

What's the Key Factor to Happiness?

Dataset

- * World Happiness Report 2024, which delivers the latest findings on happiness world-wide
- * Survey measure of subjective well-being on 100,000+ people in 130 countries, 2005 - 2023
- * Target variable: Life ladder (Cantril ladder)
- * Features: Log GDP per capita, Healthy life expectancy, Social support, Freedom to make life choices, Corruption perception, Positive affect, Negative affect, Generosity



Use Case:

Identify the feature that contributes the most to happiness in 2023

Test the result with representative countries in recent 3 years

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2363 entries, 0 to 2362
Data columns (total 11 columns):
 #   Column           Non-Null Count   Dtype  
--- 
 0   Country name     2363 non-null    object  
 1   year             2363 non-null    int64  
 2   Life Ladder      2363 non-null    float64 
 3   Log GDP per capita 2335 non-null    float64 
 4   Social support   2350 non-null    float64 
 5   Healthy life expectancy at birth 2300 non-null    float64 
 6   Freedom to make life choices 2327 non-null    float64 
 7   Generosity       2282 non-null    float64 
 8   Perceptions of corruption 2238 non-null    float64 
 9   Positive affect  2339 non-null    float64 
 10  Negative affect 2347 non-null    float64 
dtypes: float64(9), int64(1), object(1)
memory usage: 203.2+ KB
```

	Country name	year	Life Ladder	Log GDP per capita	Social support	Healthy life expectancy at birth	Freedom to make life choices	Generosity	Perceptions of corruption	Positive affect	Negative affect
0	Afghanistan	2008	3.723590	7.350416	0.450662	50.500000	0.718114	0.164055	0.881686	0.414297	0.258195
1	Afghanistan	2009	4.401778	7.508646	0.552308	50.799999	0.678896	0.187297	0.850035	0.481421	0.237092
2	Afghanistan	2010	4.758381	7.613900	0.539075	51.099998	0.600127	0.117861	0.706766	0.516907	0.275324
3	Afghanistan	2011	3.831719	7.581259	0.521104	51.400002	0.495901	0.160098	0.731109	0.479835	0.267175
4	Afghanistan	2012	3.782938	7.660506	0.520637	51.700001	0.530935	0.234157	0.775620	0.613513	0.267919
...
2358	Zimbabwe	2019	2.693523	7.697755	0.759162	53.099998	0.631908	-0.050874	0.830652	0.658434	0.235354
2359	Zimbabwe	2020	3.159802	7.596050	0.717243	53.575001	0.643303	0.002848	0.788523	0.660658	0.345736
2360	Zimbabwe	2021	3.154578	7.656878	0.685151	54.049999	0.667636	-0.079007	0.756945	0.609917	0.241682
2361	Zimbabwe	2022	3.296220	7.670073	0.666172	54.525002	0.651987	-0.072935	0.752632	0.640609	0.191350
2362	Zimbabwe	2023	3.572386	7.678590	0.693817	55.000000	0.734613	-0.068837	0.757494	0.609752	0.178953

2363 rows x 11 columns

Data Cleaning

Methods to impute missing values:

I. Time series interpolation

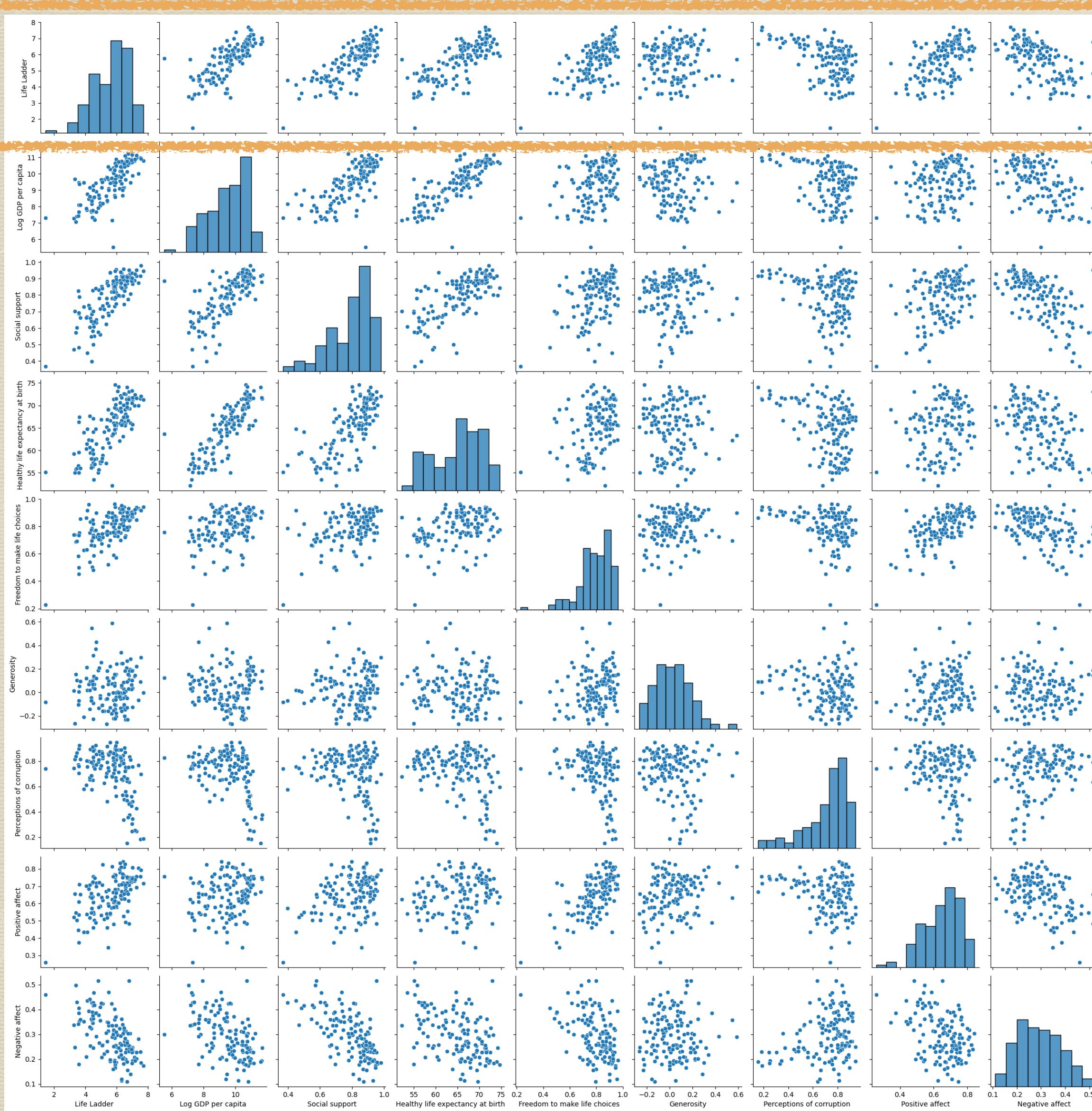
2. KNN imputer (best n value = 1)

Low variance

Relatively normal distribution

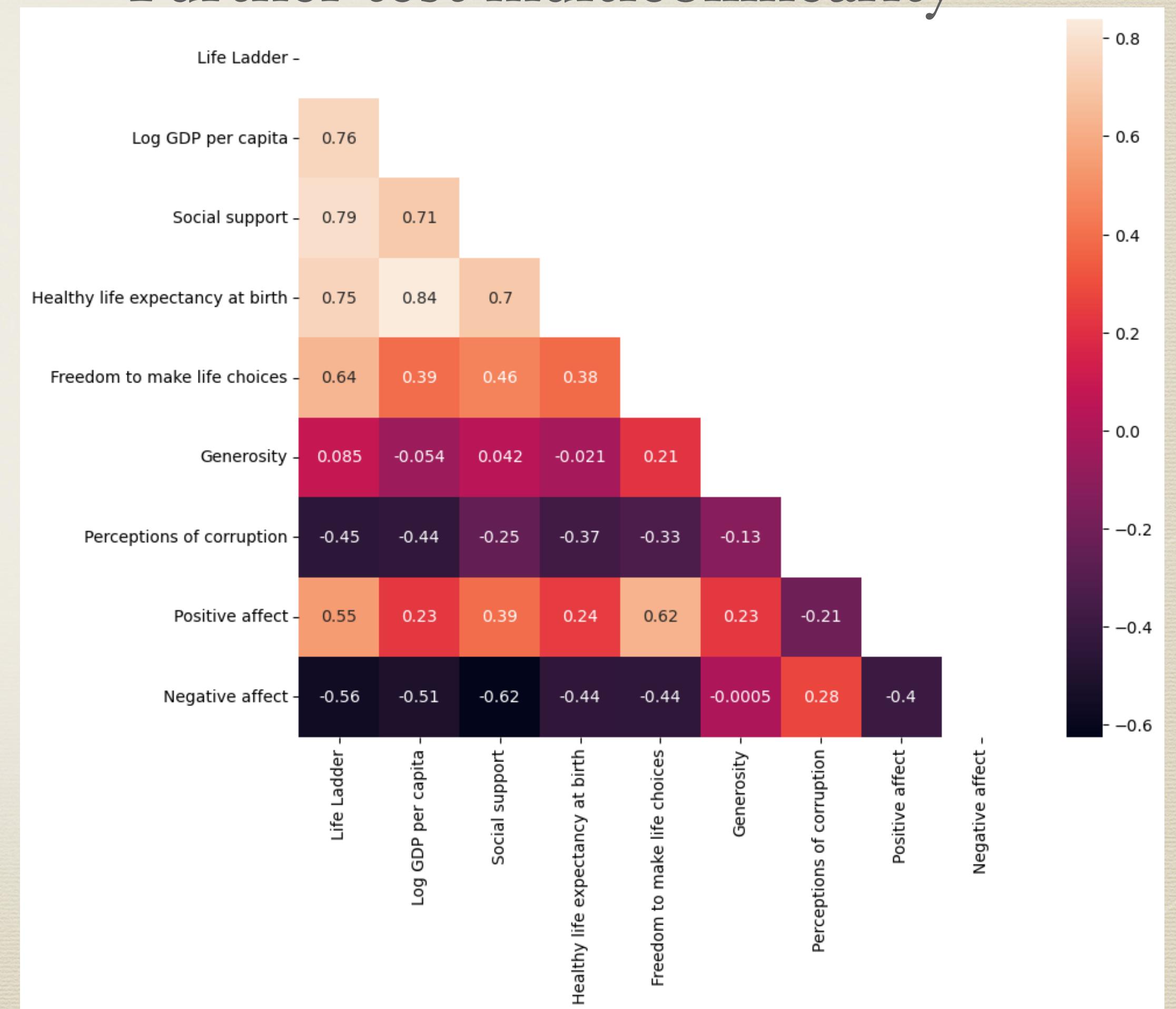
	year	Life Ladder	Log GDP per capita	Social support	Healthy life expectancy at birth	Freedom to make life choices	Generosity	Perceptions of corruption	Positive affect	Negative affect
count	138	138.000000	138.000000	138.000000	138.000000	138.000000	138.000000	138.000000	138.000000	138.000000
mean	2023-01-01 00:00:00	5.620811	9.495747	0.790988	65.157971	0.791379	0.031157	0.714204	0.652119	0.293461
min	2023-01-01 00:00:00	1.445909	5.526723	0.368478	52.200001	0.228301	-0.268211	0.152543	0.260513	0.110535
25%	2023-01-01 00:00:00	4.679697	8.612232	0.701845	60.700001	0.734933	-0.081641	0.650091	0.580944	0.229711
50%	2023-01-01 00:00:00	5.862690	9.643922	0.829442	66.099998	0.810746	0.022123	0.765596	0.668374	0.285380
75%	2023-01-01 00:00:00	6.486904	10.524409	0.889664	69.574999	0.876399	0.132979	0.836950	0.735437	0.357242
max	2023-01-01 00:00:00	7.698929	11.675588	0.978789	74.599998	0.964832	0.589554	0.947856	0.843147	0.515916
std	NaN	1.139478	1.221816	0.129663	5.554347	0.120220	0.160009	0.177828	0.109041	0.088834

Correlation

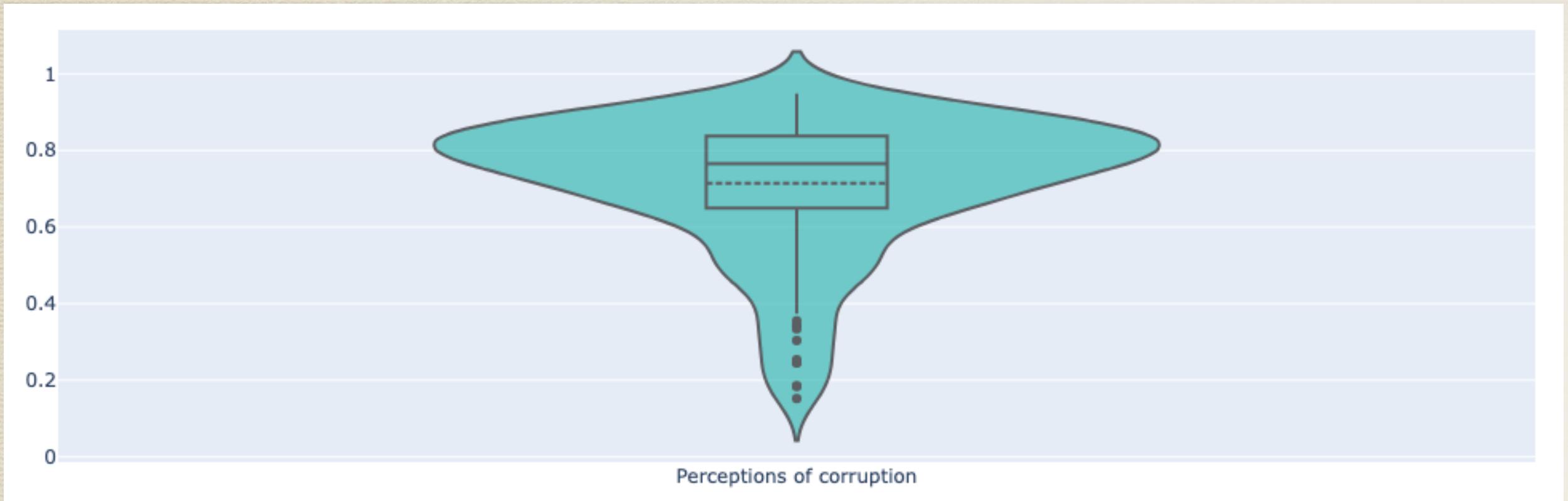


Safe to perform linear regression

Further test multicollinearity



Outliers



10 outliers within “Perceptions of corruption”:

1. Half of them among the top 10 happy countries
2. Test the regression results separately

	year	Country name	Life Ladder	Log GDP per capita	Social support	Healthy life expectancy at birth	Freedom to make life choices	Generosity	Perceptions of corruption	Positive affect	Negative affect
1851	2023-01-01	Singapore	6.653942	11.590220	0.916326	74.000000	0.861233	0.088438	0.152543	0.667157	0.190486
556	2023-01-01	Denmark	7.504193	10.996293	0.915670	71.500000	0.922739	0.089434	0.183567	0.757006	0.229489
680	2023-01-01	Finland	7.698929	10.808451	0.946865	71.300003	0.943418	-0.001130	0.185229	0.716778	0.172830
1594	2023-01-01	Norway	7.248675	11.124797	0.951586	71.599998	0.937649	0.219493	0.244850	0.756484	0.228338
2017	2023-01-01	Switzerland	6.968833	11.169257	0.903757	73.300003	0.890972	0.103554	0.246744	0.745368	0.184582
2004	2023-01-01	Sweden	7.160777	10.902137	0.926618	72.400002	0.926379	0.146668	0.253074	0.738826	0.194445
1515	2023-01-01	New Zealand	6.975955	10.719568	0.932887	70.500000	0.876608	0.181392	0.303656	0.738133	0.228583
649	2023-01-01	Estonia	6.429873	10.516722	0.958171	70.099998	0.914573	0.031501	0.333731	0.764900	0.182296
1249	2023-01-01	Luxembourg	7.016062	11.648711	0.879195	71.699997	0.910813	0.033440	0.343211	0.750700	0.194495
2206	2023-01-01	United Arab Emirates	6.728384	11.235544	0.775808	66.599998	0.885922	0.155217	0.355116	0.654696	0.304355

Regression

Linear: Standard scaler + no outlier (median)			
	mean squared error	mean absolute error	R squared
Train	0.2287	0.3393	0.7815
Test	0.3017	0.4443	0.7833

Linear: Robust scaler + outlier			
	mean squared error	mean absolute error	R squared
Train	0.2289	0.3443	0.8131
Test	0.2291	0.3826	0.8354

Lasso: Standard scaler + no outlier (median)			
	mean squared error	mean absolute error	R squared
Train	0.2602	0.3761	0.7514
Test	0.3302	0.4669	0.7628

Lasso: Robust scaler + outlier			
	mean squared error	mean absolute error	R squared
Train	0.2728	0.3868	0.7772
Test	0.2747	0.4216	0.8027

Linear: PolynomialFeatures + no outlier (median)			
	mean squared error	mean absolute error	R squared
Train	0.1332	0.2918	0.8728
Test	0.5525	0.5995	0.6032

Including outliers

Robust scaler

Linear regression

Feature Importance

	features	weights
1	Social support	0.510208
0	Log GDP per capita	0.382589
2	Healthy life expectancy at birth	0.277162
6	Positive affect	0.275915
3	Freedom to make life choices	0.266713
7	Negative affect	0.125020
4	Generosity	-0.007471
5	Perceptions of corruption	-0.163675

From linear regression:

Social support contributes
the most to happiness

Logistic Regression

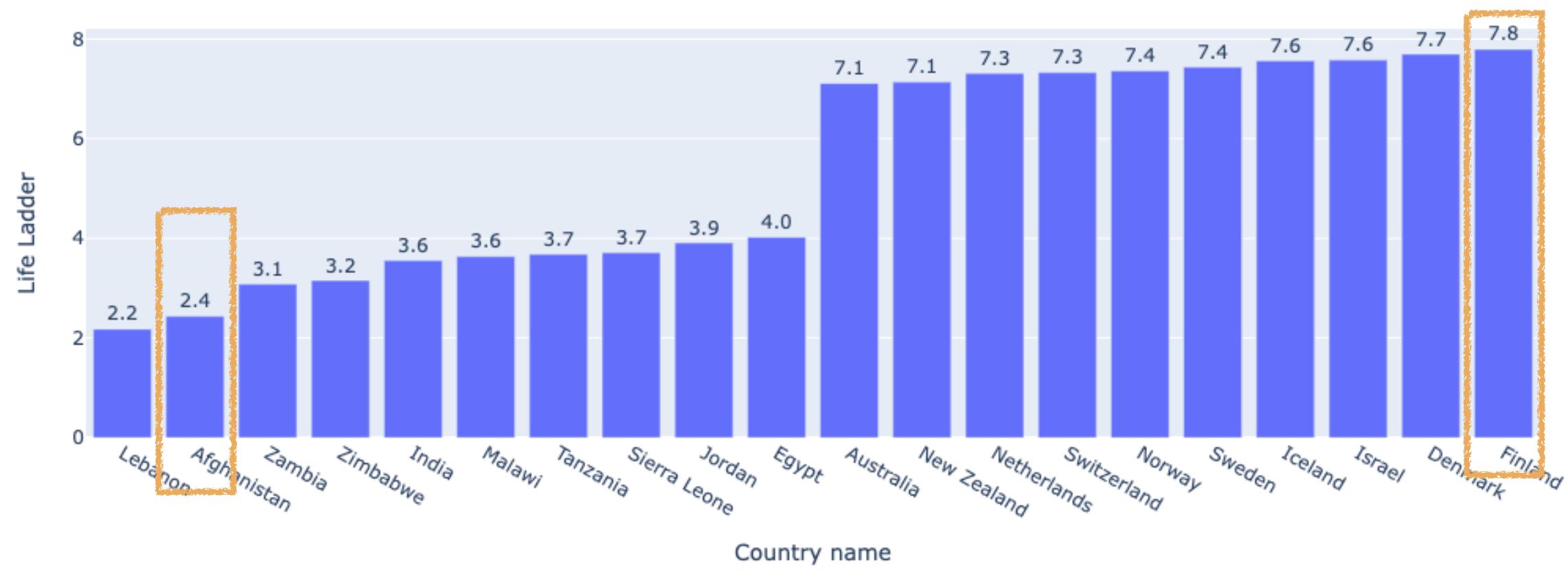
	precision	recall	f1-score	support
0	0.94	0.94	0.94	16
1	0.95	0.95	0.95	19
accuracy			0.94	35
macro avg	0.94	0.94	0.94	35
weighted avg	0.94	0.94	0.94	35
	precision	recall	f1-score	support
0	0.90	0.95	0.93	78
1	0.95	0.90	0.92	78
accuracy			0.92	156
macro avg	0.92	0.92	0.92	156
weighted avg	0.92	0.92	0.92	156

	features	weights
1	Social support	1.635080
2	Healthy life expectancy at birth	1.408721
0	Log GDP per capita	0.949141
3	Freedom to make life choices	0.863854
6	Positive affect	0.408012
4	Generosity	-0.115899
7	Negative affect	-0.825096
5	Perceptions of corruption	-1.198265

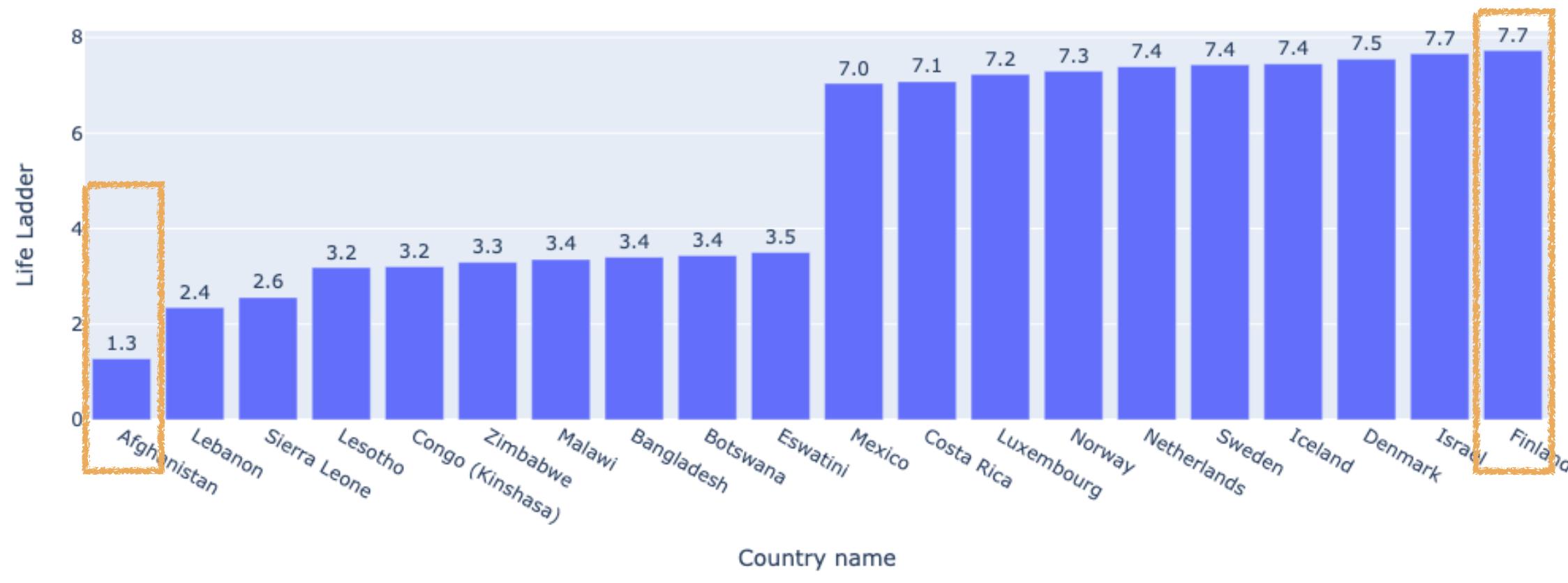
Confidence Interval		
0.95	0.6260	0.7798
0.9	0.6385	0.7673
0.85	0.6466	0.7592

Representative Countries

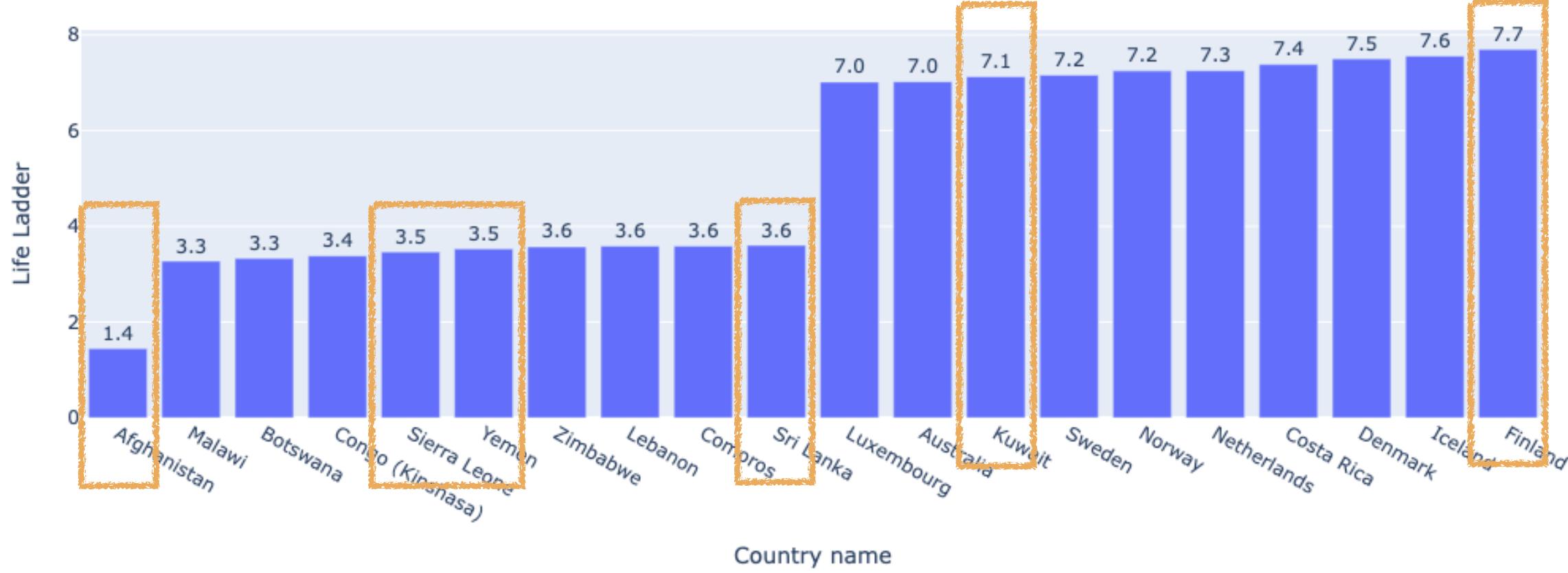
Happiness Ranking 2021



Happiness Ranking 2022

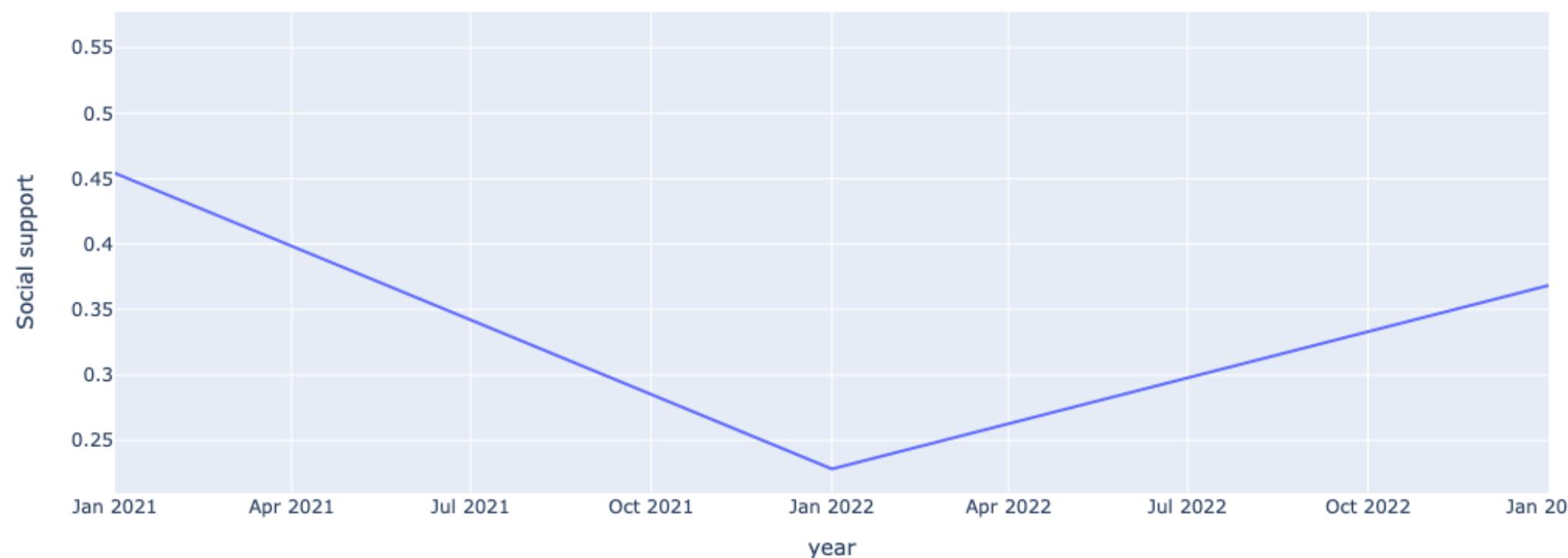


Happiness Ranking 2023

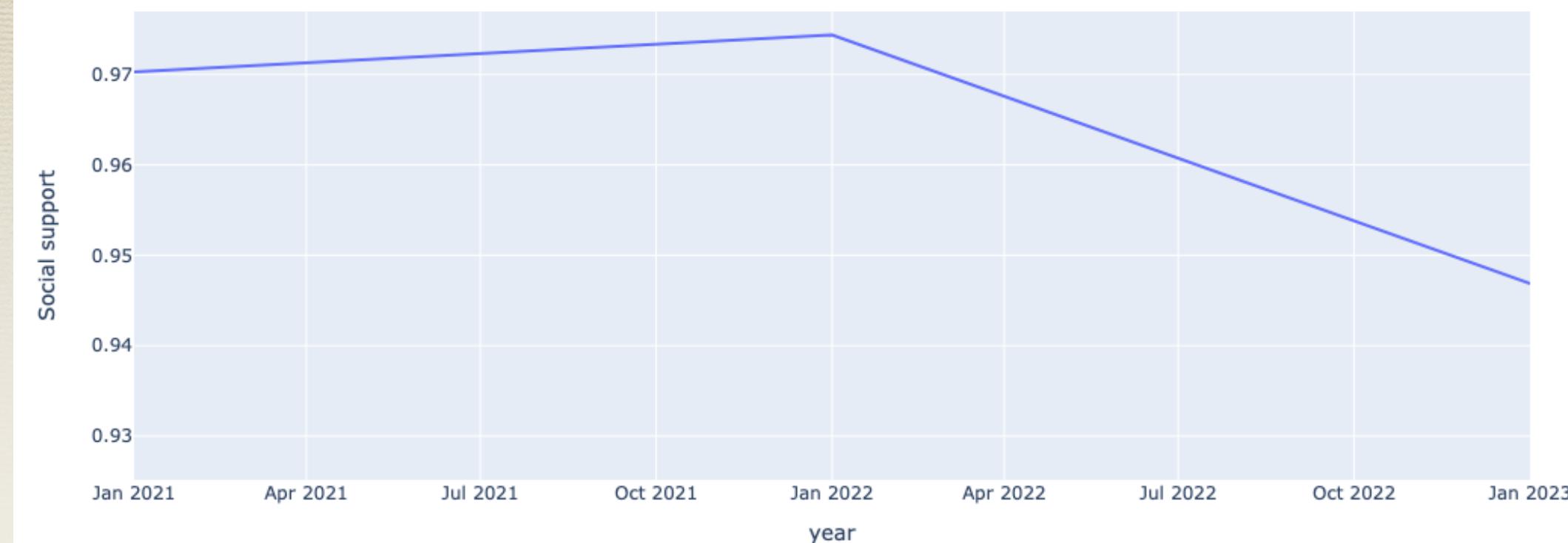


- * Afghanistan (at bottom in 3 yr)
- * Finland (at top in 3 yr)
- * Kuwait (rise to top in 3 yr)
- * Yemen, Comoros & Sri Lanka (drop to bottom in 3 yr)

Social support - Afghanistan



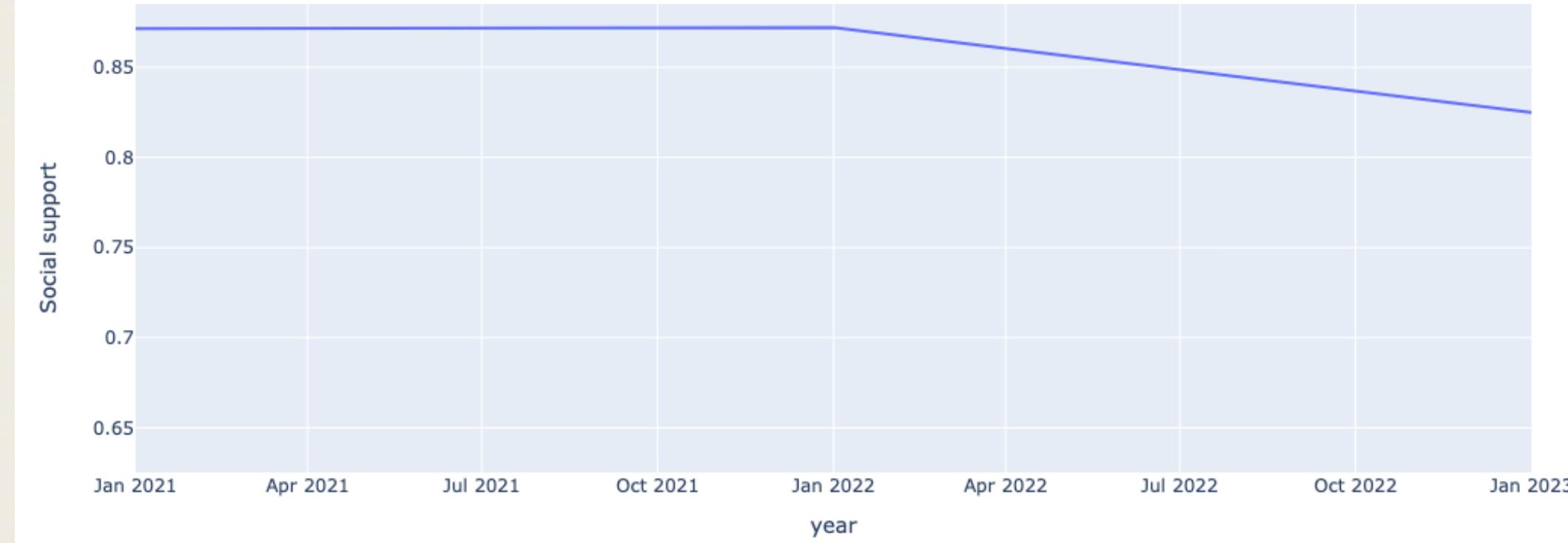
Social support - Finland



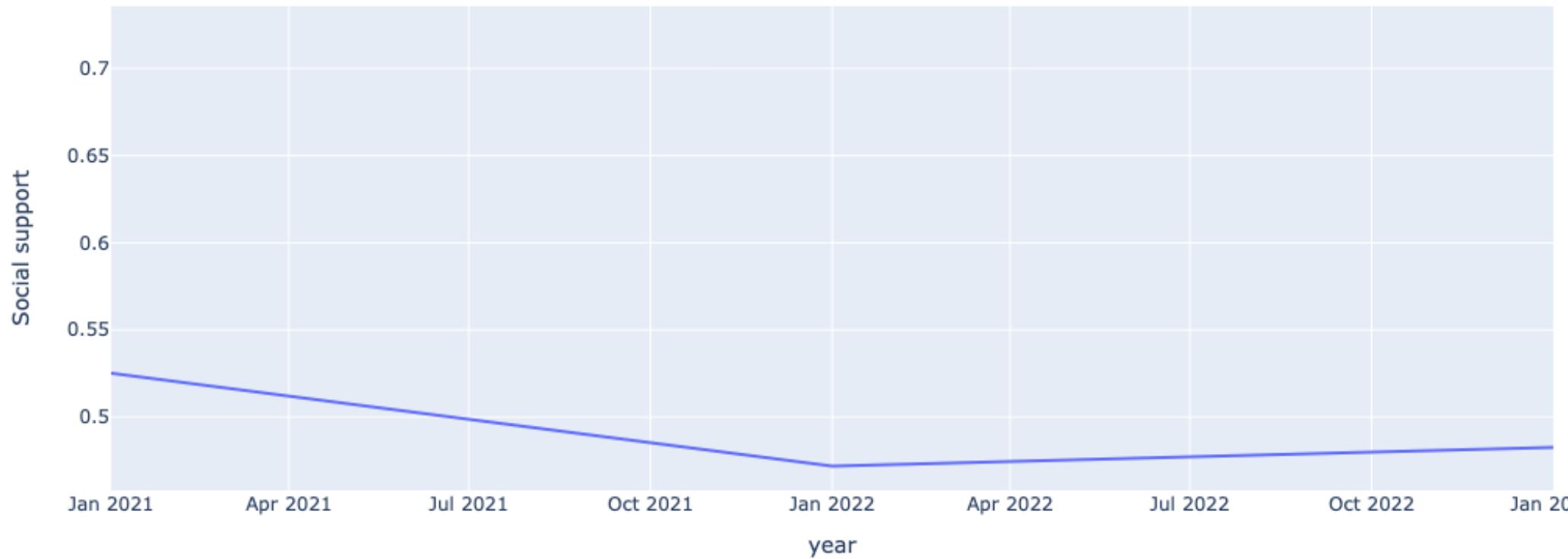
Social support - Kuwait



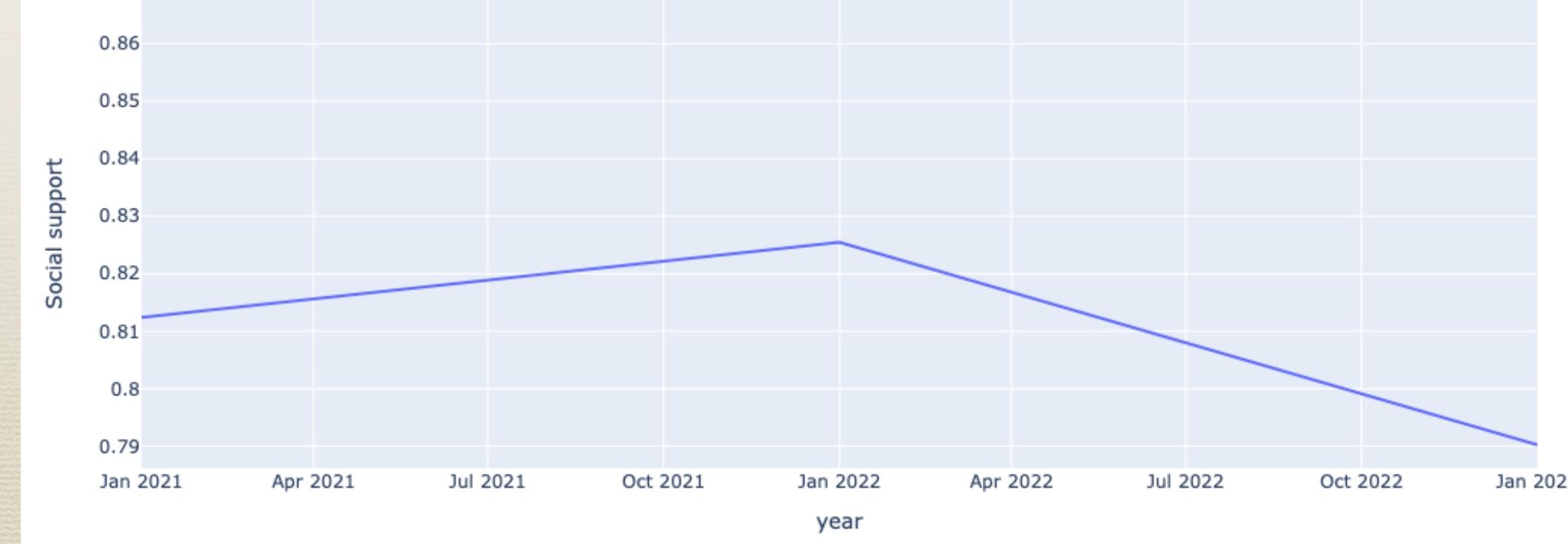
Social support - Yemen



Social support - Comoros



Social support - Sri Lanka



THANK YOU