Computational Analysis of the Transformation of Former Colonial Countries

Abstract

This paper does a computational analysis to understand the extent to which former colonies have been transformed due to colonial rule. This extent is represented by the response variable: Colonial Transformation Score, from a scale of 0 to 100. This paper uses multivariate ridge regression analysis to identify the significant variables impacting the transformation and to what degree these variables are influencing. This paper also considers how the duration of colonialism impacts the level of transformation – with the KMeans Clustering algorithm. Interestingly, we identified that there were two distinct waves of colonialism, one lasting for 100-200 years and other from 300-400 years. This aligns with the literature suggesting there were two eras of colonialism, the Mercantilist wave and the Imperialist wave with different colonial duration.Finally, using the ANOVA test, this paper compares the difference in transformation by the French versus British colonial rule. There was no significant difference found within the groups, suggesting the extent transformation is not determined by who performs the act, but more so by how they executed it during their rule.

This is a link to an interactive visual created for the purposes of data exploration to see how specific colonies have been impacted during colonial rule and what their individual observations have within them https://public.tableau.com/app/profile/aditi3951/viz/TheTruthofBritishColonialism/Dashboard_js4?publish=yes

I. INTRODUCTION

This research investigates the following question: How can we compare the economic performance of Colonised countries during periods of colonialism and now. Colonialism is an issue that has been prominent not just in history, but it is relevant to the economic performances of countries today (Hutcheon). For example, through British curriculum still being the dominant education systems or the obsession of fair-skin in previously colonized countries, it's clear the impacts of colonialism have yet to fade away. During periods of colonialism, many countries were stripped of their natural resources, exploited of their manpower and forbidden the right of an education, which is why these countries are seen to be 'less well off' or as 'developing' today.

However, these countries have a reason for their stunted development, and this reason is going to be explored in this paper. This research paper will use the Colonial Transformation Dataset from the Harvard Dataverse to assess the transformation a country went through during its period of colonialism – socially, politically and economically. Broadly speaking, this paper will explore whether colonialism was actually helpful or was it harmful to the economies of the colonised countries. The larger impact of this exploration is that in history, oftentimes, there is a lack of quantitative analysis and there is a large emphasis on stories which have been passed down – often solely by the winners. But that isn't always reality. This project will explore and show us what the reality of colonialism was for these counties, and seek to understand how we should engage with these topics in the present.

II. LITERATURE REVIEW

This literature review outlines the background of colonialism and attempts to explain the everlasting impact of colonizers on previously colonized countries till this day. To do so, this literature review first explains what colonialism is. This section elaborates on what exactly colonialism is and how prevalent it was in the past. From there, this paper will explain what the economic impact of colonialism was in the past. This section explains how colonialism was operational by the colonizers to exploit colonized countries for their resources and how this degradation of resources has impacted the economic development of these countries in the long term. And the last section within the literature review will be to understand the legacy of colonialism – how it's present today despite it being a historical event of the past. This looks into facets of education, economic and social norms that are present till this day but can be traced back to colonial times. From there, this paper establishes why this form of historical research is relevant to the present world today and discusses its implications on the larger world today. Ultimately, we are understanding the past of colonialism, attempting to explain the present, and learning from this journey as we design future policies for education and institutions moving forward.

A. *Understanding Colonialism*

Colonialism is a complex, complicated and nuanced event that has not just been an event of the past, but one that is present till today. When attempting to understand it, we can think of colonialism as a practice of domination, which involves the subjugation of one people to another. (Kohn et al., 2006). This involves the political and economic control over a dependent territory and this is not a modern phenomenon. In fact, most of world history has been full of colonization, where one society gradually expands by incorporating another territory and settling its people on newly conquered territory. While it's been around since the ancient Greeks, modern European colonialism moved large numbers of people across the ocean and forced locals from the conquered to act as slaves for their own wars. Therefore modern colonialism can be defined as the conquest and control of other people's lands and goods. (Loomba, Ania, 2007).

Ultimately, modern colonialism did more than just extract goods and wealth from the countries it conquered, it restructured their economies of the latter, drawing them into a complex relationship of their own, so there was a flow of human and natural resources between them. This flow went back and forth, where the slaves, labor and natural resources from the colony were transported to the coloniser to manufacture goods and provide for the European markets. This flow consisted of people and profits, all for the benefit of the colonizer at the expense of the colony's culture, economy and political structures. In order for this to take place, modern colonizers used domination techniques (often with militarisation) and produced economic imbalance and dependency within the colony so it could no longer stand on its own. Ultimately, it is the life of the colony that funded the colonial expansion and development of modern Europe today.

B. Understanding the Economic Impact of Colonialism

Given the nuanced explanation of colonialism above, it is clear that colonialism was a phenomenon that encompassed socio-political and economic thoughts – ultimately colonialism began as trade. And then evolved to wealth, developing larger markets with greater access to resources and labor. It all started with economics and ended with a humanitarian and social dent on most colonized countries. If we zoom into the economic inequality we observe in the world today, we know that it didn't happen from recent times. In fact, it is the path dependent outcome on many historical processes, with the most important being European colonialism. (Acemoglu et al., 2017). When we go back to pre-colonial times, we see little inequality between rich and poor countries (a factor of four), and now, post colonialism, there is a difference of a factor of 40 when comparing the richest to the poorest countries in the world. And this can be traced to the exploitation of resources from the colonized countries in the past to the developed countries of the present.

The main economic impact on the colonized countries is the 'drain of wealth', expropriation, the control over production and trade, the exploitation of natural resources, and the improvement of infrastructure – done for the benefit of transportation for the colonizers. (Carey and Simon, 2012). This had led to the outflow of financial resources lasting the entire period of colonialism and one that has never been paid back or given reparations for till the present day. According to the economic historian (Maddison, 2008) , if these funds had been invested in India, this would have made a significant contribution to raising income levels and the GDP per capita in India at present time.

C. The Legacy of Colonialism

Western colonialism still places significant influence on many communities around the world. (Olsson and Ola, 2009). According to these seminar works, economic literature traces the fundamental reason for persistent underdevelopment and stagnant economic growth back to weak institutions that countries inherited from colonial times. Clearly, despite colonialism being an event from the past, its impacts are present till this day in many parts of the world – economically and politically. When looking at the colonial determinants of democracy, it shows that the general positive relationship between colonial duration and current levels of democracy is mainly driven by British former colonies and by countries colonized after 1850. It further shows that the colonized countries have a lower level of democracy, possibly because colonial influence left many countries with a Western mindset of political structure with no time or resources to develop it once their colonial rule ended. This is a clear example of how we can back issues of underdevelopment and stunted political growth back to colonial times, and the fact that this issue hasn't been resolved and continues to be prominent is the very legacy of colonial rule.

Even outside of just economic and political thought, colonialism continues to influence our social norms. For instance in India, the love and idolization of "white as good and beautiful", where fair skin is the epitome of beauty can be traced back to the time of the British Raj. (Majidi and Khesraw, 2020). They write that the British colonialist agents deep-rooted the Western values, cultures and beauty ideals in Indian society by restoring and weakening the historical Indian values, cultures and practices that lasted in India for centuries. And till this day, their colonial influence and legacy lives on.

Based on the literature review and the discussion of its findings above, we can conclude that colonialism and its influence is still present today. To begin with, colonialism is a definition that's complicated – because it's more than just territory, it's about the domination and imperialism placed from one community onto another. But this domination isn't just political, it's economic exploitation and often involves the social conditioning of the colonized. We see from the sources above that colonialism had not just degraded the colonized in the past, but continues to do so for countries today. Many developing countries facing issues of economic inequality and political instability can be traced back to colonial times – where precolonial their economic performance was healthy and growing, after the colonial duration, their structures collapsed and till this day are trying to be rebuilt. This legacy of degradation economically, socially and politically that colonialism has left behind is still clear in sight. And with these theories and understandings, my research aims to find the data to quantify these claims, as well as explore relationships between variables to understand how colonialism impacted different countries.

The findings of this research will certainly have larger implications because if we can find or develop a quantifiable relationship between colonial times and the current economic standings of countries today, this historical research and pattern can be used to develop present institutional policies. For instance, by showing that economic development has been stunted by colonial rule in the past, perhaps countries can ask for reparations to invest in their current economies. Alternatively, by seeing how colonial times influenced political systems in the past, countries can reevaluate their systems in place, and decide for themselves what political system they choose to live in, rather than resorting to what was left to them by their colonial rulers.

III. Methods

By using the colonial dataset, I had created a dataframe using Pandas I had cleaned that dataset where each observation will be information about a previously colonised country. There will be variables detailing information about its current economic performance as well as variables showing information about their colonial past (such as Colonial Duration, Who was the Colonizer, Social Transformation score, Political Transformation Score, Colonial Transformation Score). The variables detailing the scores of to what extent colonialism transformed the country after the colonizer had left (by looking at political structures, economic growth and social change). The key variables within each observation would be the Colonial Transformation, Country Name, Social, Political and Economic Transformation Scores. I have chosen to look into these scores because this assessment of socio-political and economic transformation was created by historians in Harvard who had been the ones to create this dataset themselves. Since they were the ones who collected, transformed and recorded the initial primary sources into the dataset being currently used, I feel confident using these assessments of transformations as they are the creators and subject matter experts of this field of study. This dataset will include different countries and rather than time, the duration of colonial rules endured by them. As explained earlier, this research will be tracing back the historical roots of colonised countries and evaluating their economic performance as presence. This explores a potential causal relationship. This will contain all observational data from what has been collected from previous secondary sources.

A detailed list of the variables used and their descriptions can be seen below in Table 1.

Regarding the ethical concerns coming from the data, while this has all been vetted by historians who created this dataset, ultimately the Colonial Transformation dataset is a collection of many primary sources that have been quantified by Historians. In this process and given the nature of the discipline of History, there is inherent human bias that must have been brought into the process of quantification of the dataset by the historians. Ultimately, it was still Western historians who are traced back to being colonizers evaluating these sources. And many of the historical sources related to colonialism have been erased or written solely by winners. Given this nature of historical data and research, there are undeniable biases we need to take into account while interpreting and implementing this research. Other than that, the paragraph above has detailed the variables we

Variable	Data Type	Description
COLYEARS	Numerical	The duration of colonial rule of the country
Violent Colonization, Wars of Defence etc;	Numerical	Scale from 0-2 to determine how violent colonial rule
Form of Colonial Domination	Numerical	Scale from 0-4 for the intensity of colonial domination
Colonial Border Split	Numerical	Scale from 0-4 determining the intensity of colonial split
Gold/Silver mining during Colonial Rule	Numerical	Scale from 0-2 to measure mineral exploitation
Colonial violence total	Numerical	Assessment of how much a colonial country endured violence
Social Transformation Score (0-100)	Numerical	Assessment of change in a country's Social structures
Political Transformation Score (0-100)	Numerical	Assessment of change in a country's Political structures
Economic Transformation Score (0-100)	Numerical	Assessment of change in a country's Economic structures
Colonial Transformation Score (0-100)	Numerical	Assessment of how overall colonial change in a country

TABLE I

VARIABLES USED AND THEIR DESCRIPTIONS

will be using. These variables have been chosen through background research and historical theories suggesting which variables have been the most influential in the preservation of colonialism. Through phases of data exploration, I hope to find more variables that are significant and make an accurate assessment.

Given these considerations, the methods used in this research can be explained through data exploration, the analysis using different machine learning and regression techniques, followed by its results and the discussion of its findings.

IV. DATA EXPLORATION

As explained under the methods section, this dataset deals with former colonies across the globe by different colonizers. Given the background information and the historical theories explained under the domain review, the colonizer plays an important role in how much a country was impacted by colonization. To begin parsing through the dataset, figure 1 shows a map of all the different former colonies in the regions of Africa and Asia in its history.

From this image, we can see the different regions and we can find data for former colonies – in this case, the color encoding shows the different colonial transformation scores. The countries with a darker color show a greater transformation by colonization in their present state. We can break down this regional analysis by comparing what were the different regions taken by different colonisers – in this dataset, the data is primarily present for former British and French colonies. Therefore, figure 2 and 3, show the different regions of colonies by the French and British, once again

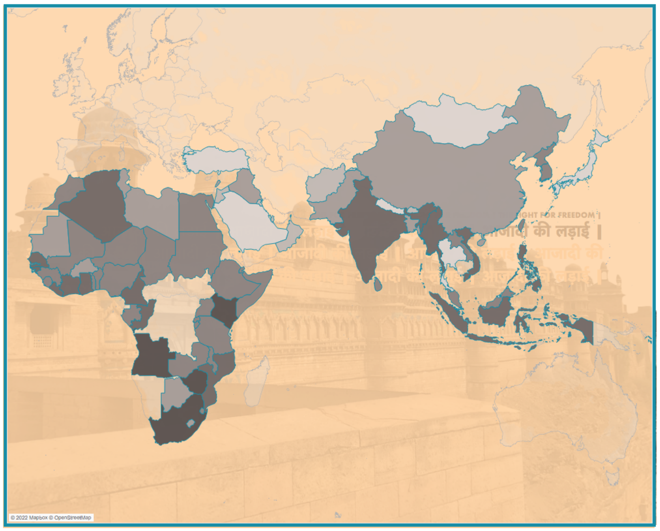

Fig. 1. The regional spread of Colonies across the African and Asian regions

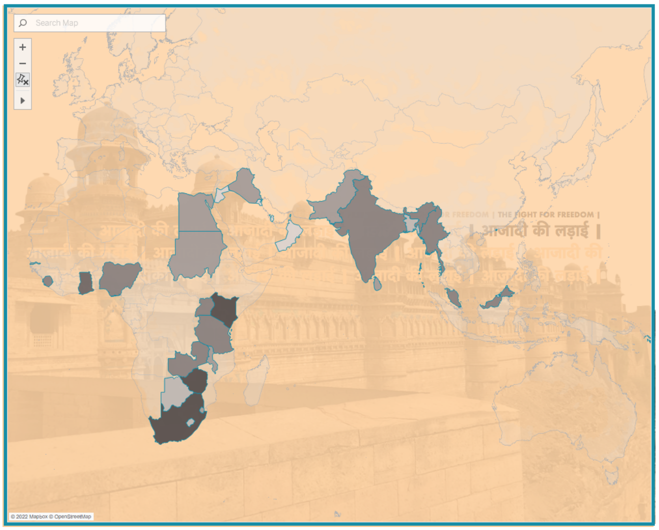
Search Map
THE FIGHT FOR FREEDOM | आज़ादी की लड़ाई |
© 2022 Mapbox © OpenStreetMap

Fig. 2. The regional spread of British Colonies

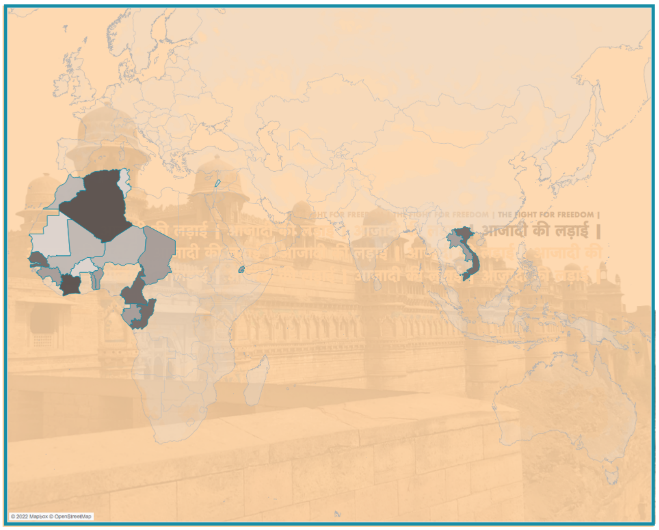
THE FIGHT FOR FREEDOM | आज़ादी की लड़ाई |
© 2022 Mapbox © OpenStreetMap

Fig. 3. The regional spread of French Colonies

encoding by color to show transformation.

We can see from these images that the French had a great regional influence over the African countries and the British with Asian countries. By taking our data and splitting the data points into former British and French colonies, under the Data Analysis section we have conducted an ANOVA test to compare the two samples – to explore if the extent of colonial transformation was propelled by the British or the French.

For further exploration of the data – looking at how each colonies political, economic and social structures were impacted through the duration of colonialism, there is an interactive link through the Tableau public placed in the abstract of the first page. By clicking on a specific country, we can explore its associated variables to get an understanding of its historical past. This tool allows us not to just stick to the patterns we notice through our analysis in this paper, but highlights the story and significance of each country's history and implication from its colonial rule.

Mentioned under methods, my response variable from the dataset is the Colonial Transformation Score, an assessment of how much a colony has been impacted. In order to identify the independent variables that are influencing this response variable, this paper uses the Pearson Correlation Coefficients to create a correlation metric to identify variables which are correlated to my response variable. With the correlation coefficients, we can measure how strong the relationship is between the response variable and other independent variables. This study uses Pearson's correlation because the independent variables identified will then be used for the multivariate regression analysis. This ranges from -1 and 1, with -1 being the strongest negative correlation and 1 is the strongest positive correlation. The Pearson correlation coefficient can be calculated using Equation 1 below.

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}} \tag{1}$$

From these calculations, figure 4 shows the correlation coefficient metric of the different independent variables against the response variable.

After looking at the different correlation coefficients and its relation to the response variable, the next section explains the analysis.

We can also see in figure 5 the distribution of the response variable values. As seen from figure 5, it is a bi-normal distribution value. This distribution is often used to describe a set of correlated real-value random variables each of which clusters around a mean value.

V. Results and Discussion

A. Multivariate Linear Regression Analysis

For the first part of the analysis, this paper has conducted a multivariate linear regression analysis.

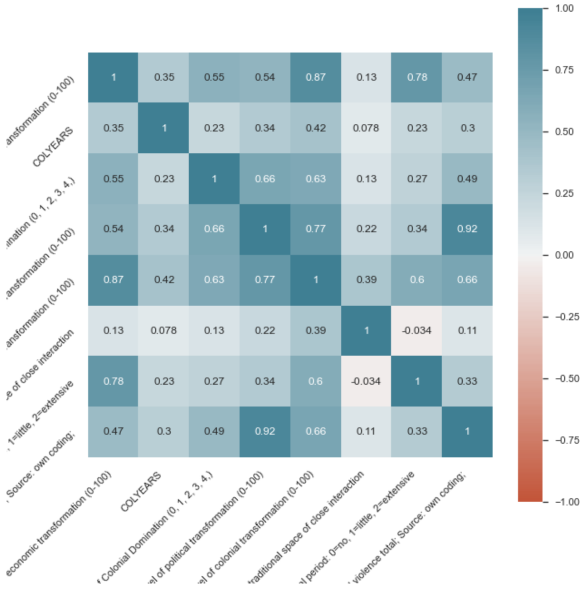

Fig. 4. Correlation Metric to Identify Variables for OLS Regression Model

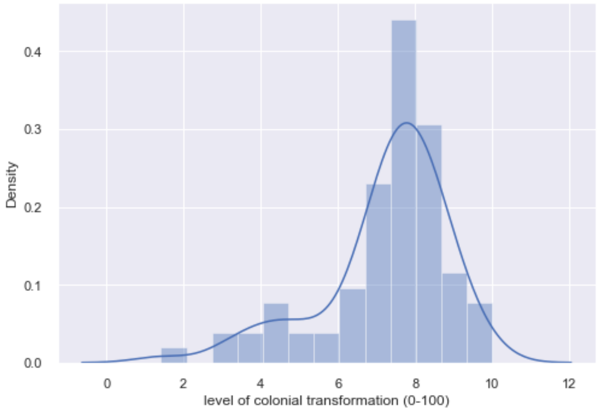

Fig. 5. Bi-normal Distribution of Colonial Transformation Scores

This is a technique that uses a single regression model with more than one independent variable. In this case there are multiple predictor variables (as identified from the correlation matrix), validating the use of the multivariate multiple regression. With this model, we can use multiple independent variables and therefore have multiple coefficients to determine the response variable. Equation 2 below shows the equation of a multivariate linear regression that is being used for the computation of results.

$$Y_i = \alpha + \beta_1 x_1 + \beta_2 x_2 + ...\beta_n x_n \tag{2}$$

In equation 2 Y_i is the estimate of the i^{th} component of dependent variable y, where we have n independent variables and x_n shows the n^{th} component of the independent variables. With this from of regression, we have selected the best possible independent variables that contribute to the dependent variable. With the help of the correlation matrix, we get a value which gives us an idea about which variable is significant and by what factor. From this, we have picked the independent variables in decreasing order of correlation value and then run the regression mode. Variables are as follows :

Independent Variable	Correlation Coefficient
Economic Transformation	0.87
Political Transformation	0.77
Colonial Violence	0.66
Colonial Domination	0.63
COLYEARS	0.47
Colonial Border Split	0.39

Table 2:Correlation Coefficient of Significant Independent Variables

By running the model in this order, we can minimize the error function of the model.From here, we use the Ordinary Least Square (OLS) regression, which is a statistical method of analysis that estimates the relationship between one or more independent variable and a dependent variable by minimizing the sum of the squares on the difference between the observed and predicted values of the dependent variable. With this OLS multivariate model we can have multiple independent variables as explained above.

Statistical Test	Result
Adj. R-squared	0.934
F-statistic	186
Prob (F-statistic)	9.46e-42
Prob(Omnibus)	0.034
Durbin-Watson	1.70
Jarque-Bera (JB)	6.19
Prob(JB)	0.045

Table 3: OLS Regression Results

When analysing this OLS Regression model, we can interpret the statistical results given. With an adjusted R-squared value of 0.934, we can see the robustness of this model is relatively high. This value indicates the goodness-of-fit measure for a linear regression model by indicating the percentage of the variance in the dependent variable that the independent variable explains collectively. With a high percentage of 93.4, this indicates the model explains the variability of the response data around its mean well. Next, the Prob(f-statistic) tells us the overall significance of the regression to assess the significant level of all the variables together. With interpreting this statistical value, the null hypothesis is that "all the regression coefficients are equal to zero". With this analysis, we get Prob(f-statistic) being $9.46 * 10^{-42}$, which is very close to zero, implying that the overall regression is meaningful. Additionally, looking at the Durbin-Watson value, this gives us the implication of whether the variance of errors is constant or not. With OLS having an assumption of homoscedasticity, constant variance in errors is an assumption we have made. With a value of 1.701, this implies that the regression results are reliable from the interpretation side of this metric. Another assumption made with OLS regression is that errors are normally distributed, and we can use the omnibus test, with the null hypothesis that the errors are normally distributed. With the Prob(Jarque-Bera) test giving a value of 0.0451, a relatively small value, this indicated that the errors are normally distributed. As seen earlier from figure 5, since the distribution of the response variable values led to a bi-normal distribution.

Since there are multiple variables predicting the outcome of the response variable, it is important for us to consider the issue of multicollinearity. When we have a multiple regression and want to test the effect of multiple factors on a particular outcome (in this case, the predictor being the *colonial transformation score*), we need to ensure these independent variables measuring are not correlated amongst themselves or measure almost the same thing. Because then the underlying effect is that what they measure gets accounted for twice across the variables and ultimately, it becomes difficult to say each exact variable is really influencing the independent and dependent variables. With multicollinearity in a multiple regression, it indicates that these variables are not actually independent and it produces estimates of the regression coefficient that are not statistically significant. This is because with high multicollinearity, it makes it difficult to estimate how much the combination of the independent variables affects the dependent variable within the regression model.

With that being said, to ensure that our model is functioning correctly, we can test for multicollinearity using the Variance Inflation Factor (VIF). This tool will help us identify the degree of multicollinearity within the model so that we can adjust the model. This is done by measuring how much the behaviour of an independent variable is influenced by its interaction with the other independent variables. Since the variances of the estimated coefficients are inflated when multicollinearity exists, the VIF for the estimated coefficient, b_k is simply the factor by which the variance is inflated. To further explain and validate the use of VIF, let's take a model where x_k is the only predictor as follows: $y_i = \beta_0 + \beta_k x_{ik} + \epsilon_i$ From this, the variance of the estimated coefficient b_k is from the following formula:

$$Var(b_k)_{min} = \frac{\sigma^2}{\Sigma(x_{ik} - \bar{x_k})^2} \tag{3}$$

This denotes that this is the smallest the variance can be. Now, we can being in a model with multiple correlated predictor:
$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + ...\beta_{p-1} x_{i,p-1} + \epsilon_i$

From the following equation of the model, if the predictors are correlated with the predictor x_k, then the variance b_k is inflated. This leads us to the following equation:

$$Var(b_k) = \frac{\sigma^2}{\Sigma(x_{ik} - \bar{x_k})^2} * \frac{1}{1 - R_k^2} \tag{4}$$

In this case, the R_k^2 is the R^2 value calculated by regressing the k^{th} predictor on the remaining predictors. From here we know that if we see a greater linear dependence on the predictor x_k and the other predictors, then we observe a larger value of R_k^2. This in turn gives us a larger variance of b_k. In order to calculate how much larger the variance is, we take the ratio of the two variances as shown below:

$$\frac{Var(b_k)}{Var(b_k)_{min}} = \frac{\frac{\sigma^2}{\Sigma(x_{ik}-\bar{x_k})^2} * \frac{1}{1-R_k^2}}{\frac{\sigma^2}{\Sigma(x_{ik}-\bar{x_k})^2}} = \frac{1}{1-R_k^2} \tag{5}$$

Therefore, we see that the Variance Inflation Factor for the k^{th} predictor is given by the following equation:

$$VIF_k = \frac{1}{1-R_k^2} \tag{6}$$

With this mathematical background, we can calculate the VIF for each of the independent variables from the OLS model above. Since this is a measure of how much the variance of the estimated coefficient is inflated by the multicollinearity, we are taking the rule of thumb that a VIF higher than 10 indicates multicollinearity within the variables. Table 4 shows the VIF values for the variables from the OLS regression.

Feature	VIF Factor
Economic transformation	1.8
Political Transformation	12.4
Colonial Border Split	1.1
Colonial Violence	9.5
Colonial Domination	2.5
Colonial Duration	1.2

Table 4:VIF factor for predictor variables from OLS Regression

From what we can see, we have the issue of multicollinearity specifically with the variable to do with 'level of political transformation' and 'Colonial Violence' since these are values nearing 10 and indicate high levels of correlation. In order to address this issue we will use the Ridge model to evaluate the residual plot. This is a technique for analysing multiple regression data that suffer from multicollinearity. The Ridge Regression performs a L2 regularization, meaning that it adds a penalty equivalent to square the magnitude of the coefficients. With this method, we keep all the predictors in the model and we minimize the sum of square of coefficients to reduce the impact of correlated predictors. We will use a trend line that over-fits the training data and therefore has a much higher variance than the OLS. Essentially, with this Ridge Regression, we will introduce a certain amount on bias into the new trend line by finding the coefficients that minimize the sum of error squares by applying a penalty to these coefficients. This will reduce the standard error and make the estimates from the model more reliable.

In practice, we will introduce this bias as λ , known as the penalty term. It is important to note that λ is represented as an alpha parameter in the Ridge Regression function, and by changing alpha, we are changing the penalty term. Therefore if λ is zero, we get the classical regression equation. And if we have a higher alpha value, then we have a created penalty, and the size of the coefficients is reduced. By reducing these parameters in the model, we are able to prevent the multicollinearity in this model unlike what was in the OLS Regression. The process of this correction is shown in the equation below:

$$SSE_{L2} = \Sigma_{i=1}^{n}(y_i - \bar{y_i})^2 + \lambda\Sigma_{j=1}^{P}\beta_j^2 \tag{7}$$

From equation 8, we see the first part is the classical regression calculation. This is added to the regression calculation including the β values that are added up. With the λ parameter, we standardize the value and have the correction described earlier. With this equation, we have created a model using Ridge Regression with the following steps.

1) A setting value of alpha is initially determined by the user. For this model, we took the initial alpha to be 5.
2) The beta coefficients are calculated in the data set
3) We created a list of random alpha values to find what would be the optimal value from this list.
4) To build the Ridge model, we created a set of empty coefficients and fit each alpha value from the list generated earlier. For each alpha value, we added the calculated coefficient value that came from it to a set of all the coefficients. To create this model, we labelled our X and Y variables (as the same from the OLS Regression)
5) With this model, we calculated the Root Mean Square Value both before and after the cross-validation technique was used.
6) Based on these values, we determined our optimal alpha, and tuned the model accordingly to calculate the error and accuracy that

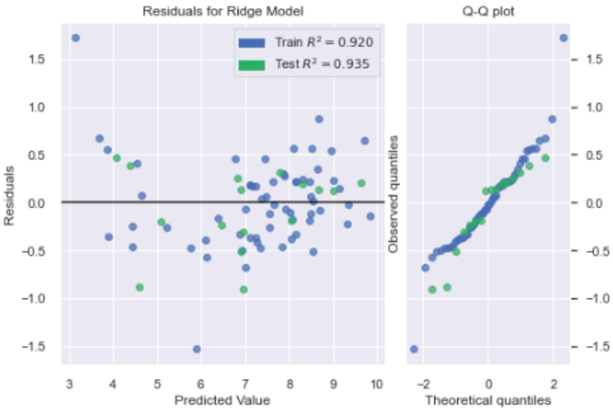

Fig. 6. Plot of Ridge Regression Residuals

this model has.

While this code is outlined in greater detail with the code book attached to this project, we see that the Root Mean Square Error (RMSE) value before the Cross Validation is 4.58 and after is 5.26. This statistical test is simply the standard deviation of the residuals and how spread out they are. It tells us how concentrated the data is around the line of best fit. From here we see that with the use of Cross Validation, the RMSE decreases, giving us better verification of the model we are creating to use. This k-fold Cross Validation being used in this model is simply a resampling procedure used to evaluate machine learning models on a limited data sample. By using this cross validation technique to resample the dataset, we were able to find the optimal value of alpha to be 3. By using $\lambda = 3$, and tuning this parameter into our model, we get a $R^2 = 0.95$. With this alpha value, we have created a model with a high accuracy value, validating the use of this model.

To further validate the use of this model, we can look at the distribution of residuals of this model as shown in figure 6.

Figure 6 shows the diagnostic plot with the QQ plot alongside the distribution of residuals. A residual is simply the difference between an observed value of the response variable and the value of the response variable predicted from the regression line. With this diagnostic, we can see the graph of residuals versus the expected order statistics of the standard normal distribution. As we can see, the plots lie on the line of the QQ-plot, indicating a match to the standard normal distribution.

With this model having a high R^2 and having accoutting for the multicolliarity from the OLS Regression, the table below shows the new coefficients of the features once multicollionairy using the Ridge Regression has been accounted for.

Independent Variable	Coefficient (OLS)	Coefficient (Ridge)
Economic Transformation	0.514	0.513
Political Transformation	0.503	0.487
Colonial Violence	-2.73	-2.47
Colonial Domination	-0.910	-0.803
COLYEARS	0.011	0.012
Colonial Border Split	5.09	4.85

Table 5:Coefficients of Independent Variables with OLS and Ridge Regression Models

From this table, we can see that the new coefficients for the variables have been scaled down in order to account for the multicollinearity. With the Ridge Regression used and the validation of this model explained earlier, the coefficients from this model are the best we have to use for regression analysis.

B. *KMeans Clustering Analysis*

The next part of this analysis uses the K-Means Clustering Analysis. This form of cluster analysis is a set of data reduction techniques designed to group similar observations in a dataset. Unlike other data reduction methods like Factor Analysis or Principal Component Analysis, cluster analysis groups observations by similarities across rows. This is done by minimizing the Euclidean distances between the clusters. This distance is computed by taking the difference between two observations on two variables to solve for the shortest distance between two points, which can be extended into n-dimensions. This requires variables to be numerical and continuous for this analysis.

For this analysis, it explores in-depth the relationship between the duration of colonialism (Co-lYears being the x-variable) with the level of colonial transformation of the country (Colonial Transformation Score being the y-variable). This analysis therefore explores how different duration of colonialism may have led to different extents

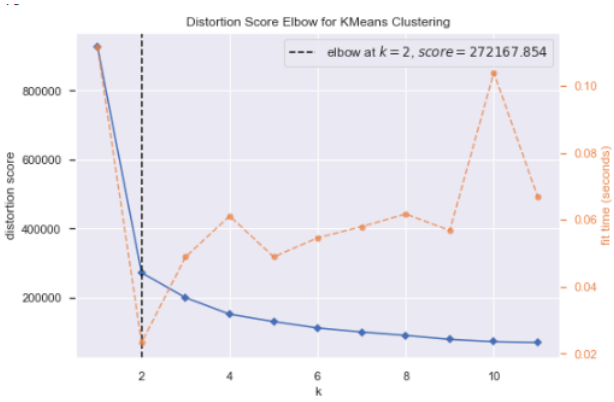

Fig. 7. Elbow Method: Identifying number of clusters, k

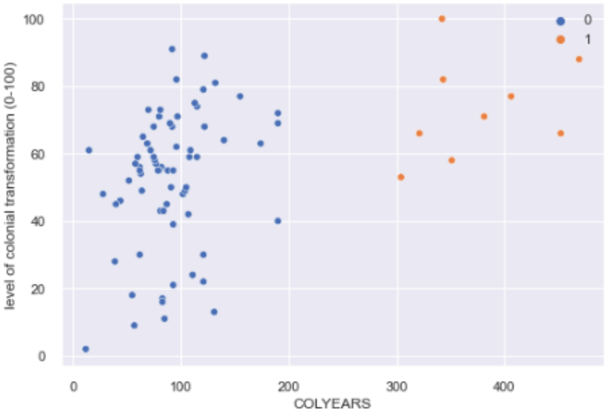

Fig. 8. KMeans Clustering Analysis: Two distinct eras of Colonialism identified. Cluster 0: Mercantilist era, Cluster 1: Imperialist era

of transformation. The reason we have chosen duration specifically is because as mentioned in the domain review, literature suggests that the duration of colonialism is one of the influential factors in how much a country is impacted by colonialism. With this information, this cluster analysis explores this previously standing research to a greater extent. The general premise of how this analysis works is under these steps:

- Specify the number of clusters, k. To find the number of clusters to assign in this analysis, we have utilized the Elbow Method. This is an empirical method used to find the best value for k that would boost model performance. It calculates the sum of the square of the points and calculates the average distance between. As the value of k increases, the within-cluster sum of square values decreases. With this method, we can plot a graph between k-values and within cluster sum of the square to get the k value. As seen in figure 7, at k=2, the graph decreases abruptly – like an elbow– indicating that the best k-value for this data would be at two clusters.
- The next step is to allocate objects to the clusters. In our case, the objects have been randomly assigned to the two clusters as with different colors.
- From here, we computed the cluster means. For each cluster, the average value is computed for each of the variables. In figure 8 , the average value of the dots is represented by the x-variable, ColYears, and the variable in the vertical dimension is the response variable, the level of colonial transformation.
- Steps 2 and 3 are repeated until the solution converges – meaning that reallocating observations and updates means cannot improve the solution.

With this method, we get figure 8.

Essentially, this algorithm is used to find groups which have not been explicitly labeled in the data and understand the dataset itself better. As seen from the graph, we have two distinct groups of clusters, one in the range of 0-200 years of colonialism and the other in the range of 300-400 years of colonialism. The first cluster has a greater number of datapoints compared to the second. We also see a greater range of colonial transformation scores in the first cluster. The second cluster however, most of the data points are clustered towards the top, indicating a higher degree of colonial transformation. This visual implies that with a greater duration of colonialism, there is a greater degree of transformation – validating the literature. With a greater number of data points we may have seen a greater difference in the clustering. While this does give us a general overview of the relationship between these variables, the clusters are not close together or distinctly disjoint from each other, indicating that while duration does make a difference, it's difficult to say the extent of it.

However it is interesting to see that the clusters formed due to colonial duration split leads to two distinct clusters – one of 0-200 year rule, and the other from 300-400 year rule. According to our literature review above, we have seen historians often split the history of colonial rule into two separate eras. And our splitting here suggests that the heterogeneous era of colonialisation should in fact be separated into an early 'mecantilist' wave (primarily led by the Spanish) and a much later 'imperialist' wave (led by the British), each as we know from our literature review to be different. The greatest difference besides the origin of the colonisers is the time frame of the eras – Mercantilist era had an active colonization 1715-1820, however the later imperialist waves reaching its peak in 1880-1900. We see this distinction of colonial duration (the first wave lasting for around 200 years from the 18^{th} and 19^{th} century), but imperial colonial rule lasted around 300-400 years (from 1450-1950). These different duration periods are shown by our clustering analysis above. This quantifies the historical theories discussed in literature using data and visuals which further validates these proposed theories.

C. ANOVA Analysis

The final part of the analysis in this paper utilizes an Analysis of Variance Test, known as ANOVA. This is essentially a t-test between two or more groups and is used to compare the means of a condition between the groups. In this paper, the first groups is a set of data points which are countries of former British colonies, and the second group is a set of data points with countries of former French colonies. This analysis is to explore how different colonizers may have impacted the overall colonial transformation for a country – does who colonized the country make a difference to the colonial change endured by the country?

To run an ANOVA test, we will start with defining our null and alternative hypotheses:

- Null Hypothesis: Both group means of former French and British colonies are equal
- The group means of former French and British colonies is not equal.

ANOVA looks into not just variation between the sample means, but also the variation within the mean itself. Equation 2 shows the formula used in one-way ANOVA tests.

$$F = \frac{\frac{SS_B}{(k-1)}}{\frac{SS_W}{N-k}} \tag{8}$$

Where, $SS_B = \Sigma n_i(\bar{y_i} - \bar{y})^2$ and $SS_W = \Sigma(\bar{y_{ij}} - \bar{y_i})^2$ with the following parameters: y_i= sample mean in the i group, n_i = total number of observation in group, $\bar{y}$ = total mean of the data, k= total number of the groups, y_{ij} = j observation in the out of k groups, N= Overall sample size.

With this formula, we can compute the ANOVA table to give us information about the different groups and determine the variability between samples and within samples. This is shown in Table 2.

Variable	sum sq	df	F	PR(>F)
C(Coloniser)	393.024	3.0	0.5419	0.587

Table 6: Overall model F(2, 31) = 0.542, p = 0.5870

For this analysis, we have done a one-way ANOVA test with just having one independent variable which is the colonial transformation score. This analysis tells us how different colonies have been impacted by colonialism by comparing the French and British colonists. Before interpreting the results, here are the following assumptions made with ANOVA:

- The observations are obtained independently and randomly from the population defined by the factor level
- The data for each factor level is normally distributed
- Independence of cases: the sample cases should be independent of each other
- Homogeneity of variance: Homogeneity means that the variance among the groups should be approximately equal

To ensure the normality of the distribution of scores and validate the use of the ANOVA test, we can look at the distribution of residuals using a QQ plot and conduct a Shapiro-Wilk test. To begin validating these assumptions, we can look at figure 9 to see the distribution of residuals. As seen in this plot, the points are coming from the set of former French colonies and the other

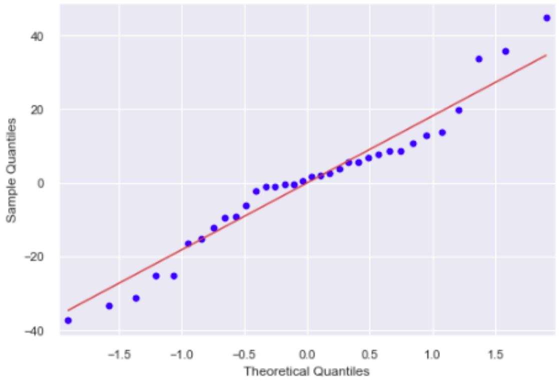

Fig. 9. Distribution of Residuals from ANOVA Model

sample is coming from a set of former British colonies. As seen, the values lie on top of each other, indicating there is normal variance and distribution between and within the samples. With the Shapiro Wilk test for normality, we can further validate the use of ANOVA.

This is a goodness-of-fit test that examines how close the sample data fits to a normal distribution. And this is done by ordering and standardizing the sample by converting the data to a distribution with the mean equalling zero and the standard deviation being one. This is essentially a measure of how well the ordered and standardized sample qualities fit the standard normal quantifies, with a scale of 0 to 1 and 1 being a perfect match. In order to conduct this test, we have the following hypothesis:

Null hypothesis: Sample used comes from a normal distribution Alternate hypothesis: Sample used does not come from the normal distribution

We can compute the critical values. We get W= 0.960and a p-value of 0.254. Since W > then the critical value, we do not reject the null hypothesis. This implies that the sample used does come from a normal distribution and validates the use of ANOVA.

From the results in Table 3, we can see the results of our ANOVA test. As seen, we get a p-value of 0.586 which is greater than 0.05. This means that we fail to reject the null hypothesis, which implies that there is no difference in the mean of colonial transformation between countries of former French and British colonization. This implies that those who colonized the country are not influential when looking at the transformation brought upon the colony.

VI. Conclusion

As seen from the analysis, understanding colonialism and conducting historical research is a complex process. With the data itself being historical, there is a lot to fill in to contextualize what the data really means. As with historical research, the data points are not just objective numbers, but each have an implication and significance to a country's history. With research within colonialism using analytic being limited, there was not much literature to guide the analytic performed in this paper. Instead, most literature to do with colonialism is theoretical and heavy with historical, political and social theories and schools of thoughts. With that being a caveat of doing research in an unexplored field, the analysis this paper brings relies on implications, contextualisation of a country's history and an understanding of colonial theories. Once again, because this dataset has been a recreation of primary sources by historians in Harvard, it is important to note their inherent bias brought into the designing of this dataset – therefore, despite having analysis and discussions performed earlier, there are limitations to this research as with the nature of historical research. With that being established, this paper has performed computational analysis of how former colonies have been transformed by their colonial rule in the following ways:

- Starting with multivariate linear regression analysis and then having developed that to a Ridge Regression model, we have created a robust model in predicting how different independent variables can lead to the colonial transformation of a country. From the analysis above, here are the following variables that are significant in determining the factors of colonial transformation (in decreasing importance) : Economic transformation, Political transformation, colonial violence, colonial domination, duration of colonization, colonial border split. This means that these variables play a significant role in determining the extent to which a country is impacted by the colonizer. This saying is more likely to be transferred if the colonists had greatly changed their economic and political structures. With economic and political structures being the core to a functioning country and democracy, the way in which colonists impact these systems leave long lasting effects. For instance, in India, when the British left and they determined Kashmir to be undisputed land, their decisions have led to the longest war over land till present. It's instances like this where playing a country's power structures greatly can influence their post-colonial transformation. This model also tells us when the colonizers used violence and domination – through military weapons, threats, slavery, rape, etc – they have once again impacted the transformation of a post-colonial nation. With violence and domination, the colonizer can easily abuse their powers and invoke fear for their purposes. The reality of colonialism was brutal, with the violence and domination leading to the loss of many lives. With violence, it's difficult for these nations to have a peaceful independence. Therefore the more violence and dominating the rule was, the more a colony suffered negatively – inhibiting travel ,education, health and development of their nation. Additionally, the longer the rule was, the more violence and domination could take place by the colonizers, making duration a significant influence in the model. Lastly, colonial border split was seen as an important factor. This is because when colonists left the land, they had the power to decide who gets, and how they separate it. Oftentimes the land was separated by religion, race or other identifying factors, leading to underlying tensions with the former colony. By dividing the border using identity, they were able to completely transform the social, cultural and political norms of a nation leading to a colony's overall transformation. There are many factors influencing the extent of transformation endured by a colony. By assessing these factors we can start to look at these countries today, and be able to trace their current standings to their colonial past. This gives us the ability not just to educate ourselves on a nation's history, but it gives us an understanding of the origins, intentions and use of social and political structures in place today. And with this understanding, evaluate whether they were created with the right intentions, or if it's time for a present day upgrade.

- Next, with the clustering analysis, we were able to explore the relationship between the duration of colonization and the overall transformation of a former colony. As explained earlier under the results, while duration does lead to a greater time of exploitation and violent domination negatively impacting the colony, it's not the most significant in the transformation. This is possibly because while time elongated the exploitation, it was the people in power and their intentions specifically that were more harmful to the nation.
- Finally, with the ANOVA test we were exploring whether there is a difference in the transformation endured by the colony based on who was dominating over them – the French versus the British. As explained in the results section, there seemed to be no statistical difference between these groups, implying that the colonial transformation was not based on who exploited who. This is interesting because while the style of dictatorship and colonialism does impact the social norms of a colony, the domination and violence is an issue of human life regardless of who is performing the act. This shows us that when understanding the impact of colonialism, who these negative impacts are coming from is not important because regardless of who performs it, the domination, violence, and transformation is endured. This suggests that while implementing research, it's not about who performed the colonialism, but about how it was executed and the last impacts from it.

This research is just a start – the data is still being collected and developed by the Harvard dataverse, but these initial findings give us insights on what variables we should be exploring further for a deeper historical understanding within colonialism.

VII. Further Research

While this paper is still being developed and there is more historical data being made available over time, this paper brings insights that can be used by policymakers especially when designing education curriculum. Many schools till present use former colonizer education curricula, where history is often taught by the winners and relays an inaccurate depiction of the past. With this research, we can use quantitative data to present the reality of what colonial transformation has been for former colonies and build robust arguments to not romanticize colonialism. Instead, this gives an opportunity to policymakers to understand the depth at which colonialism has impacted the development of countries and potentially build arguments to have former colonizer countries give reparations to former colonial countries. Since many colonizers have shaped the economic, social and political development of present day countries, the cost the former colonies endured could be accounted for and reparations could help with the development of these nations. With more data, analysis and evaluation, this paper can be used as a starting point into this realm of historical research.

VIII. Bibliography

Acemoglu, Daron, and James A. Robinson. "The economic impact of colonialism." The Long Economic and Political Shadow of History Volume I. A Global View 81 (2017).

This journal explains the concept that the economic inequality that prevails in the world today is the result of acts over a century ago. This result stems from European colonialism. If we go back 500 years, there are little inequality and small differences between the rich and poor countries (by a factor of 4). Now the differences are more than a factor of 40 when comparing the richest to the poorest country. This journal traces back into history and explains the role of European colonialism on the income inequality present today.

[2] Maseland, Robbert. "Is colonialism history? The declining impact of colonial legacies on African institutional and economic development." Journal of Institutional Economics 14.2 (2018): 259-287. This paper explores the claim that colonial history has left an enduring impact on Africa's institutional and economic development. This paper shows that the relevance of colonial legacies to institutional quality and to per capita income is rapidly disappearing in Africa . It concludes that while colonialism affected African institutional and economic development, the impact is not persistent.

[3] Alam, Manzurul, Stewart Lawrence, and Ruvendra Nandan. "Accounting for economic development in the context of post-colonialism: the Fijian experience." Critical Perspectives on Accounting 15.1 (2004): 135-157. This paper is a study of change in the operations and accounting practices within the Fijian Development Bank. This is a detailed outline of how the Bank is faced with the task of reconciling a history of colonisation and racial discrimination with the forces of globalisation and the want for economic development.

[4] Bruhn, Miriam, and Francisco A. Gallego. "Good, bad, and ugly colonial activities: do they matter for economic development?." Review of economics and statistics 94.2 (2012): 433-461.

This journal looks into the development and its variation within countries in the Americas. It argues that part of the variation can be traced back to the colonial era – when colonizers engaged in different economic activities in different regions of a country. This paper shows us that ‘bad economic activities’ led to lower economic development today than ‘good activities’.

[5] Ashcroft, Bill. Post-colonial transformation. Routledge, 2013. This book explores how post-colonial societies have responded to colonial control. This book details what kinds of transformations have taken place, investigates their strategy with political and literary transformation and proposes a new post-colonial theory of development.

[6] Ziltener, Patrick, Daniel Künzler, and André Walter. "Research note: Measuring the impacts of colonialism: A new data set for the countries of Africa and Asia." Journal of World-Systems Research 23.1 (2017): 156-190. This is a dataset comprising 15 indicators for the political, economic and social impact of colonialism. This gives us the opportunity to directly compare the levels of colonial transformation by looking at indicators for economic distortion and redirection. It evaluates the implication of economic transformation by further measuring social transformation and political liberation. This is the dataset used in this research paper.

[7] Walter, André, et al. “Measuring the Impacts of Colonialism: A New Data Set for the Countries of Africa and Asia.” Harvard Dataverse, Harvard Dataverse, 13 Nov. 2019, https://dataverse.harvard.edu/dataset.xhtml This is the dataset used in this research paper.

[8] “12.4 - Detecting Multicollinearity Using Variance Inflation Factors | STAT 501.” PennState: Statistics Online Courses, online.stat.psu.edu/stat501/lesson/12/12.4. Accessed 2 May 2022. This was used to understand the creation and use of the VIF implemented following the OLS regression and aided with the development of the Ridge Regression

[9] McDonald, Gary C. "Ridge regression." Wiley Interdisciplinary Reviews: Computational Statistics 1.1 (2009): 93-100. Used for the development and implementation of Ridge Analysis

[10] Kim, Hae-Young. "Analysis of variance (ANOVA) comparing means of more than two groups." Restorative dentistry endodontics 39.1 (2014): 74-77. Development and implementation of the ANOVA test used in this paper.

[11] Olsson, Ola. "On the democratic legacy of colonialism." Journal of Comparative Economics 37.4 (2009): 534-551. This looks at the history of colonialism, this paper explains the different waves of colonial eras – the mercantilist and the imperialist waves.

[12] Olsson, Ola. "On the institutional legacy of Mercantilist and Imperialist Colonialism." rapport nr.: Working Papers in Economics 247 (2007). Further explanation of the different waves of colonialism by the Spanish and British colonisers.

[13] Greenland, Sander, et al. "Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations." European journal of epidemiology 31.4 (2016): 337-350. Reference book used to understand the apply the different statistical tests used throughout this paper.

[14] Meijer, Rosa J., and Jelle J. Goeman. "Efficient approximate k-fold and leave-one-out cross-validation for ridge regression." Biometrical Journal 55.2 (2013): 141-155. Reference used to understand and implement the k-Fold Cross Valiadation technique to use for the development of the Ridge Model.

[15] Shrestha, Noora. "Detecting multicollinearity in regression analysis." American Journal of Applied Mathematics and Statistics 8.2 (2020): 39-42. Reference used to understand the use of VIF and detect multicollinearity.

[16] Kaul, Suvir. "Indian Empire (and the case of Kashmir)." Economic and political weekly (2011): 66-75. Used this source the contextulise the history of India and the way the British have left their impact on it.

[17] Wolpert, Stanley A., and Stanley Wolpert. Shameful flight: The last years of the British Empire in India. Oxford University Press, 2009. Evaluating the tactics of the British in terms of colonial border split, and how that has impacted the violence of independence for various colonies.

IX. Appendix

A. Ethical Statement

This project aims to explore how colonized countries have been performing since their period of colonial rule to present, and questioning the role of colonialism through it. While considering the ethics of this project, since the dataset I will be using comes from the Harvard dataverse.Moving on, as I work on this, I don't see there to be any conflicts of interest in my work with this project – however, I do acknowledge that I am an Indian citizen, making me a person who is a product of the country that has been colonized. While that is the case, the analysis I am conducting is looking at objective analysis using data analytics techniques, and the background research I will bring in for historical context comes from academic journals and historians. Therefore, my history and background remain separate from both the quantitative and qualitative research being conducted in this capstone project. This project is ambitious with what it's trying to achieve. I don't foresee many potential harms coming from this project but I see the benefits having broader implications within the education system that has been placed in British curricula. This research hopes to explore and research colonialism through an objective lens rather than just as a collection of stories passed from generations. This way, more accurate accounts of history can be taught in education settings and students can have a better understanding of this past. However, the only harm here is that while this dataset is credited through Harvard, this dataset is just a collection of different secondary sources that have been quantified by historians from Harvard. Therefore, no matter what analysis is done and shown by the project, the dataset comes from historians at Harvard who have chosen the sources, which means there is an inherent bias with the creation of the dataset itself. However, this tends to be the case with most data coming from historical research, and therefore is not a large harm inhibiting the progression of this project. Other than that, the data, its code and the metacode will be shared with the class and professor on a regular basis with constant documentation to explain each step to ensure transparency. For model validation, I will be reading on the analysis and results from other academic journals also exploring colonialism and from there, use cross-validation techniques to evaluate my findings. Since this data is open source and already available to everyone, I feel comfortable sharing my findings with my classmates and the general public. The data does not need to be kept securely since it is not sensitive information. Moreover, since it does not have to do with human-centric research, there is no IRB necessary.

B. Details about the Political,Social and Economic Transformation

Overview On the basis of these 15 political, economic and social indicators for the effects of colonialism, we can construct indices for the political, economic and social impact as well as the combined "total" impact of colonialism. This enables us to compare the levels of colonial transformation of the countries of Africa/Asia in a quantitative manner.

The three indices correlate significantly:

The Political Transformation (PT)-Index with the ET-Index 0.50. The PT-Index with the Social Transformation (ST)-Index 0.44. The Economic Transformation (ET)-Index with the ST-Index 0.51. This means that, in general, political domination led to economic and social transformation. However, the correlations are not that strong, which indicates that the three dimensions should be measured separately.

Colonies in sub-Saharan Africa were more likely to experience a higher level of transformation, politically, economically and socially (correlations with PTI 0.37, ETI 0.38, STI 0.58), than the Asian/North African ones. The smallest difference to be found is regarding political transformation. That means that, although political domination was not much less intensive in Asia/North Africa, these economies and societies were more difficult to transform through colonialism.

There are no highly significant differences between British and French colonies regarding the level of colonial transformation. However, British colonies seem to have been transformed a little less than others, politically and economically.

C. Details on : Duration of Colonialism

The Length of Colonial Domination (COLYEARS) It is common to declare the year of the formal declaration of a colony or protectorate as starting point of colonialism. We think this legalistic approach is not adequate to the problem. If political domination is crucial to colonialism, that its onset should be the point in time when political sovereignty was de facto significantly reduced by a foreign power over a significant part of the territory and/or population. This is more often than not before any de jure declaration, and by contrast in certain cases even significantly after this point. As Lange et al. (2006: 1418) point out: “India was clearly under the grip of the English East India Company by the 1750s, but it was not proclaimed a colony under control from London until 1857”. Because Muscat/Oman has never been a de jure colony, Price (2003: 481f) and others consider the country “without colonial heritage”, although there was a Portuguese occupation from the beginning of 16th to mid-17th-century and a de facto British control from mid-19th century on. As colonialism can be a gradual and informal process , its onset might be a unequal treaty called “treaty of amity and trade” with a more or less subtle loss of sovereignty (including e.g. extraterritoriality of foreign citizens, loss of control over foreign policy), the foundation of a major settlement against the will of the local population and/or rulers or the gradual gain of control over government institutions. In Egypt, the UK and France initiated 1876 a stewardship of the public finances that should be considered as a joint form of colonization, even before the country was militarily occupied in 1882. The Ottoman Empire/Turkey is widely considered as historically non-colonized, although it had lost considerable sovereignty through the gradual extension of the “capitulations”-system , the Anglo-Ottoman commercial treaty of Balta Liman in 1838, and, from 1881 on, through the foreign run Public Debts Administration which controlled major portions of Ottoman revenue, there¬by constituting “an enormous incursion on Ottoman sovereignty” (Horowitz 2004). A similar strategy was followed by the British in the case of Persia. Persia and Turkey are discussed in Cain/Hopkins’ British Imperialism: Innovation and Expansion, 1688–1914 (1999) under the title „management without development“ (p. 419ff) – we would speak of semi-colonialism in the political sphere and of financial colonialism as the mechanism of taking over government functions in order to ensure (and to maximize) debt payment (“debt trap”).

For political domination, a certain degree of enduring control over significant parts of the autochthon population is important. Single military attacks with plundering and retreat without the erection of permanent fortresses are thus not coded as the beginning of colonialism. Likewise, a simple trading station or the colonization of an isolated area as e.g. an island offshore or in the delta of a river is not considered as onset. The occupation of James Island in nowadays Gambia by the Baltic Duchy of Courland and later by the British (who even declared in 1760 a British Province of the Senegambia) did not lead to any political domination of a significant inland population and is thus not coded as onset of colonialism. The arrival of the Portuguese in the now Indonesian archipelago in early 16th century cannot be considered as onset of colonialism because they did not manage to establish political control over the “spice islands”. In contrast, on the Malay Peninsula, the Portuguese conquest of the great emporium of Malacca in 1511 clearly signified the onset of colonialism, since it remained despite many wars uninterruptedly under European control well into the 20th century and had a lasting impact on trade flows. Similarly, in Indonesia, colonialism began with the foundation of Dutch-Batavia in 1619 which became the colonial center of trade and administration until independence of the country in 1949. In our sample, colonialism started in 11 countries already in the 16th century, nine countries followed in the next two centuries and most countries followed only in the 19th and some even in the 20th century. However, in most of the latter cases there were earlier contacts with European powers.

Similar to our variable ONSET, we define the end of colonialism (COLEND) as the point in time where the vast majority of the autochthon population regained full sovereignty over internal and foreign affairs, with or without the participation of foreign settlers. At that moment, it should be in principle possible for a country to conclude alliances with whatever foreign power it wants.

What sovereignty means is clearly expressed in the words of the Afghan ruler Amanullah in 1919:

"I have declared myself and my country entirely free, autonomous and independent both internally and externally. My country will hereafter be as independent a state as the other states and powers of the world are. No foreign power will be allowed to have a hair's breadth of right to interfere internally and externally with the affairs of Afghanistan, and if any ever does I am ready to cut its throat with this sword."

Of course, sovereignty does not automatically mean the end of all political and/or economic dependencies, e.g. in foreign trade and direct investment. It is not important whether foreign administrators are present or not, but whether this presence is decided by the colonial power or by a sovereign government. Foreign military bases, semi-autonomous oil fields or other foreign enclaves tolerated by a sovereign government are for our purposes not considered a constraint of sovereignty. Egypt, again, is a special case: With British troops controlling the most important shipping infrastructure, the Suez Canal, de facto independence came only with the final withdrawing of all troops and Egyptian takeover in 1956. Difficult to assess are the cases where the anticolonial struggle developed into a war of independence against a post-WWII superpower. Vietnam's colonial period ended 1956, although complete independence and the restoration of sovereignty came only in 1975. All countries in our sample acquired full sovereignty in the 20th century.

There is no significant correlation between the colonizing country (British vs. French) and the length of domination for the countries of our sample (see 'Descriptive Statistics'). Also, there is no significant difference between the length of colonialism in African (Sub-Sahara) and Asian/North African countries. But the length of colonial domination is related to some economic and social indicators of colonial transformation (level of colonial violence, of investment in infrastructure and of work immigration, the significance of plantations and the success of missionary activities. In short, a longer colonial period means more colonial violence, more investment in infrastructure and more plantations, more work immigration and more religious conversions.