Ada - Aitor - Meysam

### Introduction **(everyone)**

In this section we will explain the basics of the project: the **description** of the problem, our personal **motivation** for it and our **main objectives**.

#### Description **(everyone)**

*The Heritage Foundation* is a think tank focused towards public policy and it’s known for their distribution of global statistics and political science research. Its most important publication is the annual **Index of Economic Freedom** which will be our project basis.

Economic freedom is defined as the “the fundamental right of every human to control his or her own labor and property” [heritage.org]. The Economic Freedom dataset compiles many economic measures like GDP as well as different expert-crafted “indexes” that sum up fundamental indicators like *fiscal health* of 186 countries in the world.

We have the following variables per country:

* **Indexes** ⇒ Economic Freedom, Property Rights, Judicial Effectiveness, Gov Integrity, Tax Burden, Gov Spending, Fiscal Health, Business Freedom, Labor Freedom, Monetary Freedom, Trade Freedom, Investment Freedom and Financial Freedom.
* **Measures** ⇒ Tax Burden, Gov Expenditure, Population, GDP, GDP Growth, PPP, Unemployment, Inflation, FDI and Public Debt.
* **Others** ⇒ Name and Region.

The complete dataset can be found at <https://www.heritage.org/index/download>.

First of all, the dataset that we have chosen is the The Economic Freedom Index Data Set. It can be found in the Kaggle Repository , and also it is available online on the following URL (The link is the original one of this data set):

https://www.heritage.org/index/excel/2019/index2019\_data.xls

We have data associated with The Index of Economic Freedom which is the product of strong association with societies and organizations around the world. Advantages of this approach can be upgrading in economic freedom, success and promote these concepts in their homes, institutes, and communities over two decades. “the Index evaluates economic policy developments in 186 countries. Countries are classified and graded on 12 measures of economic freedom which are quantitative and qualitative factors”[heritage.org].

The Economic Freedom Index Dataset has 34 explanatory variables (29 of which are continuous and 5 categorical ). These explanatory variables are:

1- CountryID (cat.)

2- Country Name (cat.)

3- WEBNAME (cont.)

4- Region (cat.)

5- World Rank (cont.)

6- Region Rank (cont.)

7- 2019 Score (cont.)

8- Property Rights (cont.)

9- Judical Effectiveness (cont.)

10- Government Integrity (cont.)

11- Tax Burden (cont.)

12- Gov't Spending (cont.)

13- Fiscal Health (cont.)

14- Business Freedom (cont.)

15- Labor Freedom (cont.)

16- Monetary Freedom (cont.)

17- Trade Freedom (cont.)

18- Investment Freedom (cont.)

19- Financial Freedom (cat.)

20- Tariff Rate (%) (cont.)

21- Income Tax Rate (%) (cont.)

22- Corporate Tax Rate (%) (cont.)

23- Tax Burden % of GDP (cont.)

24- Gov't Expenditure % of GDP (cont.)

25- Country (cat.)

26- Population (Millions) (cont.)

27- "GDP (Billions, PPP)" (cont.)

28- GDP Growth Rate (%) (cont.)

29- 5 Year GDP Growth Rate (%) (cont.)

30- GDP per Capita (PPP) (cont.)

31- Unemployment (%) (cont.)

32- Inflation (%) (cont.)

33- FDI Inflow (Millions) (cont.)

34- Public Debt (% of GDP) (cont.)

We have 186 observations and there are some missing values (they are just in continuous variables).

#### Motivation **(everyone)**

We have chosen this dataset because we’re very interested in understanding how the economy works in different countries and how different factors influence it. We’re also curious to know if geographical regions have similar characteristics within them or if there is more variety than meets the eye. Each one of us is from a different part of the world and has travelled to many other places so we’d love to test our experiences with actual data.

Furthermore it’s also a very interesting dataset practically. Even though we don’t have many samples (we’re limited by the number of countries), it’s very high quality data and its conclusions can be very relevant if the analysis are done carefully.

We have chosen this dataset because it is an interesting classification problem related to economic. The problem is not an easy one neither a very complex, so we think that with it, we will apply some of the multivariate techniques explained in class and learn how to reveal the hidden information contained in the dataset. Also,We found this dataset in kaggle repository, but then decide to search and find the most original one of this dataset which comes from the official website of it (www.heritage.org), so it is a well-studied problem and it is useful for academic purposes. The dataset has enough explanatory variables (34) and the number of observations is enough for this type of problem (187), so we think that we will not have any problem with this dataset.

Nowadays, almost all the economic systems are based on capitalism. Capitalism is synonymous with the economic freedom and economic freedom is measured by the index of freedom. The coherence of the information, the usefulness and the importance of the economical index of freedom is what motivated us to choose this dataset. Another interesting motivation point might be the emigration. If any day we decide to move to another country we have to have some information about the economy. If we want to have the freedom to choose what we want to do as private owners, we have to refer to this dataset. By analyzing this dataset we can have a clear picture about what to expect from the country we have decided to move on or even more, we can choose the best place to live in and give life to our economical projects or businesses. This dataset provides useful information about different factors that directly or indirectly have a significant impact on economy. In other words, economy is equal to quality of living.

#### Objective **(everyone)**

…

### Pre-processing

…

Summary **(Aitor - Week 20-26)**

e.g. using summary(D)

#### Missing Values **(Meysam - Week 20-26)**

North Korea, Liechtenstein, Syria and Somalia => remove them, too many NA

Yemen => try to use imputation

#### Outliers **(Ada - Week 20-26)**

Check if there are outliers and decide if we have to exclude them from the analysis.

#### Feature Selection **(Aitor - Week 20-26)**

Check correlations with objective. Remove variables that have close to 0 correlation and no common-sense relationship.

#### Feature Extraction **(Ada - Week 20-26)**

Try to find a relationship between variables that is meaningful. (best if the models end up using it)

### Validation

…

### Exploratory Analysis

Using all columns? Using only “economical indexes”? (orange columns)

#### PCA **(Meysam)**

...

#### Clustering

...

#### Interpretation

...

### Methodology

(e.g. Train-Test Discussion)

### Modeling

We will try to predict the region for a given input.

#### Decision Trees

...

#### Multinomial Logistic Regressionç

...

#### … (?)

### Results

(i.e. generalization)

### Conclusions

...