

Dữ liệu ngân hàng thế giới

Dữ liệu năm 2021 , chỉ số lạm phát theo giảm phát GDP (biên phụ thuộc) và các biến độc lập (GDP, xuất khẩu, lượng cung tiền, thay đổi chỉ số PPP theo GDP, tài khoản vãng lai và tổng chi tiêu quốc gia)

Таблица данных									
	Country	экономика	Инфляция	ВВП,%	Экспорт,%	Денежная	Динамики	Баланс текуще	Валовые
1	Albania	ALB	5.93	14.95	55.04	8.60	2.97	4.19	9.80
2	Algeria	DZA	15.74	19.79	83.90	13.18	11.12	-75	8.01
3	Angola	AGO	33.62	35.09	67.81	-9.3	28.29	863.33	20.54
4	Argentina	ARG	54.15	70.18	84.76	51.25	48.00	114.93	69.10
5	Armenia	ARM	6.87	12.96	34.10	13.10	2.60	7.81	11.45
6	Australia	AUS	2.80	5.10	-3.4	6.70	-0.44	56.33	4.56
7	Azerbaijan	AZE	21.16	27.94	67.59	18.71	16.32	-3.7e+03	5.69
8	Bangladesh	BGD	4.12	11.35	13.69	9.18	-0.033	-1.4e+03	12.89
9	Belarus	BLR	13.05	15.65	36.69	6.99	8.54	-1.3e+03	10.95
10	Belize	BLZ	3.96	19.78	50.29	12.39	-0.19	26.98	23.50

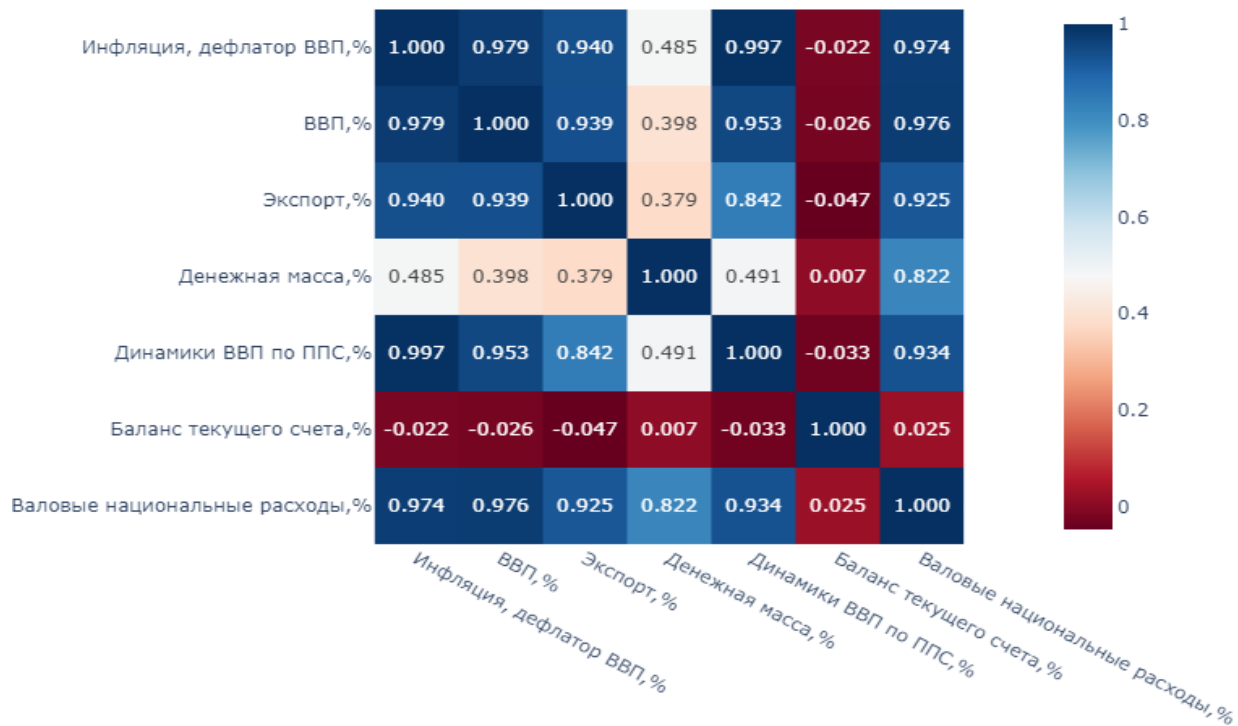
101 rows x 9 columns

101 rows x 9 columns

**Tính Hệ số tương quan ( Correlation coefficient) giữa các biến.**

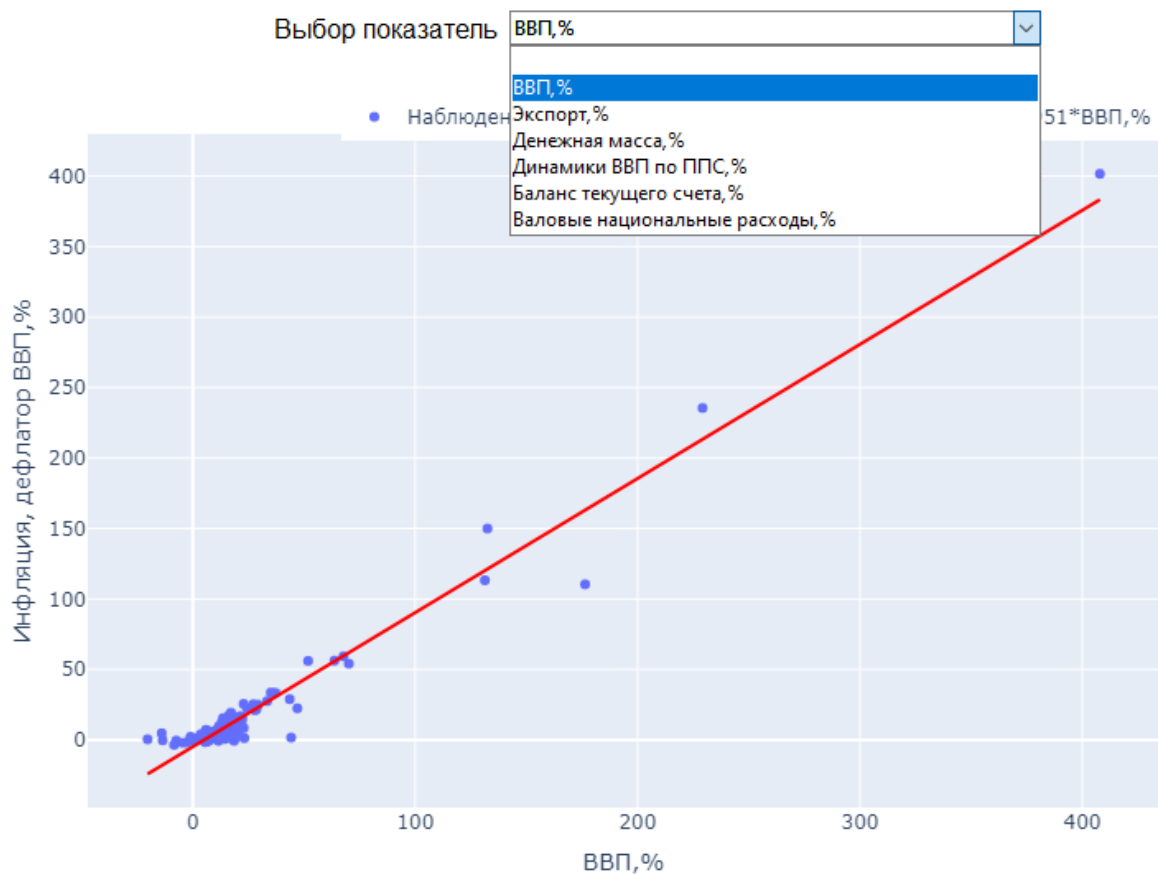
**Hệ số tương quan** là chỉ số thống kê đo lường mức độ mạnh yếu của mối quan hệ giữa hai biến số.

**Коэффициент корреляции Пирсона**



Hồi quy tuyến tính đơn giữa biến phụ thuộc và lần lượt các biến phụ thuộc. Hồi quy tuyến tính nó được dùng cho các bài toán về dự đoán, dự báo (prediction).

## Простая линейные регрессия - Simple Linear Regression

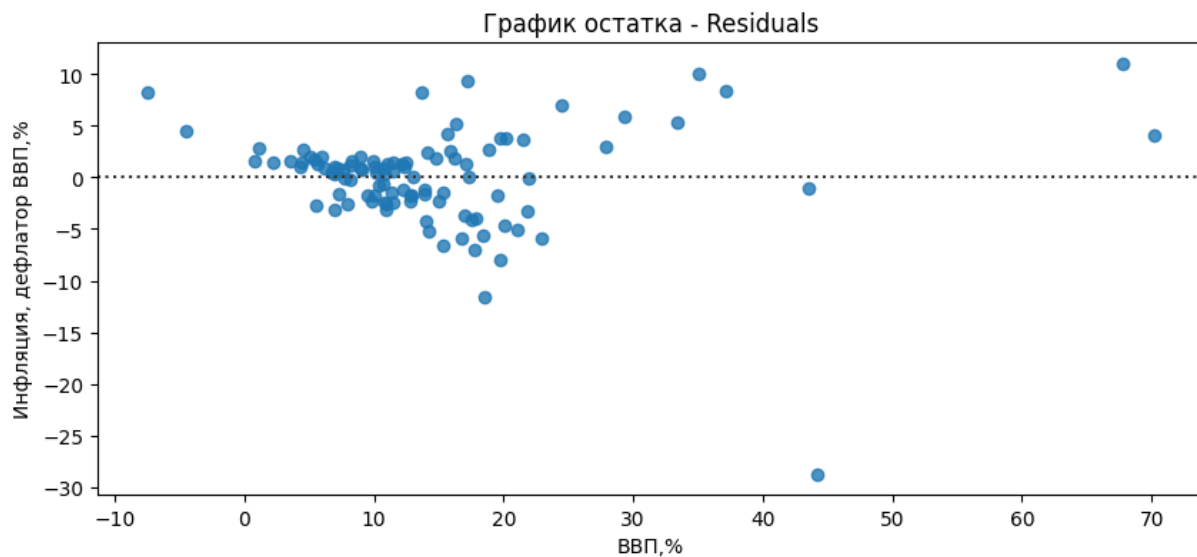


OLS Regression Results						
Dep. Variable:	Инфляция, дефлятор ВВП,%			R-squared:	0.959	
Model:	OLS			Adj. R-squared:	0.958	
Method:	Least Squares			F-statistic:	4493.	
Date:	Wed, 23 Aug 2023			Prob (F-statistic):	4.02e-136	
Time:	12:24:52			Log-Likelihood:	-671.82	
No. Observations:	196			AIC:	1348.	
Df Residuals:	194			BIC:	1354.	
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Коэффициент	-4.7145	0.592	-7.960	0.000	-5.883	-3.546
ВВП,%	0.9513	0.014	67.032	0.000	0.923	0.979
Omnibus:	110.840		Durbin-Watson:	1.834		
Prob(Omnibus):	0.000		Jarque-Bera (JB):	2159.515		
Skew:	-1.651		Prob(JB):	0.00		
Kurtosis:	18.922		Cond. No.	46.2		

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

## Phân dư



So sánh kết quả hồi quy tuyến tính giữa các biến độc lập để xem biến nào sẽ phù hợp nhất để dự báo chỉ số lạm phát, giảm phát trong tương lai.

R<sup>2</sup> càng lớn càng tốt và RMSE càng bé càng tốt.

	Показатель	ВВП,%	Экспорт,	Денежная	Динамики	Баланс те	Валовые
1	R2 (R-квадрат)	0.9586109689123123	0.883210	0.235207	0.993705	0.000481	0.948897
2	Среднеквадратическая ошибка RMSE	4.789158793907314	8.043115	8.841172	2.249167	9.891240	8.413894
3	Формула уравнения	-4.714 + 0.951*ВВП,%	3.338 + 0	2.595 + 0	3.583 + 1	9.394 + -	-2.8 + 0.

Нồi quy tuyến tính đa biến.

## Множественная линейные регрессия - Mutiple Linear Regression

R2 (R-квадрат) = 0.9562930660796466

Среднеквадратическое отклонение RMSE = 2.071279279820735

Форма уравнения  $y = 3.00 + 0.20 \cdot x_1 + -0.01 \cdot x_2 + -0.03 \cdot x_3 + 0.89 \cdot x_4 + 0.0003 \cdot x_5 + -0.06 \cdot x_6$

OLS Regression Results						
Dep. Variable:	Инфляция, дефлятор ВВП,%	R-squared:	0.956			
Model:	OLS	Adj. R-squared:	0.954			
Method:	Least Squares	F-statistic:	342.8			
Date:	Wed, 23 Aug 2023	Prob (F-statistic):	1.37e-61			
Time:	12:22:41	Log-Likelihood:	-216.86			
No. Observations:	101	AIC:	447.7			
Df Residuals:	94	BIC:	466.0			
Df Model:	6					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	3.0025	0.507	5.916	0.000	1.995	4.010
ВВП,%	0.1973	0.057	3.438	0.001	0.083	0.311
Экспорт,%	-0.0150	0.012	-1.240	0.218	-0.039	0.009
Денежная масса,%	-0.0329	0.030	-1.098	0.275	-0.092	0.027
Динамики ВВП по ППС,%	0.8898	0.046	19.380	0.000	0.799	0.981
Баланс текущего счета,%	0.0003	0.000	1.145	0.255	-0.000	0.001
Валовые национальные расходы,%	-0.0581	0.037	-1.579	0.118	-0.131	0.015
Omnibus:	88.918	Durbin-Watson:	2.283			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1765.516			
Skew:	2.453	Prob(JB):	0.00			
Kurtosis:	22.886	Cond. No.	2.14e+03			

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.14e+03. This might indicate that there are strong multicollinearity or other numerical problems.

Vì phần note của hồi quy tuyến tính đa biến có viết khả năng xảy ra đa cộng tuyến(multicollinearity) nên sẽ tính yếu tố lạm phát phương sai (variance inflation factor) > 10 thì xảy ra đa cộng tuyến sẽ giảm độ chính xác của mô hình và làm tăng sai số trong dự đoán, cũng như làm giảm khả năng kiểm định sự ảnh hưởng của từng biến độc lập đối với biến phụ thuộc.

Коэффициент инфляции дисперсии в модель с обычным методом наименьших квадратов - Variance inflation factor

	Показатель	VIF
1	ВВП, %	23.14
2	Экспорт, %	6.50
3	Денежная масса, %	4.04
4	Динамики ВВП по ППС, %	3.58
5	Баланс текущего счета, %	1.07
6	Валовые национальные расходы, %	6.66

$vif > 10$  nên xảy ra đa cộng tuyến → thay đổi mô hình

## Hồi quy Ridge

xác định với vài biến alpha để xem trường hợp nào tốt nhất. Vì  $r^2$  ở các trường hợp không cách nhau nhiều nên sẽ chọn RMSE nhỏ nhất

### Гребневая регрессия - Ridge Regression

Коэффициент инфляции дисперсии в гребневой регрессии с несколько alpha и соответственно среднеквадратическое отклонение (RMSE)

	alpha	ВВП, %	Экспорт, %	Денежная масс	Динамики ВВП	Баланс текущег	Валовые нацио	RMSE
1	0.0001	9.17156187	3.45841625	1.53187036	3.97258036	1.08053093	2.46717223	1.14074393
2	0.001	9.17156187	3.45841625	1.53187036	3.97258036	1.08053093	2.46717223	1.14074416
3	0.01	9.17156187	3.45841625	1.53187036	3.97258036	1.08053093	2.46717223	1.14074649
4	0.1	9.17156187	3.45841624	1.53187036	3.97258035	1.08053093	2.46717223	1.14076974
5	10.00	9.17152070	3.45831967	1.53186966	3.97251427	1.08052732	2.46713848	1.14340775
6	100.00	9.16760296	3.44983996	1.53180365	3.96667692	1.08023379	2.46427758	1.17371320

Результат в гребневой регрессии - Ridge Regression с  $\alpha = 0.0001$

$R^2$  (R-квадрат) = 0.9518165553543839

Среднеквадратическая ошибка RMSE = 1.1407439301674924

Формула уравнения :  $y = 3.07 + 0.19 \cdot x_1 + -0.01 \cdot x_2 + -0.04 \cdot x_3 + 0.89 \cdot x_4 + 0.0004 \cdot x_5 + -0.05 \cdot x_6$

## Tương tự hồi quy Lasso

Коэффициент инфляции дисперсии в Лассо с несколько alpha и соответственно среднеквадратическое отклонение (RMSE)

	alpha	ВВП, %	Экспорт, %	Денежная масс	Динамики ВВП	Баланс текущег	Валовые нацио	RMSE
1	0.0001	9.17156187	3.45841625	1.53187036	3.97258036	1.08053093	2.46717223	1.14075108
2	0.001	9.17156186	3.45841625	1.53187036	3.97258033	1.08053093	2.46717222	1.14081550
3	0.01	9.17156099	3.45841609	1.53187031	3.97257806	1.08053093	2.46717116	1.14146528
4	0.1	9.17147305	3.45839991	1.53186675	3.97235099	1.08053092	2.46706546	1.14833530
5	10.00	8.40028143	3.32495329	1.50115396	3.69265479	1.08049030	2.29571954	1.67379225
6	100.00	2.00574715	2.07585951	1.00346507	1.31324701	1.07651036	1.00166150	6.70582639

## Результат в регрессии Лассо - Lasso Regression с alpha = 0.0001

$R^2$  (R-квадрат) = 0.9518165553422577

Среднеквадратическая ошибка RMSE = 1.1407510832275918

Формула уравнения :  $y = 3.07 + 0.19 \cdot x_1 + -0.01 \cdot x_2 + -0.04 \cdot x_3 + 0.89 \cdot x_4 + 0.0004 \cdot x_5 + -0.05 \cdot x_6$

## So sánh kết quả của các mô hình hồi quy đa biến

Результат сравнения множественных регрессиональных модели

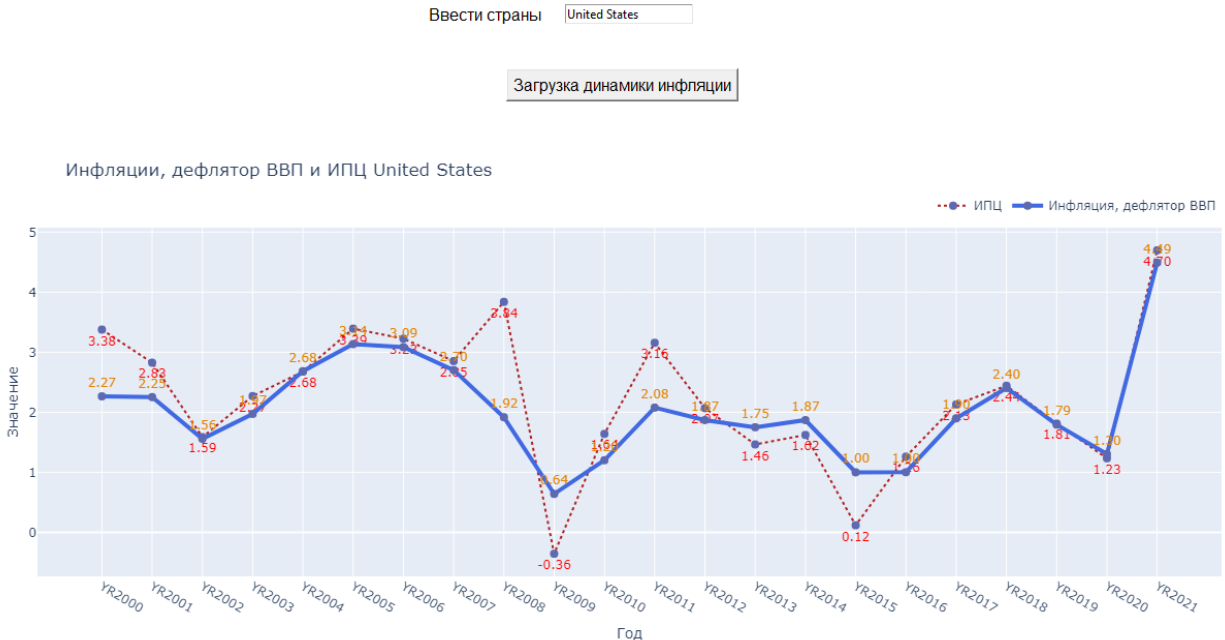
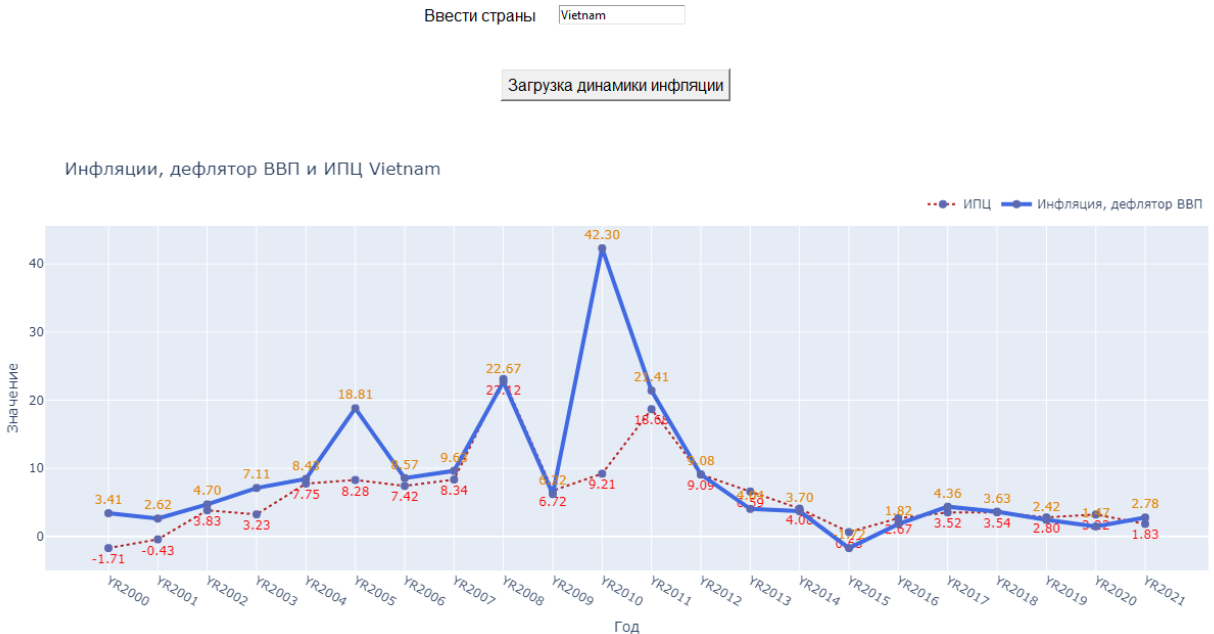
	Показатель	Множественная линейная регрессия	Гребневая регрессия	Регрессия Лассо
1	$R^2$ (R-квадрат)	0.9562930660796466	0.9518165553543839	0.9518165553422577
2	Среднеквадратическая ошибка RMSE	2.071279279820735	1.1407439301674924	1.1407510832275918
3	Формула уравнения	$3.00 + 0.20 \cdot x_1 + -0.01 \cdot x_2 + -0.03 \cdot x_3 + 0$	$3.07 + 0.19 \cdot x_1 + -0.01 \cdot x_2 + -0.04 \cdot x_3 + 0$	$3.07 + 0.19 \cdot x_1 + -0.01 \cdot x_2 + -0.04 \cdot x_3 + 0$

## Дữ liệu lạm phát theo giảm phát GDP từ 2000-2021 của các đất nước

Таблица по инфляции, дефлятор ВВП 2000-2021 всех странах

	Country	YR2000	YR2001	YR2002	YR2003	YR2004	YR2005	YR2006	YR2007	YR2008	YR2009	YR2010	Y
6	Argentina	1.04	-1.1	30.56	10.50	18.36	10.32	13.74	14.94	23.17	15.38	20.92	2
7	Armenia	-1.4	4.03	2.36	4.60	6.28	3.21	4.62	4.28	5.99	2.50	7.77	4
8	Aruba	1.03	-2.8	4.45	3.03	2.82	5.06	3.49	5.16	4.26	1.70	-1.2	4
9	Australia	2.56	4.62	2.84	2.97	3.24	3.82	5.12	4.98	4.62	5.00	1.16	6
10	Austria	1.36	1.95	1.15	1.31	1.74	2.54	1.89	2.22	1.96	1.89	0.87	7
11	Azerbaijan	12.49	2.52	4.21	6.96	9.25	14.72	11.30	21.03	27.78	-19	13.55	2
12	Bahamas, The	0.92	0.35	3.96	1.15	1.19	5.06	0.83	2.95	1.49	-1	-0.39	-
13	Bahrain	29.99	-3.4	3.41	8.60	10.99	13.73	8.84	8.44	11.37	-13	7.44	9
14	Bangladesh	3.45	3.26	3.89	5.82	4.56	4.59	5.88	6.47	7.86	6.76	7.14	7
15	Barbados	-1.3	2.26	0.91	1.12	5.83	6.67	4.06	8.57	1.72	-1.7	3.90	3
16	Belarus	185.29	79.53	44.89	30.69	22.68	18.97	10.75	12.87	21.21	9.26	11.29	7
17	Belgium	2.01	1.98	1.64	1.85	1.91	2.08	2.26	1.93	1.91	0.53	1.89	7
18	Belize	1.29	-0.79	0.86	-3.7	2.31	2.92	3.43	3.87	3.52	-2.4	2.17	4
19	Benin	4.34	1.97	3.55	3.06	0.77	4.29	2.09	0.39	6.43	2.54	0.88	3

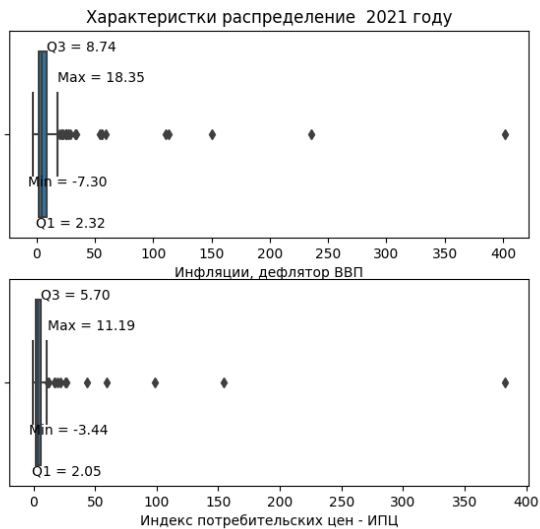
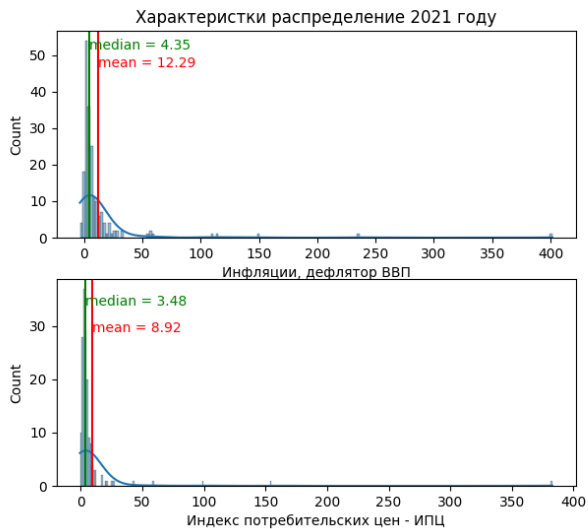
So sánh lạm phát theo giảm phát GDP và lạm phát theo CPI của từng nước và dưới đây là Việt Nam và Mỹ có thể đổi nước khác trong bộ dữ liệu



Hình dạng phân phối của dữ liệu năm 2021 và có thể tìm hiểu các năm từ 2000-2021

2021

### Характеристики распределения



Biểu đồ top 15 các nước có lạm phát(theo giảm phát GDP) và lạm phát theo CPI lớn nhất và nhỏ nhất

2020

Загрузка результат визуализации

