

The Criminalization of Abortion in Asia

**A Project Report submitted for the partial fulfillment of the requirements of the
Higher National Diploma in Data Science (Full Time) Programme.**

SENULI TARINYA LINARA RATNAYAKE

COHNDDSFT211F – 001

HIGHER NATIONAL DIPLOMA IN DATA SCIENCE

NATIONAL INSTITUTE OF BUSINESS MANAGEMENT

COLOMBO, SRI LANKA

30th MAY 2023

Declaration

I hereby declare that the work presented in this project report was carried out independently by myself and have cited the work of others and given due reference diligently.

Date: 30/05/2023

.....

Senuli Tarinya Linara Ratnayake

Signature of the Candidate

I certify that the above student carried out his/her project under my supervision and guidance.

Date:

.....

Thurairasa Balakumar

Signature of the Supervisor

Acknowledgement

I would like to express my profound gratitude to Mr. Thuraiasa Balakumar, the course director and my mentor, for his time and contributions to the completion of my project titled “The Criminalization of Abortion in Asia”. Your useful advice and suggestions were really helpful to me during the project’s completion. In this aspect, I am eternally grateful to you.

Executive Summary

This study aims at understanding the causes and consequences of the criminalization of abortion in Asia, and what can possibly aid the situation. The liberalization of criminal abortion laws is imperative to protect the health and equality of people. The determinants of unwanted pregnancies and the constraints against the liberalization of the abortion laws will be heavily explored. In theory, the conclusion should be that by merely raising awareness about abortion, it could lead to the introduction of proper sex education in schools in developing countries. Although, with further data analysis, more recommendations and solutions could be ventured.

Table of Contents

Declaration.....	ii
Acknowledgement	iii
Executive Summary	iv
List of Tables	vii
List of Figures	vii
Chapter 1: Introduction	1
1.1 Background.....	1
1.2 Research Problem	1
1.3 Objectives of the Project.....	2
1.4 Research Questions.....	2
1.5 Scope of the Research.....	2
1.6 Justification of the Research	3
1.7 Expected Limitations	3
Chapter 2: Literature Review	4
2.1 Introduction to the Research Theme	4
2.2 Theoretical Explanation about the Key Words in the Topic.....	4
2.3 Findings by other Researchers	5
2.4 The Research Gap	6
2.5 Table for Variables & their Definitions	6
Chapter 3: Data Preparation Process - Data Preprocessing & Data Wrangling.....	9
3.1 Data Cleaning.....	9
3.2 Data Integration	9
3.3 Data Transformation	10
3.4 Data Reduction.....	10
Chapter 4: Methodology	11
4.1 Introduction.....	11
4.2 Population, Sample and Sampling Technique	11
4.3 Type of Data Collected and Data Sources	11
4.4 Data Collection Tools and Plan	12
4.5 Methods, Techniques and Tools	12
Chapter 5: Data Analysis, Visualization and Interpretation.....	13
5.1 Data Analysis	13

5.2 Data Visualization.....	14
5.3 Analytics Models and Algorithms	16
5.3.1 Rural Areas vs Urban Areas.....	16
5.3.2 Poorest Wealth Quintile vs Richest Wealth Quintile.....	17
5.3.3 Formerly In Union vs In Union vs Never In Union	18
5.3.4 Total Pregnancies Ending in Abortion by Region vs by Legality.....	19
Chapter 6: Advanced Analysis.....	21
6.1 Regression Model	21
6.1.1 Exploratory Data Analysis.....	21
6.1.2 Feature Selection.....	22
6.1.3 Model Evaluation.....	24
Chapter 7: Discussion and Recommendations.....	26
7.1 Discussion.....	26
7.2 Recommendations.....	27
7.3 Conclusions.....	28
List of References	29
Appendix.....	31

List of Tables

- Table 2.5 – Table of variables and their definitions.
- Table 6.1.2 – Table displaying the feature importance in descending order.
- Table 7.1 – Table of reasons behind the request for abortion and their respective gestational time period.

List of Figures

- Fig 3.2.1 – Final dataset after undergoing data integration.
- Fig 3.3.1 – Original format of the variables in the dataset.
- Fig 3.3.2 – New variables of the dataset after undergoing data transformation.
- Fig 5.1.1 – Statistical description of all numerical variables in the dataset.
- Fig 5.2.1 – Pie chart representing the Abortion Laws around the World.
- Fig 5.2.2 – Scatter plot representing the average percentage of pregnancies ending in abortion in all regions across the world, by legality.
- Fig 5.2.2 – Scatter plot representing the average percentage of pregnancies ending in abortion in the five sub-regions of Asia, by legality.
- Fig 5.3.1 – Distribution of the percentage of women in rural [left] and urban [right] areas (with and without the use of contraception) by region, who want to avoid pregnancy.
- Fig 5.3.2 - Distribution of the percentage of women in the poorest [left] and richest [right] wealth quintiles (with and without the use of contraception) by region, who want to avoid pregnancy.
- Fig 5.3.3 - Distribution of the percentage of women who were formerly in union [top left], in union [top right] and never in union [bottom left], (with and without the use of contraception) by region, who want to avoid pregnancy.
- Fig 5.3.4.1 – Distribution of the percentage of total pregnancies ending in abortion, by region.
- Fig 5.3.4.2 – Distribution of the percentage of total pregnancies ending in abortion, by legality.

- Fig 6.1.1 – Categorical variables in the dataset have numeric representation.
- Fig 6.1.2 – Heatmap used to visualize the correlation between the variables.
- Fig 6.1.3.1 – Calculation of model accuracy on the training dataset.
- Fig 6.1.3.2 – Predicted vs Actual values plotted against each other.
- Fig 6.1.3.3 – Residuals vs Actual values plotted against each other.

Chapter 1: Introduction

1.1 Background

A global experience most women share is unwanted pregnancy, a fraction of these women seek to terminate the pregnancy by safe medical means or in most cases, by any means necessary. The termination of a pregnancy is known as an “abortion”.

The subject of abortion in Asia is considered taboo and isn’t discussed enough to raise awareness to the effects of criminalizing abortion. Unlike other health services, abortion is commonly monitored under criminal law, in addition to the regulation under health-care law. Out of the 48 countries in Asia, abortion is completely prohibited in Iraq, the Philippines and Laos. The majority of women in Asia live under unconventional abortion laws. Abortion laws in Asia reflect demographic trends, profound ethical convictions, strict religious and social beliefs, all of which out-weigh the possibility of decriminalizing abortion.

Criminalization contributes to the lower availability of health care professionals who are trained to provide safe abortions, which in turn has a negative effect on health workers who do provide abortion and can cause an increase in stricter protocol within health care establishments. The subject of this study is to investigate the reasons behind the act and criminalization of abortion, the effects of criminalizing and the possibility of decriminalizing abortion, as abortion is a fundamental human right.

1.2 Research Problem

This study aims at understanding the causes and consequences of criminalizing abortion in Asia. The focused sample in this study is Asia and its five sub-regions (Central Asia, East Asia, South Asia, Southeast Asia, and Western Asia). Abortion laws and policies are significantly more restrictive in developing countries, which lead to unsafe measures to expel the foetus. At the rare instance that a safe abortion is induced at the request of the

woman, the chances of survival are still at risk. The primary aim of this study is to investigate what can be done to prevent such risks and how progressive changes could be made to result in decriminalizing abortion.

1.3 Objectives of the Project

The main objectives of this project are to utilize and process the chosen dataset, to identify the possible relationships between variables to prove its correlation to why abortion is criminalized to begin with and which factors carry the weight of the cause. Societal attitudes toward abortion will be examined along with the growth and development of societal norms in general. The data will be used to visualize the comparison of abortion rates in the sub-regions of Asia and analyse the repercussions of the strict laws, majority of the countries in Asia need to abide by. The analysis will enable a better understanding of the severity of the growing situation.

1.4 Research Questions

The criminalization of abortion in Asia raises a number of questions regarding the factors responsible for it and the consequence that follows. This research poses the same questions. Why does a pregnancy lead to abortion? What is a socially accepted reason for abortion? What are the risks of inducing an abortion, let alone an unsafe abortion? In what ways can society collectively progress? Should abortion be legalized? Questions such as these will be explored in this paper.

1.5 Scope of the Research

Though the criminalization of abortion is the primary subject of this research, investigating the varying degrees of criminal law under which abortion is regulated globally, would almost overlook the underlying issue that most women in developing countries still face. Therefore, the focused group in this research is Asia (examples of cases in developed countries may be used for comparison). However, the research does

not investigate the laws in each sub-region in depth, instead on what grounds abortion is legal, if at all. The laws vary and the conditions for Abortion on Request are as follows: to save a woman's life, no restriction, varies by state, on social or economic grounds, on request (gestational limits vary), to preserve physical/mental health, prohibited altogether.

1.6 Justification of the Research

The ultimate goal of this research is to raise awareness and help shed light on a situation, that is overlooked by so many who are oblivious to the effects of criminalizing a basic human right. The mere fact that legal abortions on request that are allowed in some countries in Asia, must be justified, should be reason enough to raise concern. With the newer generations showing interest in progressing and being more open-minded, it is important to strive for positive changes and changing conservative attitudes on abortion.

1.7 Expected Limitations

It is important to work with updated data as it plays an important role in displaying the progress and helps convey the reasoning behind the research. As mentioned, the topic of abortion and all the subjects that surround it are barely ever discussed with any sense of importance. Out of the 195 countries in the world, only records of 163 countries were available in the dataset used. Since abortion is legally restricted in many countries, it makes it challenging to measure and document. Therefore, there was a deficit in relevant updated datasets or information from verified sources. A few examples used from the recent years are sourced from social media, as there was an insufficient amount of resources.

Chapter 2: Literature Review

2.1 Introduction to the Research Theme

A research theme is meant to express and outline the main goals of the research. The theme in this study mainly revolves around “social, cultural, environmental, and population research”. Although the access to legitimate records on abortion is restricted, the purpose is to perform a thorough analysis with the data available to gain useful insight. Therefore, majority of the work referenced in this study have proven to be extremely resourceful.

2.2 Theoretical Explanation about the Key Words in the Topic

The following definitions are referenced from Oxford Languages:

Abortion - The deliberate termination of a human pregnancy, most often performed during the first 28 weeks of pregnancy.

Contraceptives - A device or drug serving to prevent pregnancy.

Criminalization - The act of turning an activity into a criminal offence by making it illegal.

Decriminalization - The action or process of ceasing to treat something as illegal or as criminal offence.

Female-Selective Abortion (FSA) - It is the practice of terminating a pregnancy based upon the predicted sex of the infant.

Gestational time period - This prescribes the time period of developing inside the womb between conception and birth.

Liberalization - The removal or loosening of restrictions on something, typically an economic or political system.

Low- and middle-income countries (LMICs) - The countries that are classified by the World Bank as having a gross national income per capita of \$1,085 or less in 2021, for the current 2023 fiscal year.

Miscarriage - The unplanned expulsion of a foetus from the womb before it is able to survive independently.

Stillbirth - The birth of an infant that has died in the womb, after having survived at least the first 28 weeks of pregnancy.

2.3 Findings by other Researchers

The Guttmacher Institute published *Facts on abortion in Asia* in 2015, it states that in 2014 at least 6% (approximately 5,400 deaths) of all maternal deaths were due to unsafe abortions.

In 2019, with reference to *Fighting for Safe Abortion Access in Sri Lanka*, Gunasekera. S. was interviewed about access to vital healthcare services in Sri Lanka and its impact. She concludes that the discrimination against women who seek medical help, causes them to be discouraged and can be extremely difficult and stressful when healthcare professionals disregard their medical needs.

A recent podcast from Nikkei Asia (2022) called “*Asia Stream: Abortion in Asia -- Beyond Taboo*” that a few of the most unsafe places in Asia to get an abortion were in South Central Asia, which include India, Pakistan, Bangladesh, Afghanistan, etc.

A demographic fact sheet published in 2014, *Selective Abortion in Asia*, in which it states that selective abortion in Asia is used as a means to avoid having a girl child (due to the continuing preference for boys). Hence, regardless of whether the access to abortion is legal or illegal, it socially accepted to do so and is still practiced.

Center for Reproductive Rights is a global human rights organization. *The World's Abortion Laws* is “the definitive record of the legal status of abortion in countries across the globe, updated in real time”, which states that legal restrictions on abortion do not

result in fewer abortions, instead it forces women to put their lives at risk in order to seek unsafe or “back-door” abortions.

The Council of Foreign Relations (from Women & Foreign Policy Program) notes that between the years 1990-94 and 2015-19, the average abortion rate in countries with generally legal abortion declined by 43% (excluding India and China). It was also found that, the average abortion rate increased by approximately 12% in countries with severe restrictions on abortion, in comparison to the rest.

Abortion in Asia: The limits of choice by Nikkei Asia highlights how restrictions on abortion are more severe in oppressed countries with strict religious and social beliefs, where although abortion maybe illegal, back street (unsafe) abortions are very common.

In 2021 Kamdar. B. writes that the lack of access to safe and legal abortions and contraceptives is a leading reason for the high maternal mortality in South Asia. Other barriers such as lack of awareness of the legal status of abortion and the fear of being discriminated also poses as an issue. For example, the article also states that despite the legalization of abortion since 2002, only four in ten women in Nepal are aware that abortion is legal.

2.4 The Research Gap

Due to the time restrictions, this study does not contain any records of primary sourced data. With reference to the published literature from section 2.3 and the dataset used for this research, although resourceful, there is still a lack of accurate and updated data required to perform a data analysis from a new perspective. Using the available data and raising awareness is the attempt at bridging this research gap.

2.5 Table for Variables & their Definitions

The data in the chosen dataset ranges from the years 2015 to 2019 and the control (a control is an element that remains unchanged or unaffected by other variables) that has

been applied to the entire dataset is the age group of women (15 - 49), all of whom want to avoid pregnancy.

Variable	Definition
Country	The specific country in which data was collected from.
Region	The region that each country is situated in.
% rural, no contraception	Among women from rural areas, the percentage of cases where no method of contraception was used.
% rural, with contraception	Among women from rural areas, the percentage of cases where a traditional method of contraception was used.
% urban, no contraception	Among women from urban areas, the percentage of cases where no method of contraception was used.
% urban, with contraception	Among women from urban areas, the percentage of cases where a traditional method of contraception was used.
% formerly in union, no contraception	Among women formerly in union, the percentage where no method of contraception was used.
% formerly in union, with contraception	Among women formerly in union, the percentage where a traditional method of contraception was used.
% poorest wealth quintile, no contraception	Among the women in the poorest wealth quintile, the percentage where no method of contraception was used.
% poorest wealth quintile, with contraception	Among the women in the poorest wealth quintile, the percentage where a traditional method of contraception was used.
% richest wealth quintile, no contraception	Among the women in the richest wealth quintile, the percentage where no method of contraception was used.
% richest wealth quintile, with contraception	Among the women in the richest wealth quintile, the percentage where a traditional method of contraception was used.
% in union, no contraception	Among the women in union, the percentage where no method of contraception was used.
% in union, with contraception	Among the women in union, the percentage where a traditional method of contraception was used.

% never in union, no contraception	Among the women never in union, the percentage where no method of contraception was used.
never in union, with contraception	Among the women never in union, the percentage where a traditional method of contraception was used.
% avg pregnancies ending in abortion	The average annual percentage of all pregnancies that ended in abortion.
% avg unintended pregnancies ending in abortion	The average annual percentage of all unintended pregnancies that ended in abortion.
Legality	The state of abortion being in accordance with the law, by country.

Table 2.5 – Table of variables and their definitions.

Chapter 3: Data Preparation Process - Data Preprocessing & Data Wrangling

3.1 Data Cleaning

In the chosen dataset the values “u” are denoted by “unavailable”, due to the lack of updated records. Although these values aren’t recognized as null values, the average value of the column cannot be used to fill these values, as each record vastly differs and would defeat its purpose when used for comparisons, since the average abortion rate diverges between countries with fewer restrictions and those with more.

3.2 Data Integration

Two datasets were used for this research, one with the statistics and the other with the abortion laws that applied to each country. The datasets were merged into a single dataset and was used for the data analysis.

	Country	Region	% rural, no contraception	% rural, with contraception	% urban, no contraception	% urban, with contraception	% formerly in union, no contraception	% formerly in union, with contraception	% poorest wealth quintile, no contraception	% poorest wealth quintile, with contraception	% richest wealth quintile, no contraception
0	Afghanistan	South Asia	55.0	4.0	39.0	9.0	55.0	1.0	61.0	2.0	35.0
1	Albania	Central and Eastern Europe	28.0	65.0	27.0	63.0	61.0	26.0	32.0	60.0	20.0
2	Algeria	North Africa	15.0	9.0	14.0	11.0	100.0	0.0	16.0	7.0	13.0
3	Angola	Southern Africa	92.0	2.0	56.0	3.0	50.0	1.0	94.0	1.0	39.0
4	Argentina	South America	20.0	6.0	12.0	4.0	15.0	2.0	19.0	6.0	8.0

% richest wealth quintile, with contraception	% in union, no contraception	% in union, with contraception	% never in union, no contraception	% never in union, with contraception	% avg pregnancies ending in abortion	% avg unintended pregnancies ending in abortion	Legality
9.0	51.0	5.0	17.0	11.0	0.0	0.0	To save a woman's life
70.0	27.0	65.0	24.0	55.0	13.0	59.0	No restriction
11.0	13.0	10.0	46.0	8.0	0.0	0.0	To preserve physical/mental health
3.0	70.0	3.0	52.0	2.0	13.0	28.0	Prohibited altogether
2.0	13.0	5.0	11.0	1.0	28.0	48.0	To preserve physical health

Fig 3.2.1 – Final dataset after undergoing data integration.

3.3 Data Transformation

The downloaded dataset contained very detailed and lengthy variable names.

	Country	Region	Among rural women aged 15–49 who want to avoid pregnancy, % not using any method of contraception, 2019 [1]	Among rural women aged 15–49 who want to avoid pregnancy, % using a traditional method of contraception, 2019 [2, 1]	Among urban women aged 15–49 who want to avoid pregnancy, % not using any method of contraception, 2019 [1]	Among urban women aged 15–49 who want to avoid pregnancy, % using a traditional method of contraception, 2019 [2, 1]	Among women aged 15–49 formerly in union who want to avoid pregnancy, % not using any method of contraception, 2019 [1]	Among women aged 15–49 formerly in union who want to avoid pregnancy, % using a traditional method of contraception, 2019 [2, 1]	Among women aged 15–49 in the poorest wealth quintile who want to avoid pregnancy, % not using any method of contraception, 2019 [1]	Among women aged 15–49 in the poorest wealth quintile who want to avoid pregnancy, % using a traditional method of contraception, 2019 [2, 1]	Among women aged 15–49 in the richest wealth quintile who want to avoid pregnancy, % not using any method of contraception, 2019 [1]
0	Afghanistan	South Asia	55	4	39	9	55	1	61	2	
1	Albania	Central and Eastern Europe	28	65	27	63	61	26	32	60	

Fig 3.3.1 – Original format of the variables in the dataset.

For convenience during the data analysis phase, the variable names were shortened.

	Country	Region	% rural, no contraception	% rural, with contraception	% urban, no contraception	% urban, with contraception	% formerly in union, no contraception	% formerly in union, with contraception	% poorest wealth quintile, no contraception	% poorest wealth quintile, with contraception	% richest wealth quintile, no contraception
0	Afghanistan	South Asia	55	4	39	9	55	1	61	2	35
1	Albania	Central and Eastern Europe	28	65	27	63	61	26	32	60	20
2	Algeria	North Africa	15	9	14	11	100	0	16	7	13
3	Angola	Southern Africa	92	2	56	3	50	1	94	1	39
4	Argentina	South America	20	6	12	4	15	2	19	6	8

Fig 3.3.2 – New variables of the dataset after undergoing data transformation.

Apart from the variable names being changed, the “u” values were replaced with 0’s, using Microsoft Excel, in order for the values to be identified as integer or float values, instead of string values, when carrying out the data analysis.

3.4 Data Reduction

Since the focused sample is Asia and its sub-regions, majority of the data analysis was done using only the respective data, however, none of the data was particularly reduced, but was sampled.

Chapter 4: Methodology

4.1 Introduction

The methods of theoretical research comprise of the use of case histories and document analysis. The data for this project was obtained using secondary data collection methods, through social media monitoring and online tracking, via authentic sites where the data is made available for public use.

4.2 Population, Sample and Sampling Technique

In order to obtain statistically accurate results, the target population is abortion in Asia, more specifically its 5 sub-regions: Central Asia, East Asia, South Asia, Southeast Asia, and Western Asia. The standardized data generated help gather information from a representative sample of the target population. The sampling techniques used, narrow down the demographic groups which help gather more accurate results to put all the findings into perspective. The types of sampling used consist of convenience, stratified and cluster sampling.

4.3 Type of Data Collected and Data Sources

The main dataset used for this research was obtained from the Guttmacher Data Center, which helps source data selectively from the Guttmacher Institute and other trusted sources. The dataset consists of 18 variables (2 categorical, 16 numerical) that were specifically chosen to fit the theme of this paper and are mentioned in detail in section 2.5 of the report. The second dataset that will be merged with the main dataset, is important to show the correlation between the legality of abortion according to the country and the statistics presented in the first dataset. The mentioned datasets were sourced from:

<https://data.guttmacher.org/regions>

<https://worldpopulationreview.com/country-rankings/countries-where-abortion-is-illegal>

4.4 Data Collection Tools and Plan

Both qualitative and quantitative data were collected with the use of documents and records, case studies, surveys that were used in articles, etc.

A thorough exploratory data analysis will be conducted to explore, and be familiarized with, the contents of the dataset. The conditions applied for the analysis will be useful when answering the research questions posed in section 1.4 of the proposal. After the appropriate methodology is applied, the findings will be interpreted in context of the problem.

4.5 Methods, Techniques and Tools

The techniques and statistical models that will be used to analyze and visualize the data include descriptive (graphical techniques and numeric measures), and inferential (correlation and regression analysis) statistics. Both descriptive and inferential statistic methods will be applied with the use of the multiple libraries embedded in Python and R (ggplot2, matplotlib, seaborn, numpy, pandas, etc.).

The advanced statistical analysis is performed by applying supervised machine learning concepts like linear regression, with the help of another important library embedded in Python known as scikit-learn. It offers functions that enable users to evaluate regression models with three main metrics: Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Squared Error (RMSE).

Chapter 5: Data Analysis, Visualization and Interpretation

5.1 Data Analysis

	% in union, no contraception	% in union, with contraception	% never in union, no contraception	% never in union, with contraception	% avg pregnancies ending in abortion	% avg unintended pregnancies ending in abortion
	35.00	35.00	35.00	35.00	35.00	35.00
	17.71	8.66	27.97	9.34	15.29	40.94
	15.70	8.50	23.73	8.47	13.29	33.76
	0.00	0.00	0.00	0.00	0.00	0.00
	2.50	1.00	3.50	1.00	0.00	0.00
	18.00	5.00	28.00	9.00	16.00	61.00
	22.00	16.50	38.00	13.00	27.50	68.50
	60.00	25.00	87.00	29.00	34.00	83.00

	% rural, no contraception	% rural, with contraception	% urban, no contraception	% urban, with contraception	% formerly in union, no contraception	% formerly in union, with contraception	% poorest wealth quintile, no contraception	% poorest wealth quintile, with contraception	% richest wealth quintile, no contraception	% richest wealth quintile, with contraception
count	35.00	35.00	35.00	35.00	35.00	35.00	35.00	35.00	35.00	35.00
mean	18.51	7.34	16.54	10.26	15.74	2.37	21.66	6.06	15.69	11.34
std	16.71	7.73	14.15	9.71	20.83	3.02	19.49	6.73	13.54	10.61
min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25%	3.50	0.50	2.50	2.00	0.00	0.00	3.00	0.50	2.50	2.00
50%	18.00	5.00	16.00	8.00	10.00	1.00	20.00	3.00	15.00	9.00
75%	22.50	11.50	23.00	20.00	23.00	4.00	28.50	8.50	22.00	22.50
max	61.00	26.00	59.00	34.00	100.00	12.00	69.00	20.00	58.00	34.00

Fig 5.1.1 – Statistical description of all numerical variables in the dataset.

As mentioned in section 2.5 of the proposal, the data in the chosen dataset ranges from the years 2015 to 2019, and the control that has been applied to the entire dataset is the age group of women (15 - 49), all of whom want to avoid pregnancy, and all the numerical variables have been measured as a percentage. The focused population of the dataset is Asia, unless mentioned otherwise. From all the experimental variables, the highest and lowest mean, with a value of 27.97% and 2.37% respectively, belong to the portion of women that have never been in union (no contraception) and the women who have formerly been in union (with contraception). The average values of unintended (40.94%) and normal (15.29%) pregnancies that end in abortion, are inclusive of all types

of cases, although if the count (35) is taken into account, the values could be more accurate with a larger population.

5.2 Data Visualization

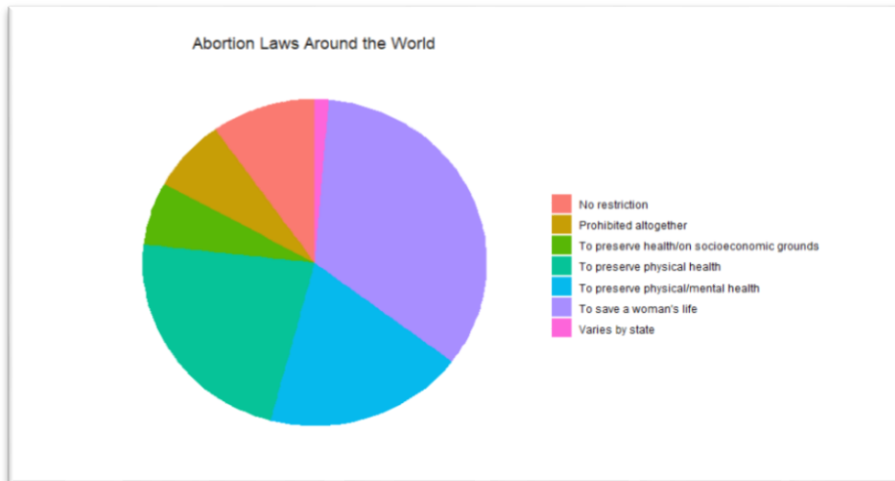


Fig 5.2.1 – Pie chart representing the Abortion Laws around the World.

The abortion laws around the world, that have been included in this dataset, are: (1) No restriction, (2) Prohibited altogether, (3) To preserve health/on socioeconomic grounds, (4) To preserve physical health, (5) To preserve physical/mental health, (6) To save a woman's life, (7) Varies by state. Evidently, the majority is to save a woman's life, while the minority varies by state.

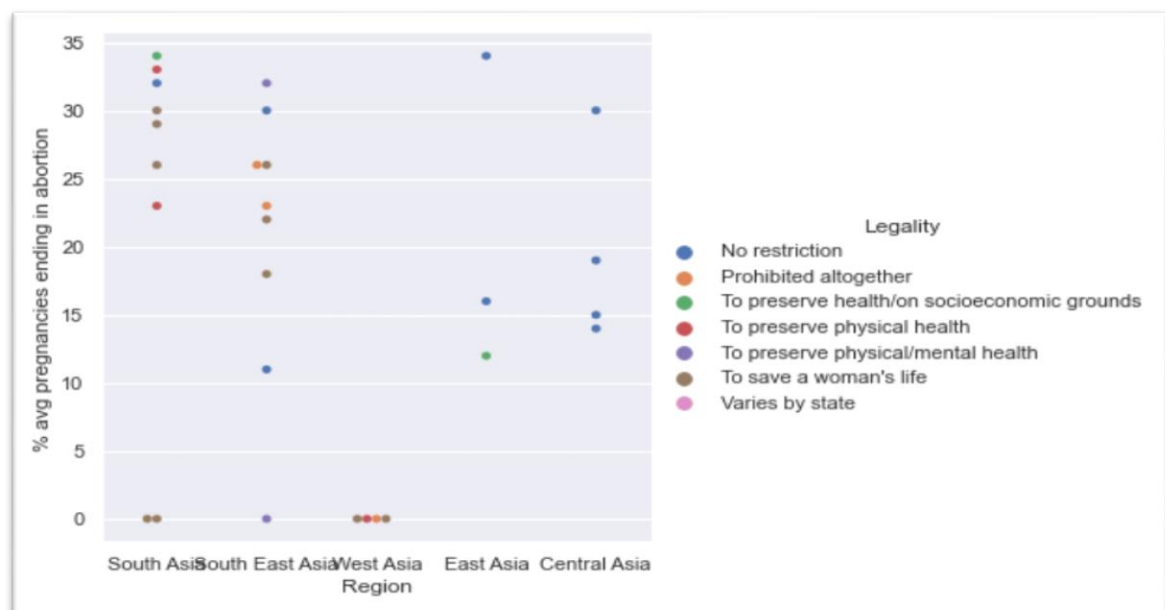


Fig 5.2.3 – Scatter plot representing the average percentage of pregnancies ending in abortion in the five sub-regions of Asia, by legality.

The scatter plot displays the individual Asian countries by region and legality. Most of the points are represented by the “No restriction” and “To save a woman’s life” laws, where majority falls under Central Asia and South East Asia, respectively. While East and West Asia contain countries where abortion is “Prohibited altogether”. As displayed, abortions take place regardless of the grounds of abortion. In actuality the problem lies in the measures that can be taken to prevent the unsafe abortions.

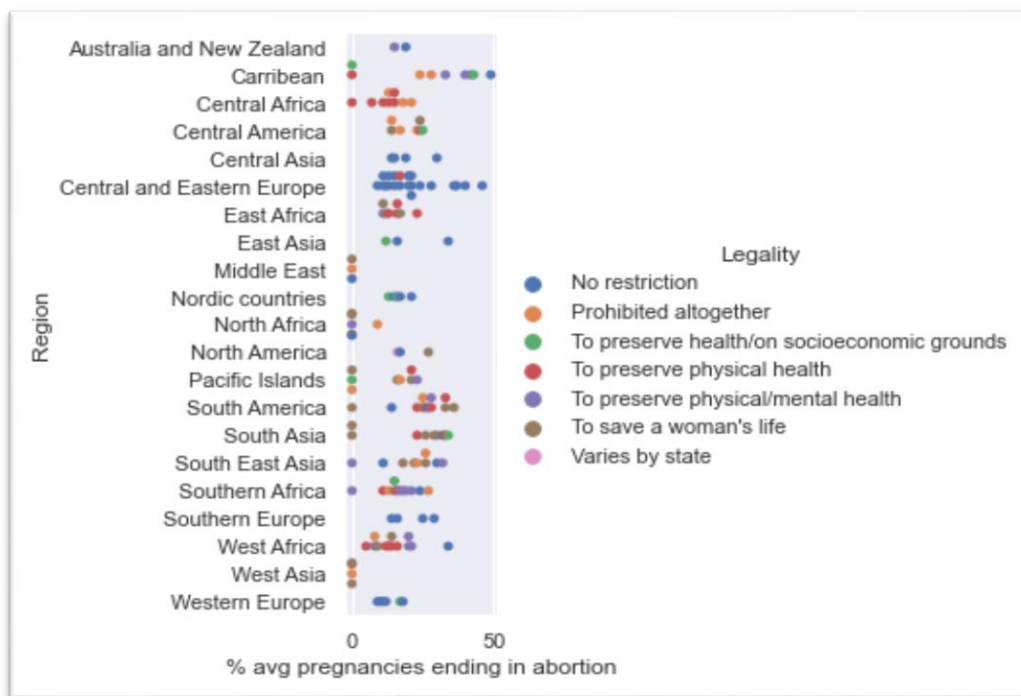


Fig 5.2.2 – Scatter plot representing the average percentage of pregnancies ending in abortion in all regions across the world, by legality.

In reference to the plot above, majority of the blue points are represented by countries that are in different regions of Europe, which abide by the “No restriction” abortion law. Which helps understand that abortion is legal in most European countries, where most of the capitalist and industrialized first world countries are situated in Europe. In contrast, there is no fixed trend in the rest of the laws dispersed among all the regions in the world.

5.3 Analytics Models and Algorithms

To understand which external factors are of more importance, and affect the average percentage of pregnancies and unintended pregnancies ending in abortion in different regions, the distributions of similar groups will be analysed.

5.3.1 Rural Areas vs Urban Areas

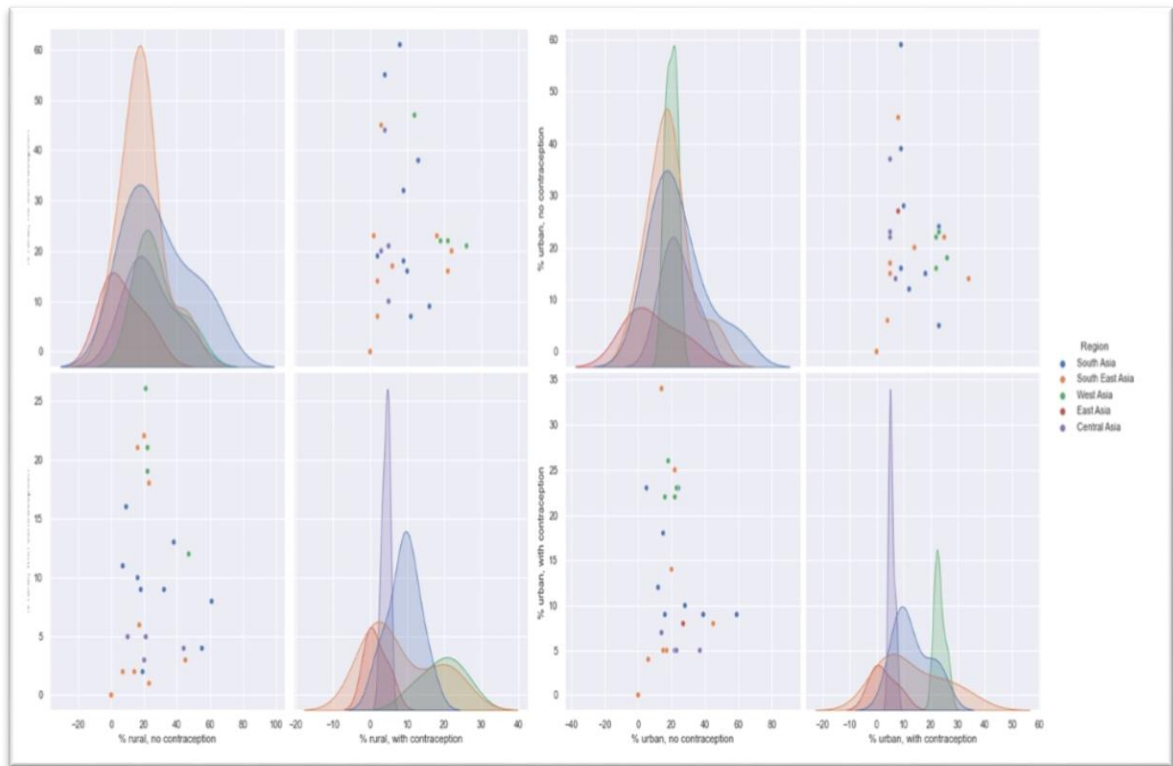


Fig 5.3.1 – Distribution of the percentage of women in rural [left] and urban [right] areas (with and without the use of contraception) by region, who want to avoid pregnancy.

The pairplot displays a pairwise relationship between the rural and urban population of the data, by region. The distributions of both areas where contraceptives have been used are quite similar, where Central Asia peaks at an average of less than 35%. Whereas, in rural areas with no contraception, majority is represented by Southeast Asia with an average of 60%. The lowest amount of abortions in both rural and urban areas, with and without contraception, belong to East Asia with an average of less than 10%. Amongst some Southeast Asian countries, abortion laws can be divided into two groups – (1) Indonesia, Brunei, Cambodia, Malaysia, Thailand, Singapore and Myanmar can have

legal abortions to protect the life of a pregnant woman, (2) Laos and The Philippines where abortion is not permitted for any reason. In Vietnam, there are no laws to prosecute those who their pregnancy at all.

The abortion rate being higher in rural areas than in urban areas could be due to a number of reasons ranging from financial stability to sex selection (FSA), which is very common in many Asian cultures.

5.3.2 Poorest Wealth Quintile vs Richest Wealth Quintile

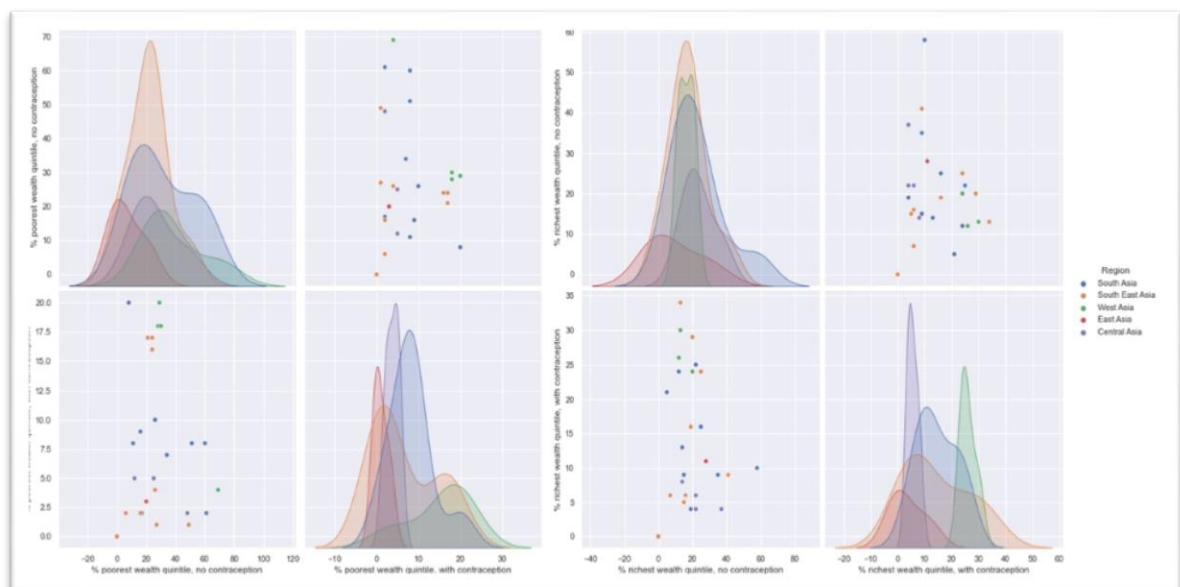
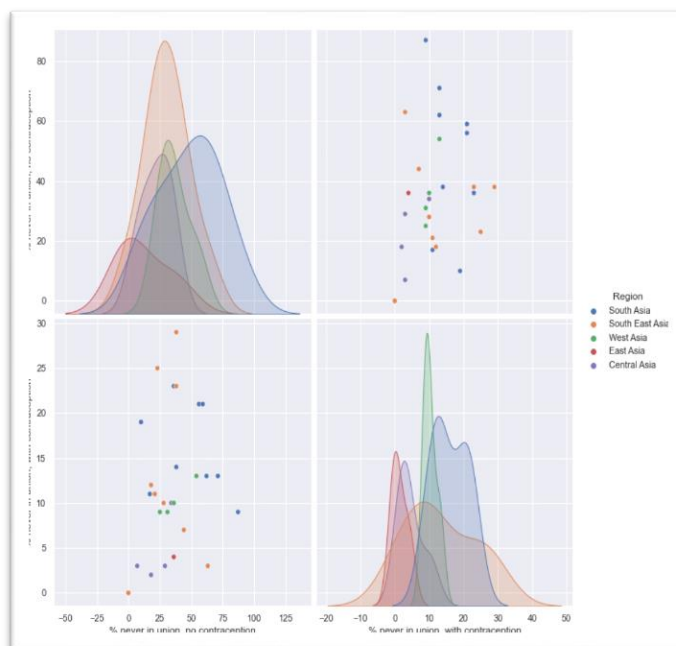
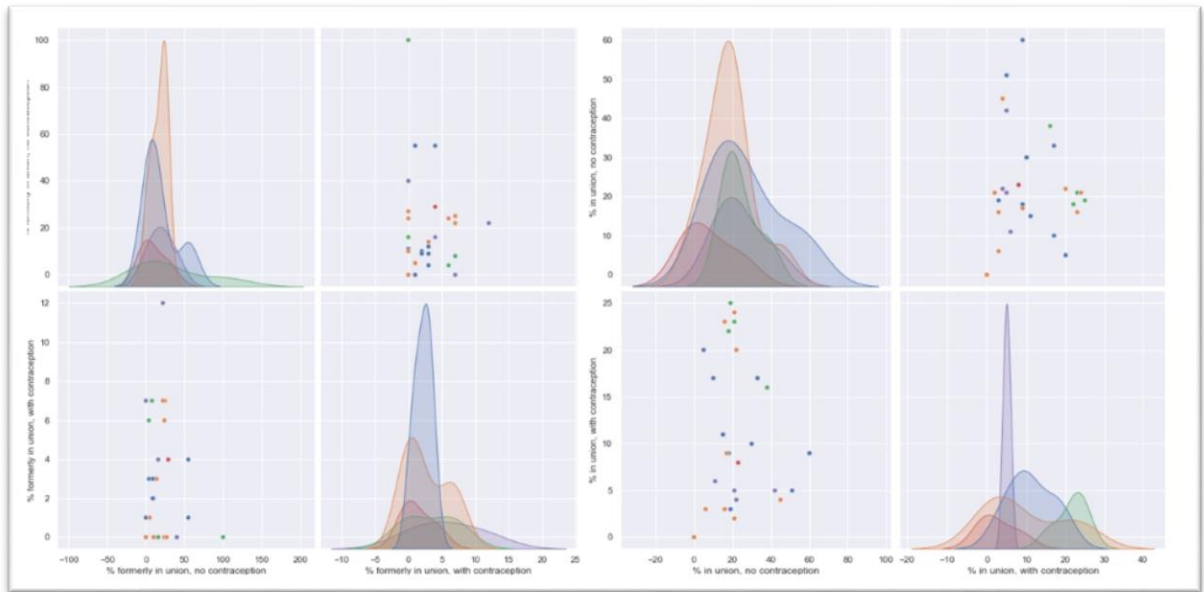


Fig 5.3.2 - Distribution of the percentage of women in the poorest [left] and richest [right] wealth quintiles (with and without the use of contraception) by region, who want to avoid pregnancy.

Similar to rural and urban areas, with an average of almost 70% and 60%, the poorest and richest wealth quintiles (without contraception) respectively, seek abortion in Southeast Asia. Although, the variation for both quintiles (with contraception) is almost fairly distributed with Central Asia peaking at an average of 20% and 33% for the poorest and richest quintiles respectively. Comparatively the overall rate is higher for the poorest wealth quintile which could be due to the likelihood that an increase in economic burden resulting in considerable costs, would affect more poor families than rich.



5.3.3 Formerly In Union vs In Union vs Never In Union

Fig 5.3.3 - Distribution of the percentage of women who were formerly in union [top left], in union [top right] and never in union [bottom left], (with and without the use of contraception) by region, who want to avoid pregnancy.

The variation for the women who were formerly in union in

comparison, to the women who are in union and have never been in union, is relatively symmetric with the highest peak values at 12% and 100%, with and without the use of contraceptives, respectively. The overall percentage of women in union, who have used contraceptives, in all regions except for Central Asia, is relatively lower than that of the portion that have not used contraceptives. Whereas the distribution for the women who have never been in union (with and without contraception) is relatively high and evenly

distributed among most regions, which helps understand that in most cases, unstable unions are associated with more abortions, therefore relationship stability could aid the decisive process around abortions, to overcome this limitation.

5.3.4 Total Pregnancies Ending in Abortion by Region vs by Legality

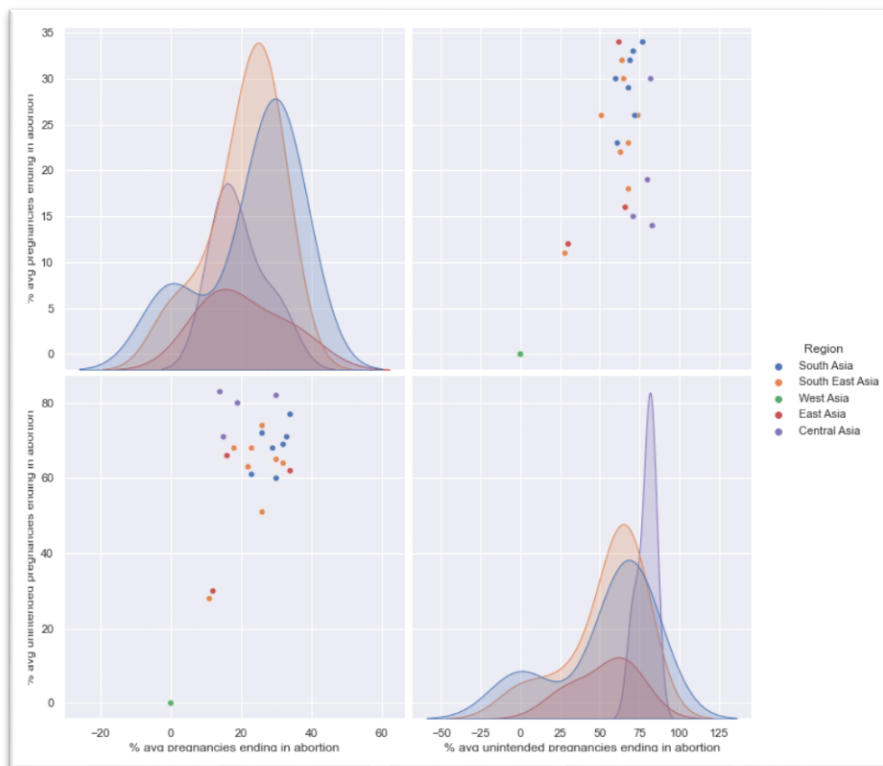


Fig 5.3.4.1 – Distribution of the percentage of total pregnancies ending in abortion, by region.

Both distributions for the average pregnancies and unintended pregnancies

ending in abortion are negatively skewed, which means the median (with reference to *fig 5.1.1* - 16% for average pregnancies and 61% for average unintended pregnancies) is greater than the average, due to the significantly lower values that reduce the mean. Comparatively, the variation of the average normal pregnancies is higher than the average unintended pregnancies ending in abortion. Evidently, majority of the abortions occurred in Southeast and Central Asia. According to the World Health Organization, around 45% of all abortions are unsafe, of which 97% take place in developing countries. The number of countries in each sub-region of Asia is as follows: (1) South Asia – 9, (2) Southeast Asia – 11, (3) West Asia – 18, (4) East Asia – 6, (5) Central Asia – 5, of which 8 are developing countries.

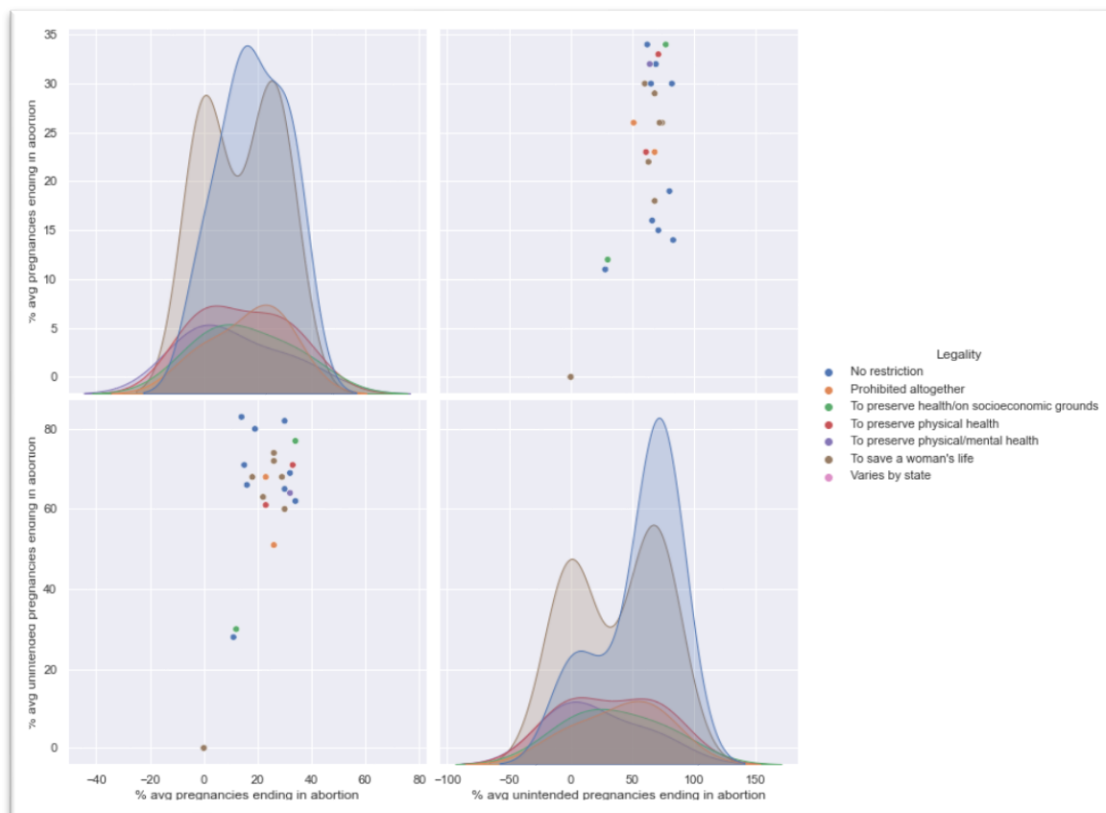


Fig 5.3.4.2 – Distribution of the percentage of total pregnancies ending in abortion, by legality.

As displayed above, most abortions occur in sub regions that abide by “No restriction” and “To save a woman’s life”. Approximately six out of ten unintended pregnancies end in an induced abortion. It should also be noted that due to the severity of the consequences of abortion in countries where it is “Prohibited altogether”, there is a limitation of records. The distributions are also double-peaked due to the lack of records that are registered as null values.

Asia consists of regions where abortion is legal, broadly legal and completely restricted, out of which statistically almost 90% of abortions in regions with liberal abortion laws are considered safe, compared with just 25% of abortions in regions where abortion is banned.

Chapter 6: Advanced Analysis

6.1 Regression Model

With reference to the findings thus far (5.3 Analytics Models and Algorithms), it can be concluded that comparatively, some factors affect the total percentage of abortions that occur. To further specify which factors are most likely to continue affecting the total percentage of abortions, a predictive analysis, called linear regression, is performed. The conditions applied to this algorithm will be, the total percentage of unintended pregnancies ending in abortion, and the variables where no contraception was used.

As mentioned in section 5.4, the metrics that will be used to evaluate the regression model are:

Mean Absolute Error (MAE) – It measures the average variance of the actual data and the predicted data.

Root Mean Squared Error (RMSE) – The MAE fails to punish big error terms, due to this RMSE is used to show how far the predicted values are measured to the actual values using Euclidean distance (the length of a line segment between two points).

6.1.1 Exploratory Data Analysis

Country	% rural, no contraception	% urban, no contraception	% formerly in union, no contraception	% poorest wealth quintile, no contraception	% richest wealth quintile, no contraception	% in union, no contraception	% never in union, no contraception	% avg pregnancies ending in abortion	% avg unintended pregnancies ending in abortion	Region_Central Asia
Afghanistan	55.0	39.0	55.0	61.0	35.0	51.0	17.0	0.0	0.0	0
Bangladesh	16.0	12.0	10.0	16.0	14.0	15.0	71.0	30.0	60.0	0
Bhutan	19.0	17.0	11.0	17.0	19.0	19.0	36.0	29.0	68.0	0
Cambodia	16.0	14.0	24.0	21.0	13.0	16.0	23.0	30.0	65.0	0
China	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0

5 rows × 21 columns

Fig 6.1.1 – Categorical variables in the dataset have numeric representation.

Furthering the exploratory data analysis, which is a part of the data preprocessing phase, categorical variables are converted to dummy/numeric variables to represent the subgroups of the sample data.

6.1.2 Feature Selection

Feature selection is used to reduce the amount of input variables to a regression model to eliminate the noise in data. This can be done by calculating the correlation and feature importance.

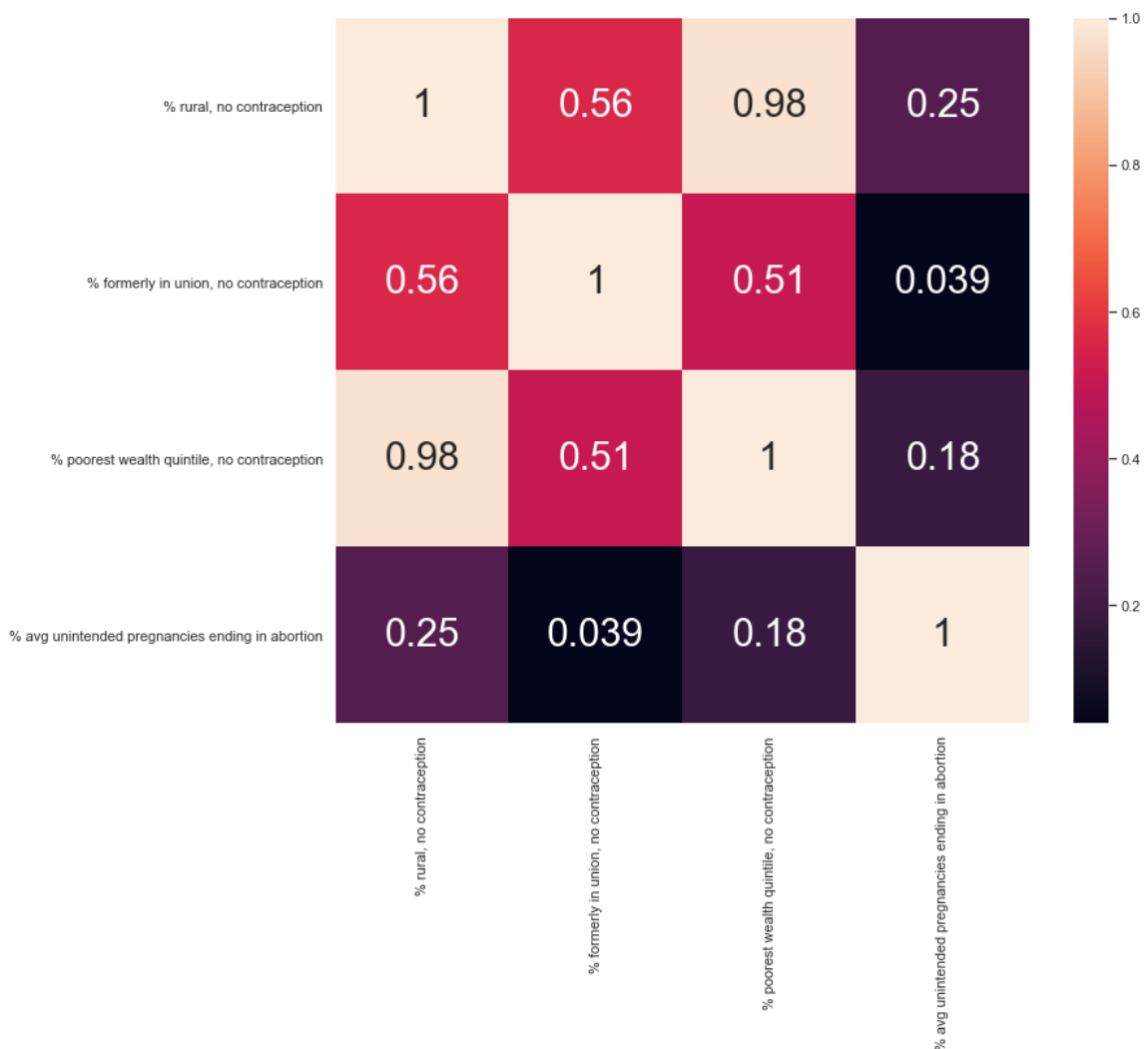


Fig 6.1.2 – Heatmap used to visualize the correlation between the variables.

The higher the correlation (indicated by the darker colors), the more other variables tend to increase/decrease. Correlation measures the strength of a linear relationship, although it does not necessarily guarantee causation; other factors may also be applicable.

According to *fig 6.1.2*, a few variables such as those of which who were formerly in union and the poorest wealth quintile, have no correlation and whose correlation value is near 0.

index		Feature	Feature Importance
0	1	% formerly in union, no contraception	0.367022
1	0	% rural, no contraception	0.141094
2	5	Region_South Asia	0.107337
3	8	Legality_No restriction	0.107034
4	6	Region_South East Asia	0.089520
5	3	Region_Central Asia	0.075275
6	4	Region_East Asia	0.064589
7	10	Legality_To preserve health/on socioeconomic g...	0.032086
8	2	% poorest wealth quintile, no contraception	0.016043
9	7	Region_West Asia	0.000000
10	9	Legality_Prohibited altogether	0.000000
11	11	Legality_To preserve physical health	0.000000
12	12	Legality_To preserve physical/mental health	0.000000
13	13	Legality_To save a woman's life	0.000000
14	14	Legality_Varies by state	0.000000

Table 6.1.2 – Table displaying the feature importance in descending order.

As shown above, the feature importance is calculated and displayed in descending order, a higher score does not imply the specific feature will have a larger effect on the model. A tree-based estimator is run prior to calculating the feature importance, to split the sample data using the significant features.

6.1.3 Model Evaluation

The model is run by first, splitting the raw data, and then training the regression model. After creating and fitting the model, the accuracy, RMSE, MAE, coefficients and intercept are calculated.

```
The Accuracy on the training dataset is: 0.8012615976134841
The Accuracy n2 on the training dataset is: 0.8012615976134841

The Accuracy on the testing dataset is: -0.17363296518260896

The RMSE on the training dataset is: 14.850377722650157
The RMSE on the testing dataset is: 34.73394283437151

The MAE on the training dataset is: 11.195450969976402
The MAE on the testing dataset is: 27.333420983686356

Coefficients: [ 0.79376451 -0.28266691 -0.62005272 74.32740981 39.45180777
 52.91986344 71.82585791 12.94762233 5.91545764 -0.65914417
 3.79369875 7.4336003 -5.81789334 -10.66571918 0. ]

Intercept: 1.693614032951558
```

Fig 6.1.3.1 – Calculation of model accuracy on the training dataset.

The accuracy score represents the coefficient of determination (R^2). According to the findings presented above, 80% of both the training and testing data are accurate, this represents the variance in the outcome that the model is capable of predicting based on its features. This confirms the validity and usefulness of the model.

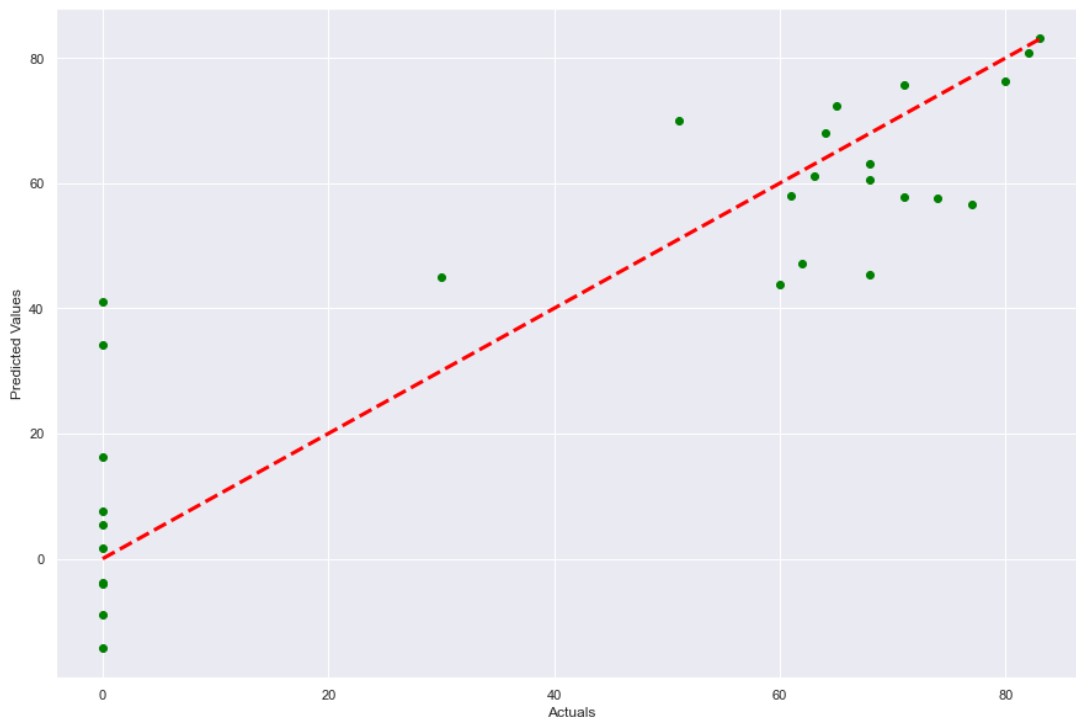


Fig 6.1.3.2 – Predicted vs Actual values plotted against each other.

As shown in *fig 6.1.3.2*, the amount of data that falls on the regression line is conveyed through the R-squared value of 80%.

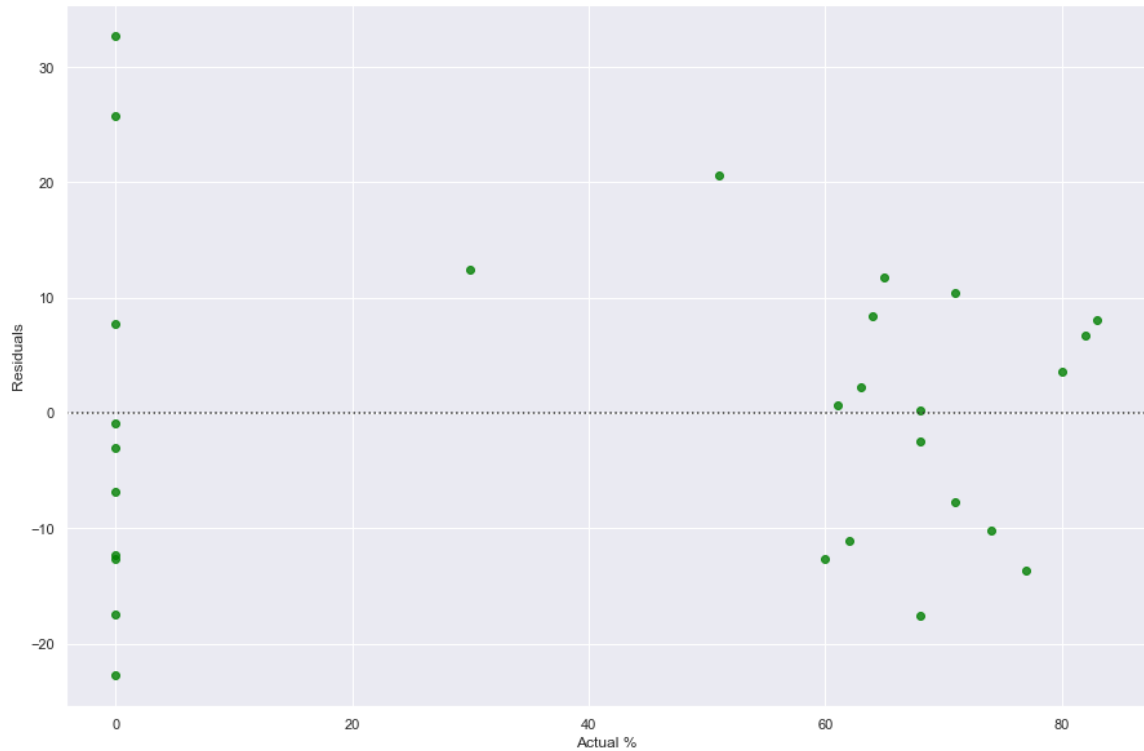


Fig 6.1.3.3 – Residuals vs Actual values plotted against each other.

The root mean squared error (RMSE) is the standard deviation of the residuals (the difference between the predicted value and the regression line). The lower the MAE and RMSE values, 11.2 and 14.1 respectively (*with reference to fig 6.1.3.1*), implies a higher accuracy of the regression model.

Chapter 7: Discussion and Recommendations

7.1 Discussion

The table below displays a few examples of reasons behind the request for abortion, from all around the world, and their relative gestational time period. The severity of the cases ranges from being “socially acceptable” for an abortion, to extreme cases that cause controversy and judgement. [**These examples were sourced from a video on a social media platform (www.tiktok.com), however, the video can no longer be found as the video was reported and taken down for “violating guidelines”.*]

	Reasoning behind the Request for Abortion	Gestational Time Period
1.	Showed development of the foetus without life sustaining organs, after an anatomy scan.	20 weeks
2.	Impregnated by an abusive husband.	15 weeks
3.	Raped in a foster home.	11 weeks
4.	Had been trying to become pregnant for 15 years, only to get pregnant with five foetuses at once, putting the health of all of them at peril.	10 weeks
5.	Found out about the pregnancy at the age of 52, approaching menopause.	9 weeks
6.	Incarcerated till days before she found out about her pregnancy.	13 weeks
7.	Not financially, emotionally or physically ready to raise a child.	n/a
8.	Receives asylum from Myanmar and her interpreter walked out on her first appointment requiring a delay.	14 weeks
9.	The egg never made it out of the fallopian tube, either she terminates the pregnancy or risks dying from internal bleeding.	13 weeks
10.	She had undergone IVF, ending up with 6 viable implanted eggs, requiring selective reduction in order to ensure her safety and a safe amount of foetuses.	n/a

11.	Partner removed condom during sexual intercourse, got rid of her birth control pills and forcefully pulled out her IUD.	8 ½ weeks
12.	Contraceptive that was used did not prevent the pregnancy.	6 ½ weeks
13.	Had an accidental pregnancy and was beaten up by her father unrecognizable.	9 weeks
14.	Sexually assaulted on her way home from work, only to find out she had been impregnated by her assailant.	4 weeks

Table 7.1 – Table of reasons behind the request for abortion and their respective gestational time period.

As displayed in the table, the reasoning behind a request for abortion can vary over a range of situations. The online platform, this data was sourced from, did not state if they were granted approval for an abortion or not. When people with unintended pregnancies are faced with difficulties in attaining a safe, affordable and unbiased abortion, they often resort to unsafe abortions.

7.2 Recommendations

Most cases of abortions are due to a woman being impregnated when they did not intend to. From the results gathered, the use of contraceptives is the most effective way to prevent unintended pregnancies. Therefore, schools and programs should focus on educating teenagers, young adults and adults, and improving their knowledge on access to and the use of contraceptive methods, in order to reduce the need for an abortion.

Governments must improve their sexual and reproductive healthcare and family planning services, so its women are no longer caught between the lack of the services and unsafe abortions. They must also make sure of making this information accessible and affordable, regardless of which tax bracket majority of the population resides in.

Sex selection, financial hardship, concern for the mother's life, or a teenage girl not ready for motherhood, no matter what the reason, many women who terminate their pregnancies say they often have to face harsh judgments about their choices. In order to

prevent the controversy surrounding this subject, quality sex education should be provided and more importantly, should be accessible.

Studies show that where abortion is permitted on broad legal grounds, it is generally safe, and where it is highly restricted, it is typically unsafe. To reduce the number of nonsurgical procedures, the grounds for legal abortion must be expanded, and accessibility to safe abortion services and post-abortion care should be improved for women. The providers for these services must also be heavily trained.

7.3 Conclusions

All types of people get abortions for all types of reasons, all are valid, all deserve the choice. Restrictive laws will bring government mandated pregnancies for women and girls of all ages. Studies have discovered that, the enforced restriction of abortion is not supportive to humans' freedom and is a major cause for illegal abortion which could threaten the physical/mental health, and most often the lives of pregnant women.

With reference to chapter 6, creating a functional regression model results in the ability to analyse historical healthcare data and create a predictive tool to estimate [1] which sub-regions/countries require immediate healthcare services, or [2] calculate the accurate cost of health insurance for each demographic in order for healthcare to be accessible, etc.

Human rights are in favor of decriminalizing and ensuring access to abortion, it also requires that abortion services should be available to all, irrespective of their specific circumstances.

List of References

Center for Reproductive Rights (May 18, 2019). *World's Abortion Laws Map Progress* [Video]. YouTube. https://youtu.be/_IeifPzqh8

Center for Reproductive Rights. *The World's Abortion Laws*.
<https://reproductiverights.org/maps/worlds-abortion-laws/>

Council of Foreign Relations (from Women & Foreign Policy Program). *Abortion Law: Global Comparisons*.
<https://www.cfr.org/article/abortion-law-global-comparisons>

Desai. S., Hyunh. M. & Jones. H. E. (June 8th, 2021) | *Differences in Abortion Rates between Asia Populations by Country of Origin and Nativity Status in New York City, 2011-2015*. | National Library of Medicine. (2021). | doi: [10.3390/ijerph18126182](https://doi.org/10.3390/ijerph18126182)

D. Ismi, S. Kiran and K. Arisa (June 15, 2022) | *Abortion in Asia: The limits of choice*
<https://asia.nikkei.com/Spotlight/The-Big-Story/Abortion-in-Asia-The-limits-of-choice>

Global Abortion Policies Database (GAPD) - World Health Organization (WHO)
<https://abortion-policies.srhr.org/>

Gunasekera. S. (2019). *Fighting for Safe Abortion Access in Sri Lanka*.
<https://www.ippf.org/blogs/fighting-safe-abortion-access-sri-lanka>

Guttmacher Institute. (2015). *Facts on abortion in Asia*.
https://www.guttmacher.org/sites/default/files/factsheet/ib_aww-asia_0.pdf

Ined. (2014). *Selective Abortion in Asia*.
https://www.ined.fr/en/everything_about_population/demographic-facts-sheets/focus-on/selective-abortion-Asia/

Kamdar. B. (June 15th, 2021). *Poor Access to Safe Abortions is Killing South Asian Women*. | The Diplomat

<https://thediplomat.com/2021/06/poor-access-to-safe-abortions-is-killing-south-asian-women/>

Law & Policy Recommendation 1: Criminalization (2.21) | Abortion Care Guideline

<https://srhr.org/abortioncare/chapter-2/recommendations-relating-to-regulation-of-abortion-2-2/law-policy-recommendation-1-criminalization-2-2-1/>

Nikkei Asia. (2022). *Asia Stream: Abortion in Asia -- Beyond Taboo*.

<https://asia.nikkei.com/Spotlight/Podcast/Asia-Stream-Abortion-in-Asia-Beyond-Taboo>

Vivanco, J. M. (January 11th, 2021). *Criminalization of Abortion Has Failed. It's Time to End it*.

<https://www.hrw.org/news/2021/01/11/criminalization-abortion-has-failed-its-time-end-it>

World Population Review. *Countries where Abortion is Illegal 2022*.

<https://worldpopulationreview.com/country-rankings/countries-where-abortion-is-illegal>

Appendix

With reference to Chapter 3:

1. Data Cleaning

```
In [13]: df.isna().sum()
Out[13]: Country      0
Region      0
% rural, no contraception  0
% rural, with contraception  0
% urban, no contraception  0
% urban, with contraception  0
% formerly in union, no contraception  0
% formerly in union, with contraception  0
% poorest wealth quintile, no contraception  0
% poorest wealth quintile, with contraception  0
% richest wealth quintile, no contraception  0
% richest wealth quintile, with contraception  0
% in union, no contraception  0
% in union, with contraception  0
% never in union, no contraception  0
% never in union, with contraception  0
% avg pregnancies ending in abortion  0
% avg unintended pregnancies ending in abortion  0
dtype: int64
```

2. Data Integration

```
In [7]: df = pd.merge(countries, legality, on='Country', how='right')
df.head()
```

3. Data Transformation

```
In [14]: df.columns = ['Country',
                        'Region',
                        '% rural, no contraception',
                        '% rural, with contraception',
                        '% urban, no contraception',
                        '% urban, with contraception',
                        '% formerly in union, no contraception',
                        '% formerly in union, with contraception',
                        '% poorest wealth quintile, no contraception',
                        '% poorest wealth quintile, with contraception',
                        '% richest wealth quintile, no contraception',
                        '% richest wealth quintile, with contraception',
                        '% in union, no contraception',
                        '% in union, with contraception',
                        '% never in union, no contraception',
                        '% never in union, with contraception',
                        '% avg pregnancies ending in abortion',
                        '% avg unintended pregnancies ending in abortion']
df.head()
```

Out[14]:

Country	Region	% rural, no contraception	% rural, with contraception	% urban, no contraception	% urban, with contraception	% formerly in union, no contraception
---------	--------	------------------------------	--------------------------------	------------------------------	--------------------------------	---

4. Data Reduction

```
In [12]: df2 = df[df['Region'].str.contains('Asia', na=True)]
df2 = df2.set_index('Country')
print(df2.shape)
print(df2.head())
df2.head()

(163, 19)
(35, 18)
```

With reference to Chapter 5:

1. Data Analysis

```
In [17]: round(df2.describe(),2)
```

2. Data Visualization

```
14 ~~~{r}
15 pie = ggplot(df, aes(x="", y=as.character(legality), fill=legality)) +
16   geom_bar(stat="identity", width=1) +
17   coord_polar("y", start=0)
18
19 pie = pie + labs(x = NULL, y = NULL, fill = NULL, title = "Abortion Laws Around the world")
20
21 pie = pie + theme_classic() + theme(axis.line = element_blank(),
22   axis.text = element_blank(),
23   axis.ticks = element_blank(),
24   plot.title = element_text(hjust = 0.5, color = "#1c1a19"))
25 pie
26 ~~~
```

```
In [20]: sns.catplot(data=df, y='Region', x="% avg pregnancies ending in abortion",
hue="Legality", kind="swarm");
```

```
In [19]: sns.catplot(data=df2, x='Region', y="% avg pregnancies ending in abortion", hue="Legality", kind="swarm");
```

3. Analysis Models and Algorithms

```
In [20]: rural = sns.pairplot(df2[['% rural, no contraception',
'% rural, with contraception',
'Region']],hue='Region',height=5)
```

```
In [21]: urban = sns.pairplot(df2[['% urban, no contraception',
'% urban, with contraception',
'Region']],hue='Region',height=5)
```

```
In [22]: poorest_wealth_quintile = sns.pairplot(df2[['% poorest wealth quintile, no contraception',
'% poorest wealth quintile, with contraception',
'Region']],hue='Region',height=5)
```

```
In [23]: richest_wealth_quintile = sns.pairplot(df2[['% richest wealth quintile, no contraception',
'% richest wealth quintile, with contraception',
'Region']],hue='Region',height=5)
```



```

In [24]: formerly_in_union = sns.pairplot(df2[['% formerly in union, no contraception',
                                             '% formerly in union, with contraception',
                                             'Region']],hue='Region',height=5)

In [25]: in_union = sns.pairplot(df2[['% in union, no contraception',
                                       '% in union, with contraception',
                                       'Region']],hue='Region',height=5)

In [26]: never_in_union = sns.pairplot(df2[['% never in union, no contraception',
                                             '% never in union, with contraception',
                                             'Region']],hue='Region',height=5)

In [27]: pregnancies_ending_in_abortion = sns.pairplot(df2[['% avg pregnancies ending in abortion',
                                                            '% avg unintended pregnancies ending in abortion',
                                                            'Region']],hue='Region',height=5)

In [27]: pregnancies_ending_in_abortion = sns.pairplot(df2[['% avg pregnancies ending in abortion',
                                                            '% avg unintended pregnancies ending in abortion',
                                                            'Legality']],hue='Legality',height=5)

```

With reference to Chapter 6:

1. Exploratory Data Analysis

```

In [59]: import os
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score, explained_variance_score, mean_absolute_error, mean_squared_error
from math import sqrt

%matplotlib inline
sns.set(rc={'figure.figsize':(12,10)})

```

2. Feature Selection

```

In [99]: nc = df2[['% rural, no contraception',
                  '% formerly in union, no contraception',
                  '% poorest wealth quintile, no contraception',
                  '% avg unintended pregnancies ending in abortion']].corr()

In [100]: nocontraception = sns.heatmap(nc, annot = True, annot_kws={'size':30})

In [101]: df3 = df2[['% rural, no contraception',
                    '% formerly in union, no contraception',
                    '% poorest wealth quintile, no contraception',
                    '% avg unintended pregnancies ending in abortion',
                    'Region','Legality']]
df3.head()

```

```
In [103]: X = newdf.drop('% avg unintended pregnancies ending in abortion', axis = 1).values
          y = newdf['% avg unintended pregnancies ending in abortion']

          print(X.shape)
          print(y.shape)
```

```
X
(35, 15)
(35,)
```

```
In [68]: X = newdf.drop('% avg unintended pregnancies ending in abortion', axis = 1).values
          X2 = newdf.drop('% avg unintended pregnancies ending in abortion', axis = 1)
          y = newdf['% avg unintended pregnancies ending in abortion']

          print(X.shape)
          print(y.shape)
```

```
X
(35, 20)
(35,)
```

```
In [104]: dt = DecisionTreeClassifier(random_state=15,criterion='entropy',max_depth=10)
          dt.fit(X,y)
```

```
Out[104]: DecisionTreeClassifier(criterion='entropy', max_depth=10, random_state=15)
```

```
In [44]: f_col = []
          f = []

          for i,column in enumerate(newdf.drop('% avg unintended pregnancies ending in abortion',axis=1)):
              print('the feature importance for {} is: {}'.format(column,dt.feature_importances_[i]))

              f_col.append(column)
              f.append(dt.feature_importances_[i])
```

```
In [53]: f_col
          f
          f_df = zip(f_col,f)
          f_df = pd.DataFrame(f_df,columns=['Feature','Feature Importance'])
          f_df

          f_df = f_df.sort_values('Feature Importance',ascending=False).reset_index()

          columns_to_keep = f_df['Feature'][0:40]
          f_df
```

```
In [106]: X_train, X_test, y_train, y_test = train_test_split(X, y,
                                                              train_size=0.80, test_size = 0.2, random_state=15)

          print(X_train.shape)
          print(X_test.shape)
          print(y_train.shape)
          print(y_test.shape)
```

```
(28, 15)
(7, 15)
(28,)
(7,)
```

```
In [107]: lm = LinearRegression(fit_intercept = True)
lm.fit(X_train, y_train)

y_pred = lm.predict(X_train)
y_pred
```

3. Evaluating the Model

```
In [108]: print('The Accuracy on the training dataset is: ',
              lm.score(X_train, y_train) )
print('The Accuracy n2 on the training dataset is: ',
      r2_score(y_train,y_pred) )

print("")
# Model Accuracy on testing dataset
print('The Accuracy on the testing dataset is: ',
      lm.score(X_test, y_test) )

print("")
# The Root Mean Squared Error (RMSE)
print('The RMSE on the training dataset is: ',
      sqrt(mean_squared_error(y_train,y_pred)))
print('The RMSE on the testing dataset is: ',
      sqrt(mean_squared_error(y_test,lm.predict(X_test))))

print("")
# The Mean Absolute Error (MAE)
print('The MAE on the training dataset is: ',
      mean_absolute_error(y_train,y_pred))
print('The MAE on the testing dataset is: ',
      mean_absolute_error(y_test,lm.predict(X_test)))

print("")
# Coefficients
print('Coefficients: ', lm.coef_ )

print("")
# The Intercept
print('Intercept: ', lm.intercept_)
```

```
In [109]: plt.figure(figsize=(15,10))

plt.scatter(y_train, y_pred, c='green')
plt.plot([y_train.min(), y_train.max()], [y_train.min(),
      y_train.max()], 'k--', c='red', lw=3)
plt.xlabel('Actuals')
plt.ylabel('Predicted Values')
plt.title('Actuals Vs Predicted Values');
```

```
In [110]: plt.figure(figsize=(15,10))

sns.residplot(y_train, y_pred, color='green')
plt.xlabel('Actual %')
plt.ylabel('Residuals')
plt.title('Actuals Vs Residuals');
```