

Inlämningsuppgift i kursen Programmera mera i Python A266TG februari 2023.

Syfte och mål

Syftet med inlämningsuppgiften är att du ska visa, inte minst för dig själv, att du kan lösa ett antal uppgifter med hjälp av Python och några av dess tilläggsmoduler och där du får använda en del av de kunskaper som du fått i denna kurs. Målet med inlämningsuppgiften är att du ska få en insikt i att Python och (ett par) av dess tilläggsmoduler erbjuder ett kraftfullt verktyg för databehandling och presentation av analyserad data.

Förutsättningar:

Uppgiften består av fem uppgifter. Varje uppgift kommer att bedömas var för sig och för att få betyget godkänt måste alla deluppgifterna vara godkända. En mera utförlig information om det administrativa kring inlämningsuppgiften finns att läsa på kursens hemsida i Canvas.

OBS!

Uppgifterna ska lösas med hjälp av tilläggsmodulerna *Pandas* och *matplotlib.pyplot*.

Några enkla regler som skall följas:

- koden ska vara genomtänkt, lättläst och välkommenterad.
- skriv varför en kodrad finns och inte vad den gör.
- variabler ska ha meningsfulla namn.
- du får inte lämna in kod som genererar felmeddelande eller varningar. Då blir det omedelbart retur.
- du får inte lämna in bortkommenterad kod.

För att få en hanterbar och smidig rättning måste ovanstående regler följas. (Grova) avsteg från dessa regler innebär retur.

Tänk på att innehållet i csv-filen kan importeras till Excel eller ett motsvarande kalkylprogram. Där kan du analysera data och kontrollera att lösningen på dina uppgifter är korrekta. Gör gärna denna kontroll innan du lämnar in inlämningsuppgiften så att du slipper få retur på grund av felaktiga resultat i dina lösningar. Dina lösningar skriver du i den svarsfilen som finns att ladda ned på samma ställe som detta PM. Eftersom att kamratgranskningen är anonym bör du undvika att ha ditt namn

med med i filnamnet på svarsfilen. Rättningen av dina uppgifter kommer att göras i senaste Spyder IDE-version.

Uppgiftsbeskrivning

I denna inlämningsuppgift ska du med hjälp av modulerna *Pandas* och *matplotlib.pyplot* analysera och vidareförädla data från amerikanska Central Intelligence Agency (CIA) och från SimpleMaps. CIA tillhandahåller en referensdatabas - CIA World Factbook (<https://www.cia.gov/the-world-factbook/>) - där det finns en mängd information som fritt kan användas. Samma gäller för SimpleMaps som också tillhandahåller en mängd information som fritt kan användas (se <https://simplemaps.com/data/world-cities>).

Nedan följer en kort beskrivning av innehållet i de filer som kommer att användas. Ladda gärna in dessa csv-filen i Excel eller ett motsvarande program och bekanta dig med innehållet innan du börjar att lösa uppgifterna.

cia_factbook.csv:

Denna csv-fil innehåller information från World Factbook om olika länders befolkning, area, ekonomi, geografi, internationella relationer med mera.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Index	country	area	birth_rate	death_rate	infant_mortality	internet_usage	life_expectancy	maternal_mortality	net_migration	population	population_growth_rate		
2	1	Russia	17098242	11.87	13.83	7.08	40853000	70.16	34	1.69	1.42E+08	-0.03		
3	2	Canada	9984670	10.29	8.31	4.71	26960000	81.67	12	5.66	34834841	0.76		
4	3	United States	9826675	13.42	8.15	6.17	2.45E+08	79.56	21	2.45	3.19E+08	0.77		
5	4	China	9596960	12.17	7.44	14.79	3.89E+08	75.15	37	-0.32	1.36E+09	0.44		
6	5	Brazil	8514877	14.72	6.54	19.21	75982000	73.28	56	-0.15	2.03E+08	0.8		
7	6	Australia	7741220	12.19	7.07	4.43	15810000	82.07	7	5.74	22507617	1.09		
8	7	India	3287263	19.89	7.35	43.19	61338000	67.8	200	-0.05	1.24E+09	1.25		
9	8	Argentina	2780400	16.88	7.34	9.96	13694000	77.51	77	0	43024374	0.95		
10	9	Kazakhstan	2724900	19.61	8.31	21.61	5299000	70.24	51	0.42	17948816	1.17		
11	10	Algeria	2381741	23.99	4.31	21.76	NA	76.39	97	-0.93	38813722	1.88		
12	11	Congo, Dem.	2344858	35.62	10.3	73.15	290000	56.54	540	-0.33	77433744	2.5		
13	12	Greenland	2166086	14.53	8.38	9.42	36000	71.82	NA	-5.98	57728	0.02		
14	13	Saudi Arabia	2149690	18.78	3.32	14.58	9774000	74.82	24	-0.59	27345986	1.49		

worldcities.csv

Denna csv-fil innehåller extraherad information från databasen World Cities Database från SimpleMaps och innehåller data om världens städer (se <https://simplemaps.com/data/world-cities>).

	A	B	C	D	E	F	G	H	I	J	K
1	city	city_ascii	lat	lng	country	iso2	iso3	admin_na	capital	populatio	id
2	Tokyo	Tokyo	356.839	1.397.744	Japan	JP	JPN	TÅkyÅ	primary	39105000	1.39E+09
3	Jakarta	Jakarta	-62.146	1.068.451	Indonesia	ID	IDN	Jakarta	primary	35362000	1.36E+09
4	Delhi	Delhi	286.667	772.167	India	IN	IND	Delhi	admin	31870000	1.36E+09
5	Manila	Manila	146.000	1.209.833	Philippine	PH	PHL	Manila	primary	23971000	1.61E+09
6	SÃ£o Paul	Sao Paulo	-235.504	-466.339	Brazil	BR	BRA	SÃ£o Paul	admin	22495000	1.08E+09
7	Seoul	Seoul	375.600	1.269.900	South Kor	KR	KOR	Seoul	primary	22394000	1.41E+09
8	Mumbai	Mumbai	190.758	728.775	India	IN	IND	MahÃrÃsh	admin	22186000	1.36E+09
9	Shanghai	Shanghai	311.667	1.214.667	China	CN	CHN	Shanghai	admin	22118000	1.16E+09
10	Mexico Ci	Mexico Ci	194.333	-991.333	Mexico	MX	MEX	Ciudad de	primary	21505000	1.48E+09
11	Guangzho	Guangzho	231.288	1.132.590	China	CN	CHN	Guangdon	admin	21489000	1.16E+09
12	Cairo	Cairo	300.444	312.358	Egypt	EG	EGY	Al QÃhira	primary	19787000	1.82E+09
13	Beijing	Beijing	399.040	1.164.075	China	CN	CHN	Beijing	primary	19437000	1.16E+09
14	New York	New York	406.943	-739.249	United Sta	US	USA	New York		18713220	1.84E+09
15	KolkÃta	Kolkata	225.727	883.639	India	IN	IND	West Benj	admin	18698000	1.36E+09
16	Moscow	Moscow	557.558	376.178	Russia	RU	RUS	Moskva	primary	17693000	1.64E+09
17	Bangkok	Bangkok	137.500	1.005.167	Thailand	TH	THA	Krung The	primary	17573000	1.76E+09

worldpubind.csv

Denna csv-fil innehåller information från SimpleMaps över länders befolkningsförändringar och är extraherad från följande källor: NGIA, US Geological Survey, US Census Bureau och NASA (se <https://simplemaps.com/data/world-cities>).

worldpubind - Excel																										
Arktiv Start Infoga Sidlayout Formler Data Granska Visa Hjälp Berätta vad du vill göra																										
Klistra in Kopiera Hämta format Tecken Justering Radrytt text Centra över kolumner Allmänt Villkorstyd format Formatera som tabell Cellformat Infoga Bort Format Autosumma Fyll Sortera och filtera Sök och markera Radera Redigering																										
BN267 15993524																										
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	
1	Country	N	Country	C	Indicator	Indicator	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979
2	Aruba	ABW	Populatio	SP	POP	TC	54608	55811	56682	57475	58178	58782	59291	59522	59471	59330	59106	58816	58855	59365	60028	60715	61193	61465	61738	62006
3	Africa East	AFE	Populatio	SP	POP	TC	1.31E+08	1.34E+08	1.38E+08	1.42E+08	1.46E+08	1.54E+08	1.58E+08	1.63E+08	1.68E+08	1.72E+08	1.78E+08	1.83E+08	1.88E+08	1.94E+08	1.99E+08	2.05E+08	2.11E+08	2.17E+08	2.24E+08	2.31E+08
4	Afghanistan	AFG	Populatio	SP	POP	TC	8622466	8790140	8969047	9157465	9355514	9565147	9783147	10010030	10247780	10494489	10752971	11015857	11286753	11575305	11869879	12157386	12425267	12687301	12938862	12986369
5	Africa West	AFW	Populatio	SP	POP	TC	97256290	99314028	1.01E+08	1.04E+08	1.06E+08	1.08E+08	1.11E+08	1.13E+08	1.16E+08	1.19E+08	1.21E+08	1.24E+08	1.27E+08	1.31E+08	1.34E+08	1.38E+08	1.41E+08	1.45E+08	1.49E+08	1.53E+08
6	Angola	AGO	Populatio	SP	POP	TC	5357195	5441333	5521400	5599827	5673159	5736582	5787044	5827503	5868203	5928386	6029700	6177049	6364731	6576230	6802494	7032713	7266780	7511895	7771590	8043216
7	Albania	ALB	Populatio	SP	POP	TC	1608800	1659800	1711319	1762621	1814135	1864791	1914573	1965598	2022272	2081695	2135479	2187853	2243126	2296752	2350124	2404831	2458526	2513546	2566266	2617832
8	Andorra	AND	Populatio	SP	POP	TC	9443	10216	11014	11839	12690	13563	14546	15745	17079	18449	19860	21322	22832	24393	26003	27640	29294	30949	32574	34142
9	Arab World	ARB	Populatio	SP	POP	TC	93359407	95760348	98268683	1.01E+08	1.04E+08	1.06E+08	1.09E+08	1.12E+08	1.16E+08	1.19E+08	1.22E+08	1.26E+08	1.3E+08	1.33E+08	1.37E+08	1.42E+08	1.47E+08	1.52E+08	1.57E+08	1.62E+08
10	United Arab	ARE	Populatio	SP	POP	TC	133426	140984	148877	157006	165305	173797	182509	191404	213582	253261	298084	344513	392368	441546	491955	543394	614177	706861	805231	908452
11	Argentina	ARG	Populatio	SP	POP	TC	20349744	20680653	21020359	21364017	21708487	22053661	22403116	22757014	23112971	23472028	23842803	24223379	24612794	25020588	25449754	25875558	26290257	26713780	27146121	27584134
12	Armenia	ARM	Populatio	SP	POP	TC	1904148	1971530	2039346	2106142	2171029	2234051	2295496	2355997	2416461	2476359	2534377	2590923	2648048	2706852	2767646	2830798	2896311	2965523	3038224	3097058

Uppgifter

Viktig information innan du startar att lösa uppgifterna: Skriv programmet till samtliga uppgifter i samma svarsfil. Ange tydligt med en (eller flera) kommentarsrader där programmet till en ny uppgift startar. Efterhand som du löser uppgifterna kan du lämpligen låta de uppgifter som du inte arbetar med vara bortkommenterade.

1. Skapa DataFrame-objekten *df_cia_factbook*, *df_worldcities* och *df_worldpubind* som innehåller en kopia av innehållet i csv-filerna *cia_factbook.csv*, *worldcities.csv* och *worldpubind.csv*. Dessa DataFrame-objekt använder du sedan när du löser nedanstående uppgifter. Obs! Separationstecknet i csv-filen är semikolontecknet (;).
2. I denna uppgift ska du träna på att vidareförädla och presentera information som finns i *df_cia_factbook*.
Uppgiften går ut på att du ska bestämma befolkningstätheten för ett angivet antal länder och plotta resultatet i ett stapeldiagram. Programmet ska ha en inmatningsrad (du behöver inte lägga ned stor möda med omfattande felkontroller – vi utgår ifrån att användaren skriver in korrekt information) där man anger hur många länder som ska presenteras i stapeldiagrammet (max 10 länder) alternativt så anger man ett enskilt land som ska presenteras. Matar man in 7+ ska de sju länderna med störst befolkningstäthet plottas i stapeldiagrammet, matar du in 5- ska de fem länderna med lägst befolkningstäthet plottas osv. Programmet behöver inte kunna presentera länder som ligger utanför de tio med högst respektive lägst befolkningstäthet. Anger man Sweden som inmatning (alltså det land som står angivet i kolumnen *country*) ska Sveriges befolkningstäthet skrivas ut på skärmen. I stapeldiagrammet ska namnet på länderna stå angivna utmed x-axeln och befolkningstätheten (antal invånare/km²) på y-axeln. Befolkningstätheten som ska presenteras ska läsas från kolumnen *density* som skapas som en ny kolumn i *df_cia_factbook* som kvoten mellan kolumnen *population* och kolumnen *area*. Tyvärr så förekommer det både oändligheten (inf) och NaN i denna kolumn. Vi löser detta genom att vi bortser från de (få) länder där man inte får ett numeriskt värde för *density* genom att ta bort dessa länder ur *df_cia_factbook*.

3. I denna uppgift ska du fortsätta att träna på att extrahera, vidareförädla och presentera information som finns lagrad i *df_cia_factbook* enligt nedanstående beskrivningar. Programmet ska bestå av en meny där man väljer ett av nedanstående fyra menyalternativ med passande beskrivningar på menyalternativen (du får själv tänka ut lämpliga beskrivningar på menyalternativen som extra träning). När ett menyalternativ är utfört ska menyn åter komma upp och ett nytt alternativ ska kunna väljas osv. Menyalternativ 4 avslutar programmet.

Menyalternativ 1:

Detta menyalternativ ska skriva ut en tabell (på skärmen) med kolumnerna *country*, *area*, *birth_rate* och *life_exp_at_birth* för det land (de länder) som uppfyller följande kriterier:

- befolkningen ska vara större än medelvärdet av befolkningen över alla länder.
- *area* ska vara mindre än medelarean över alla länder.
- *birth_rate* (som är angivet i csv-filen som personer per 1000 invånare) ska ligga mellan 15 och 24.
- *life_exp_at_birth* ska vara större än 70 år.

(Länder som har NaN i någon av kolumnerna i tabellen ska inte finnas med i tabellen)

Menyalternativ 2:

I detta menyalternativ ska du undersöka internetmognaden hos befolkningen i de länder som finns i *df_cia_factbook*. Gör så här:

Skapa en ny kolumn *internet_user_density* som innehåller antalet internetanvändare per 100.000 invånare. Skapa därefter en tabell som består av kolumnerna *country*, *population* och *internet_user_density* och som innehåller de fem länder med lägst antal internetanvändare per 100.000 invånare och de fem länder med flest antal internetanvändare per 100.000 invånare.

(Länder som har NaN i kolumnen *internet_user_density* ska inte visas i tabellen)

Menyalternativ 3:

I detta menyalternativ ska vi göra en uppskattning om länderna i *df_cia_factbook* har en positiv eller negativ befolkningsförändring uttryckt som procent av folkmängden för det år som beräkningen omfattar.

Netto befolkningsförändring per 1000 invånare får vi genom beräkningen

$population_growth_rate = (birth_rate - death_rate + net_migration_rate)$.

Använd sedan *population_growth_rate* för att beräkna befolkningsförändringen uttryckt i procent av folkmängden. Om resultatet av beräkningen är ett positivt tal kan vi anta en ökande befolkning annars en minskande befolkning.

Presentationen av ovanstående beräkningar ska bestå av en tabell som innehåller de fem länder med mest negativ befolkningsförändring och de fem länder med mest positiv befolkningsförändring (alltså totalt 10 länder i tabellen). Tabellen ska innehålla kolumnerna *country*, *birth_rate*, *death_rate*, *net_migration_rate* och *population_change*. De tio länderna ska också presenteras i ett stapeldiagram där namnet på länderna finns på x-axeln och befolkningsförändringen finns på y-axeln.

(Länder som har NaN i kolumnen *population_change* ska inte visas i tabellen)

Menyalternativ 4:

Avsluta programmet

4. I denna uppgift ska du skriva ett program som analyserar den procentuella befolkningsutvecklingen mellan åren 1960 och 2021 för de länder som finns i DataFrame *df_worldpubind* genom att utföra nedanstående uppgifter:

Uppgift a:

Skapa en tabell (på skärmen) som innehåller den procentuella befolkningsutvecklingen för de 5 länder med mest negativ befolkningsförändring (dvs befolkningsminskning) och de 5 länder med störst positiv befolkningsändring (dvs befolkningsökning) mellan år 1960 och år 2021. Visa också samma information grafiskt i två stapeldiagram där det första stapeldiagrammet visar de fem länderna med den negativa befolkningsutvecklingen och det andra stapeldiagrammet visar de länder med den positiva befolkningsutvecklingen. På x-axeln visas länderna och på och y-axeln den procentuella befolkningsutvecklingen. Vi definierar den procentuella befolkningsutvecklingen från år 1960 till år 2021 enligt formel: $((antal_invånare_2021 - antal_invånare_1960) / antal_invånare_1960) * 100$

Uppgift b:

Skapa ett program som först frågar efter ett land och därefter beräknar och presenterar den procentuella förändringen i folkmängd år för år under tidsperioden 1961 – 2021 i ett linjediagram. Den procentuella befolkningsförändringen mellan år (x-1) och år x där x är ett årtal mellan 1961 och 2021 beräknas enligt formeln:

$((antal_invånare_år_x - antal_invånare_år_x-1) / antal_invånare_år_x-1) * 100$. Resultatet av beräkningarna ska läggas i en ny kolumn med namnet *population_change*.

Linjediagrammets x-axel innehåller årtalen och den vänstra y-axeln anger den procentuella befolkningsförändringen. Som en sista del i denna uppgift ska du lägga till en höger y-axel i diagrammet och plotta ett linjediagram över antalet invånare per år.

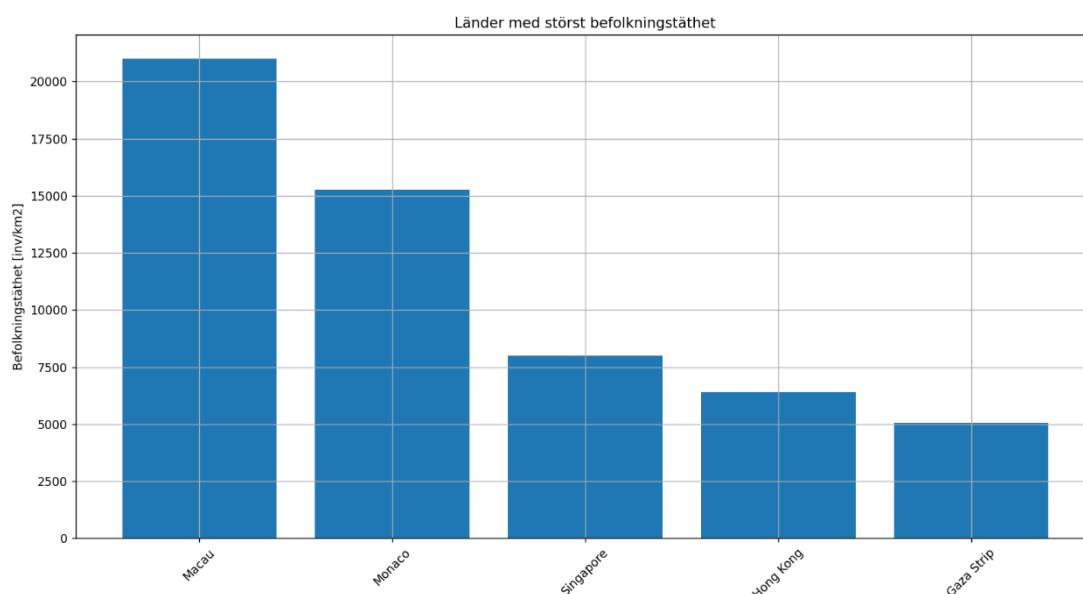
5. I denna uppgift ska du extrahera och presentera data från DataFrame *df_worldcities* enligt följande beskrivningar:
- Skapa en tabell som innehåller de 10 länder med flest antal städer och där tabellkolumnerna består av landet, antal städer i landet, namnet på den stad i landet som har flest invånare och antalet invånare i den staden. Tabellen ska vara sorterad i fallande ordning efter antal städer. Plotta därefter två stapeldiagram som är placerade över varandra med hjälp av subplot-kommandot (som finns i matplotlib.pyplot). Det övre stapeldiagrammet ska visa tabellens länder (x-axeln) och antal städer (y-axeln) i respektive land. Det undre stapeldiagrammet ska visa den största staden från respektive land (x-axeln) och antalet invånare (y-axeln) i dessa städer.

Bilaga: Exempel från programkörningar

Nedanstående programkörningar visar exempel på hur presentationen kan utformas för respektive uppgifter. Dina presentationer ska likna nedanstående presentationer, men de behöver inte vara exakt lika.

(OBS! värden i tabeller och grafer i nedanstående presentationer kan vara exempelvärden – du kan alltså inte förlita dig på att innehållet i nedanstående presentationer är 'facit')

Uppgift 2:



Uppgift 3 – menyalternativ 1:

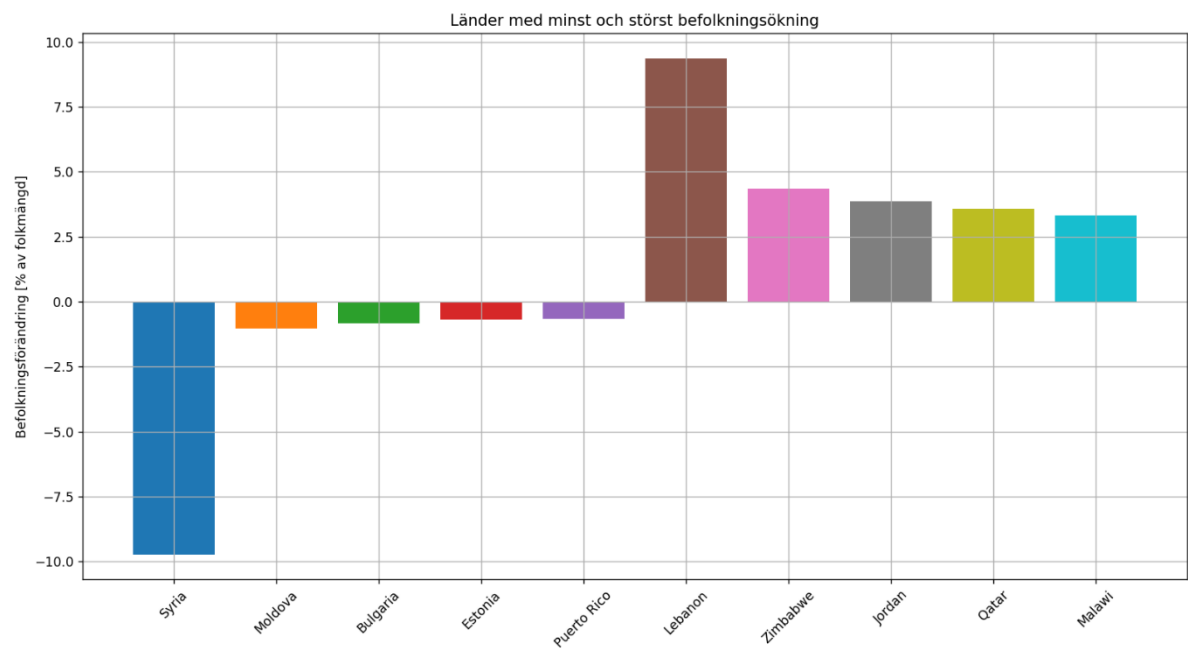
Land	Area [km2]	Antal födselar [per 1000 inv]	Livslängd [år]
Vietnam	331210	16	72
Philippines	2970000	21	74

Uppgift 3 – menyalternativ 2:

Land	Folkmängd	Antal internetanvändare [per 100.000 inv]
Länder med lägst antal internetanvändare per 100.000 invånare:		
Timor-Leste	1201542	174.8
Burma	55746253	197.3
Sierra Leone	5743725	259.4
Bangladesh	166280712	371.2
Congo, Democratic Republic of the	77433744	374.5
Länder med högst antal internetanvändare per 100.000 invånare:		
Iceland	317351	95036.7
Netherlands	16877351	88118.1
Saint Lucia	163362	87474.4
Sweden	9723809	86365.3

Uppgift 3 – menyalternativ 3:

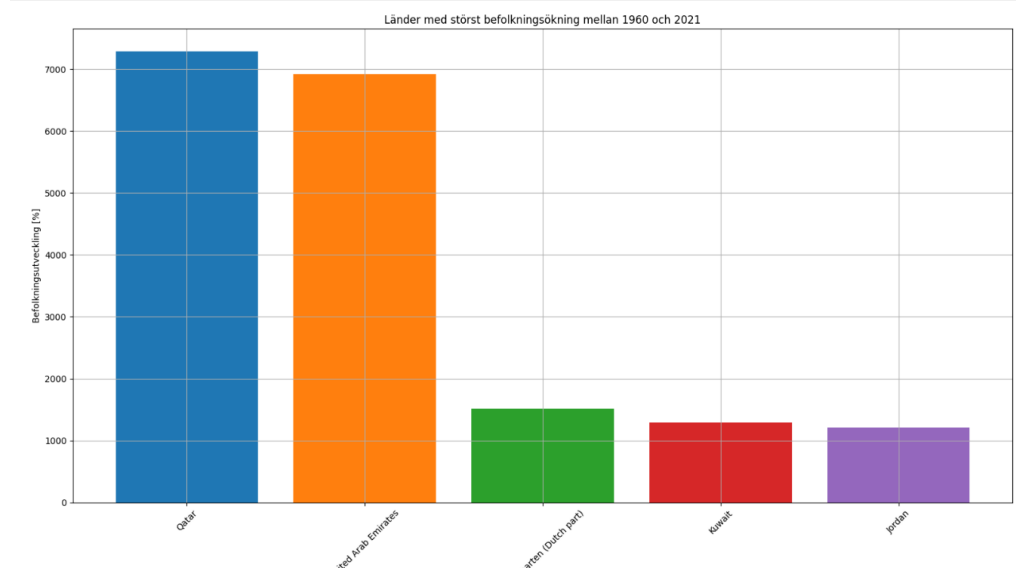
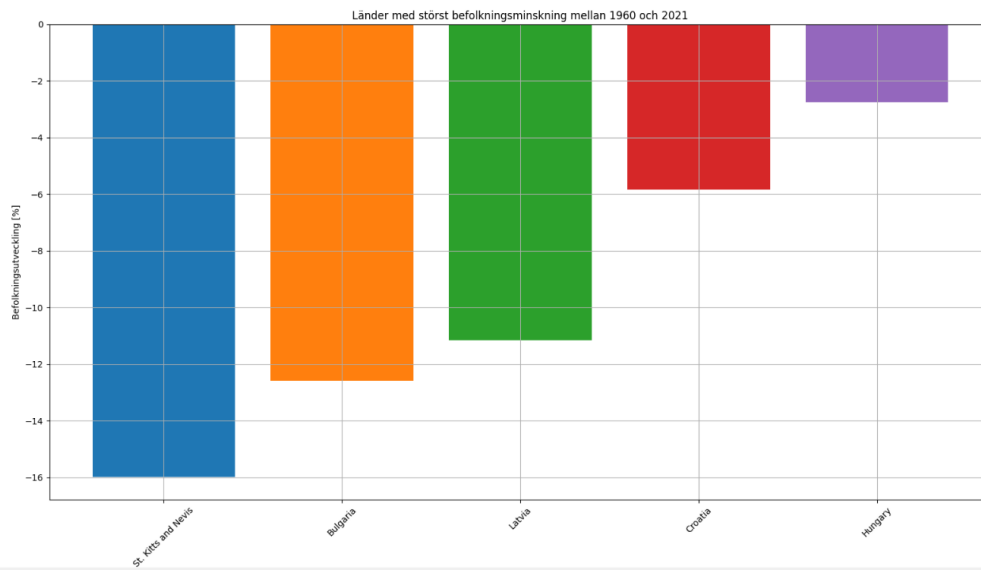
Land	Antal födselar [per 1000 inv]	Antal döda [per 1000 inv]	Netto migration [per 1000 inv]	Befolkningsförändring [% av folkmängd]
Länder där befolkningen minskar mest :				
Syria	22	6	-113	-9.73
Moldova	12	12	-9	-1.02
Bulgaria	8	14	-2	-0.83
Estonia	10	13	-3	-0.68
Puerto Rico	10	8	-8	-0.65
Länder där befolkningen ökar mest:				
Lebanon	14	4	83	9.37
Zimbabwe	32	10	21	4.36



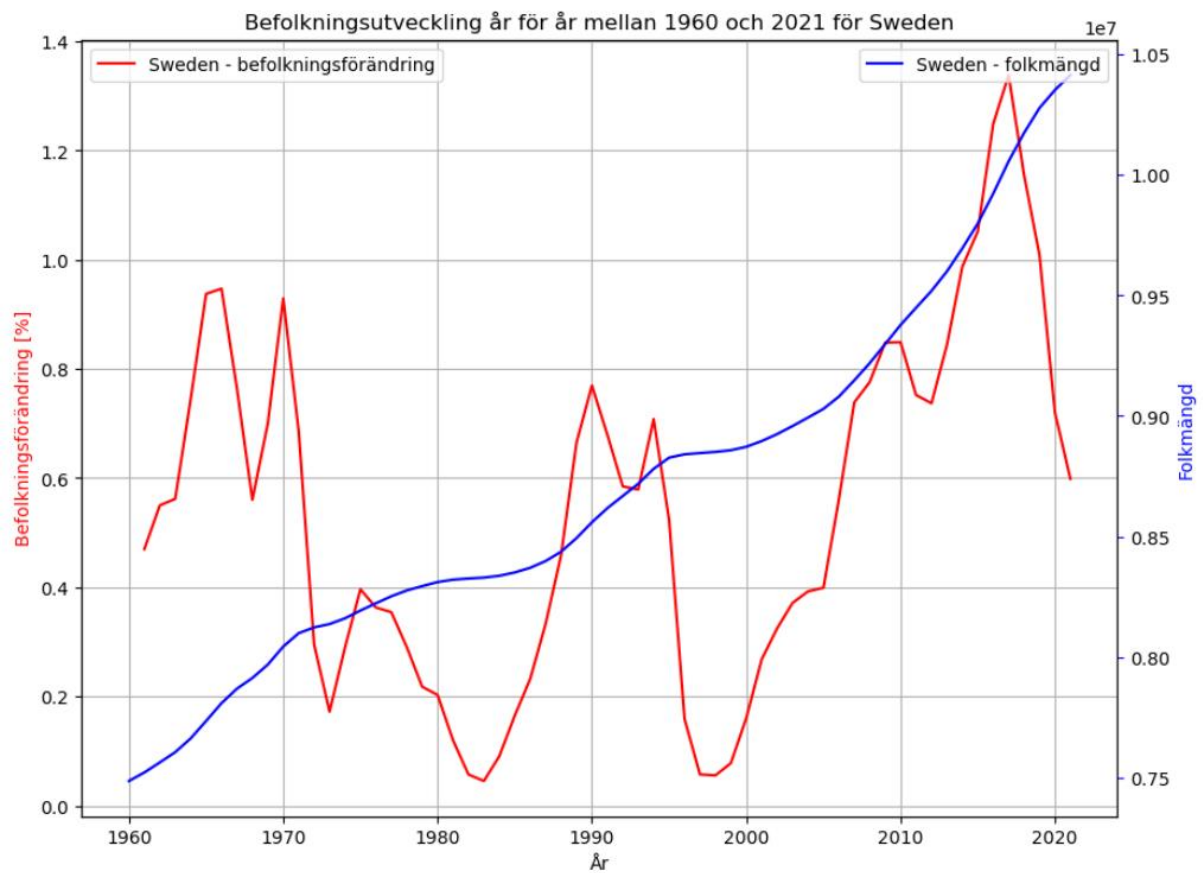
Uppgift 4 a:

Befolkningsutvecklingen hos ett antal länder under åren 1960-2021	
Land	Befolkningsutveckling mellan 1960 och 2021 [%]
De fem länder som minskat sin folkmängd mest:	
St. Kitts and Nevis	-15.98
Bulgaria	-12.58
Latvia	-11.15
Croatia	-5.83
Hungary	-2.75
De fem länder som ökat sin folkmängd mest:	
Qatar	7288.31
United Arab Emirates	6918.98

Uppgift 4 a:



Uppgift 4 b:



Uppgift 5:

De 10 länder med flest antal städer			
Land	Antal städer	Största stad	Antal inv i största stad
United States	7824	New York	18713220
Brazil	3604	São Paulo	22495000
Germany	2643	Berlin	3664088
Italy	2140	Rome	2872800

