Univerzitet u Nišu  
Elektronski fakultet  
Katedra za računarstvo

Aleksandar Stamenković

**Interna statistika koju Microsoft SQL Server održava**

**Seminarski rad**

**Master akademske studije**

**Predmet: Sistemi za upravljanje bazama podataka**

Smer: Računarstvo i informatika  
Modul: Bezbednost računarskih sistema

Student:

Aleksandar Stamenković  
Br. indeksa: 1403

Niš, april 2022. god.

**Sadržaj**

[1 Uvod 3](#_Toc101983282)

[2 Kreiranje statistike 3](#_Toc101983283)

[2.1 Histogram u statistici 4](#_Toc101983284)

[3 Kreiranje tabele i popunjavanje podacima 4](#_Toc101983285)

[4 Korišćenje SQL Server statistike 5](#_Toc101983286)

[5 Automatsko ažuriranje statistike 7](#_Toc101983287)

[6 Korisnički definisane statistike 9](#_Toc101983288)

[6.1 Kreiranje statistike 9](#_Toc101983289)

[6.2 Modifikovanje statistike 9](#_Toc101983290)

[6.3 Brisanje statistike 10](#_Toc101983291)

[7 Zaključak 10](#_Toc101983292)

[8 Literatura 11](#_Toc101983293)

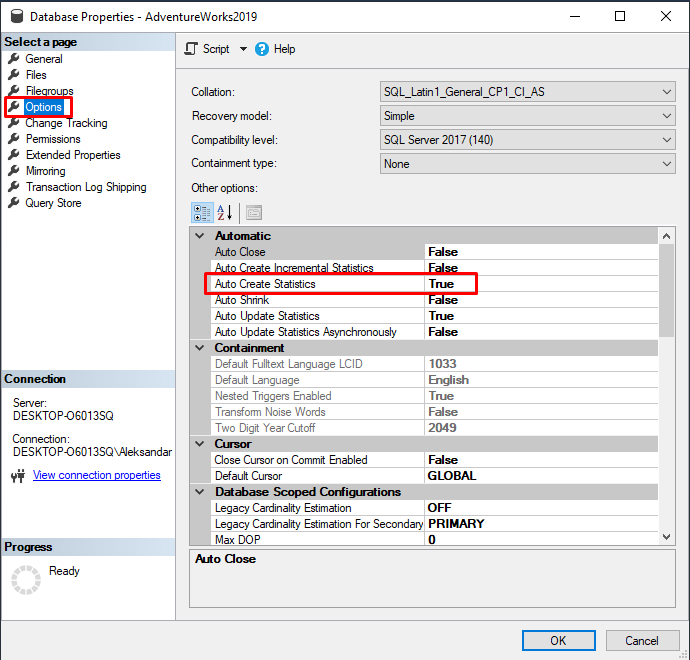
# Uvod

Statistika koju Microsoft SQL Server održava je jedna od glavnih stvari za optimizatora upita koji se izvršava za vreme generisanja plana upita. Statistika se koristi od strane optimizatora da bi procenila koliko redova će upit vratiti pa na taj način izračunava cenu plana upita koristeći tu procenu. Zahtevi CPU, U/I i memorije se kreiraju na osnovu ovih procena, tako da precizna i ažurirana statistika igra važnu ulogu u kreiranju efikasnijih planova upita. Sem toga, preciznost statistike pomaže u odgovarajućim planiranjem resursa za upite. Tako da može se reći da statistika igra ključnu ulogu u performansama izvršenja upita.

Kao klijent preko koga će se vršiti konekcija na bazu, koristiće se SQL Server Management Studio, a baza koja će se koristiti u svrhu demonstracije je opšte poznata baza Adventure works koja se nalazi na sledećem linku: <https://docs.microsoft.com/en-us/sql/samples/adventureworks-install-configure?view=sql-server-ver15&tabs=ssms>.

# Kreiranje statistike

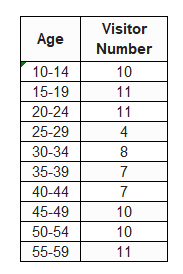
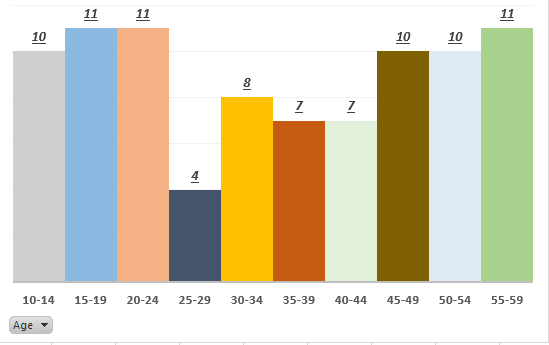
SQL Server automatski kreira statistiku kada se kreira indeks nad tabelom ili indeksirani pogled. Istovremeno, optimizator upita može odlučiti da li će da kreira statistiku za neindeksirane individualne kolone za vreme izvršenja upita. Kako god, ova opcija može da radi ukoliko omogućeno automatsko kreiranje statistike (eng. Auto Create Statistics). Opcija za automatsko kreiranje statistike se može naći pod tabom Options u svojstvima konkretne baze podataka, kao što je prikazano na sledećoj slici:



Slika 1 Automatsko kreiranje statistike

## Histogram u statistici

Histogram je klasifikacija vrednosti u grupi podataka. Histogram omogućava da se jasno vide učestanosti distribucije skupa podataka. Na primer, na sledećoj slici je prikazan histogram koji je kreiran na osnovu raspodele godina starosti među ljudima koji posećuje neki muzej u jednom danu. Histogrami se generalno koriste radi vizualizacije podataka.

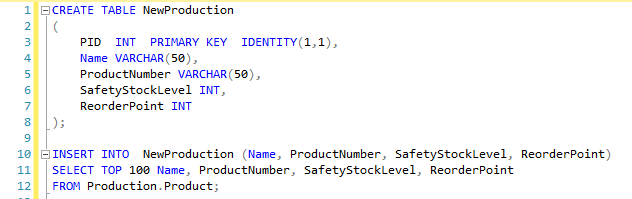
 

Slika 2 Histogram

SQL server statistika se skladišti kao distribucija vrednosti kolona u histogram i čuva jedinstvene vrednosti kao vektor gustine. Ova dva meta podatka koristi optimizator upita da bi izračunao koliko redova će upit vratiti. Razumevanje koncepta statistike SQL Servera će biti mnogo jasnije uz primere. Tako da, kao što je rečeno i u uvodu ovog rada, u nastavku će biti korišćena opšte pozana baza podataka Adventure works.

# Kreiranje tabele i popunjavanje podacima

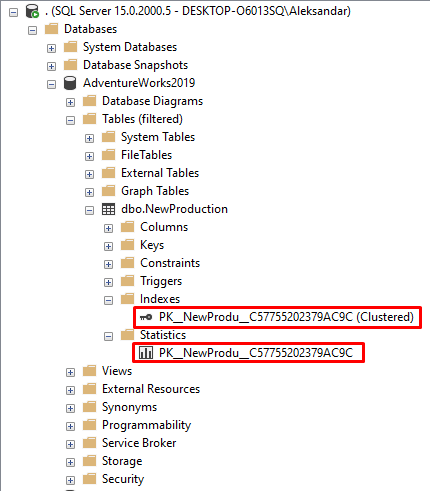
Konkreno će biti korišćena tabela Product iz ove baze podataka. Koristićemo sledeći upit da bismo napravili novu tabelu NewProduction i napunili je podacima:



Slika 3 SQL upit za kreiranje tabele i njeno popunjavanje

# Korišćenje SQL Server statistike

Kao što je već rečeno, nakon kreiranja indeksa, SQL Server automatski kreira statistiku za taj indeks. Na osnovu ovoga, možemo očekivati da SQL Server kreira statistiku za primarni ključ kreirane tabele. Podrazmevano, kada se kreira ograničenje primarnog ključa, SQL Sever automatski kreira jedinstveni klasterovani indeks. Koristeći SQL Server Management Studio, možemo videti SQL Server statistiku pod folderom sa tabelama (eng. Tables) kao što je prikazano na sledećoj slici:

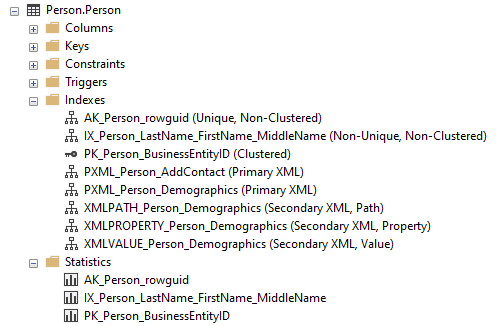


Slika 4 Indeksi i statistike tabele dbo.NewProduction

Komanda koja se koristi da bi se prikazala statistika je sledeća:

DBCC SHOW\_STATISTICS ('Object\_Name', 'Statistic\_Name')

Pogledajmo najpre statistiku za već postojeću tabelu, recimo neka je to tabela Person.Person. Ona ima sledeće kreirane indekse i njohove odgovarajuće statistike:

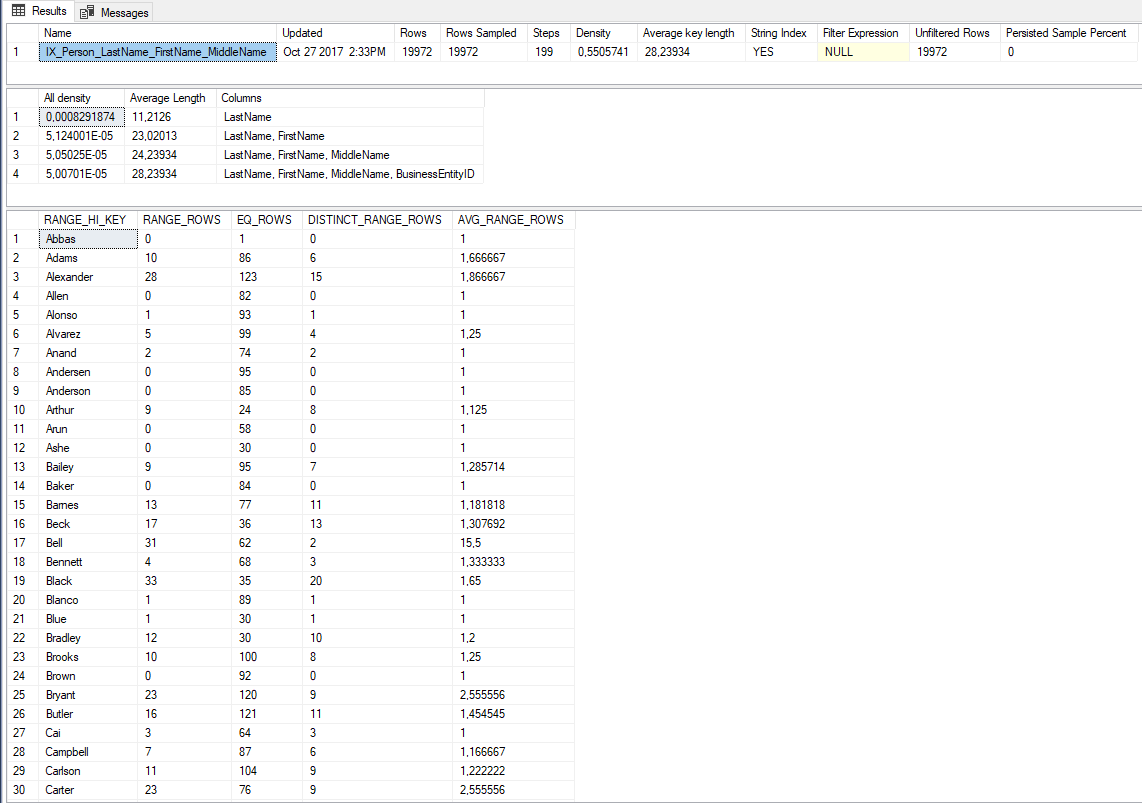


Slika 5 Indeksi i statistike tabele Person.Person

Da bismo pogledali statistiku nad indeksom za ime, prezime i srednje ime iskoristićemo sledeću komandu:

DBCC SHOW\_STATISTICS ('Person.Person','IX\_Person\_LastName\_FirstName\_MiddleName')

Nakon čega dobijamo sledeći izlaz:



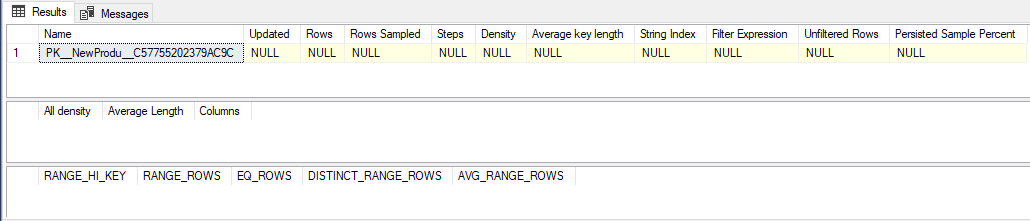
Slika 6 Izgled statistike

Izlaz koji se dobija se sastoji iz 3 dela, a to su redom s vrha naniže:

* Zaglavlje statistike (eng. Statistic header)
* Vektor gustine (eng. Density vector)
* Histogram

Ukoliko bismo želeli da pogledamo statistiku za našu novokreiranu tabelu dobili bismo prazan izlaz, kao što je to prikazano na slici ispod (iako tabela već ima podatke):

**Komanda**: DBCC SHOW\_STATISTICS ('dbo.NewProduction','PK\_\_NewProdu\_\_C57755202379AC9C')

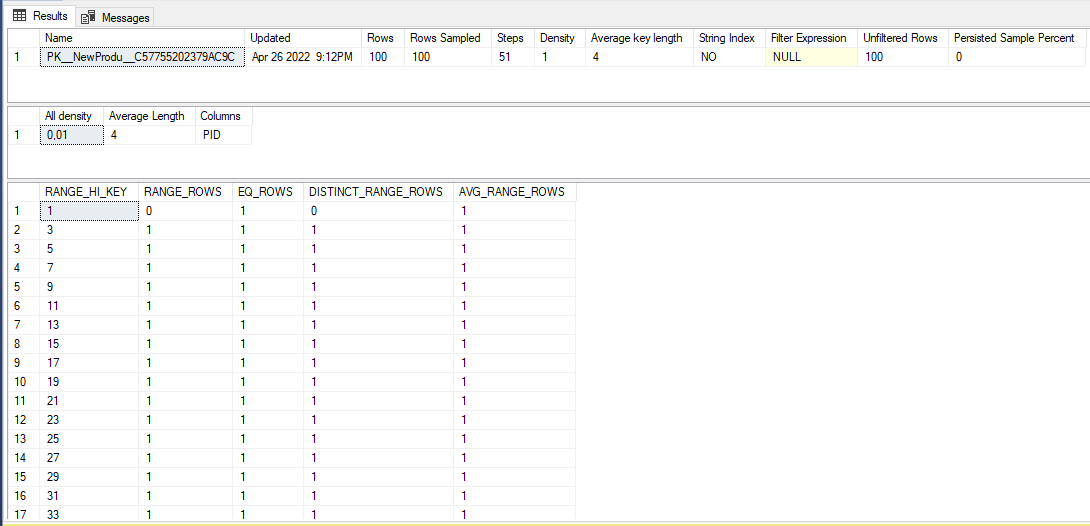


Slika 7 Prazna statistika

Statistički podaci će biti ažurirani kada se izvrši upit koji zahteva statističke podatke za klasterovani indeks. Primer takvog upita je:

SELECT \* FROM NewProduction WHERE PID > 24

Dakle, dovoljno je u where klauzuli navesti kolonu koja predstavlja podatak nad kojim je kreirana statistika. Nakon ovoga možemo da pogledamo statistiku:



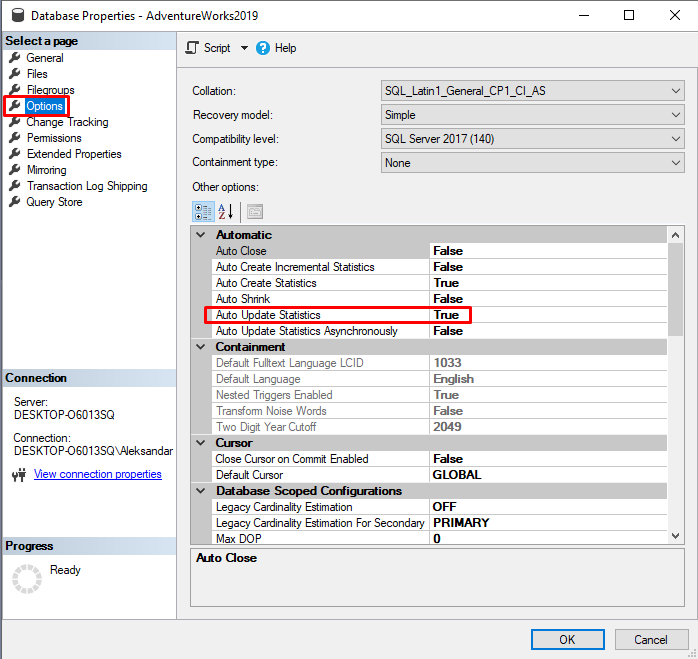
Slika 8 Statistika novokreirane tabele dbo.NewProduction

S druge strane, SQL Server statistika može postati zastarela nakon neke modifikacije (insert, update, delete) tabele zato što će se distribucija vrednosti kolona promeniti. U ovom slučaju, optimizator upita proverava brojač modifikacija i ukoliko on pređe vredost praga, statistika će se ažurirati. Nakon ove operacije, optimizator generiše novi plan upita zbog toga što je moguće da se povećala šansa da će se generisati efikasniji plan upita sa ažuriranom statistikom. Optimizator upita odlučuje da li će ažurirati statistiku na osnovu sledećih pravila:

* Broj vrsta u tabeli je prešao sa 0 na broj veći od 0,
* Tabela je imala 500 ili manje redova kada je statistika poslednji put bila uzorkovana i od tada ima više od 500 modifikacija,
* Tabela je imala više od 500 redova kada se statistika poslednji put ažurirala i broj modifikacija redova je veći od nakon što je statistika poslednji put uzorkovana.

# Automatsko ažuriranje statistike

Opcija automatskog ažuriranja statistike se može naći pod tabom Options u svojstvima baze podataka., kao što je prikazano na sledećoj slici:



Slika 9 Automatsko ažuriranje statistike

Sada ćemo da pokušamo da ubacimo 500.000 novih redova u našoj tabeli, a onda ćemo izvršiti isti upit. Tako da očekujemo da će se statistika ažurirati.

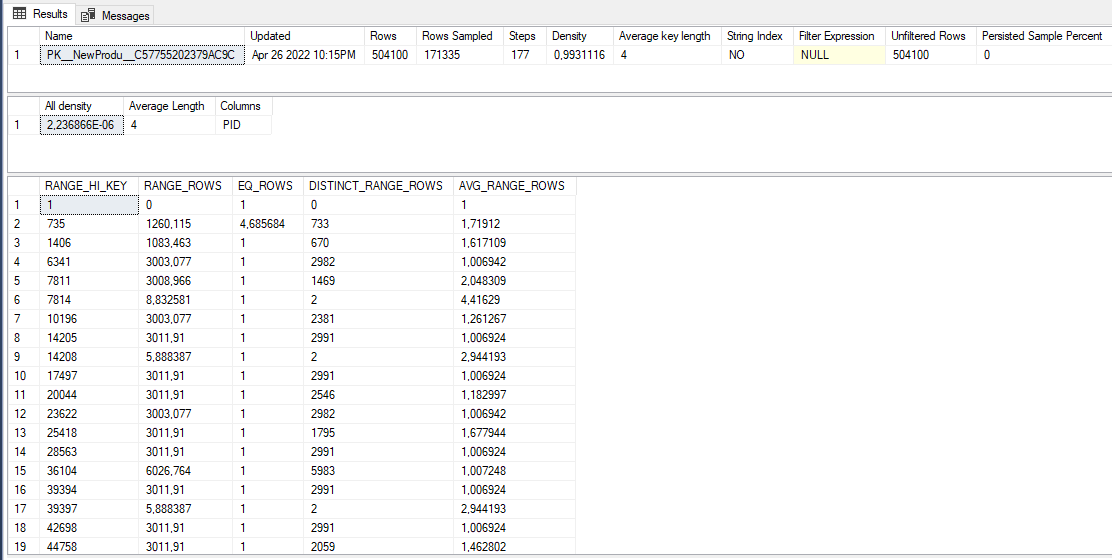
INSERT INTO  NewProduction ( Name,ProductNumber,SafetyStockLevel,ReorderPoint)

SELECT    Name,ProductNumber,SafetyStockLevel,ReorderPoint FROM Production.Product

GO 1000

SELECT \* FROM NewProduction WHERE PID > 24

Kao što se sa sledeće slike može videti, statistika je ažurirana (u zaglavlju statistike se povećala vrednost za Rows i Rows Sampled):



Slika 10 Izgled statistike nakon dodavanja podataka

# Korisnički definisane statistike

## Kreiranje statistike

Osim što SQL Server sam može kreirati statistike, moguće je i definisanje statistika od strane korisnika. Postoji poseban SQL upit kojim se to može uraditi. Sintaksa je sledeća:

CREATE STATISTICS <ImeStatistike>

ON Person.Person (<ImeKolone> [,<ImeKolone>])

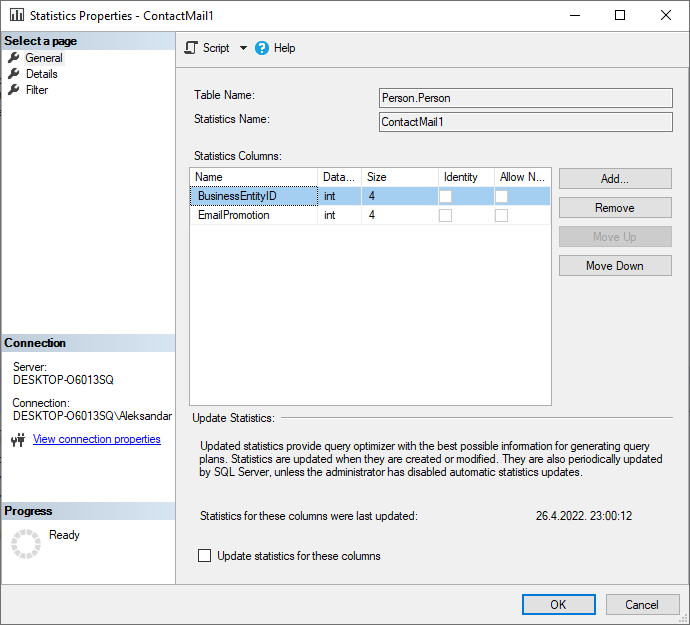
Na primer, napravićemo statistiku nad tabelom Person.Person na sledeći način:

CREATE STATISTICS ContactMail

ON Person.Person (BusinessEntityID, EmailPromotion)

## Modifikovanje statistike

Da bi se modifikovala postojeća statistika neophodno je desnim klikom miša kliknuti na statistiku koja se želi modifikovati, pa zatim se klikne na Properties. Na taj način se otvara sledeći prozor gde je moguće izvršiti modifikaciju:



Slika 11 Modifikovanje statistike

Modifikacija statistie preko SQL-a nije moguća. Alternativni pristup je da se najpre obriše statistika koja se želi pormeniti i zatim se kreira nova.

## Brisanje statistike

Sintaksa za brisanje postojeće statistike je sledeća:

DROP STATISTICS <ImeTabele>.<ImeStatistike> [,<ImeTabele>.<ImeStatistike>]

Primer brisanja kreirane statistike je:

DROP STATISTICS Person.Person.ContactMail1

# Zaključak

Optimizator upita koristi statistiku da bi napravio plan upita (eng. query plan) koji će poboljšati performanse upita. Za najveći broj upita optimizator već generiše neophodnu statistiku za visoko kvalitetan plan upita. Međutim, u nekim slučajevima poželjno je kreiranje dodatne statistike ili izmenu upita da bi se dobili najbolji rezultati.

Statistika za optimizator predstavlja velike binarne objekte (eng. Binary large object – BLOB) koji sadrže informacije o distribuciji vrednosti u jednoj ili više kolona tabele ili indeksiranog pogleda. Optimizator koristi ovi statistiku da bi procenio kardinalnost (broj redova) u rezultatu upita. Kardinalnost omogućava optimizatoru da kreira upite visokog kvaliteta. Tako na primer, u zavisnosti od predikata koji se koriste u upitu, optimizator može iskoristiti kardinalnost i izabrati indeks seek operator umesto indeks scan operatora koji je zahtevniji u pogledu resursa.

# Literatura

[1] Microsoft Docs Statistics <https://docs.microsoft.com/en-us/sql/relational-databases/statistics/statistics?view=sql-server-ver15>

[2] AdventureWorks sample databases <https://docs.microsoft.com/en-us/sql/samples/adventureworks-install-configure?view=sql-server-ver15&tabs=ssms>

[3] Fundamentals of SQL Server Statistics <https://www.sqlshack.com/fundamentals-of-sql-server-statistics/>

[4] Deep Dive into SQL Server Statistics <https://www.sqlshack.com/deep-dive-into-sql-server-statistics/>