# **Preuzimanje baye NYC Taxi**

<https://docs.microsoft.com/en-us/sql/machine-learning/tutorials/demo-data-nyctaxi-in-sql?view=sql-server-ver15>

# **Dodavanje ID-a u postojeću tabelu**

Alter Table NYCTaxi\_Sample

Add Id\_new Int Identity(1, 1)

Go

# **Analiza baze podataka NYC Taxi**

Pre bilo kakvog normalizovanja ili modelovanja baze ili kreiranja skladišta podataka, potrebno je razumeti bazu, podatke i početne zavisnosti između entiteta. Pošto je baza denormalizovana, odnosno zato što postoji samo jedan entitet, moramo razumeti značenje svih atributa i njihovu povezanost.

|  |  |
| --- | --- |
| Polje | Značenje |
| medallion | Vozač |
| hack\_license | Licenca vozila |
| vendor\_id | Dobavljač od kojeg je dobijena informacija:   * **CMT** – Creative Mobile Technologies * **VTS** – VeriFone, Inc. * **DDS** – Digital Dispatch Systems |
| rate\_code | Tarifa vožnje:   1. Standardna cena 2. JFK (aerodrom) 3. Newark (aerodrom) 4. Nassau ili Westchester 5. Cena po dogovoru 6. Grupna vožnja |
| store\_and\_fwd\_flag | Da li je vožnja naknadno dostavljena serveru zbog problema u konekciji (N / Y) |
| pickup\_datetime | Datum i vreme početka vožnje |
| dropoff\_datetime | Datum i vreme završetka vožnje |
| passenger\_count | Broj putnika |
| trip\_time\_in\_secs | Trajanje vožnje u sekundama |
| trip\_distance | Dužina pređenog puta u miljama (informacija dobijena od taksimetra) |
| pickup\_longitude | Geografska dužina početka vožnje |
| pickup\_latitude | Geografska širina početka vožnje |
| dropoff\_longitude | Geografska dužina završetka vožnje |
| dropoff\_latitude | Geografska širina završetka vožnje |
| payment\_type | Način plaćanja:   * **CRD** – kreditna kartica * **CSH** – u kešu * **NOC** (no charge) – bez naplate * **DIS** (dispute) – spor oko naplate * **UNK** (unknown) – nije poznato |
| fare\_amount | Iznos vožnje izračunat od strane taksimetra na osnovu dužine pređenog puta i utrošenog vremena |
| surcharge | Totalna suma dodatnih taksi i putarina (MTA\_Tax, Tolls\_Amount) |
| mta\_tax | Automatski dodat „Porez na gradski prevoz putnika“ (*Metropolitan Commuter Transportation Mobility Tax*) na osnovu tarife |
| tolls\_amount | Putarina za korišćenje mostova ili tunela |
| total\_amount | Totalna suma naplaćena od putnika, ne uključuje bakšiš ostavljen u kešu |
| tip\_amount | Bakšiš ostavljen isključivo kreditnom karticom |
| tipped | Da li je ostavljen bakšiš (0 / 1) |
| tip\_class | Klasa kojoj bakšiš pripada |