

Tijana Lazić IN 3/2021  
Savo Savić IN 50/2021



**Turbulencije u rasporedu:  
Zašto avioni kasne**

# Sadržaj:

- Karakteristike baze podataka
- Semantika obeležja
- Primer širenja podataka
- Predlog logičke šeme za MongoDB
- Predlog agregacija - 10 pitanja



# Karakteristike baze podataka

**Tema:** Analiza kašnjenja civilnih letova u SAD-u 2023. godine, uzimajući u obzir meteorološke uslove i karakteristike aviona

**Izvor podataka:** Kaggle datasetovi „[2023 US Civil Flights Delay, Meteo and Aircraft](#)” i “[Airports Data](#)”

**Veličina skupa podataka:** Cancelled\_Diverted\_2023.csv (18.69 MB), US\_flights\_2023.csv (1.15 GB), airports\_geolocation.csv (26.77 kB), weather\_meteo\_by\_airport.csv( 7.41 MB), airports.csv ((9.1 MB)), airport-frequencies.csv (1.22 MB), runways.csv (3.05 MB)

# Karakteristike baze podataka

## Struktura podataka:

**Cancelled\_Diverted\_2023.csv**: Informacije o otkazanim i preusmerenim letovima (broj redova: 104.488, broj kolona: 23)

**US\_flights\_2023.csv**: Detalji o letovima, uključujući kašnjenja, avione i tipove kašnjenja (broj redova: 6.743.404, broj kolona: 24)

**Airports\_Geolocation.csv**: Geografske informacije o aerodromima (broj redova: 364, broj kolona: 7)

**Weather\_Meteo\_by\_Airport.csv**: Meteorološki podaci po aerodromima (broj redova: 132.860, broj kolona: 10)

**airports.csv**: Sadrži detalje o aerodromima širom sveta, uključujući tip aerodroma, lokaciju, nadmorsku visinu, i usluge (broj redova: 61.221, broj kolona: 18)

**airport\_frequencies.csv**: Sadrži podatke o komunikacionim frekvencijama aerodroma (broj redova: 28.927, broj kolona: 6)

**runways.csv**: Sadrži podatke o pistama na aerodromima, uključujući dužinu, širinu, površinu, i osvetljenost (broj redova: 41.761, broj kolona: 20)

# Semantika obeležja

## Cancelled\_Diverted\_2023.csv

**FlightDate:** Datum leta (npr. 2023-01-01)

**Day\_Of\_Week:** Dan u nedelji (1–7, gde je 1 ponedeljak)

**Airline:** Naziv avio-kompanije

**Tail\_Number:** Jedinstveni identifikator aviona

**Cancelled:** Indikator otkazivanja (0/1)

**Diverted:** Indikator preusmeravanja (0/1)

**Dep\_Airport, Arr\_Airport:** IATA kodovi polaznog i dolaznog aerodroma

**Dep\_CityName, Arr\_CityName:** Gradovi polaska i dolaska

# Semantika obeležja

## Cancelled\_Diverted\_2023.csv

**DepTime\_label:** Vremenski interval polaska (npr. jutro, popodne)

**Dep\_Delay, Arr\_Delay:** Kašnjenje polaska/dolaska (u minutima)

**Dep\_Delay\_Tag, Arr\_Delay\_Tag:** Indikator kašnjenja (0/1)

**Dep\_Delay\_Type, Arr\_Delay\_Type:** Tip kašnjenja (npr. early, on-time, delayed)

**Flight\_Duration:** Trajanje leta (u minutima)

**Distance\_type:** Kategorija distance (npr. kratka, srednja, duga)

**Delay\_Carrier, Delay\_Weather, Delay\_NAS, Delay\_Security, Delay\_LastAircraft:** Kašnjenja uzrokovana avio-kompanijom, vremenom, NAS-om, sigurnošću ili prethodnim letom (u minutima)

# Semantika obeležja

## US\_flights\_2023.csv

**FlightDate:** Datum leta (format: YYYY-MM-DD, npr. 2023-01-01)

**Day\_Of\_Week:** Dan u nedelji kada je let obavljen (1–7, gde je 1 ponedeljak, 2 utorak, itd.)

**Airline:** Naziv avio-kompanije koja upravlja letom (npr. American Airlines, Delta)

**Tail\_Number:** Jedinostveni identifikator aviona (npr. N123AA)

**Dep\_Airport:** IATA kod polaznog aerodroma (npr. JFK)

**Dep\_CityName:** Naziv grada polaznog aerodroma (npr. New York)

**DepTime\_label:** Vremenski interval polaska (npr. Morning, Afternoon, Evening)

# Semantika obeležja

## US\_flights\_2023.csv

**Dep\_Delay:** Kašnjenje polaska u minutima (pozitivno za kašnjenje, negativno za raniji polazak)

**Dep\_Delay\_Tag:** Indikator kašnjenja polaska (0 za na vreme ili ranije, 1 za kašnjenje)

**Dep\_Delay\_Type:** Tip kašnjenja polaska (npr. Early, On-time, Delayed)

**Arr\_Airport:** IATA kod dolaznog aerodroma (npr. LAX)

**Arr\_CityName:** Naziv grada dolaznog aerodroma (npr. Los Angeles)

**Arr\_Delay:** Kašnjenje dolaska u minutima (pozitivno za kašnjenje, negativno za raniji dolazak)

**Arr\_Delay\_Type:** Tip kašnjenja dolaska (npr. Early, On-time, Delayed)

**Flight\_Duration:** Trajanje leta u minutima



# Semantika obeležja

## US\_flights\_2023.csv

**Distance\_type:** Kategorija distance leta (npr. Short, Medium, Long)

**Delay\_Carrier:** Kašnjenje uzrokovano avio-kompanijom (u minutima)

**Delay\_Weather:** Kašnjenje uzrokovano vremenskim uslovima (u minutima)

**Delay\_NAS:** Kašnjenje uzrokovano Nacionalnim sistemom vazdušnog saobraćaja (NAS) (u minutima)

**Delay\_Security:** Kašnjenje uzrokovano sigurnosnim procedurama (u minutima)

**Delay\_LastAircraft:** Kašnjenje uzrokovano kašnjenjem prethodnog leta aviona (u minutima)

**Manufacturer:** Proizvođač aviona (npr. Boeing, Airbus)

**Model:** Model aviona (npr. 737, A320)

**Aircraft\_age:** Starost aviona u godinama

# Semantika obeležja

## Airports\_Geolocation.csv

**IATA\_CODE:** IATA kod aerodroma

**AIRPORT:** Naziv aerodroma

**CITY:** Grad

**STATE:** Država

**COUNTRY:** Zemlja aerodroma

**LATITUDE, LONGITUDE:** Geografske koordinate aerodroma

# Semantika obeležja

## **Weather\_Meteo\_by\_Airport.csv**

**time:** Datum i vreme meteorološkog merenja

**tavg, tmin, tmax:** Prosečna, minimalna i maksimalna temperatura (°C)

**prcp:** Padavine (mm)

**snow:** Sneg (mm)

**wdir:** Smer vetra (stepeni)

**wspd:** Brzina vetra (km/h)

**pres:** Atmosferski pritisak (hPa)

**airport\_id:** IATA kod aerodroma

# Semantika obeležja

## airports.csv

**id:** Jedinstveni identifikator aerodroma

**ident:** ICAO kod aerodroma (npr. KJFK za JFK aerodrom)

**type:** Tip aerodroma (npr. large\_airport, medium\_airport, small\_airport)

**name:** Puno ime aerodroma (npr. John F Kennedy International Airport)

**latitude\_deg:** Geografska širina aerodroma

**longitude\_deg:** Geografska dužina aerodroma

**elevation\_ft:** Nadmorska visina aerodroma u stopama (može uticati na performanse aviona)

**continent:** Kontinent na kojem se nalazi aerodrom (npr. NA za Severnu Ameriku)

**iso\_country:** ISO kod države (npr. US)

# Semantika obeležja

## airports.csv

**iso\_region:** ISO kod regiona (npr. US-NY za Njujork)

**municipality:** Grad ili opština gde se aerodrom nalazi

**scheduled\_service:** Da li aerodrom ima redovne komercijalne letove (yes/no)

**gps\_code:** GPS kod aerodroma (može biti sličan ICAO kodu)

**iata\_code:** IATA kod aerodroma (npr. JFK), ključ za povezivanje sa US\_flights\_2023

**local\_code:** Lokalni kod aerodroma (često za male aerodrome)

**home\_link:** URL zvanične web stranice aerodroma.

**wikipedia\_link:** URL Vikipedija stranice aerodroma.

**keywords:** Ključne reči povezane sa aerodromom

# Semantika obeležja

## **airport\_frequencies.csv**

**id:** Jedinstveni identifikator frekvencije

**airport\_ref:** Referenca na id iz Airports tabele

**airport\_ident:** ICAO kod aerodroma (povezuje se sa ident iz Airports)

**type:** Tip frekvencije (npr. TWR za kulu, APP za prilaz)

**description:** Opis frekvencije (npr. Tower, Approach)

**frequency\_mhz:** Frekvencija u megahercima

# Semantika obeležja

## runways.csv

**id:** Jedinstveni identifikator piste

**airport\_ref:** Referenca na id iz Airports tabele

**airport\_ident:** ICAO kod aerodroma (povezuje se sa ident iz Airports)

**length\_ft:** Dužina piste u stopama

**width\_ft:** Širina piste u stopama

**surface:** Materijal piste (npr. Asphalt, Concrete)

**lighted:** Da li je pista osvetljena (1 = da, 0 = ne)

**closed:** Da li je pista zatvorena (1 = da, 0 = ne)

**le\_ident:** Identifikator nižeg kraja piste (npr. 04L)

**le\_latitude\_deg:** Geografska širina nižeg kraja piste

# Semantika obeležja

## runways.csv

**le\_longitude\_deg:** Geografska dužina nižeg kraja piste

**le\_elevation\_ft:** Nadmorska visina nižeg kraja piste

**le\_heading\_degT:** Azimut nižeg kraja piste

**le\_displaced\_threshold\_ft:** Pomeraj praga nižeg kraja piste

**he\_ident:** Identifikator višeg kraja piste (npr. 22R)

**he\_latitude\_deg:** Geografska širina višeg kraja piste

**he\_longitude\_deg:** Geografska dužina višeg kraja piste

**he\_elevation\_ft:** Nadmorska visina višeg kraja piste

**he\_heading\_degT:** Azimut višeg kraja piste

**he\_displaced\_threshold\_ft:** Pomeraj praga višeg kraja piste



# Primer prostiranja podataka kroz datoteke

## Proces širenja

Pronalazim let sa kašnjenjem ( $\text{Dep\_Delay} > 0$  ili  $\text{Arr\_Delay} > 0$ ) iz `US_flights_2023.csv` za datum 2023-12-31.

Povezujem sa `Cancelled_Diverted_2023.csv` koristeći `FlightDate`, `Airline`, `Dep_Airport`, `Arr_Airport`.

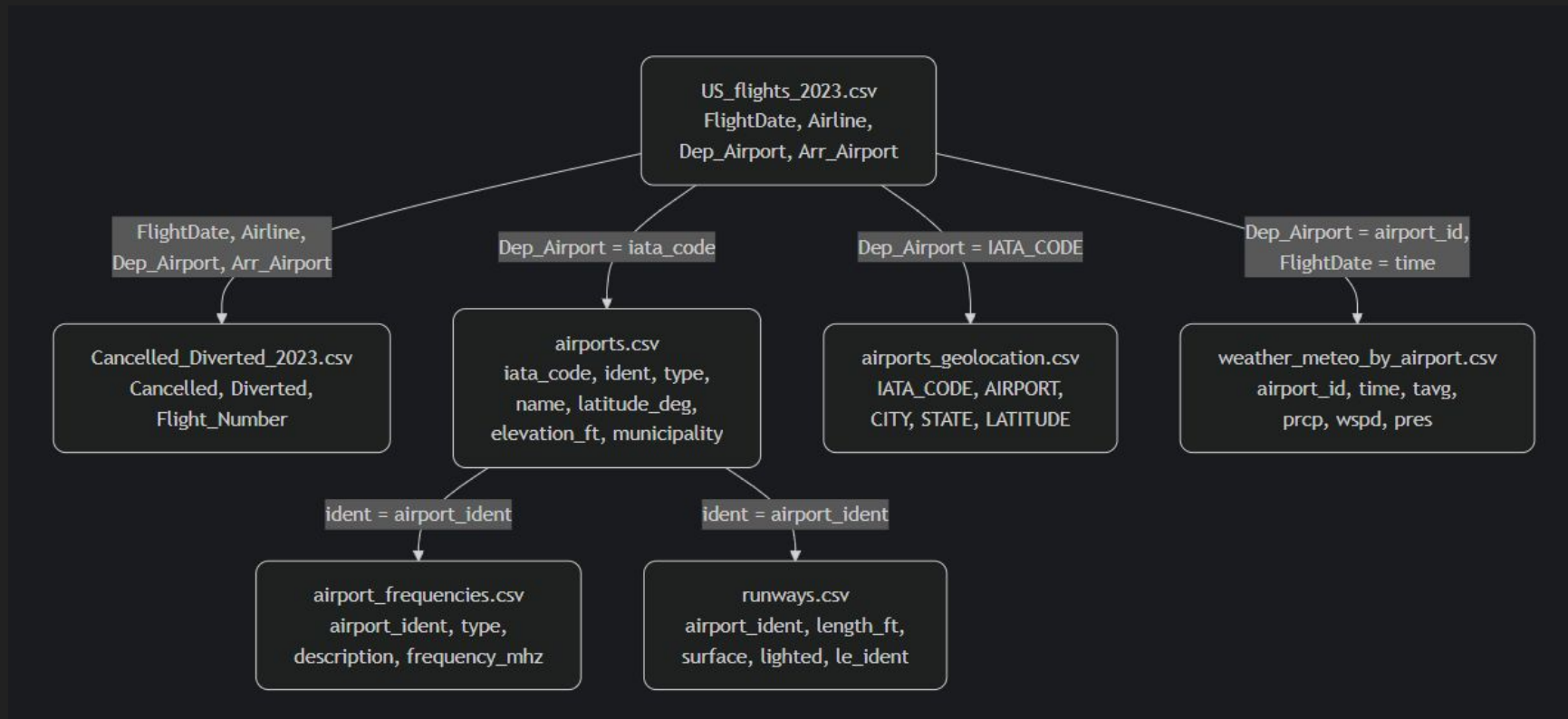
Povezujem polazni aerodrom (`Dep_Airport`) sa `airports.csv` preko `iata_code`.

Koristim `ident` iz `airports.csv` za povezivanje sa `airport_frequencies.csv` i `runways.csv` preko `airport_ident`.

Povezujem vremenske uslove iz `weather_meteo_by_airport.csv` koristeći `airport_id` (koji odgovara `Dep_Airport`) i `time` (koji odgovara `FlightDate`).

Povezujem geolokaciju iz `airports_geolocation.csv` koristeći `IATA_CODE` (koji odgovara `Dep_Airport`).

# Primer prostiranja podataka kroz datoteke



# Primer prostiranja podataka kroz datoteke

Let (iz US\_flights\_2023.csv):

- FlightDate: 2023-12-31
- Airline: Delta Air Lines
- Dep\_Airport: ATL
- Arr\_Airport: LAX
- Dep\_Delay: 20 (minuta)
- Arr\_Delay: 15 (minuta)
- Tail\_Number: N456DL
- Flight\_Duration: 300 (minuta)
- Manufacturer: Boeing
- Model: 737-900
- Aircraft\_age: 7.5
- Dep\_CityName: Atlanta
- Arr\_CityName: Los Angeles
- DepTime\_label: Afternoon
- Dep\_Delay\_Tag: 1
- Dep\_Delay\_Type: Moderate
- Arr\_Delay\_Type: Minor
- Distance\_type: Long
- Delay\_Carrier: 5.0
- Delay\_Weather: 10.0
- Delay\_NAS: 5.0
- Delay\_Security: 0.0
- Delay\_LastAircraft: 0.

# Primer prostiranja podataka kroz datoteke

## Status leta (iz Cancelled\_Diverted\_2023.csv):

- FlightDate: 2023-12-31
- Airline: Delta Air Lines
- Dep\_Airport: ATL
- Arr\_Airport: LAX
- Cancelled: 0
- Diverted: 0
- Flight\_Number: 789

## Frekvencije za ATL (iz airport\_frequencies.csv):

- id: 67890
- airport\_ref: 12345
- airport\_ident: KATL
- type: TWR
- description: Tower
- frequency\_mhz: 119.5

# Primer prostiranja podataka kroz datoteke

## Polazni aerodrom (ATL) (iz airports.csv):

- id: 12345
- ident: KATL
- type: large\_airport
- name: Hartsfield-Jackson Atlanta International Airport
- latitude\_deg: 33.6367
- longitude\_deg: -84.4281
- elevation\_ft: 1026
- continent: NA
- iso\_country: US
- iso\_region: US-GA
- municipality: Atlanta
- scheduled\_service: yes
- gps\_code: KATL
- iata\_code: ATL
- local\_code: ATL
- home\_link: <http://www.atl.com>
- wikipedia\_link: [https://en.wikipedia.org/wiki/Hartsfield-Jackson\\_Atlanta\\_International\\_Airport](https://en.wikipedia.org/wiki/Hartsfield-Jackson_Atlanta_International_Airport)
- keywords: Atlanta, Hartsfield

# Primer prostiranja podataka kroz datoteke

## Piste za ATL (iz runways.csv):

- id: 23456
- airport\_ref: 12345
- airport\_ident: KATL
- length\_ft: 11889
- width\_ft: 150
- surface: Concrete
- lighted: 1
- closed: 0
- le\_ident: 08L
- le\_latitude\_deg: 33.634
- le\_longitude\_deg: -84.435
- le\_elevation\_ft: 1015
- le\_heading\_degT: 89.0
- le\_displaced\_threshold\_ft: 0
- he\_ident: 26R
- he\_latitude\_deg: 33.639
- he\_longitude\_deg: -84.420
- he\_elevation\_ft: 1020
- he\_heading\_degT: 269.0
- he\_displaced\_threshold\_ft: 0

# Primer prostiranja podataka kroz datoteke

## Vremenski uslovi za ATL (iz weather\_meteo\_by\_airport.csv):

- airport\_id: ATL
- time: 2023-12-31
- tavg: 12.0 (°C)
- tmin: 8.0 (°C)
- tmax: 15.0 (°C)
- prcp: 2.5 (mm)
- snow: 0.0 (mm)
- wdir: 270.0 (stepeni)
- wspd: 18.0 (km/h)
- pres: 1010.5 (hPa)

## Geolokacija aerodroma (ATL) (iz airports\_geolocation.csv):

- IATA\_CODE: ATL
- AIRPORT: Hartsfield-Jackson Atlanta International Airport
- CITY: Atlanta
- STATE: GA
- COUNTRY: US
- LATITUDE: 33.6367
- LONGITUDE: -84.4281

# Predlog logičke šeme za MongoDB

## Flights

```
{
  "flight_id": "AA123_2023-01-01",
  "flight_date": "2023-01-01",
  "day_of_week": 1,
  "airline": "American Airlines",
  "tail_number": "N123AA",
  "flight_number": 1234,
  "status": {
    "cancelled": false,
    "diverted": false
  },
  "departure": {
    "airport": "JFK",
    "city": "New York",
    "time_label": "Morning",
    "delay": 15,
    "delay_tag": 1,
    "delay_type": "Delayed"
  },
}
```

```
  "arrival": {
    "airport": "LAX",
    "city": "Los Angeles",
    "delay": 10,
    "delay_type": "Delayed"
  },
  "flight_details": {
    "duration": 360,
    "distance_type": "Long",
    "delays": {
      "carrier": 5,
      "weather": 0,
      "nas": 10,
      "security": 0,
      "late_aircraft": 0
    }
  },
  "aircraft": {
    "manufacturer": "Boeing",
    "model": "737",
    "age": 5
  }
}
```



# Predlog logičke šeme za MongoDB

## Airports

```
{
  "iata_code": "JFK",
  "ident": "KJFK",
  "type": "large_airport",
  "name": "John F Kennedy International Airport",
  "elevation_ft": 13.0,
  "continent": "NA",
  "iso_country": "US",
  "iso_region": "US-NY",
  "municipality": "New York",
  "scheduled_service": "yes",
  "gps_code": "KJFK",
  "local_code": null,
  "home_link": "http://www.panynj.gov/airports/jfk.html",
  "wikipedia_link":
  "https://en.wikipedia.org/wiki/John_F._Kennedy_International_Airport",
  "keywords": "JFK, Kennedy"
}
```

## Weather

```
{
  "airport_id": "JFK",
  "time": "2023-01-01",
  "weather": {
    "tavg": 5,
    "tmin": 2,
    "tmax": 8,
    "prcp": 0,
    "snow": 0,
    "wdir": 270,
    "wspd": 10,
    "pres": 1013
  }
}
```

# Predlog logičke šeme za MongoDB

## Runways

```
{
  "airport_ident": "KJFK",
  "runway_id": "123456",
  "length_ft": 14572.0,
  "width_ft": 200.0,
  "surface": "Asphalt",
  "lighted": 1,
  "closed": 0,
  "low_end": {
    "ident": "04L",
    "latitude_deg": 40.629,
    "longitude_deg": -73.785,
    "elevation_ft": 12.0,
    "heading_degT": 44.0,
    "displaced_threshold_ft": 0.0
  },
  "high_end": {
    "ident": "22R",
    "latitude_deg": 40.645,
    "longitude_deg": -73.771,
    "elevation_ft": 13.0,
    "heading_degT": 224.0,
    "displaced_threshold_ft": 0.0
  }
}
```

# Predlog logičke šeme za MongoDB

## Frequencies

```
{
  "airport_ident": "KJFK",
  "frequency_id": 123,
  "type": "TWR",
  "description": "Tower",
  "frequency_mhz": 119.1
}
```

## Geolocation

```
{
  "iata_code": "JFK",
  "name": "John F Kennedy International
Airport",
  "city": "New York",
  "state": "NY",
  "country": "US",
  "location": {
    "latitude": 40.639801,
    "longitude": -73.7789
  }
}
```

# Predlog agregacija – 10 pitanja

1. Koji su aerodromi tipa 'large\_airport' sa prosečnim kašnjenjem polaska većim od 15 minuta, kada su padavine bile veće od 10 mm, i imaju piste duže od 10,000 stopa?
2. Koje avio-kompanije imaju najviše kašnjenja uzrokovanih vremenskim uslovima (vremensko kašnjenje > 10 minuta), za letove sa aerodroma sa nadmorskom visinom većom od 500 stopa i više od 5 komunikacionih frekvencija?
3. Koja je prosečna starost aviona starijih od 10 godina za letove sa kašnjenjem uzrokovanim nacionalnim avio-sistemom (NAS kašnjenje > 5 minuta), na aerodromima sa asfaltnom pistom i osvetljenjem?
4. Koji gradovi imaju najviše preusmerenih letova kada je smer vetra veći od 180 stepeni i brzina vetra preko 20 km/h?
5. Koji modeli aviona imaju najviše kašnjenja uzrokovanih kasnim avionom (kašnjenje > 15 minuta), na aerodromima tipa 'medium\_airport' sa više od 3 piste?

# Predlog agregacija – 10 pitanja

6. Koji je prosečni delay polaska za jutarnje letove (`DepTime_label = 'Morning'`) na aerodromima gde je bilo snega (`snow > 0`) i koriste frekvenciju tipa 'TWR'?
7. Koji proizvođači aviona imaju najviše otkazanih letova na aerodromima koji imaju lokalni kod?
8. Koji aerodromi imaju najveći udeo kašnjenja uzrokovanih NAS-om (`Delay_NAS`) tokom određenih dana u nedelji?
9. Koji su najčešći modeli aviona za letove na duge distance (`Distance_type = 'Long'`) koji poleću sa aerodroma u Njujorku sa pomerajem praga piste (`le_displaced_threshold_ft > 0`)?
10. Koji aerodromi imaju najduže prosečno trajanje leta za letove sa kašnjenjem uzrokovanim kasnim avionom (`Delay_LastAircraft > 10` minuta), u uslovima niskog pritiska (`pres < 1000` hPa), i `home_link`?