

Uticaj GDP-a na uspesnost na Ol

Autor:

Data Analyst kurs 2024

1. Uvod

Olimpijske igre su najveći svetski sportski događaj, na kojem učestvuju zemlje sa ciljem demonstracije sportskih veština i nacionalnog ponosa. Uspeh na ovom takmičenju nije samo rezultat sportskih veština, već i šireg društvenog i ekonomskog konteksta. Razumevanje kako ekonomski pokazatelji utiču na sportske rezultate može pružiti važne uvide o ulozi ekonomskih resursa u postizanju olimpijskog uspeha.

Kratak opis cilja analize:

Ovaj izveštaj analizira učinak zemalja na Olimpijadi 2024. godine, istražujući vezu između ekonomskih pokazatelja zemalja (kao što su bruto domaći proizvod – GDP, populacija i GDP po glavi stanovnika) i njihovih sportskih performansi, izraženih kroz broj osvojenih medalja. Takođe, analiza se fokusira na razliku između "skupljih" i "jeftinijih" sportova, analizirajući kako ekonomski resursi utiču na uspeh u različitim sportovima.

Ključna pitanja analize:

- Da li zemlje sa višim GDP-om osvajaju više medalja?
- Da li su sportovi sa većim kapitalnim ulaganjima (skupi sportovi) koncentrisani u ekonomski jakim zemljama (visok GDP po glavi stanovnika), dok su jeftini sportovi uspešni u ekonomski slabijim zemljama?

Ovaj izveštaj će istražiti navedena pitanja analizom medalja i ekonomskih pokazatelja zemalja, kao i identifikovanjem povezanosti između resursa i uspeha na Olimpijadi.

2. Pregled podataka

2.1. Opis dataset-a

Za potrebe analize, korišćena su dva osnovna skupa podataka:

1. Podaci o ekonomskim pokazateljima zemalja:

- Ovaj skup podataka sadrži informacije o bruto domaćem proizvodu (GDP), populaciji i GDP-u po glavi stanovnika za 177 zemalja.
- Podaci su preuzeti sa sajta Worldometers, web izvora ekonomskih i demografskih statistika.
- Ključne kolone uključuju:
 - **country**: Naziv zemlje.
 - **gdp**: Ukupna vrednost GDP-a (u američkim dolarima).
 - **population**: Broj stanovnika zemlje.
 - **gdp_per_capita**: GDP po glavi stanovnika (u američkim dolarima).

2. Podaci o osvajačima medalja:

- Ovaj skup podataka sadrži informacije o osvajačima medalja na Olimpijskim igrama 2024. godine, uključujući detalje o medaljama, disciplinama i zemljama.
- Podaci su preuzeti sa stranice Kaggle.
- Glavne kolone uključuju:
 - **medal_type**: Tip medalje (zlato, srebro, bronza).
 - **name**: Ime sportiste ukoliko je pojedinačan sport u pitanju ili naziv države ukoliko je grupni.
 - **discipline**: Sport u kome se sportista takmicio.
 - **country**: Zemlja sportiste.
 - **event**: Disciplina u kojoj se sportista takmicio.

3. Lista "skupih" i "jeftinih" sportova

- Ove liste su napravljene na osnovu opštog znanja o sportovima.
- Faktori za klasifikaciju su bili: troškovi opreme i infrastrukture, pristupačnost i dugoročni troškovi obuke.
- U skupe sportove spadaju: golf, jedrenje, konjički sportovi, tenis, biciklizam
- U jeftine sportove spadaju: atletika, boks, džudo, tekvondo i rvanje

2.2. Obrada podataka

Kako bi se osigurala tačnost i upotrebljivost podataka, sprovedeni su sledeći koraci obrade:

1. Provera nedostajućih vrednosti:

- U skupu podataka o zemljama nije pronađeno nijedno polje sa nedostajućim vrednostima.

- U skupu podataka o medaljama identifikovano je nekoliko nedostajućih vrednosti u kolonama `medal_code` i `url_event`. Kolona `medal_code` je popunjena na osnovu kolone `medal_type` dok je kolona `url_event` ignorisana jer nije ključna za analizu.
- Čišćenje i standardizacija:**
 - GDP vrednosti su konvertovane iz string formata (sa znakovima poput "\$" i ",") u numeričke vrednosti za lakšu analizu.
 - Nazivi zemalja su standardizovani kako bi se omogućilo povezivanje skupova podataka.
 - Zemlje koje su osvajaci medalja, a imaju upitan status po politickom pitanju i ne nalaze na listi zemalja na sajtu Worldometers su isključene iz analize.
 - Grupisanje podataka:** skup podataka o medaljama je grupisan po drzavama, da bi se video ukupan broj medalja, a kasnije po disciplini da bi se video prosečan GDP osvajaca medalja u odredjenom sportu.
 - Spajanje podataka:**
 - Oba skupa podatka su povezana na osnovu kolone `country`, čime je omogućena analiza odnosa između ekonomskih pokazatelja i broja osvojenih medalja.

2.3. Primer podataka

Podaci o zemljama:

country	gdp	population	gdp_per_capita
United States	\$25,462,700,000,000	341,534,046	\$74,554
China	\$17,963,200,000,000	1,425,179,569	\$12,604

Podaci o medaljama:

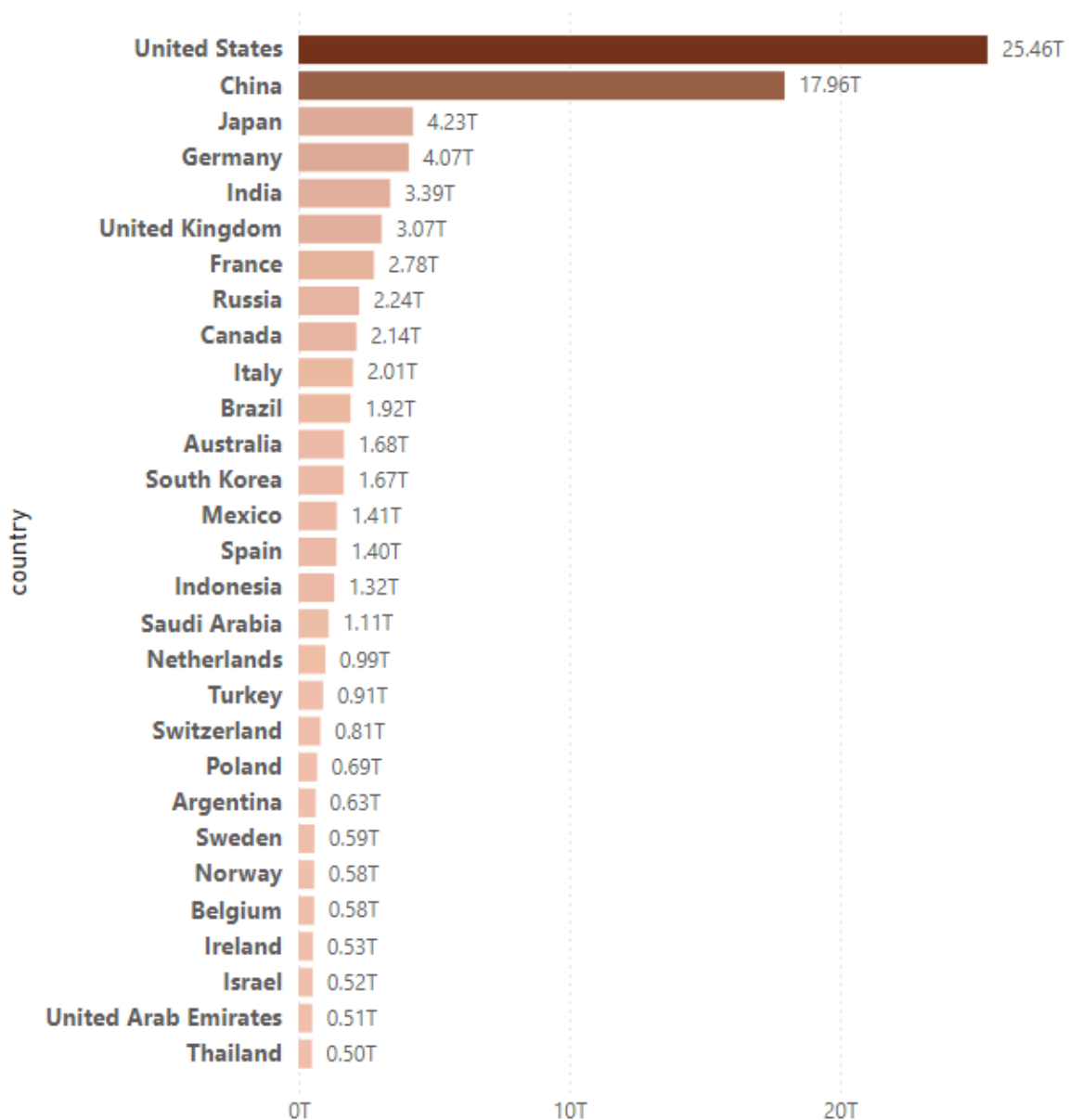
medal_type	name	discipline	country	medal_date
Gold Medal	Remco Evenepoel	Cycling Road	Belgium	2024-07-27
Silver Medal	Filippo Ganna	Cycling Road	Italy	2024-07-27

3. Analiza podataka

3.1. Analiza zemalja i medalja pojedinačno

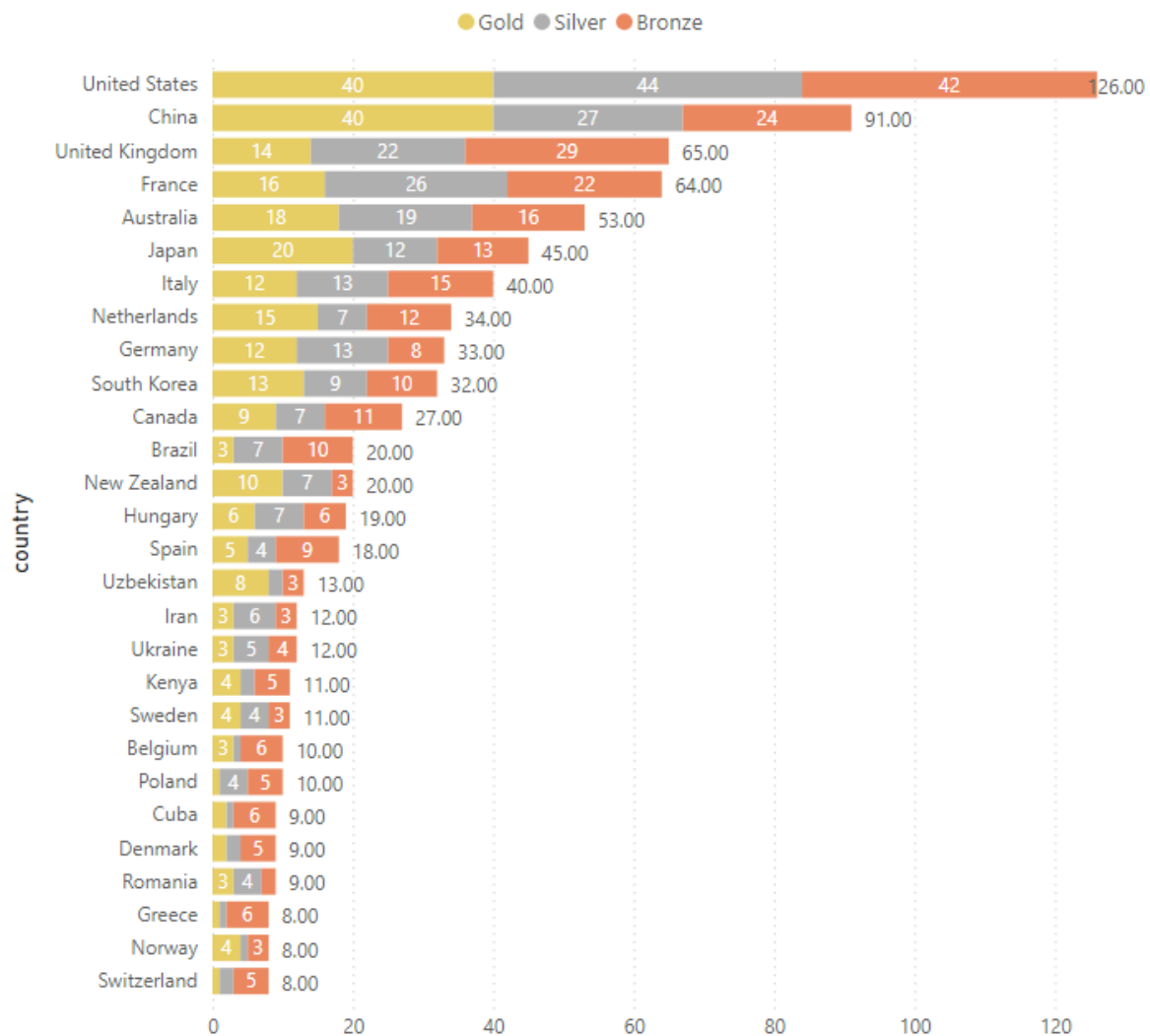
Graficki prikaz vrednosti GDP-a po zamljama. Prema ovom grafiku moze se uvideti da Kina i Sjedinjene americkie drzave imaju daleko vecu vrednost GDP-a u odnosu na sve ostale zemlje.

Gdp by Country



Na slici ispod je graficki prikaz medalja po drzavama (zlatona medalja, srebrna medalja, bronzana medalja, kao i ukupan broj medalja. Prema ovom grafiku, takodje vidimo da Kina i Amerika dominiraju u ukupnom broju medalja.

Medals by Country

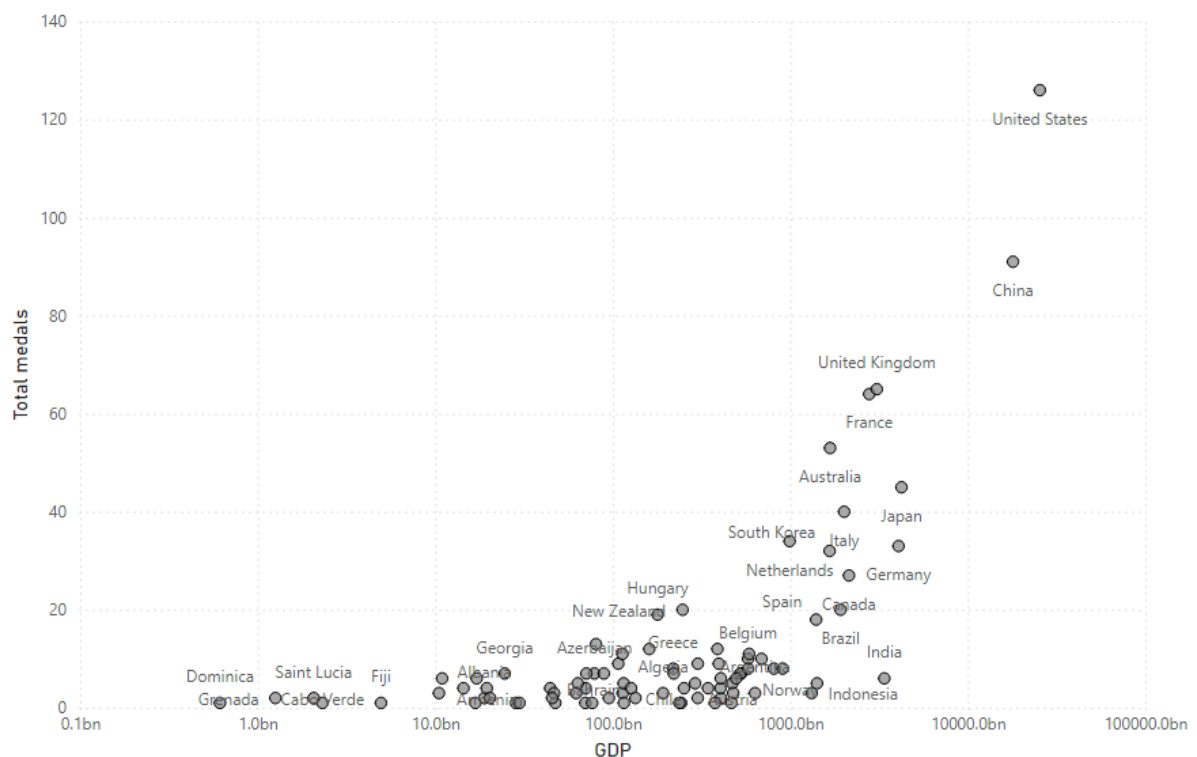


3.2. Korelacija između GDP-a i broja osvojenih medalja

- Koeficijenta korelacije između broja osvojenih medalja i populacije iznosi: **0.39**.
- Koeficijenta korelacije između broja osvojenih medalja i GDP-a po stanovniku iznosi: **0.34**.
- Koeficijenta korelacije između broja osvojenih medalja i GDP-a: **0.85**.

Kako GDP kao mera objedinjuje i populaciju i ekonomiju jedne zemlje, samo nju ćemo koristiti u ovom delu istraživanja. Prikazani scatter plot pokazuje odnos između broja osvojenih medalja u odnosu na GDP. Zaključak je da zemlje sa višim GDP-om imaju veći broj medalja.

GDP vs Total medals

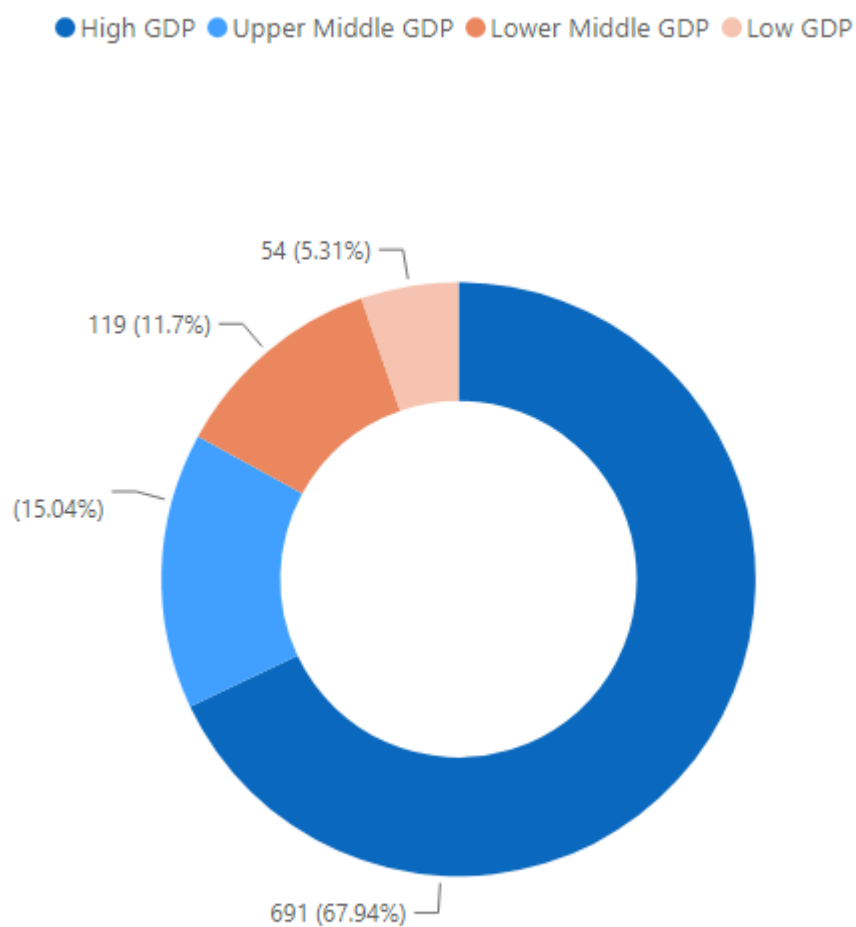


3.3. Kategorizacija države po GDP-u i ukupan broj medalja po kategoriji

Na grafiku je prikazana raspodela medalja prema GDP kategoriji kojoj pripadaju države. Kategorije su napravljene prema državama osvajacima medalja i to:

- kategorija 'Low GDP' su pripale države koje imaju GDP niži od $6 \cdot 10^{10}$
- kategorija 'Lower Middle GDP', za GDP niži od $2 \cdot 10^{11}$
- kategorija 'Upper Middle GDP', za GDP niži od $6 \cdot 10^{11}$
- kategorija 'High GDP', za sve ostale

Total Medals by GDP Category



3.4. Prosecan GDP po glavi stanovnika osvajaca medalja po sportu

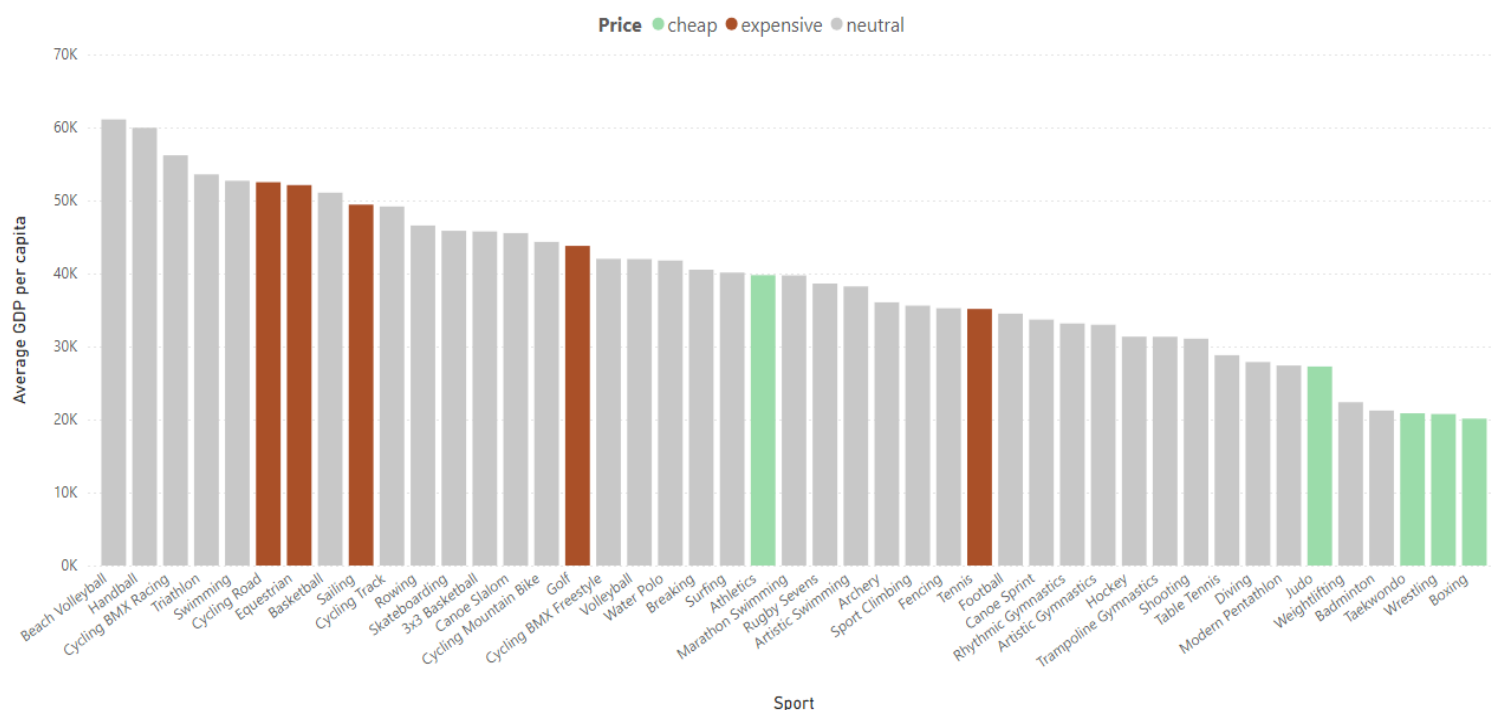
U ovoj sekciji analiziramo povezanost između tipova sportova i ekonomskih resursa zemalja. Imamo dve grupe:

- **Skupi sportovi:** Sportovi koji zahtevaju visoka ulaganja u infrastrukturu i opremu, kao što su golf, tenis, jedrenje, biciklizam i konjični sport
- **Jeftini sportovi:** Sportovi koji zahtevaju manja ulaganja, kao što su trčanje, boks i borilačke veštine.

Ostali sportovi ce ostati neutralni.

Na grafiku su prikazani sportovi i prosečna vrednost GDP-a po glavi stanovnika od osvajaca medalja u datom sportu. Crvenom bojom su prikazani “skupi” sportovi, dok su zelenom “jeftini”. Ovaj grafik pokazuje da postoji određena zavisnost između datih promenljivih, ali nije isključiva kao što vidimo na primeru tenisa i atletike. Ono što je značajno je da su borilački sportovi pri samom dnu lestvice..

Average of GDP per Capita by Sport



4. Ključni nalazi

- Zemlje sa visokim GDP imaju tendenciju da osvajaju više medaljana Olimpijskim igrama.
- Zemlje sa nižim GDP-om, imaju niži broj osvojenih medalja.
- Prosečan GDP po stanovniku osvajača medalja po sportu pokazuje da **neki** od “skupih” sportovi često donose medalje ekonomski jakim zemljama, dok su **neki** od jeftiniji sportova (borilačke veštine) popularni u ekonomski slabijim zemljama.

5. Zaključci i preporuke

- **Za velike ekonomije:**
 - Nastaviti sa masovnim ulaganjima u sportsku infrastrukturu i programe.
- **Za male ekonomije:**
 - Fokusirati se na specijalizovane sportske discipline gde imaju konkurentsku prednost.
- **Za globalne organizacije:**
 - Podsticanje ravnoteže u finansiranju sporta kako bi manje razvijene zemlje dobile priliku za uspeh.

6. Prilozi

Link ka Power BI dashboardu