Prediction of Vietname future economic development

TRUNG QUOC NGO

June 30, 2020

1. Introduction

1.1. Background

Vietnam is among one of the fastest-growing economies in the world, with the forecasted annual GDP growth rate of 5%. In the recent year, Vietnam has managed to improve its domestic market to attract foreign invesment. In practice, modernizing the infrastructure, extensive market reforms while continued reducing public debt and tightening credit policies to allow friendlier and more competitive business environment. For month prior to the COVID-19 pandemic, companies have already begun to look for alternative logistics hubs, towards cost-effective and flexibility. As the Coronavirus wreak economic turmoil around the world, we expect to see a leap change in the global supply chain infrastructure. The interesting question is whether or not Vietnam has the capability to serve as a promising destination on regional scale amid this rapid and challenging transformation.

1.2. Objective

The objective of this project is to predict the future economic growth of Vietnam using the historial GDP rate, public debt, population size, population age, labour force, foreign trade, investment from 2009 to 2019 as the basis of model. Foursquare data is used to futher refine study and showcase attractiveness to investment of different regions in Vietnam.

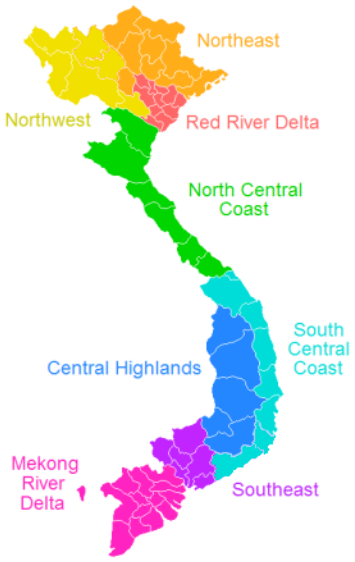
1.3. Interest

This project will be of interest for companies looking to invest in Vietnam, those who plan to be expatriates in South East Asia and local Vietnamese who would like to obtain a thorough look at the country development trend and potential in the near future.

2. Data acquisition and cleaning

2.1. Data sources

According to "List of regions of Vietnam", Vietnam government groups the country into 8 macro regions and 63 provinces.



This dataset from wikipedia also comprised of "Area (km2)", "Population (2015 census)".

I scraped the additional Lat, Long data from simplemaps, provincial population from pro_pop (2019 census) and merge them to the original dataset.

I use Vietnam Geojon combined with above merge data set to chonopleth map displaying population by province.

Furthermore, I use the main source from data.workbank.org website for the demographic, industrial sector development, GDP, foreign investment trends from 2009 to 2019 to analyse along side with the Foursquare data. This helps to gain insight into the general development trend of the country based on common indicators. To aid in the reliability of the dataset, I use two other sources, namely adb and trading economics as reference check.

2.2. Data cleaning

The List of regions of Vietnam, simplemaps, pro_pop share common keys ("province name") so I used "province name" as primary key to combine these 3 table in to the same dataframe. I thus they don't require additional effort to. Additional effort required to convert them to UTF-8 code for consistency purpose.

The original data.workbank.org dataframe consists of 1432 rows x 12 columns. Dataset contains a total of 1431 social and economic indicators that spans from 1960 to 2019. Missing values are commonly found between 1960 and early 90s and typically the year 2019, therefore I scraped the latest value for 2019 from "data.abd.org" and "tradingeconomics.com". I started the data cleaning process by removing columns "Country Name", "Country Code", "Indicator Code" and "columns: 1960:2009". To avoid ambiguity, I decided to use cut off margin to remove indicators with less than 100% data density. After initial screening the attribute numbers

reduced to 347, among these remainders I observed the duplication in the attributes due to the way how an attribute is estimated, i.e an attribute was estimated by different sources.

I then screen out all the duplications that were not considered as national estimates, not based on current US$ (2018) and not based on total population, this helps to reduce the size of dataframe to 137 rows x 12 columns.

2.3. Features selection

Upon completing the data cleaning process, I split the data.workbank.org dataframe into 4 different categories, namely industrial sectors, demographic, monetary policies and employment. These properties were contained in 6 different dataframes.

'industry_gdp' represents the sector percentage compared to total GDP. Dataframe size: 15x12

'industry_real2018' represents the contributed values from different sectors corresponding to 2018 real term. Dataframe size: 17x12

'population%' represents the working age population compared to total population. Dataframe size: 6x12

'population' represents the 2018 census data, population ages 0-14, 15-64, 65+ and movement between urban and rural populaion. Dataframe size: 8x12

'monpol' describes the interest rate, tax rate trend over time which in turn implies the improvement in monetary policies of the country. Dataframe size: 7x12

'employment' indicates the employability within the labour force, the employability of the total population in general and age dependency ratio between young and old age groups. Dataframe size: 11x12

3. Exploratory Data Analysis

3.1. Relationship between gdp and population

3.2. Relationship between growth and population
3.3. Relationship between growth and population
3.4. Relationship between growth and population
3.5. Relationship between growth and population
3.6. Relationship between growth and population
3.7. Relationship between growth and population
3.8. Relationship between growth and population

4. Predictive Modeling
4.1. Regression models
4.1.1. Applying standard algorithms and their problems
4.1.2. Solutions to problems
4.1.3. Performances of different models

4.2. Classification models

5. Recommendations

6. Conclusions