

Research on the Economics of Happiness

Владислав Гарбуз, Артём Скоркин

11 ноября 2023 г.

1 Введение

Всемирный доклад о счастье (англ. World Happiness Report) — ежегодный доклад, публикуемый подразделением ООН по поиску решений стабильного развития.

В июле 2011 года Генеральная Ассамблея ООН приняла резолюцию, призывающую страны — члены ООН, оценивать счастье своего

Первый всемирный доклад о счастье был опубликован 1 апреля 2012 года как основополагающий текст для встречи. Это привлекло международное внимание как первое глобальное исследование счастья в мире. В докладе были изложены принципы мирового счастья, причины счастья и нищеты, а также последствия, которые были освещены в тематических исследованиях. В сентябре 2013 года во втором всемирном докладе о счастье предлагалось первое международное обозрение, и с этого момента доклад публикуется каждый год. В докладе используются данные опросов института Гэллапа.

2 Постановка задачи

Наша цель:

1. Выявить зависимость и ее силу между параметрами страны (ВВП на душу населения, социальная поддержка или же семья, здоровье, свобода, щедрость, коррупция) и уровнем счастья ее жителей.
2. Построить модель, которая бы прогнозировала уровень счастья по заданным параметрам

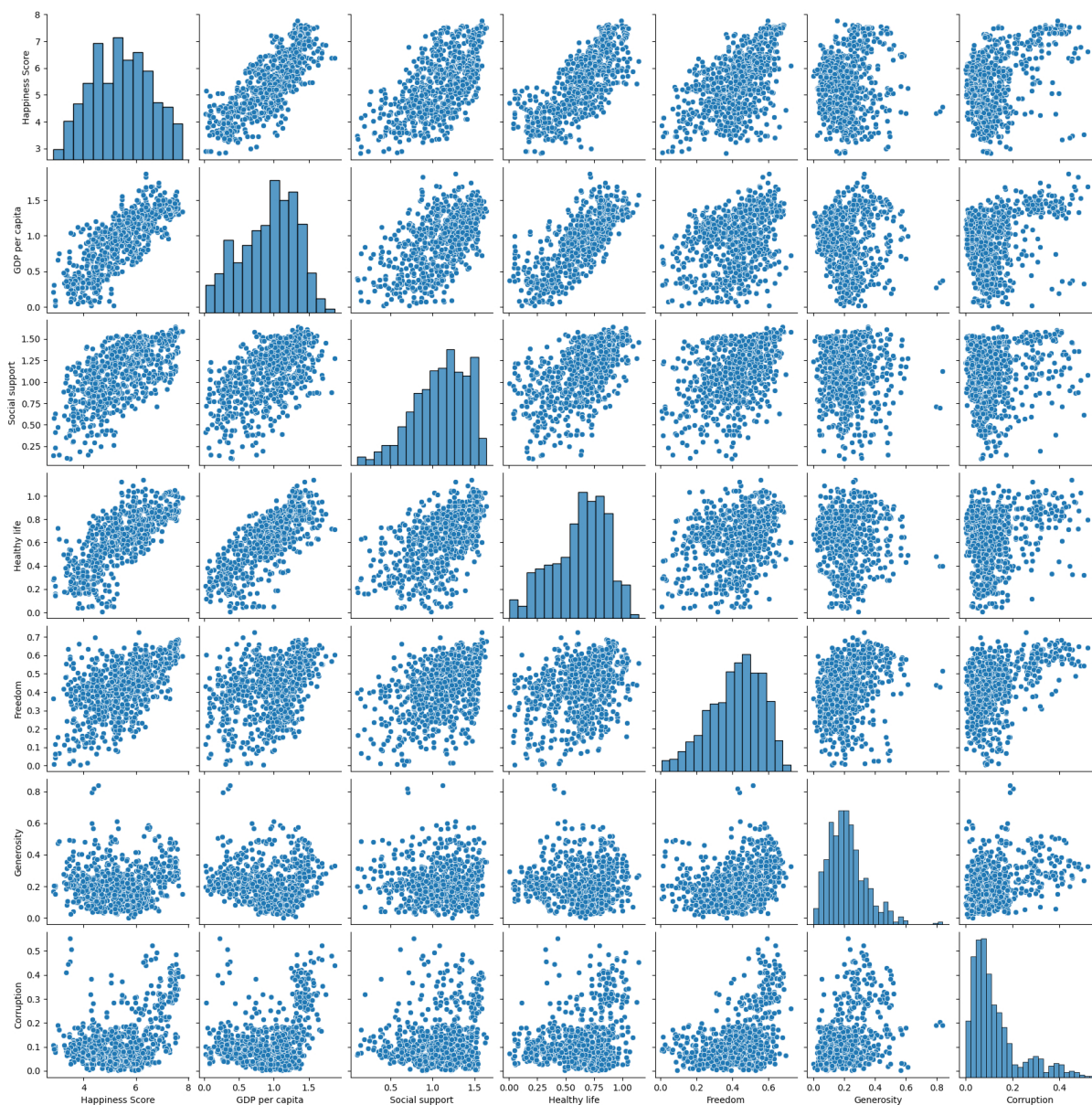
3 Используемые методы

1. Предобработка данных - избавились от пустых и пропущенных значений для более качественного анализа.
2. Разведочный анализ данных - построили парные графики и матрицу корреляции. Это помогло нам определить взаимосвязь между параметрами и уровнем счастья.
3. Построили модель линейной регрессии. С ее помощью мы определили прерменные веса каждого из параметров и научились неплохо прогнозировать уровень счастья

ПО НИМ.

4 Полученные результаты

4.1 Попарный график



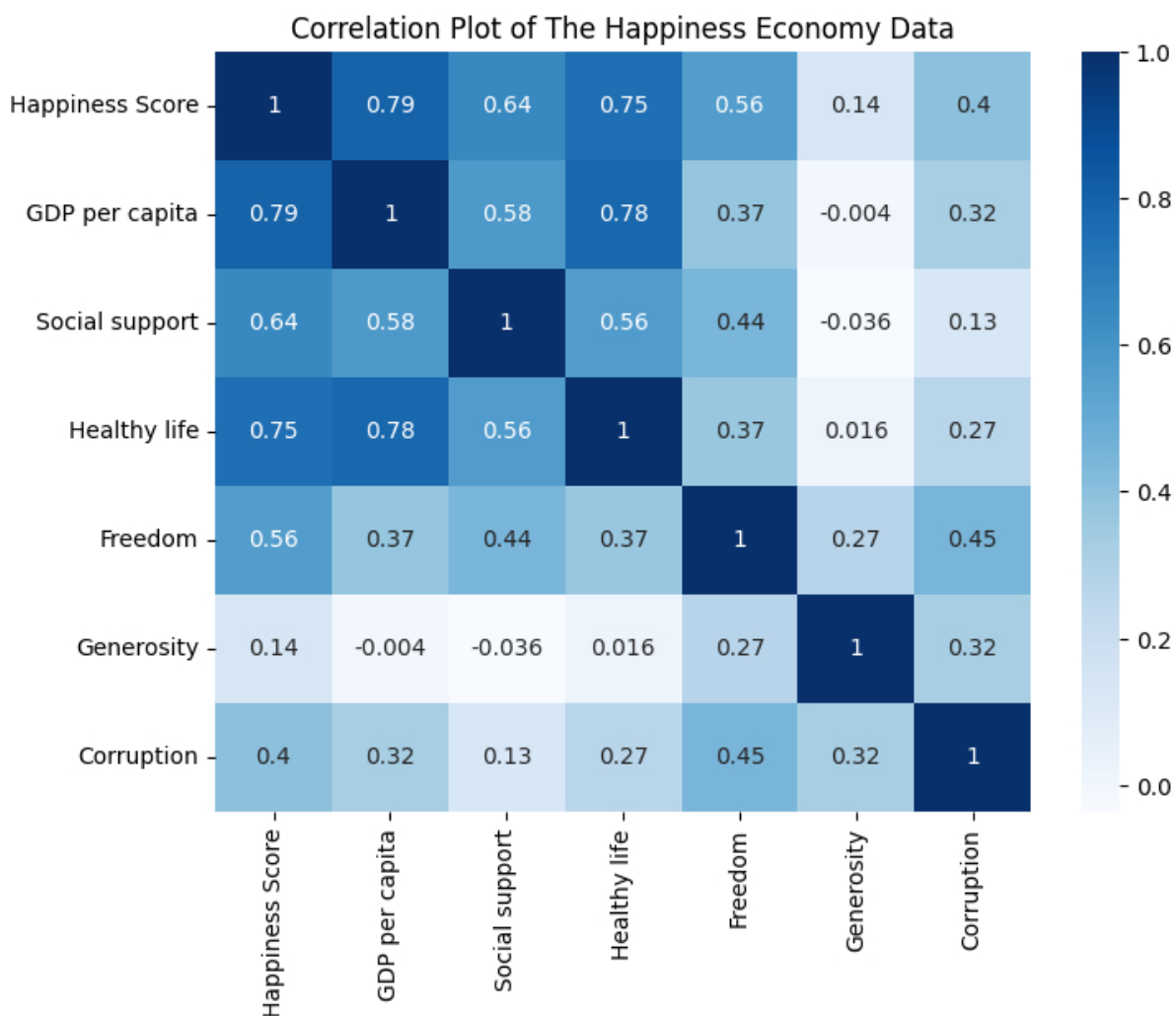
Из графиков можно сделать следующие выводы:

Наиболее чётко линейная зависимость, а именно положительная корреляция, для объясняемой переменной "Happiness Score" прослеживается с признаками "GDP per capita" и "Healthy life". Несколько хуже, но всё ещё достаточно хорошо виден подобный результат и для признака "Social support" а также "Freedom". Наконец, хуже всего положительная корреляция для объясняемой переменной наблюдается с "Generosity" и "Corruption".

Отдельно стоит отметить, что также есть место для сильной положительной корреляции между признаками "GDP per capita" и "Healthy life" и менее выраженная между, но существенная между "GDP per capita" и "Social support".

Вывод из наших наблюдений представим ниже, после построения и сопоставления результатов с корреляционной матрицей (тепловой картой).

4.2 Тепловая карта



Наиболее коррелирующими со счастьем факторами являются "GDP per capita" и "Healthy life" ($corr = 0.79$ и 0.75 соответственно), следом идёт "Social support" ($corr = 0.64$), потом "Freedom" ($corr = 0.56$), "Corruption" ($corr = 0.4$), а в самом конце списка со значительным отрывом от остальных показателей находятся "Generosity" ($corr = 0.14$). Более того, оказывается, что "GDP per capita" и "Healthy life" имеют высокую положительную корреляцию (0.78). Чуть менее выраженная, однако достаточная корреляция (> 0.5) присутствует и между "GDP per capita" и "Social support" а также между "Healthy life" и "Social support" откуда делаем вывод, что первые три по значимости фактора тесно связаны между собой.

4.3 Линейная регрессия

OLS Regression Results						
Dep. Variable:	Happiness Score			R-squared:	0.761	
Model:	OLS			Adj. R-squared:	0.759	
Method:	Least Squares			F-statistic:	315.4	
Date:	Sat, 11 Nov 2023			Prob (F-statistic):	7.60e-181	
Time:	21:16:23			Log-Likelihood:	-496.68	
No. Observations:	600			AIC:	1007.	
Df Residuals:	593			BIC:	1038.	
Df Model:	6					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	2.0728	0.096	21.674	0.000	1.885	2.261
GDP per capita	1.1482	0.096	11.960	0.000	0.960	1.337
Social support	0.6980	0.097	7.172	0.000	0.507	0.889
Healthy life	1.0035	0.157	6.383	0.000	0.695	1.312
Freedom	1.5563	0.199	7.831	0.000	1.166	1.947
Generosity	0.6240	0.205	3.042	0.002	0.221	1.027
Corruption	0.7783	0.255	3.056	0.002	0.278	1.279
Omnibus:	10.813	Durbin-Watson:		2.082		
Prob(Omnibus):	0.004	Jarque-Bera (JB):		11.443		
Skew:	-0.272	Prob(JB):		0.00327		
Kurtosis:	3.401	Cond. No.		24.0		

Коэффициент детерминации равен 0.761, что говорит о том, что наша модель достаточно хорошо может предсказывать уровень счастья в стране.

Помимо этого, выяснилось, что фактор "Freedom" с заданными значениями в модели вносит наибольший вклад из всех факторов - ≈ 1.556 . Следом идёт "GDP per capita" с весом в ≈ 1.148 и "Healthy life" со значением в ≈ 1.004 . Остальные же регрессоры имеют веса со значением меньше единицы.

Почему же получилось так, что параметр "Social support" сильно коррелирующий с уровнем счастья, имеет не слишком большой вес, а слабо коррелирующий "Freedom" - наибольший?

Ответом может служить мультиколлинеарность, которая была открыта нами ранее,

между тремя факторами с наибольшей положительной корреляцией по отношению к общей оценке счастья — они слишком сильно привязаны друг к другу (имеют большую положительную корреляцию), и часть веса из "Social support" перешла к "GDP per capita" (≈ 1.148) и "Healthy life" (≈ 1.004), оставив "Social support" с ≈ 0.698 .

Что с "Freedom"? Вспомним, что корреляция у "Freedom" с общей оценкой счастья была 0.56, в то время как у "Corruption" и "Generosity" 0.4 и 0.14 соответственно. Другими словами, два последних фактора оказались наименее статистически значимыми и их вес перешёл к "Freedom" что имеет большую значимость для определения общей оценки счастья. Вес не перешёл к другим признакам, так как те и так мультиколлинеарны.

Осталось лишь рассмотреть среднеквадратическую и среднюю абсолютную ошибки.

Для наших данных, где значения оценки счастья не превышают 7.769 и при этом являются положительными, мы имеем весьма хороший результат — для среднеквадратической ошибки получаем ≈ 0.272 , а для средней абсолютной ошибки ≈ 0.408 , что действительно мало. Получается, что наша модель линейной регрессии может хорошо предсказывать уровень счастья по заданным параметрам.