**Impact of Corruption on Development**

It is relatively safe to say that where there is power, there is corruption. While most everyone believes it's wrong, some say corruption is 'necessary evil' because the growth is too slow without it. In this analysis, we will look into the impact of corruption on the development of different countries. I used CPI data from transparency.org and World Development Indicator data from the world bank to train a machine learning model to find the essential features and analyze the relationship to the CPI scores.

**Data sources**

* Corruption Perception Index (CPI):

https://www.transparency.org/en/cpi/2020/index/nzl

* World Development Indicators (WDI):

https://datatopics.worldbank.org/world-development-indicators/

**Note:**

The first attempt, a linear regression model, failed to meet the assumptions and was replaced by a random forest classifier.

**Analysis with variables: (line500)**

1. Mean: 42.90 / Median: 38.00 - The majority of countries are severely corrupt.
2. Improved: 62/Magnitude: 108.0, Declined: 48/Magnitude: -76.0 - Overall CPI score is increasing
3. For most of the countries, CPI score does not fluctuate over 30%
4. WE/EU region has the highest CPI scores, ECA has the lowest, and AP's spread out.

**CPI score and below variables generally move together:**

1. GDP per capita

* Wealthier countries have less corruption.

1. Domestic credit to private sector by banks

* How much money the central bank pumps into the economy

1. Price level ratio of PPP conversion factor (GDP) to market exchange rate

* Purchasing power of the currency
* It tends to go down. (inflation)
* Low value usually means poor, and CPI scores are high in these countries.

1. Population density (people per sq. km of land area)

* When it is high, CPI score is also high
* When it moves the CPI score generally moves with it.

1. Consumer price index, GDP deflator

* These are related to cost of living.
* Fluctuation means unstable, and it comes with lower CPI scores

1. Refugee population by country or territory of origin

* Refugees come from countries in crisis, and these countries tend to have relatively low CPI scores.

1. Net secondary income (Net current transfers from abroad) (current US$)

* Money transfer between residents and non-residents
* Lower means more money is sent out.
* The lower the more foreigners in the countries, possibly resulting lower value
* Lower could imply the country is wealthier, and in many cases high CPI score.

**CPI scores with below have weak relation:**

1. Unemployment, total (% of total labor force) (modeled ILO estimate)