Introduction To Data Science

Assignment – 1

A detailed comparison between two districts of Andhra Pradesh; Krishna and Visakhapatnam, on the basis of their production of rice.

By: Mayank Rao and Meghana

USN: 01FB16ECS199 and 01FB16ECS204

Section: Computer Science Section D

Introduction:

Visakhapatnam is the largest city and the Financial Capital of Andhra Pradesh. It is the most populous city in the state and one of the million-plus states of India, comprising of a population of **5,340,000** in a relatively small area of **540km^2**. The economy of the city is the tenth largest in the country, with a GDP of **$ 43.5 billion**.

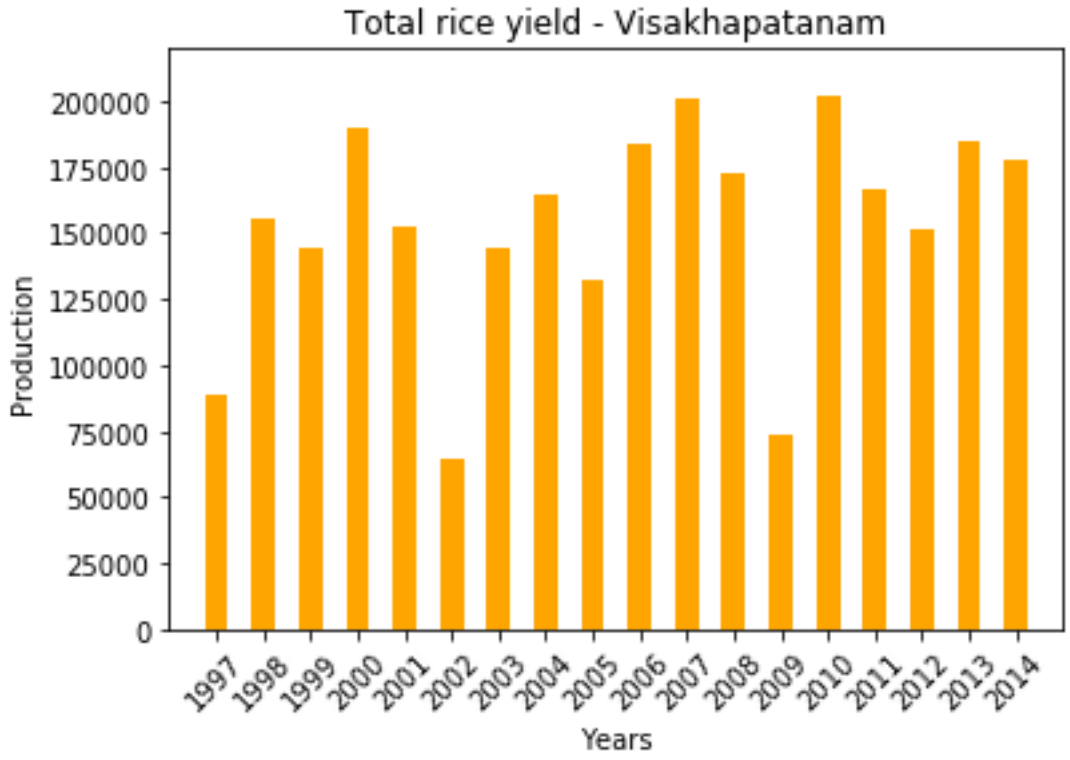
Krishna district is an administrative district in the Coastal Andhra region of the Indian state of Andhra Pradesh. It has an area of **8,727 km^2** and has a population of **45,29,009** as per **2011** census of India. Krishna district has the highest literacy rate in Andhra Pradesh.

The reason for the selection of these two particular states is due to the stark contrast between the two of them. Visakhapatnam is one of the most rapidly growing cities in the world, whereas Krishna despite having the highest literacy rate, still has a large number of poor people and is hence very slowly developing.

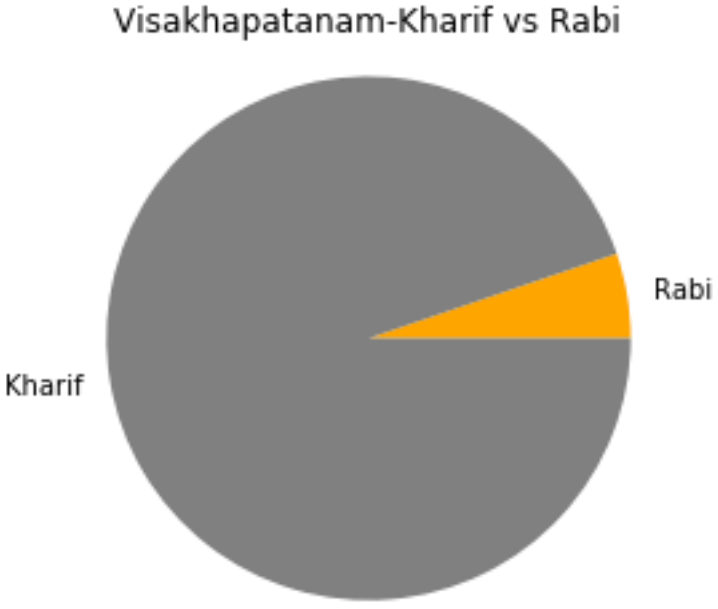
The aim of the assignment is to compare the relative yield of rice between the two districts and to analyze what factors contribute in the production of rice along with explaining the various trends the graphs indicate, as well as giving a predictive analysis of the future of these two districts with respect to rice.

Rice was the preferred selected crop due to being the most abundant crop produced in South India and a part of the staple diet of the country. Both districts produce fairly abundant amounts of the crop, and are located in ideal locations to produce it, with large flat lands fertilized by the **Krishna and Godavari rivers**, allowing production of large quantities of rice.

Analysis of Visakhapatnam:

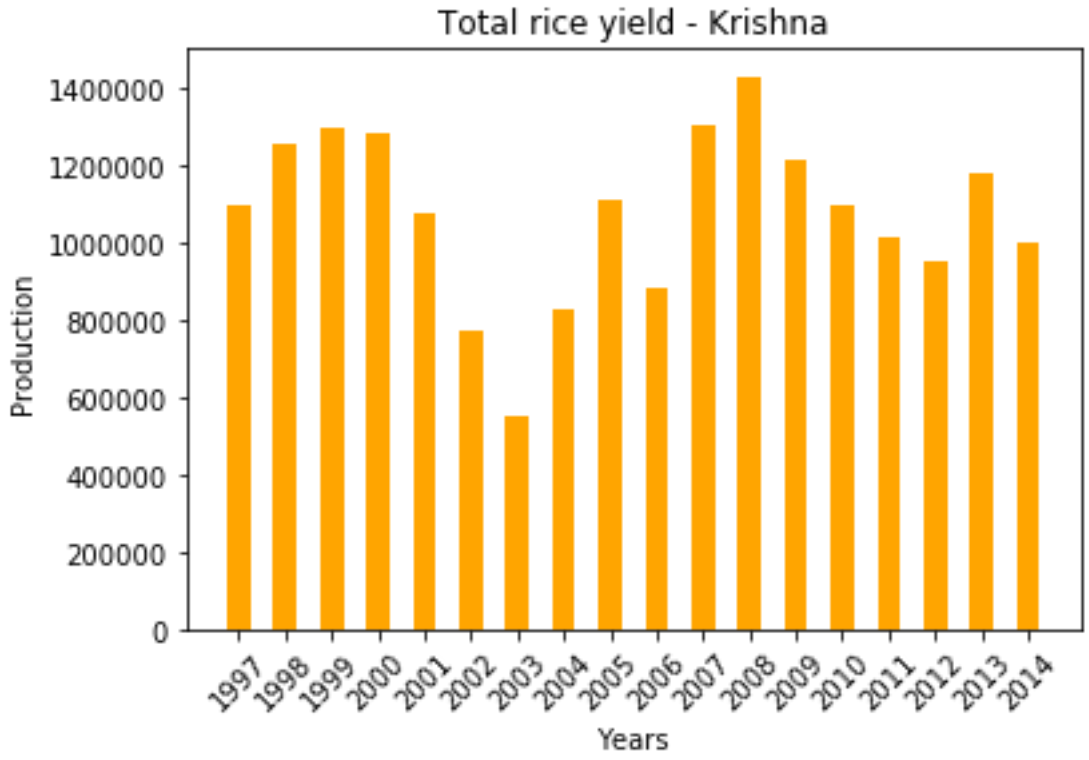


Visakhapatnam produces fairly large quantities of rice, its highest was **201,000Kg** in the year **2007** and its worst was **63,500kg** in the year **2002**.

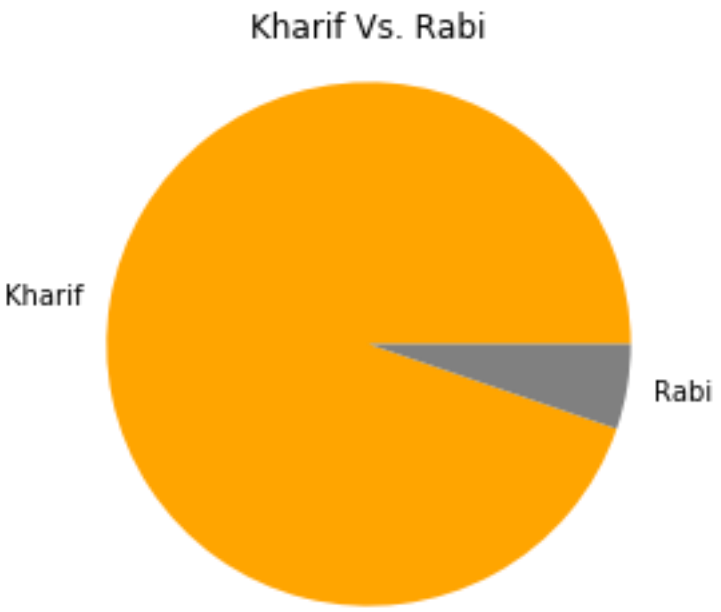


As expected, majority of the production happens during the Kharif season as rice is a Kharif crop.

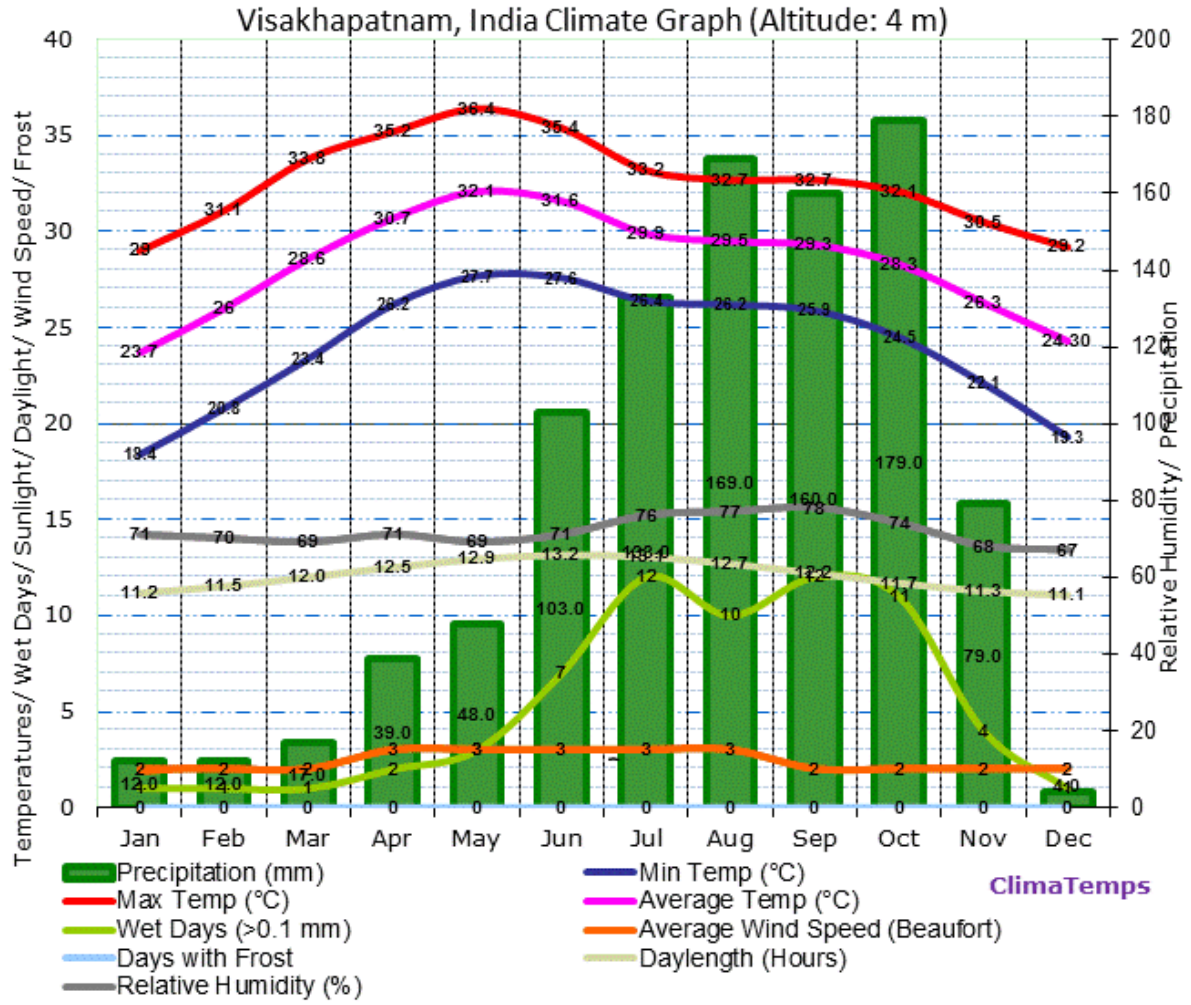
Analysis of Krishna:



Compared to Visakhapatnam, Krishna produces exponentially more rice. It produced a maximum of **1,430,000Kg** of rice in **2008** and a minimum of **545,924Kg** of rice in **2003**.


Kharif Vs. Rabi
Kharif
Rabi

Rainfall patterns and Climate of Krishna and Visakhapatnam:

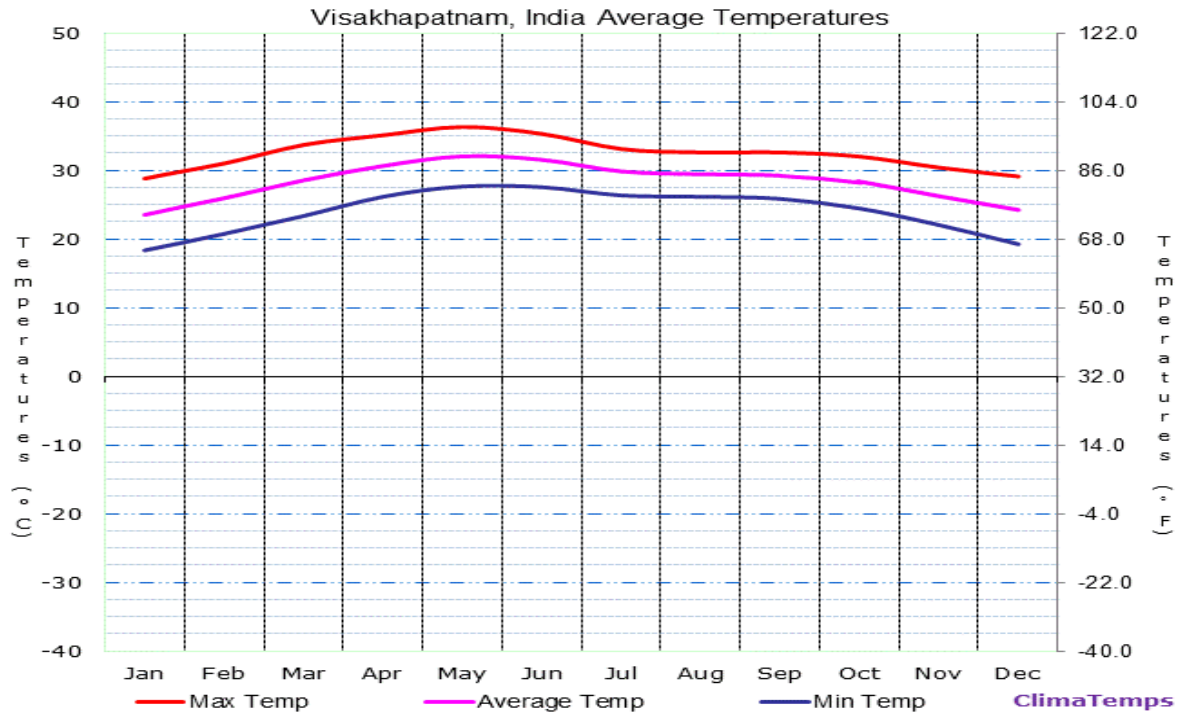


Source : ClimaTemps

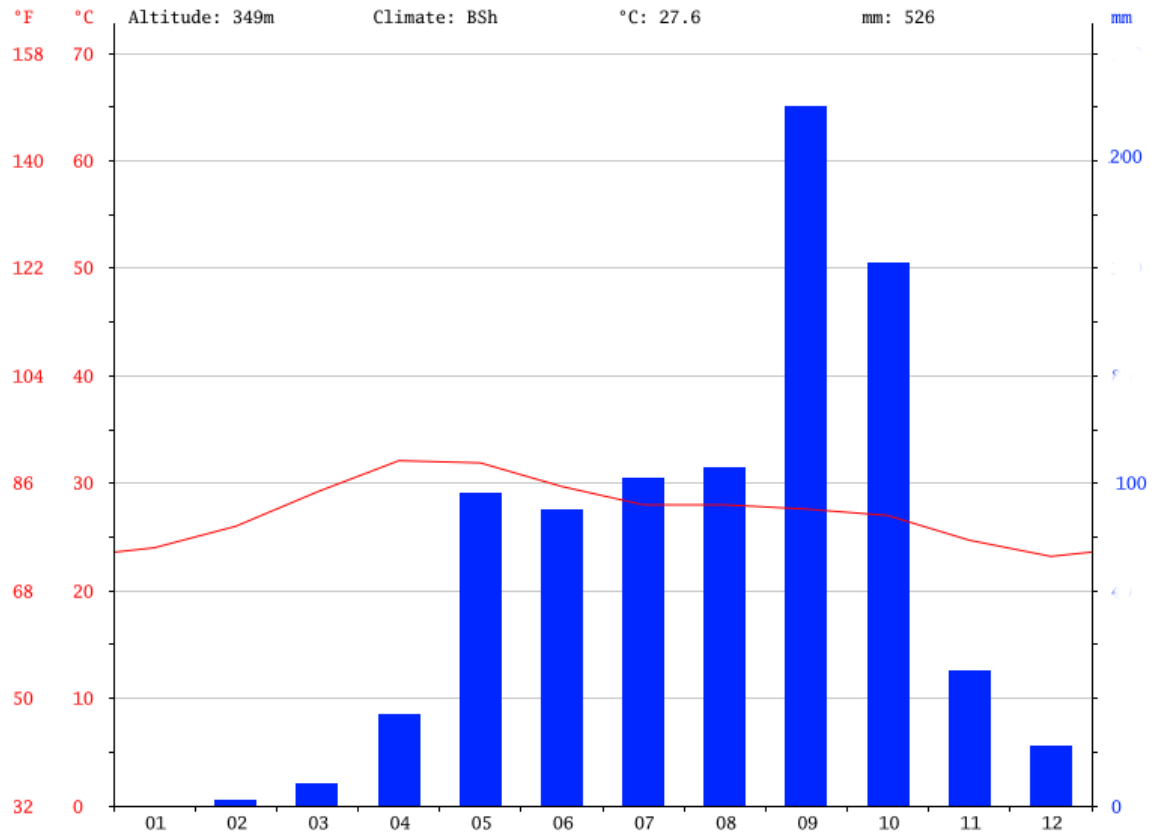
As is evident from the graph, the ideal temperature to grow rice is present only during the Summer months or Kharif season (end of May to December), combined with the ideal rainfall of around 100-200mm (source : Wikipedia)

The Graph perfectly describes this behaviour as rainfall peaks exactly as June begins and stays above it all the until November, where Kharif season is almost

at an end.



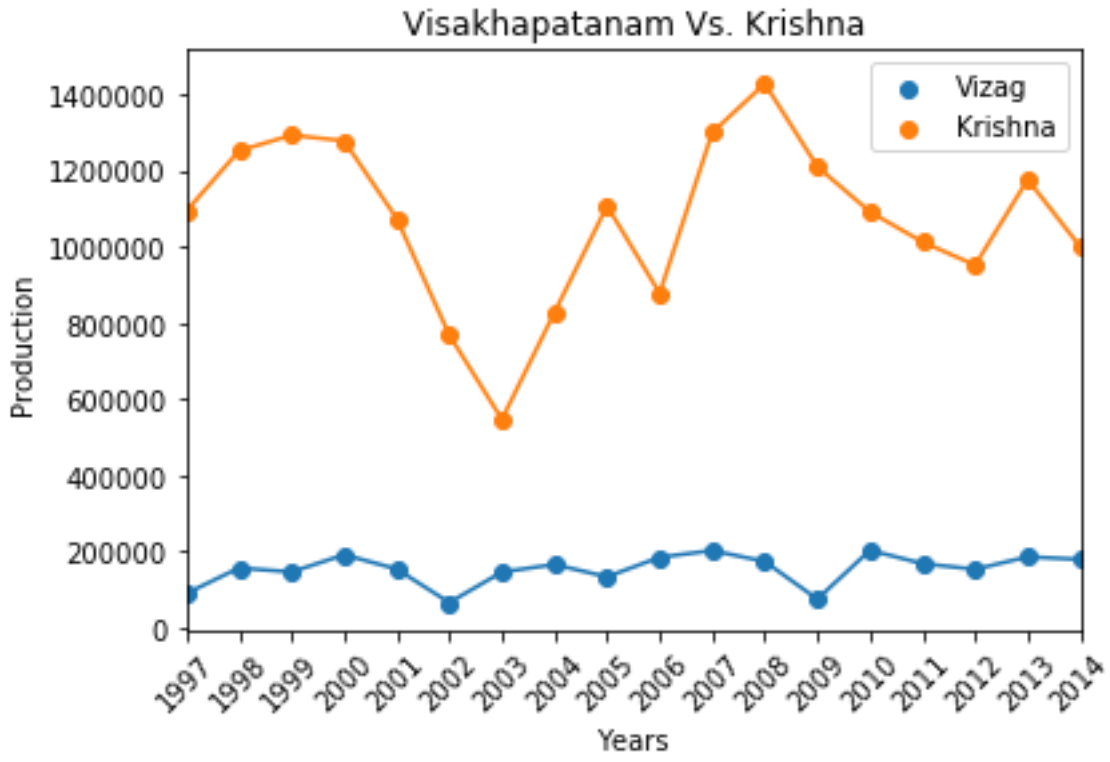
The Temperature is also nearly constant throughout the year and stays ideal for rice, from 16°C to 30°C (source : Wikipedia) and starts at around June, the Kharif season.



The Graph perfectly captures the nature of growth of rice in Krishna district as well as the drops in production during 2003-2006, during these months the rainfall was subpar and completely inefficient to grow rice.

Comparison between Visakhapatnam and Krishna:

We will now compare Krishna and Visakhapatnam in terms of their rice production



It is evident from the graph that Krishna far outstrips Visakhapatnam in terms of production of rice.

The graph also tells us a few things:

1) Visakhapatnam produces consistent amounts of rice every year. Even at their worst year, they produced almost the same amount of rice, whereas Krishna has very erratic production of rice, most likely due to their fluctuating rainfall patterns.

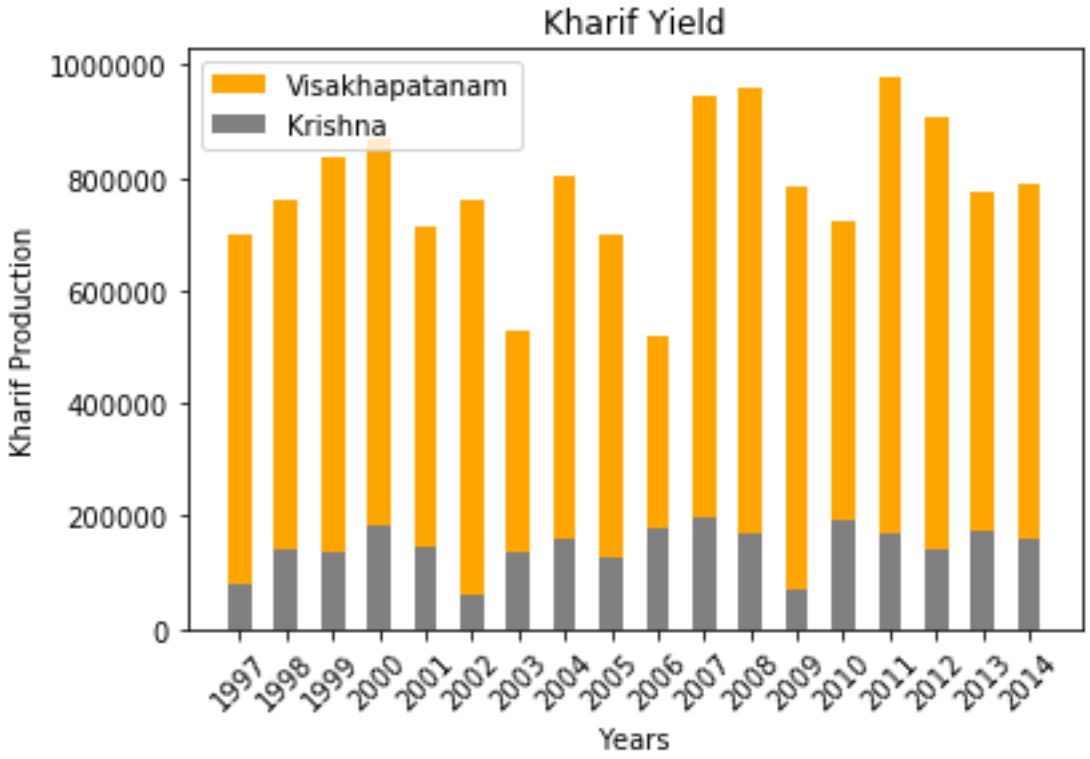
   On a good year, with adequate rainfall, Krishna produces excellent quantities of rice, however due to the rainfall fluctuations, it can cause either parched lands or submerge the crop. Rice being a primarily water grown crop is heavily dependent on correct water quantity.

2) **2002-2003** was a bad year for growth of rice, attributed to the extremely cold climate prevalent during the span of those years. Rainfall was short and unsteady and hence production was difficult.

3) Another noticeable change was that during **2009**, Visakhapatnam had very rainfall as less as 750cm, whereas Krishna had abundant rainfall over that period, contributing to the different amounts of rice produced.
s
4) Rice is not as necessary a food in Visakhapatnam , nor is it very useful in terms of their income. Being a developing district, it focuses more on secondary and tertiary sector rather than farming, which is a primary sector.

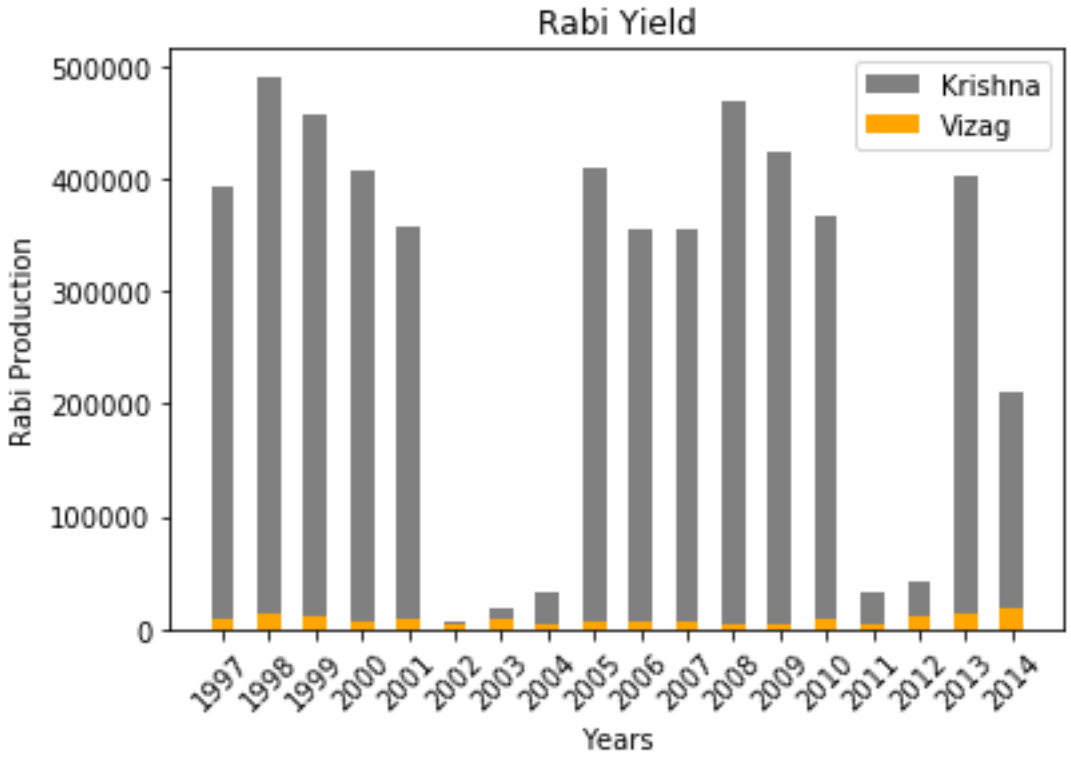
5) Visakhapatnam is a port town, and as such is more dependent on its fisheries rather than farming. It is a major port for transportation of goods to and from the rest of the world, behind only Mumbai.

6) Krishna however has a relatively unfavorable coastline compared to its neighbor Visakhapatnam and at a massive size of **8727km^2** , a large amount of its area is under cultivation and is hence suitable for producing large amounts of rice.



This graph paints a more accurate picture of the quantity of rice produced in both districts.

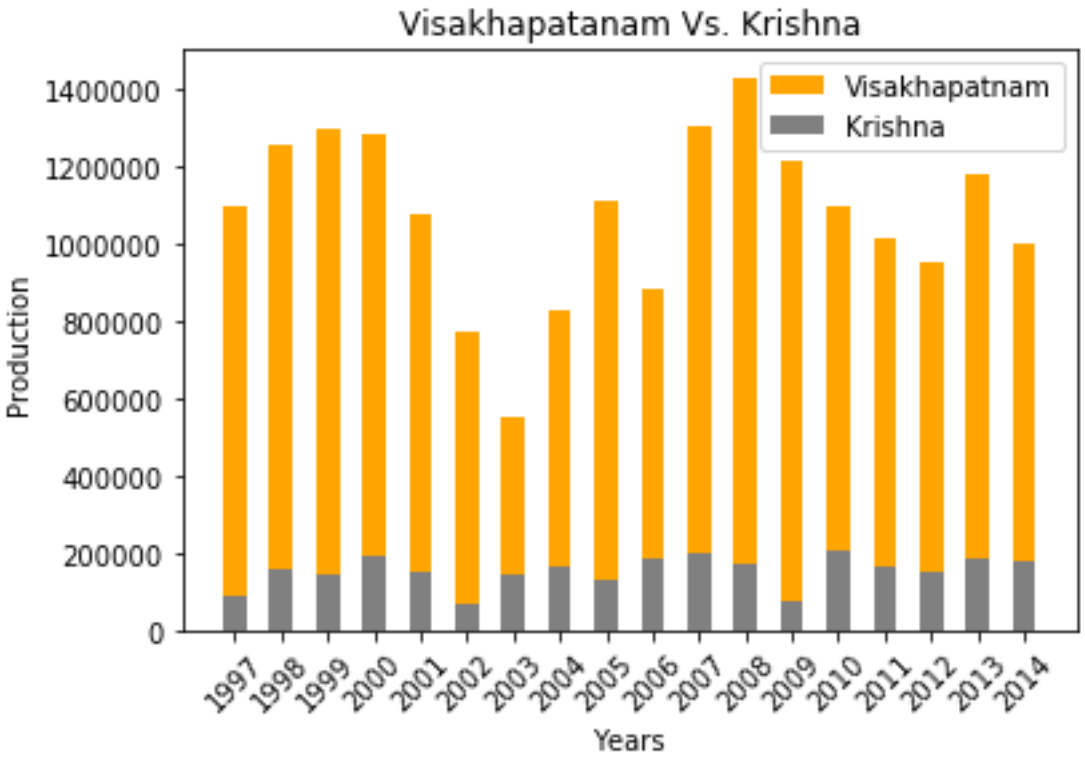
As is evident from the graph, Krishna completely dwarfs Visakhapatnam in production of rice. Not even one year has Visakhapatnam produced even half of what Krishna produces during the Kharif season.

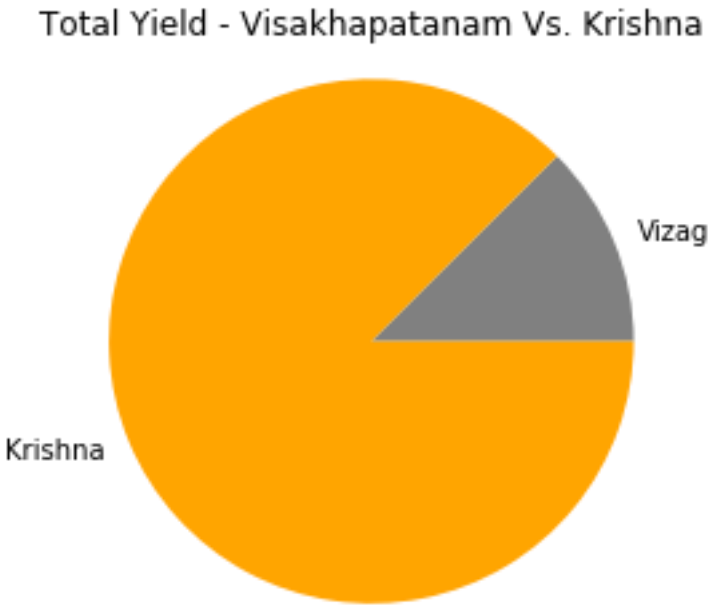


The Rabi yield however speaks a different story. Rice being majorly a Kharif crop obviously will not grow as well in Rabi season.

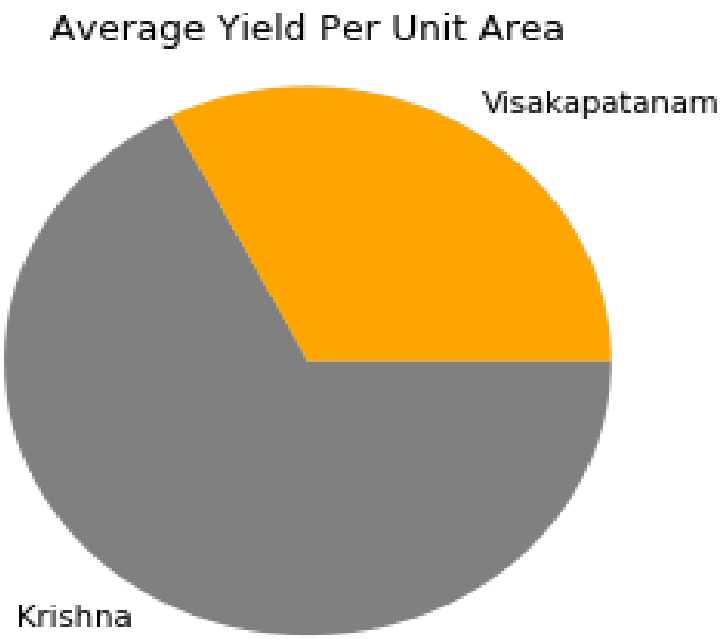
It therefore makes sense that Visakhapatnam produces even less Rabi Rice due to the fact rice is not a hugely profiting industry, and therefore it makes no sense for them to grow rice in the off season when instead fishing would be a better and profitable prospect.

But, during **2002,2003 and 2004**, Andhra Pradesh experienced the heaviest and coldest winter in a long time. Rice being an extremely sensitive crop could not take the cold temperatures and the crops perished. This also marks the first time Visakhapatnam nearly overtook Krishna briefly, albeit due to the winter not affecting Visakhapatnam as much due to moderating climate from the **Bay of Bengal**.





Nevertheless, Krishna overall still completely dominates Visakhapatnam in terms of production of rice. Even with the marginal extra production of rice during **2002,2003** and **2004**, Visakhapatnam could not compete with the sheer quantity of rice produced by Krishna.



Overall, Krishna being a still developing district focuses more on rice as a means of survival and farming in general. They've been given the tools to make farming a useful resource. The rivers, the weather, the soil and a huge profit in exporting rice through the gateway Visakhapatnam.

This trend however does not seem to be lasting. Krishna has by far the highest Literacy rate of any district in Andhra Pradesh at a whopping **73.74%.** As you can see, the production of rice is slowly reducing as the IT sector takes over due to various improvements in Educational Facilities and recently funded Government support, Krishna's production of rice may very well reduce over the coming years.

However, farming still is and always will be an important source of income to the poorer section of the society, and yet still a very profitable one.

Hypothesis Test for given data:

From the previous section we concluded that Krishna produces much higher quantity of rice as compared to Visakhapatnam. For the testing of the given sample, the **Null Hypothesis** and the **Alternate Hypothesis** must be determined.

We select the hypotheses on the basis of our conclusion, i.e. since we concluded that **Krishna produces more rice, this is our Alternate Hypothesis**.

Therefore, our Null Hypothesis should indicate that contrary to the Alternate Hypothesis, **Krishna produces lesser rice.**

Putting it in a mathematical form:

$$H_o : \mu_k - \mu_v \leq 0$$

$$H_a : \mu_k - \mu_v > 0$$

μ_k : Mean production of Krishna μ_v : Mean production of Visakhapatnam

For the analysis of the given hypotheses, we have two plausible methods.

- Z – Score method
- Student t distribution

Student t distribution requires that the sample size be less than **30 and must also be a Normal Distribution**. Our sample size is 17 which is lesser than 30 and

Is also a Normal Distribution hence we are using Student t distribution.

Within this Student t distribution, we are applying Rejection Region approach and P – value approach.

Analysis of Student t distribution using Rejection Region approach:

Significance Level (α) = 0.05

Difference of mean production of Krishna and Visakhapatnam = 9,17,877.438

Critical Value = 89,151.79

t – score corresponding to the Critical Value = 1.7396

t – score for the difference of sample mean = 17.3952

From the above statements it is clear that the Critical Value is far less than the difference of the mean production.

We can hence reject the Null hypothesis with a 5% chance of having Type 1 Error or False Positive.

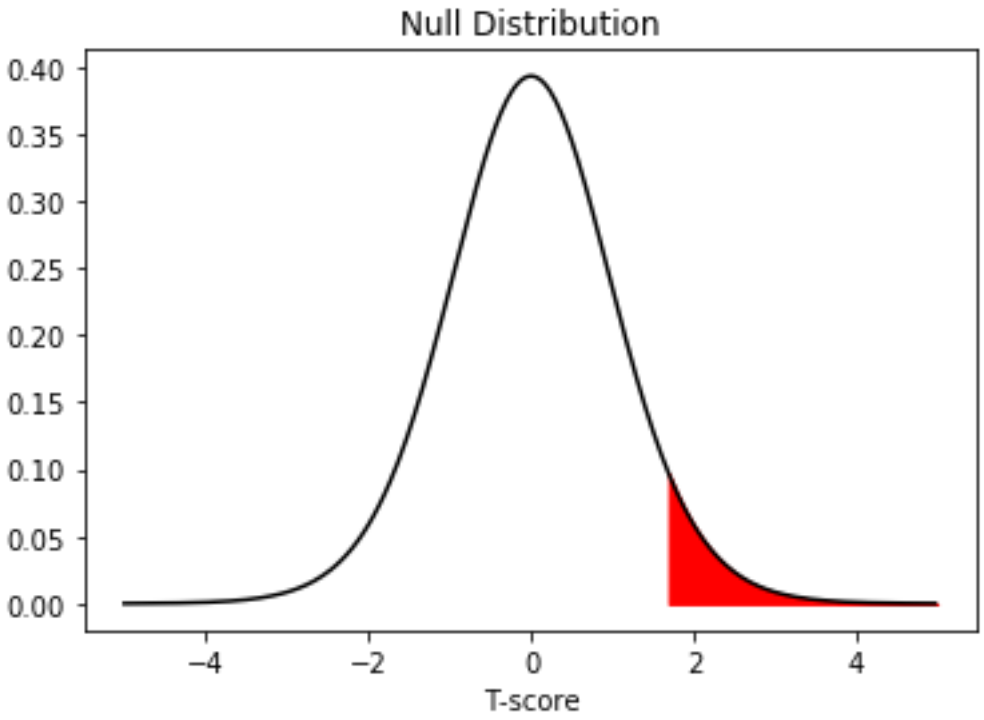
Analysis of Student t distribution using P - value approach:

For us to reject the Null Hypothesis the P - value corresponding to our difference of means should be less than 0.05 or 5%.

The P value determined from the calculations is 1.4417×10^{-12} . This value is so small we can consider it to be non – existent or 0.

It is clear from the calculated P value that the Null Hypothesis can be rejected.

Graph of the Rejection Region:



Conclusion from the Hypothesis Test:

We have rejected the Null Hypothesis, and hence we can conclude that Krishna produces more rice than Visakhapatnam, affirming the fact that the Alternative Hypothesis is true.

It is also worth noting that in both of the performed tests, the calculated value far exceeds the required threshold value, and hence we can without a doubt confirm that the Null Hypothesis is to be rejected.