JEDHA PROJECT ESSENTIALS
Deliverable

NAME: André de Oliveira Gomes

Title of the Project:

Anomalies of temperatures in RJ and rainfalls of SP

Client profile:

Our client is typically a think tank company that wants to start from the scratch to build a decision model to economise the water reservoirs of the state of Rio de Janeiro in the upcoming years. The reason is to try to avoid calamity situations such as the one that happened in SP during the period of 2014-2017 in the absence of rainfall. The critical anomaly temperature that is taken as threshold is the value 0.62. The choice was made a prior as a legitimate benchmark for a deviation of temperature since it was the value registered at 2010, the year that ends the warmest decade registered in the globe.
See https://www.ncdc.noaa.gov/sotc/global/201013.

Questions:

1. How bad were the levels of precipitation in SP (2013-2016)?
2. How comparable were the values of the temperature anomalies in SP in that period with historical records?
3. How do the temperatures on RJ relate and how to predict high occurrences of temperature anomalies bigger than 0.62?

Sources:

- Temperatures (monthly 1961-2017): https://data.giss.nasa.gov/gistemp/
- Pluviosity SP (annual 1961-2017): https://www.kaggle.com/arkanius/rain-intensitty-in-so-paulo/metadata
- CO2 (annual 1961-2017):
- https://www.kaggle.com/srikantsahu/co2-and-ghg-emission-data?select=emission+data.csv

Methodology:

1. Anomalies of temperatures: 3 data sets of different Brazilian cities were used: SP, RJ and Manaus. The benchmark temperature for each city was the median temperature registered monthly in each data set. A temperature anomaly is the absolute value of the deviations between the registered temperatures and the benchmark temperature.
2.
3. **Import and cleaning of the data.** We dealt with different data sets and there was some previous work to create two final data frames: one for the annual rainfall precipitation/mm^3 in SP, the annual averaged anomalies of temperatures registered and the global levels of emission C02 in Brazil; the second containing the anomalies of temperatures registered in SP, RJ and Manaus on a monthly basis. Some non negligible amount (as it is commented in the code) of entries on the data sets were missing and several adjustments of the data types and corresponding missing values were made during the cleaning phase.

4. **Data visualisation** using matplotlib- Pyhton combined with several graphics in Tableau. The relevant graphics are shown in the diaporama and on the code. Exploring visually the data we draw several conclusions: the critical values of rainfall in the period after 2000 were registered in 2014 with a tendency of high decrease starting in 2010. At same time the anomalies of temperatures in SP

rise with an almost monotonic tendency accomplished by the anomalies of temperatures in RJ.

5. **First statistics:** the annual rainfalls in SP and the emissions of C02 do not show correlations with the annual rainfalls registered in SP. This is a point of valid discussion such as it was made on the code file. We used Spearman coefficient calculations since we have not a very big data set of annual registers and we can not assume that the underlying distributions are Normal (we can not perform the Central Limit theorem since the observed annual values are in limited number). Nevertheless the Spearman coefficients between the the anomalies of temperatures are high as commented in the code which suggested us the use of regression techniques in order to predict models.

6. **Predictions using regressions:** we use multilinear regression between the variables anomalies of temperatures for SP,RJ and Manaus. The regression test presented high score (R^2 above 80%) and the model predicts that the anomaly temperatures of RJ follow the same speed as the anomaly temperatures of SP under this model. Please see diaporama and code.

7. **Predictions using the logistic regression.** Under the logistic model it is concluded with high (almost 90%) of precision and accuracy on the data set (see the diaporama and the code file please for the numbers and precise structure in the comments of the implementation of this algorithm) the probability of getting a anomaly temperature above 0.62 is modelled by a logarithmic relation between the temperature anomalies of SP and Manaus where the occurrence of anomalies bigger than 0.62 occur often (check confusion matrix in the code file). This means that if in the near future we expect time series of anomalies in the temperatures of Rio similar ou “very close” of parts of the data set that is treated then we can expect that under this model those anomalies are bigger than the critical value 0.62 which was the anomaly registered in 2010, a remarkable global warm year that closed the warmest decade registered on the planet.

Comments:

This deliverable is just descriptive and does not include graphics or quantitative results that were achieved during this exploratory analysis of the data. Climate (modelled in local geographical or global geographical scales) is highly nonlinear and mathematically complex, not being by nature in any way able to be captured by linear models. For the future it is relevant to rethink in the way to associate the variables of temperature anomalies in SP with the variables of C02 emissions and the annual rainfalls in different ways and using different and larger data sets. The first data frame that we used to do a first analysis is too small and also other importante variables such as deforestation area of Amazonia, atmospheric pressure and exogenous climatic variables such as proxy-data given by the gulf stream should be taken into account.