

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY
DHARWAD



ज्ञानेन विकासः

Data Analytics on the Indian Economy

BY:

ABHAY KUMAR (15CS01)

KUSHAL DEB (15CS06)

DISCIPLINE OF COMPUTER SCIENCE AND ENGINEERING

December 8, 2018

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY
DHARWAD

Data Analytics on the Indian Economy

A PROJECT REPORT

Submitted in partial fulfillment of the
requirements for the award of the degrees

of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

Submitted by:

Abhay Kumar (15CS01)
Kushal Deb (15CS06)

Guided by:

Dr. Kavi Mahesh
Director, IIIT Dharwad

December 8, 2018

CANDIDATE’S DECLARATION

We hereby declare that the project entitled “Data Analytics on the Indian Economy” submitted in partial fulfillment for the award of the degree of Bachelor of Technology in ‘Computer Science and Engineering’ completed under the supervision of Dr. Kavi Mahesh, Director, IIIT Dharwad is an authentic work.

Signature and name of the student(s) with date

CERTIFICATE by MINI PROJECT Guide(s)

It is certified that the above statement made by the students is correct to the best of my knowledge.

Signature and name of the guide(s) with date

Preface

This report on “Data Analytics on the Indian Economy” is prepared under the guidance of Dr. Kavi Mahesh.

In this project we have analysed the following facets of the Indian economy-

1. Employment
2. Tax
3. Digital Payment
4. GDP
5. Inflation

Abhay Kumar (15CS01)

Kushal Deb (15CS06)

B.Tech, 4th Year

Discipline of Computer Science and Engineering

IIIT Dharwad

Acknowledgements

We wish to thank Dr. Kavi Mahesh for his kind support and valuable guidance.

We are really thankful to him for taking out time to guide us in the project inspite of his very busy schedule. The project benefited a lot from his invaluable guidance.

Without his support this report would not have been possible.

Abhay Kumar (15CS01)
Kushal Deb (15CS06)
B.Tech, 4th Year
Discipline of Computer Science and Engineering
IIIT Dharwad

Abstract

In this project we have analyzed the following facets of the Indian economy-

1. Employment
2. Tax
3. Digital Payment
4. GDP
5. Inflation

The project will depict how the above mentioned facets of the Indian economy have changed and are changing through the years.

Contents

1	Employment	8
1.1	Part 1	8
1.1.1	Overall Population	8
1.1.2	Male Population	9
1.1.3	Female Population	10
1.2	Part 2	11
1.2.1	Employment in industries	12
1.2.2	Wage and Salaried Workers	13
1.2.3	Vulnerable Employment	14
1.3	PART 3	15
1.3.1	Employment to Population ratio	15
2	Tax Data	17
2.1	Part 1	17
2.2	Part 2	19
2.3	Part 3	25
2.4	Part 4	27
2.5	Part 5	27
3	Digital Payment	31
4	GDP Growth	33
5	Inflation	34
6	Others	35
7	Conclusion	37
8	Future Works	37
9	Dataset Sources	38
10	References	39

List of Figures

1	Overall Unemployment Percentage	9
2	Male Unemployment Percentage	10
3	Female Unemployment Percentage	11
4	Overall Employment Percentage in Industries	13
5	Overall employment Percentage for Wage and Salaried workers	14
6	Overall Vulnerable Employment Percentage	15
7	Plot of Population (Employed vs Unemployed)	16
8	Corporate and Personal Taxes	18
9	The difference between the taxes of each year	19
10	Tax collected from states in 2015 - 16	20
11	Tax collected from states in 2016 - 17	21
12	Tax collected from states in 2017 - 18*	22
13	Percentage change in tax collected per state in 2016-17	23
14	Percentage change in tax collected per state in 2017-18*	24
15	Total Tax, Direct tax and Indirect tax	25
16	Direct tax as percentage of total tax	26
17	Percentage change in Direct Tax	26
18	Relationship between GDP Growth rate and Tax Growth rate	27
19	Return filed by category.	28
20	Number of returns filed in financial year 2017 – 18	28
21	Percentage change in the number of persons filling in the different category in between 2013 – 14 to 2018 -17.	29
22	Number of taxpayers in financial year 2017 -18	30
23	Percentage change in the number of the taxpayers in the dif- ferent categories in between 2013 – 14 to 2018 -17	30
24	Usage of different modes of online transaction	32
25	GDP	33
26	Inflation as Percentage	34
27	Seasonality and trend of GDP	35
28	Time series plot of GDP	36

1 Employment

All the data for this section has been taken from World Bank.

1.1 Part 1

In this part we will see the general trend of unemployment in India from 1991 to 2017.

First we see the correlation of unemployment with respect to the population of working age.

1.1.1 Overall Population

The Pearson coefficient for correlation between population (of working age, 15-64) and unemployment is -0.39033803120875155 from the year 1991 to 2017.

We can see that the coefficient is negative which means that the correlation is negative but the magnitude of the correlation is too little to conclude anything with affirmation.

For particular years-

- 2004 to 2009 (UPA Term1) – Unemployment decreasing with Pearson coefficient factor being -0.883357686068193
- 2009 to 2014 (UPA Term 2) - Unemployment decreasing with Pearson coefficient factor being -0.7118533674591503
- 2014 to 2017 (NDA) – Unemployment increasing with Pearson coefficient factor being 0.9411527157423891

Plot of Overall Unemployment Percentage is as follows.



Figure 1: Overall Unemployment Percentage

1.1.2 Male Population

The Pearson coefficient for correlation between population (of working age, 15-64) and male unemployment is -0.5386666627079341 from the year 1991 to 2017.

We can see that the coefficient is negative which means that the correlation is negative but the magnitude of the correlation is too little to conclude that unemployment is decreasing with affirmation.

For particular years

- 2004 to 2009 (UPA Term1) – Unemployment decreasing with Pearson coefficient factor being -0.8328871401291689
- 2009 to 2014 (UPA Term 2) - Unemployment decreasing with Pearson coefficient factor being -0.4782438498399577
- 2014 to 2017 (NDA) – Unemployment has been increasing with Pearson coefficient factor being 0.9766665796899555

Plot of Male Unemployment Percentage is as follows



Figure 2: Male Unemployment Percentage

1.1.3 Female Population

The Pearson coefficient for correlation between population (of working age, 15-64) and female unemployment is 0.3338909522033719 from the year 1991 to 2017.

We can see that the coefficient is positive which means that the correlation is positive but the magnitude of the correlation is too little to conclude that unemployment is increasing with population.

For particular years

- 2004 to 2009 (UPA Term1) – Unemployment decreasing with Pearson coefficient factor being -0.6695944187581783 .
- 2009 to 2014 (UPA Term 2) - Unemployment decreasing with Pearson coefficient factor being -0.9885937949927284 .
- 2014 to 2017 (NDA) – Unemployment has been increasing with Pearson coefficient factor being 0.5273043026065888 .

Plot of the Female Unemployment Percentage is as follows

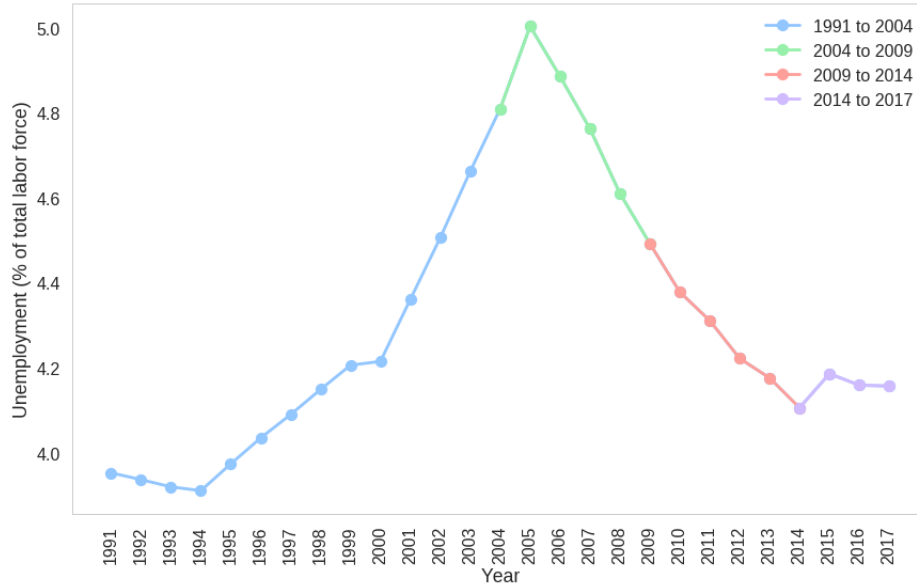


Figure 3: Female Unemployment Percentage

1.2 Part 2

In this part we will see the general trend of employment in India from 1991 to 2017 for different types of industries.

The different types of employments are-

1. **Employment in Industries** - Employment is defined as persons of working age who were engaged in any activity to produce goods or provide services for pay or profit, whether at work during the reference period or not at work due to temporary absence from a job, or to working-time arrangement. The industry sector consists of mining and quarrying, manufacturing, construction, and public utilities (electricity, gas, and water), in accordance with divisions 2-5 (ISIC 2) or categories C-F (ISIC 3) or categories B-F (ISIC 4).
2. **Wage and Salaried Workers** - Wage and salaried workers (employees) are those workers who hold the type of jobs defined as "paid employment jobs," where the incumbents hold explicit (written or oral) or implicit employment contracts that give them a basic remuneration

that is not directly dependent upon the revenue of the unit for which they work.

3. **Vulnerable Employment** - Vulnerable employment is contributing family workers and own-account workers as a percentage of total employment .

Correlation of employment with respect to the population of working age for each type of employments. The correlations have been done with the overall population.

1.2.1 Employment in industries

We see that the overall employment in industries has increased with respect to the working population from 1991 to 2017. The Pearson coefficient is 0.9650492484455186.

For particular years

- 2004 to 2009 (UPA Term1) – Employment increasing with Pearson coefficient factor being 0.9730914301293814 .
- 2009 to 2014 (UPA Term 2) - Employment increasing with Pearson coefficient factor being 0.889162692141398 .
- 2014 to 2017 (NDA) – Employment has been decreasing with Pearson coefficient factor being -0.8366886077104094 .

Plot of the Overall Employment percentage in Industries is as follows

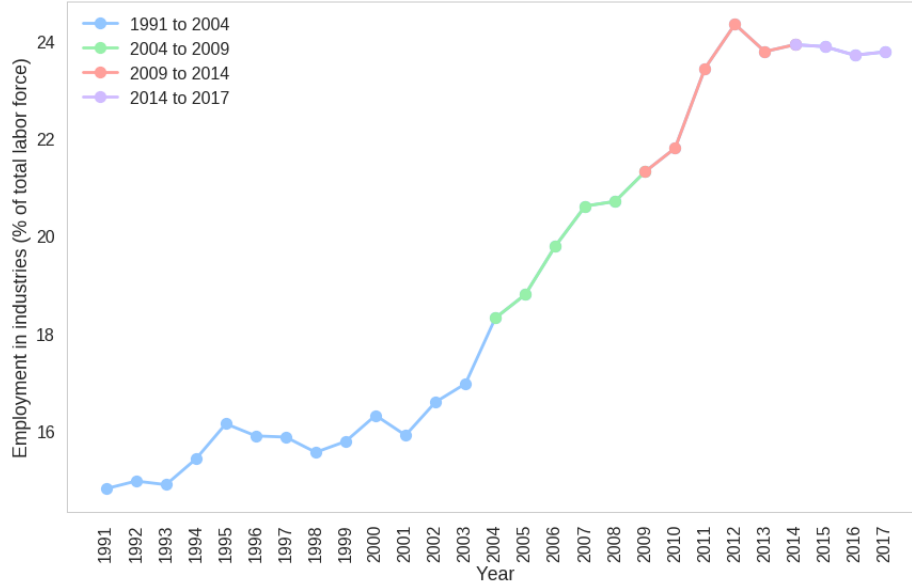


Figure 4: Overall Employment Percentage in Industries

1.2.2 Wage and Salaried Workers

We see that the Overall employment for wage and salaried workers has increased with respect to the working population from 1991 to 2017. The Pearson coefficient is 0.7732885889506631.

For particular years

- 2004 to 2009 (UPA Term1) – Employment increasing with Pearson coefficient factor being 0.9808729414448635.
- 2009 to 2014 (UPA Term 2) - Employment increasing with Pearson coefficient factor being 0.9309755309877512.
- 2014 to 2017 (NDA) – Employment has been increasing with Pearson coefficient factor being 0.9755166180580906.

Plot of the Overall Employment percentage for Wage and Salaried workers is as follows.

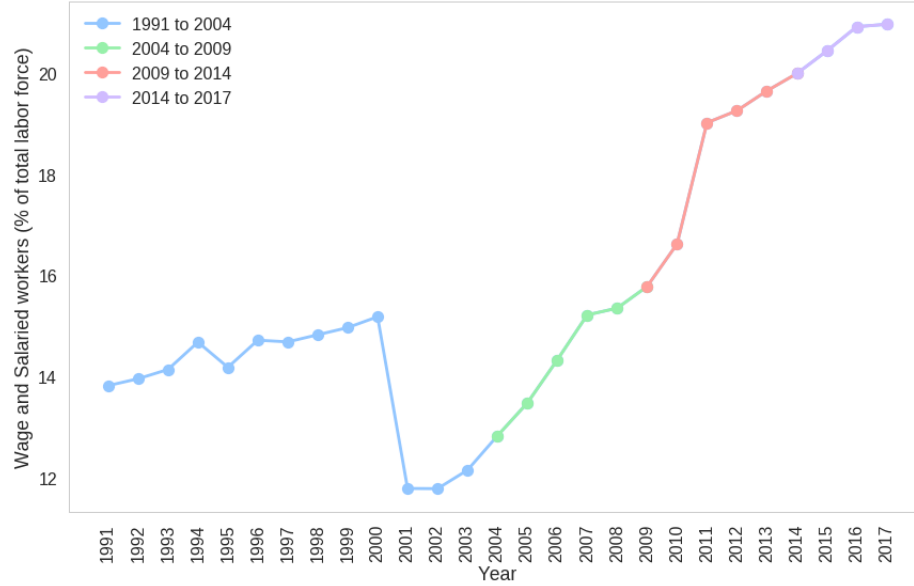


Figure 5: Overall employment Percentage for Wage and Salaried workers

1.2.3 Vulnerable Employment

We see that the Overall Vulnerable Employment has decreased with respect to the population from 1991 to 2017. The Pearson coefficient is -0.7213439890808735.

For particular years

- 2004 to 2009 (UPA Term1) – Employment decreasing with Pearson coefficient factor being -0.9791994147908917 .
- 2009 to 2014 (UPA Term 2) - Employment decreasing with Pearson coefficient factor being -0.9265044777230029 .
- 2014 to 2017 (NDA) – Employment decreasing with Pearson coefficient factor being -0.9761536812271961 .

Decreasing employment in Vulnerable sector is a good sign for the economy because it shows that more people are moving from vulnerable jobs to stable jobs.

Plot of Overall Vulnerable Employment percentage is as follows.

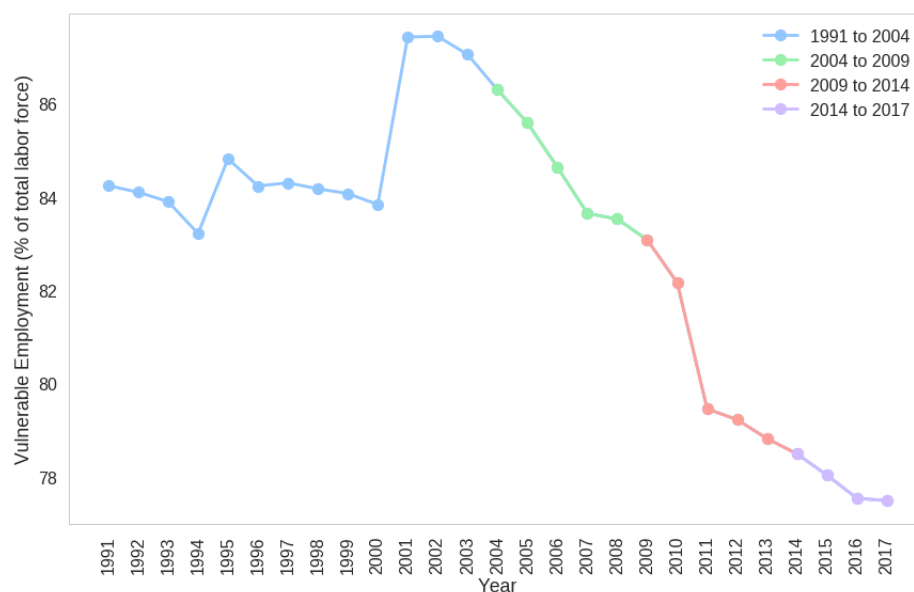


Figure 6: Overall Vulnerable Employment Percentage

1.3 PART 3

This part explores the population of people that are employed and unemployed.

1.3.1 Employment to Population ratio

Employment to population ratio is the proportion of a country's population that is employed.

Employment is defined as persons of working age who, during a short reference period, were engaged in any activity to produce goods or provide services for pay or profit, whether at work during the reference period (i.e. who worked in a job for at least one hour) or not at work due to temporary absence from a job, or to working-time arrangements. Ages 15 and older are generally considered the working-age population.

We have the total population between the working age. From that we have calculated the employed population and the unemployed population.

Employed population = Employment to population ratio * total population

Unemployed population = Total population – Employed population

Plot of population is as follows

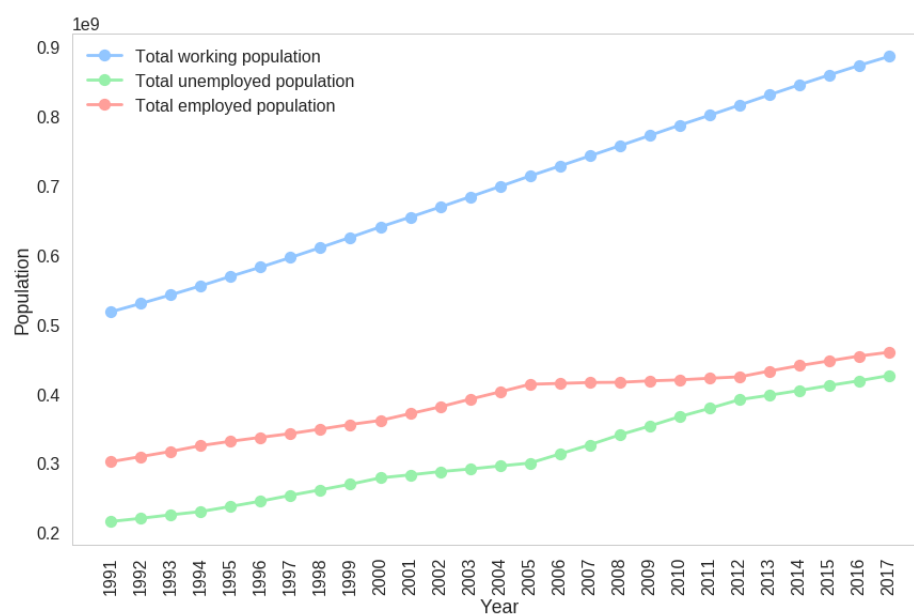


Figure 7: Plot of Population (Employed vs Unemployed)

We can see from the above plot that both the employment rate and unemployment rates are increasing at a steady rate.

2 Tax Data

2.1 Part 1

In this part we will see the different trends of taxes in each financial years through the years.

The different types of taxes are -

1. **Income Tax** - This tax is paid by the taxpayers other than companies registered under company law in India on the income earned by them. They are taxed on the basis of slabs at different rates.
2. **Corporate Tax** - This tax is paid by the companies registered under company law in India on the net profit that it makes from businesses. It is taxed at a specific rate as prescribed by the income tax act subject to the changes in the rates every year by the IT department.
3. **Personal Tax** - An income tax is a tax imposed on individuals or entities (taxpayers) that varies with respective income or profits (taxable income). Income tax generally is computed as the product of a tax rate times taxable income. Taxation rates may vary by type or characteristics of the taxpayer.

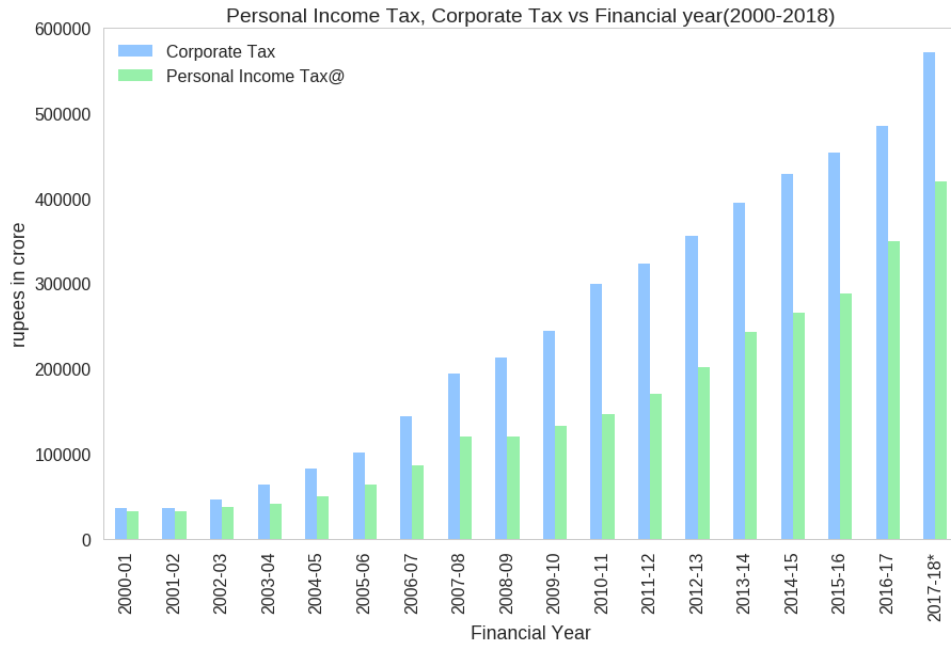


Figure 8: Corporate and Personal Taxes

We can see from the plot that the amount of taxes is increasing with each passing year.

The following plot shows us in which particular year the collection of personal tax and corporate tax was maximum.

- The maximum personal tax is : 419998.0 at 2017 – 18* financial year
- The maximum corporate tax is : 571202.0 at 2017 – 18* financial year which is the most recent financial year.

The difference between the taxes of each year from the previous year is as follows

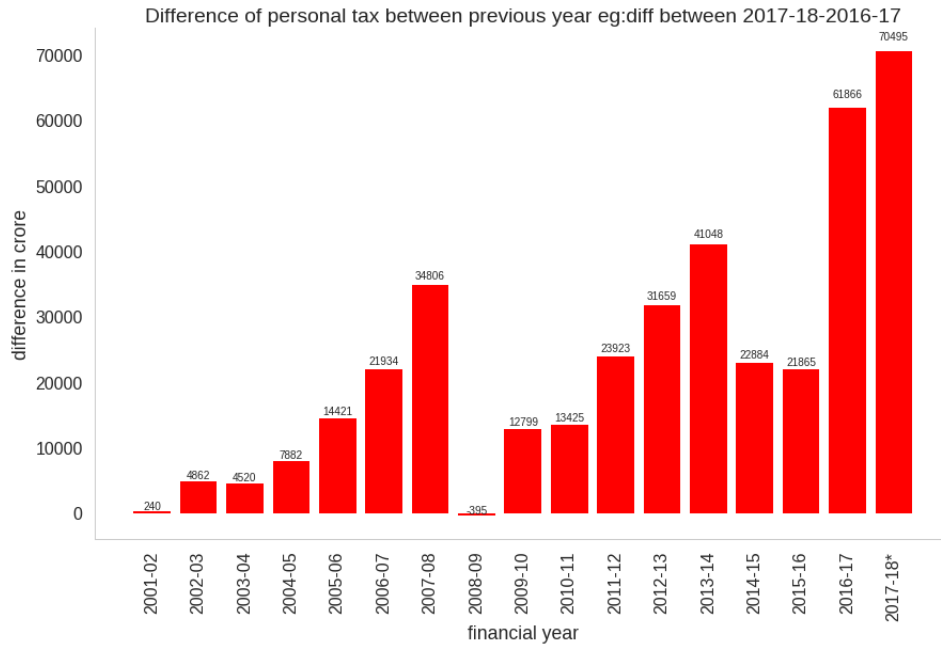


Figure 9: The difference between the taxes of each year

We can see that there is a rapid increase in the tax collection in 2016-17 and 2017-18*.

2.2 Part 2

In this part we will see the tax collection from each state. The plots of all the tax collected from states is as follows. The given years are -:

- 2015 - 16
- 2016 - 17
- 2017 - 18*

For the year 2015-16

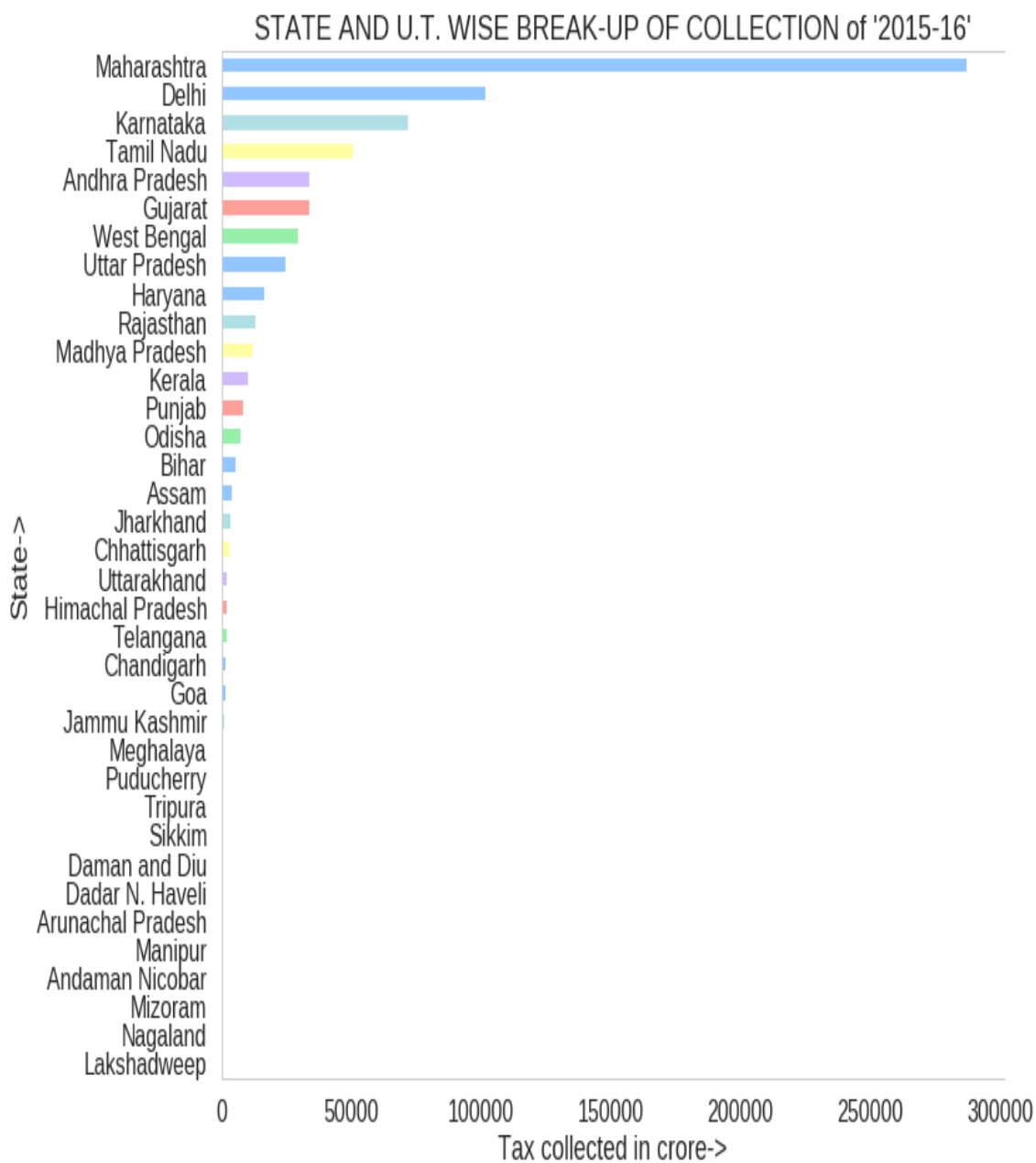


Figure 10: Tax collected from states in 2015 - 16

For the year 2016-17

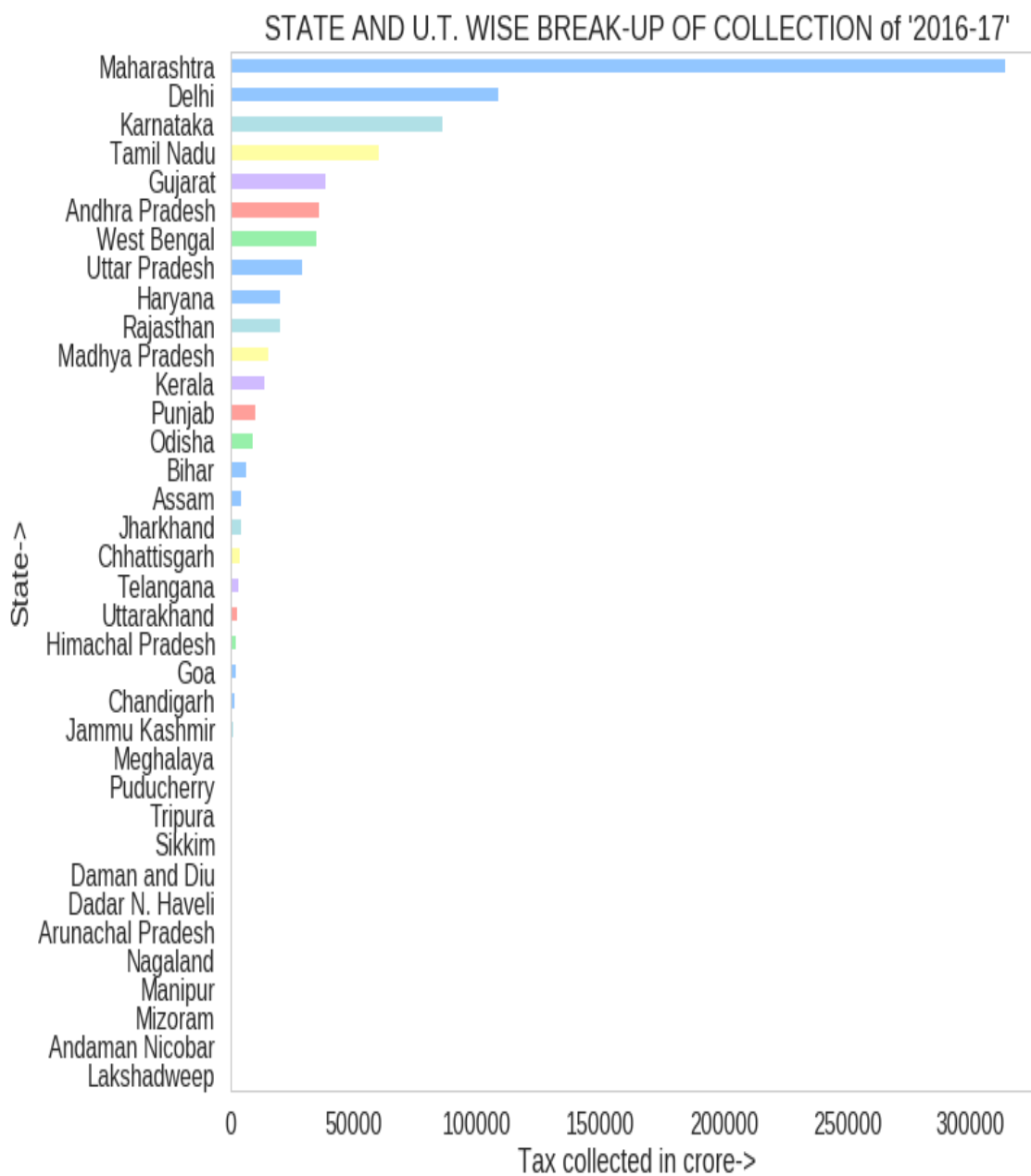


Figure 11: Tax collected from states in 2016 - 17

For the year 2017 – 18*

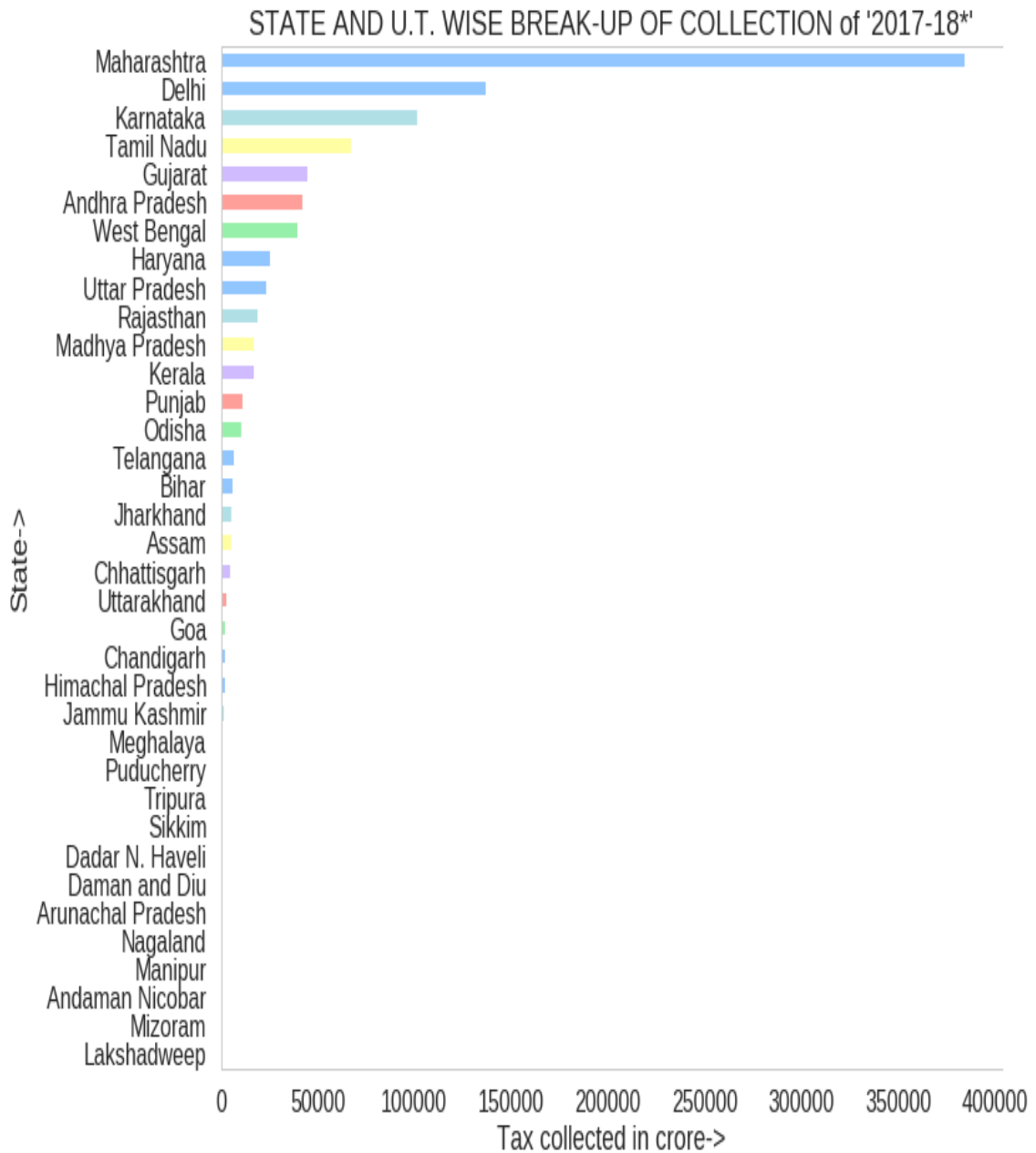


Figure 12: Tax collected from states in 2017 - 18*

Now we will see the difference in tax collected in each year with respect to the previous year.

This is calculated in terms of “Change Percentage”.

$$\text{Change} = \frac{|\text{Present Year collection} - \text{Previous Year Collection}|}{\text{Previous Year Collection}}$$

The plots for the change percentage are -
Percentage change in 2016-17

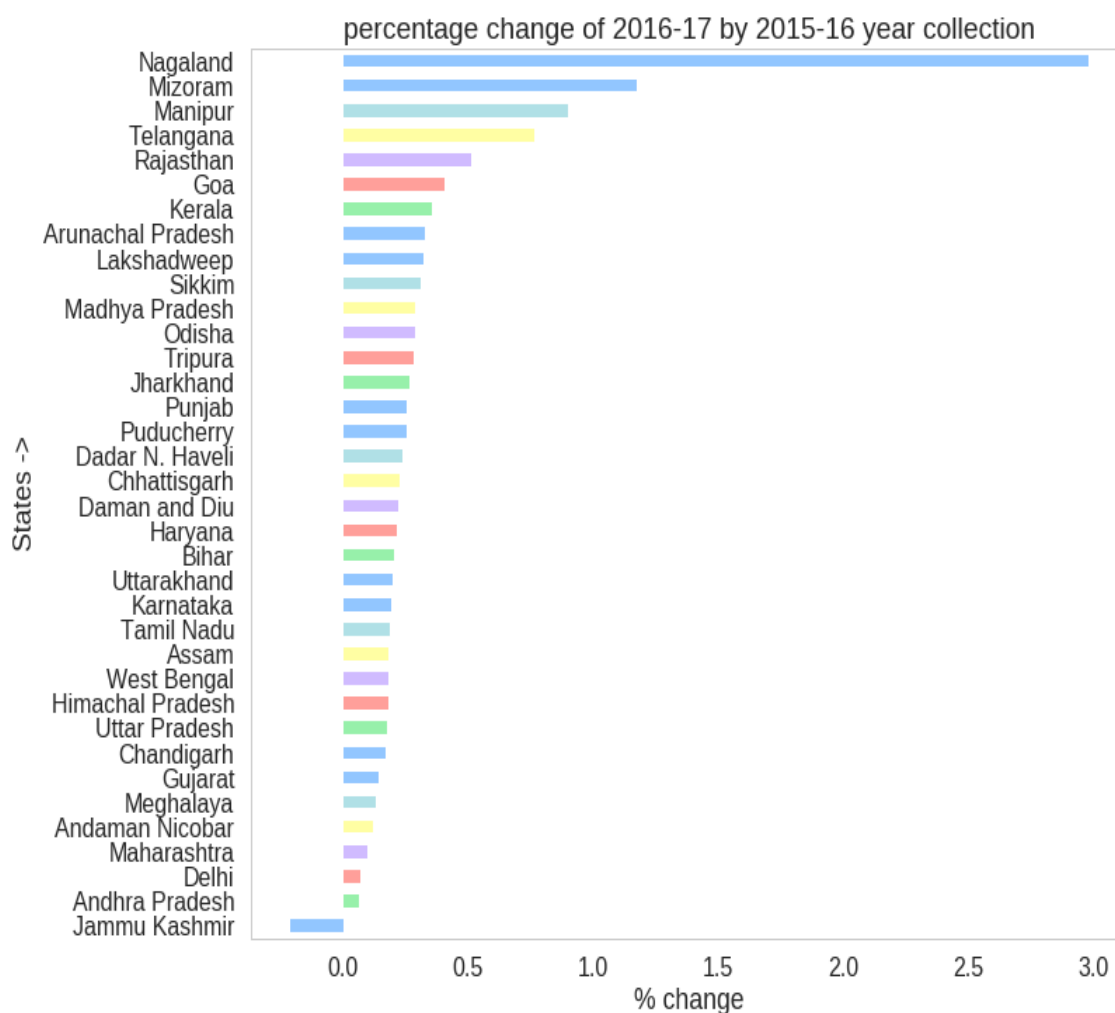


Figure 13: Percentage change in tax collected per state in 2016-17

Percentage change in 2017-18*

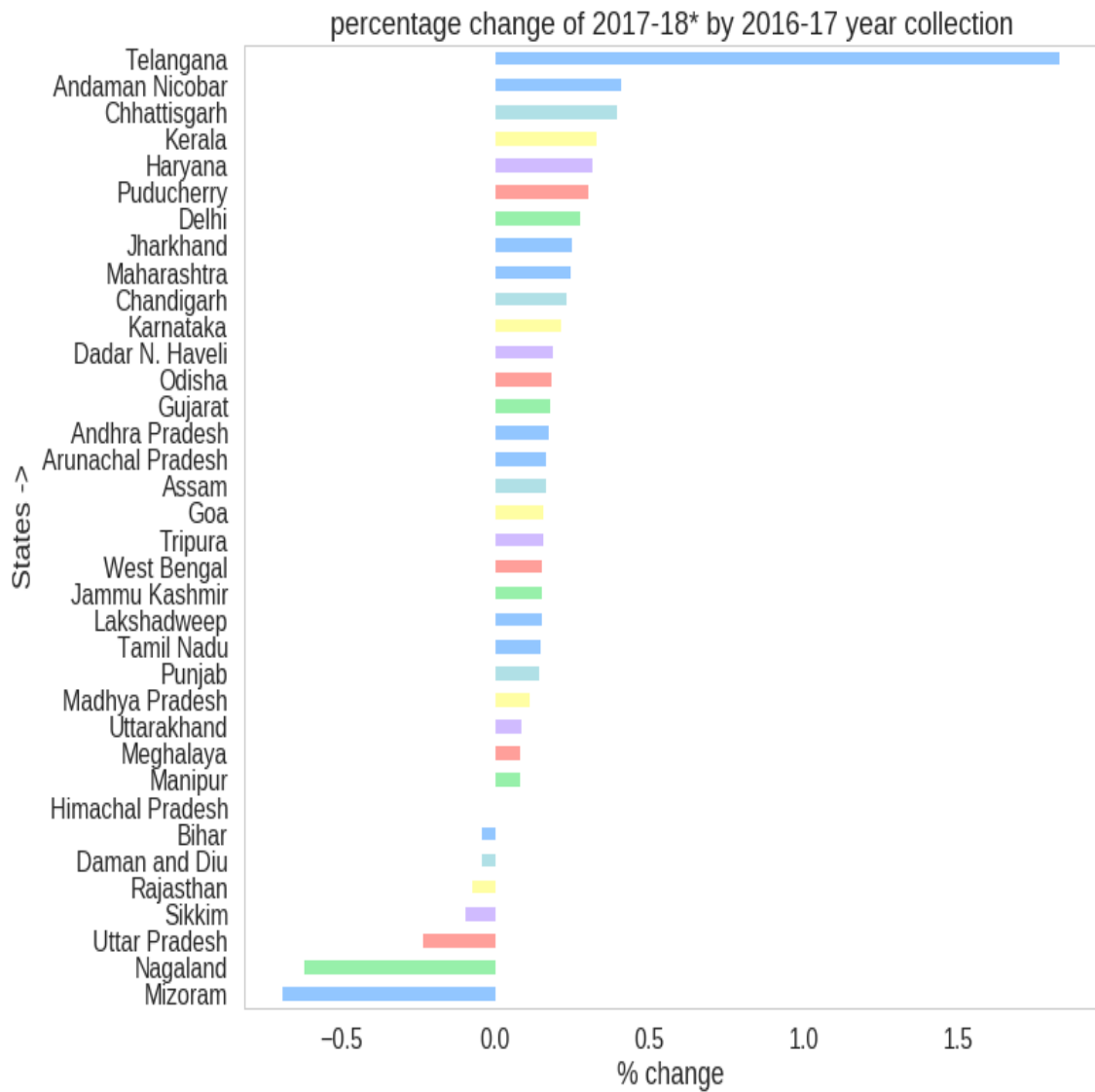


Figure 14: Percentage change in tax collected per state in 2017-18*

2.3 Part 3

In this part we see the contribution of Direct and Indirect taxes on the Total tax.

- **Direct Tax** - Direct tax is a type of tax where the incidence and impact of taxation fall on the same entity. Description: In the case of direct tax, the burden can't be shifted by the taxpayer to someone else. These are largely taxes on income or wealth.
- **Indirect Tax** - An indirect tax (such as sales tax, per unit tax, value added tax (VAT), or goods and services tax (GST)) is a tax collected by an intermediary (such as a retail store) from the person who bears the ultimate economic burden of the tax (such as the consumer).

The plot for Total tax, Direct tax and Indirect tax from the year 2000-01 to 2016-16 is as follows

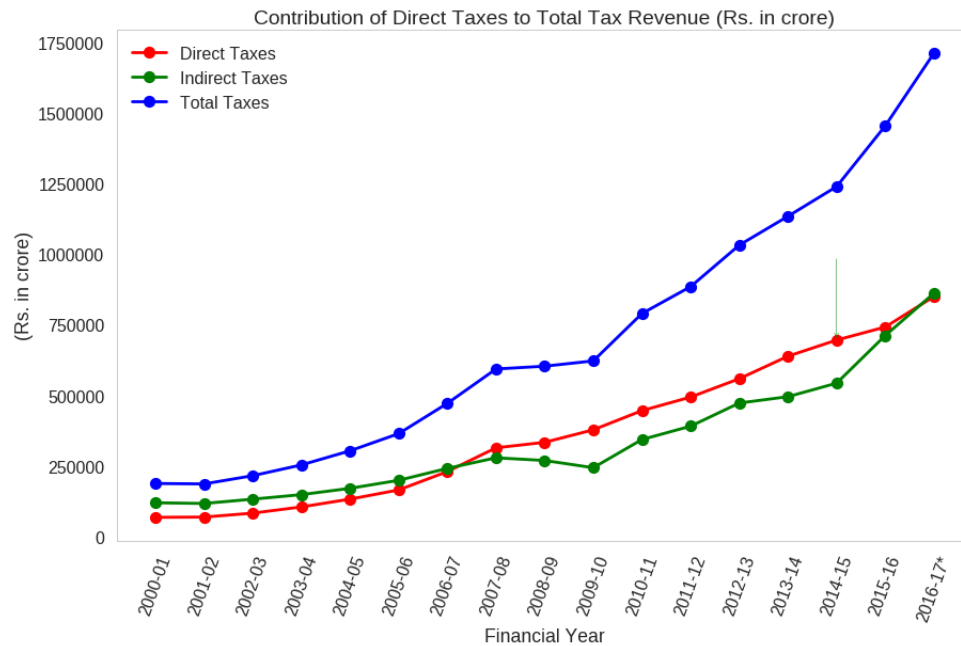


Figure 15: Total Tax, Direct tax and Indirect tax

Plot of Direct tax as percentage of total tax is as follows

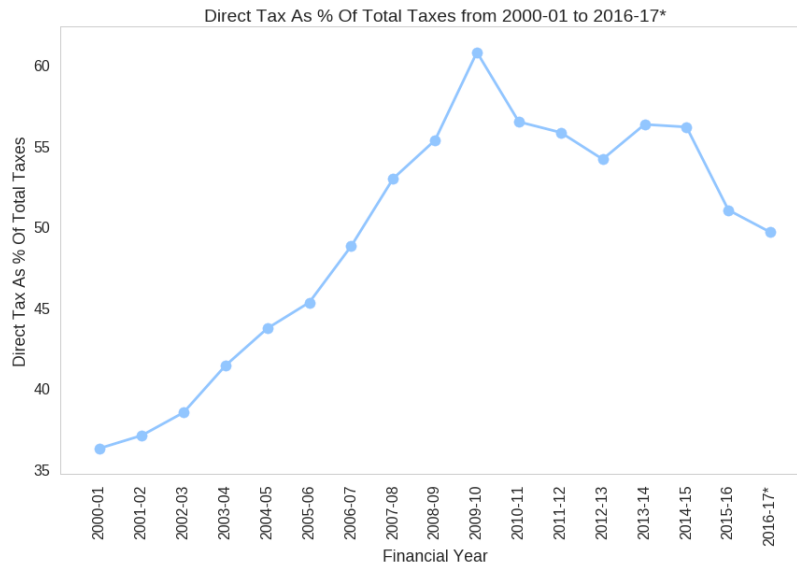


Figure 16: Direct tax as percentage of total tax

The Plot of percentage change in direct tax in each year with respect to the previous year is as follows

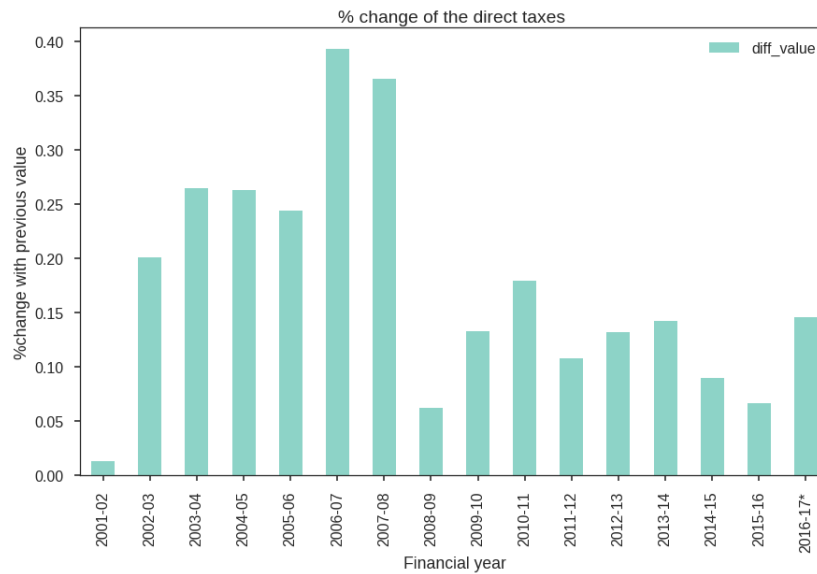


Figure 17: Percentage change in Direct Tax

2.4 Part 4

In this part we see the relationship between GDP Growth rate and Tax Growth rate.

The Pearson correlation between GDP growth rate and Tax growth rate is 0.461925.

The plot for this is as follows.

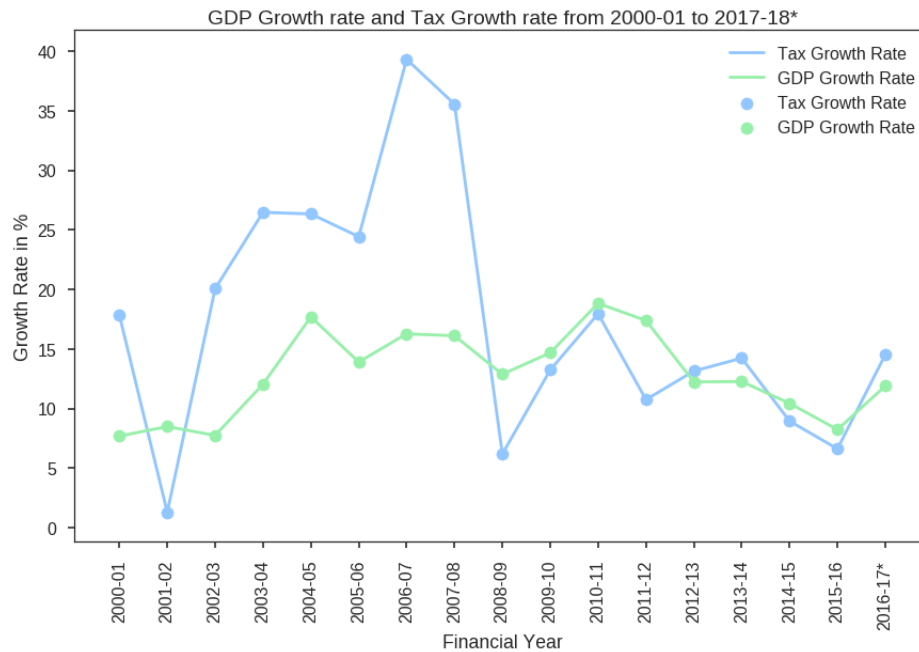


Figure 18: Relationship between GDP Growth rate and Tax Growth rate

2.5 Part 5

In this part we see the maximum return field by which category and the difference.

The plot for tax returns filed by individuals is as follows.

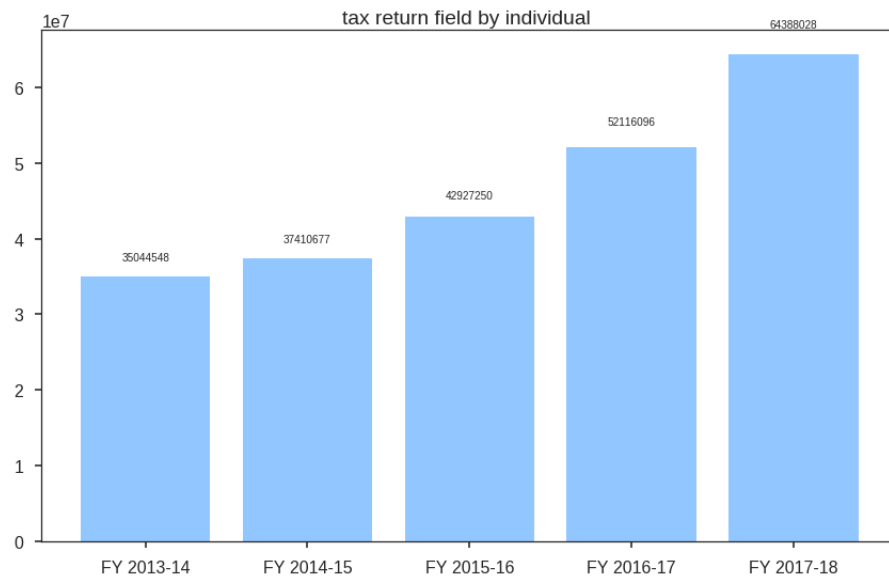


Figure 19: Return field by category.

Number of returns filed in financial year 2017 – 18

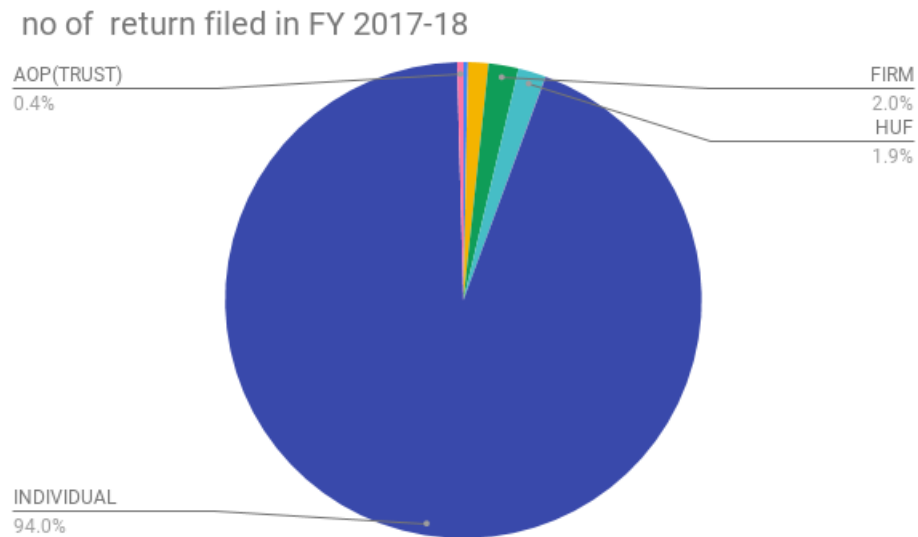


Figure 20: Number of returns filed in financial year 2017 – 18

Plot for percentage change in the number of persons filling in the different category in between 2013 – 14 to 2018 -17.

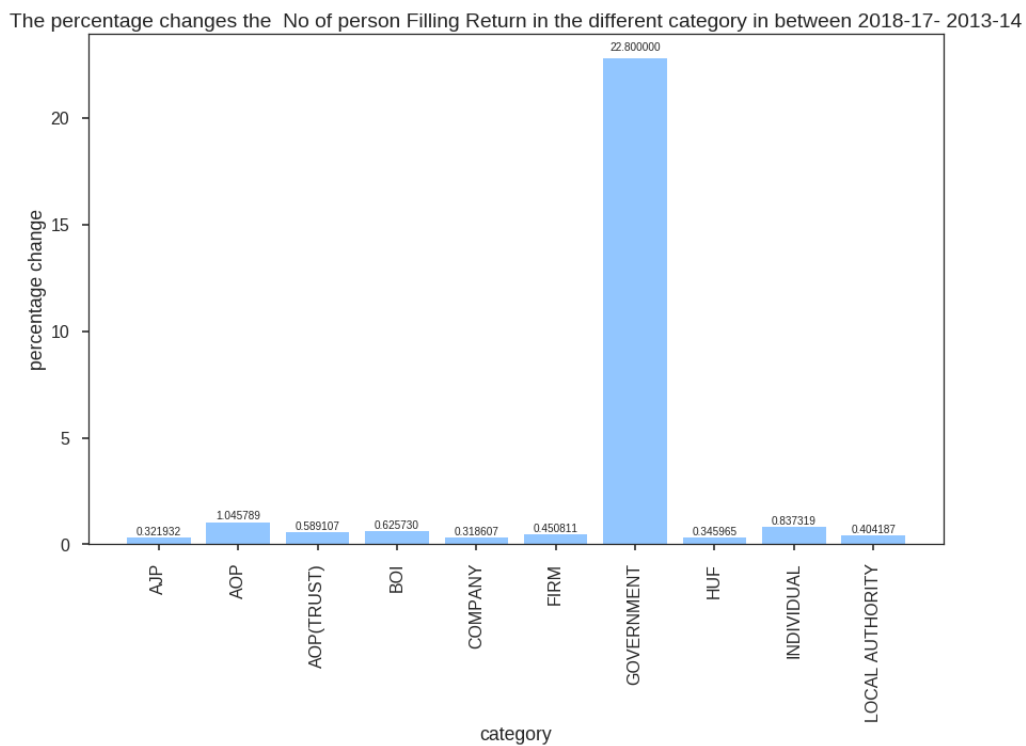


Figure 21: Percentage change in the number of persons filling in the different category in between 2013 – 14 to 2018 -17.

Number of taxpayers in financial year 2017 -18

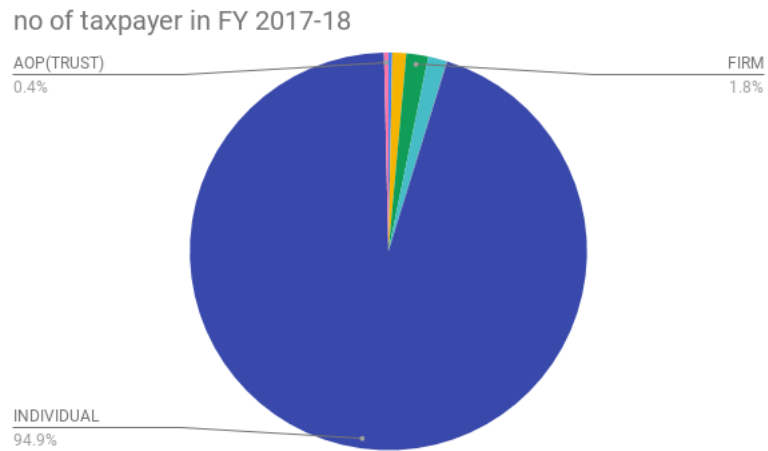


Figure 22: Number of taxpayers in financial year 2017 -18

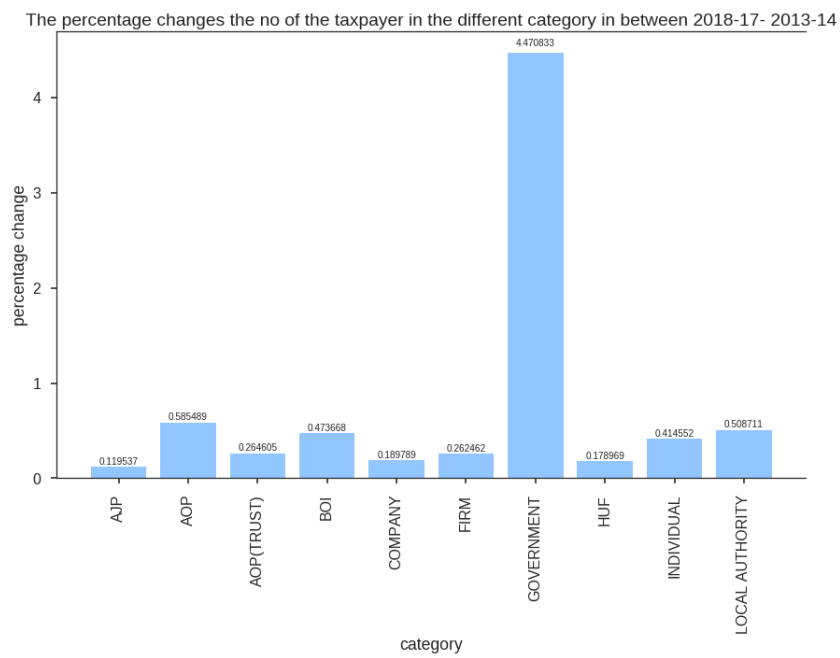


Figure 23: Percentage change in the number of the taxpayers in the different categories in between 2013 – 14 to 2018 -17

3 Digital Payment

Here, we will analyze the Digital Payment system in India. These mode of payments come under Digital Payments -

1. **Prepaid Payment Instruments** - Prepaid payment instruments are methods that facilitate purchase of goods and services against the value stored on such instruments. The value stored on such instruments represents the value paid for by the holder, by cash, by debit to a bank account, or by credit card.
2. **m-Wallet** - The mobile wallet, which is also called mWallet, digital wallet, or eWallet, refers to a mobile technology that is used similarly to a real wallet. The Mobile Wallet provides a convenient solution for any business looking to allow customers to purchase their products online with greater ease, therefore driving sales.
3. **PPI Cards** - PPI stands for 'Payment Protection Insurance'. It's designed to cover credit card repayments for a year in the event of an accident, sickness or unemployment, or sometimes just accident and sickness. Yet it's been widely mis-sold, and you could even have it without knowing.
4. **Paper Voucher** - A paper or gift voucher is a card or piece of paper that you buy at a shop and give to someone, which entitles the person to exchange it for goods worth the same amount.
5. **Mobile Banking** - Mobile banking refers to the use of a smartphone or other cellular device to perform online banking tasks while away from your home computer, such as monitoring account balances, transferring funds between accounts, bill payment and locating an ATM.

The following graph shows the usage of different types of online transaction modes.

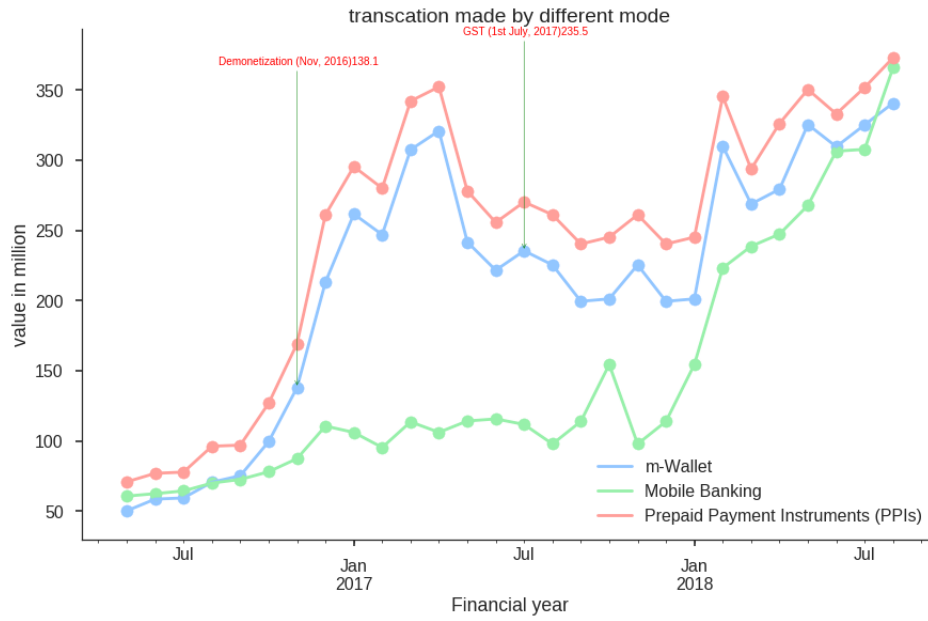


Figure 24: Usage of different modes of online transaction

From the above graph we can see that the most used mode on online payment is **PPIs**.

The **Pearson correlation** between the different online modes of payments are-

	PPIs	m-Wallet	Mobile Banking
PPIs	1.000000	0.996694	0.705593
m-Wallet	0.996694	1.000000	0.726938
Mobile Banking	0.705593	0.726938	1.000000

4 GDP Growth

Definition of GDP

Gross domestic product (GDP) is the monetary value of all the finished goods and services produced within a country's borders in a specific time period. Though GDP is usually calculated on an annual basis, it can be calculated on a quarterly basis as well (in the United States, for example, the government releases an annualized GDP estimate for each quarter and also for an entire year).

GDP includes all private and public consumption, government outlays, investments, private inventories, paid-in construction costs and the foreign balance of trade (exports are added, imports are subtracted). Put simply, GDP is a broad measurement of a nation's overall economic activity. It may be contrasted with gross national product (GNP), which measures the overall production of an economy's citizens, including those living abroad, while domestic production by foreigners is excluded.

In this part we will see the trend of GDP growth through the years. The plot for GDP is in fig 25.

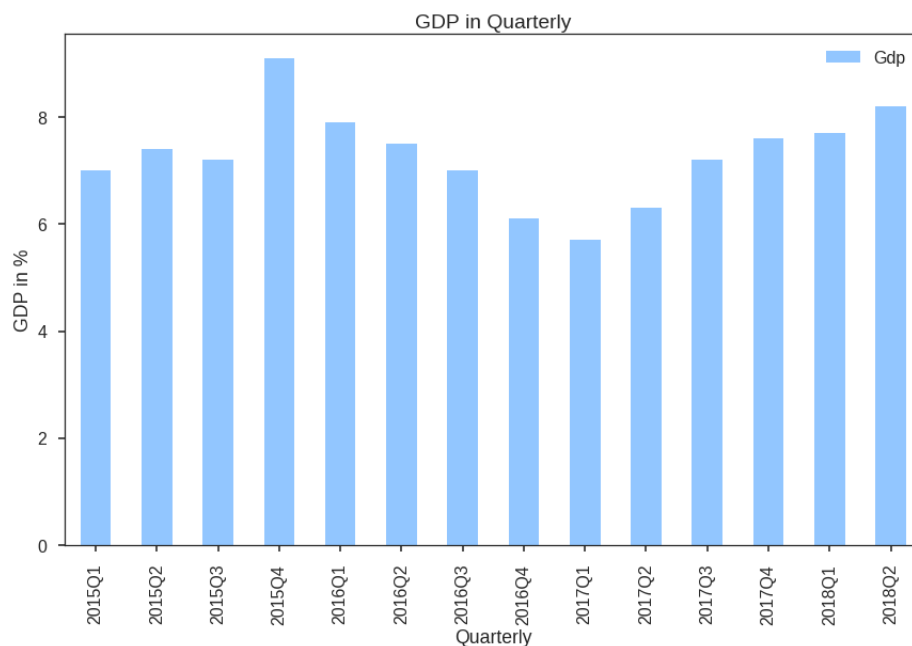


Figure 25: GDP

5 Inflation

Definition of Inflation

Inflation is a quantitative measure of the rate at which the average price level of a basket of selected goods and services in an economy increases over a period of time. Often expressed as a percentage, inflation indicates a decrease in the **purchasing power** of a nation's currency. As prices rise, they start to impact the general cost of living for the common public and the appropriate monetary authority of the country, like the **central bank**, then takes the necessary measures to keep inflation within permissible limits and keep the economy running smoothly. Inflation is measured in a variety of ways depending upon the types of goods and services considered, and is the opposite of deflation which indicates a general decline occurring in prices for goods and services when the inflation rate falls below 0 percent.

Plot of inflation as percentage through the years



Figure 26: Inflation as Percentage

6 Others

In this section we will discuss about the topics which we tried to incorporate in the project but could not do due to some factor(s).

1. Employment

- Due to lack of data regarding the employment of various sectors we could not accommodate those into the employment dataset.
- We tried calculate the rate at which we need to increase employment to accommodate the entire working population in the workforce in the next decade or so. But, we could not find a suitable algorithm/technique to do so.

2. GDP

- We don't get any seasonality component from the data only trend due the lack of monthly data. (Figure:27)
- We tried to do time series analysis on GDP but the problem we encountered is that the plot for the time series was very smooth and it did not capture the true nature of GDP. (Figure:28)

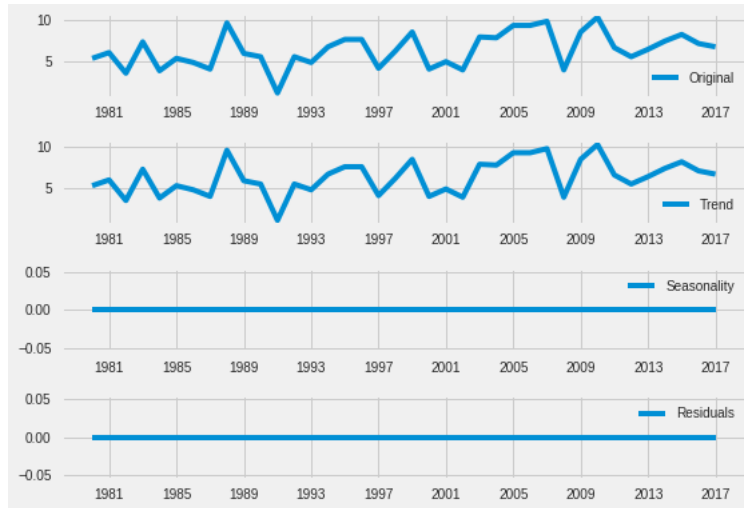


Figure 27: Seasonality and trend of GDP

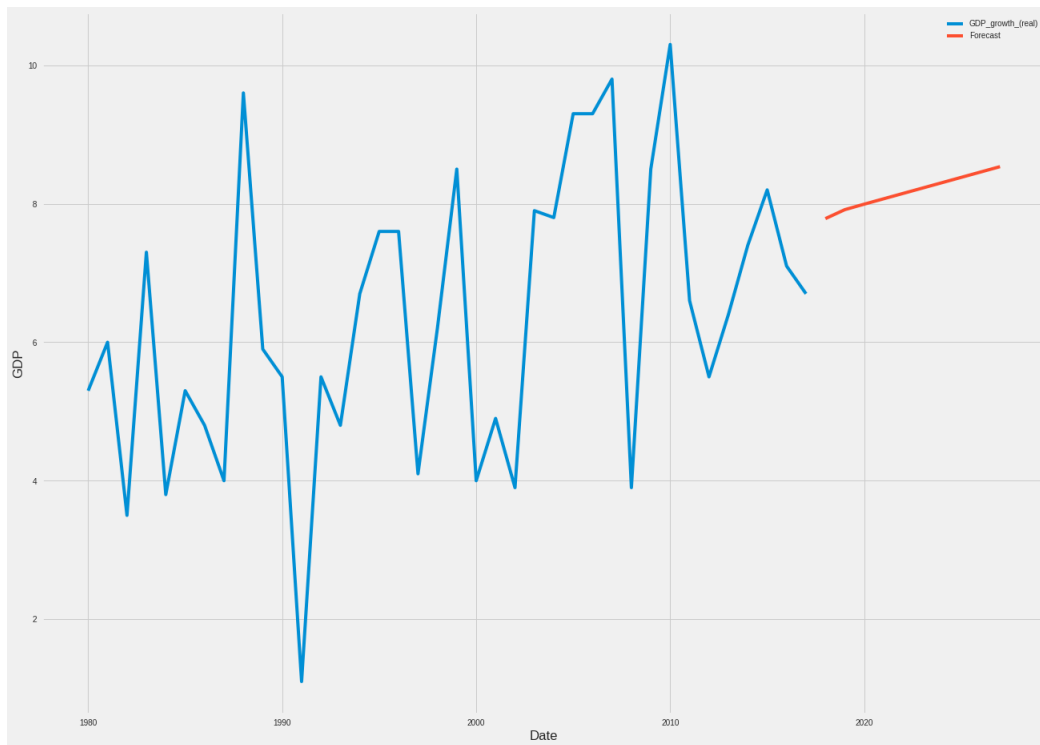


Figure 28: Time series plot of GDP

7 Conclusion

From this project we studied the following-

1. The employment curve in India.
2. Data about the tax collection in India.
3. The affect of demonetization and GST on the Indian economy.
4. Increase in the usage of digital payment modes.
5. Inflation and GDP in India.

However, From this project we can conclude that using these results we can not say for certain that which government has done a better job in elevating the economy of our country. But it can be used to initiate a healthy debate regarding politics.

8 Future Works

1. Integrating the Provident funds data in the employment dataset.
2. Including the data from the training programs launched by the present government to improve employment.
3. Study the long term effects of Demonetization and GST.

9 Dataset Sources

1. INDIA API (1960 – 2017) – World Bank
2. Tax Data – incometaxindia.gov.in
3. Digital Transactions – RBI
4. GDP – Wikipedia (web scrapping)
5. Inflation 2012 – 2018 - Bloomberg

10 References

1. <http://economictimes.indiatimes.com/news/politics-and-nation/enforcement-directorate-raids-forex-shops-making-back-date-entries/articleshow/55373554.cms>
2. <http://www.newindianexpress.com/states/andhra-pradesh/2016/nov/08/chandrababu-naidu-welcomes-demonetisation-move-1536401.html>
3. <http://www.news18.com/news/india/why-were-the-notes-scrapped-rbi-chief-and-economic-affairs-secretary-explain-1309756.html>
4. <https://www.ft.com/content/7dbe0e14-8d8a-11e7-a352-e46f43c5825d>
5. <https://timesofindia.indiatimes.com/business/india-business/gdp-grows-at-5-quarter-high-of-7-2-in-oct-dec/articleshow/63118408.cms>
6. <http://www.financialexpress.com/economy/economists-bearish-on-the-india-growth-story-predicts-sub-6-per-cent-for-fy-2017/486381>
7. <https://www.indiatoday.in/education-today/gk-current-affairs/story/demonetization-gst-slowed-down-gdp-growth-of-india-1123939-2018-01-06>
8. <http://indiatoday.intoday.in/story/parliament-live-congress-pm-modi-raincoat-dr-manmohan-singh-opposition/1/878509.html>
9. <https://www.spss-tutorials.com/pearson-correlation-coefficient>
10. <https://www.indianeconomy.net/splclassroom/what-is-tax-buoyancy-how-it-works-for-the-government>
11. <https://www.thebalance.com/what-is-inflation-how-it-s-measured-and-managed-3306170>
12. <https://www.moneycrashers.com/what-is-inflation-definition-causes-inflation-rate> tradingeconomics.com/india/gdp-growth-annual