych przewoźników. Dodatkowym celem będzie ustalenie i potencjalne określenie zależności pomiędzy poprzednimi przejazdami klientów i spodziewanym ruchem (oczekujemy, że skoro ktoś jechał z A do B, to istnieje realne prawdopodobieństwo, że będzie wracał z B do A, lub udawał się w jakieś inne miejsce C). Kolejnym elementem może być określenie wpływu faktu, że na dany kierowca operuje tylko na ograniczonym obszarze na efektywność funkcjonowania firmy i liczby wykonanych kursów (na podstawie danych od innych przewoźników).

Wykorzystywany zbiór danych:

<https://www.kaggle.com/datasets/fivethirtyeight/uber-pickups-in-new-york-city>.

Zawiera następujące dane:

- dane odnośnie przejazdów z wykorzystaniem firmy Uber od kwietnia do września 2014 roku (z podziałem na miesiące i dokładnymi danymi odnośnie lokalizacji),
- dane odnośnie przejazdów z wykorzystaniem firmy Uber od stycznia do czerwca 2015 roku (z podziałem na miesiące i dokładność informacji o lokalizacji jest niższa),
- dane dla innych firm o podobnym profilu działalności, zawierające dane odnośnie firmy, terminu wykonania usługi, lokalizacji, ew. kierowcy itd.
- zagregowane dane odnośnie ruchu pojazdów z przedsiębiorstw/firm zajmujących się przewozem na terenie Nowego Jorku (z częściowym uwzględnieniem ruchu klasycznych taksówek).