

**WORLD HAPPINESS REPORT (Data Analysis)**

**INTERNSHIP REPORT**

**Quarter IV (Year 1)**

***Submitted by***

**Nivedhidha I E0220003**

***In partial fulfilment for the award of the degree of***

**BACHELOR OF TECHNOLOGY**

**in**

**COMPUTER SCIENCE AND ENGINEERING**

**(Cyber Security & Internet of Things)**

**Sri Ramachandra Engineering and Technology,**

**Sri Ramachandra Institute of Higher Education and Research,**

**Porur, Chennai -600116**

**JULY, 2021**



**BONAFIDE CERTIFICATE**

Certified that this project report **“WORLD HAPPINESS REPORT (Data Analysis)”** is the bonafide work of **Nivedhidha I Reg No. E0220003** who carried out the internship work under my supervision.

**Signature of Faculty Mentor Signature of Vice-Principal**

|  |  |
| --- | --- |
| **Prof. Siva Shankar G**  Lecturer  Sri Ramachandra Engineering and Technology  Porur, Chennai-600116 | **Prof. M. Prema**  Vice-Principal  Sri Ramachandra Engineering and Technology  Porur, Chennai-600116 |

**Evaluation Date:**

**Table of Contents**

**Title** **Page**

1. Problem Statement 5
   1. Problem Statement 5
   2. Solution 5
2. Domain Introduction 5
   1. Data Analysis 5
   2. Data Visualization 6
   3. Web Development 6
3. Objective 6
4. Survey 7
   1. World Happiness Report 7
   2. The Dataset 7
5. Workflow 8
   1. Data Analysis 8
   2. Web Development 9
   3. Website Layout 10
6. Technology Used 11
   1. Data Analytics 11
   2. Web Development 12
7. Implementation 13
   1. Data Visualization - Tableau 13
   2. Data Analysis and Machine Learning - Python 20
   3. Website Development 20
8. Output 21
   1. Tableau 21
   2. Python 38
   3. Website 47
9. Sample Code 51
10. Conclusion 51
11. Target Audience 52
12. Future Scope 52
13. References 53

**ACKNOWLEDGEMENT**

I express my sincere gratitude to our Chancellor, Vice-Chancellor and our sincere gratitude to our Provost **Dr. V. Raju** and our Vice-Principal **Prof. M. Prema** for their support and for providing the required facilities for carrying out this study.

I wish to thank my faculty mentor, **Prof. Siva Shankar G**, Department of Computer Science and Engineering, Sri Ramachandra Engineering and Technology for extending help and encouragement throughout the project. Without his/her continuous guidance and persistent help, this project would not have been a success for me.

I am grateful to the Department of Computer Science and Engineering, Sri Ramachandra Engineering and Technology, our beloved parents and friends for extending the support, who helped us to overcome obstacles in the study.

1. **PROBLEM STATEMENT**
   1. **Problem**
      1. A country’s growth depends on its people’s happiness.
      2. It is important for the government to know about how we can help in improving the country’s citizen’s happiness.
   2. **Solution**
      1. We have addressed the important questions and done an analysis to give an insight into how the happiness of a country can be improved.
2. **DOMAIN INTRODUCTION**

Over time, happiness has been reducing around us, due to various factors mainly being stress. We wanted to know about the world situation regarding this and suggest the major factors in which countries could focus on, to make them a happy one. Hence, we decided to work with this dataset.

1. **Data Analytics:**

Data analytics is the process of examining data sets to analyse, find trends and conclude the insights that it contains. Increasingly data analytics is used with the aid of specialized systems and software.

Data Analytics Technologies and techniques are widely used in commercial industries to enable organizations to make more informed business decisions. It is also used by scientists and researchers to verify or disprove scientific models, theories and hypotheses.

1. **Data Visualization:**

Data visualization is the graphical representation of information and data by using visual elements like charts, graphs and maps. Data visualization tools provide a more approachable way to see and understand trends, outliers and patterns found in the data.

In the world of Big Data, data visualization tools are essential to analyze massive amounts of information and make data-driven decisions.

## **Web Development:**

Web development refers to the building, creating and maintaining of websites. It includes aspects such as web design, web hosting, web programming, and database management.

1. **OBJECTIVE**

The Objective of the Project was to learn the basic concepts of Data Analysis and Frontend Website Development. Then to perform Data Visualization on our Dataset by forming questions and using Tableau software to create the required graphs and plots. Also to create storyboards for providing a proper understanding of the viewers.

Then create a website using a software of our choice, learn about AWS(Amazon Web Services) and do hosting the website in AWS so that anyone can access it.

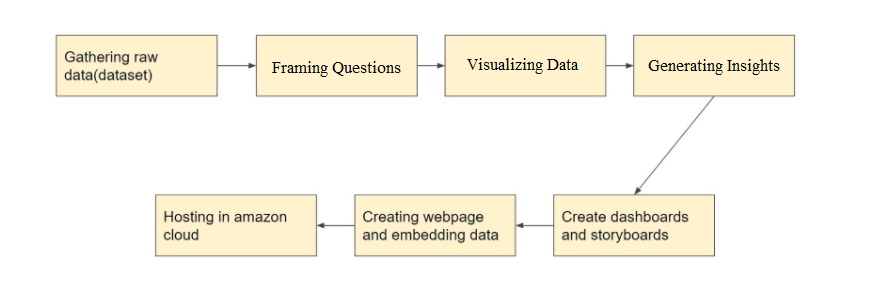
1. **SURVEY**

**World Happiness Report**

1. The Index reflects a new worldwide demand for more attention to happiness as a criterion for government policy.
2. It is also an eye-opener for those who think that the pursuit of a healthy GDP number alone is sufficient to mark the happiness of a nation.
3. For the making of the World Happiness Report, nationally representative samples of respondents are asked to rate their own current lives on a 0 to 10 scale.
4. The data used to rank countries in each report are drawn from the Gallup World Poll, as well as other sources such as the World Values Survey, in some of the reports.
5. Experts in fields including economics, psychology, survey analysis, and national statistics, describe how measurements of well-being can be used effectively to assess the progress of nations.

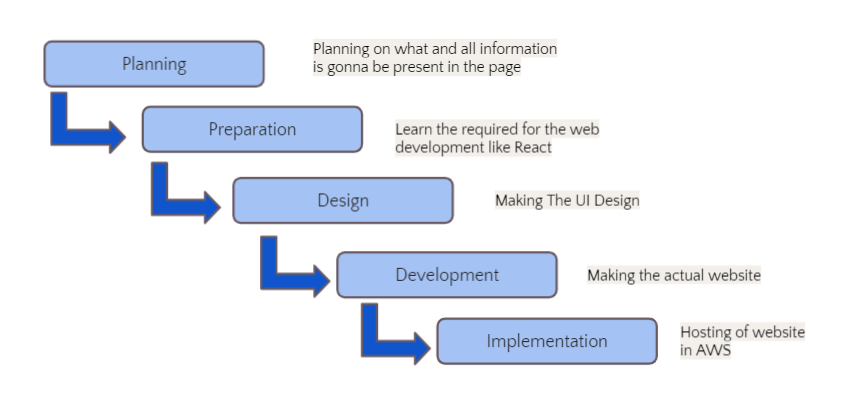
**The Dataset**

1. The dataset provides the happiness rank for nearly 150 countries for the 5 years.
2. The rank is calculated based on various parameters like GDP and the rank is given by people of the sample for different questions asked.
3. The questions were mainly based on what rank they would give on the factors for a contribution towards their happiness.
4. The factors were:
   1. Social Support
   2. Economy(GDP per capita)
   3. Freedom
   4. Trust Towards Government
   5. Health (Life Expectancy)
   6. Generosity
5. **WORKFLOW**
   1. **Data Analysis**

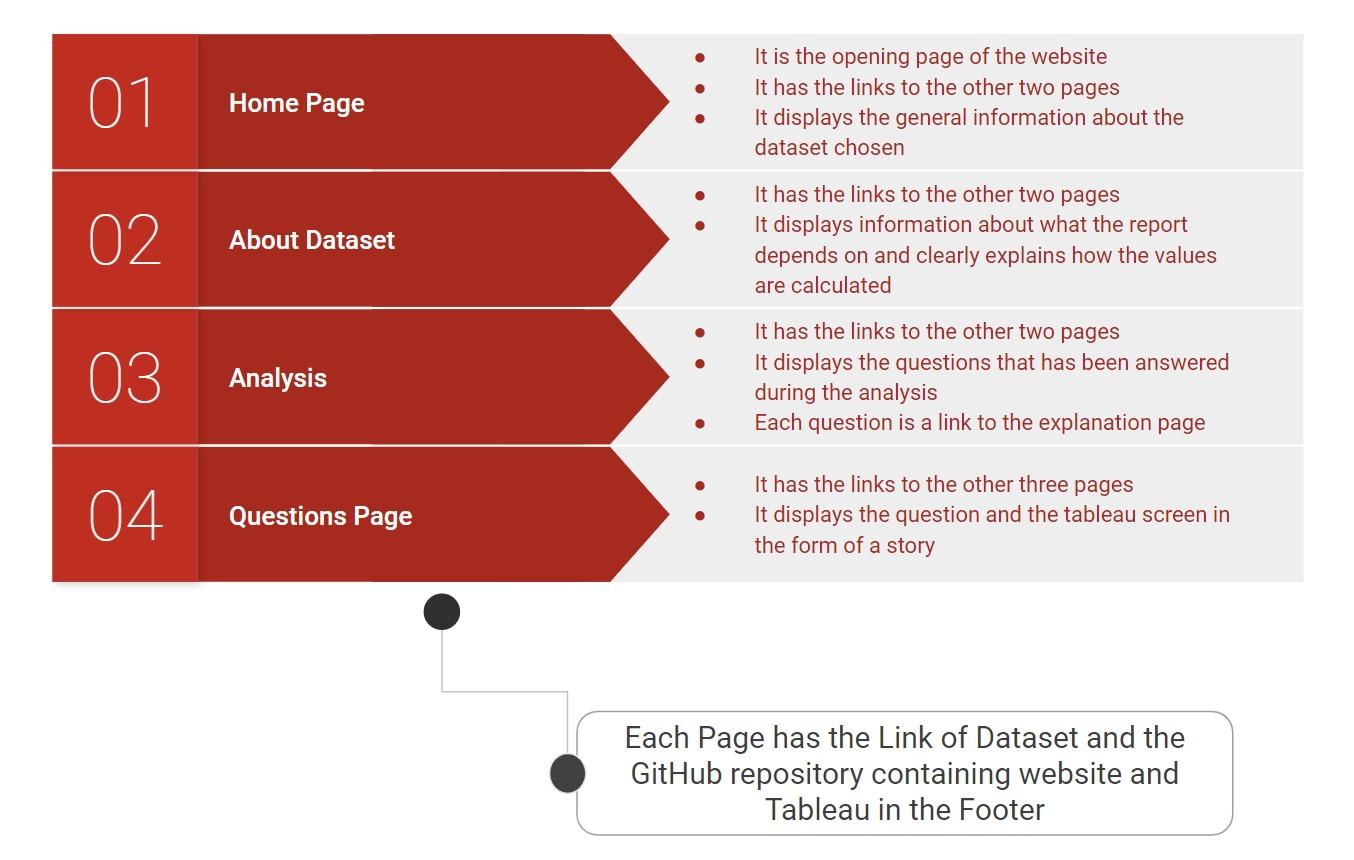


1. Gathering Raw Data - First we should find a proper dataset which is dependable and should do the required cleansing to make it ready for analysing.
2. Framing Questions - Next we should frame questions from the dataset which we want to answer through our analysis.
3. Visualizing Data - Next we should visualize the dataset by creating various graphs.
4. Generating Insights - Next we should generate insights from the created graphs and find out the answers for questions formed before.
5. Creating Dashboards and Storyboards - Next we should create dashboards or storyboards as per requirement for creating a storyline to explain how we got our insight to others.
6. Creating Website and Embedding Data - Next we should create a website using an IDE to make the analysis available for everyone.
7. Hosting in AWS - Finally we should host the website in AWS to make the website available for anyone interested in the analysis.

* 1. **Web Development**



* + 1. Planning - First we should plan on what and all information is going to be present on the website.
    2. Preparation - Then we should start preparing by learning the required IDE like Angular and React JS.
    3. Design - Next we should make the UI design of the website.
    4. Development - Then we should make the website using HTML, CSS, Angular and React.
    5. Implementation - In the end we should host the website so that it becomes available for anyone interested in the movie.
  1. **Website Layout**



* There are totally 4 types of web pages present in the website.
* Each page has the links of GitHub repository and Kaggle dataset.
* Home page contains general information about the world happiness report.
* The About Dataset page has the details about each factor of the dataset.
* Analysis page has the questions which have links to each question of the analysis page.
* There are totally 8 Question pages containing a tableau screen in the form of a story.

1. **TECHNOLOGY USED**

## **Data Analytics**

1. Tableau

Tableau is a visual analytics platform transforming the way we use data to solve problems - empowering people and organizations to make the most of their data. It is one of the fastest evolving data visualization tools. It is very fast to deploy, easy to learn and very intuitive to use for a customer.

### MySQL

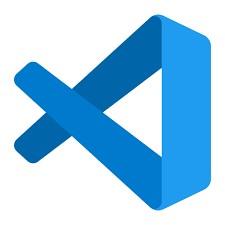
MySQL can cost-effectively help you deliver high performance, scalable database applications. It is a relational DBMS based on SQL. It is open-source and free software under the GNU license. It is supported by Oracle Company.

### JUPYTER NOTEBOOK

Jupyter is a free, open-source, interactive web tool known as a computational notebook, which researchers can use to combine software code, computational output, explanatory text and multimedia resources in a single document. It can be used for all sorts of data science tasks including data cleaning and transformation, numerical simulation, exploratory data analysis, data visualization, statistical modelling, machine learning, deep learning, and much more.

## **Web Development**

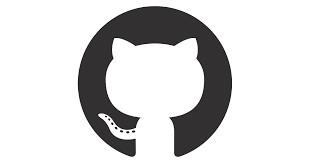
### VISUAL STUDIO CODE

Visual Studio Code is a source code editor that can be used with a variety of programming languages, including Java, JavaScript, Go, Node. js, Python and C++. It is based on the Electron framework, which is used to develop NodeJS Web applications that run on the Blink layout engine.

### REACT JS

### ReactJS is an open-source JavaScript library that is used for building user interfaces specifically for single-page applications. It's used for handling the view layer for web and mobile apps. React also allows us to create reusable UI components.

### GITHUB



GitHub is a code hosting platform for version control and collaboration. It lets you and others work together on projects from anywhere.

#### 

#### AWS (AMAZON WEB SERVICES)

Amazon Web Services (AWS) is a secure cloud services platform, offering computing power, database storage, content delivery and other functionality to help businesses scale and grow. Deliver static and dynamic files quickly around the world using a Content Delivery Network (CDN).

1. **IMPLEMENTATION**

## **Data Visualization - Tableau**

1. What is the overall situation in the world regarding happiness?

* First, we plotted a graph that shows the regions and the happiness scores of the countries. We analyzed that darker shaded countries are mostly a part of North America and ANZ and Western Europe which implies that they have higher happiness scores in comparison to the other regions.
* Next, we plotted a treemap and from that, we got to know that happiness score generally was affected by two major factors, ie., Social Support and Economy(GDP per capita) throughout the world. And Generosity and Trust were the bottom 2 factors accepting the Happiness Score
* Then we compared the top 2 factors with happiness scores with respect to regions. From that graph, we came to a conclusion that Economy and GDP per capita equally contributed to the Happiness Score.
* So, this is the overall situation of the world in regard to happiness and its score.

1. Which countries are better positioned in each of these aspects?

* For this analysis, we plotted a bar graph of the Top 10 countries with respect to all the 6 factors on the basis of their Happiness Scores.
* Economy(GDP Per Capita): Through this graph, we can view that Qatar has the highest score of 1.7 followed by Luxembourg and Singapore. And we came to a conclusion that, due to Qatar's high GDP, it is one of the richest countries and also ranks a high position in the world happiness report.
* Freedom: Through this graph, we can view that Uzbekistan has the highest score but has one of the least happiness scores and ranks. On the other hand, Norway, the 2nd highest country in regard to freedom ranks the highest position in the world happiness report. This meaning Freedom is not one of the main factors for happiness.
* Generosity: Through this graph, we can view that Myanmar has the highest score of 0.72. Though the score is less when compared to the other factors, the score of the 2nd highest country is 0.5 which is very less. So on average Myanmar has the highest score in terms of generosity according to people's opinion.
* Health(Life Expectancy): Through this graph, we can see that Singapore and Hong Kong have the highest scores amongst the other countries. This implies that these countries have very good health infrastructure and higher life expectancy.
* Social Support: Through this graph, we can view that Iceland has the highest score of social support in terms of peoples opinion which implies that people in Iceland have enhanced quality of life in adverse life situations due to the support from their family and friends.
* Trust(Government Corruption): Through this graph, we can view that Rwanda and Singapore have the highest score. Another important point to notice is that though these countries have the highest position, the score is less than 0.5 which is very less when compared to all the other factors. So we can understand that on average the trust in government is low throughout the world due to activities like corruption.

1. Finding the reason why happiness rank changed over the years in different regions?

* In this analysis, we looked upon the countries that we’re improving and deteriorating over the years.
* The first table displays the list of countries that are improving and deteriorating throughout all the five years (2015-2019) on the basis of the happiness rank. Then we plotted a bubble chart for both the variations and we observed that social support and economic GDP makes some changes in the improving and deteriorating countries compared to other aspects. The improving and deteriorating were grouped using the create group function available in Tableau.
* Then we plotted a dual-axis line graph to compare on the basis of the 3 main factors: Social Support, Economy (GDP per capita) and Generosity.

1. In what ways, Generosity and Health play a role in impacting the world happiness score?

* This Analysis was made to find whether there is a relationship found between Generosity and Health. Generally, studies say that people who are generous are said to be happier and eventually have a higher life expectancy. So we checked whether that was true.
* First, we plotted a histogram for Health(Life Expectancy) and the Count of Countries given the particular score. We were able to clearly see that most of the countries have given their score between 0.50 to 0.90, we can say that the Health(Life Expectancy)score on an average is high in most of the countries as per people's opinion.
* Then we did the same with respect to Generosity. In this graph we were able to see that most of the countries have given their score between 0.05 to 0.30, we can say that the generosity score on an average is low in most of the countries as per people's opinion.
* Now to compare both the factors side by side we plotted a box-plot. From these boxplots, we can see that the score given by people for Generosity is really less than Health. North America and ANZ are at the top in both the factors, whereas Sub-Saharan Africa is in the least in Health(Life Expectancy). Central and East Europe is the bottom-most region in terms of Generosity. And in a few years like 2015 and 2019 Sub-Saharan Africa is outside the boxplot meaning they did not follow the trend and are called outliers. One more important thing to notice is that The score of health is much higher than generosity.
* So now to check if a correlation exists we plotted a linear regression trend line over a scatter plot. Through this graph, we can understand that there is some kind of correlation found between health and generosity as most of the scatter points lie near the trendline. The trendline shows an inclination which implies that they have a positive correlation. And, since the inclination angle is less, it follows a weak positive correlation trend. This means that there is a correlation between both of the factors.
* So, we can come to a conclusion that Generosity and Health are co-related with respect to happiness score.

1. Does a country's GDP affect the happiness rate?

* In this analysis, we checked whether the Economy of the country affects the happiness score/rank of the same.
* First, we plotted a graph that shows a comparison between Happiness score and Economy(GDP per capita). A linear trendline has been plotted on the scatter plot, in which the line shows a large inclination and points are closer to the line. This shows that there is a positive correlation between both factors.
* Then we plotted a treemap with respect to the GDP and Happiness Score in regard to each region. From the treemap, we can analyze that North America and ANZ and Western Europe have the highest score. On the other hand, South Asia and Sub-Saharan Africa have the least score in both factors. So then we created a group on the above basis.
* The scatter plot splits all the countries into 3 categories namely top, bottom and middle with respect to region. If we see the countries in the top group we can see that there is a linear trend line that has a large inclination towards the top showing a strong positive correlation, meaning that the economy has played a major role towards the happiness score in the list of top countries. On the other hand, the countries in the least group also show an inclination towards the top but the angle is too low which implies a weak positive correlation meaning a low economy is not the only reason for their least happiness score. And the mid-range countries have a mediocre inclination towards the top. So, we can understand from the graph that the Economy(GDP per capita) is one of the main reasons for the happiness score but the impact is different for each region.
* In the next map, we can visualize that the top countries are following the same pattern in all the years, by seeing their colour codes. On the other hand, the bottom countries are continuously varying. Hence we can conclude that the countries at the top are more stable than the ones at the bottom.

1. What is the trend followed by India’s happiness score in the past 5 years?

* As we live in India we wanted to analyze India’s situation with regard to happiness
* First, we plotted 2 line graphs, one with respect to happiness rand and the other with respect to the happiness score of India. From the first graph, we were able to see that the happiness rank and score of India has been decreasing every year, for the past 5 years. And in the second graph, we can see that the rank is given based on which position India is present in the world regarding happiness, whereas the score would give a much more clear picture regarding the changes that would have happened in India throughout the years.
* In the next graph, we can see the factors that would have affected India to have such a low happiness score. The graph clearly shows Economy and Social Support were the 2 main factors contributing to India’s score.
* Then we compared all the factors throughout all the years.
* Trust(Government Corruption): This graph shows the change in
* Trust(government corruption) yearwise. Trust has continuously increased from 2015 to 2018 and had a sudden downfall in the year 2019. We can see that in the year 2019, one of the main reasons for the decrease in happiness score would be the people's trust towards the government. Studies say that the Central Government was accused of destroying democratic institutions and processes after the election was conducted.
* Freedom: This graph shows the change in Freedom yearwise. Freedom has continuously increased from 2015 to 2018 and had a sudden downfall in the year 2019. We can see that in the year 2019, one of the main reasons for the decrease in happiness score would be the People's opinion on Freedom.
* Social Support: This graph shows the change in Social Support yearwise. Social Support score has decreased in the year 2016 and has had an increase for the next 3 years. We can see that in the year 2016, one of the main reasons for the decrease in happiness score would be the Social Support score.
* Generosity: This graph shows the change in Generosity yearwise. Generosity score has continuously decreased till the year 2018 and had an increase in the year 2019. We can see that in the first 4 years, one of the main reasons for the decrease in happiness score would be the Generosity score.
* Health(Life Expectancy): This graph shows the change in Health(Life Expectancy) yearwise. Health has continuously improved from 2015 to 2017 and had a sudden downfall in the year 2018 and again raised in the year 2019. We can see that in the year 2016, one of the main reasons for the decrease in happiness score would be due to Health. The lower level of Health can be attributed to the low level of public investments in preventive health facilities such as sanitation and waste management, as well as in medical care facilities such as primary health centres and health professionals.
* Economy(GDP per Capita): This graph shows the change in
* Economy(GDP per Capita) yearwise. The GDP has continuously increased from 2015 to 2017 and had a sudden downfall in the year 2018 and again raised in the year 2019. We can see that in the year 2018, one of the main reasons for the decrease in happiness score would be the GDP value. The lower GDP growth figures are attributed to weak domestic consumption so, the Indian economy reported a continuous decline in GDP growth rate over the year 2018.

1. Which country had more trust in the government with respect to freedom?

* This analysis was visualized to know which country has more trust in government with respect to freedom.
* First, we plotted a histogram for Trust and the Count of Countries given the particular score. We were able to clearly see that most of the countries have given their score between 0.02 to 0.14, we can say that the Trust(Government Corruption) score on average is low in most of the countries as per people's opinion. And we did the same even for freedom where we can view that most of the countries have given their score between 0.30 to 0.60, we can say that the Freedom score on an average is high in most of the countries as per people's opinion.
* Then we plotted dual axes between line and bar graph. From which we can conclude that the factors didn’t have any correlation with each other. Also, according to people's opinion, the sense of Freedom is higher when compared to the trust towards the government. Another point to note is Trust towards the government has negligible change, whereas Freedom is varying highly. And the region of North America and ANZ has the highest score in both the factors, hence has the highest happiness score. There is a big gap between the factors for all the regions.
* And then found that Denmark has the highest trust towards the government.

1. An analysis to find which are the factors causing fewer happiness scores in a few regions and prediction on how it would be if each factor is improved.

* Here for this analysis, we are finding the factors that are causing the lesser happiness score in a few regions i.e., bottom-most countries.
* So, first, we plotted a bar graph showing the happiness scores of the bottom-most countries which we are going to study.
* From the pie chart, we were able to analyze that these countries are supposed to focus more on their Freedom, Trust and Generosity to go to better positions on the world scale. Even though the Economy and Health are higher compared to other factors, it still would need concentration since it is lower when compared to other countries.

## **Data Analysis and Machine Learning - Python**

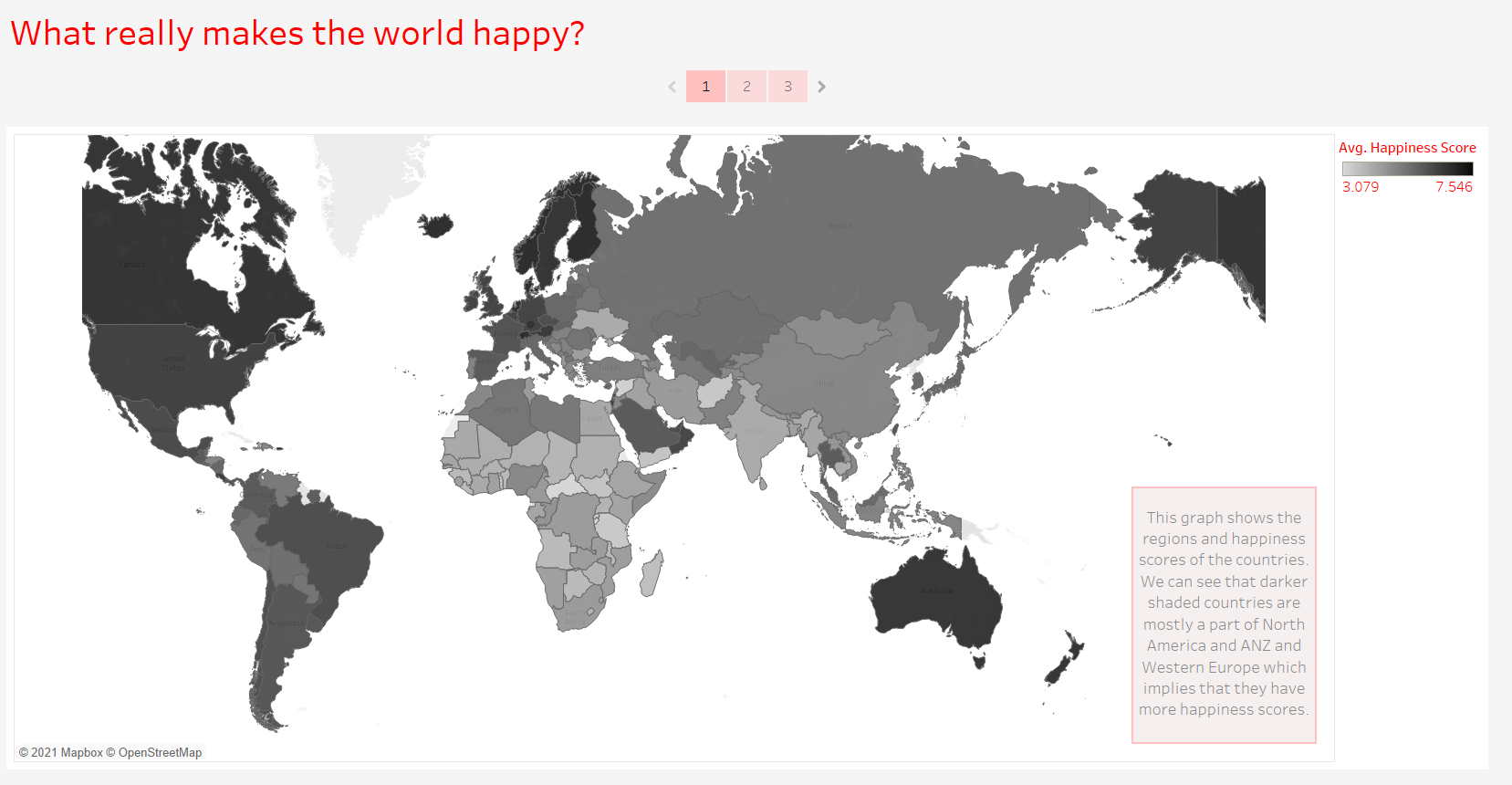
A prediction was made in python. We created a country called Custom Land meaning you can input your values for all the above factors and get to know what the happiness score might be.

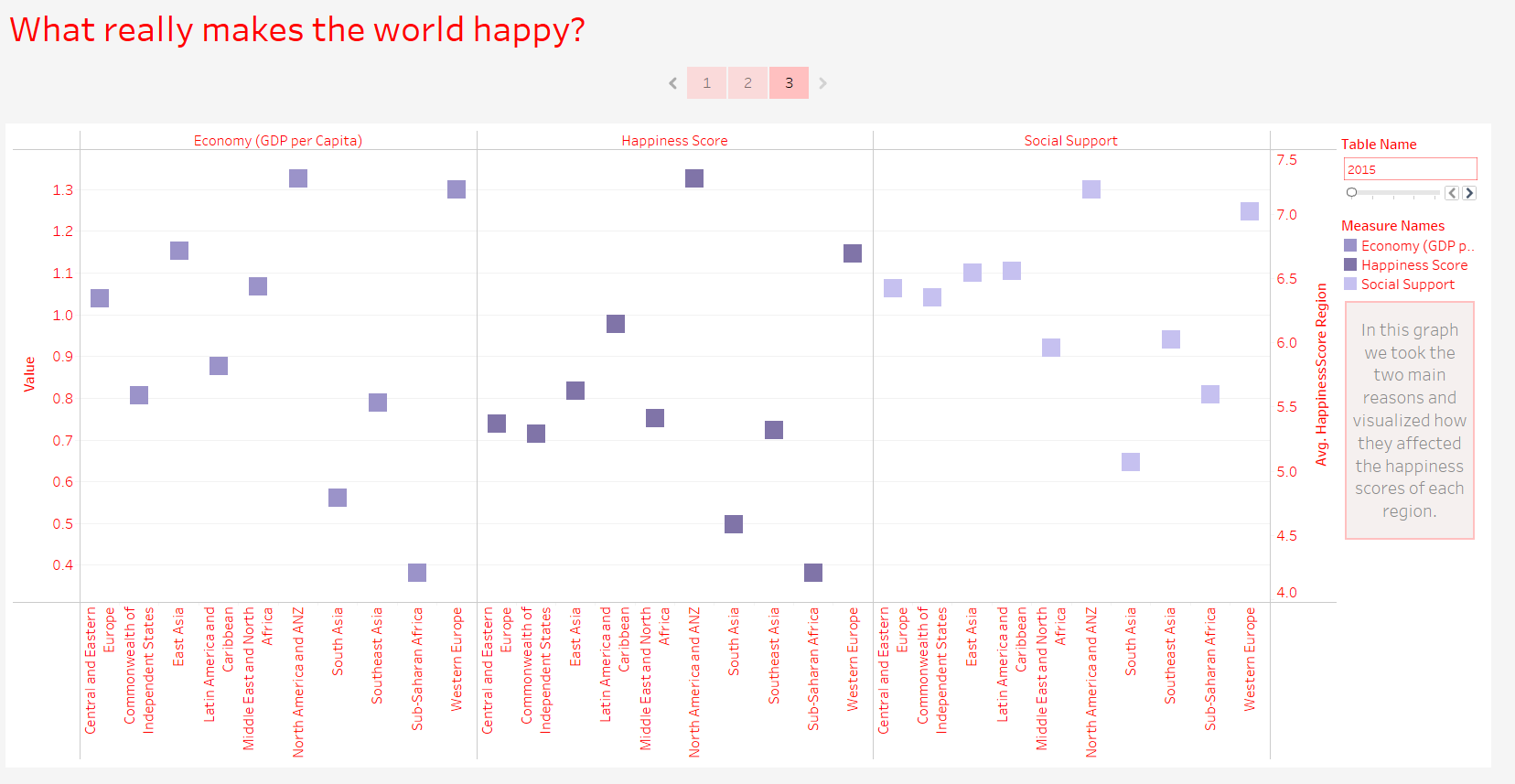
## **Website Development**

We created a static angular website with a few web pages in it and all the Storyboards of the visualizations we made are embedded in it.

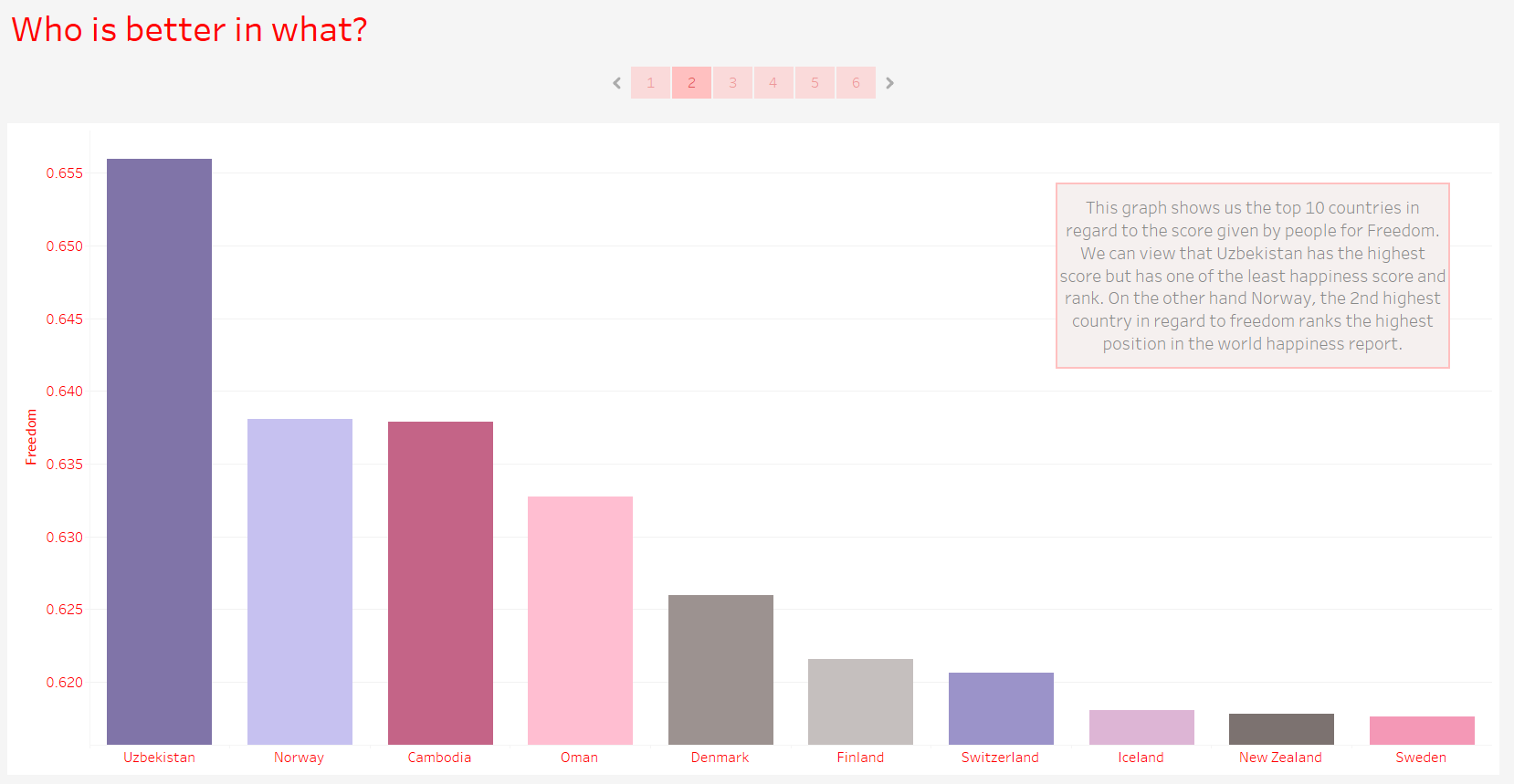
1. **OUTPUT**
   1. **Tableau**

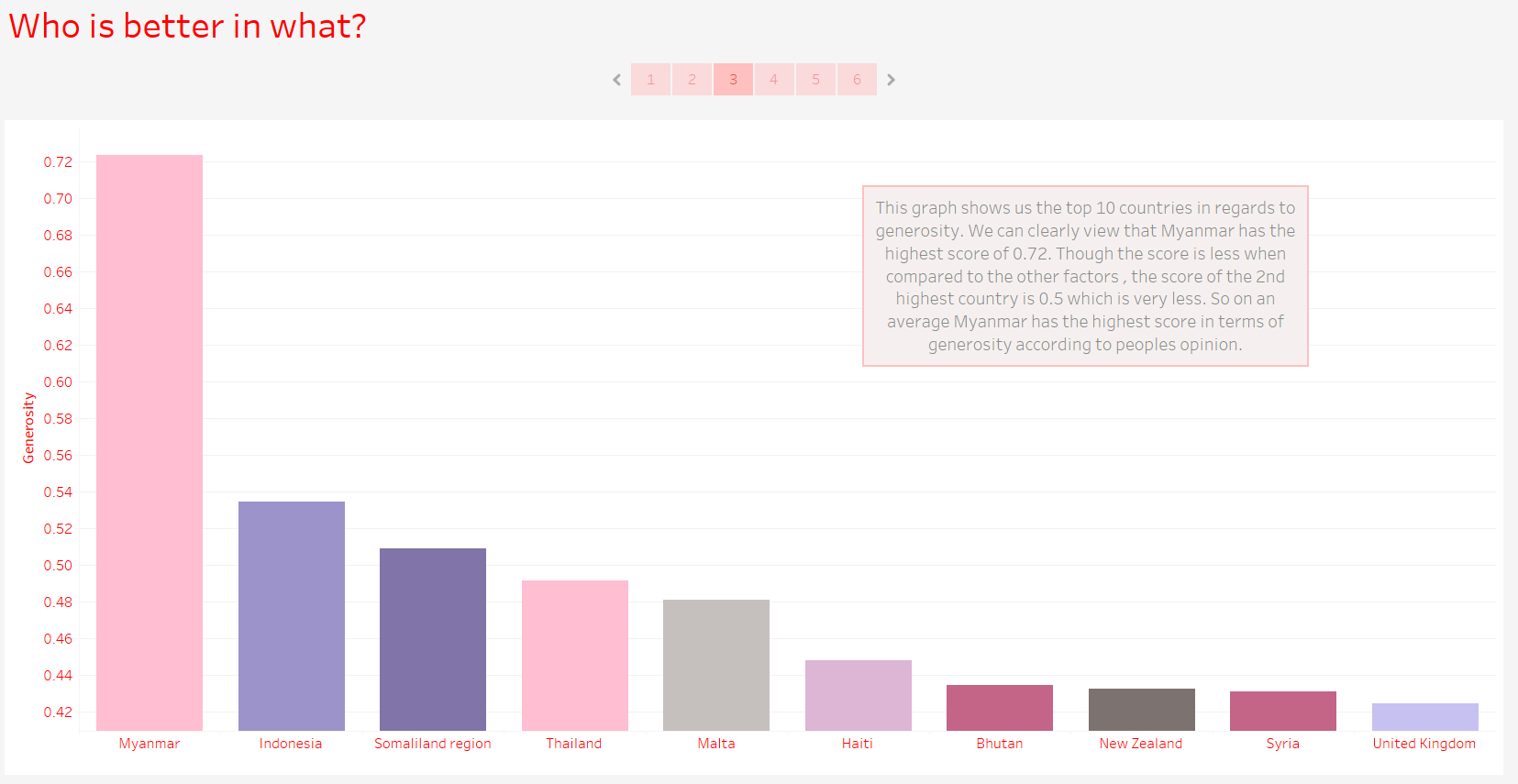
Q0:

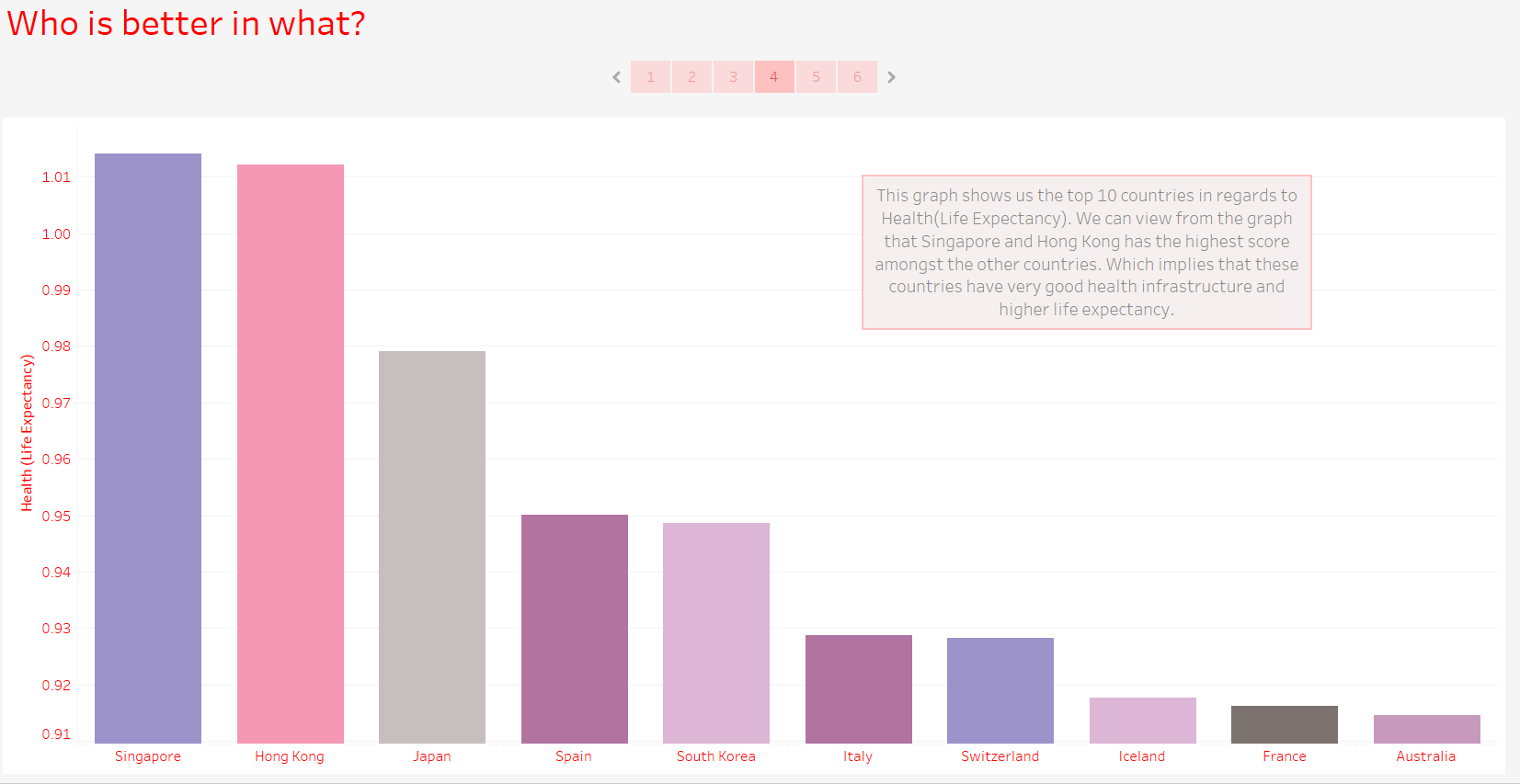
****

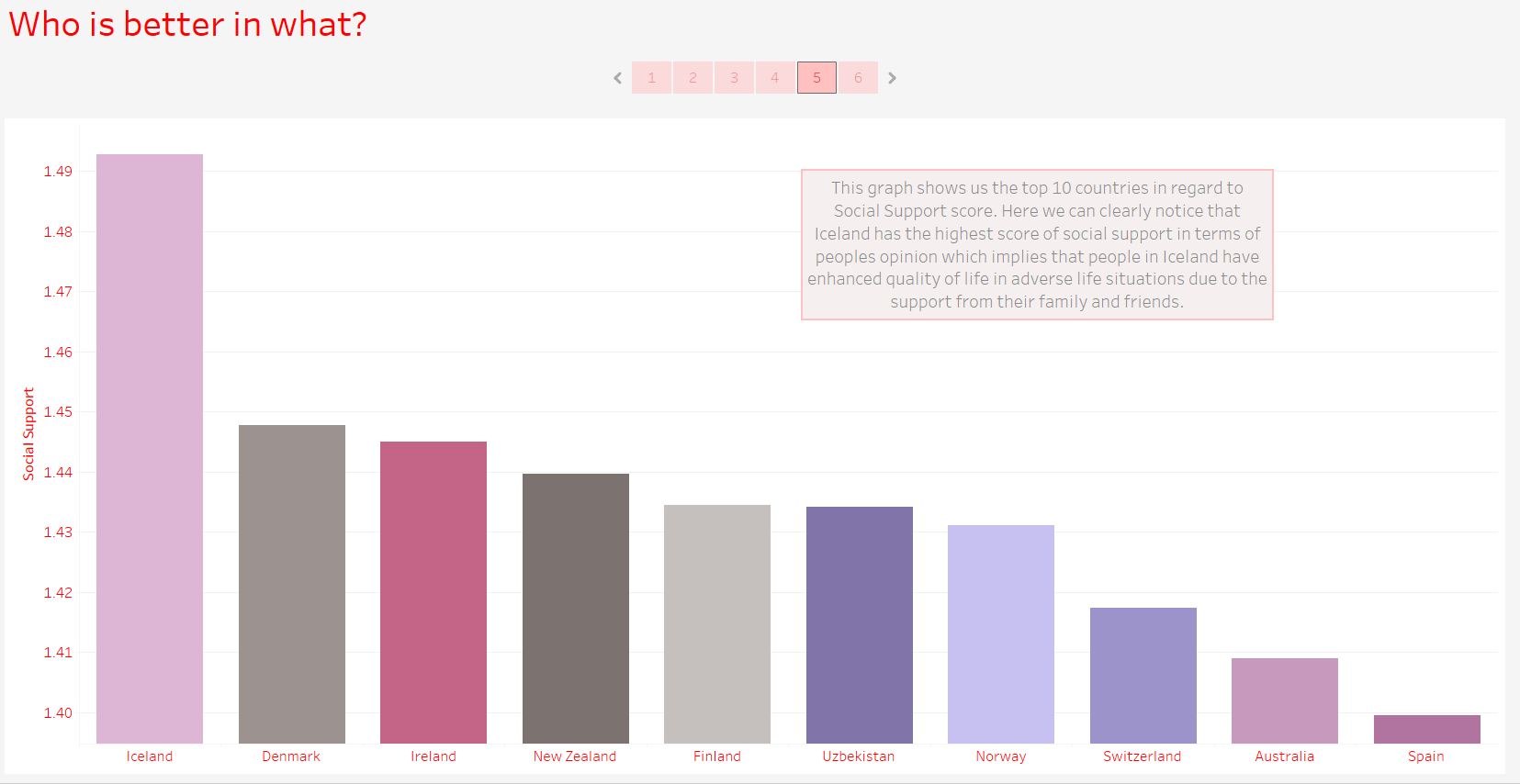
****

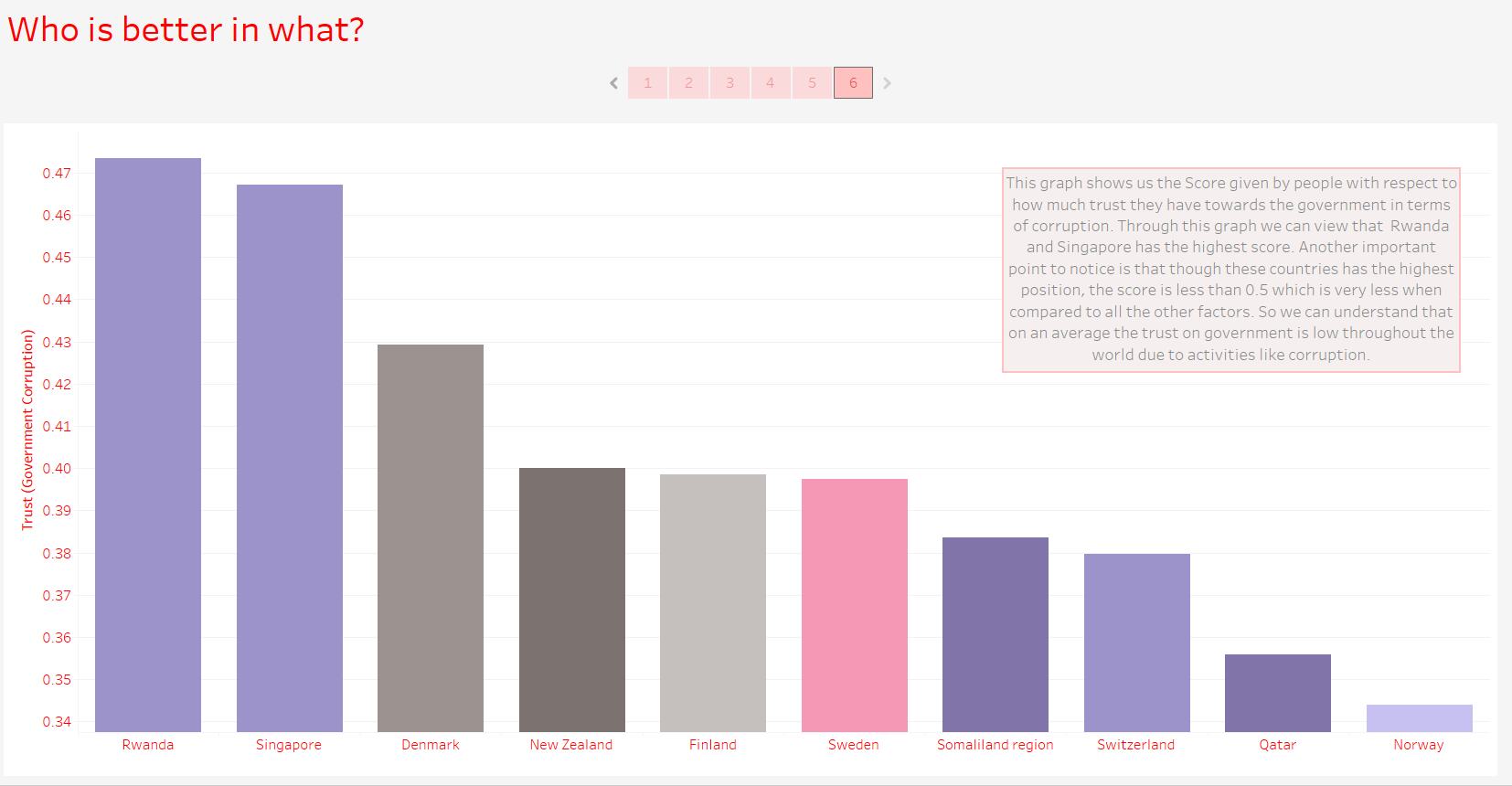
Q1:



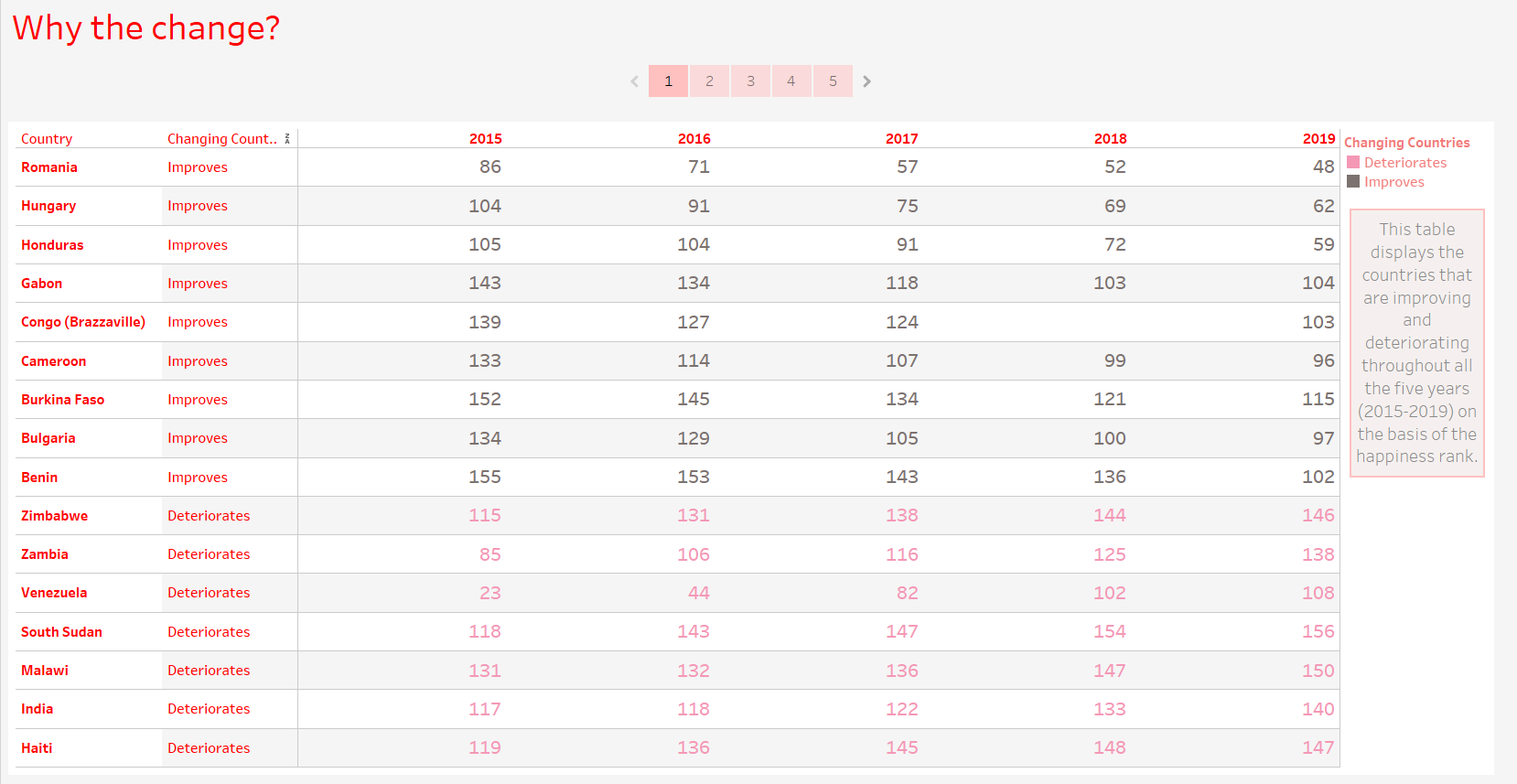


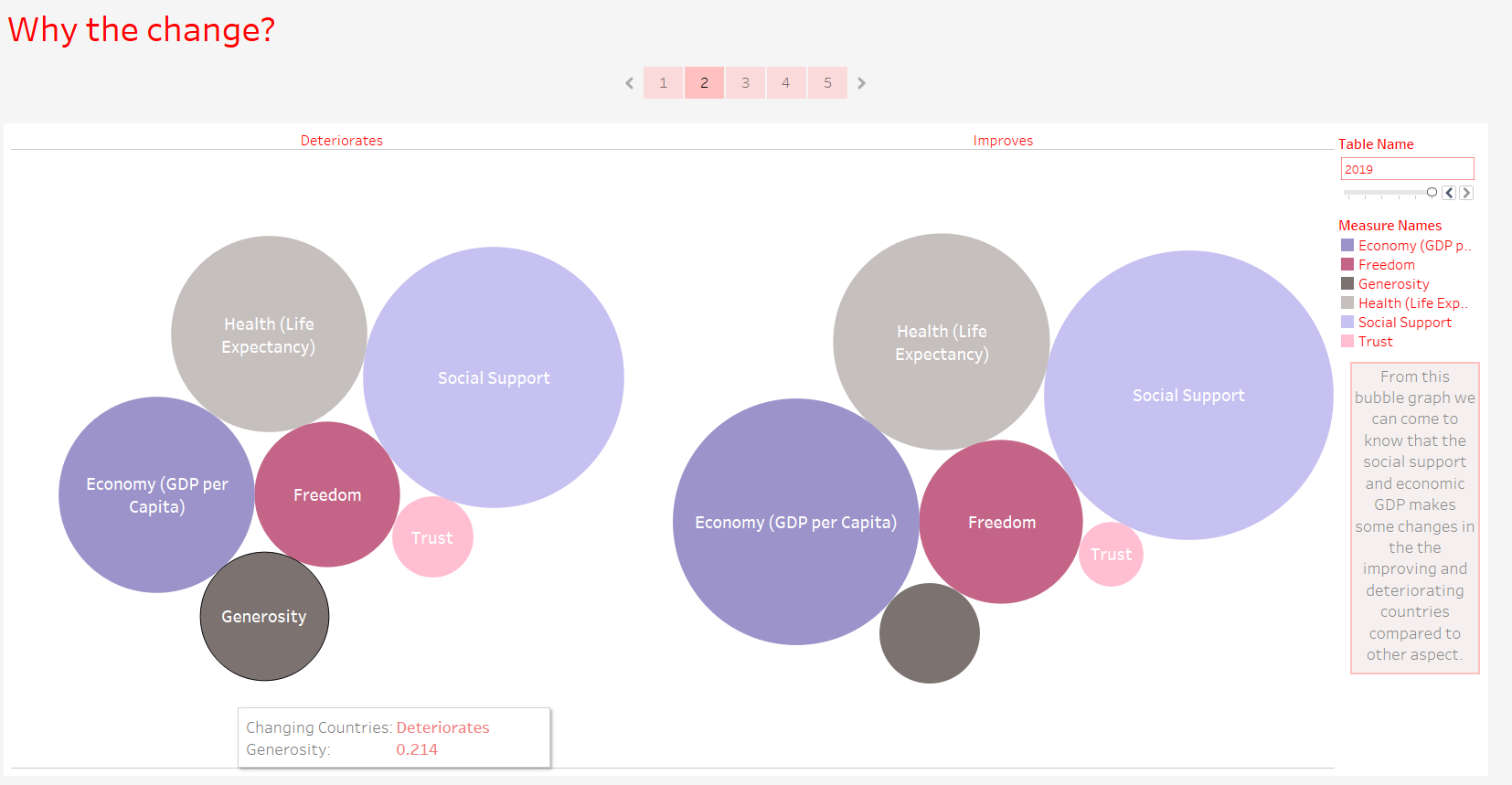


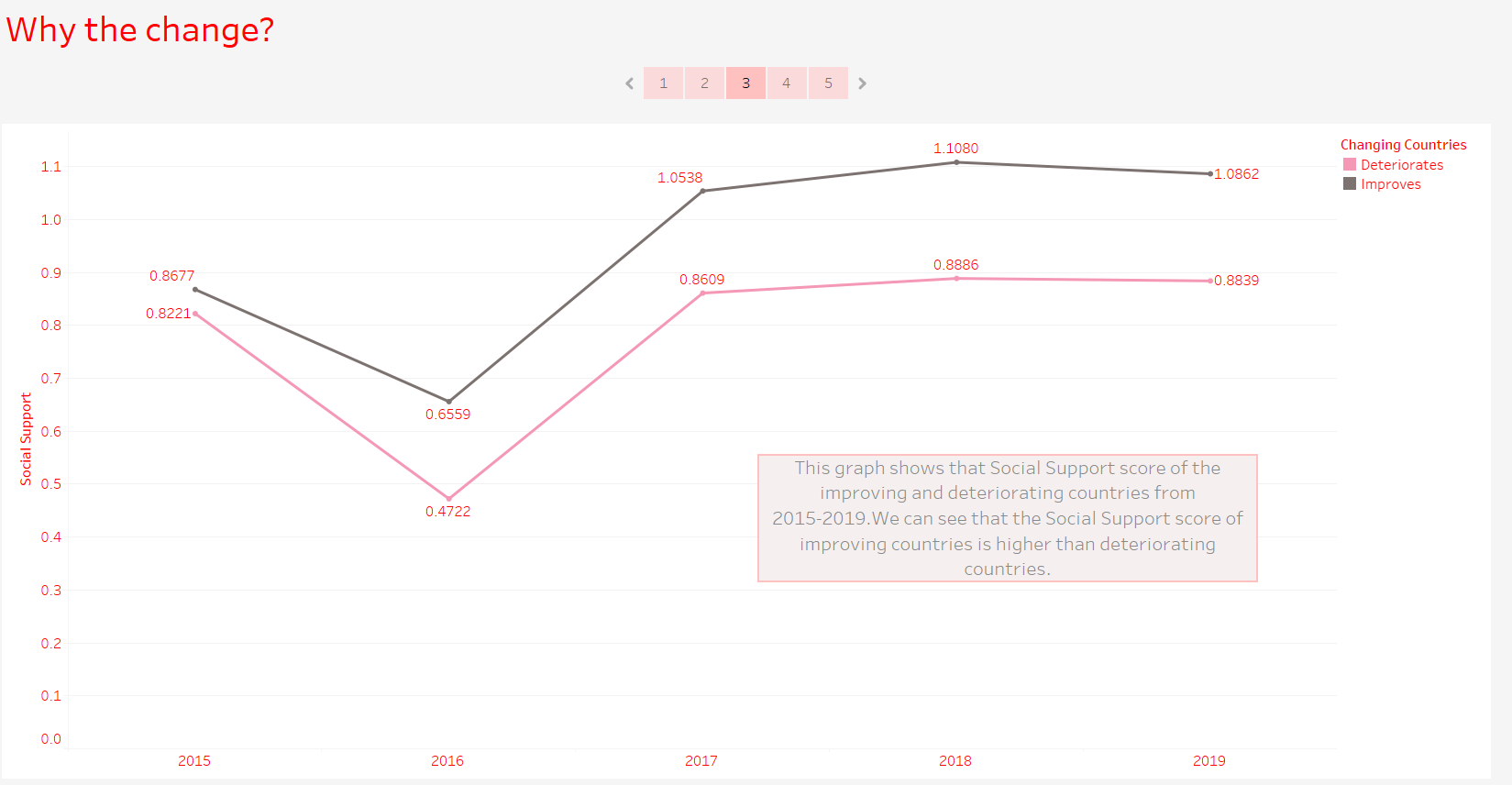


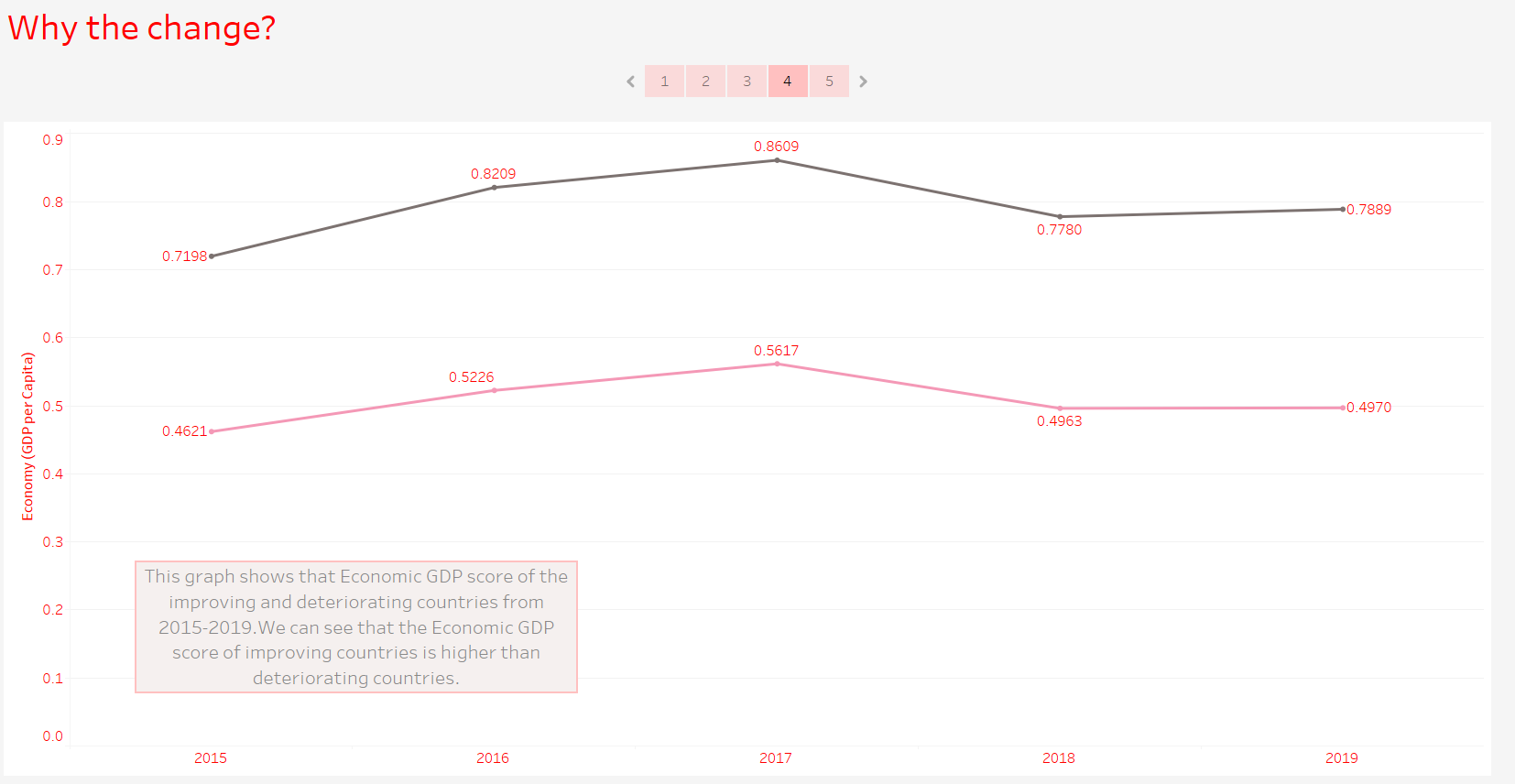


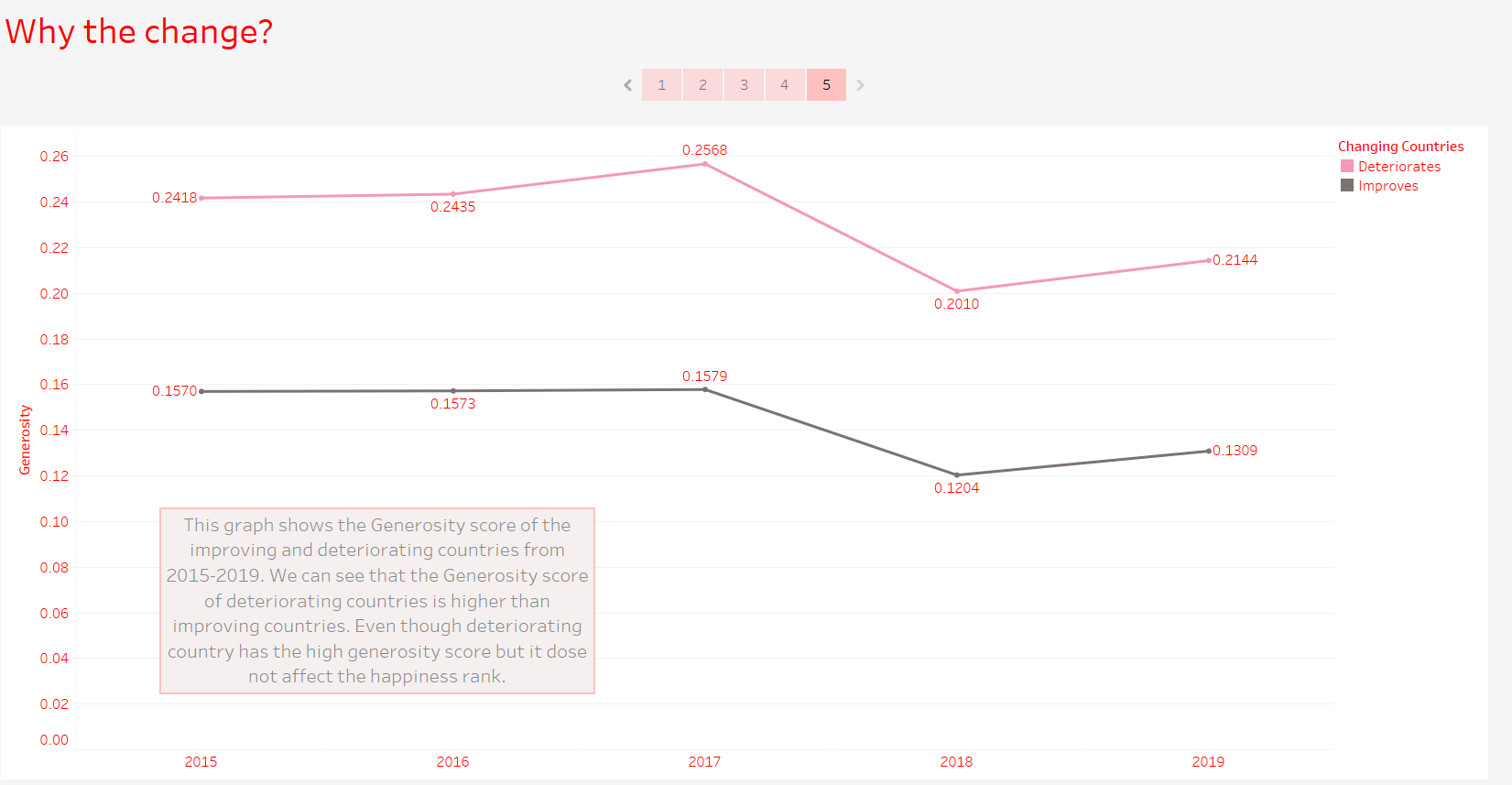
Q2:



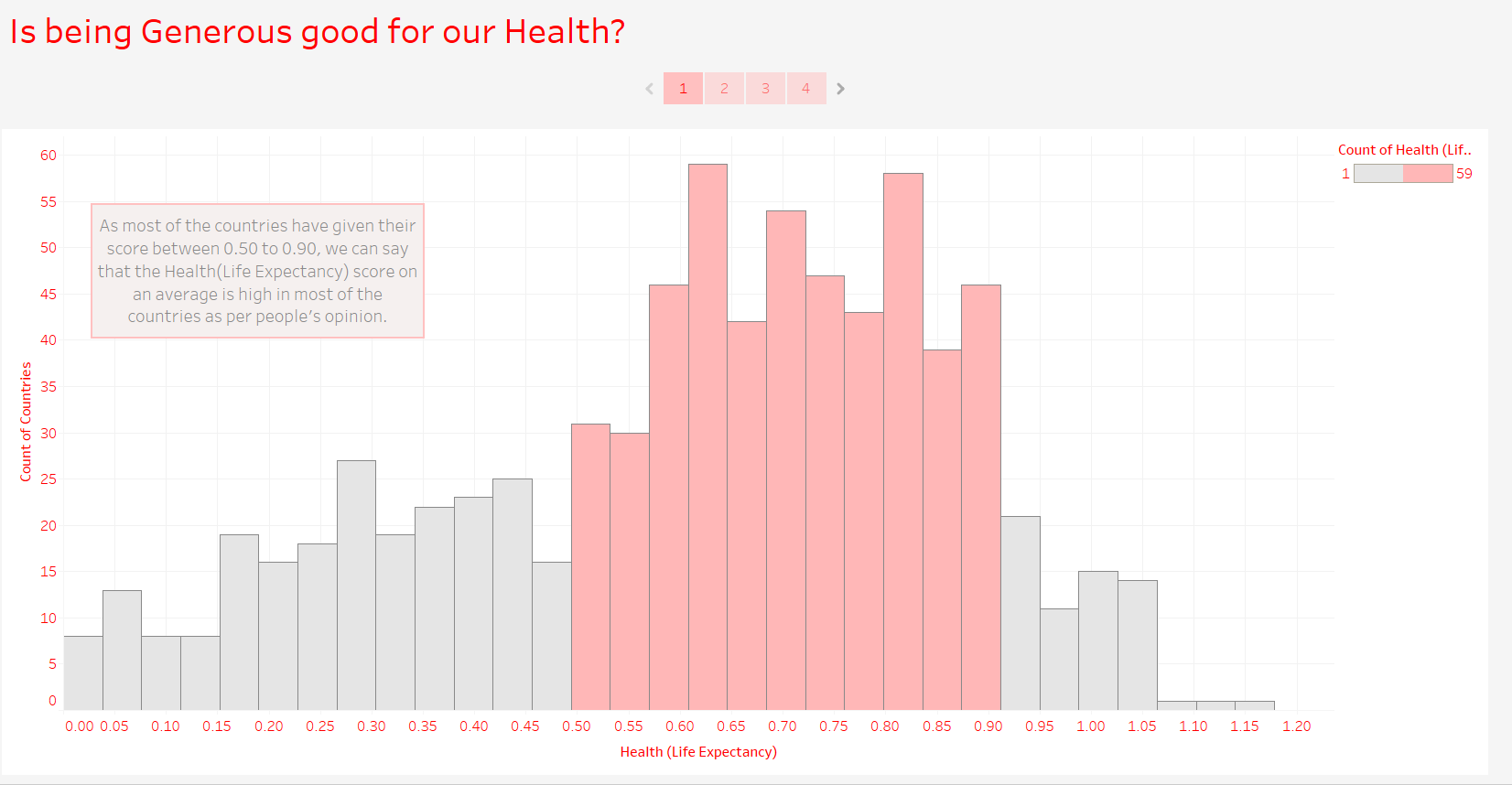


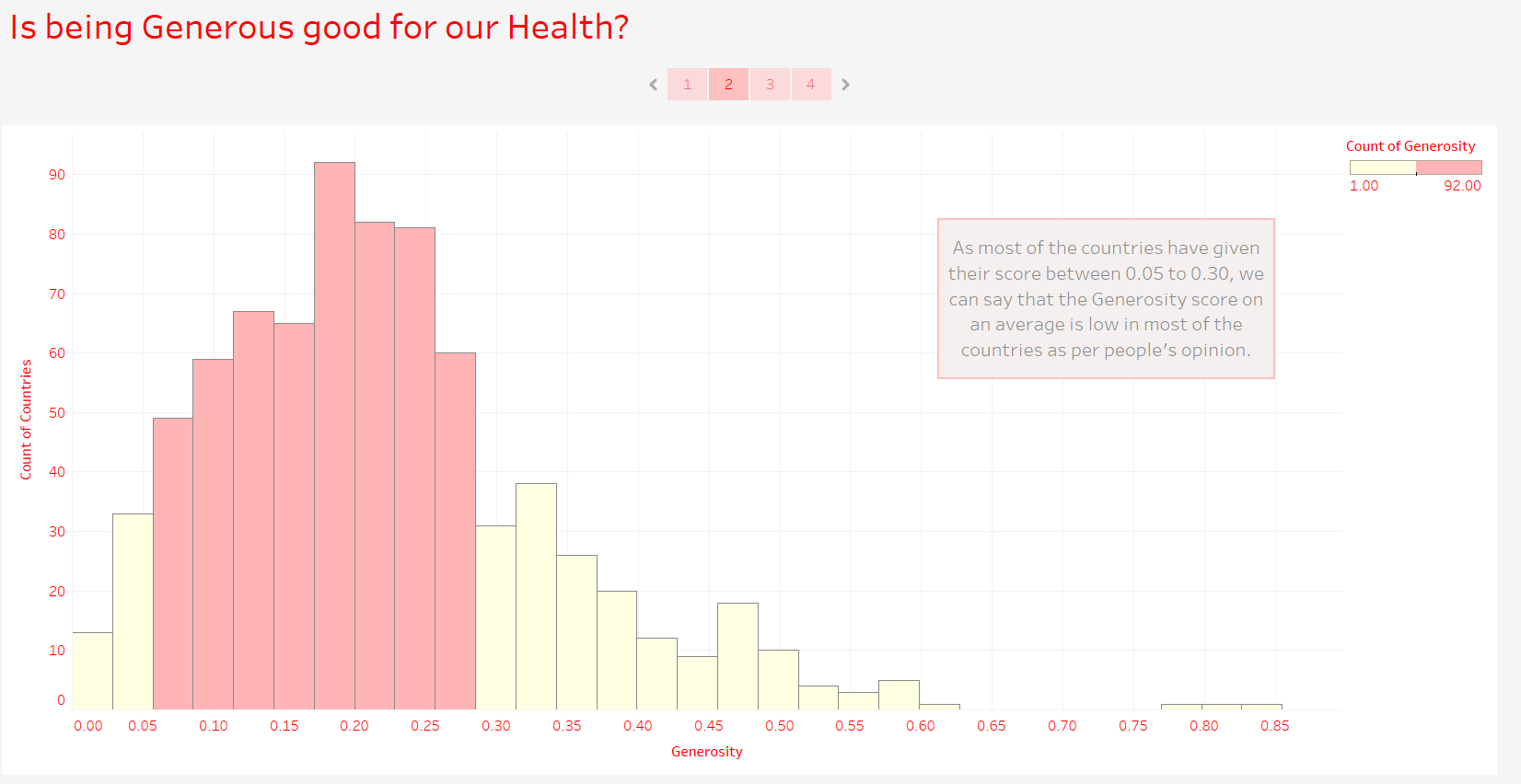


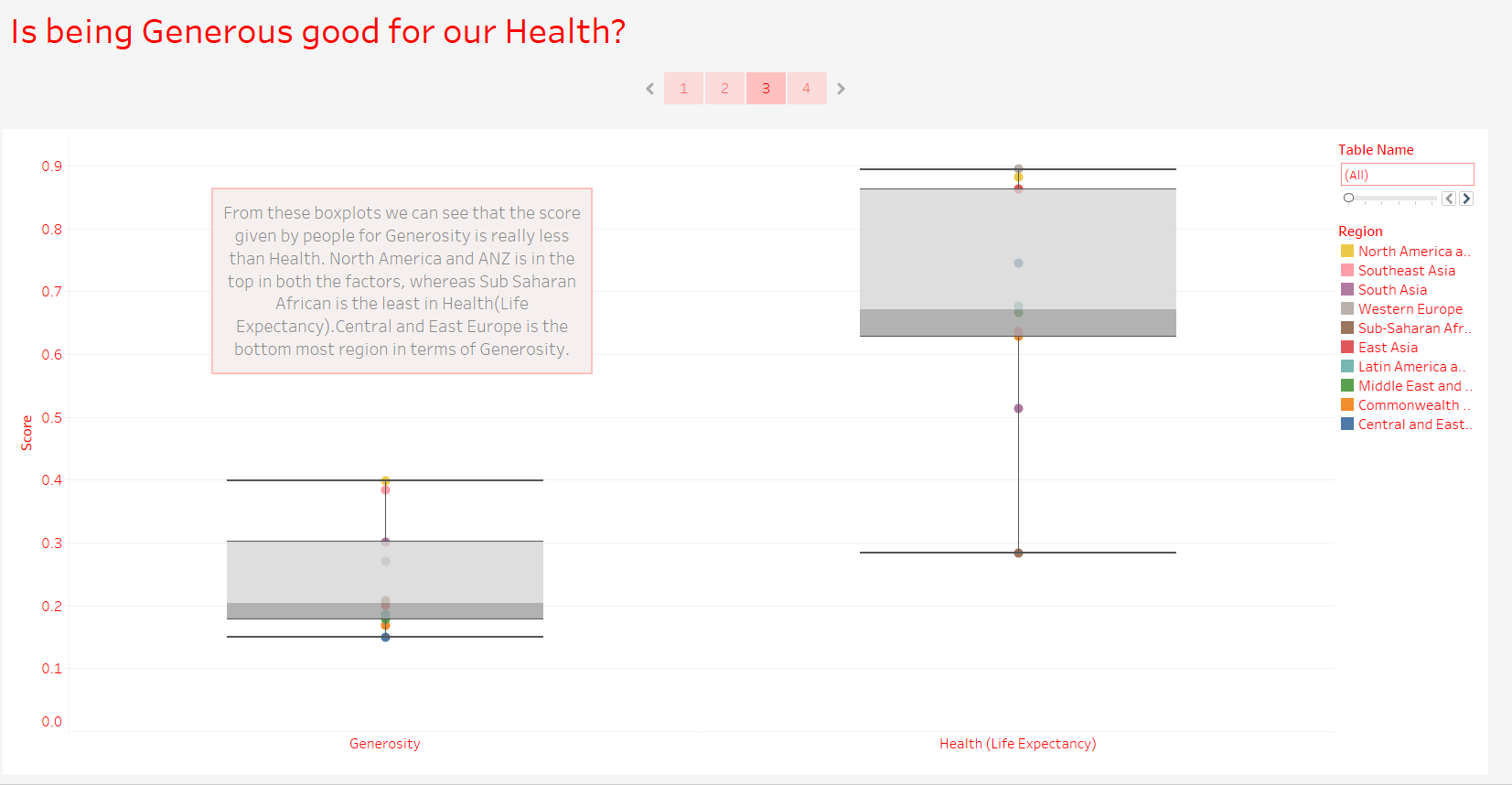


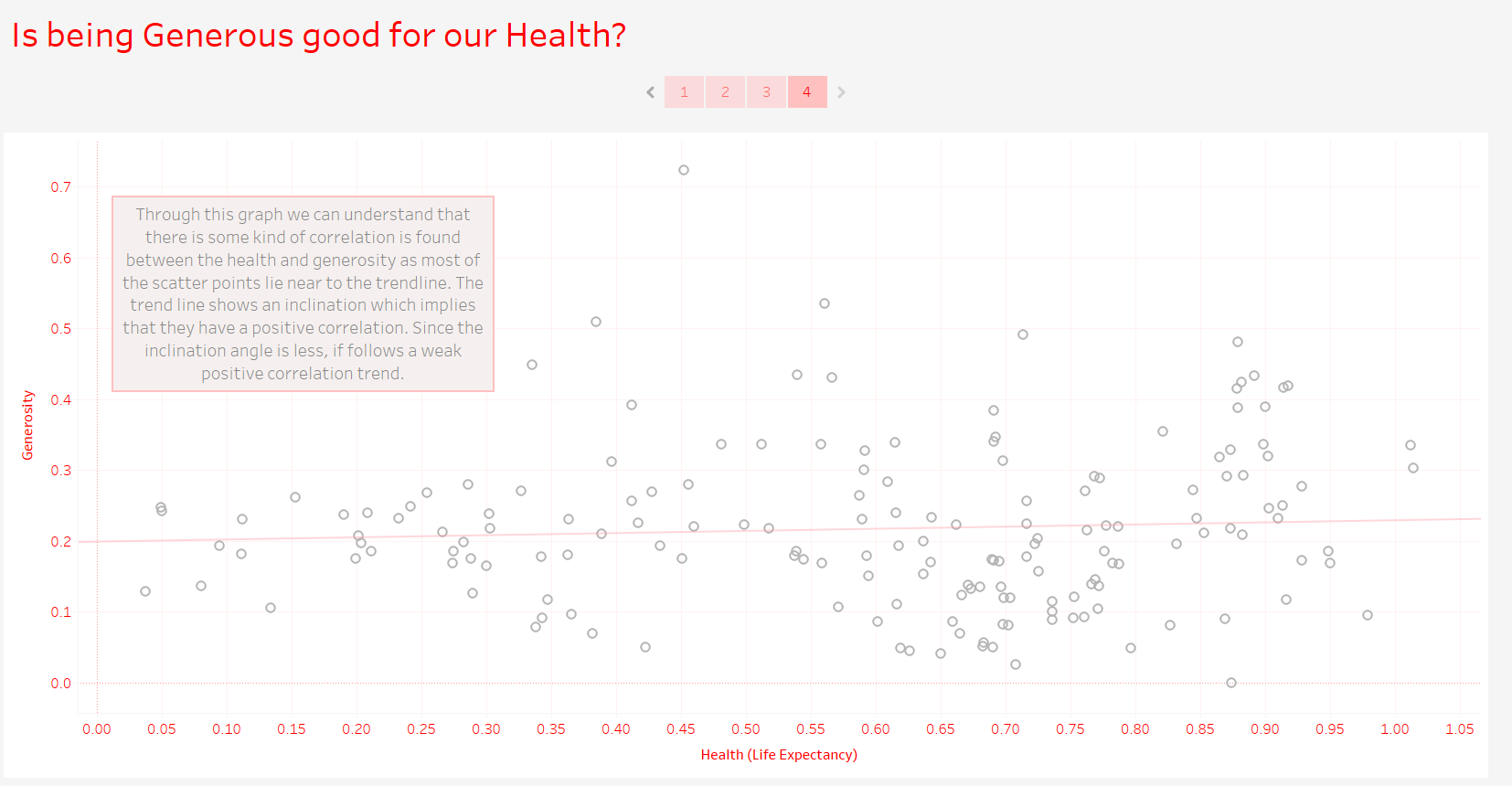


Q3:



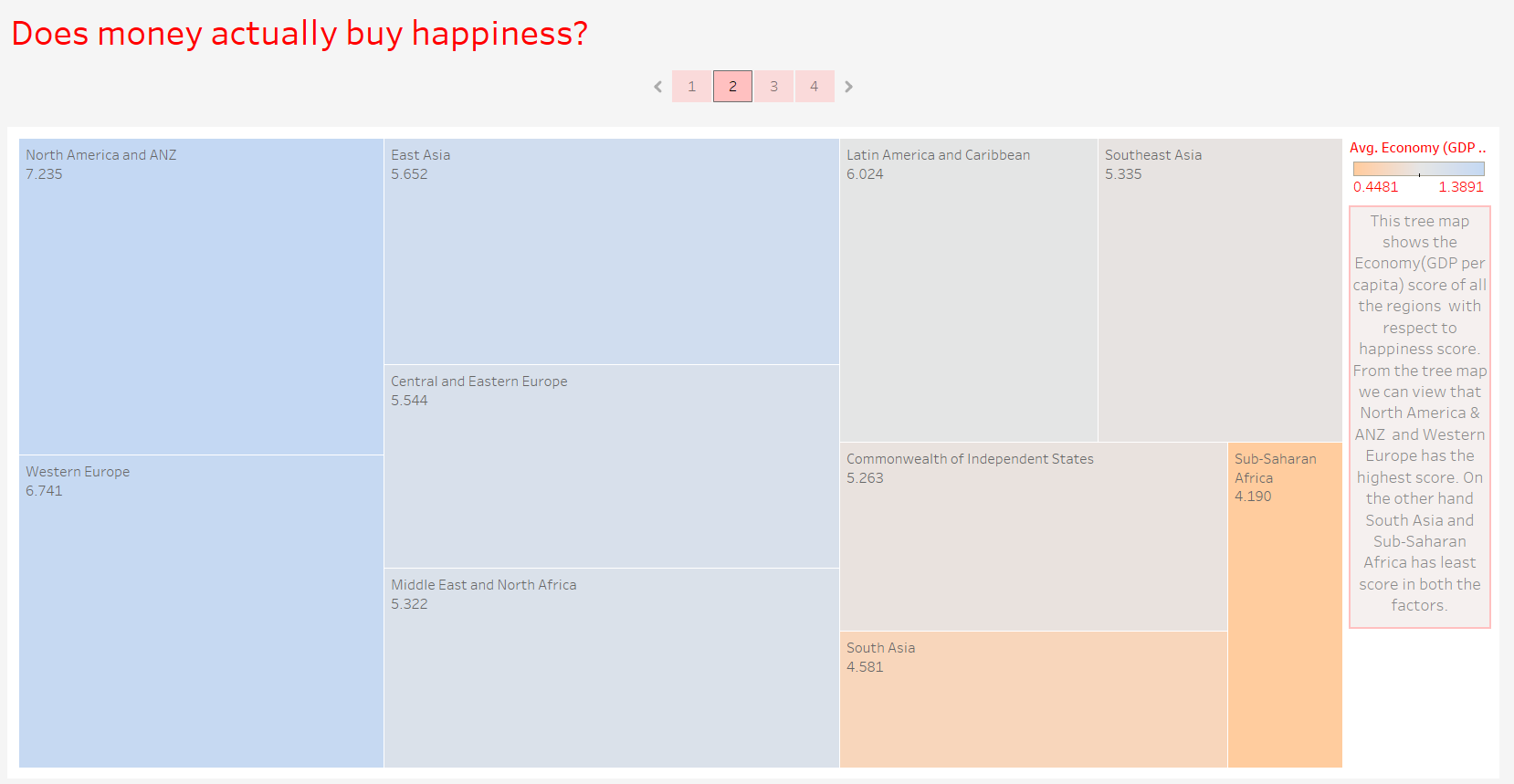


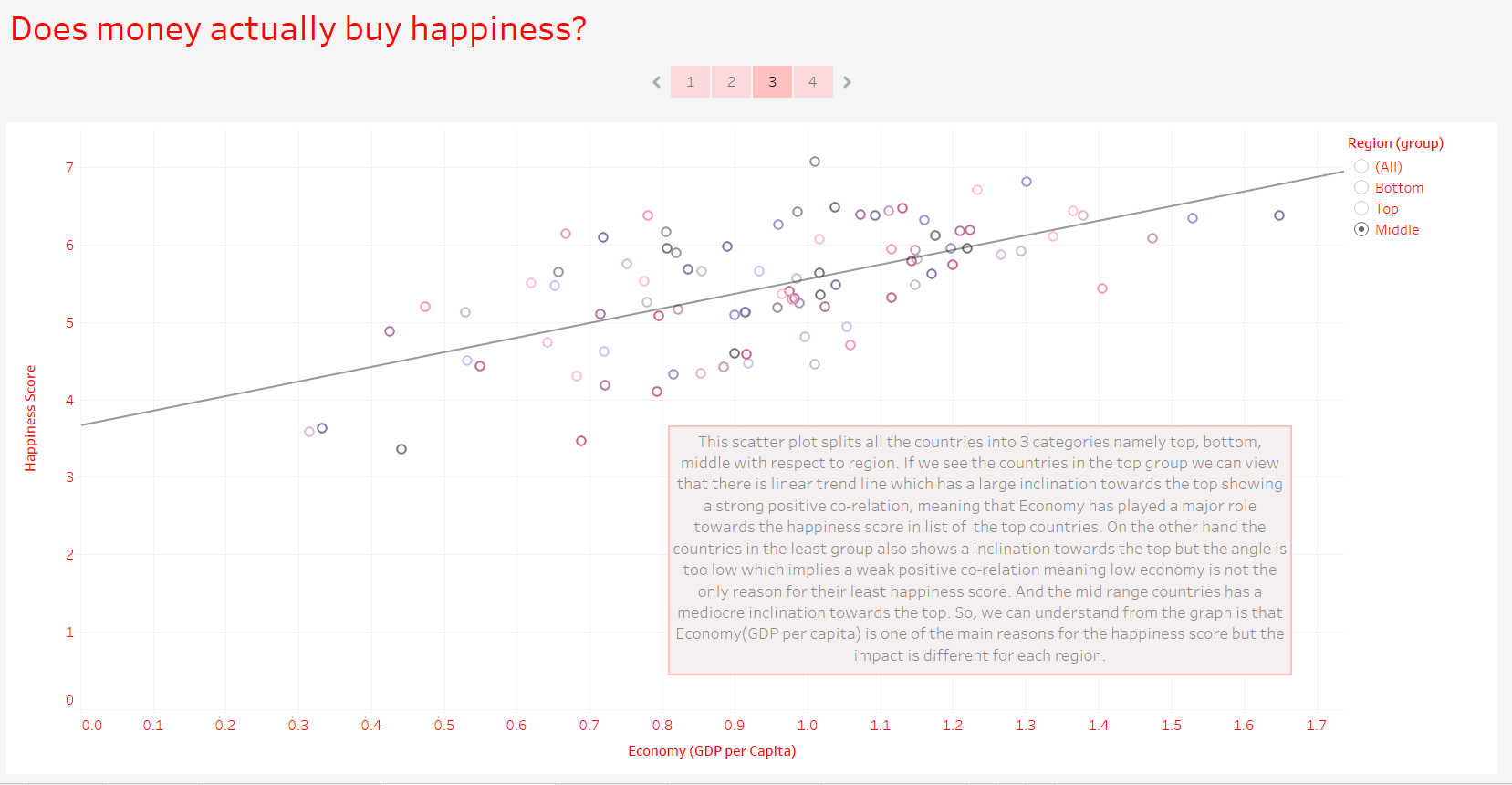


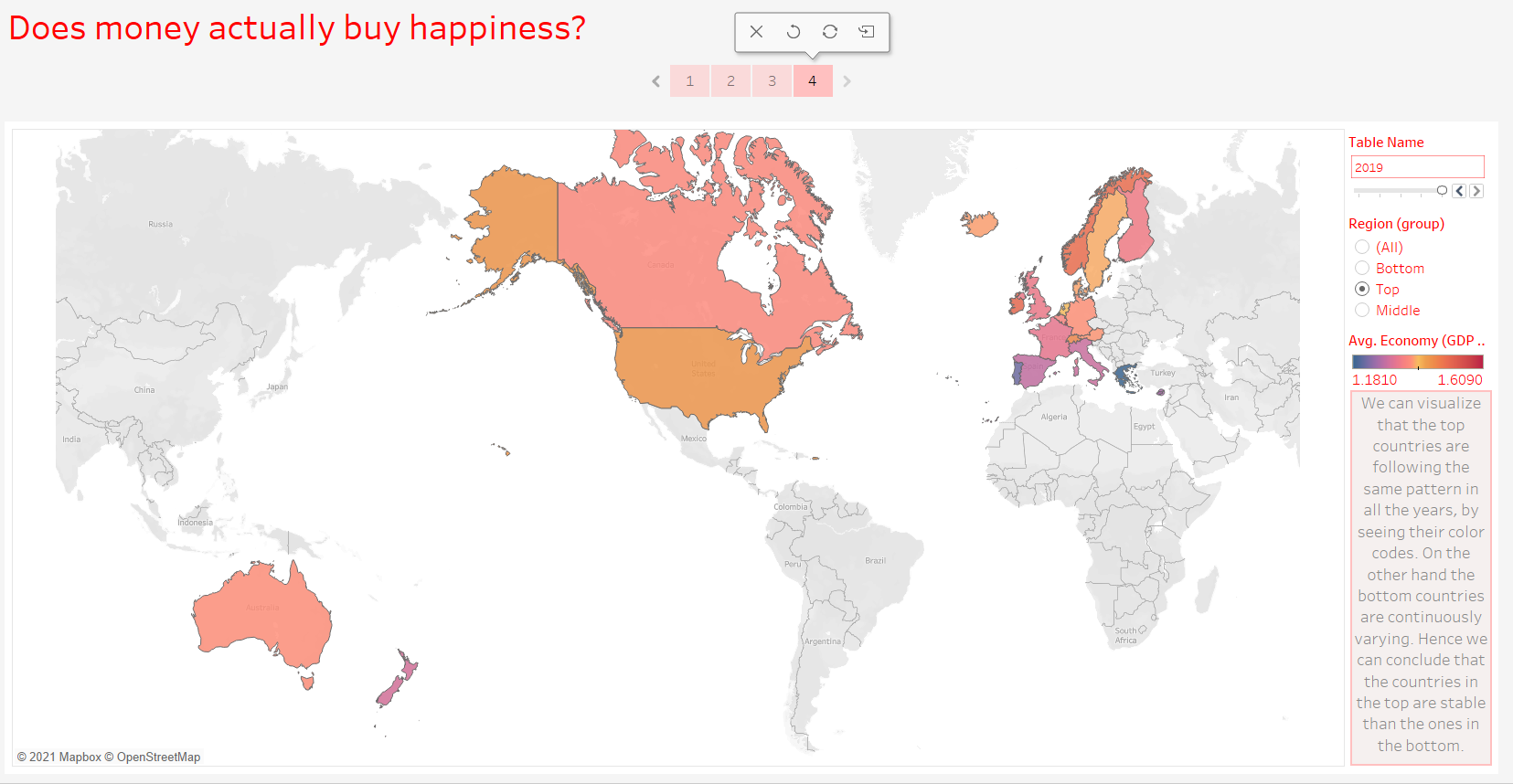


Q4:







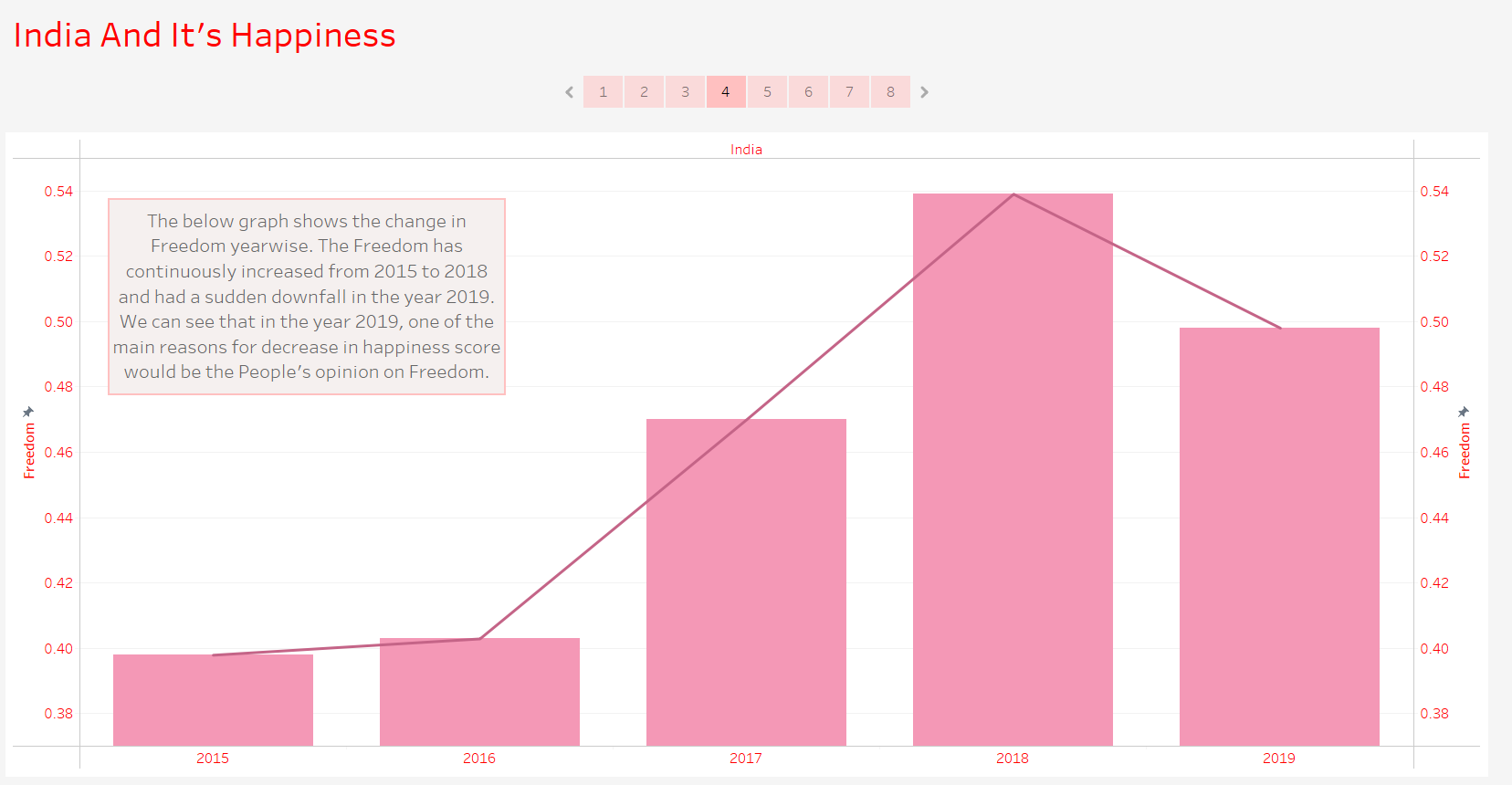


Q5:

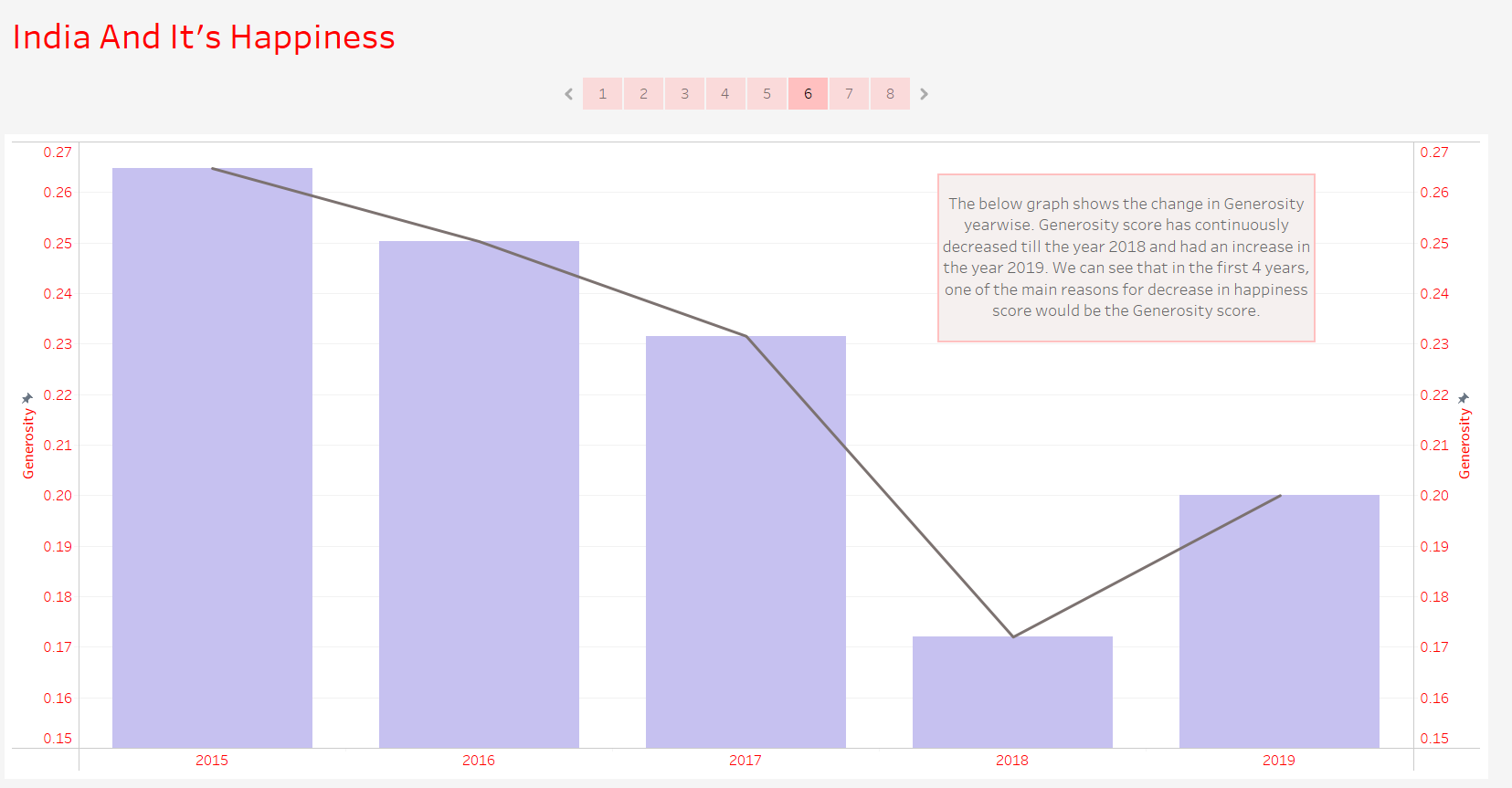


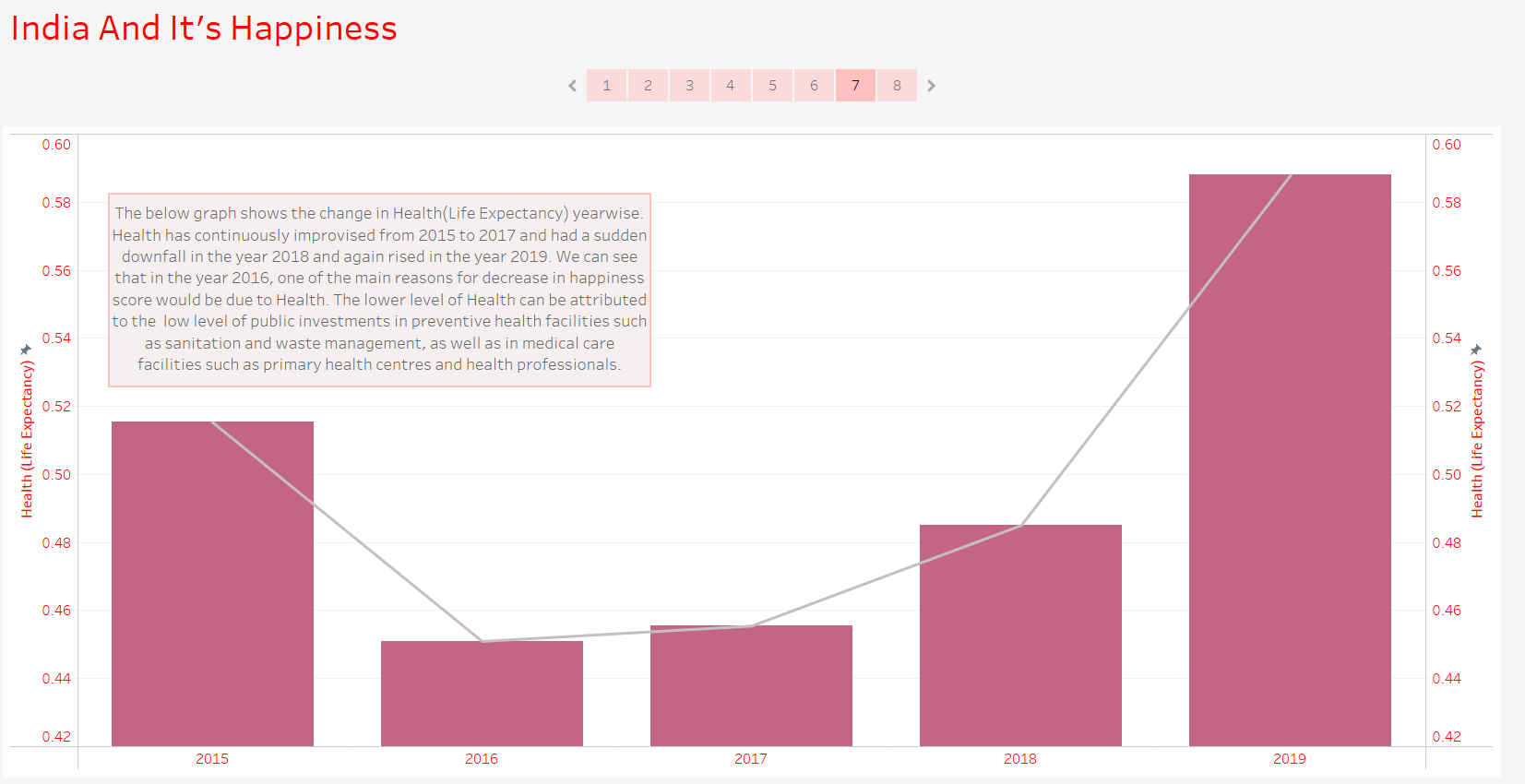


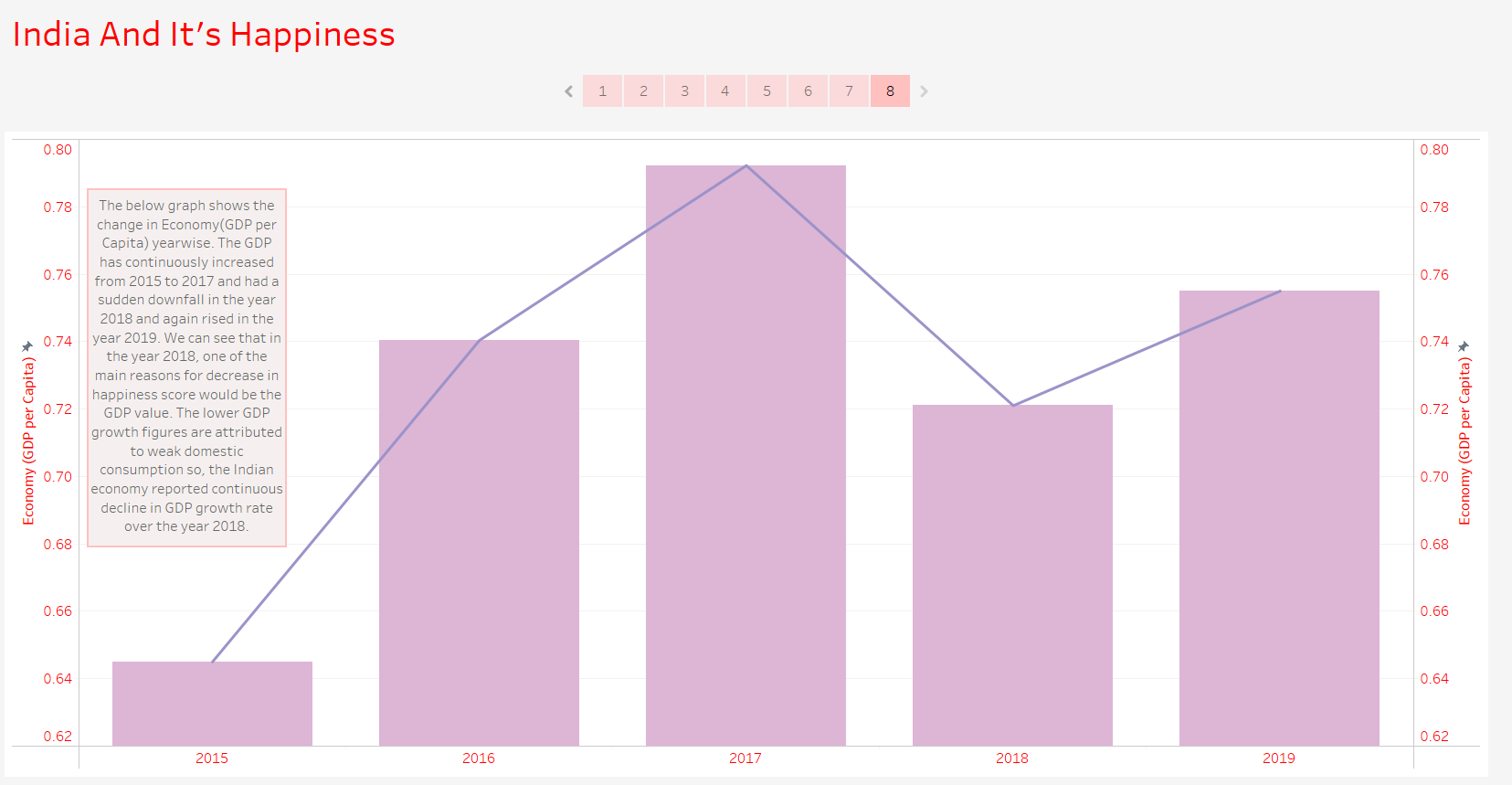




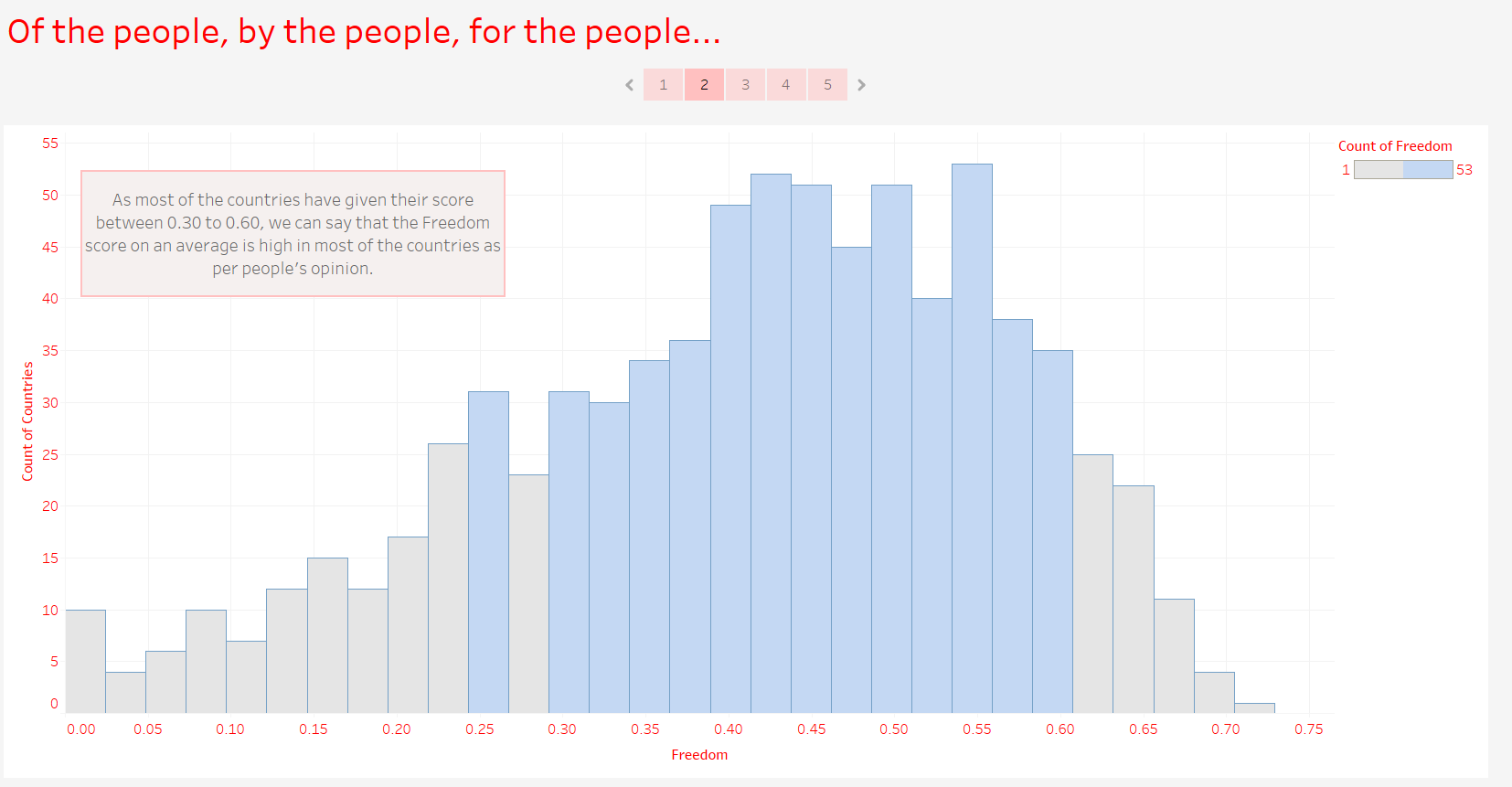




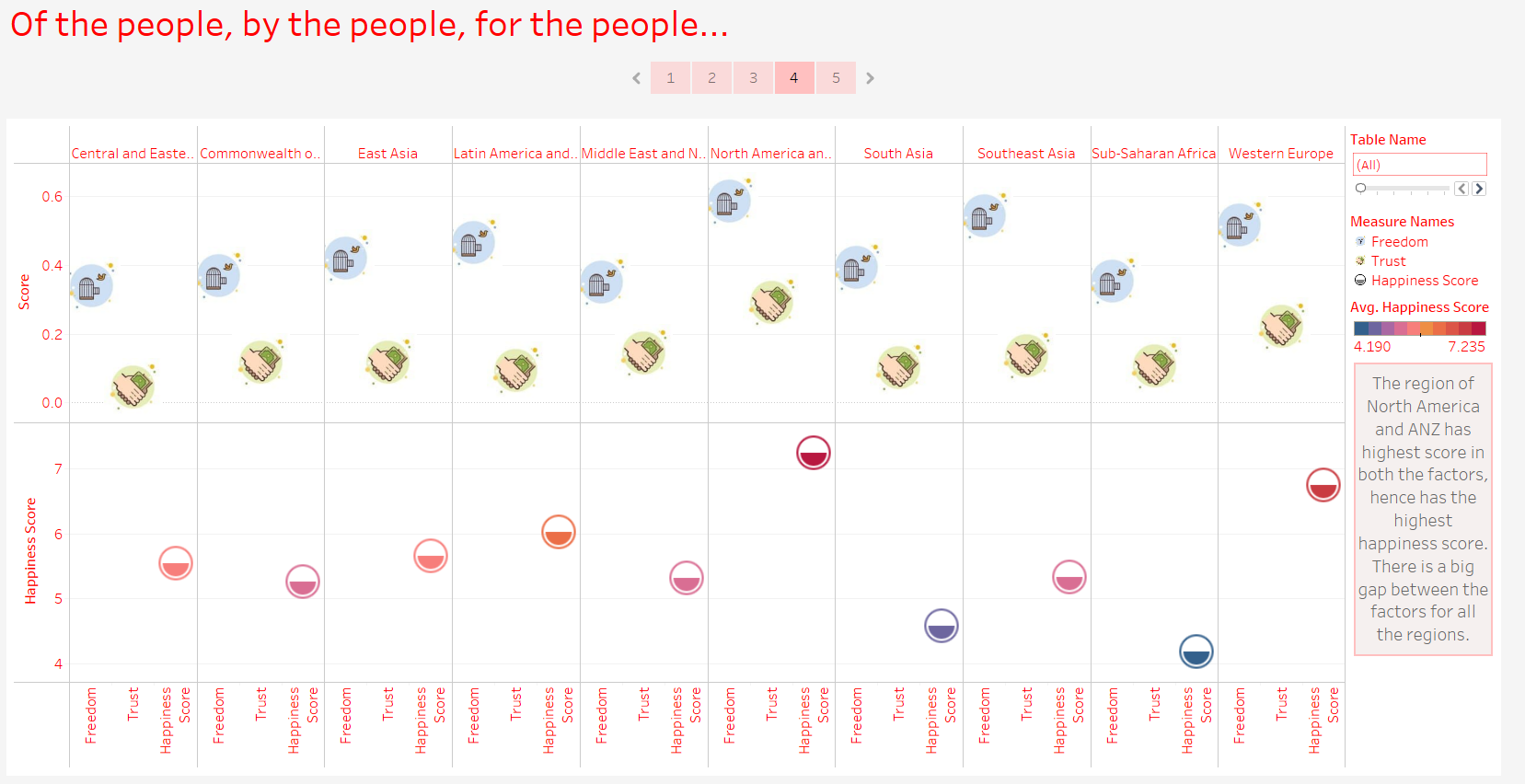


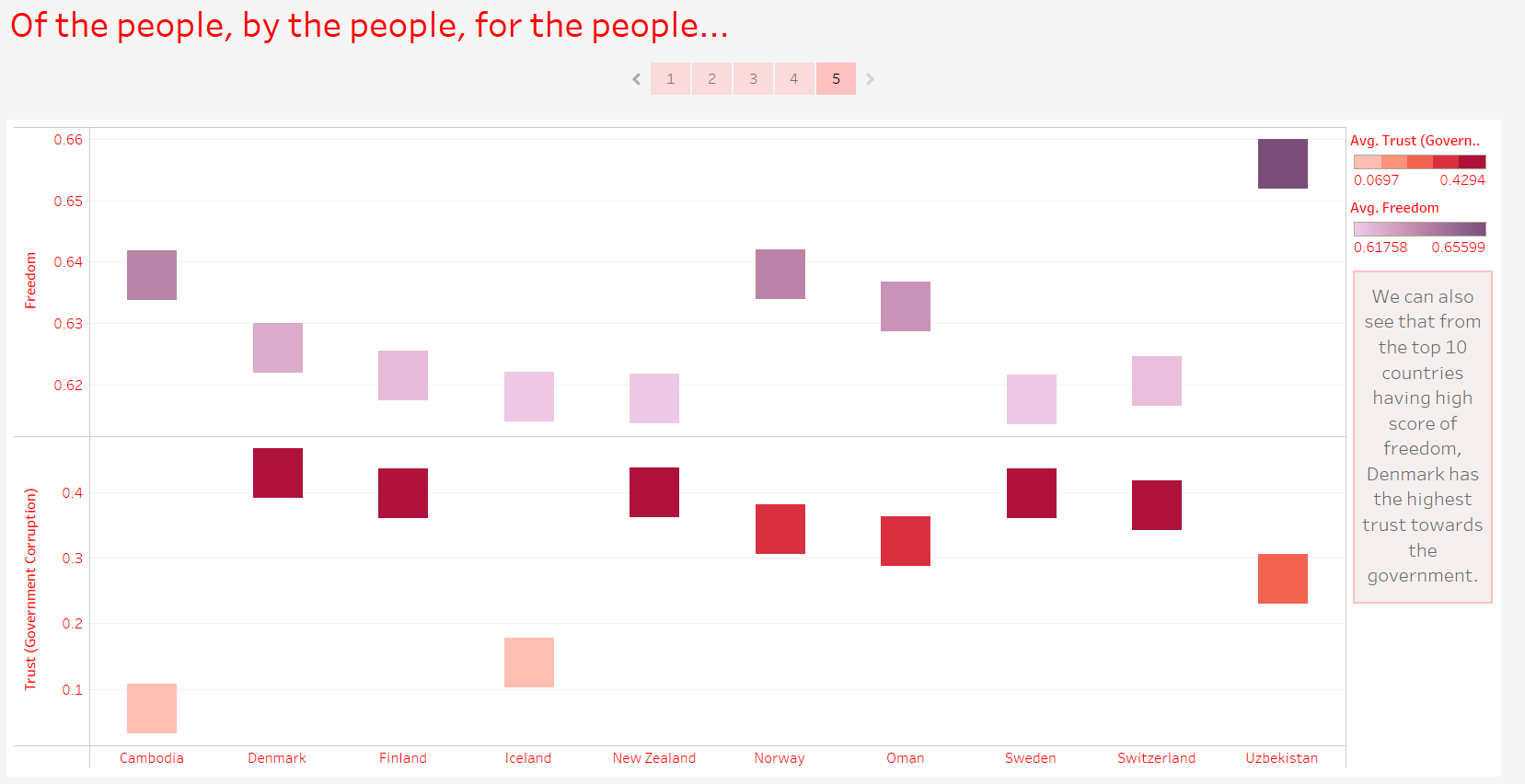


Q6:

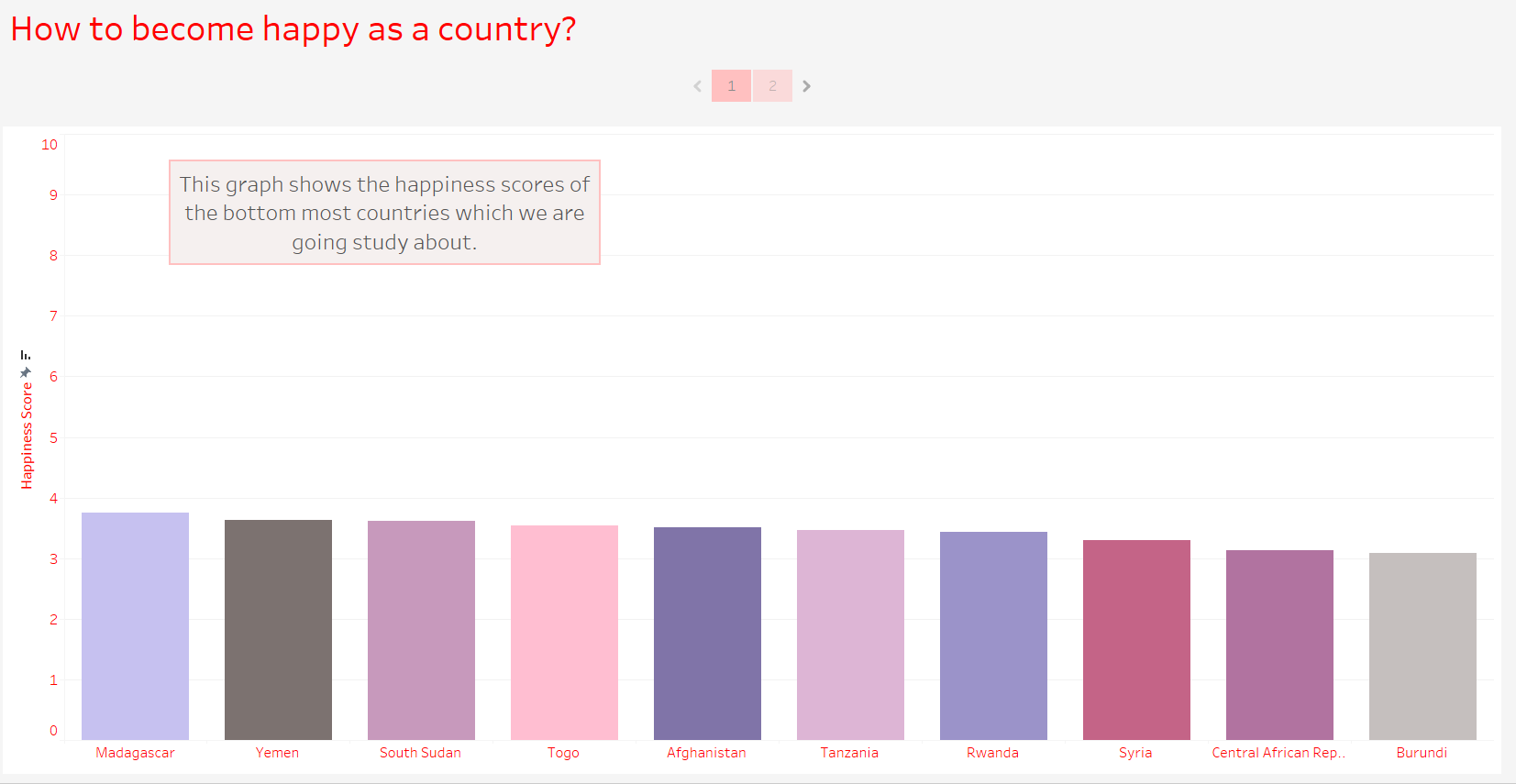


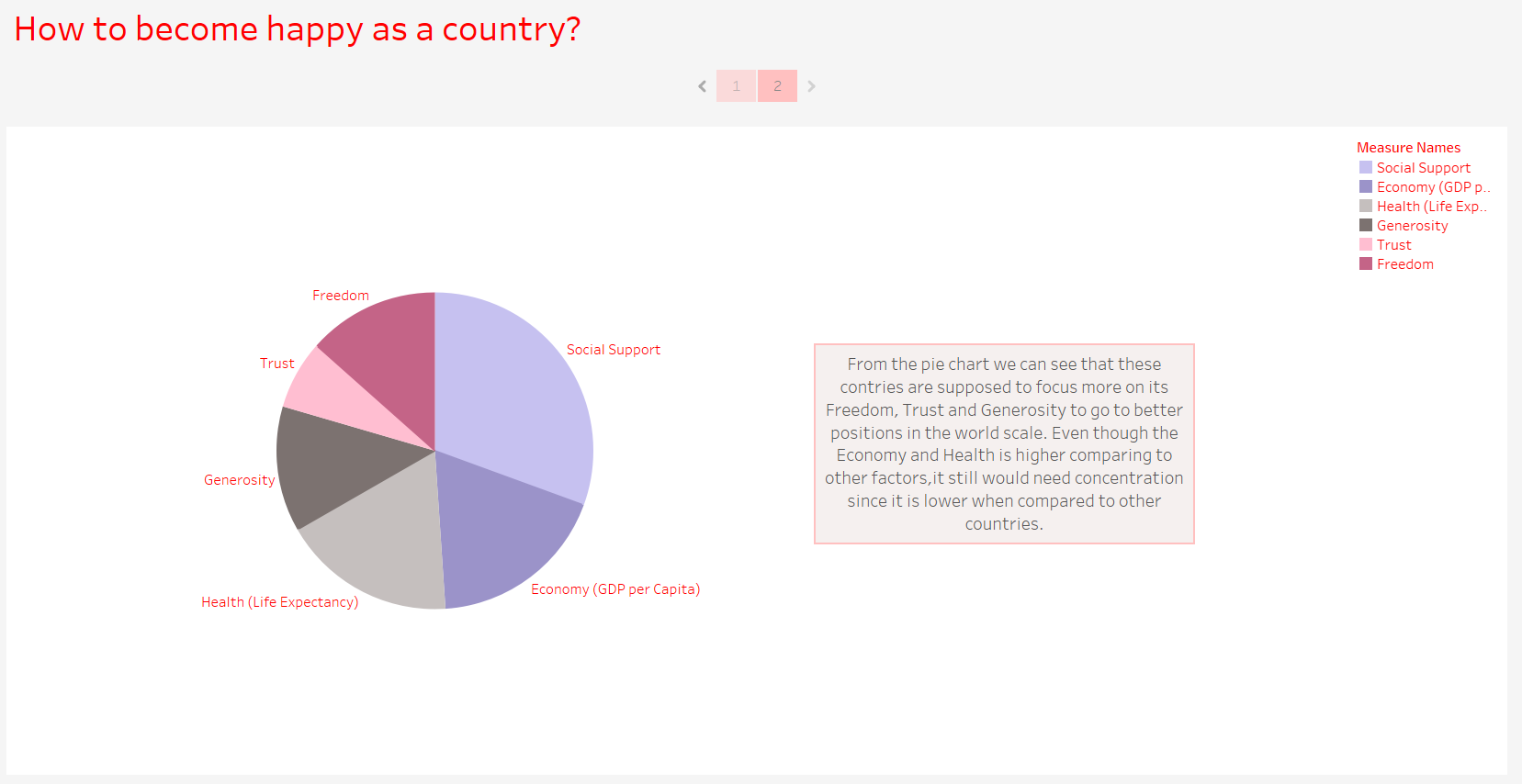






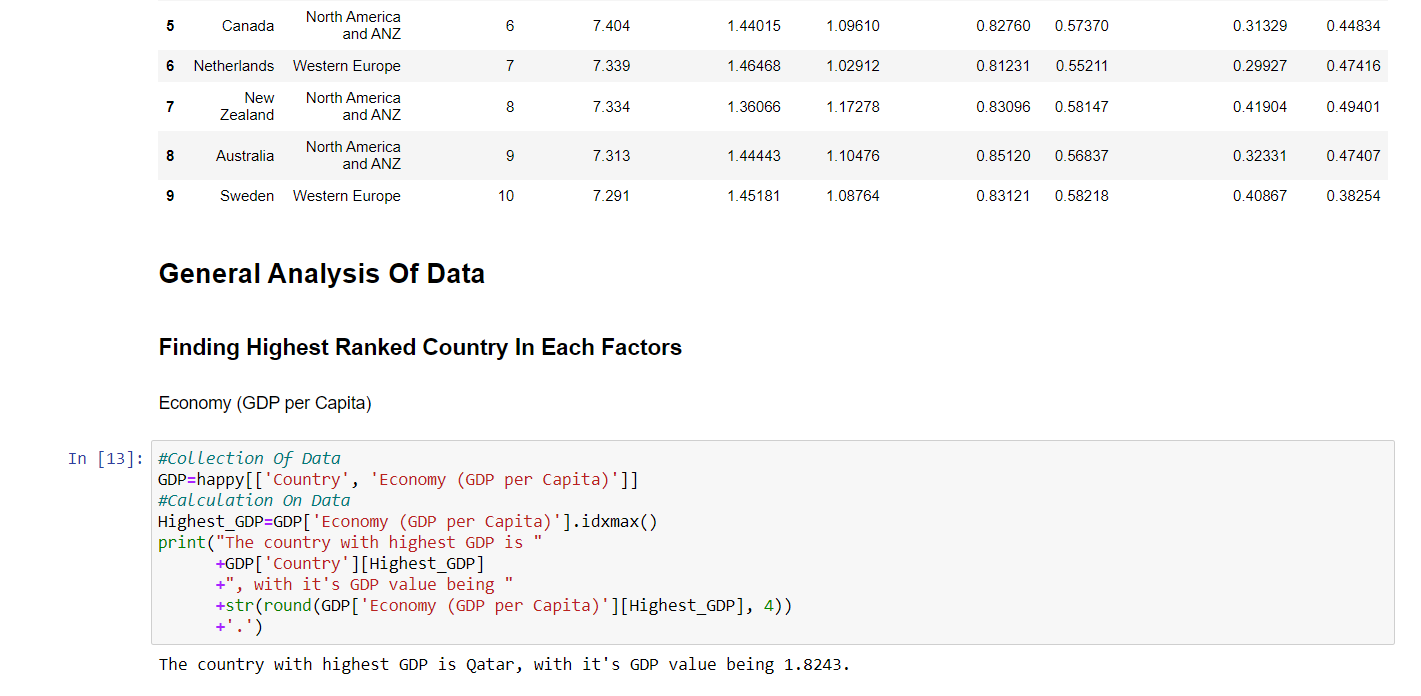
Q7:



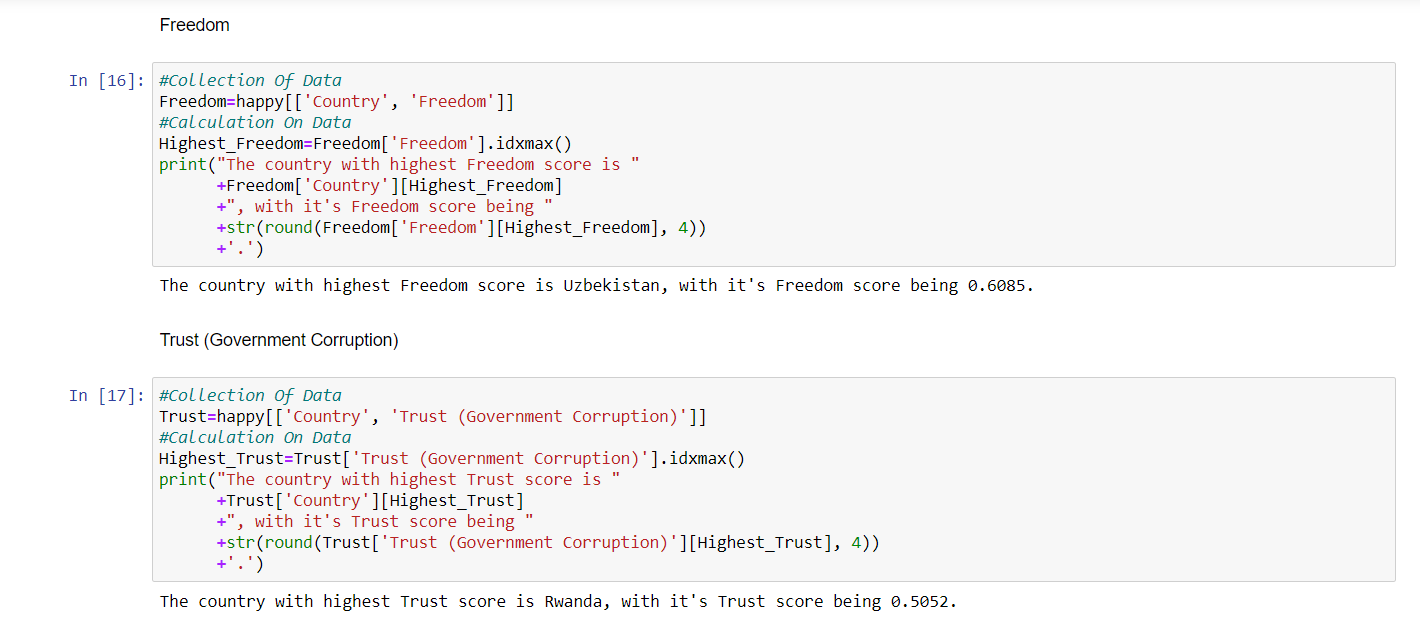


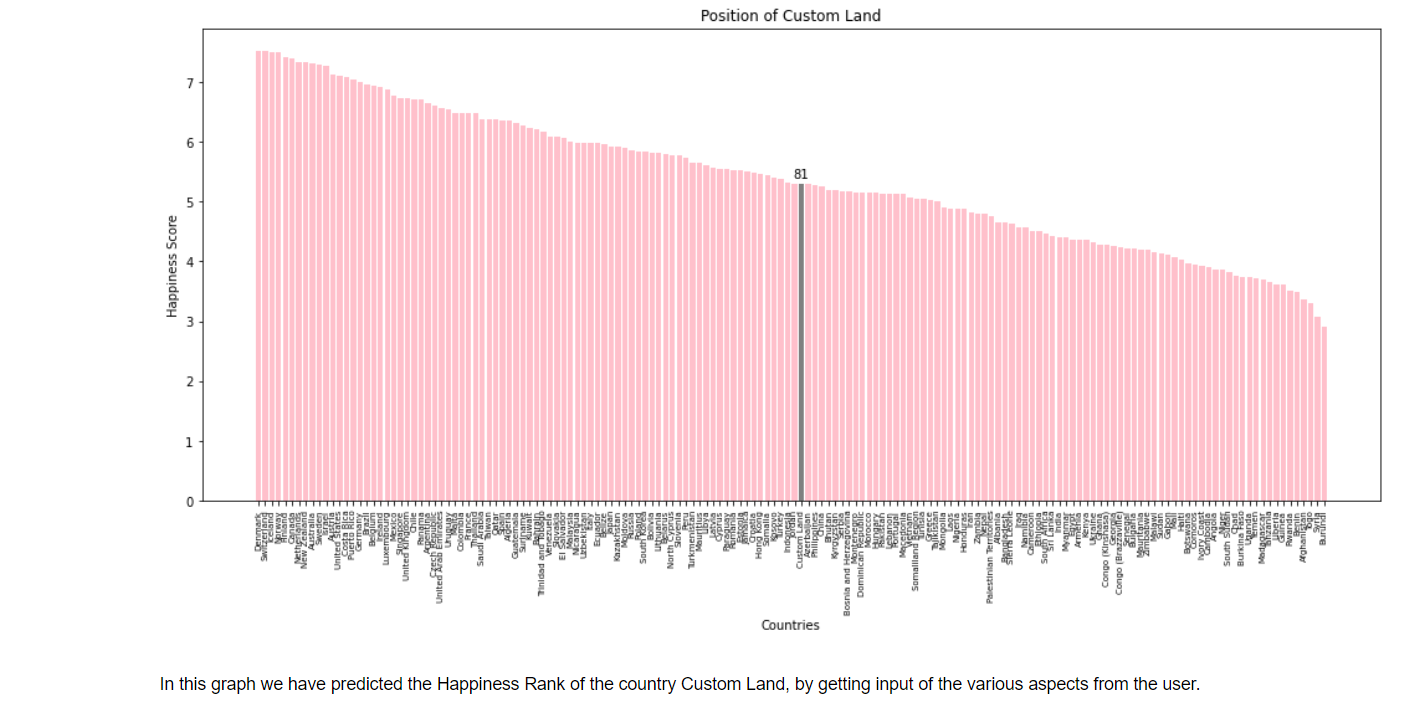
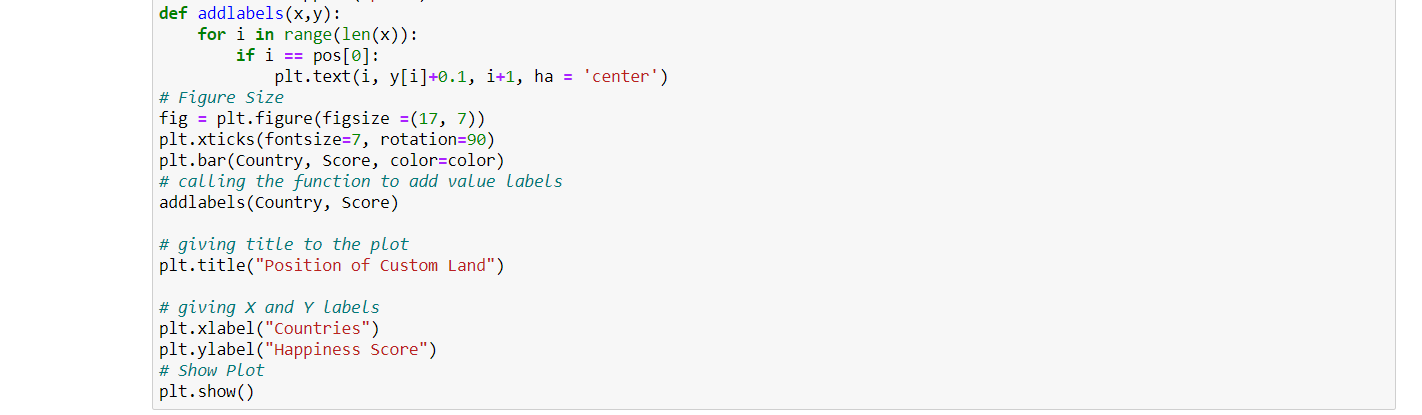
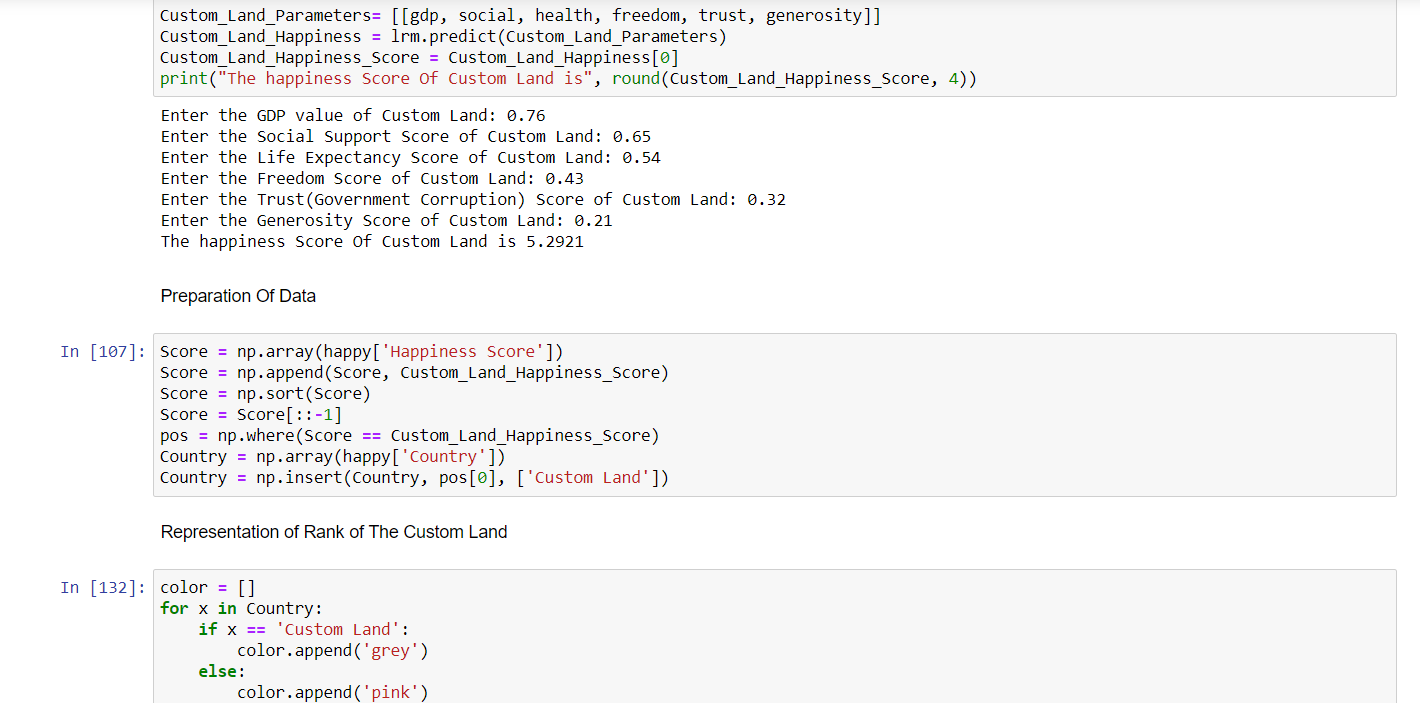
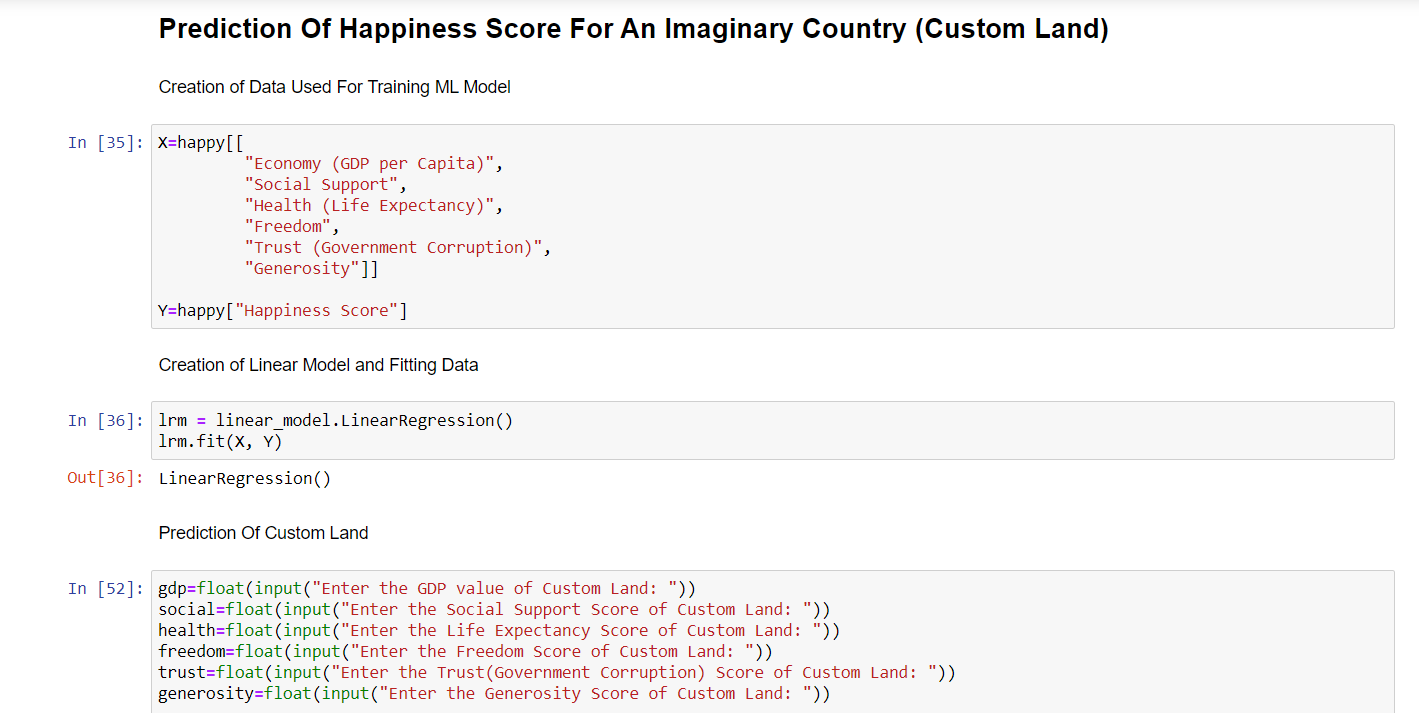
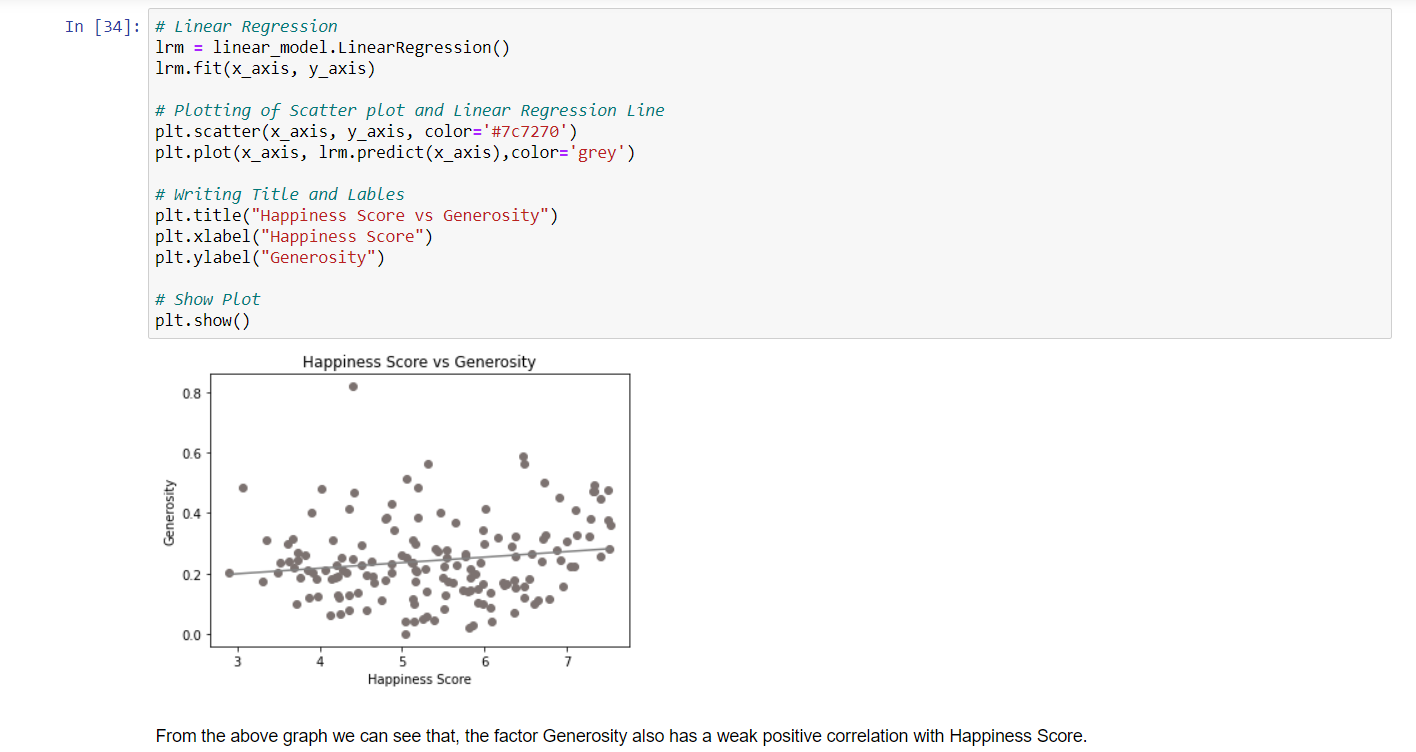
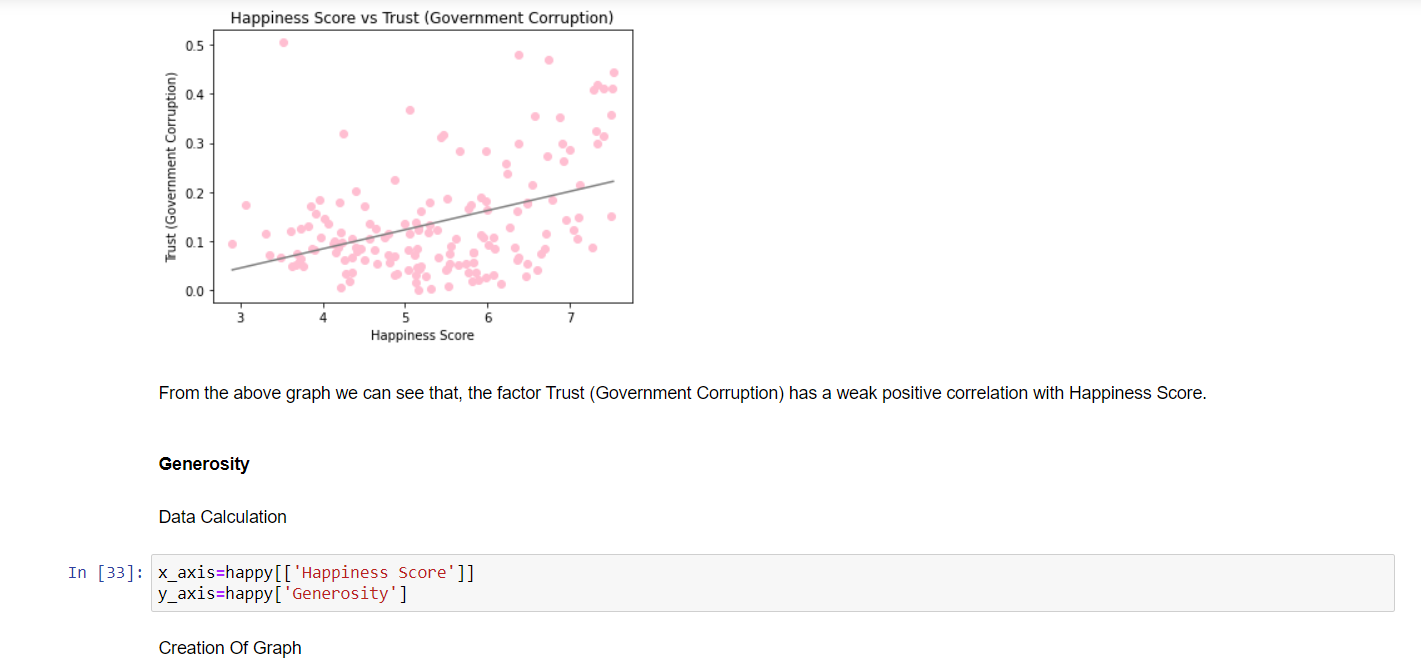
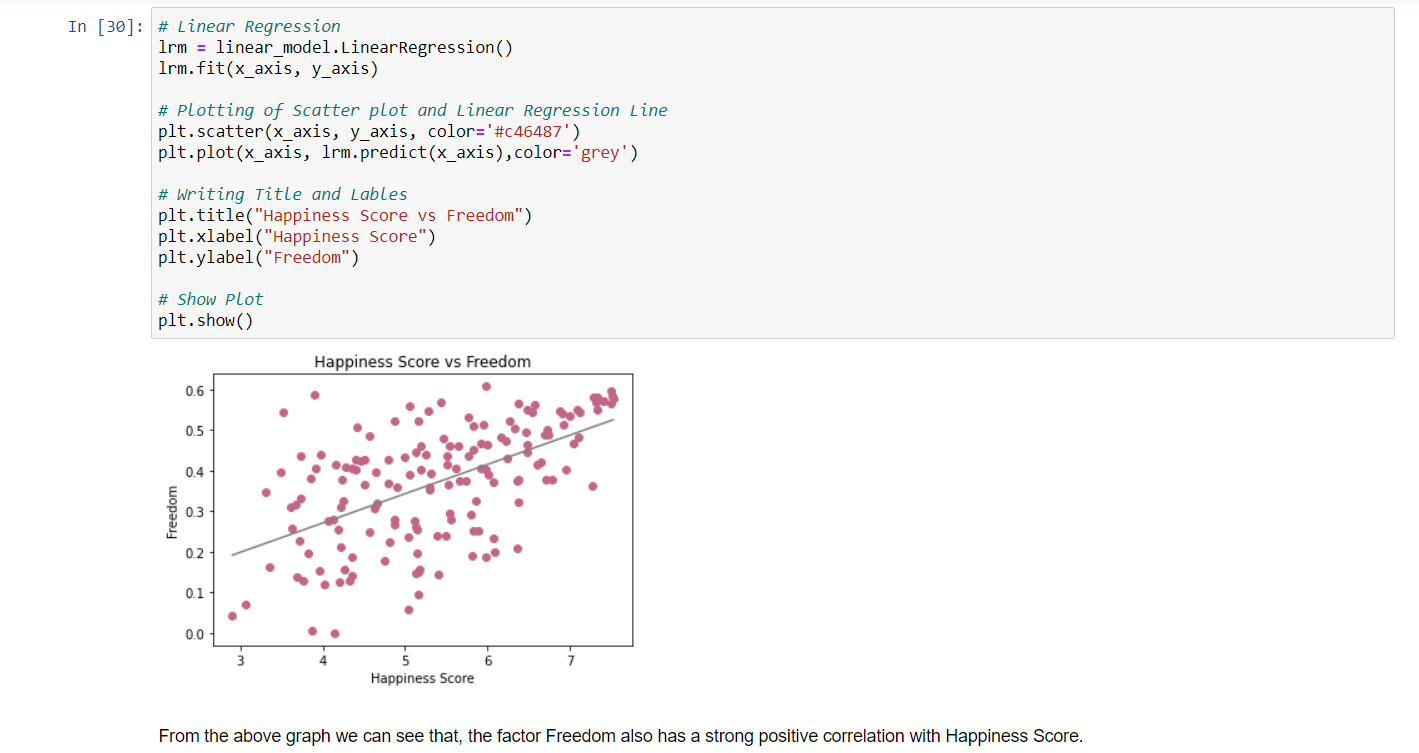
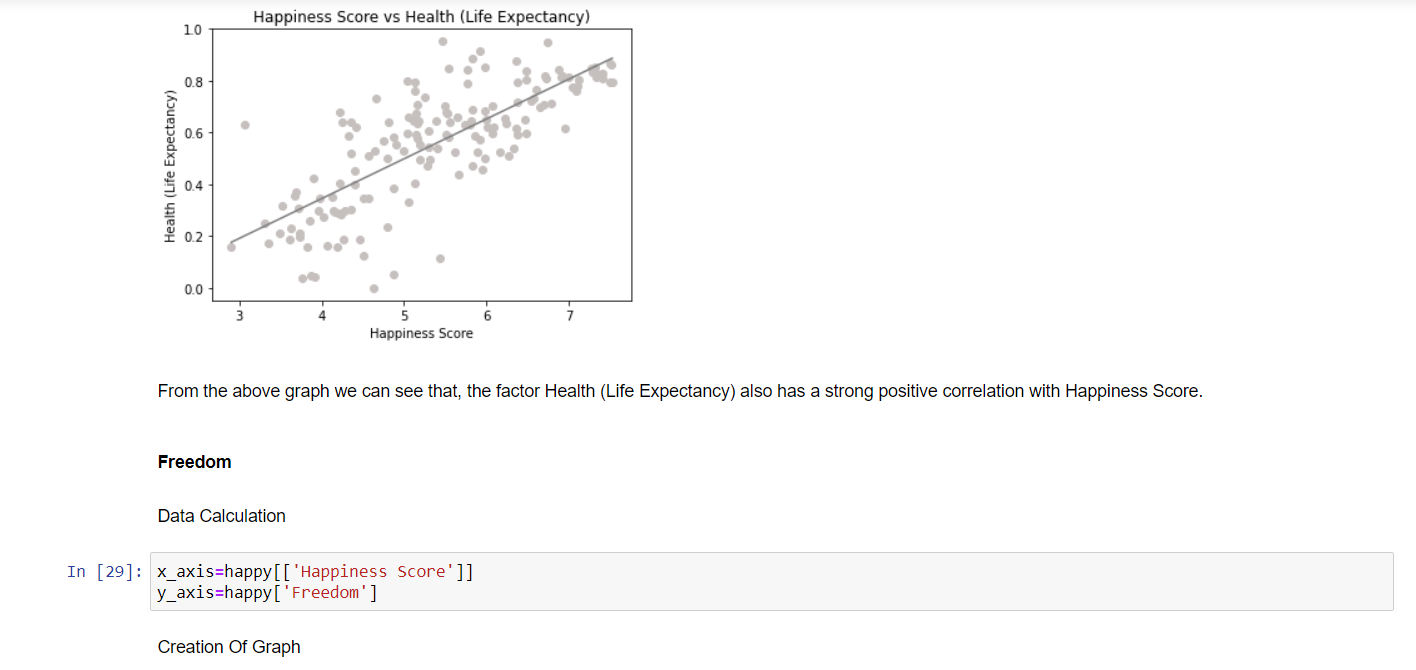
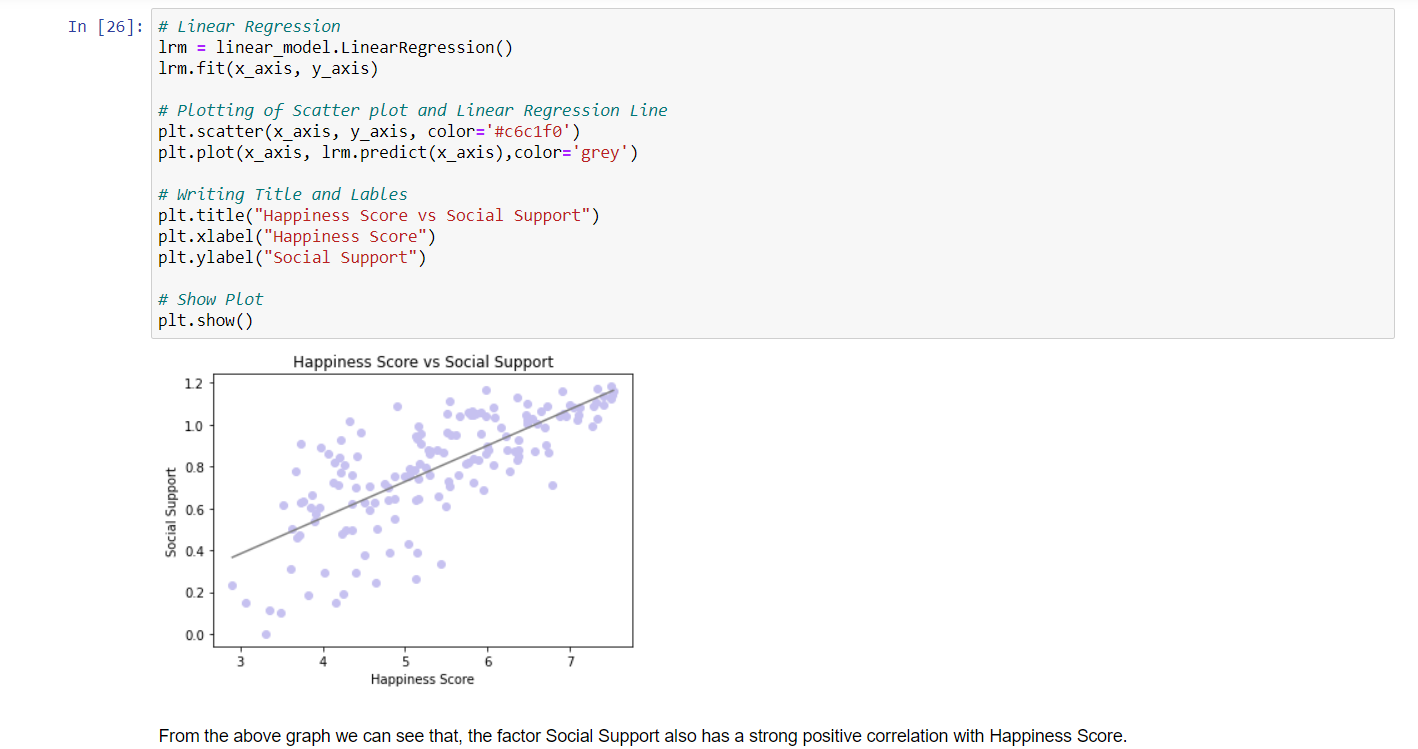
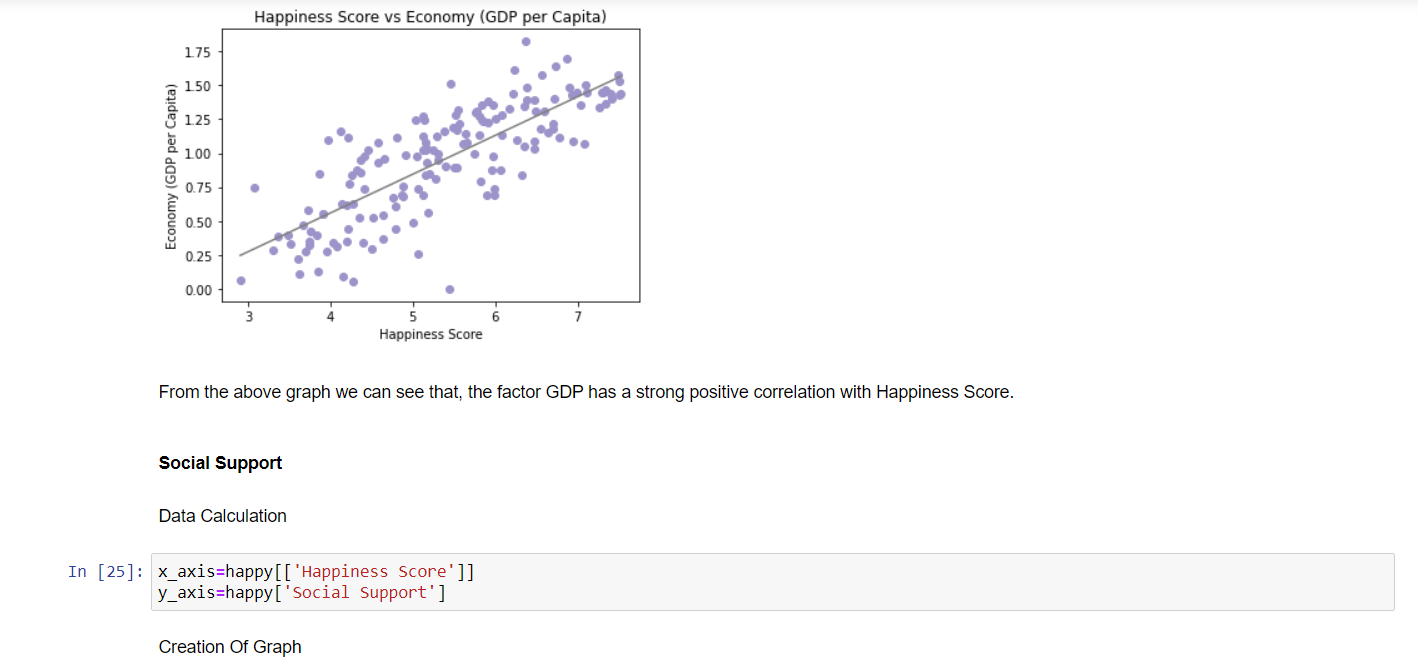
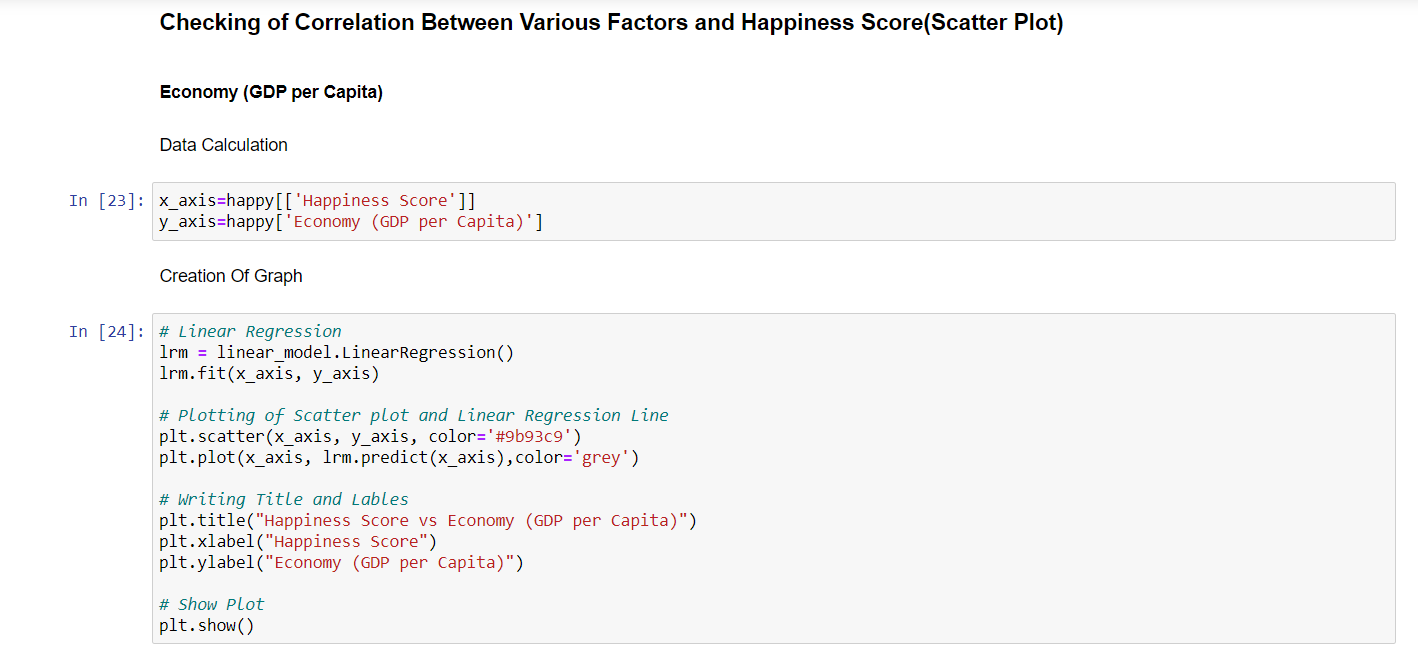
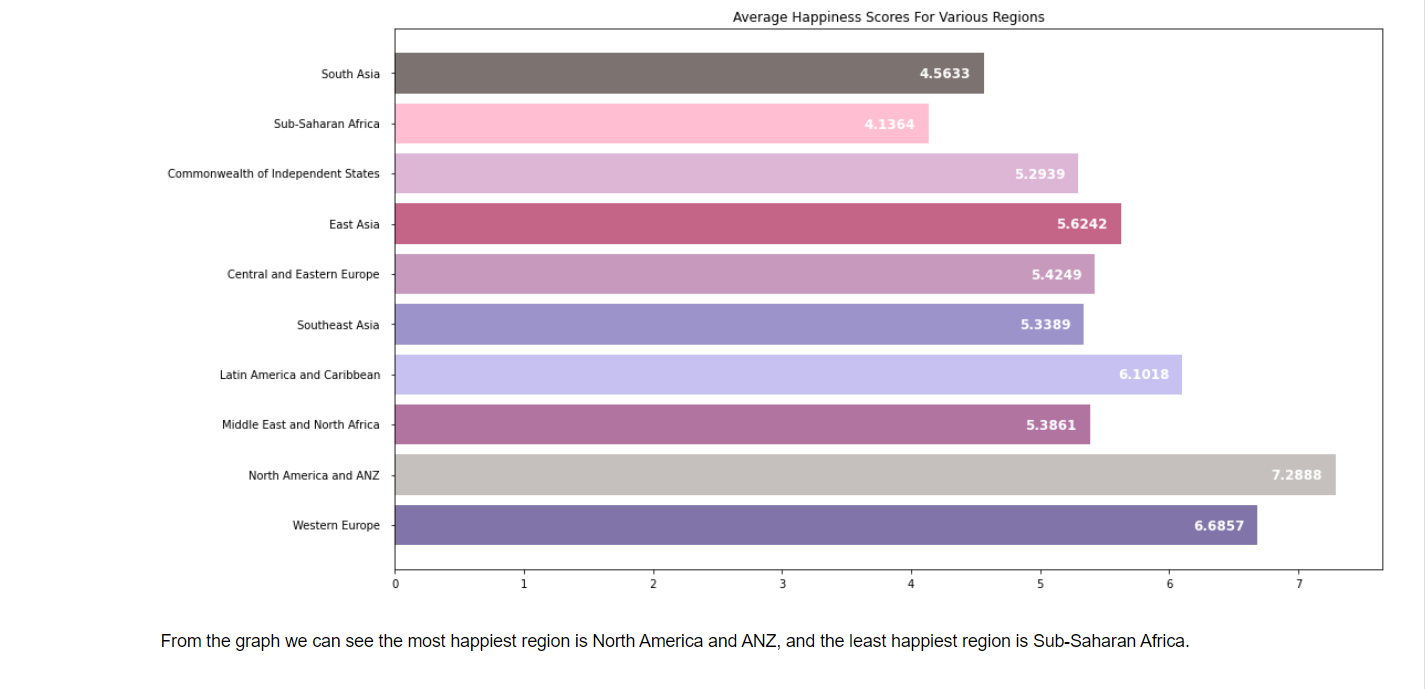
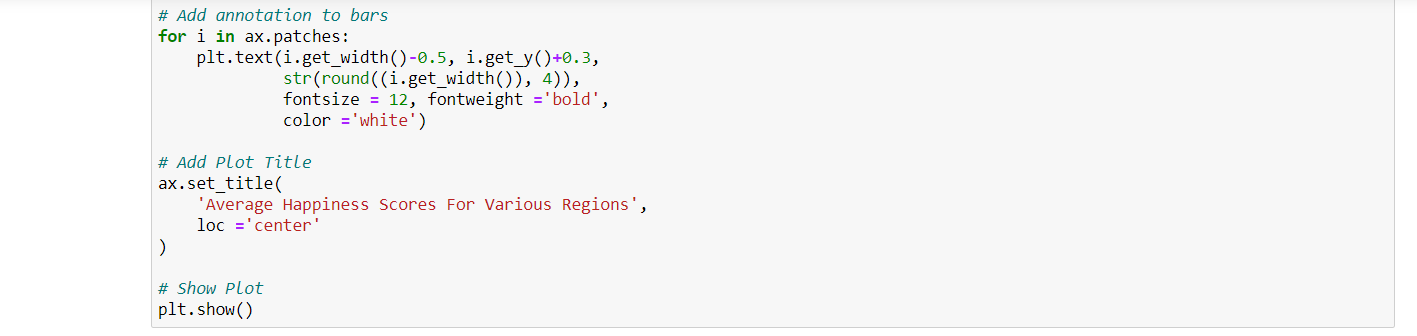
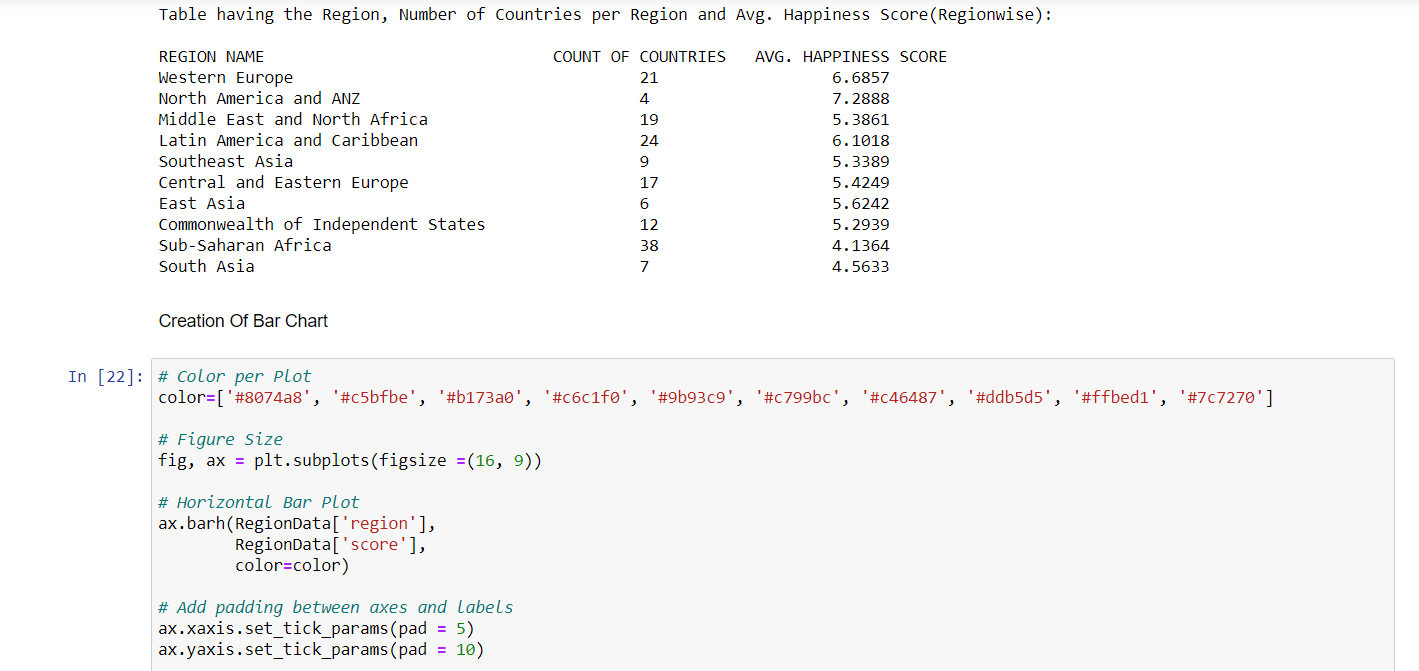
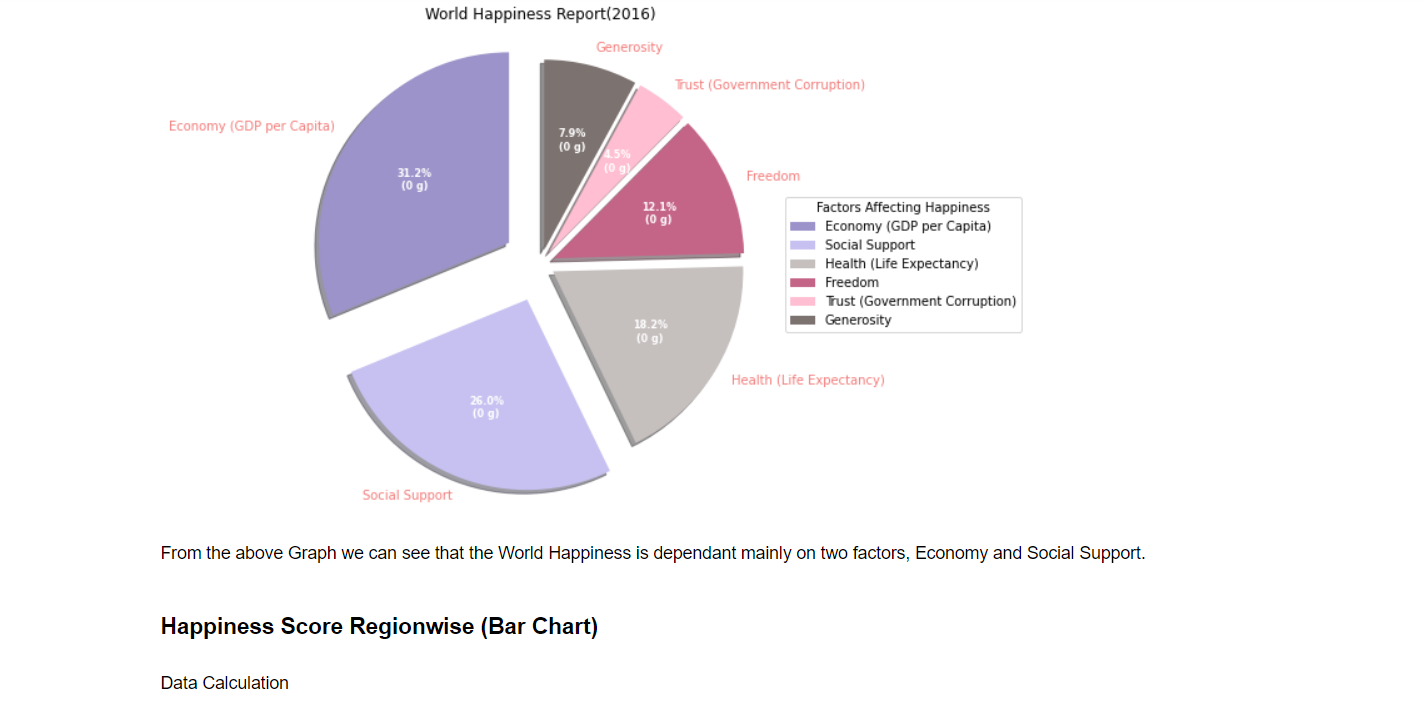
* 1. **Python**

****

****

****

****

****

* 1. **Website**

1. Home Page

****

1. About Dataset Page



1. Questions Page



1. Analysis Page Sample



1. **SAMPLE CODE**

**function App() {**

**return(**

**<Router>**

**<div>**

**<Header/>**

**<Switch>**

**<Route exact path='/'><Home/></Route>**

**<Route path='/about'><About/></Route>**

**<Route exact path='/analysis'><Question/></Route>**

**<Route path='/analysis/q0'><Q0/></Route>**

**<Route path='/analysis/q1'><Q1/></Route>**

**<Route path='/analysis/q2'><Q2/></Route>**

**<Route path='/analysis/q3'><Q3/></Route>**

**<Route path='/analysis/q4'><Q4/></Route>**

**<Route path='/analysis/q5'><Q5/></Route>**

**<Route path='/analysis/q6'><Q6/></Route>**

**<Route path='/analysis/q7'><Q7/></Route>**

**</Switch>**

**<Footer/>**

**</div>**

**</Router>**

**);**

**}**

1. **CONCLUSION**

The world happiness report depends on the average of scores that people gave, based on their opinions about how different factors affect their happiness. From the overall analysis of the World Happiness Report, we could come to a conclusion that the happiness scores vary based on different factors in different regions.

On the basis of the analysis that we made, we found that Economy(GDP per capita) and Social Support are the main factors on which happiness score depends on.

1. **TARGET AUDIENCE**

* People who are interested in learning about World Happiness Report
* People who want to know about how the Report is made and based on what factors scores are given
* People who want to know about how different factors influence the happiness of the country
* People who are learning data analysis
* Data analysis enthusiasts

1. **FUTURE SCOPE (Questions Which Can Be Addressed Later)**
2. What is the percentage of each factor, while taken for calculating the happiness score?
3. What impact does Social Support Score have on the Happiness Rank?
4. Is there any correlation between Social Support and Health?
5. How Covid has brought a major change in people’s Health and did that affect the Happiness Score? (If we get the 2021 year dataset!)
6. **REFERENCE**

* [**https://www.kaggle.com/mathurinache/world-happiness-report?selec t=2015.csv**](https://www.kaggle.com/mathurinache/world-happiness-report?select=2015.csv) **-** for Collecting Dataset
* [**https://en.wikipedia.org/wiki/World\_Happiness\_Report**](https://en.wikipedia.org/wiki/World_Happiness_Report)- for

Understanding the Dataset

* [**https://www.youtube.com/watch?v=aHaOIvR00So**](https://www.youtube.com/watch?v=aHaOIvR00So)- for Learning Tableau
* [**https://m.youtube.com/watch?v=CI2kX2EFhWc**](https://m.youtube.com/watch?v=CI2kX2EFhWc)- for Learning React JS
* [**https://www.npmjs.com/package/tableau-react-embed#usage**](https://www.npmjs.com/package/tableau-react-embed#usage) **-** for Integration of Tableau into Website
* [**https://www.datacamp.com/community/tutorials/git-push-pull**](https://www.datacamp.com/community/tutorials/git-push-pull) **-** for Using GitHub Repository
* [**https://youtu.be/lB4DTqMEumY**](https://youtu.be/lB4DTqMEumY) **-** for Learning Hosting in AWS