Analiza radova na putu u SAD (2016 - 2021)

SISTEMI BAZA PODATAKA

IN45/2019 OGNJEN POZNANOVIĆ

IN35/2019 LAZAR FRATRIĆ

Predlozi agregacija

Ognjen

1. Pronaći 10 država sa najdužim prosečnim vremenom trajanja radova.
2. Naći 10 država sa sa najvećim temperaturama za svaku godinu.
3. Naći za svaku državu po godinama koliko je radova obavljeno na levoj i desnoj strani.
4. Naći prosečnu temperaturu, vlažnost i pritisak za sve vremenske zone po godinama.
5. Naći koliko je svaka država obnovila milja puteva za svaku kategoriju radova kada je vreme izvođenja manje od godinu dana.

Predlozi agregacija

Lazar

6. Prikazati broj radova i prosečan broj milja po radu za svaki grad u radijus od 500km od NYC, sortirani u opadajućem redosledu.
7. U kojoj saveznoj državi, se najviše izvodilo radova po kategorijama, po godinama?
8. Koje je prosečno vreme trajanja radova u blizini pešačkog prelaza, za svaki grad?
9. Pronaći šifru radova koji su odradili najviše milja po gradovima.
10. Koja je ukupna dužina radova na autoputu za grad?

Inicijalna logička šema


```
{
    "_id" : ObjectId("647f4779a3cd254107b612eb"),
    "ID" : "C-2079942",
    "Severity" : NumberInt(4),
    "Start_Time" : ISODate("2020-11-09T05:46:55.000+0000"),
    "End_Time" : ISODate("2020-11-16T07:59:00.000+0000"),
    "Start_Lat" : 40.740911,
    "Start_Lng" : -96.644236,
    "End_Lat" : "40.74262",
    "End_Lng" : "-96.644244",
    "Distance(mi)" : 0.118081396671,
    "Description" : "Closed road from Pine Lake Rd to Madalyn Rd due to roadwork.",
    "Number" : null,
    "Street" : "S 56th St",
    "Side" : "R",
    "City" : "Lincoln",
    "County" : "Lancaster",
    "State" : "NE",
    "Zipcode" : "68516",
    "Country" : "US",
    "Timezone" : "US/Central",
    "Airport_Code" : "KLNK",
    "Weather_Timestamp" : ISODate("2020-11-09T09:54:00.000+0000"),
    "Temperature(F)" : "66.0",
    "Wind_Chill(F)" : "66.0",
    "Humidity(%)" : "78.0",
    "Pressure(in)" : "28.5",
    "Visibility(mi)" : "10.0",
    "Wind_Direction" : "S",
    "Wind_Speed(mph)" : "23.0",
    "Precipitation(in)" : "0.02",
    "Weather_Condition" : "Cloudy / Windy",
    "Amenity" : false,
    "Bump" : false,
    "Crossing" : true,
    "Give_Way" : false,
    "Junction" : false,
    "No_Exit" : false,
    "Railway" : false,
    "Roundabout" : false,
    "Station" : false,
    "Stop" : false,
    "Traffic_Calming" : false,
    "Traffic_Signal" : true,
    "Turning_Loop" : false,
    "Sunrise_Sunset" : "Night",
    "Civil_Twilight" : "Night",
    "Nautical_Twilight" : "Night",
    "Astronomical_Twilight" : "Day",
    "Start_day" : 9.0,
    "Start_month" : 11.0,
    "Start_year" : 2020.0,
    "End_day" : 16.0,
    "End_month" : 11.0,
    "End_year" : 2020.0
```

Izmene nad bazom

-Iz baze su izbačeni svi dokumenti koji su imali null vrednost u nekom od polja

-Polja Start_time i End_time su razbijeni na polja Start_year, Start_month, Start_day i End_year, End_month, End_day

Agregacije

Primer agregacije

Naći 10 država sa najdužim prosečnim vremenom trajanja radova - vreme izvršavanja 55s

```
db.radovi.aggregate([
  {
    $project: {
      State: 1,
      trajanje: {
        $subtract: [
          { $add: [{ $multiply: ["$End_year", 360] }, { $multiply: ["$End_month", 30] }, "$End_day"] },
          { $add: [{ $multiply: ["$Start_year", 360] }, { $multiply: ["$Start_month", 30] }, "$Start_day"] }
        ]
      }
    }
  },
  {
    $group: {
      _id: "$State",
      prosecno_trajanje: { $avg: "$trajanje" }
    }
  },
  {
    $sort: {
      prosecno_trajanje: -1
    }
  },
  {
    $limit: 10
  },
  {
    $project: {
      _id: 0,
      drzava: "$_id",
      prosecno_trajanje: { $round: ["$prosecno_trajanje", 2] }
    }
  }
])
```

Primer agregacije

Naći 10 država sa sa najvećim temperaturama za svaku godinu - vreme izvršavanja 9s

```
db.radovi.aggregate([
  {
    $group: {
      _id: {
        State: "$State",
        Start_year: "$Start_year"
      },
      maxTemperatura: { $max: "$Temperature(F)" }
    }
  },
  {
    $group: {
      _id: "$_id.State",
      maxTemperature: {
        $push: {
          Start_year: "$_id.Start_year",
          maxTemperatura: "$maxTemperatura"
        }
      }
    }
  },
  {
    $project: {
      _id: 0,
      State: "$_id",
      maxTemperature: 1
    }
  },
  {
    $sort: {
      "maxTemperature.maxTemperatura": -1
    }
  },
  {
    $limit: 10
  }
])
```

Primer agregacije

Naći za svaku državu po godinama koliko je radova obavljeno na levoj i desnoj strani.

```
db.radovi.aggregate([
  {
    $group: {
      _id: {
        State: "$State",
        Start_year: "$Start_year",
        strana: {
          $cond: {
            if: { $eq: ["$Side", "L"] },
            then: "Leva",
            else: "Desna"
          }
        }
      },
      brojRadova: { $sum: 1 }
    }
  },
  {
    $group: {
      _id: {
        State: "$_id.State",
        Start_year: "$_id.Start_year"
      },
      radovi: {
        $push: {
          strana: "$_id.strana",
          brojRadova: "$brojRadova"
        }
      },
      ukupanBrojRadova: { $sum: "$brojRadova" }
    }
  },
  {
    $sort: {
      "_id.State": 1
    }
  }
])
```

Primer agregacije

Naći prosečnu temperaturu, vlažnost i pritisak za sve vremenske zone po godinama

```
db.radovi.aggregate([
  {
    $group: {
      _id: {
        Timezone: "$Timezone",
        Start_year: "$Start_year"
      },
      avgTemperatura: { $avg: "$Temperature(F)" },
      avgHumidity: { $avg: "$Humidity(%)" },
      avgPressure: { $avg: "$Pressure(in)" }
    }
  },
  {
    $group: {
      _id: "$_id.Timezone",
      vrmenski_uslovi: {
        $push: {
          Start_year: "$_id.Start_year",
          avgTemperatura: "$avgTemperatura",
          avgHumidity: "$avgHumidity",
          avgPressure: "$avgPressure",
        }
      }
    }
  },
```

```
  {
    $project: {
      _id: 0,
      TImeZone: "$_id",
      vrmenski_uslovi: 1
    }
  },
  {
    $unwind: "$vrmenski_uslovi"
  },
  {
    $sort: {
      "vrmenski_uslovi.avgTemperatura": -1
    }
  }
])
```

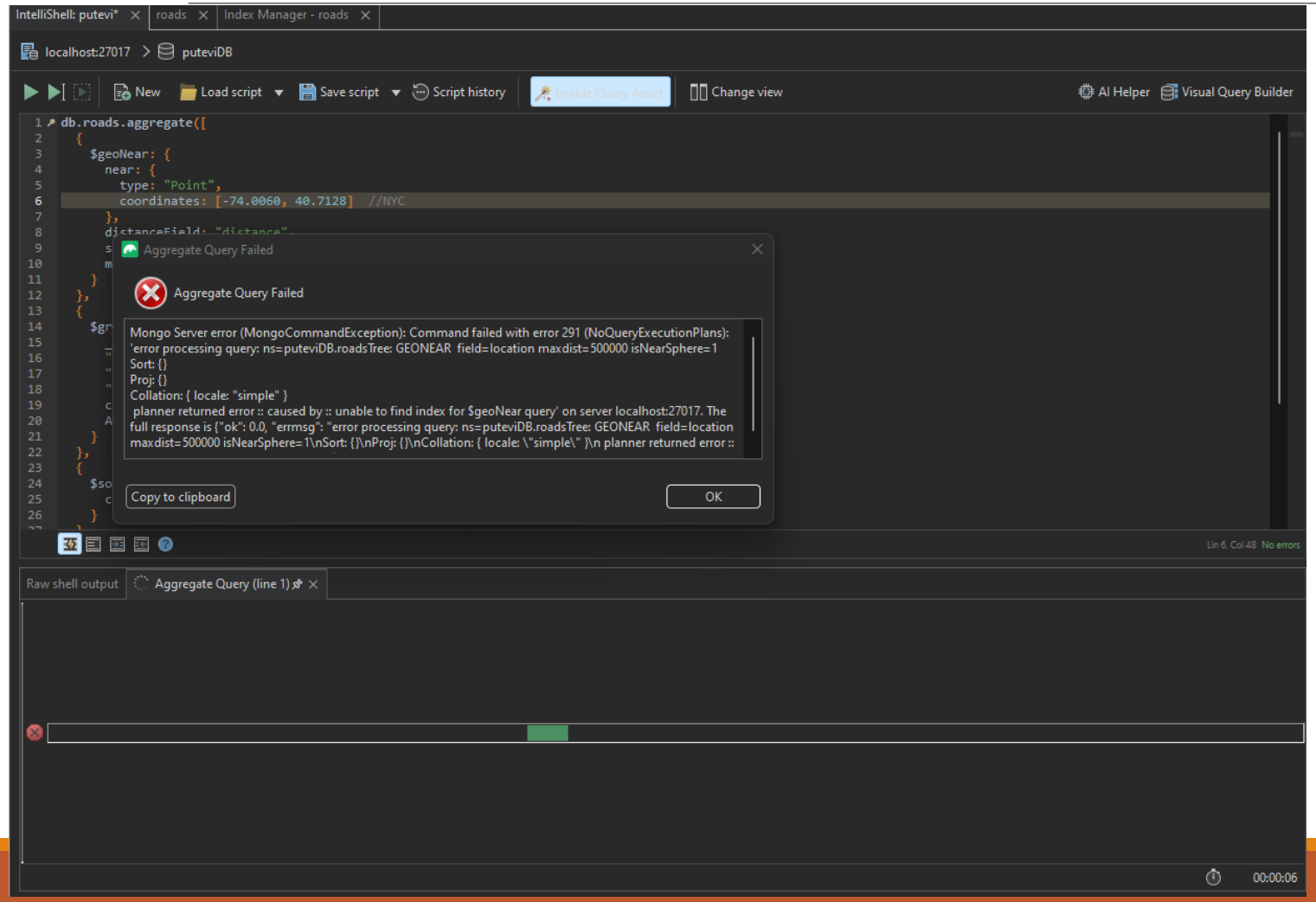
Primer agregacije

Naći koliko je svaka država obnovila milja puteva za svaku kategoriju radova kada je vreme izvođenja manje od godinu dana - vreme izvršavanja 52s

```
db.radovi.aggregate([
  {
    $project: {
      State: 1,
      Severity: 1,
      "Distance(mi)": 1,
      trajanje: {
        $subtract: [
          { $add: [{ $multiply: ["$End_year", 360] }, { $multiply: ["$End_month", 30] }, "$End_day"] },
          { $add: [{ $multiply: ["$Start_year", 360] }, { $multiply: ["$Start_month", 30] }, "$Start_day"] }
        ]
      }
    }
  },
  {
    $match: {
      trajanje: { $lt: 366 }
    }
  },
```

```
  {
    $group: {
      _id: {
        State: "$State",
        Severity: "$Severity"
      },
      uradjeno_milja: {$sum: "$Distance(mi)"},
    }
  },
  {
    $group: {
      _id: "$_id.State",
      radovi: {
        $push: {
          Severity: "$_id.Severity",
          uradjeno_milja: "$uradjeno_milja"
        }
      }
    }
  },
    {
    $project: {
      _id: 0,
      State: "$_id",
      radovi: 1
    }
  }
])
```

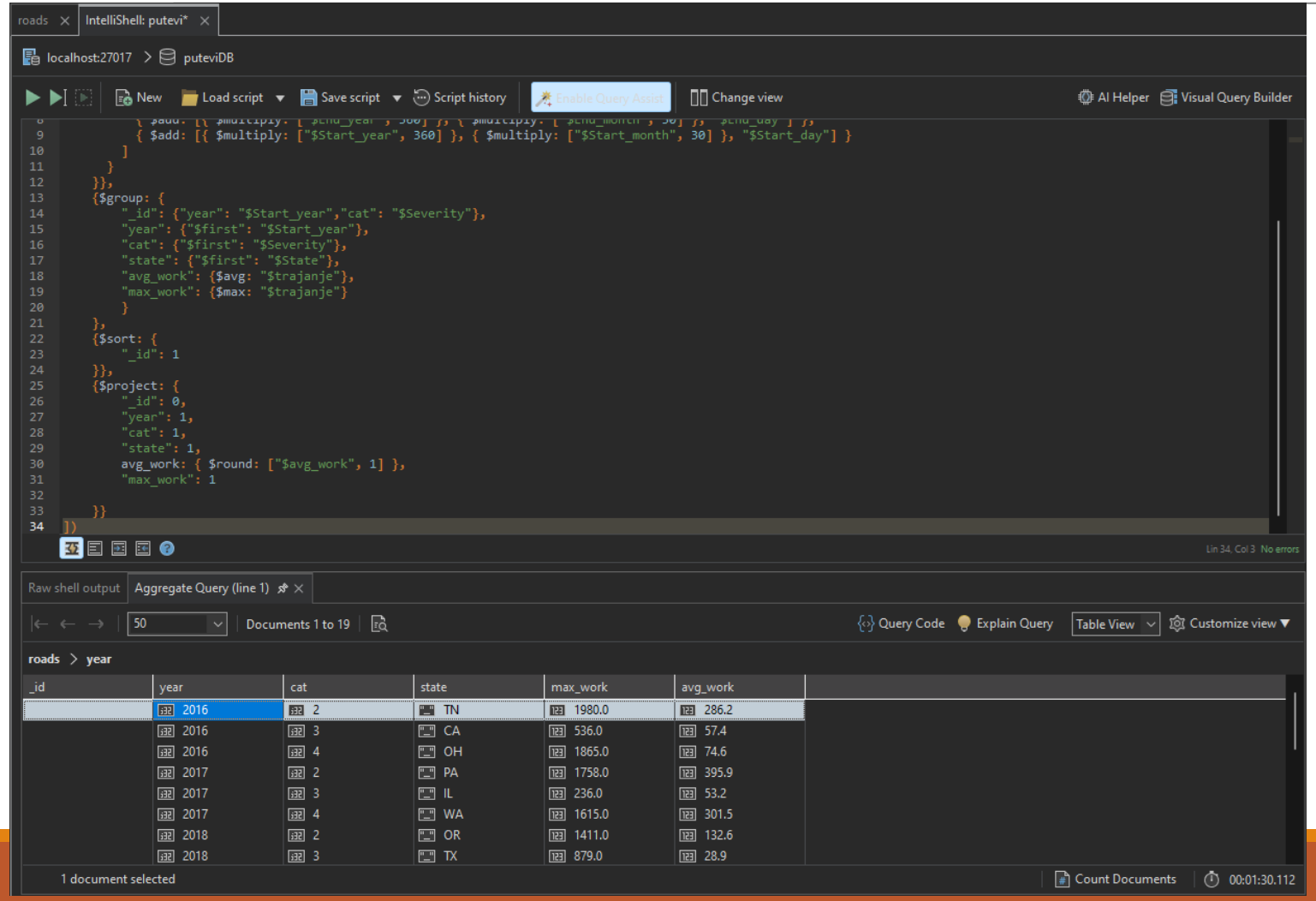
Primer agregacije



6. Prikazati broj radova i prosečan broj milja po radu za svaki grad u radijus od 500km od NYC, sortirani u opadajućem redosledu.

Vreme izvršavanja: Upit se ne može izvršiti bez dodavanja indeksa!

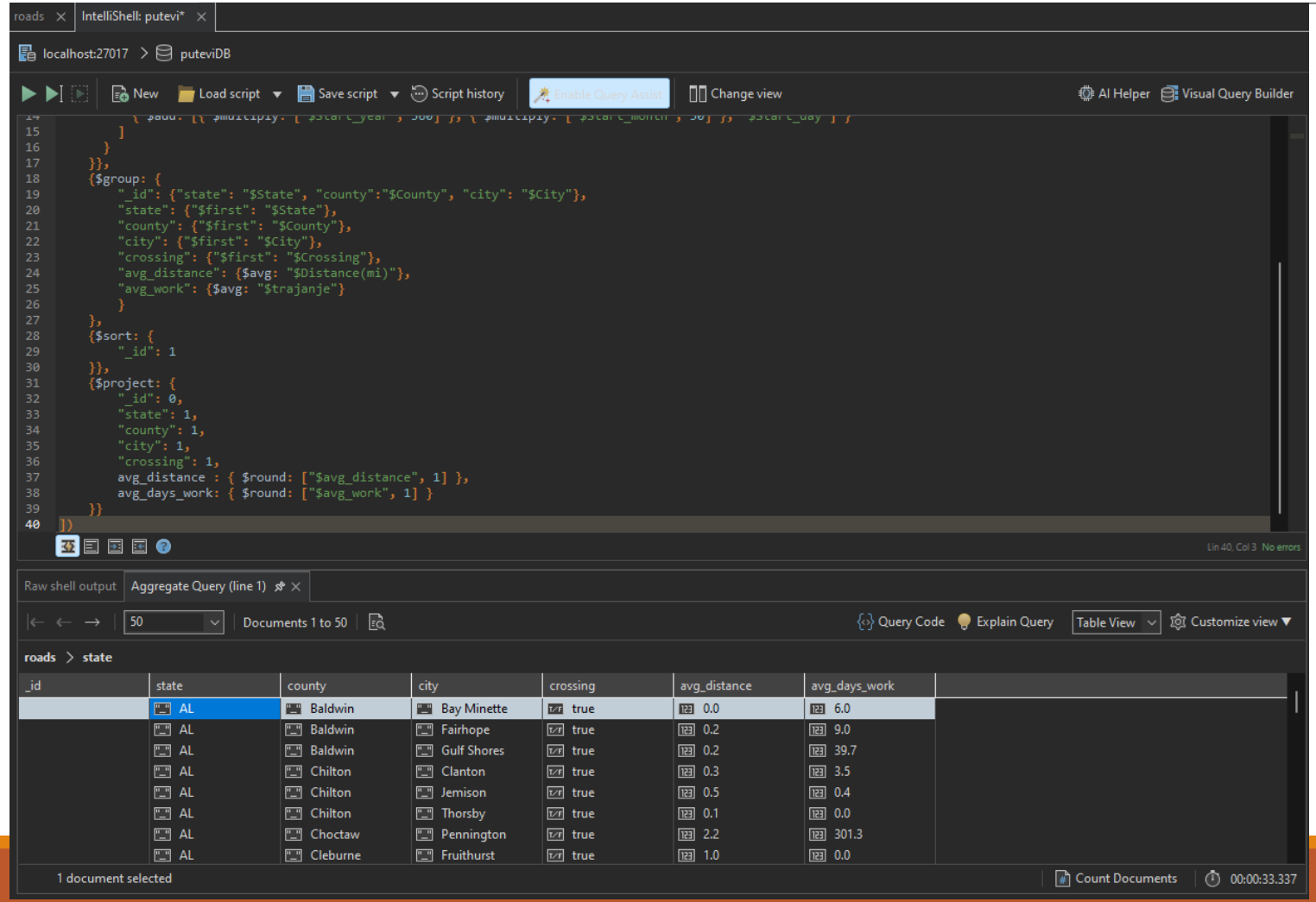
Primer agregacije



7. U kojoj saveznoj državi, se najviše izvodilo radova po kategorijama, po godinama?

Vreme izvršavanja: 90s

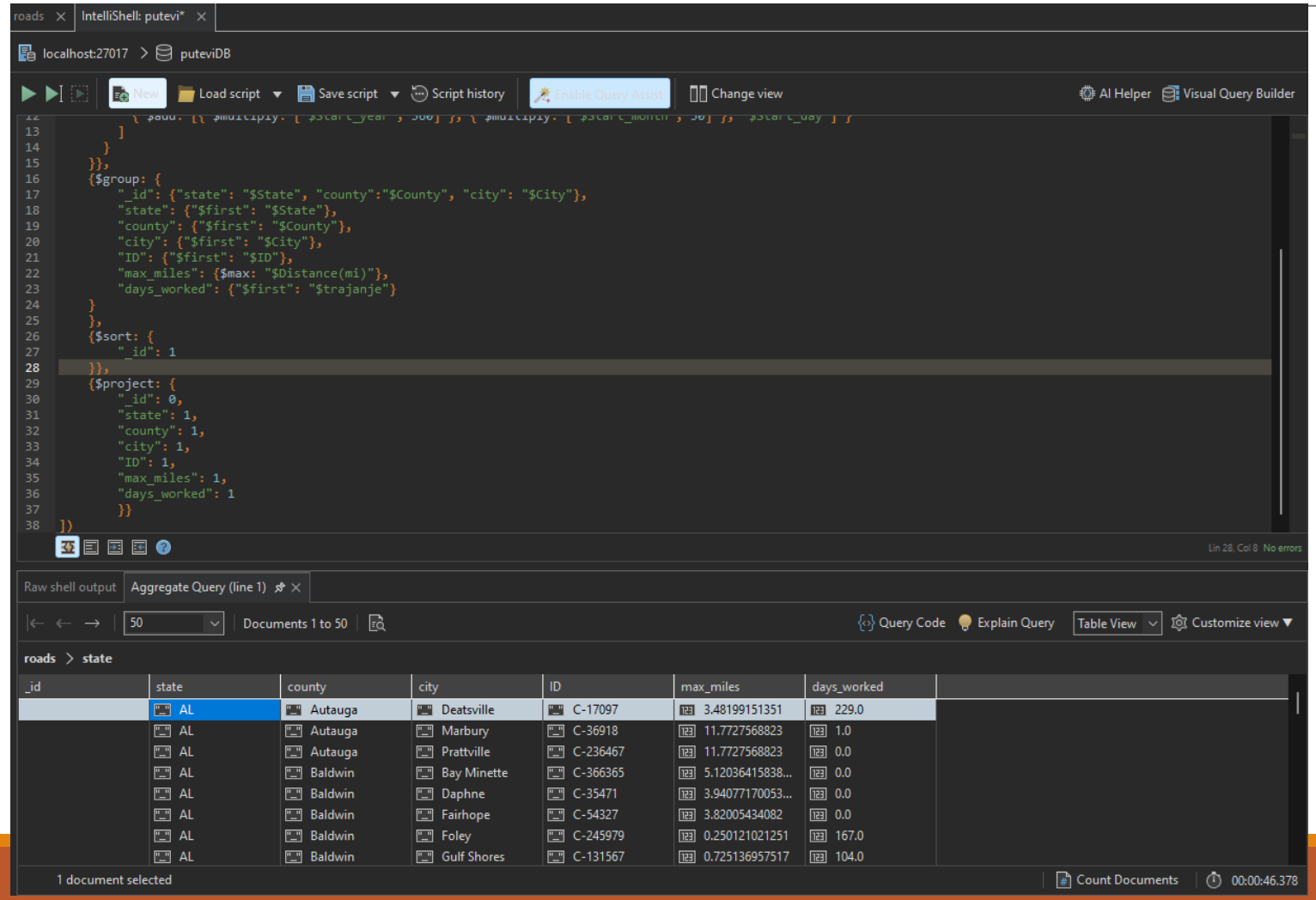
Primer agregacije



8. Koje je prosečno vreme trajanja radova u blizini pešačkog prelaza, za svaki grad, po okrugu, po saveznoj državi?

Vreme izvršavanja: 33s

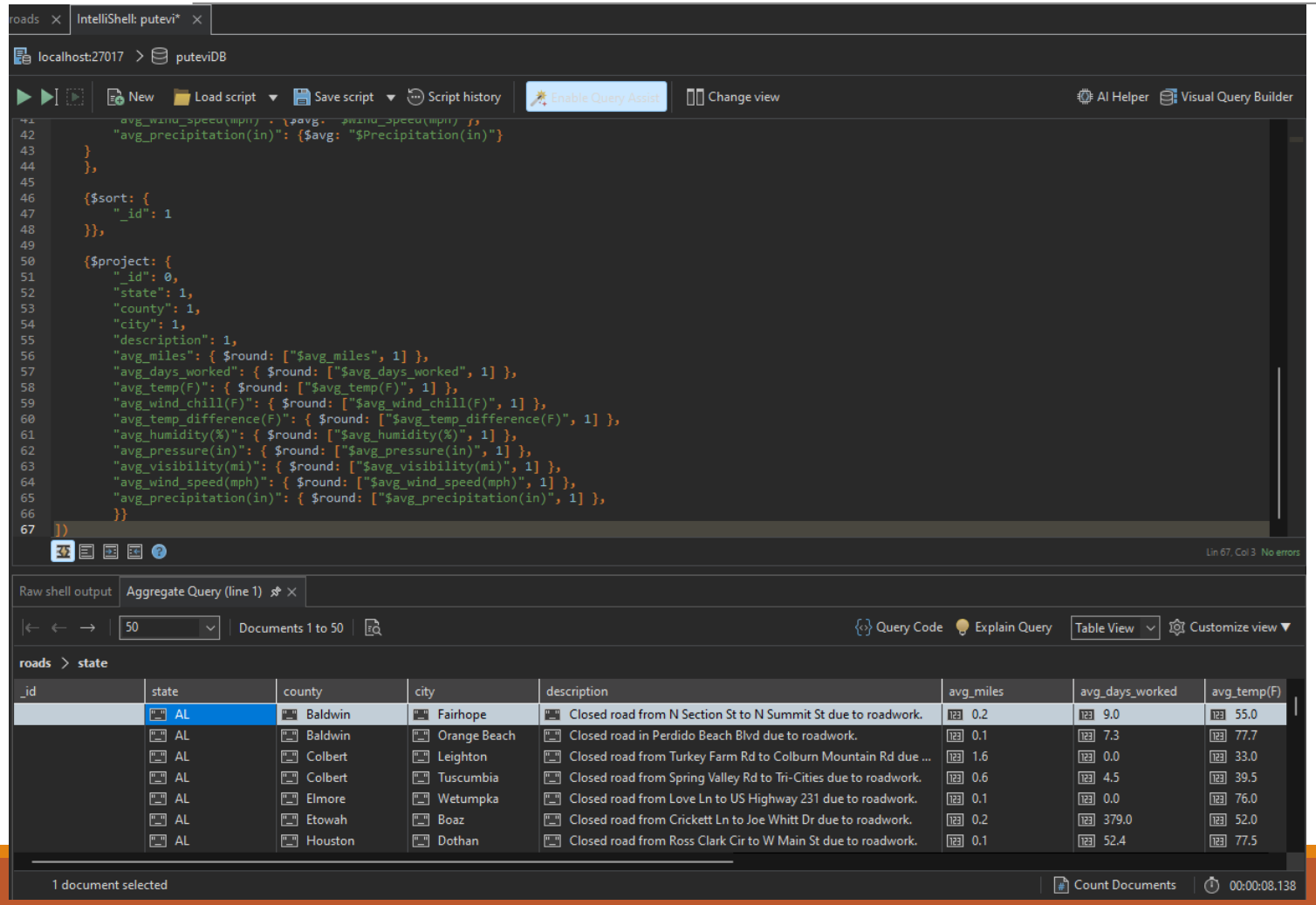
Primer agregacije



9. Pronaci šifru radova na kojima se odradilo najviše milja u tom gradu, po okrugu, po saveznoj državi.

Vreme izvršavanja: 46s

Primer agregacije



10. Koja je ukupna dužina radova na autoputu za svaki grad?

Vreme izvršavanja: 10s

Logička šema prilagođena agregacija

Logička šema prilagođena agregacija

Početna kolekcija je podeljena na dve nove “roadsPrimary” i “roadsSecondary”


```
{
    "_id" : ObjectId("6486b94fd75c133bf6ffe678"),
    "Severity" : NumberInt(4),
    "Start_Time" : ISODate("2019-04-05T16:00:00.000+0000"),
    "End_Time" : ISODate("2020-09-29T11:53:57.000+0000"),
    "Start_Lat" : 32.838359999999994,
    "Start_Lng" : -93.152378,
    "End_Lat" : 32.85074,
    "End_Lng" : -93.164388,
    "Distance(mi)" : 1.10349651559,
    "Description" : "Construction on LA-534 WB near EDMONDS LOOP Road closed. Take alternate route.",
    "Number" : NumberInt(4200),
    "Street" : "Highway 534",
    "Side" : "R",
    "City" : "Haynesville",
    "County" : "Claiborne",
    "State" : "LA",
    "Zipcode" : "71038-7130",
    "Country" : "US",
    "Timezone" : "US/Central",
    "Airport_Code" : "KMNE",
    "Weather_Timestamp" : ISODate("2019-04-05T15:55:00.000+0000"),
    "Temperature(F)" : NumberInt(75),
    "Wind_Chill(F)" : NumberInt(75),
    "Humidity(%)" : NumberInt(58),
    "Pressure(in)" : NumberInt(29),
    "Visibility(mi)" : NumberInt(10),
    "Wind_Direction" : "S",
    "Wind_Speed(mph)" : NumberInt(3),
    "Precipitation(in)" : NumberInt(0),
    "Weather_Condition" : "Fair",
    "Start_day" : NumberInt(5),
    "Start_month" : NumberInt(4),
    "Start_year" : NumberInt(2019),
    "End_day" : NumberInt(29),
    "End_month" : NumberInt(9),
    "End_year" : NumberInt(2020),
    "Days_worked" : NumberInt(534),
    "ID" : "C-1"
}
```



```
{
    "_id" : ObjectId("6486d3f1caef744b7e429b36"),
    "Amenity" : false,
    "Bump" : false,
    "Crossing" : false,
    "Give_Way" : false,
    "Junction" : false,
    "No_Exit" : false,
    "Railway" : false,
    "Roundabout" : false,
    "Station" : false,
    "Stop" : false,
    "Traffic_Calming" : false,
    "Traffic_Signal" : false,
    "Turning_Loop" : false,
    "Sunrise_Sunset" : "Day",
    "Civil_Twilight" : "Day",
    "Nautical_Twilight" : "Day",
    "Astronomical_Twilight" : "Day",
    "ID" : "C-1"
}
```

Logička šema prilagođena agregacija

- Dodato polje Days_worked
- Dodato polje location
- Dodat Geospatial indeks
- Dodat indeks nad poljem Days_worked i Temperature(F)

Logička šema prilagođena agregacija

Dodato polje Days_worked i location

- Polje Days_worked je dodato na osnovu polja Start_year, Start_month, Start_day i End_year, End_month, End_day

- Polje location:

```
db.roadsPrimary.aggregate([
  {
    $addFields: {
      location: {
        type: "Point",
        coordinates: [
          "$Start_Lng",
          "$Start_Lat"
        ]
      }
    }
  },
  {
    $out: "roadsPrimary"
  }
])
```

Logička šema prilagođena agregacija

Dodat Geospatial indeks, indeks nad “Days_worked”, “State” i “Temperature(F)”

```
1  db.roadsPrimary.createIndex({ location: "2dsphere" })
2
3
```

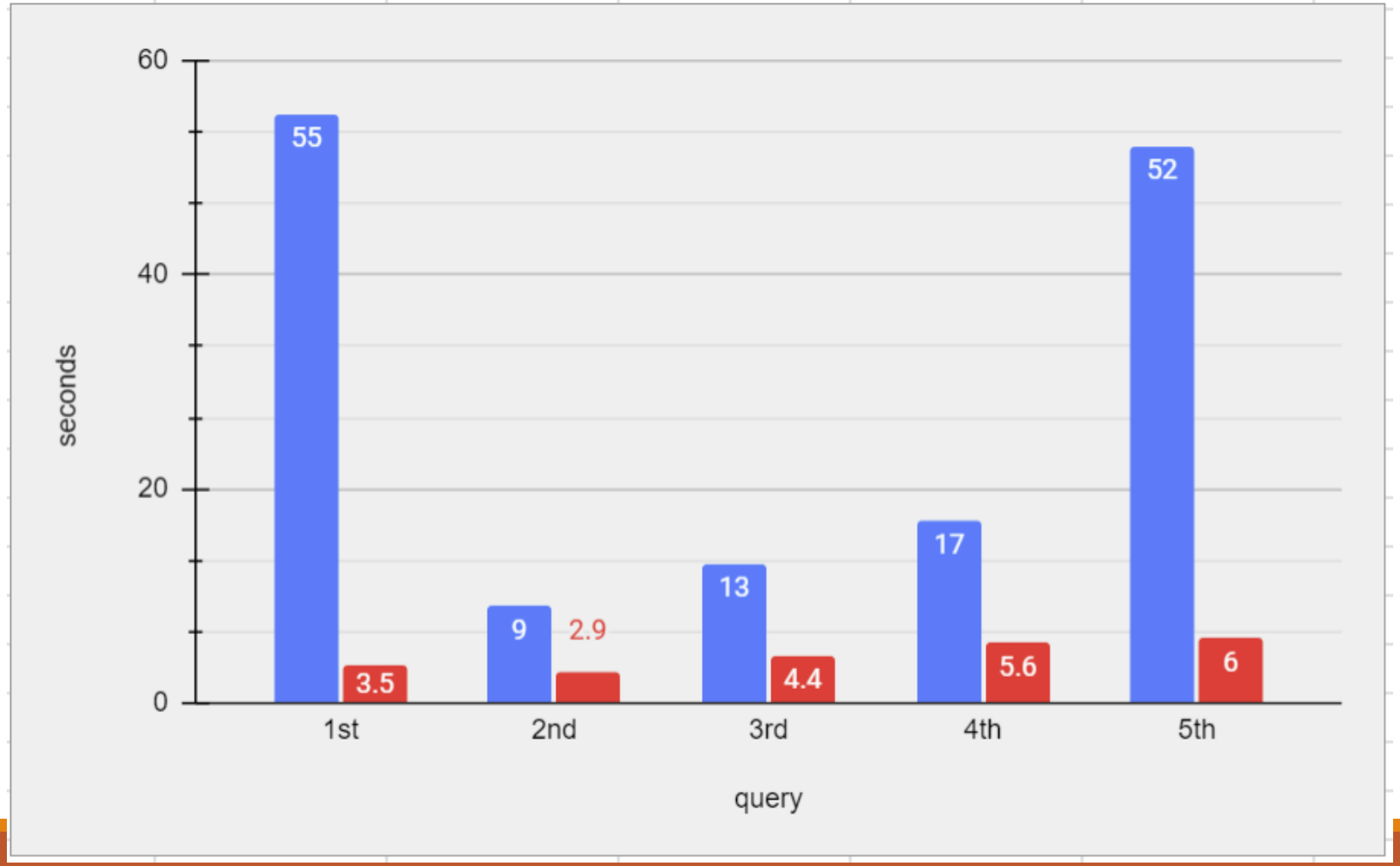
```
db.roadsPrimary.createIndex({ Days_worked: -1 })
```

```
1  db.roadsPrimary.createIndex({ State: -1 })
2
```

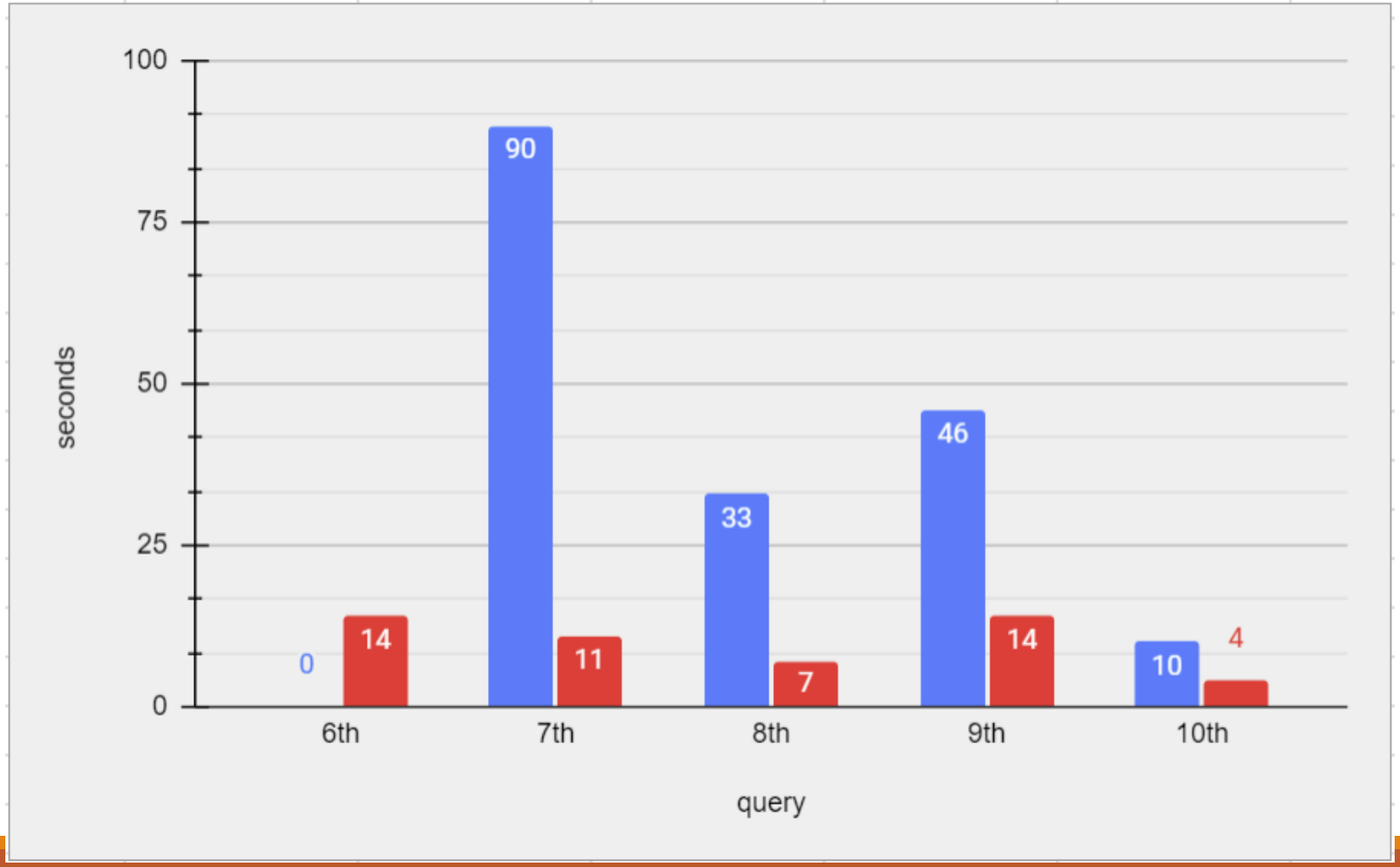
```
db.roadsPrimary.createIndex({ Temperature(F) : -1 })
```

Poređenje performansi

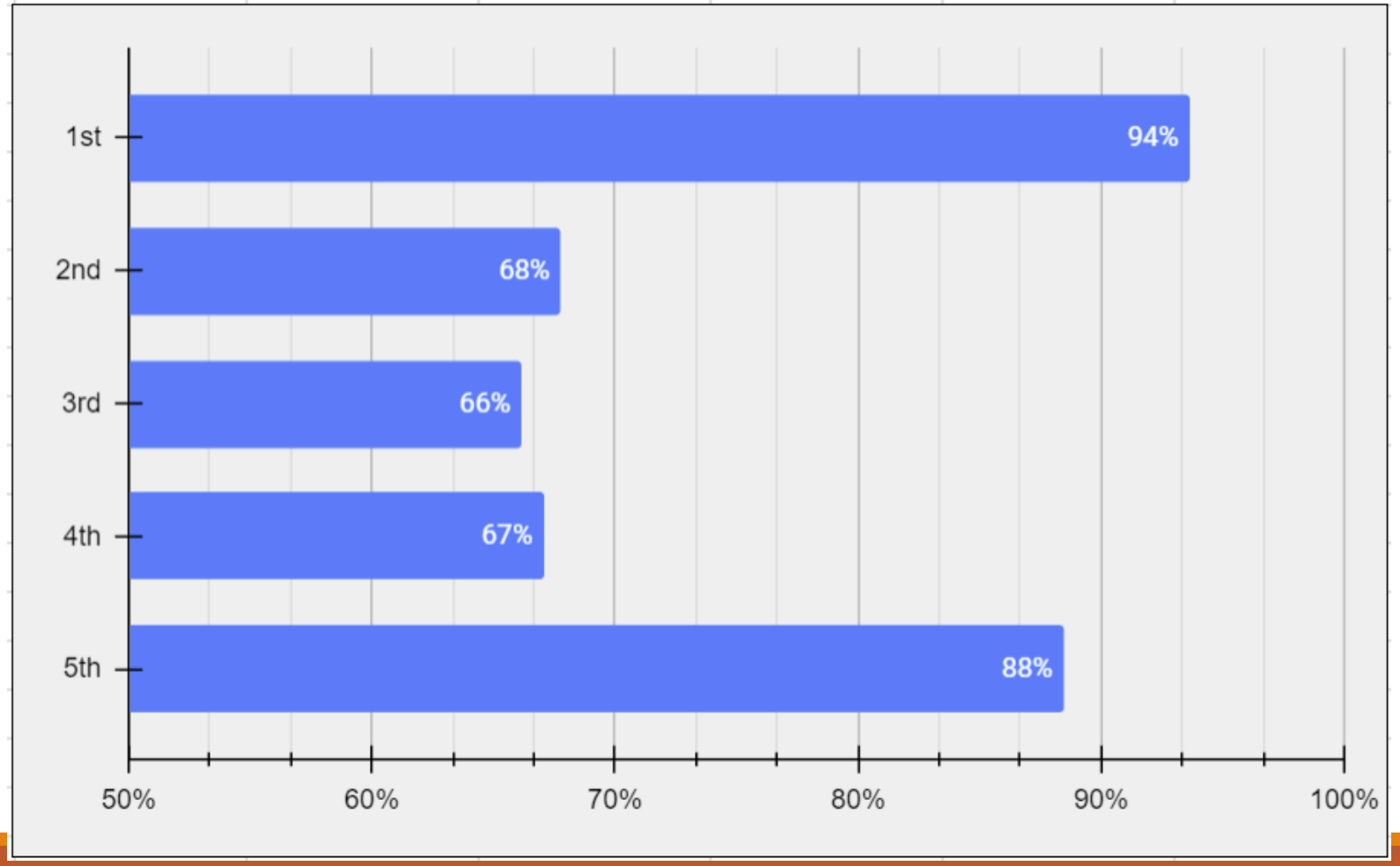
Poređenje performansi – u sekundama



Poređenje performansi – u sekundama



Poređenje performansi – u procentima



Poređenje performansi – u procentima