The file **index2022\_data.xls** contains data about 184 countries.

Methodology for calculating the values of variables:

<https://www.heritage.org/index/pdf/2022/book/02_2022_IndexOfEconomicFreedom_METHODOLOGY.pdf>

**Variables description**

* CountryID - ID of the Country
* CountryName - Name of the Country
* HighIncome - Belonging to high income group of countries (1 - belongs, 2 - no)
* Income - Belonging to the income group (low income, middle income, high income
* Development - Is the country developed or developing
* Region - Region of the country's location
* WorldRank - World Rank
* RegionRank - Region Rank
* 2022Score - 2022 Score
* PropertyRights - Property Rights
* JudicialEffectiveness - Judicial Effectiveness
* GovernmentIntegrity - Government Integrity
* TaxBurden - Tax Burden
* GovernmentSpending - Government Spending
* FiscalHealth - Fiscal Health
* BusinessFreedom - Business Freedom
* LaborFreedom - Labor Freedom
* MonetaryFreedom - Monetary Freedom
* TradeFreedom - Trade Freedom
* InvestmentFreedom - Investment Freedom
* FinancialFreedom - Financial Freedom
* TariffRate - Tariff Rate (%)
* IncomeTaxRate - Income Tax Rate (%)
* CorporateTaxRate - Corporate Tax Rate (%)
* TaxBurdenofGDP - Tax Burden % of GDP
* GovernmentExpenditurofGDP - Gov't Expenditure % of GDP
* Population - Population (Millions)
* GDP - GDP (Billions, PPP)
* GDP\_GrowthRate - GDP Growth Rate (%)
* 5Year\_GDP\_GrowthRate - 5 Year GDP Growth Rate (%)
* GDPperCapita - GDP per Capita (PPP)
* Unemployment - Unemployment (%)
* Inflation - Inflation (%)
* FDIInflow - FDI Inflow (Millions)
* PublicDebt - Public Debt (% of GDP)

| **Student** | **Linear regression: Dependent variable** | **Linear regression: Predictors** | **Cluster analysis** |
| --- | --- | --- | --- |
| Бабаев Расул Ильхам оглы | PropertyRights | HighIncome, Development, Region + not less than 5 continuous  variables which have the highest correlation with PropertyRights | PropertyRights + not less than 5 continuous variables which have the correlation with PropertyRights less than 0,5 |
| Баталова Лиана Ильмировна | JudicialEffectiveness | Income, Development, Region + not less than 5 continuous  variables which have the highest correlation with JudicialEffectiveness | JudicialEffectiveness + not less than 5 continuous variables which have the correlation with JudicialEffectiveness less than 0,5 |
| Белавенцев Валерий Евгеньевич | GovernmentIntegrity | HighIncome, Development, Region + not less than 5 continuous  variables which have the highest correlation with GovernmentIntegrity | GovernmentIntegrity + not less than 5 continuous variables which have the correlation with GovernmentIntegrity less than 0,5 |
| Беляева Сабина Александровна | TaxBurden | Income, Development, Region + not less than 5 continuous  variables which have the highest correlation with TaxBurden | TaxBurden + not less than 5 continuous variables which have the correlation with TaxBurden less than 0,5 |
| Бутаков Дмитрий Викторович | GovernmentSpending | HighIncome, Development, Region + not less than 5 continuous  variables which have the highest correlation with GovernmentSpending | GovernmentSpending + not less than 5 continuous variables which have the correlation with GovernmentSpending less than 0,5 |
| Волков Андрей Андреевич | BusinessFreedom | Income, Development, Region + not less than 5 continuous  variables which have the highest correlation with BusinessFreedom | BusinessFreedom + not less than 5 continuous variables which have the correlation with BusinessFreedom less than 0,5 |
| Гадиев Михаил Искандерович | LaborFreedom | HighIncome, Development, Region + not less than 5 continuous  variables which have the highest correlation with LaborFreedom | LaborFreedom + not less than 5 continuous variables which have the correlation with LaborFreedom less than 0,5 |
| Гончаров Дмитрий Андреевич | MonetaryFreedom | Income, Development, Region + not less than 5 continuous  variables which have the highest correlation with MonetaryFreedom | MonetaryFreedom + not less than 5 continuous variables which have the correlation with MonetaryFreedom less than 0,5 |
| Данилов Алексей Андреевич | TradeFreedom | HighIncome, Development, Region + not less than 5 continuous  variables which have the highest correlation with TradeFreedom | TradeFreedom + not less than 5 continuous variables which have the correlation with TradeFreedom less than 0,5 |
| Джамбонг Тенке Ханк-Дебэн - | InvestmentFreedom | Income, Development, Region + not less than 5 continuous  variables which have the highest correlation with InvestmentFreedom | InvestmentFreedom + not less than 5 continuous variables which have the correlation with InvestmentFreedom less than 0,5 |
| Донская Софья Алексеевна | FinancialFreedom | HighIncome, Development, Region + not less than 5 continuous  variables which have the highest correlation with FinancialFreedom | FinancialFreedom + not less than 5 continuous variables which have the correlation with FinancialFreedom less than 0,5 |
| Жданов Вадим Борисович | TariffRate | Income, Development, Region + not less than 5 continuous  variables which have the highest correlation with TariffRate | TariffRate + not less than 5 continuous variables which have the correlation with TariffRate less than 0,5 |
| Зубарев Антон Анатольевич | IncomeTaxRate | HighIncome, Development, Region + not less than 5 continuous  variables which have the highest correlation with IncomeTaxRate | IncomeTaxRate + not less than 5 continuous variables which have the correlation with IncomeTaxRate less than 0,5 |
| Калинин Антон Игоревич | CorporateTaxRate | Income, Development, Region + not less than 5 continuous  variables which have the highest correlation with CorporateTaxRate | CorporateTaxRate + not less than 5 continuous variables which have the correlation with CorporateTaxRate less than 0,5 |
| Карелина Елена Андреевна | TaxBurdenofGDP | HighIncome, Development, Region + not less than 5 continuous  variables which have the highest correlation with TaxBurdenofGDP | TaxBurdenofGDP + not less than 5 continuous variables which have the correlation with TaxBurdenofGDP less than 0,5 |
| Колчанова Алина Викторовна | GovernmentExpenditurofGDP | Income, Development, Region + not less than 5 continuous  variables which have the highest correlation with GovernmentExpenditurofGDP | GovernmentExpenditurofGDP + not less than 5 continuous variables which have the correlation with GovernmentExpenditurofGDP less than 0,5 |
| Костюченко Илья Игоревич | Inflation | HighIncome, Development, Region + not less than 5 continuous  variables which have the highest correlation with Inflation | Inflation + not less than 5 continuous variables which have the correlation with Inflation less than 0,5 |
| Кугай Виктор Евгеньевич | GDP\_GrowthRate | Income, Development, Region + not less than 5 continuous  variables which have the highest correlation with GDP\_GrowthRate | GDP\_GrowthRate + not less than 5 continuous variables which have the correlation with GDP\_GrowthRate less than 0,5 |
| Кумар Махеш - | 5Year\_GDP\_GrowthRate | HighIncome, Development, Region + not less than 5 continuous  variables which have the highest correlation with 5Year\_GDP\_GrowthRate | 5Year\_GDP\_GrowthRate + not less than 5 continuous variables which have the correlation with 5Year\_GDP\_GrowthRate less than 0,5 |
| Мохаммад Али - | Unemployment | HighIncome, Development, Region + not less than 5 continuous  variables which have the highest correlation with Unemployment | Unemployment + not less than 5 continuous variables which have the correlation with Unemployment less than 0,5 |
| Редникина Дарья Юрьевна | GDPperCapita | Income, Development, Region + not less than 5 continuous  variables which have the highest correlation with GDPperCapita | GDPperCapita + not less than 5 continuous variables which have the correlation with GDPperCapita less than 0,5 |
| Савельев Семен Юрьевич | Inflation | Income, Development, Region + not less than 5 continuous  variables which have the highest correlation with Inflation | Inflation + not less than 5 continuous variables which have the correlation with Inflation less than 0,5 |
| Соколов Семен Константинович | FDIInflow | HighIncome, Development, Region + not less than 5 continuous  variables which have the highest correlation with FDIInflow | FDIInflow + not less than 5 continuous variables which have the correlation with FDIInflow less than 0,5 |
| Суворов Николай Михайлович | PublicDebt | Income, Development, Region + not less than 5 continuous  variables which have the highest correlation with PublicDebt | PublicDebt + not less than 5 continuous variables which have the correlation with PublicDebt less than 0,5 |
| Филимонов Артём Александрович | GDP | HighIncome, Development, Region + not less than 5 continuous  variables which have the highest correlation with GDP | GDP + not less than 5 continuous variables which have the correlation with GDP less than 0,5 |
| Шорина Анна Алексеевна | FiscalHealth | HighIncome, Development, Region + not less than 5 continuous  variables which have the highest correlation with FiscalHealth | FiscalHealth + not less than 5 continuous variables which have the correlation with FiscalHealth less than 0,5 |

Please select variables for the analysis based on the information from the above table and consider the following tasks.

**Task 1**

* Do the statistical descriptive analysis of all the variables which will be used in regression analysis. Each quantitative output should be accompanied with your textual interpretation of the obtained statistics.
* Analyze all pairwise relationships between the variables which will be used in regression analysis based on the suitable statistical methods. Make conclusions about each obtained result.
* Do preliminary factor analysis based on continuous predictors. If it’s not applicable for the selected data explain why. If it is applicable save factors, describe them and give them names. Use factors as predictors in regression analysis instead of original continuous variables.
* Do the linear regression analysis. Formulate all research hypotheses which will be verified during the data analysis.
* Specify the regression equation. Assess the goodness-of-fit of the model.
* Describe the relationships between each predictor and dependent variable in detail. Explain the meaning of each regression coefficient.
* Perform the model diagnostics.
* Make final conclusions about the results of analysis.

**Task 2**

* Do the statistical descriptive analysis of all the variables which will be used in cluster analysis.
* Select a suitable cluster algorithm. Explain your selection.
* Define the number of clusters. Explain your decision.
* Describe the clusters. Each cluster should have its profile: name and detailed description in textual form based on the calculated statistical characteristic. Explain your selection of names.

**Task 3**

We are now concerned with the issue of increasing the number of highly qualified IT-specialists in the country. One of the initiatives is to develop training of IT-specialists at different educational levels. The objective is to both increase the number of students in IT-specialties and the quality of their training.

Propose a methodology of a quantitative research on the basis of which some recommendations could be given on how to achieve the above mentioned tasks. Formulate research questions and hypotheses. How the research sample could be formed? What kind of variables could be used in the analysis (list possible variables)? What methods of quantitative data analysis could be applied?

Please send the Jupyter Notebook with comments and answers to amelikyan@hse.ru. The task should be done **individually**.The deadline for submitting the task is **16 April 16:00.**