VIETNAM NATIONAL UNIVERSITY - HO CHI MINH CITY
HO CHI MINH UNIVERSITY OF TECHNOLOGY
FACULTY OF COMPUTER SCIENCE AND ENGINEERING

John von Neumann Institute

HACKATHON 2019
QUALIFICATION ROUND

Problem: Vietnam Macroeconomics Analysis

Team: **noobk**

Candidates: Vũ Đào Anh Tuấn
Đinh Minh Tân
Trần Trương Tuấn Phát

Contents

List of Figures

List of Tables

1 Introduction

1.1 Problems

Exploratory data analysis on the dataset and give us some useful insights as described following those steps:

1. Data Imputation
2. Data Exploration
3. Features Extraction
4. Discussion

1.2 Overview about data

We are provided **Vietnam macroeconomic data** which has:

1. 3 rows:
   - Year, measured in integer, from 1980 to 2018
   - GDP, current prices (Billions of U.S. dollars), measured in float
   - Unemployment rate (Percent), measured in float
2. 39 columns:
   - from 1980 to 2018
   - **10 columns "no data"** value from 1980 to 1989

2 Approaches

2.1 Data Imputation

After measuring Covariance and Correlation, we see there is a negative correlation between GDP and Unemployment rate.

```
In [26]: nonna_df.cov()
```

Out[26]:

	GDP	UR
GDP	5406.645724	-174.571634
UR	-174.571634	8.151059

```
In [27]: nonna_df.corr()
```

Out[27]:

	GDP	UR
GDP	1.000000	-0.831577
UR	-0.831577	1.000000

Figure 1: Covariance and Correlation between GDP and Unemployment rate

Hence, we can fill those missing values by prediction results from **linear regression**.

2.2 Data Exploration

There is a negative correlation between GDP and Unemployment Rate.

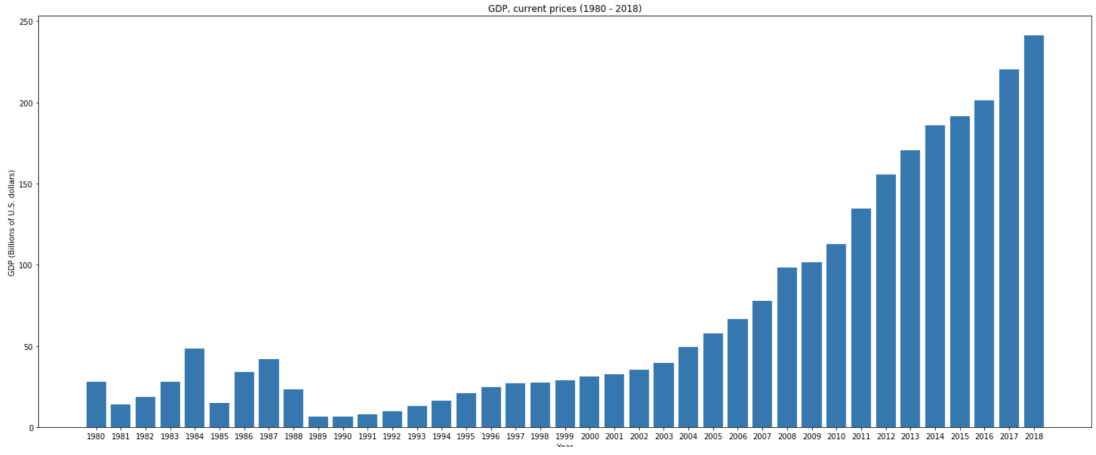

Figure 2: GDP is increasing year after year

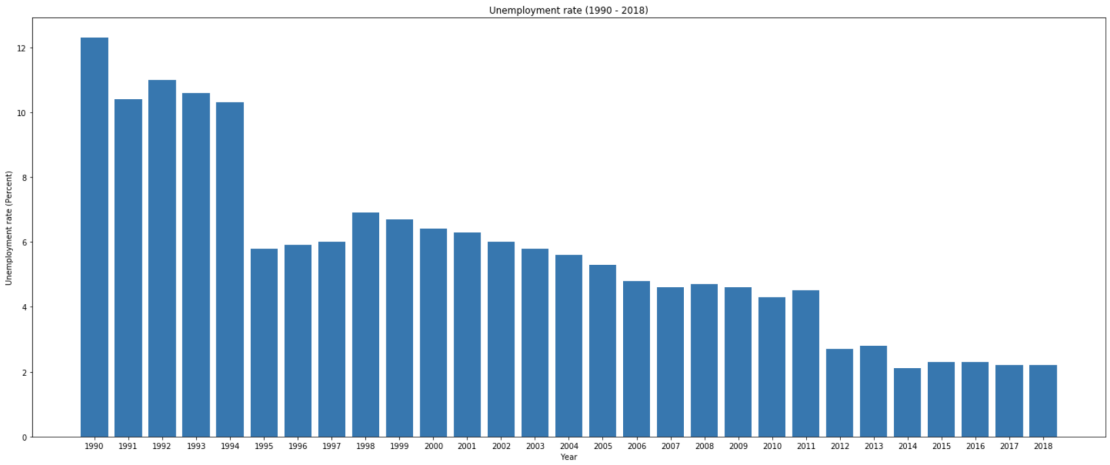

Figure 3: Unemployment Rate is decreasing year after year

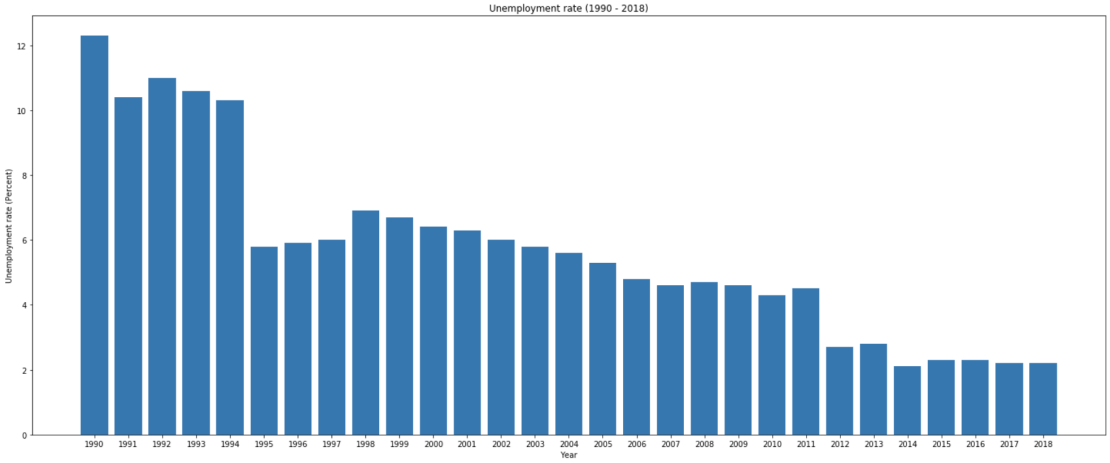

Figure 4: Unemployment Rate is decreasing year after year

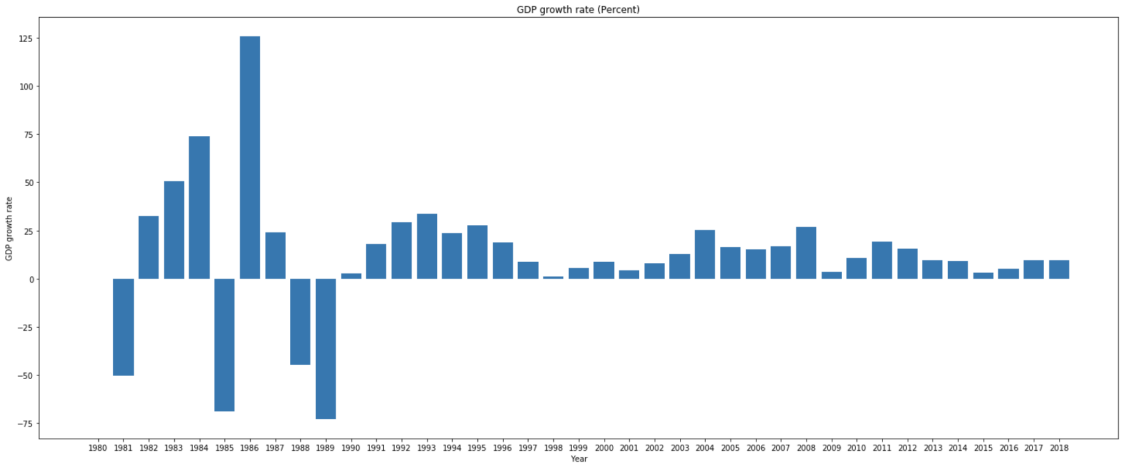

Figure 5: GDP growth rates are positive year after year

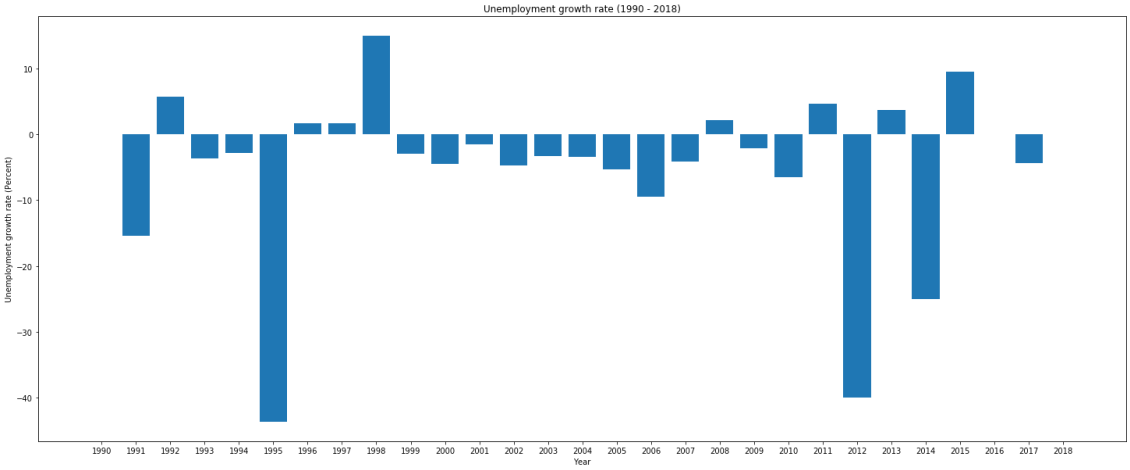

Figure 6: Unemployment Rate are negative year after year

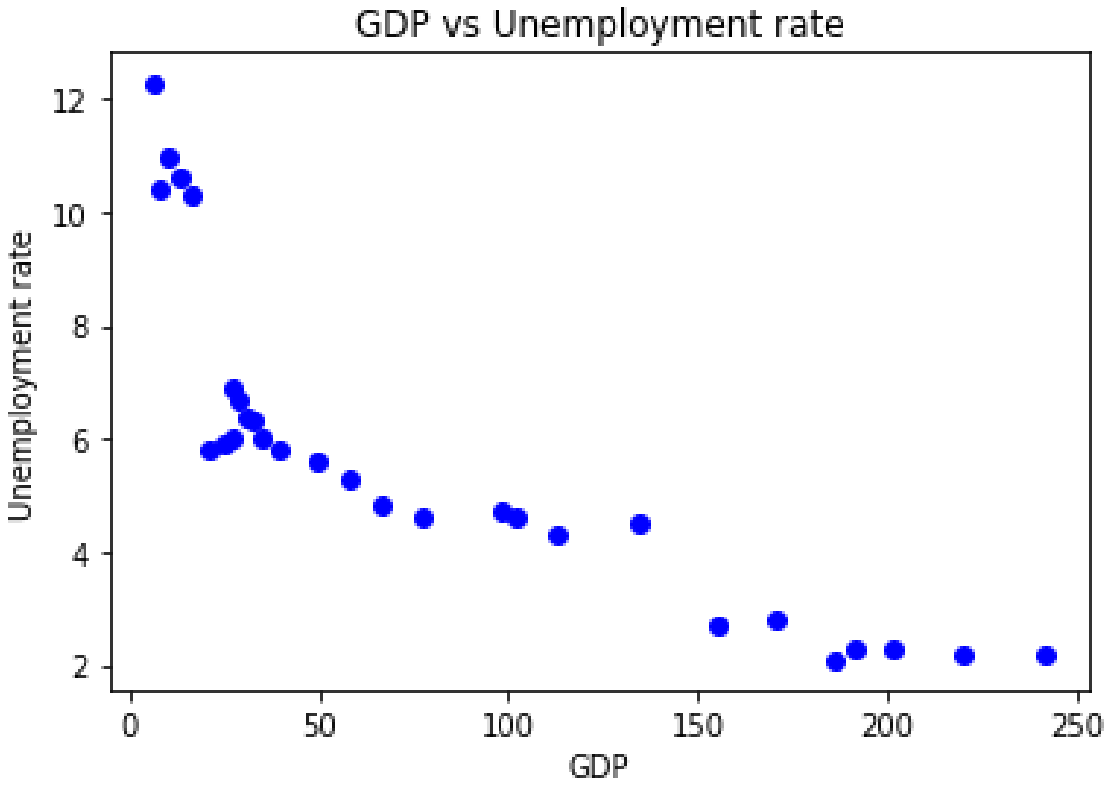

Figure 7: There can be **Linear Regression model** between GDP and Unemployment Rate.

```
unemployment_rate = -0.0311*gdp + 0.0308*gdp_growth_rate + 7.8384
```

Figure 8: Linear Regression model.

- GDP growth rate is always positive over the years
- Unemployment growth rate is always negative over the year
- From 1980 to 1990, GDP had been unstable
- In 2012 and 2015, Unemployment rate decreased strongly.

2.3 Features Extraction

Some features can be extracted from dataset:

- GDP
- Unemployment rate

3 Discussion

After reviewing the dataset set, the group has some options about the data set given as follows:

- The number of features is too small, only 3 rows
- The number of columns is also quite small
- Number of missing values' portion is still large.

For GDP, it should be detailed like real GDP, nominal GDP, and components of GDP like:

- Consumption (C)
- Investment (I)
- Government purchases (G)
- Net Exports (NX)

With unemployment rate need for information such as:

- Total population
- Portions of elderly people, children, workers