



World Bank's 10-Year Program  
Poverty Reduction in Africa

DatAnalytics Consulting Corp.



ALFONSO VILLEGAS NARVÁEZ  
SPARK  
MBD, SECTION 2

## Introduction

The World Bank is an international organization that was founded in 1944 to re-build Europe after World War II. It is one from a variety of organizations that help to shape and balance the world economy. Today, its main function is to fight poverty by offering developmental assistance to middle-income and low-income third world countries.

According to the World Bank itself, the organization has two very specific goals by 2030:

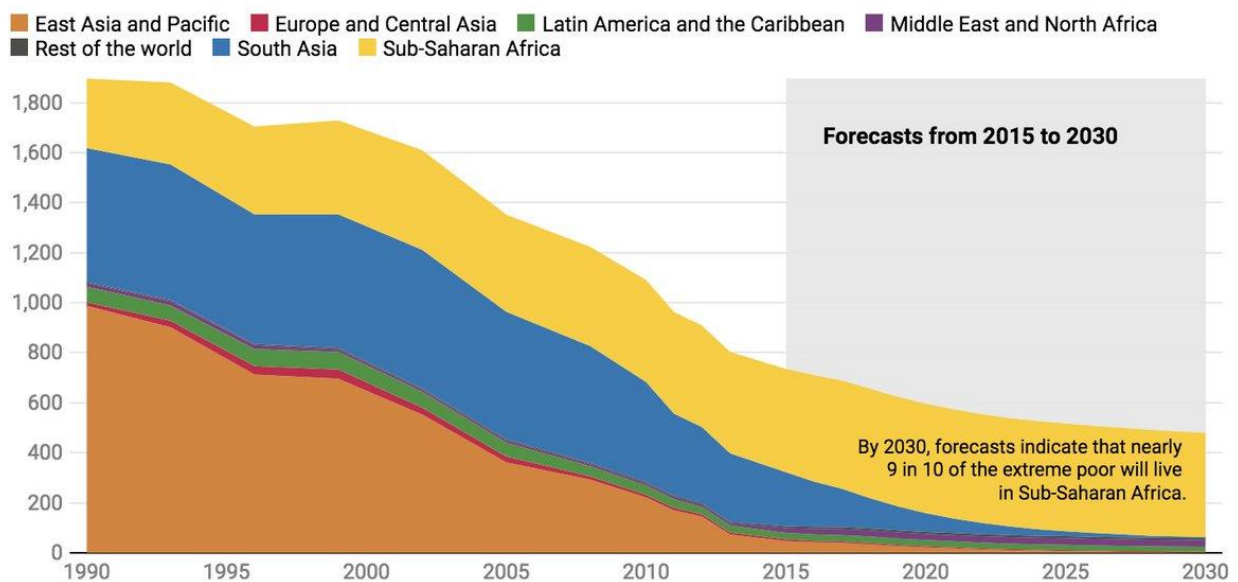
- *End extreme poverty by decreasing the percentage of people living on less than \$1.90 a day to no more than 3%.*
- *Promote shared prosperity by fostering the income growth of the bottom 40% for every country.*

<https://www.worldbank.org/en/about/what-we-do>

The specific problem that is going to be tackled with this 10-year program, has been established in the previous months as a result of the World Bank and IMF annual meetings in Washington D.C. on the last weeks of 2019:

Results have shown that there are more poor people in Africa now than 10 years ago, and most importantly, projections show that if the actual tendency does not change, by 2030 Africa may have 90% of the world's poor. While the poverty rate in Africa, which is defined as the percentage of people living on less than \$1.90 per day, fell between 1990 and 2015, rapid population growth resulted in the number of poor people on the continent increasing to more than 416 million, from 278 million over the same period.

People in extreme poverty (millions)



Source: World Bank PovcalNet and Poverty & Equity Data Portal

This is obviously a problem that needs to be tackled with urgency, not only because of the economic and social impact that it may have, but because it is within the main scope of the World Bank's primary objectives for the following years.

**DatAnalytics Consulting Corp.** has been entrusted to carry on an extensive analysis for the Financial Management area of the World Bank. The scope of this assay is to generate insights that will help both the Financial Management and Loan Committee areas to have a better understanding of the past and current situations regarding contracts that have been awarded by the organization by year, region and sector worldwide, but most specifically on Africa. This information will aid them as a guideline to develop a 10-year program that will have poverty reduction in Africa as its primary objective.

The main input for our analysis will be the World Bank's Major Contracts Dataset, which contains the banks major contract commitments from 2004 to 2017 and which will be explained thoroughly in the following section of this report.

As a second stage of the project, we will develop and train a model that will help them classify the reviewed contracts for each sector, based on the social and economic impact that they may have.

### World Bank's Major Contracts Dataset

This set of contract awards includes data on commitments against contracts that were reviewed and awarded by the Bank (prior-reviewed Bank-funded contracts) under IDA/IBRD investment projects. This dataset does not list all contracts awarded by the Bank, but should be viewed as a guide to determine the distribution of major contract commitments among the Bank's member countries. The data has been extracted and generated on September 6, 2017, and includes contracts from the fiscal years 2004 to 2018.

The original data set file can be downloaded [here](#).

The dataset is composed of the following variables (note that some of them are for banks internal use, provide no value for our analysis and therefore will be excluded furtherly):

1. As of Date: 6-Sep-17, date when the file was generated.
2. Fiscal Year: fiscal year when the actual loan has been implemented.
3. Region: geographical area from the country for which the contract was awarded.
4. Borrower Country: country for which the contract was awarded.
5. Borrower Country Code: country code for which the contract was awarded.
6. Project ID: project's internal registry number.
7. Project Name: investment project registered name.
8. Procurement Type: specific area for which the loan will be utilized.
9. Procurement Category: broader category in which the loan will be utilized.
10. Procurement Method: type of bidding and selection.
11. Product Line: internal are in which the contract has been granted.
12. Major Sector: economic sector to be supported by the contract implementation.

13. WB Contract Number: contract's internal registry number.
14. Contract Description: short description of each contract.
15. Contract Signing Date: date when the contract has been formally agreed.
16. Supplier: primary supplier working on the project.
17. Supplier Country: country of supplier registration.
18. Supplier Country Code: country code of supplier registration.
19. Total Contract Amount (USD): total worth of the contract in USD.
20. Borrower Contract Reference Number: borrower country contract number.

## Tool Selection

The dataset that we are using for the analysis contains 154,811 observations (rows) from the previously described 20 variables (columns). As agreed with World Bank's business users, on the following days we will get an updated version containing the contracts for 2019, which consists of somewhere close to 175,000 reviewed and awarded contracts.

We have chosen **Spark** as our Data Processing and Analytics engine for this task. The main reasons for it are enlisted as follows:

- **Speed:** By exploiting in memory computing along with a number of different optimizations, Spark provides very fast response for large scale data processing. Spark loads chunks of data to RAM memory, computes and does not move data or write to disk unless it is absolutely necessary or has been requested by the user.
- **Flexibility and Support:** as stated in the design meetings with business, choosing a flexible tool that provides easy access and fast queries for both business users and developers is essential, mainly because the World Bank has more than 10,000 employees in more than 120 offices worldwide. Programming languages supported by Spark include Java, Python, Scala and R. Additionally, there are libraries for SQL and Machine Learning that provide functionalities for fast queries using SQL syntax, model training and interactive analytics, which will be utilized in the second stage of this project.
- **Simplicity:** Spark is accessible via a set of APIs designed for interacting quickly and easily with data. These APIs are well documented and structured in such a way that it will be very easy for business users and developers to put it to work in any of the worldwide offices.
- **Security:** Spark provides fault tolerance and security in the following way: when a file is uploaded into the cluster, it is splitted in chunks, distributed amongst the data nodes and replicated across the cluster. Computation is also splitted and executed in parallel on computer clusters. Both data and computation can tolerate failures by recurring to another node for data or processing.
- **Capability:** due to the previously mentioned structure, Spark is able to handle huge amounts of data at a time with very fast processing and response time.

As we can see, Spark provides a lot of benefits due to its ability to combine very different techniques, languages and processes together into a single easy to use engine. We are sure that it has been the right

choice for this program, and that it will continue to provide added value functionalities in the next stages of the project.

## Analysis

Find below a brief summary of the content and analysis contained in the Markdown file that complements this report:

- **Data Understanding:** The first actions are focused on gaining a complete overview of the information contained in the Contract's dataset: How much data do we have and what variables? Relevant features for our analysis, data cleaning and preparation.

The World Bank Contracts dataset has 20 columns  
The World Bank Contracts dataset has 154811 rows

- **Initial Analysis:** Understanding the overall extent of contracts that have been awarded worldwide by the world bank: how much has been awarded per year? Which regions have benefited the most? Keeping in mind that Africa has been the top priority for poverty reduction, we would expect it to be the region with the highest number and highest amount of contracts awarded.

Region	SumPerRegion
AFRICA	42351174243
SOUTH ASIA	33324159096
EAST ASIA AND PACIFIC	30746079077
EUROPE AND CENTRAL ASIA	28299487362
LATIN AMERICA AND CARIBBEAN	23974043540
MIDDLE EAST AND NORTH AFRICA	11101262642
OTHER	8468552

- **African Region:** After getting an overall picture, we will now concentrate on the African region, which is the main focus of our analysis:
  - Data cleaning and selection, which features do we need to work with to do a faster and more user friendly analysis?
  - Overall understanding of the numbers: how much and how many contracts have been awarded to this region per year? Which are the years with the highest and lowest awards?
  - Which are the sectors that have been supported and to what extent?
  - What are the 10 African countries that have received the highest number and amount of contracts?

Year	Country	Category	Sector	TotalAmount
2009	Africa	CONSULTANT SERVICES	Public Administration	184365
2009	Burkina Faso	CONSULTANT SERVICES	Public Administration	258926
2009	Guinea	GOODS	Public Administration	468583
2009	Madagascar	CONSULTANT SERVICES	Health	54788
2009	Madagascar	CONSULTANT SERVICES	Social Protection	101750
2009	Sao Tome and Principe	GOODS	Public Administration	945657
2009	Tanzania	CONSULTANT SERVICES	Water, Sanitation and Waste Management	354833
2009	Zambia	CONSULTANT SERVICES	Agriculture, Fishing and Forestry	40000
2009	Africa	CONSULTANT SERVICES	Information and Communications Technologies	178201
2009	Burundi	CIVIL WORKS	Transportation	533286

only showing top 10 rows

As a last step, we have generated a new version of the file encoding numeric values to some of the string categories when necessary.

This file will serve as an input for the second stage of this project, which is: to train a model that will help our business users to classify the reviewed contracts for each sector, based on the social and economic impact that they may have for the future.

Year	Country	Category	Sector	TotalAmount	NumCategory
2009	Africa	CONSULTANT SERVICES	Public Administra...	184365	3
2009	Burkina Faso	CONSULTANT SERVICES	Public Administra...	258926	3
2009	Guinea	GOODS	Public Administra...	468583	1
2009	Madagascar	CONSULTANT SERVICES	Health	54788	3
2009	Madagascar	CONSULTANT SERVICES	Social Protection	101750	3
2009	Sao Tome and Prin...	GOODS	Public Administra...	945657	1
2009	Tanzania	CONSULTANT SERVICES	Water, Sanitation...	354833	3
2009	Zambia	CONSULTANT SERVICES	Agriculture, Fish...	40000	3
2009	Africa	CONSULTANT SERVICES	Information and C...	178201	3
2009	Burundi	CIVIL WORKS	Transportation	533286	2
2009	Burundi	CIVIL WORKS	Transportation	861451	2
2009	Burundi	CIVIL WORKS	Transportation	677151	2
2009	Ethiopia	GOODS	Financial Sector	180496	1
2009	Gabon	CONSULTANT SERVICES	(Historic)Health ...	75562	3
2009	Gabon	CONSULTANT SERVICES	(Historic)Health ...	76433	3
2009	Gambia, The	GOODS	Education	47454	1
2009	Lesotho	GOODS	(Historic)Health ...	742349	1
2009	Madagascar	CIVIL WORKS	Energy and Extrac...	433589	2
2009	Nigeria	CONSULTANT SERVICES	Financial Sector	1071612	3
2009	Cameroon	CONSULTANT SERVICES	Public Administra...	109062	3

only showing top 20 rows

## Conclusions

Although it is not within the scope of **DatAnalytics Consulting Corp.** to generate business conclusions from the generated data, we can see that Africa has been the main focus for the World Bank's contracts on the last years. This was the expected, since as stated before, Africa is the region with the highest poverty growth rate in the world and therefore, has been the main focus of the World Bank over the last years.

With the insights generated from the third section of these documents, the World Bank will be able to carry on specific analysis as to: which African countries have received the most contracts and why, is there any bias on the contract awarding process, why is the given support not having the expected results on poverty reduction, and lastly, look for signs or evidence of corruption that is affecting the development of each supported sector.

## Sources

<https://www.worldbank.org/en/about/what-we-do>

<https://www.kaggle.com/theworldbank/world-banks-major-contracts>

<https://todaynewsafrika.com/podcast-more-poor-people-in-africa-now-than-10-years-ago-world-bank-africa-top-economists-give-damning-verdict-of-last-decade>

<https://www.bloomberg.com/news/articles/2019-10-09/africa-may-have-90-of-the-world-s-poor-by-2030-world-bank-says>

<https://databricks.com/spark/about>