

# Druga domaća zadaća iz Podatkovnog inženjerstva

Vaš je zadatak povezati se s bazom podataka i stvoriti nekoliko grafikona na temelju podataka koji se nalaze u bazi, prikazujući ih na "nadzornu ploču" t.j. na dashboardu.

## Baze Podataka

Za zadatak baze podataka i pohranjivanja radit ćemo s PostgreSQL-om, točnije koristeći njegovo proširenje TimescaleDB jer je ova baza optimizirana za pohranu podataka vremenskih serija.

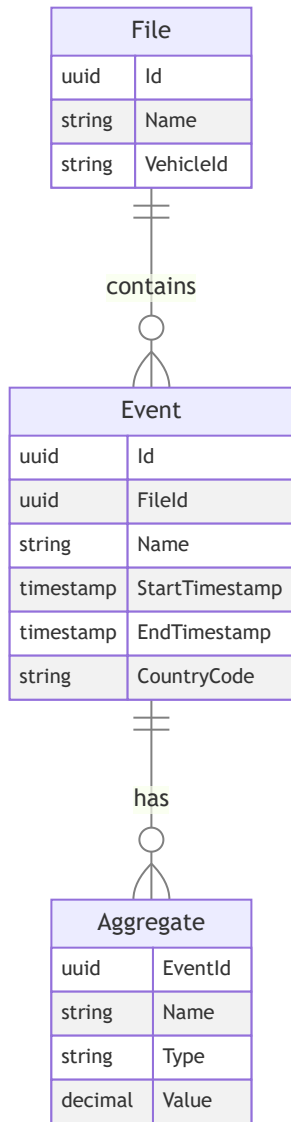
### Sadržaj:

- [Podatci](#)
- [Baza Podataka](#)
  - [Povezivanje na bazu](#)
- [Apache superset](#)
  - [Pokretanje aplikacije](#)
  - [Rad s dockeriziranim Supersedom](#)
  - [Dodavanje nove baze podataka](#)
  - [Stvaranje prvog upita \(query\)](#)
- [Dashboard u Supersetu \(zadatci\)](#)

## Podatci

Podaci koje ste obradili u prethodnom zadatku mogu se predstaviti ER dijagramom na idućoj stranici:

- Jedna datoteka može sadržavati više događaja,
  - Datoteka je predstavljena svojim jedinstvenim identifikatorom i imenom
- Jedna instanca događaja nalazi se u jednoj datoteci
  - Događaj je predstavljen svojim jedinstvenim identifikatorom, imenom te vremenskim oznakama početka i završetka
    - Vremenske oznake početka i završetka predstavljaju kada je događaj počeo i kada je završio
  - Jedan događaj ima više izračunatih agregacija
    - Događaj mora sadržavati jedinstvenu kombinaciju Naziv agregacije i Tip agregacije, npr. samo jedna agregacija ('SPEED', 'max') moguća je za jedan događaj
- Jedna agregacija povezana je s jednim događajem
  - Agregacija je definirana svojim imenom, tipom i vrijednošću



## Baza Podataka

Model baze podataka s kojim ćemo raditi prikazan je u odjeljku [Podatci](#), zajedno s načinom na koji ćemo stvoriti tablice.

### Pristupni podaci za povezivanje na bazu podataka (credentials)

- **host** -> db-pmf-lab.postgres.database.azure.com
- **port** -> 5432
- **database** -> pmf\_lab\_db
- **schema** -> public
- **username** -> pmf\_lab\_reader
- **password** -> S0oIdwSI7k
- **display name** -> Možete sami odabrati, ovo je prikazano ime za Superset.

# Apache Superset

[Službena web stranica](#) za Apache Superset sadrži opću dokumentaciju i smjernice o tome kako raditi s alatom.

## Pokretanje aplikacije

Postoji nekoliko načina na koje možemo instalirati i koristiti Apache Superset, a mi ćemo koristiti Docker sliku jer je jednostavno postaviti je lokalno i pokrenuti je ako je Docker instaliran na računalu.

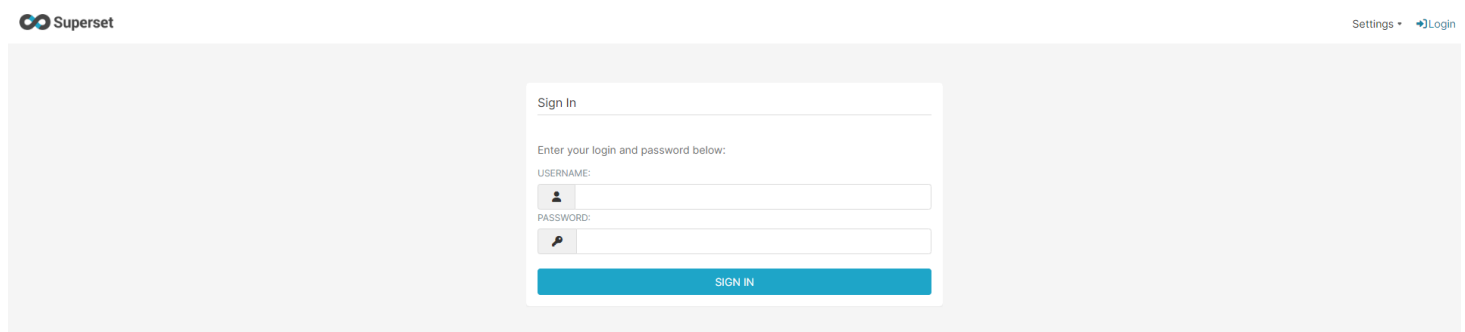
Kako bismo olakšali, pripremili smo skriptu za instalaciju koja će povući sliku i inicijalizirati je s potrebnom konfiguracijom, kao i skripte za pokretanje kontejnera Superseta. Skripta, kao i dodatne informacije, dostupna je u mapi `superset/scripts/local-docker`; za više informacija o procesu instalacije pogledajte datoteku `readme-docker.md`.

Da biste pokrenuli bash skriptu u Unix okruženju izvršavanjem u mapu koja je sadrži, unesite sljedeće:

```
./installation.sh
```

Ako koristite Windows okruženje, umjesto toga možete koristiti drugu skriptu za instalaciju `installation.ps1`.

Kada skripta završi s izvođenjem, možete otvoriti aplikaciju u vašem pregledniku na adresi <http://localhost:8080> i trebala bi se pojaviti stranica za prijavu



Ako niste mijenjali ništa u instalacijskoj skripti, korisnički podaci za prijavu trebali bi biti sljedeći:

- **Korisničko ime** -> admin
- **Lozinka** -> admin

U slučaju da ste promijenili bilo što u instalacijskoj skripti u vezi s donjim odlomkom koda, korisničko ime i lozinka bit će one koje ste postavili ovdje.

```
# execute user creation in the 'superset' container to be able to login in the app
docker exec -it superset superset fab create-admin \
    --username admin \
    --firstname Superset \
    --lastname Admin \
    --email admin@superset.com \
    --password admin
```

Možete također stvoriti dodatnog admin korisnika i koristiti tog umjesto ovoga, vaš je izbor no zadane postavke bi trebale biti u redu jer se koristi na vašem lokalnom računalu.

## Rad s dockeriziranim Supersetom

**⚠ UPOZORENJE** Sva konfiguracija aplikacije koju ova instanca Superseta koristi pohranjena je u bazi podataka koja je uključena u docker sliku i kada uklonite ili zaustavite kontejner vaš će rad također biti uklonjen.

Kako biste ublažili problem slučajnog gubitka vašeg rada, možete koristiti sljedeći pristup nakon početne instalacije docker slike.

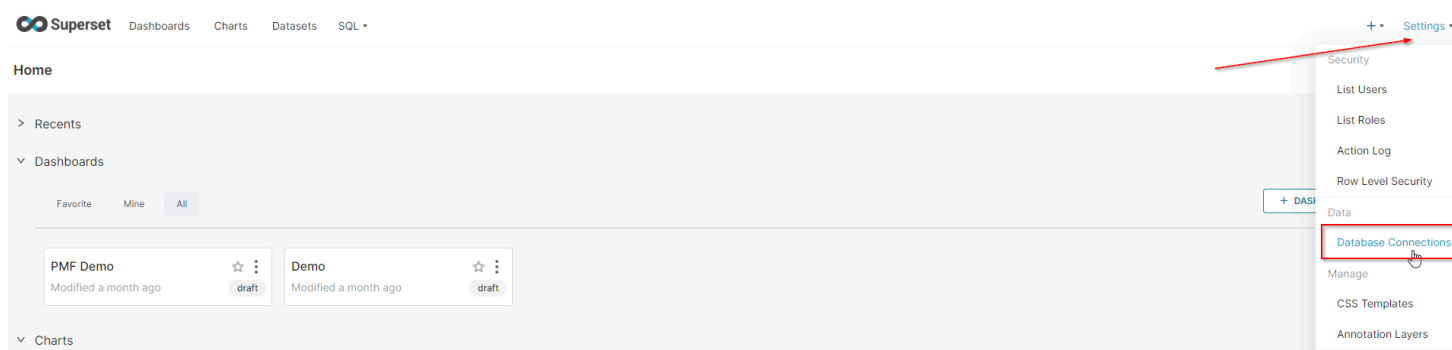
Za pokretanje aplikacije možete koristiti skriptu `start-superset.sh` koja se nalazi u mapi `superset/scripts/local-docker`.

Za zaustavljanje aplikacije možete koristiti skriptu `stop-superset.sh` koja se nalazi u mapi `superset/scripts/local-docker`.

## Dodavanje nove baze podataka

Pristupni podaci za bazu podataka su isti oni koje ste primili u odjeljku [Pristupni podaci za povezivanje na bazu podataka \(credentials\)](#)

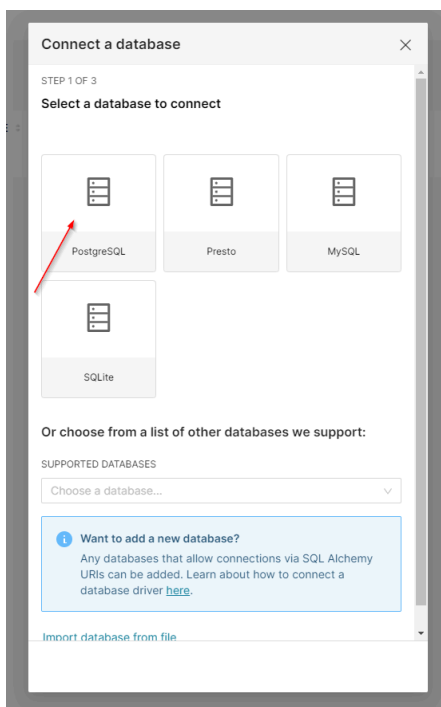
Prije nego što stvorimo i pokrenemo prvi upit (query), moramo konfigurirati vezu s bazom podataka PostgreSQL.



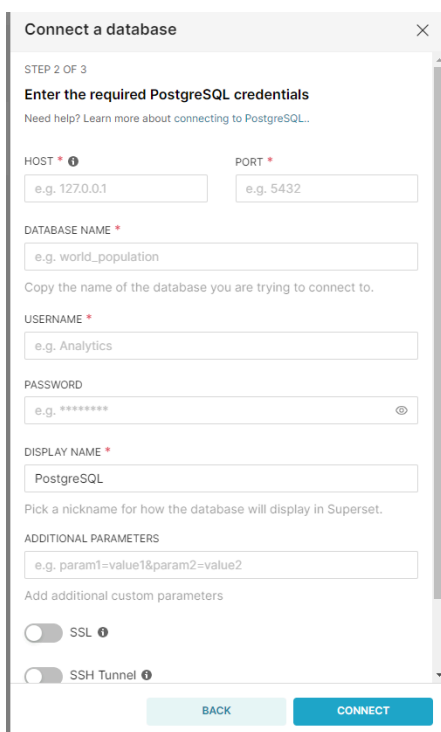
Kliknite na izbornik `Settings` i odaberite stavku `Database Connections` kako biste otvorili stranicu s dostupnim vezama.



Kliknite na gumb **+Database** kako biste otvorili dijalog za postavljanje veze s bazom podataka.



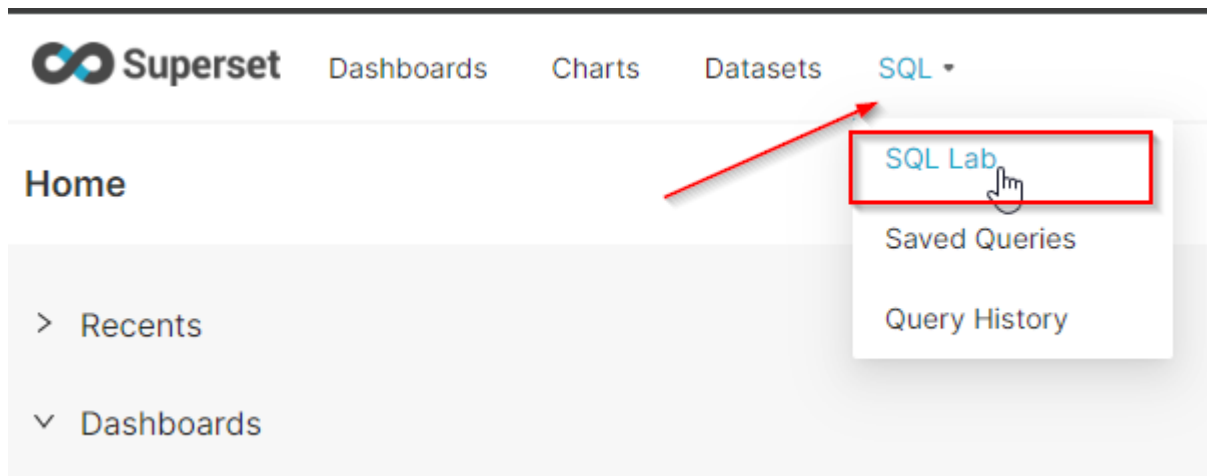
U prozoru odaberite vrstu baze podataka **PostgreSQL**.



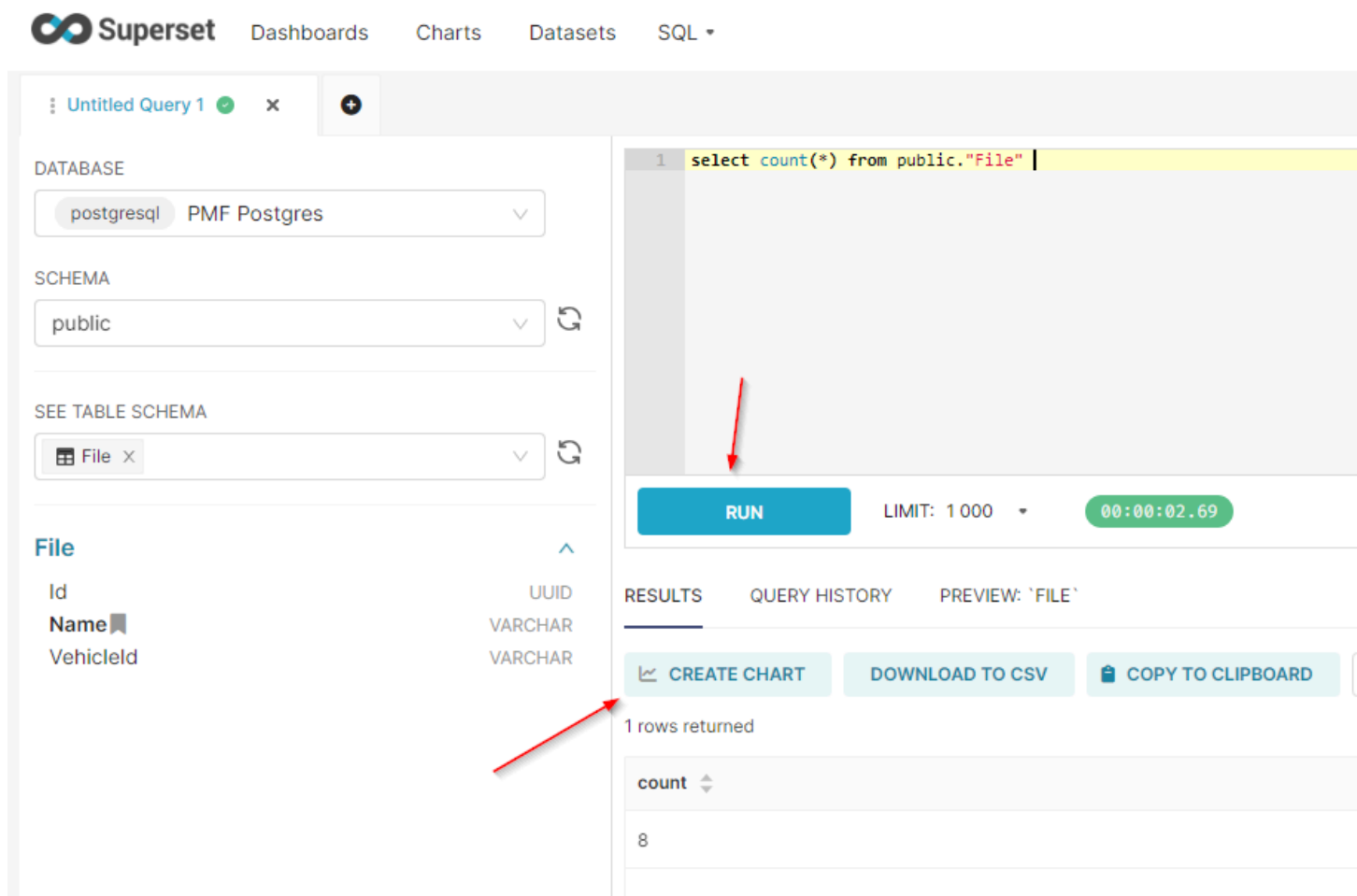
Klikom na **connect** trebalo bi testirati vašu vezu s bazom podataka za vašeg korisnika, na posljednjem koraku nema potrebe za izmjenama, samo potvrdite.

## Stvaranje prvog upita (query)

Dobit će te shemu baze podataka s kojom ćete se povezati s vašim korisnikom. Imat ćete pristup samo toj shemi, u primjerima ćemo koristiti shemu `public`.



Postoje različiti prikazi gdje možete napisati query, ali preporučeni pristup bio bi otići na `SQL Lab` kao u primjeru.



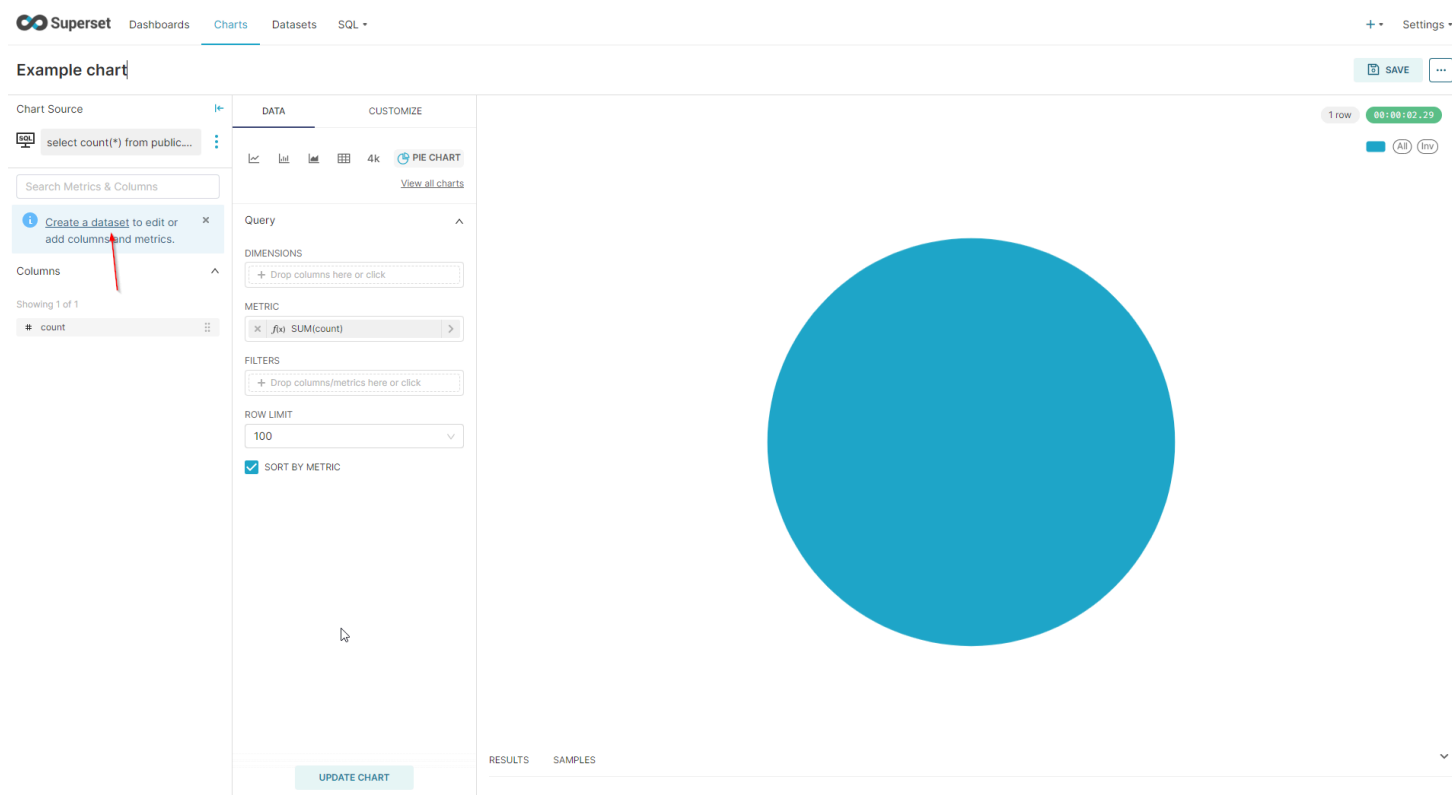
U `SQL Lab` na lijevoj strani nalazi se selektor `Database` i `schema` gdje biste trebali odabrati vezu s bazom podataka koju ste prethodno stvorili. Padajući izbornik `See Table Schema` mogao bi biti koristan

ako želite dodatno pregledati stvorene tablice ili ih koristiti kao podsjetnik na podatke koje sadrže.

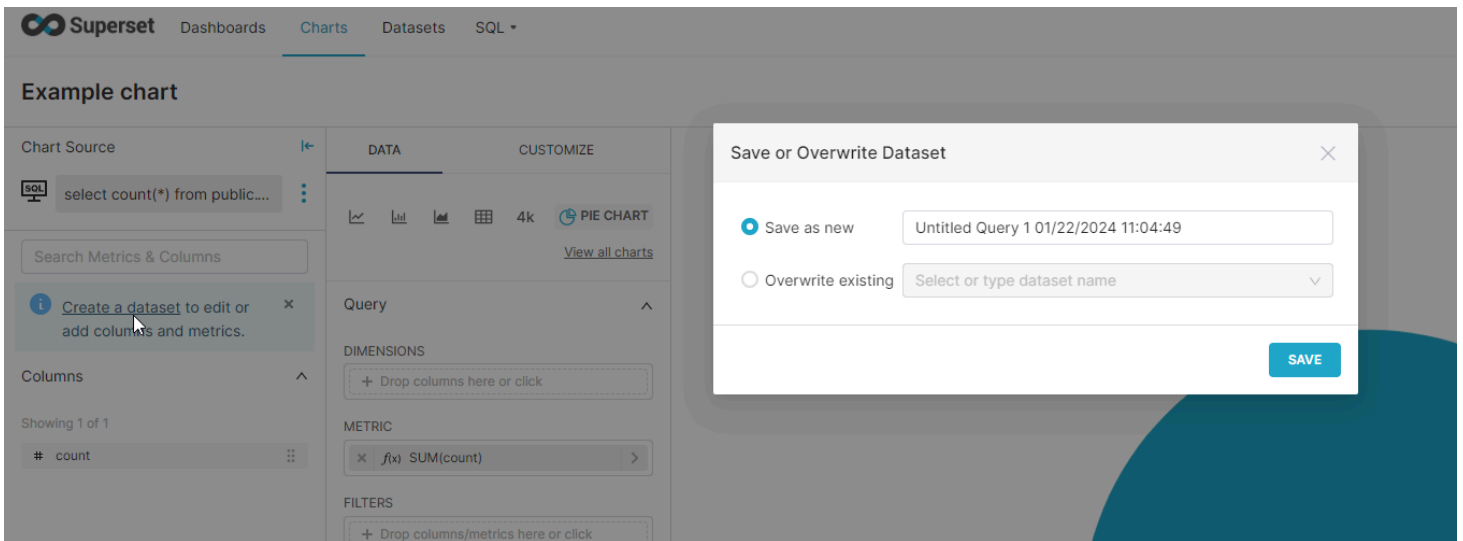
U dijelu urednika možete napisati testni query koristeći shemu koju ste dobili. Uspoređujući s primjerom, vaš upit bi bio:

```
select count(*) from pmf_lab."File"
```

Za pokretanje querya kliknite na gumb `Run`, a nakon što query prođe, trebali biste vidjeti rezultat u tablici ispod. Za stvaranje vašeg prvog grafikona kliknite na gumb `create chart`, što će otvoriti sljedeću stranicu.



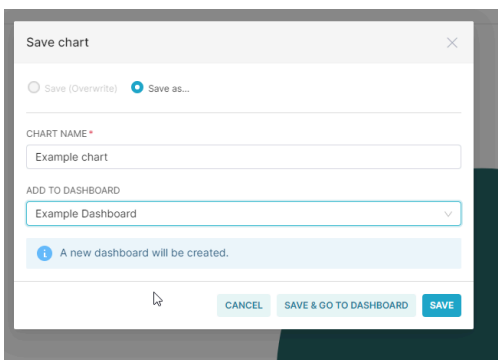
Na ovoj stranici jedna stvar koju bismo preporučili je stvaranje skupa podataka ( `Create a dataset` ), tako da možete urediti upit kasnije ako je potrebno, bez stvaranja cijelog grafikona od početka. Klikom na taj link otvara se sljedeći dijalog.



Imenujte skup podataka kako god želite, pod uvjetom da ima smisla, najvjerojatnije bi naziv trebao odražavati namjeru querya.

Na stranici `chart` postoji mnogo opcija prilagodbe koje možete odabrati, od odabira vrste grafikona do dodatnih agregacija na skupu podataka.

Za spremanje grafikona kliknite na gumb `save` u gornjem desnom kutu, što će otvoriti dodatni dijalog



Ovdje trebate nazvati grafikon i odabrati "Nadzornu ploču" ( `Dashboard` ) na kojoj želite spremiti grafikon. Prilikom prvog stvaranja grafikona, nadzorna ploča možda još ne postoji, pa će aplikacija stvoriti jednu za nas.

## Dashboard u Supersetu (zadatci)

Vaš je zadatak povezati se s bazom podataka koristeći gore navedene upute i stvoriti nekoliko grafikona na temelju podataka koji se nalaze u bazi podataka, prikazujući ih na "nadzornoj ploči" t.j. na dashboardu.

Spremite i prikažite sljedeće grafikone:

1. Prikažite broj događaja po imenu događaja u obliku kružnog dijagrama (pie plot).



2. Prikažite na tabličan način sve događaje čije je trajanje duže od 40 sekundi. Uz informacije o događajima, prikazite izračunato trajanje događaja.
  - Trajanje događaja je razlika u vremenu između završetka i početka događaja.
  - Za usporedbu vremenske razlike koristite [INTERVAL data type](#) npr.  
`timestamp1 - timestamp2 > INTERVAL '3 hours' .`
3. Prikažite na vizualizaciji karte (map visualization) broj događaja koji su se dogodili po zemljama bojenjem zemlje različitim bojama.
4. Prikažite broj događaja po identifikatorima vozila u obliku kružnog dijagrama (pie plot).
5. Prikažite broj događaja u određenim vremenskim intervalima na trakastom dijagramu (bar chart), grupirajući događaje prema imenu događaja. Vremenski intervali koje trebate koristiti su:
  - `'< 30 s'`
  - `'30s - 120s'`
  - `'> 120s'`
6. Prikažite na trakastom dijagramu (bar chart) koliko je događaja započelo u intervalima od jednog sata koristeći funkciju [time\\_bucket](#) iz TimescaleDB.
  - Da biste koristili funkciju `time_bucket`, trebate dodati prefiks `public`, pa bi trebalo biti `public.time_bucket`.
7. Prikažite na tabličan način sve događaje koji imaju agregat (`'SPEED', 'max'`) u rasponu od `[50,60]`.
8. Prikažite vrijednosti agregata (`'T_OIL', 'min'`) u scatter plotu samo za događaje koji imaju agregat2 (`'SPEED', 'max'`) u rasponu od `[50, 60]`, dodatno grupirajući događaje po vremenskom intervalu od jednoga sata i po imenu događaja.
  - Za agregaciju vrijednosti koristite odgovarajuću agregaciju:
    - tip = 'max' -> agregirajte koristeći `'max(a.Value)'`
    - tip = 'min' -> agregirajte koristeći `'min(a.Value)'`
  - X os treba biti vremenski interval po satima, a Y os agregirani agregat.
  - Koristite ime događaja za dodatno kategoriziranje podataka u više grupa na grafikonu, npr. svako ime događaja trebalo bi biti drugačije obojano.
9. Prikažite na tabličan način neke statistike o tome kako su podaci tablice događaja pohranjeni kao hiper-tablica. Za to koristite funkciju [chunks\\_detailed\\_size](#) iz TimescaleDB i vizualizirajte samo stupce `chunk_name` i `table_bytes`, ostali nisu važni ovdje. Budući da stupac `table_bytes` nije stvarno ljudski čitljiv budući da je u bajtima, koristite funkciju [pg\\_size\\_pretty](#) kako biste ga prikazali na čitljiviji način. Poredajte dijelove od najvećeg do najmanjeg, koristeći `table_bytes` za poredanje.
  - Da biste koristili funkciju `chunks_detailed_size`, trebate dodati prefiks `public`, pa bi trebalo biti `public.chunks_detailed_size`.