**CHATGPT u pronalasku podataka**

Pri radu na aplikaciji za filtriranje i rangiranje sveučilišta susreli smo se s problemom teškoće pronalaska podataka te manjkom podataka. Sami smo pronašli podatke za 120 sveučilišta, prvo smo odlučili provjeriti s CHATGPT-om kakve će on podatke uspjeti pronaći te koliko će se razlikovati od naših podataka. Ako bismo se uvjerili da stavlja poprilično točne podatke proširili bismo naš skup podataka na 1000 sveučilišta generiranih od njega i na temelju tih podataka odradili filtriranje i rangiranje sveučilišta.

Na početku pokušaj da CHATGPT odmah iz generira cijelu tablicu (u promptu su opisani svi potrebni parametri te države koje nas zanimaju):

1. **PROMPT**

I am working on an application in which I need accurate dataset containing universities with parameters that consist of:

* university (name of the university)
* collegeRank (global ranking of university)
* tuition (EUR/year) (how much is paid each year to study at the university, price needs to be in euros)
* percOfIntStud (percentage of international students based on all students)
* acceptance rate (how many people manage to get into the university in percentage)
* avgSafetyIndex (how secure is the country, how many crimes they have, and based on that index in the range from 0 to 100)
* cost of living (this consists of how much monthly persons pays on groceries, restaurants, transportation, and utilities excluding rent, price needs to be in euros)
* rent (monthly average for rent in euros)
* groceries (monthly average for groceries in euros)
* recreationCost (EUR/month) (monthly average for recreation and culture in euros)
* healthcare price (average out of the pocket spend on healthcare per year in euros) –
* avgMntTransportCost (average transportaion price in euros per month)
* link (link for the university)
* and each of the columns: Computer Science, Business, Economics, Psychology, Biology, Law, Medicine, Mathematics, Art, Physics they have values 1 or 0 based on the fact if university has these majors

The parameters avgSafetyIndex, cost of living, rent, groceries, recreationCost, healthcare price, avgMntTransportCost are based on the city in which university is located.

I want you to find me data for these countries: …

Albania

Algeria

Argentina

Armenia

Australia

Austria

Azerbaijan

Bahamas

Bahrain

Bangladesh

Barbados

Belarus

Belgium

Bolivia

Bosnia And Herzegovina

Botswana

Brazil

Bulgaria

Cambodia

Cameroon

Canada

Chile

China

Colombia

Costa Rica

Croatia

Cuba

Cyprus

Czech Republic

Denmark

Dominican Republic

Ecuador

Egypt

El Salvador

Estonia

Fiji

Finland

France

Georgia

Germany

Ghana

Greece

Guatemala

Hong Kong

Hungary

Iceland

India

Indonesia

Iran

Iraq

Ireland

Israel

Italy

Jamaica

Japan

Jordan

Kazakhstan

Kenya

Kosovo

Kuwait

Kyrgyzstan

Latvia

Lebanon

Libya

Lithuania

Luxembourg

Malaysia

Malta

Mauritius

Mexico

Moldova

Montenegro

Morocco

Nepal

Netherlands

New Zealand

Nigeria

North Macedonia

Norway

Oman

Pakistan

Palestine

Panama

Paraguay

Peru

Philippines

Poland

Portugal

Puerto Rico

Qatar

Romania

Russia

Saudi Arabia

Serbia

Singapore

Slovakia

Slovenia

South Africa

South Korea

Spain

Sri Lanka

Sweden

Switzerland

Syria

Taiwan

Tanzania

Thailand

Trinidad And Tobago

Tunisia

Turkey

Uganda

Ukraine

United Arab Emirates

United Kingdom

United States

Uruguay

Uzbekistan

Venezuela

Vietnam

Zimbabwe

For each country find the best university based on ranking and based on chosen university for each country write down other parameters for it.

Make me an excel file for this dataset.

1. **ODGOVOR**

Slika na kojoj se prikazuje tekst, Font, snimka zaslona, pismo

Opis je automatski generiran

U prvom odgovoru ne odgovara s popunjenom tablicom već navodi izvore u kojima trebamo potražiti navedene podatke.

Također postavlja brojne linkove:

* <https://www.usnews.com/education/best-colleges/slideshows/universities-with-the-highest-international-student-acceptance-rates>
* <https://www.numbeo.com/cost-of-living/>
* <https://www.topuniversities.com/student-info/student-finance/tuition-fees-worlds-top-universities>
* <https://www.topuniversities.com/university-rankings-articles/world-university-rankings/top-universities-international-students>
* <https://terratern.com/blog/most-expensive-countries-in-the-world/>
* <https://www.worldtrips.com/resources/affordable-us-colleges-for-international-students>
* <https://fastercapital.com/content/Cost-of-Living--Cost-of-Living-Index-and-Factors-for-Different-Countries-and-Cities.html>
* <https://fastercapital.com/topics/comparing-the-cost-of-living-across-countries.html/1>
* <https://www.shiksha.com/studyabroad/universities-with-the-highest-percentage-of-international-students-articlepage-184784>
* <https://www.globalpassport.ai/blog/costa-rica-cost-of-living-expert-guide-2024>
* <https://www.timeshighereducation.com/student/best-universities/international-student-table-2018-top-200-universities>
* <https://housinganywhere.com/Spain/cost-of-living-in-spain-compared-to-the-world>
* <https://educationdata.org/average-cost-of-college-by-country>
* <https://www.movingto.io/it/cost-of-living-italy>
* <https://www.niche.com/colleges/search/top-colleges-for-international-students/>
* <https://www.gooverseas.com/blog/10-universities-where-you-can-study-abroad-in-english>
* <https://www.dewittmove.com/cost-of-living-in-switzerland/>
* <https://www.statista.com/statistics/297132/top-host-destination-of-international-students-worldwide/>
* <https://internationalliving.com/countries/spain/cost-of-living-in-spain/>

Najviše se usredotočio na traženje izvora vezanih uz troškove života, a od država se usredotočio na Costa Ricu, Španjolsku, Italiju i Švicarsku.

1. **PROMPT**

**Slika na kojoj se prikazuje tekst, snimka zaslona, Font, dokument

Opis je automatski generiran**

Šalje praznu Excel tablicu s gore napisanim parametrima.

1. **PROMPT**

**Slika na kojoj se prikazuje tekst, snimka zaslona, Font, dokument

Opis je automatski generiran**

Ništa ne šalje samo govori da će započeti proces prikupljanja podataka.

1. PROMPT

Slika na kojoj se prikazuje tekst, snimka zaslona, Font

Opis je automatski generiran

1. PROMPT

Slika na kojoj se prikazuje tekst, Font, snimka zaslona, dokument

Opis je automatski generiran

1. PROMPT

Slika na kojoj se prikazuje tekst, snimka zaslona, Font

Opis je automatski generiran

1. **KONVERZACIJA**

Nakon još nekoliko neuspješnih promptova u kojima je samo odgovarao kako će sakupljati podatke te da već skuplja podatke. Na prompt šta je do sada uspio napraviti šalje Excel tablicu University\_Dataset\_Partial u kojoj se nalazi svega 5 države koje su: Australija, Austrija, Kanada, Njemačka i Japan. Iako je prije slao linkove za države Costa Ricu, Španjolsku, Italiju i Švicarsku za njih se na kraju nije odlučio pri biranju prvih 5 država za obradu.

Podaci koji se nalaze u tablici su uvelike slični našima osim kod parametra safety indeks gdje on stavlja poprilično velike brojeve. Svi njegovi brojevi su zaokruženi (na primjer za Austriju za University of Vienna smo pronašli da je školarina 1453 eura, on to zaokružuje i stavlja na 1500 eura). Također linkovi koji predstavljaju odlazak na stranicu pojedinog sveučilišta uspješno rade.

1. **KONVERZACIJA**

Slika na kojoj se prikazuje tekst, snimka zaslona, Font, broj

Opis je automatski generiran

Uviđamo da mu je cijela tablica preveliki zadatak jer govori „I'll notify you“ što CHATGPT nije u mogućnosti raditi.

Nakon prompta gdje ga ispitujemo za cijelu tablicu uviđamo da očito nije u mogućnosti odmah pronaći sve podatke pa smanjujemo naš zahtjev na 20 država. Nije u mogućnosti pronaći odmah za 20 država nego samo šalje za 2, a šalje tablicu naziva University\_Dataset\_20\_Countries iako se u tablici nalazi samo 7 država sveukupno (uz početnih 5 dodaje Ujedinjeno Kraljevstvo i Indiju).

PROMPT:

Slika na kojoj se prikazuje tekst, snimka zaslona, Font

Opis je automatski generiran

Nakon što smo uvidjeli da nije u mogućnosti prikazati u tablici sveučilišta, odlučujemo se isprobati kako će odgovarati ako ga pitamo da odgovori s tekstom umjesto s tablicom. U ovom slučaju odmah postavlja traženih 5 sveučilišta sa svim parametrima. Nastavljamo s ovim načinom pokušavajući povećati broj sveučilišta po odgovoru na 20, ali on još uvijek vraća 10 sveučilišta tako da onda tako nastavljamo dok ne dođemo do svih država.

Slika na kojoj se prikazuje tekst, snimka zaslona, Font, dokument

Opis je automatski generiran

Provjera od kuda je izvukao podatke, ponavlja izvore ista kao i s početka konverzacije.

**KRAJ**

U krajnjoj tablici nedostaju države: Albania, Algeria, Argentina, Armenia, Azerbaijan, Bahamas, (Cuba) za njih je potrebno poslati još jedan prompt. Države koje je dodao bez da su rečene su: Yemen, Zambia.

KREIRANJE 1000 SVEUČILIŠTA

Uvidjevši da je uspješno generirao podatke za početnih 120 država, kako bismo napunili tablicu nastavljamo dalje sa sveučilištima, ali ovaj puta po ranking. Problem kod rankinga je što je uzimao ranking za različith stranica pa se neka sveučilišta pojavljuju više puta. Najviše ponovljena sveučilišta su bila:

University of Cape Town (5)

University of Vienna (5)

University of Antwerp (5)

University of Oslo (5)

University of Porto (5

University of Pisa (6)

University of Warsaw (5)

University of Buenos Aires (5)

KoĂ§ University (5)

EĂ¶tvĂ¶s LorĂˇnd University (5)

To je bilo do rankinga 350. Uvidjevši da puno ponavlja sveučilišta odlučujemo mu dati izvor od kojeg će uzimati rankinge sveučilišta:

Slika na kojoj se prikazuje tekst, snimka zaslona, Font, broj

Opis je automatski generiran

Slika na kojoj se prikazuje tekst, snimka zaslona, broj, paralelno

Opis je automatski generiran

Daje u potpunosti različita sveučilišta od onih koja su navedena na stranici. Misleći da možda nije dovoljno dobro specificiran prompt (nedovoljno specifirano šta označava 350) ponavljamo ga s dodatkom da pogleda prema world rank koji je prvi parameter u tablici.

Slika na kojoj se prikazuje tekst, snimka zaslona, broj, Font

Opis je automatski generiran



Opet nije dobro posložio sveučilišta, čak je i u odnosu na prethodni prompt drugačije posložio sveučilišta i dalje krivo. Iako je krivo posložio sveučilišta ona i dalje postoje na danoj stranici.

Kreiraju se sveučilišta prema rankinga, ali zbog nekonzistentnosti podataka koje pronalazi za ranking sveučilišta, chatgpt određena sveučilišta ponavlja više od jednom, ona koja su najviše ponovljena su:

University of Toronto (3), rankovi: 24,

University of Illinois Urbana-Champaign (3), rankovi: 42,

University of Tokyo (3), rankovi: 43,

University of Zurich (4), rankovi: 55, 73,

University of Helsinki (3), rankovi: 57,

University of Cape Town (5), rankovi: 60, 86, 113

University of Vienna (5), rankovi: 61, 82, 103, 117, 137

University of Adelaide (4), rankovi: 64, 83, 104, 248

University of Geneva (3), rankovi: 65

University of Groningen (3), rankovi: 66

University of Liverpool (3), rankovi: 68,

University of Calgary (4), rankovi: 69, 105, 121, 532

University of Antwerp (5), rankovi: 70, 108, 125, 204, 291, 361

University of Ottawa (4), rankovi: 71, 110, 352, 458

University of Cologne (4), rankovi: 72, 114

University of Oslo (5), rankovi: 74, 97, 118, 152,

University of Gothenburg (3), rankovi: 75,

University of Porto (5), rankovi: 78, 147, 180, 199

University of Barcelona (3), rankovi: 81

University of Sao Paulo (3), rankovi: 88, 156, 188

University of St Andrews (3), rankovi: 93

University of Reading (4), rankovi: 112, 249,

University of Pisa (6), rankovi: 122, 184, 201

University of Warsaw (5), rankovi: 123, 167, 187,

University of Buenos Aires (5), rankovi: 131, 150, 164, 179, 195

University of Pretoria (3), rankovi: 132,

University of Malaya (3), rankovi: 133,

University of Tartu (3), rankovi: 134,

KoĂ§ University (5), rankovi: 148, 168, 189, 224,

Charles University (3), rankovi: 151,

EĂ¶tvĂ¶s LorĂˇnd University (5), rankovi: 153, 171, 186, 225

National and Kapodistrian University of Athens (3), rankovi: 154,

Hebrew University of Jerusalem (3), rankovi: 155,

Chulalongkorn University (3), rankovi: 172,

Masaryk University (3), rankovi: 203, 221

Federal University of Minas Gerais (3), rankovi: 208, 306, 617

University of Miami (3), rankovi: 247,

University of Delaware (3), rankovi: