Asistent:
Elena Akik

Nikola Pantić IN 40/2020
Srđan Petrović IN 47/2020

Preporuka Steam Igara

Sistemi baza podataka

Link do repozitorijuma: https://github.com/ssspetrovic/steam-recommendations

Tema

- Kao temu, odabrali smo "Game recommendations on Steam" skup podataka
- Steam - najveća svetska platforma igara na svetu
- Ovaj skup podataka fokusira se na preporuke korisnika na Steam platformi.
- Sadrži informacije o različitim igrama, njihovim karakteristikama i interakcijama korisnika.
- Preporuke se temelje na ponašanju korisnika, kao što su ocene igara, vreme igranja, dodavanje u listu želja itd.

Datoteke i broj zapisa

- **games.csv** : Informacije o igrama - **50.872** zapisa
- **recommendations.csv**: Informacije o recenzijama - **41.154.794** zapisa
- **users.csv**: Informacije o igracima - **14.306.064** zapisa

Semantika kolona

- **games.csv**:
  - app_id: Jedinstveni identifikator igre.
  - title: Naslov igre.
  - date_release: Datum izdavanja igre.
  - rating: Ocena igre.
  - positive_ratio: Procenat pozitivnih recenzija u odnosu na ukupan broj recenzija.
  - user_reviews: Broj recenzija koje je igra dobila od korisnika.
  - win: Da li je igra dostupna za Windows platformu (true/false).
  - mac: Da li je igra dostupna za Mac platformu (true/false).
  - linux: Da li je igra dostupna za Linux platformu (true/false).
  - steam_deck: Da li je igra dostupna za Steam Deck platformu (true/false).
  - price_final: Konačna cena igre.
  - price_original: Originalna cena igre pre popusta.
  - discount: Procenat popusta na cenu igre.

- **recommendations.csv**:
  - user_id: Jedinstveni identifikator korisnika koji je ostavio preporuku.
  - review_id: Jedinstveni identifikator recenzije.
  - app_id: Jedinstveni identifikator igre na koju se odnosi preporuka.
  - funny: Broj korisnika koji su označili recenziju kao smešnu.
  - helpful: Broj korisnika koji su označili recenziju kao korisnu.
  - date: Datum kada je preporuka ostavljena.
  - is_recommended: Da li je igra preporučena (true/false).
  - hours: Broj sati koje je korisnik proveo igrajući igru.

- **users.csv**:
  - user_id: Jedinstveni identifikator korisnika.
  - products: Može predstavljati proizvode koje je korisnik kupio, registrovao ili koristi.
  - reviews: Broj recenzija koje je korisnik napisao.

Predlog logičke šeme baze podataka

1. Kolekcija: Games
{
  "_id": ObjectId("..."),
  "app_id": "54321",
  "title": "Amazing Game",
  "date_release":
ISODate("2022-03-15T00:00:00Z"),
  "rating": 4.7,
  "positive_ratio": 0.92,
  "user_reviews": 1500
}

2. Kolekcija: Platform
{
  "_id": ObjectId("..."),
  "app_id": "54321",
  "windows": true,
  "mac": true,
  "linux": false,
  "steam_deck": true
}

3. Kolekcija: Price
{
  "_id": ObjectId("..."),
  "app_id": "54321",
  "price_final": 29.99,
  "price_original": 49.99,
  "discount": 40
}

4. Kolekcija: Recommendations:
{
  "_id": ObjectId("..."),
  "review_id": "67890",
  "user_id": "12345",
  "app_id": "54321",
  "date":
ISODate("2023-05-10T00:00:00Z"),
  "is_recommended": true,
  "hours": 35.0
}

5. Kolekcija: Review Feedback
{
  "_id": ObjectId("..."),
  "review_id": "67890",
  "funny": 2,
  "helpful": 25
}

6. Kolekcija: Users
{
  "_id": ObjectId("..."),
  "user_id": "12345",
  "products": 50,
  "reviews": 20
}

Agregacije nad podacima - Nikola

1. Najigranije plaćene igre na Linuxu koje su izašle do 2010. godine:

Agregacije nad podacima - Nikola

2. Koliko bi koštalo da se kupe sve igre za svaku od platformi?

```
 1  ▾ [
 2  ▾    {
 3  ▾      $group: {
 4            _id: "$app_id",
 5            app_id: {
 6              $first: "$app_id"
 7            }
 8          }
 9        },
10  ▾    {
11  ▾      $lookup: {
12            from: "Platform",
13            localField: "app_id",
14            foreignField: "app_id",
15            as: "platforms"
16          }
17        },
18  ▾    {
19  ▾      $lookup: {
20            from: "Price",
21            localField: "app_id",
22            foreignField: "app_id",
23            as: "prices"
24          }
25        },
26  ▾    {
27          $unwind: "$platforms"
28        },
29  ▾    {
30          $unwind: "$prices"
31        },
32  ▾    {
33          $facet: {
34            windows: [
35  ▾          {
36  ▾            $group: {
37                  _id: null,
38                  total_price: {
39  ▾                $sum: {
40  ▾                  $cond: [
41  ▾                    {
42                         $eq: [
```

PIPELINE OUTPUT

Sample of 4 documents

platform : "Steam Deck"
total_price : 5905.89

platform : "Windows"
total_price : 5905.89

platform : "Mac"
total_price : 1451.99

platform : "Linux"
total_price : 825

OUTPUT OPTIONS ▾

Agregacije nad podacima - Nikola

3. Najviše recenzirane igre za svaki kvartal:

```
1  ▼ [
2  ▼   {
3      $lookup: {
4        from: "Recommendations",
5        localField: "app_id",
6        foreignField: "app_id",
7        as: "reviews"
8      }
9    },
10 ▼   {
11      $unwind: "$reviews"
12    },
13 ▼   {
14      $group: {
15        _id: {
16          title: "$title",
17          year: {
18            $year: "$reviews.date"
19          },
20          quarter: {
21            $concat: [
22              {
23                $toString: {
24                  $year: "$reviews.date"
25                }
26              },
27              ", ",
28              "Q",
29              {
30                $cond: [
31                  {
32                    $lte: [
33                      {
34                        $month: "$reviews.date"
35                      },
36                      3
37                    ]
38                  },
39                  "1",
40                  {
41                    $cond: [
```

PIPELINE OUTPUT
Sample of 30 documents

OUTPUT OPTIONS ▼

```
total_reviews : 510
quarter : "2015, Q1"
title : "Counter-Strike: Global Offensive"

total_reviews : 1748
quarter : "2015, Q2"
title : "Grand Theft Auto V"

total_reviews : 488
quarter : "2015, Q3"
title : "Counter-Strike: Global Offensive"

total_reviews : 2253
quarter : "2015, Q4"
title : "Fallout 4"

total_reviews : 605
quarter : "2016, Q1"
title : "Fallout 4"

total_reviews : 1037
quarter : "2016, Q2"
title : "DARK SOULS" III"

total_reviews : 5960
quarter : "2016, Q3"
title : "No Man's Sky"
```

Agregacije nad podacima - Nikola

4. Koje igre su i dalje popularne?

```
 1 ▾ [
 2 ▾   {
 3 ▾     $lookup: {
 4           from: "Recommendations",
 5           localField: "app_id",
 6           foreignField: "app_id",
 7           as: "recommendations"
 8         }
 9       },
10 ▾   {
11         $unwind: "$recommendations"
12       },
13 ▾   {
14 ▾     $match: {
15           date_release: {
16             $lt: ISODate("2010-05-25")
17           },
18 ▾         "recommendations.date": {
19             $gte: ISODate("2020-05-25")
20           }
21         }
22       },
23 ▾   {
24 ▾     $group: {
25           _id: "$_id",
26 ▾         title: {
27             $first: "$title"
28           },
29 ▾         date_release: {
30             $first: "$date_release"
31           },
32 ▾         positive_reviews_last_year: {
33             $sum: 1
34           }
35         }
36       },
37 ▾   {
38 ▾     $sort: {
39           positive_reviews_last_year: -1
40         }
41       }
42     ]
```

PIPELINE OUTPUT

OUTPUT OPTIONS ▾

Sample of 3 documents
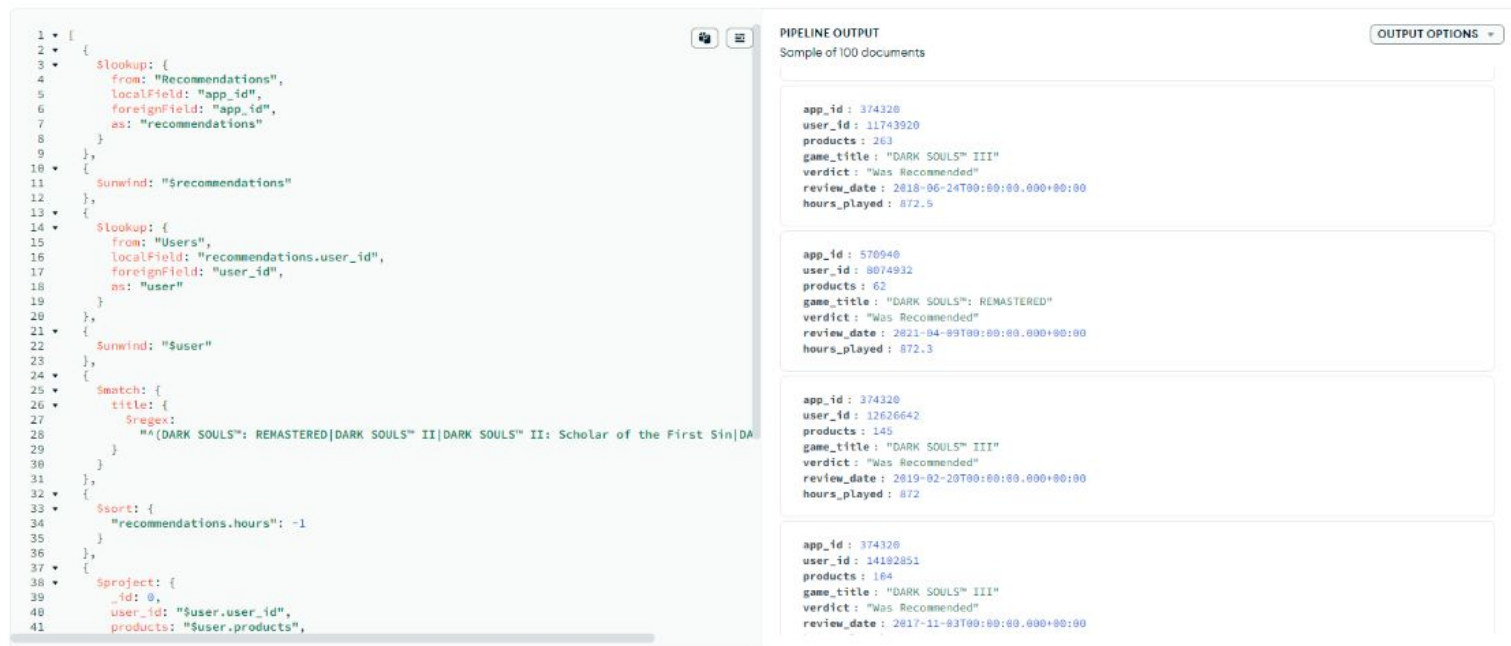
```
_id: ObjectId('666c57337620c964d5fd1ac0')
title : "Garry's Mod"
date_release : 2006-11-29T00:00:00.000+00:00
positive_reviews_last_year : 2859
```

```
_id: ObjectId('666c57337620c964d5fd10G4')
title : "Left 4 Dead 2"
date_release : 2009-11-16T00:00:00.000+00:00
positive_reviews_last_year : 584
```

```
_id: ObjectId('666c57337620c964d5fd97d1')
title : "Team Fortress 2"
date_release : 2007-10-10T00:00:00.000+00:00
positive_reviews_last_year : 117
```

Agregacije nad podacima - Nikola

5. Igrači sa najviše sati u Dark Souls igrama:

Agregacije nad podacima - Srđan

1. Recenzije koje su označene kao smešne za igre sa visokim pozitivnim rejtingom na Steam Deck platformi

Agregacije nad podacima - Srđan

2. Maksimalan broj sati koje su korisnici igrali igre za koje nisu preporučili, grupišući ih po igrama.

```
1  ▼  [
2  ▼    {
3  ▼      $lookup: {
4            from: "Recommendations",
5            localField: "app_id",
6            foreignField: "app_id",
7            as: "recommendations"
8          }
9        },
10 ▼    {
11         $unwind: "$recommendations"
12       },
13 ▼    {
14 ▼      $match: {
15           "recommendations.is_recommended": false
16         }
17       },
18 ▼    {
19 ▼      $group: {
20 ▼        _id: {
21             game_id: "$app_id",
22             game_name: "$title"
23           },
24 ▼        max_hours_played: {
25             $max: "$recommendations.hours"
26           },
27 ▼        users_with_max_hours: {
28 ▼          $push: {
29               user_id: "$recommendations.user_id",
30               hours: "$recommendations.hours"
31             }
32           }
33         }
34       },
35 ▼    {
36 ▼      $project: {
37           _id: 0,
38           game_id: "$_id.game_id",
39           game_name: "$_id.game_name",
40           max_hours_played: 1,
41 ▼        user_id: {
42             $arrayElemAt: [
```

PIPELINE OUTPUT
Sample of 5 documents

OUTPUT OPTIONS ▼

```
max_hours_played : 999.9
game_id : 386360
game_name : "SMITE®"
user_id : 9932417
```

```
max_hours_played : 999.9
game_id : 381210
game_name : "Dead by Daylight"
user_id : 13826219
```

```
max_hours_played : 999.9
game_id : 570
game_name : "Dota 2"
user_id : 10263359
```

```
max_hours_played : 999.9
game_id : 377160
game_name : "Fallout 4"
user_id : 11834030
```

```
max_hours_played : 999.8
game_id : 376210
game_name : "The Isle"
user_id : 9726305
```

Agregacije nad podacima - Srđan

3. Najveći fanovi "Dota 2" igre na osnovu maksimalnog broja sati koje su igrali.

```
1 ▼ [
2  ▼  {
3  ▼    $match: {
4        hours: { $gt: 20 },
5        is_recommended: true,
6        user_id: { $mod: [10, 2] }
7      }
8    },
9  ▼  {
10 ▼    $lookup: {
11       from: "Games",
12       localField: "app_id",
13       foreignField: "app_id",
14       as: "game_info"
15     }
16   },
17 ▼  {
18     $unwind: "$game_info"
19   },
20 ▼  {
21 ▼    $match: {
22       "game_info.title": "Dota 2"
23     }
24   },
25 ▼  {
26 ▼    $group: {
27       _id: "$user_id",
28       max_hours_played: { $max: "$hours" }
29     }
30   },
31 ▼  {
32 ▼    $sort: {
33       max_hours_played: -1
34     }
35   },
36 ▼  {
37     $limit: 10
38   },
39 ▼  {
40 ▼    $project: {
41       _id: 0,
42       user_id: "$_id"
```

PIPELINE OUTPUT
Sample of 10 documents

OUTPUT OPTIONS ▾

```
max_hours_played : 999.8
user_id : 10330952
```

```
max_hours_played : 998.4
user_id : 5528372
```

```
max_hours_played : 995.5
user_id : 1017772
```

```
max_hours_played : 991.9
user_id : 6022732
```

```
max_hours_played : 989.4
user_id : 5350942
```

```
max_hours_played : 985.6
user_id : 4917322
```

```
max_hours_played : 981.5
user_id : 10893502
```

```
max_hours_played : 978.7
user_id : 13811782
```

```
max_hours_played : 977.2
```

Agregacije nad podacima - Srđan

4. Procenat igara koje imaju vrlo pozitivan rejting, uz prosečnu cenu i prosečne korisničke recenzije.

```
 1 ▾ [
 2 ▾   {
 3 ▾     $lookup: {
 4         from: "Price",
 5         localField: "app_id",
 6         foreignField: "app_id",
 7         as: "price_info"
 8       }
 9     },
10 ▾   {
11       $unwind: "$price_info"
12     },
13 ▾   {
14       $group: {
15         _id: "$rating",
16         count: { $sum: 1 },
17         average_price: { $avg: "$price_info.price_final" },
18         average_reviews: { $avg: "$user_reviews" }
19       }
20     },
21 ▾   {
22       $group: {
23         _id: null,
24         total: { $sum: "$count" },
25         ratings: {
26 ▾         $push: {
27             rating: "$_id",
28             count: "$count",
29             average_price: "$average_price",
30             average_reviews: "$average_reviews"
31           }
32         }
33       }
34     },
35 ▾   {
36       $unwind: "$ratings"
37     },
38 ▾   {
39       $project: {
40         _id: 0,
41         rating: "$ratings.rating",
42         percentage: {
```

PIPELINE OUTPUT
Sample of 1 document

```
rating : "Very Positive"
percentage : 25.8275672275515
average_price : 10.183091559479411
average_reviews : 4082.434431844128
```

OUTPUT OPTIONS ▾

Agregacije nad podacima - Srđan

5. Najpopularnije igre na platformama Mac i Linux sa cenom manjom od 15 dolara, sortirane po broju korisničkih recenzija.
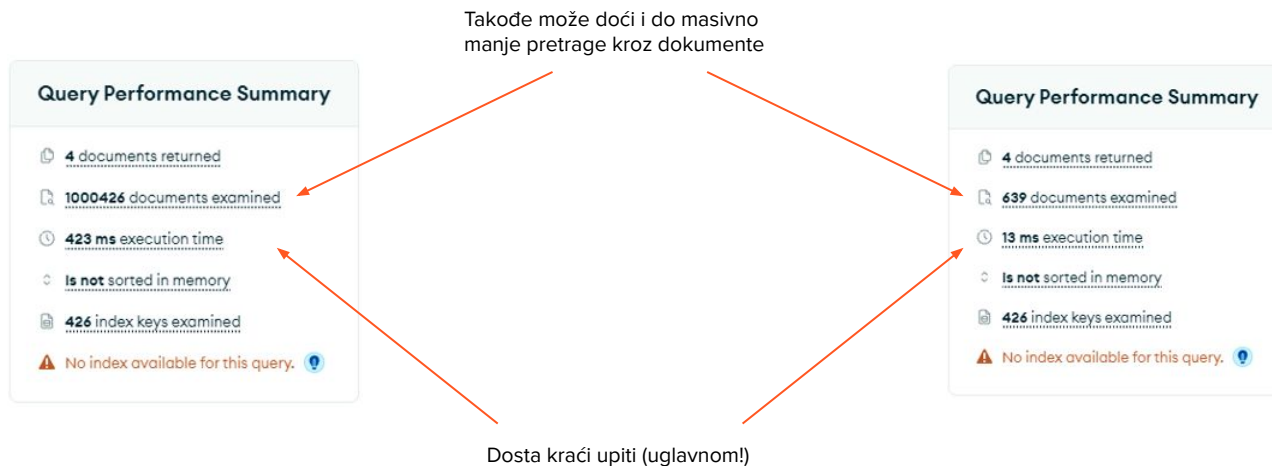
Restruktuiranje šema baze podataka

Rešenje za naš sistem baza podataka je sledeće:

- Users: ostaje isti, redudantno bi bilo da se restruktuira
- Games: uvedena je proširena referenca na cenu igre, pošto se često može upotrebiti
- Platform: odrađen je Subset pri čemu umesto da imamo podatke kao što su True i False za svaku platformu u svakom redu se ili nalazi platforma ako je igra igriva na njoj ili ne
- Price: takođe odrađen Subset, u ovom slučaju se unosi final_price uvek, ali ako je original_price isti, odbacuju se original_price i discount za taj red
- Recommendations: urađeno je baketiranje, pri čemu imamo skupljene podatke u nizove, čiji je ključ app_id
- ReviewFeedbacks: koristi se Computed restruktuiranje kako bi imali i zbir ukupnih odgovora na recenziju
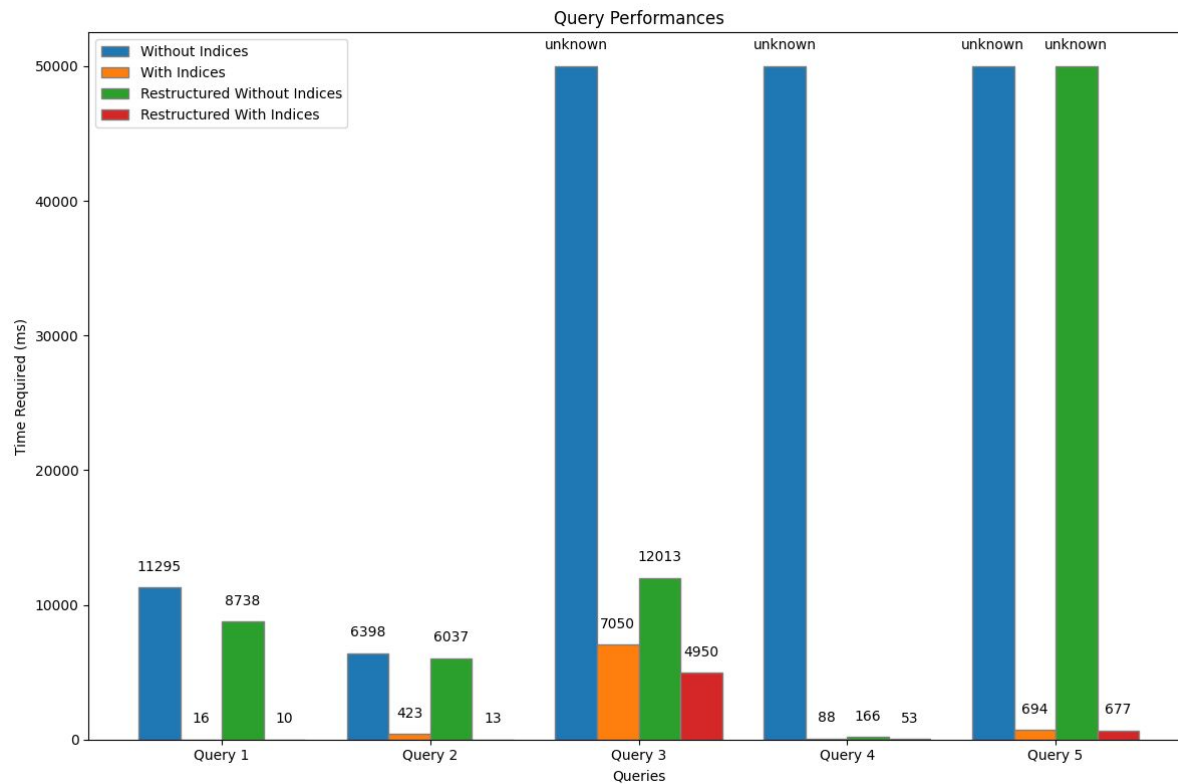
Restruktuirane šeme baze podataka

1. Kolekcija: Games
{
  "_id": {
    "$oid": "666cae77e2ba1ab06ce5c962"
  },
  "app_id": 13500,
  "date_release": {
    "$date": "2008-11-21T00:00:00.000Z"
  },
  "title": "Prince of Persia: Warrior Within™",
  "rating": "Very Positive",
  "positive_ratio": 84,
  "user_reviews": 2199,
  "price": 9.99
}

2. Kolekcija: Platform
{
  "_id": {
    "$oid": "666c40e23e3905b5a5489f23"
  },
  "app_id": 13500,
  "platforms": [
    "Windows",
    "Steam Deck"
  ]
}

3. Kolekcija: Price
{
  "_id": {
    "$oid": "666c414aaf7b1f1f9e55b957"
  },
  "app_id": 13500,
  "price_final": 9.99
}

4. Kolekcija: Recommendations:
{
  "_id": {
    "$oid": "666c5ac6933ffcb2665a82ed"
  },
  "app_id": 629730,
  "end_date": {
    "$date": "2021-12-07T00:00:00.000Z"
  },
  "recommendations": [
    {
      "review_id": 31,
      "user_id": 747452,
      "date": {
        "$date": "2021-02-10T00:00:00.000Z"
      },
      "is_recommended": true,
      "hours": 40.3
    },
    ...
  ]
}

5. Kolekcija: Review Feedback
{
  "_id": {
    "$oid": "666c75eaaec435bb146b23f0"
  },
  "review_id": 500,
  "funny": 1253,
  "helpful": 1494,
  "total_feedback": 2747
}

6. Kolekcija: Users
{
  "_id": {
    "$oid": "666c70ca0c3f3875ba4c729f"
  },
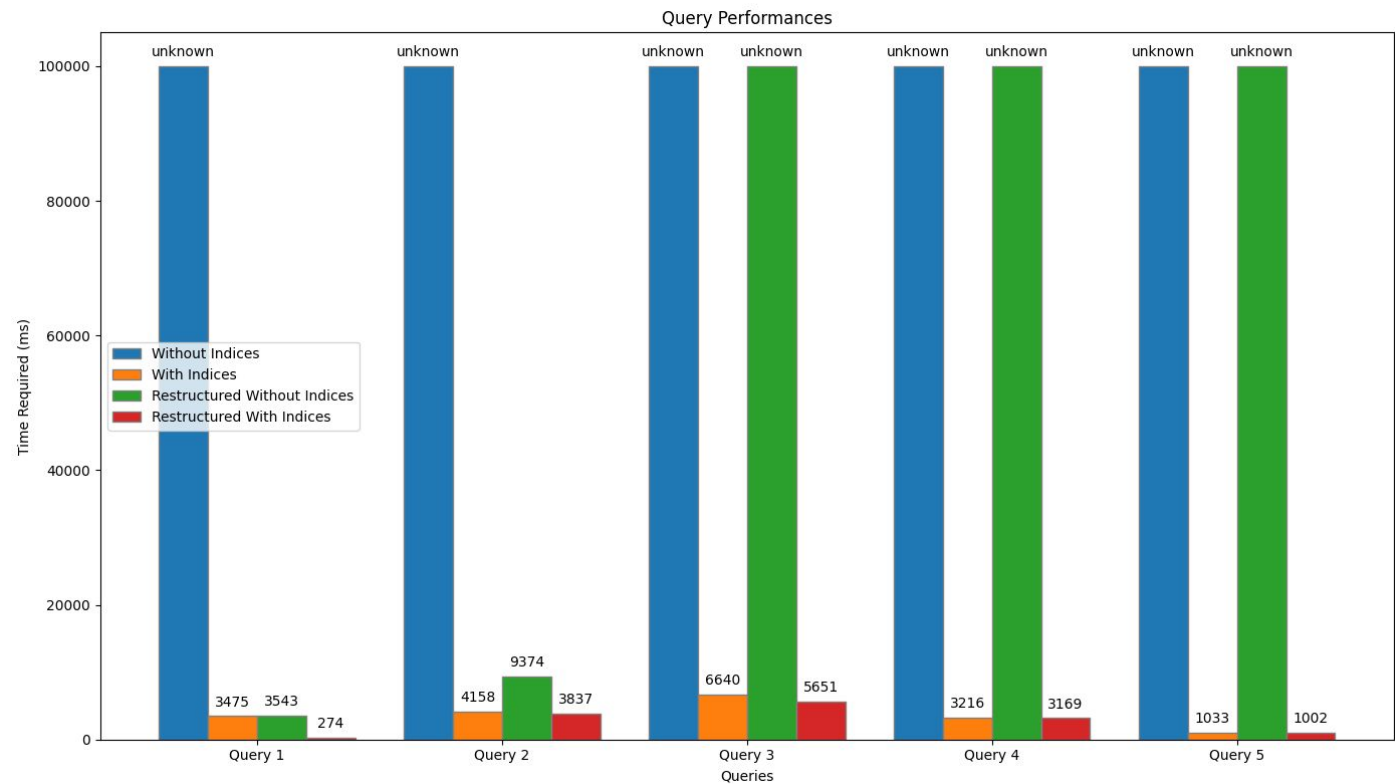 "user_id": 6366584,
  "products": 1597,
  "reviews": 129
}

Primer performansi

Takođe može doći i do masivno
manje pretrage kroz dokumente

Query Performance Summary

- 4 documents returned
- 1000426 documents examined
- 423 ms execution time
- Is not sorted in memory
- 426 index keys examined
- ⚠ No index available for this query. 💡

Query Performance Summary

- 4 documents returned
- 639 documents examined
- 13 ms execution time
- Is not sorted in memory
- 426 index keys examined
- ⚠ No index available for this query. 💡

Dosta kraći upiti (uglavnom!)

Performanse - Nikola

Query Performances

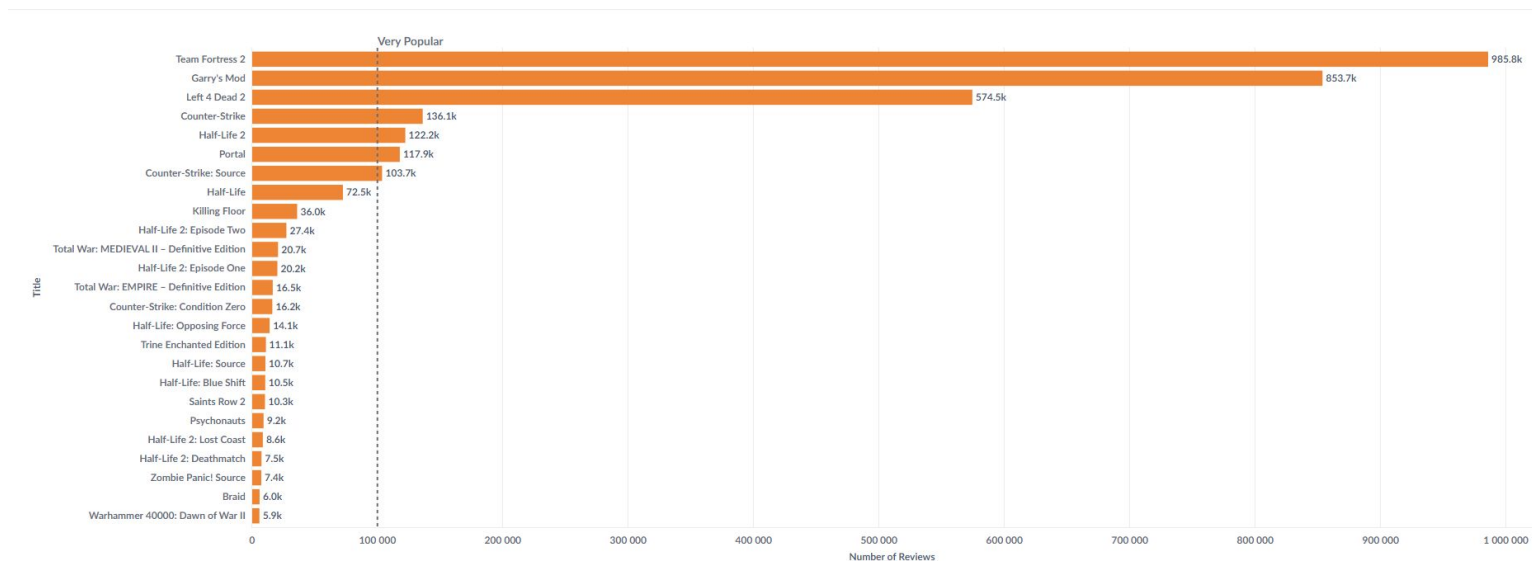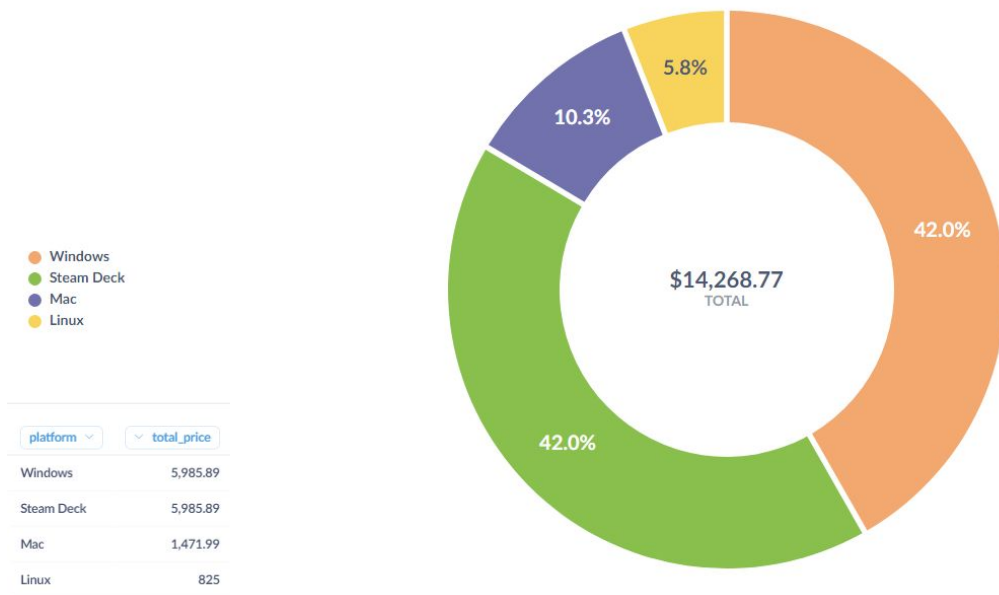# Performanse - Srđan

Query Performances

Metabase Upiti - Nikola

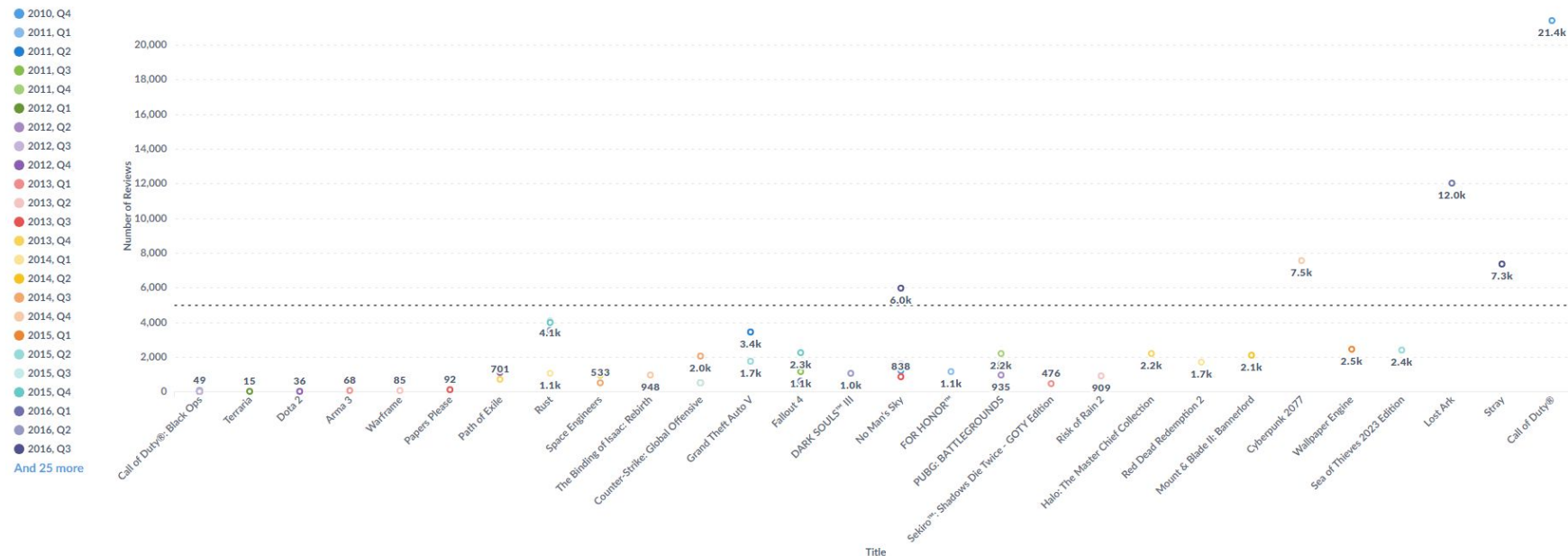**1. Najigranije plaćene igre na Linuxu koje su izašle do 2010. godine:**

Metabase Upiti - Nikola

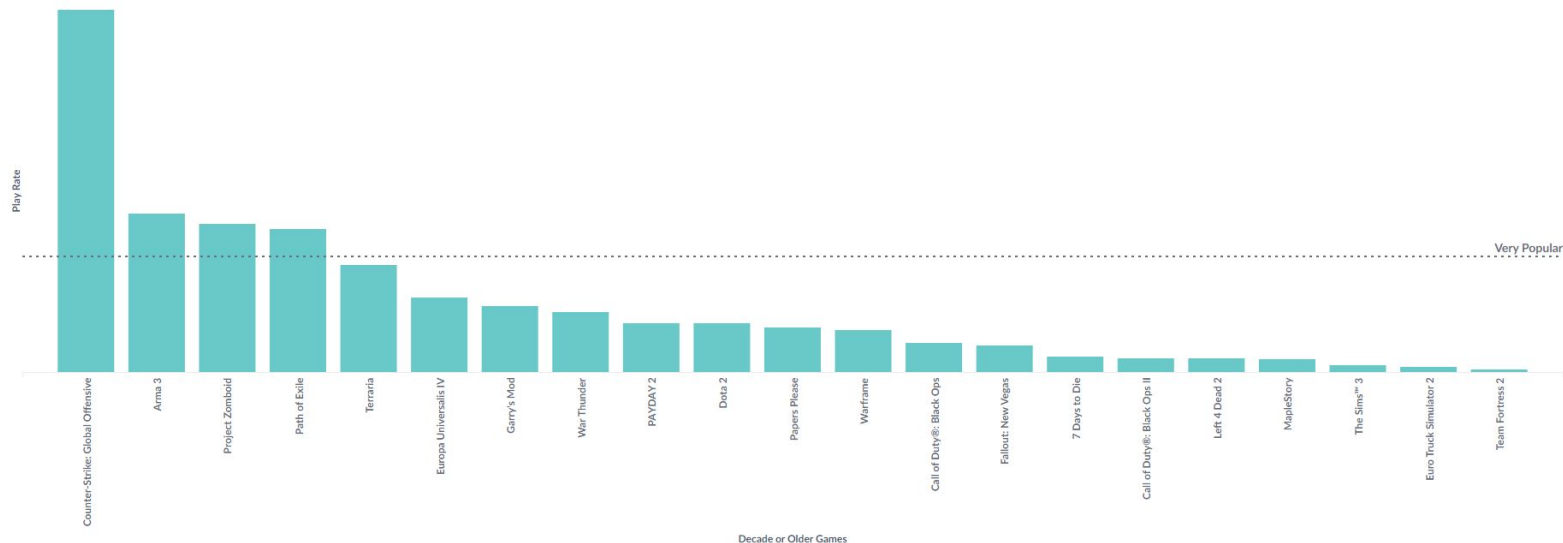**2. Koliko bi koštalo da se kupe sve igre za svaku od platformi?**

platform	total_price
Windows	5,985.89
Steam Deck	5,985.89
Mac	1,471.99
Linux	825

Legend:
- Windows
- Steam Deck
- Mac
- Linux

$14,268.77
TOTAL

42.0%
42.0%
10.3%
5.8%

Metabase Upiti - Nikola

3. Najviše recenzirane igre za svaki kvartal:
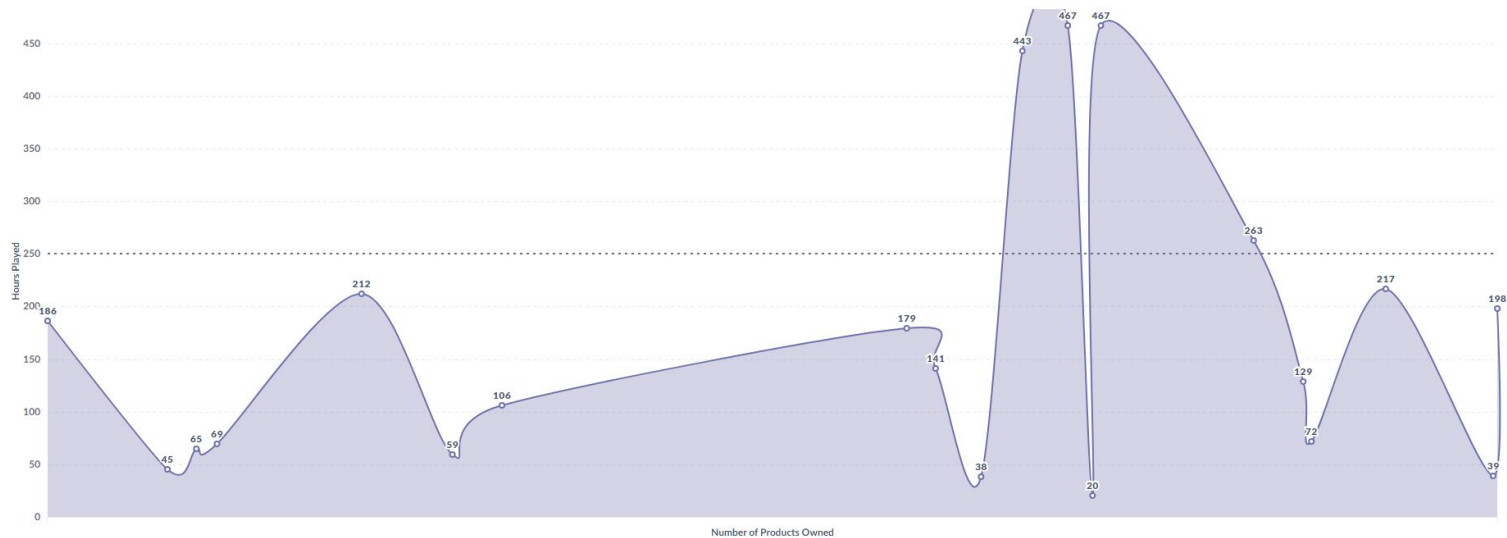
Metabase Upiti - Nikola

4. Top 20 trenutno najpopularnijih igara koje su izašle pre 2014. godine:
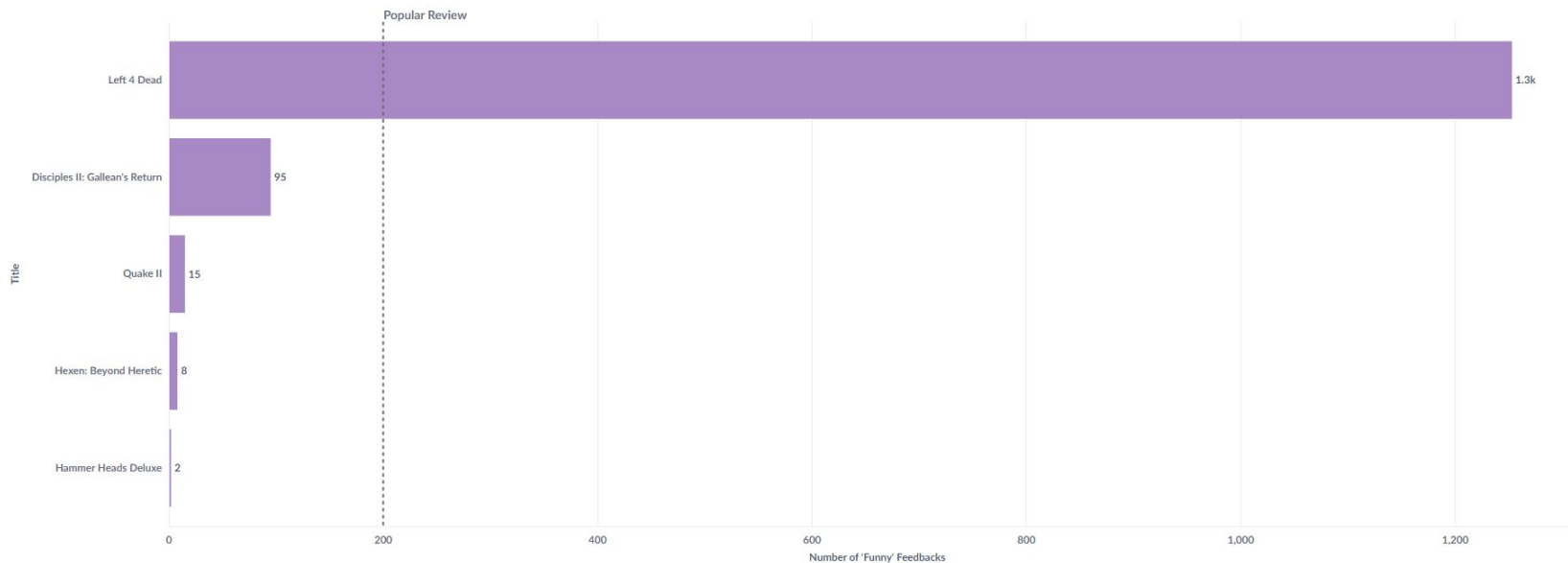

Izmenjena donja granica za bolji prikaz upita*

Metabase Upiti - Nikola

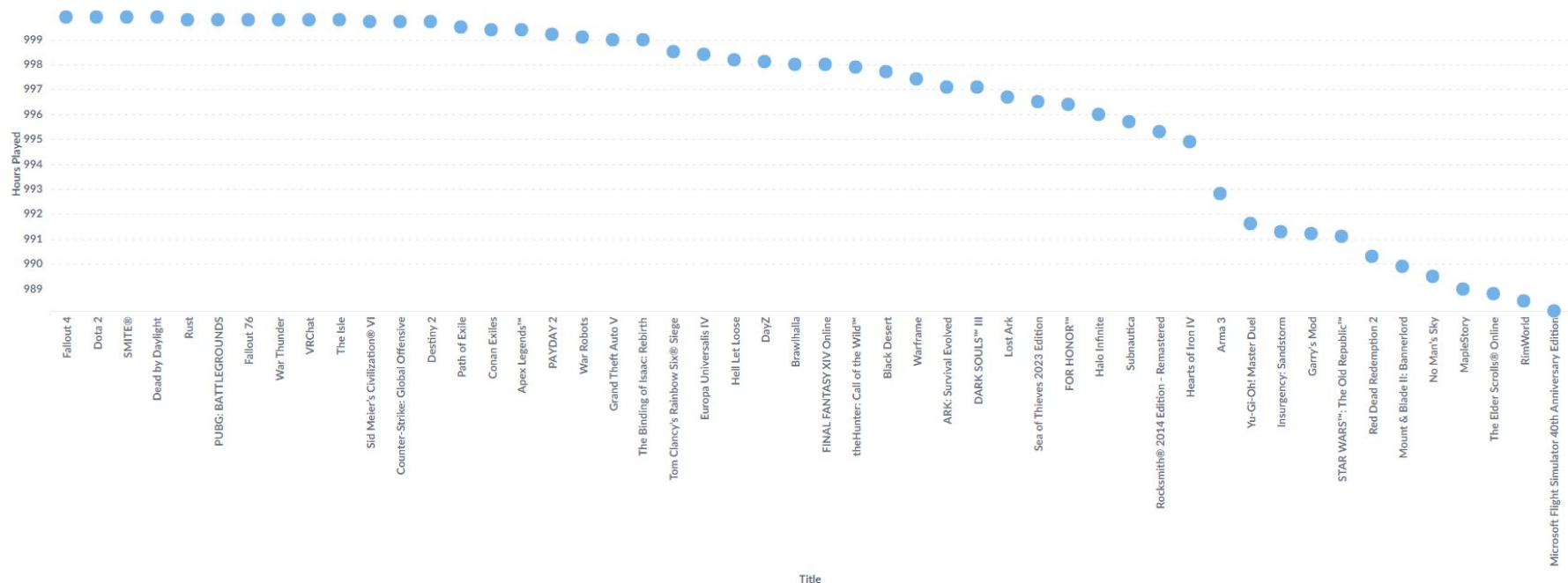**5. Koliko utiče broj posedovanih igara na broj sati u Dark Souls serijalu?**

Metabase Upiti - Srđan

1. Recenzije koje su označene kao smešne za igre sa visokim pozitivnim rejtingom na Steam Deck platformi:
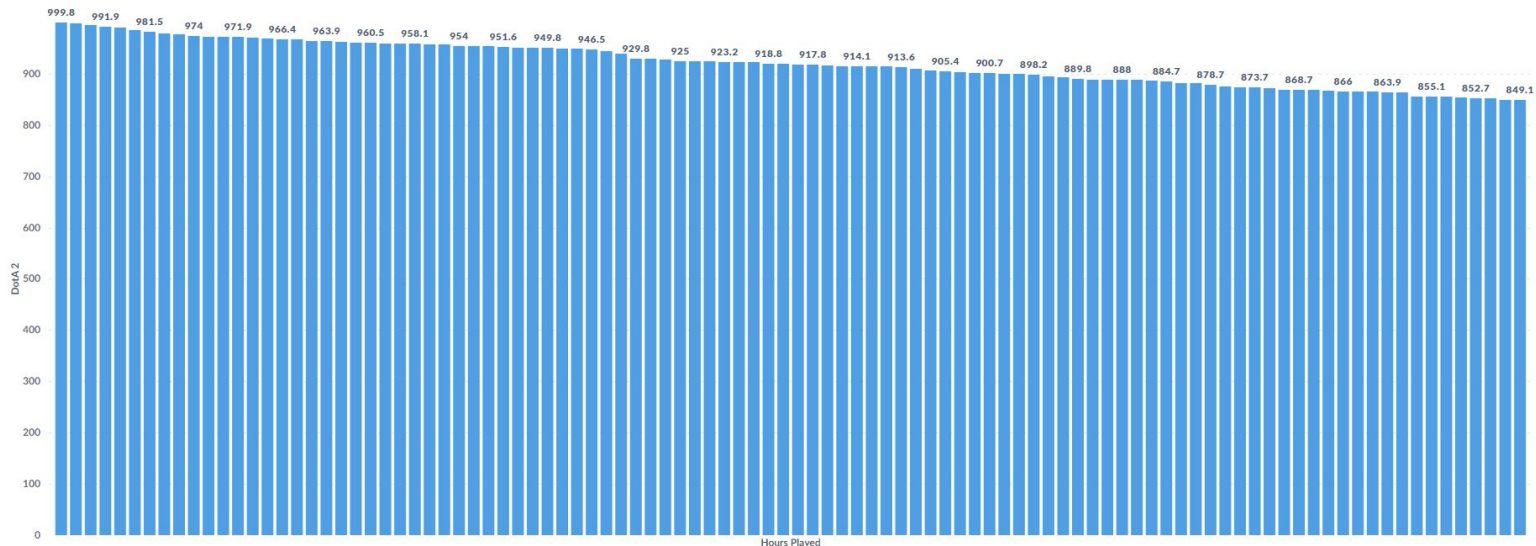
Metabase Upiti - Srđan

2. Maksimalan broj sati koje su korisnici igrali igre za koje nisu preporučili, grupišući ih po igrama.

Metabase Upiti - Srđan

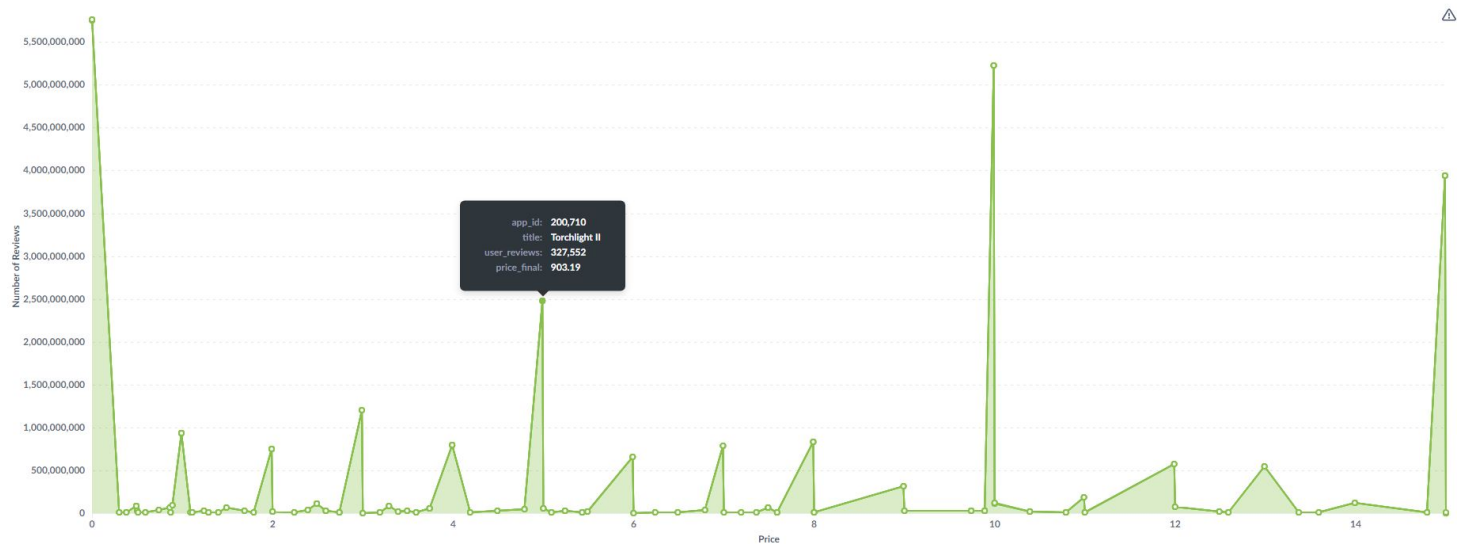**3. Najveći fanovi "Dota 2" igre na osnovu maksimalnog broja sati koje su igrali.**

Metabase Upiti - Srđan

4. Procenat igara koje imaju vrlo pozitivan rejting, uz prosečnu cenu i prosečne korisničke recenzije.

rating	Very Positive
percentage	100
average_price	10.18
average_reviews	4,082.43

Metabase Upiti - Srđan

5. Najpopularnije igre na platformama Mac i Linux sa cenom manjom od 15 dolara, sortirane po broju korisničkih recenzija.

Hvala!