**Lab 3 Anna Ślęzak**

1. Wykonaj zapytania (3p):
   1. Sprawdź dla każdego dnia tygodnia w jakiej godzinie jest najwięcej branych taksówek.

Obraz zawierający tekst, zrzut ekranu, Czcionka, dokument

Opis wygenerowany automatycznie

* 1. Znajdź przejazdy które znacząco odbiegają od ceny standardowej (outliery)

Obraz zawierający tekst, zrzut ekranu, Czcionka, numer

Opis wygenerowany automatycznie\

* 1. Znajdź średnią cenę przejazdu per osoba dla każdej firmy w zależności od liczby osób (z przedziału 1-7) w samochodzie, zlicz też liczbę przejazdów dla każdej ceny razem z procentowym udziałem we wszystkich przejazdach danej firmy.

Obraz zawierający tekst, Czcionka, zrzut ekranu, algebra

Opis wygenerowany automatycznie

1. Znajdź własny dataset w formacie innym niż Parquet, skonwertuj go do formatu Parquet, a następnie wykonaj własne nietrywialne zapytanie na obu formatach z wykorzystaniem AWS Athena i sprawdź które zapytanie działa szybciej, które procesuje więcej danych. Sprawdź równoważność zwróconych wyników. Do konwersji wystarczą 2 zapytania. (2p) <https://docs.aws.amazon.com/athena/latest/ug/ctas-examples.html#ctas-example-format>

Tabela w formacie csv:

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

Tabela w formacie parquet:

Obraz zawierający tekst, zrzut ekranu, Czcionka, numer

Opis wygenerowany automatycznie

**Wybór najpopularniejszej trasy w danym miesiącu dla określonej linii lotniczej oraz średnia cena biletu dla tej trasy.**

CSV:

Obraz zawierający tekst, zrzut ekranu, numer, Czcionka

Opis wygenerowany automatycznie

Obraz zawierający tekst, numer, Czcionka, linia

Opis wygenerowany automatycznie

PARQUET:

Obraz zawierający tekst, zrzut ekranu, Czcionka, numer

Opis wygenerowany automatycznie

Obraz zawierający tekst, numer, linia, Czcionka

Opis wygenerowany automatycznie