

# California Traffic Collision Data from SWITRS

## O Bazi

Ovi podaci potiču od kalifornijske autoputne patrola i obuhvataju sudare od 1. januara 2001. do sredine decembra 2021.

Od CHP-a (California Highway Patrol) je tražena baza podataka četiri puta, jednom u 2016, 2017, 2018, 2020, 2021. Kombinovani su skupovi podataka.

Postoje tri glavne tabele:

- collisions: Sadrži informacije o sudaru, gde se dogodio, koja su vozila bila uključena.
- parties: Sadrži informacije o grupama ljudi koji su učestvovali u sudaru, uključujući starost, pol i treznost.
- victims: Sadrži informacije o povredama određenih ljudi koji su učestvovali u sudaru.
- Postoji i tabela pod nazivom case\_ids koja je korišćena za spajanje ostalih tabela. To nam govori iz kojeg od pet originalnih skupova podataka dolazi svaki red. (2016, 2017, 2018, 2020, 2021)

# Semantika kolona



# Collisions:

1. **CASE\_ID** - Jedinstveni identifikator saobraćajne nesreće
2. **COLLISION\_DATE** - Datum kada se saobraćajna nesreća dogodila (YYYYMMDD)
3. **COLLISION\_TIME** - Vreme kada se dogodila saobraćajna nesreća
4. **OFFICER\_ID** – Broj značke policajca koji je pisao uviđaj sudara
5. **COUNTY\_LOCATION** - Grad u kome se desio sudar.
6. **PRIMARY\_ROAD** – Ime ulice u kojoj se desio sudar.
7. **SECONDARY\_ROAD** - Ime puta koji se seče sa glavnim putem gde se desio sudar.
8. **DISTANCE** - Distanca između glavnog i sporednog puta gde se desio sudar.
9. **INTERSECTION** - Da li se sudar desio na raskrsnici
10. **WEATHER\_1** - Vremenski uslovi za vreme sudara.
11. **WEATHER\_2** - Dodatni vremenski uslovi za vreme sudara.
12. **TOW\_AWAY** - Oznaka koja označava da li je vozilo odvučeno sa mesta sudara.
13. **PRIMARY\_COLLISION\_FACTOR** - Glavni razlog saobraćajne nesreće
14. **PCF\_VIOL\_CATEGORY** - Glavni razlog saobraćajne nesreće dat iz pravne sekcije.
15. **HIT\_AND\_RUN** – Da li je krivac pobegao sa mesta sudara i ukoliko jeste da li je ubio nekog.
16. **TYPE\_OF\_COLLISION** – Generalni tip kolizije.
17. **MOTOR\_VEHICLE\_INVOLVED\_WITH** - Opisuje šta je, u kombinaciji sa motornim vozilom u transportu, izazvalo prvu povredu ili događaj koji je prouzrokovao štetu, na putu ili van njega.
18. **PEDESTRIAN\_ACTION** – Radnja neposredno pre sudara prvog povređenog ili na drugi način umešanog pešaka.
19. **ROAD\_SURFACE** – Stanje površine put u vreme sudara u uključenoj saobraćajnoj traci.
20. **ROAD\_CONDITION\_1** - Stanje puta u vreme sudara u uključenoj saobraćajnoj traci.
21. **ROAD\_CONDITION\_2** - Drugo stanje puta u vreme sudara u saobraćajnoj traci/trakama.
22. **LIGHTING** - Uslovi osvetljenja za vreme i mesto udesa.
23. **ALCOHOL\_INVOLVED** - Označava da li je u nesreći učestvovala strana koja je pila.

# Parties:

1. CASE\_ID - Jedinstveni identifikator saobraćajne nesreće
2. PARTY\_NUMBER - Broj koji zajedno sa case\_id jedinstveno identifikuje stranu u sudaru.
3. PARTY\_TYPE - Tip učesnika
4. AT\_FAULT - Identifikuje ukoliko krivica učesnika.
5. PARTY\_SEX - Pol učesnika.
6. PARTY\_AGE - Godine učesnika u vreme sudara.
7. PARTY\_SOBRIETY - Da li je učesnik pod dejstvom alkohola.
8. PARTY\_DRUG\_PHYSICAL - Da li je učesnik pod dejstvom narkotika.
9. PARTY\_SAFETY\_EQUIP\_1 - Sigurnosna oprema učesnika.
10. PARTY\_SAFETY\_EQUIP\_2 - Sigurnosna oprema učesnika.
11. CELLPHONE\_IN\_USE - Da li je učesnik u trenutku sudara koristio mobilni telefon.
12. CELLPHONE\_USE\_TYPE – Na koji način je telefon bio korišćen.
13. FINANCIAL\_RESPONSIBILITY – Da li učesnik ima osiguranje i da li je predato, validno ili se može koristiti u ovoj situaciji
14. OAF\_VIOLATION\_CATEGORY - Kategorija faktora koji je doprineo udesu, ali nije bio primarni uzrok udesa.
15. OTHER\_ASSOCIATE\_FACTOR\_1 - Faktor koji je doprineo udesu, ali nije bio primarni uzrok pada.
16. MOVMENT\_PRECEDING\_COLLISION - Dejstvo vozila pre sudara i pre akcije izbegavanja.
17. VEHICLE\_YEAR - Godina modela vozila koje je učesnik vozio.
18. VEHICLE\_MAKE - Marka vozila koju je učesnik vozio.
19. STATEWIDE\_VEHICLE\_TYPE - Tip vozila učesnika prema listi koja se zove „statewide vehicle type“..
20. CHP\_VEHICLE\_TYPE\_TOWING - Tip vozila ili tegljača prema CHP priručniku HPM 110.5 Poglavlje 3, Aneks F Kodovi tipova vozila.
21. PARTY\_RACE - Rasa učesnika

## Victims:

1. CASE\_ID - Jedinstveni identifikator saobraćajne nesreće
2. PARTY\_NUMBER - Broj koji zajedno sa case\_id jedinstveno identifikuje stranu u sudaru.
3. VICTIM\_ROLE- Uloga žrtve.
4. VICTIM\_SEX – Pol žrtve.
5. VICTIM\_AGE - Koliko godina je imala žrtva za vreme nesreće.
6. VICTIM\_DEGREE\_OF\_INJURY – Težina povrede žrtve.
7. VICTIM\_SEATING\_POSITION - Položaj sedenja žrtve.
8. VICTIM\_SAFETY\_EQUIP\_1 - Sigurnosna oprema žrtve.
9. VICTIM\_SAFETY\_EQUIP\_2 - Sigurnosna oprema žrtve.
10. VICTIM\_EJECTED - Označava da li je žrtva izbačena iz vozila.

# Logička šema

# Collisions

```
{
  "_id": {
    "$oid": "665ca6d6cb238445993d51a4"
  },
  "case_id": 81715,
  "officer_id": 11342,
  "county_location": "los angeles",
  "primary_road": "PATTERSON ST",
  "secondary_road": "ATLANTIC AV",
  "distance": 0,
  "intersection": 1,
  "weather_1": "cloudy",
  "weather_2": "windy",
  "tow_away": 1,
  "primary_collision_factor": "vehicle code violation",
  "pcf_violation_category": "speeding",
  "hit_and_run": "not hit and run",
  "type_of_collision": "rear end",
  "motor_vehicle_involved_with": "other motor vehicle",
  "pedestrian_action": "no pedestrian involved",
  "road_surface": "dry",
  "road_condition_1": "normal",
  "road_condition_2": null,
  "lighting": "daylight",
  "collision_date": {
    "$date": "2020-03-14T00:00:00.000Z"
  },
  "collision_time": "07:45:00",
  "process_date": {
    "$date": "2020-06-22T00:00:00.000Z"
  }
}
```

## Victims

```
{
  "_id": {
    "$oid": "665cd81ccb2384459979cbcb6"
  },
  "case_id": 3899457,
  "party_number": 2,
  "victim_role": "passenger",
  "victim_sex": "female",
  "victim_age": 1,
  "victim_degree_of_injury": "no injury",
  "victim_seating_position": "passenger seat 6",
  "victim_safety_equipment_1": "air bag not deployed",
  "victim_safety_equipment_2": "child restraint in vehicle used",
  "victim_ejected": "not ejected"
}
```

## Party

```
{
  "_id": {
    "$oid": "665cbce5cb238445995ced44"
  },
  "id": 1,
  "case_id": 81715,
  "party_number": 1,
  "party_type": "driver",
  "at_fault": 1,
  "party_sex": "female",
  "party_age": 35,
  "party_sobriety": "not applicable",
  "party_drug_physical": "not applicable",
  "party_safety_equipment_1": "air bag deployed",
  "party_safety_equipment_2": "lap/shoulder harness used",
  "cellphone_in_use": 0,
  "cellphone_use_type": "cellphone not in use",
  "oaf_violation_category": null,
  "other_associate_factor_1": "none apparent",
  "movement_preceding_collision": "proceeding straight",
  "vehicle_year": 2007,
  "vehicle_make": "ford",
  "statewide_vehicle_type": "emergency vehicle",
  "chp_vehicle_type_towing": "police car",
  "party_race": "other"
}
```



## Agregacije:

- Koji vremenski uslovi su najsmrtonosniji?
- Koji je najcesci razlog sudara kod skolskih autobusa ?
- Koliko se svake godine povecao procenat udesa kod top 4 marke automobila?
- Koji pol je kriv za najveći broj udesa leti i za koliko posto?
- Koji je procenat alkoholicara koji su imali osiguranje ?
- Koji dan je najsmrtonosniji ?
- Koji tip sudara je najsmrtonosniji ?
- Koliko motorciklista doživi solo saobraćajnu nezgodu (jedini učesnik) godišnje ?
- Za koje vreme treba upozoriti vozace kada je najopasnije voziti i pod kojim osvetljenjem ?
- Prikazati najveći udes koji je zabeležen sa detaljima o svakom ucesniku i povredjenoj osobi.

# Optimizacija

# Bucketing:

```
_id: ObjectId('666c9d2b10e9b388a3beed80')
case_id: "3982899"
victims: Array (4)
  0: Object
    case_id: "3982899"
    party_number: "1"
    victim_role: "passenger"
    victim_sex: "female"
    victim_age: "74.0"
    victim_degree_of_injury: "no injury"
    victim_seating_position: "passenger seat 3"
    victim_safety_equipment_1: "air bag not deployed"
    victim_safety_equipment_2: "lap/shoulder harness used"
    victim_ejected: true
  1: Object
    case_id: "3982899"
    party_number: "1"
    victim_role: "passenger"
    victim_sex: "female"
    victim_age: "83.0"
    victim_degree_of_injury: "no injury"
    victim_seating_position: "passenger seat 6"
    victim_safety_equipment_1: "air bag not deployed"
    victim_safety_equipment_2: "lap/shoulder harness used"
    victim_ejected: true
  2: Object
  3: Object
total_victims: 4
total_lethal: 0
```

## Victims:

- Victims kolekcija u sebi sadrži case\_id i party\_number.
- Grupisanje po case\_id
- Case\_id + Party\_number – unique
- Nove kolone:
  - total\_victims – predstavlja broj žrtava koje su grupisane u taj case\_id.
  - total\_lethal – predstavlja broj žrtava koje nisu preživele nesreću.

```

_id: ObjectId('666c9c4610e9b388a3b08c1b')
case_id: "3899446"
▼ parties: Array (3)
  ▼ 0: Object
    case_id: "3899446"
    party_number: "1"
    party_type: "driver"
    at_fault: true
    party_sex: "male"
    party_age: "36.0"
    party_sobriety: "had not been drinking"
    party_drug_physical: ""
    party_safety_equipment_1: "air bag not dep
    party_safety_equipment_2: "lap/shoulder ha
    financial_responsibility: "proof of insura
    cellphone_in_use: true
    cellphone_use_type: "cellphone not in use"
    oaf_violation_category: ""
    other_associate_factor_1: "none apparent"
    movement_preceding_collision: "proceeding
    vehicle_year: "2004.0"
    vehicle_make: "bmw"
    statewide_vehicle_type: "passenger car"
    chp_vehicle_type_towing: "passenger ca
    party_race: "hispanic"
  ▶ 1: Object
  ▶ 2: Object
total_parties: 3

```

# Parties

- Isti pristup je korišćen i kod parties kolone.
- Grupisanje po case\_id
- Case\_id + Party\_number – unique
- Nova kolona:
  - total\_parties - broj partija vezano za taj slučaj.

# Indexi

## Collisions, Parties, Victims:

➤ case_id_1	REGULAR ⓘ	6.8 MB	82549619 (since Fri Jun 14 2024)
-------------	-----------	--------	----------------------------------

## Collisions:

➤ collision_datetime_1_case_id_1	REGULAR ⓘ	13.4 MB	147 (since Fri Jun 14 2024)
➤ datetime.season_1	REGULAR ⓘ	4.6 MB	124 (since Sat Jun 15 2024)

Veliki broj grupisanja

# **Dodatne promene strukture**

# Čuvanje podataka:

Promena načina čuvanja određenih podataka kao i grupisanje određenih kolona.

```
road_surface : "dry"  
road_condition_1 : "normal"  
road_condition_2 : ""  
lighting : "dusk or dawn"
```

```
weather_1 : "cloudy"  
weather_2 : ""
```

```
▼ conditions : Object  
  weather : "cloudy"  
  road_surface : "dry"  
  road_conditions : "normal"  
  lighting : "dusk or dawn"
```

```
collision_datetime : "2009-02-11T17:35:00"
```

```
▼ datetime : Object  
  datetime : "2009-02-11T17:35:00"  
  day_of_week : "Wednesday"  
  season : "Winter"
```