Cross Country Analysis & Emigration

Ritu Bhamrah - 015324478

Department of Applied Data Science, San Jose State University

Course: DATA 230 – Data Visualization

Instructor: Andrew H. Bond

December 12, 2021

Abstract

The emigration has been always a debatable topic to understand if it has advantages or disadvantages. This project focusses understanding the factors which could contribute the emigration. For this project, we collected the data from the online resources Kaggle and merged the different data files using tableau prep builder. The cleaning of data is also achieved using tableau prep builder. The data is analyzed using the visualization tool tableau. This visualization could help us to understand some of the factors causing emigration. The findings of this project can be leveraged to take various measures to improve the socio-economic conditions in different regions.

Table of Contents

1. Introduction

Emigration is the act of people moving from one place to another. If the act of movement is unidirectional, then it can cause some advantages and disadvantages to the origin and destination regions. There could be many reasons for people to migrate from one place to another place. For example, people move from smaller cities to urban areas for better job opportunities. This can create population density imbalance between urban and rural areas. Thus, it is important to analyze and understand the emigration rates and various factors causing emigration to address and improve the conditions in the regions with high emigrations rates.

According to Damarys et al. (2013) “Economic, demographic and sociological factors hold dominant positions in most accounts of the intention to emigrate.” Based on this idea we would collect the data for this project as below:

- Demographic data - Internet usage and birth rate for different regions.
- Sociological factors- Happiness score, corruption, life expectancy, and freedom for different regions.
- Economic factors - Gross Domestic Product (GDP) per capita and income group

The primary objective of this project is to understand the different factors that influence the people to migrate from one region to another region. Such data analysis will help to identify the socio-economic problems and to provide measurable solutions for overall growth of different countries.

1.1 Purpose of this document

The purpose of this document is to provide a detailed project description of the emigration based on demographic, economic, and sociological factors.

This document includes details as below:

- Data collection from online resources.
- Deliverables - This will show end results obtained from this project.
- Development - This includes data preparation and data visualization.
- Milestones for project organization and time plans.

1.2 Intended Audience

The data visualization gives a clear understanding of the correlation between various parameters like freedom, happiness, income, emigration etc. It offers a great pictorial representation using bar graphs, pie charts, geo-maps, etc. Using this visualized representation, some conclusions can be drawn and measures for improvements can be made. Audience for this project will be:

- Professor
- Students

1.3 Scope

The scope of this project is to analyze the emigration data region wise. The regions that the project covers are North America, Asia Pacific, Europe, Middle East, Arab States, South America, and Africa. The goal is to analyze the emigration among these regions and try to evaluate the regions with higher emigration and to understand the factors contributing to higher emigration rates. The results will be demonstrated using bar graphs, geo maps, dashboards, etc. using tableau tool.

2. Background and Objectives

Due to globalization migration of people from one region to another is more prevalent in 21st century. Many researchers have studying migration trends from very long time. The Table 1 shows some of the researchers and conclusions of their study on the immigration.

Table 1

Literature Survey for Migration Cause

Author	Causes
Mayo county library	One of the causes for emigration is the economic crisis, which also leads to chain emigration.
Castelli (2018)	The one major cause for emigration is low economic development in the origin country.
Mlambo (2018)	Discussed various factors of migration from rural to urban regions in South America. The authors highlighted the factors like health and educational services, Employment opportunities for cause of migration. The migration to urban regions caused various drawbacks like less resource availability, loss of skilled people, etc. in rural regions and over population in the urban region.

Note. The table shows the various research done by different authors and their conclusions. The above data is summary of data taken from 3 different resources as:

- https://www.mayo.ie/library/local-history/historical-events/emigration/causes-of-emigration
- https://doi.org/10.1093/jtm/tay040
- https://doi.org/10.14738/abr.64.4407

2.1 Objectives

Below are the objectives for this project:

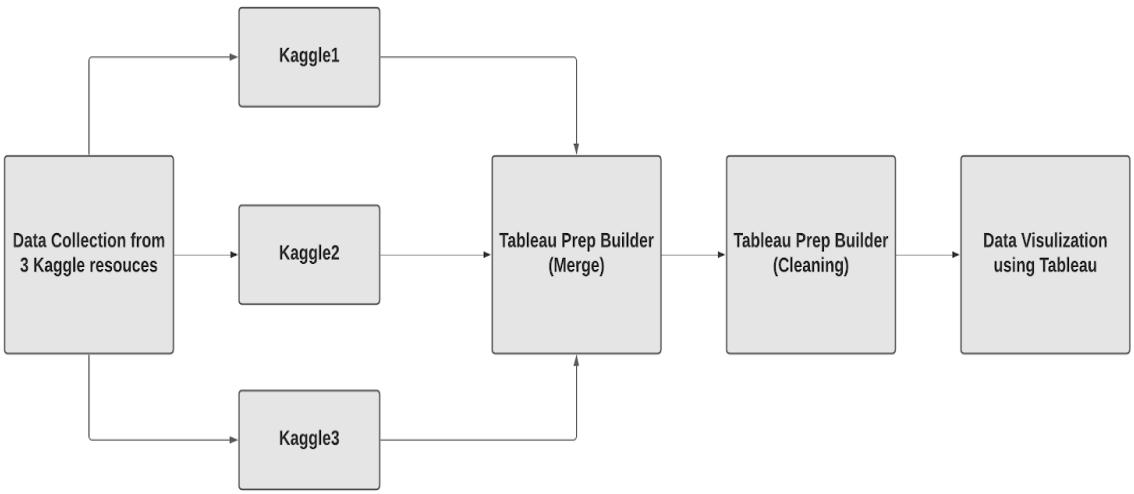
- Collect the data for the project requirement from various online resources like Kaggle.
- Merge the data using tableau prep builder.
- Visualize the emigration trends among various regions using dashboard and storytelling in tableau.
- Analyze the causes of emigration.

3. Architecture and High-Level Design

The project architecture diagram shows the different phases shown in Figure 1.

Figure 1

Project Architecture



Notes. The architecture shows the data collection, data preparation, data cleaning and visualization phases for this project. These phases can further be explained in detail as below:

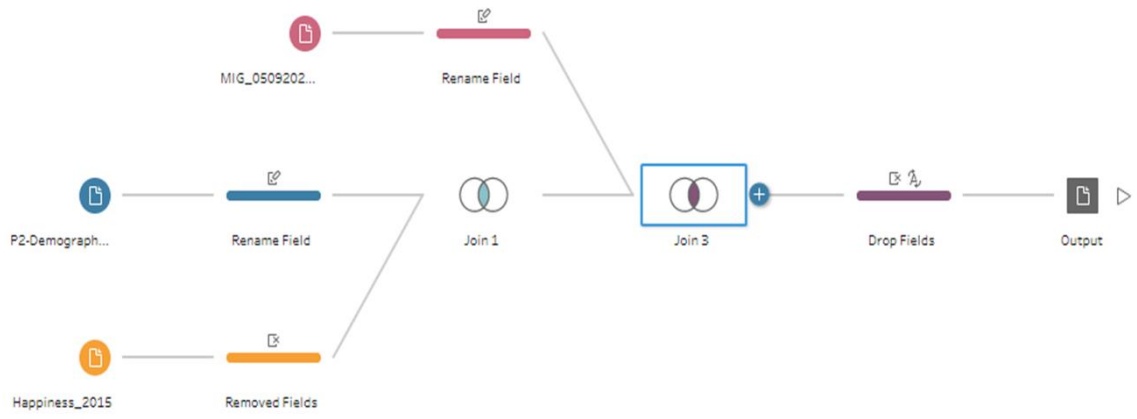
Data collection - For data collection, we have collected the data from three different Kaggle resources which have different fields as shown below:

- Demographic Data – Country name, country code, birth rate, internet users, income group.
- Happiness Report – Country, region, happiness rank, happiness score, standard Error, GDP, family, life expectancy, freedom, government trust, generosity, dystopia residual.
- Emigration Data – Emigration, country, region.

Data Preparation - The data preparation is done in tableau prep builder. This includes joining the 3 resources based on Country and cleaning of data.

Figure 2

Data Preparation Flow Chart



Notes. The Figure 2 shows the cleaning and joining steps performed on raw data. I have removed some fields from the happiness file and renamed some fields in the demographic file. After

cleaning I joined the 2 files and again applied join with the third file. The join here is done based on the country field.

Figure 3

Data After Preparation

	A	B	C	D	E	F	G	H	I	J	K	L
1	Country Code	Birth rate	Internet users	Income Group	Country	Happiness Score	GDP per Capita	Life Expectancy	Freedom	Corruption	Region	Immigration
2	AFG	35.253	5.9	Low income	Afghanistan	3.575	0.31982	0.30335	0.23414	0.42922	APAC	302327
3	AGO	45.985	19.1	Upper middle income	Angola	4.033	0.75778	0.16683	0.48827	0.12468	Africa	204
4	ALB	12.877	57.2	Upper middle income	Albania	4.959	0.87867	0.81325	0.35733	0.06413	Europe	36593
5	ARE	11.044	88	High income	United Arab Emirates	6.901	1.42727	0.80925	0.64157	0.1069	Middle Ea	56
6	ARG	17.716	59.9	High income	Argentina	6.574	1.05351	0.78723	0.44974	0.08484	South Am	17532
7	ARM	13.308	41.9	Lower middle income	Armenia	4.35	0.76821	0.7299	0.19847	0.48357	Europe	143
8	AUS	13.2	83	High income	Australia	7.284	1.33358	0.93156	0.65124	0.35637	APAC	67687
9	AUT	9.4	80.6188	High income	Austria	7.2	1.33723	0.89042	0.62433	0.18676	Europe	18976
10	AZE	18.3	58.7	Upper middle income	Azerbaijan	5.212	1.02389	0.64045	0.3703	0.16065	APAC	61041
11	BDI	44.151	1.3	Low income	Burundi	2.905	0.0153	0.22396	0.48827	0.10062	Africa	22155
12	BEL	11.2	82.1702	High income	Belgium	6.937	1.30782	0.89667	0.5845	0.48357	Europe	866
13	BEN	36.44	4.9	Low income	Benin	3.34	0.28665	0.3191	0.48827	0.12468	Africa	53323
14	BFA	40.551	9.1	Low income	Burkina Faso	3.587	0.25812	0.27125	0.39493	0.12832	Africa	2509
15	BGD	20.142	6.63	Lower middle income	Bangladesh	4.694	0.39753	0.60164	0.4082	0.12569	APAC	16775
16	BGR	9.2	53.0615	Upper middle income	Bulgaria	4.218	1.01216	0.76649	0.30587	0.00872	Europe	105
17	BHR	15.04	90.0000397	High income	Bahrain	5.96	1.32376	0.74716	0.45492	0.1175	Arab State	143990

Notes. The Figure 3 shows the CSV obtained from tableau prep builder after data preparation. It has fields shown in Figure 3 as country code, birth rate, income group, country, happiness score, GDP per capita, life expectancy, freedom, corruption, region, and emigration.

Data visualization - The data visualization is achieved using data visualization tool tableau.

4. Organization

The organization of the project covers the person working on the project and the customer who will be using this visualization.

4.1 Project group

Table 2

Person Details Working on Project

Name	Initials	Responsibility (roles)
Ritu Bhamrah	R.B.	Data collection, data cleaning, data visualization, report, and presentation.

Note. Table 2 shows the name, initials and responsibilities of the person working on this project.

4.2 Customer

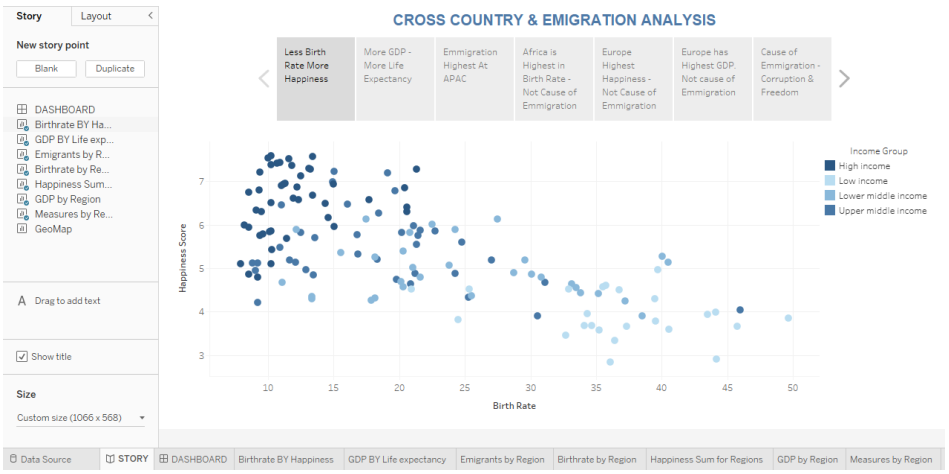
The target customers for this visualization would be anyone who is interested in analyzing emigration.

5. Development process

This project uses the tableau tool for visualization. This covers the dashboard and storytelling. The story shows all the graphs created for this visualization and the dashboard shows all the visualizations on a single page.

Figure 4

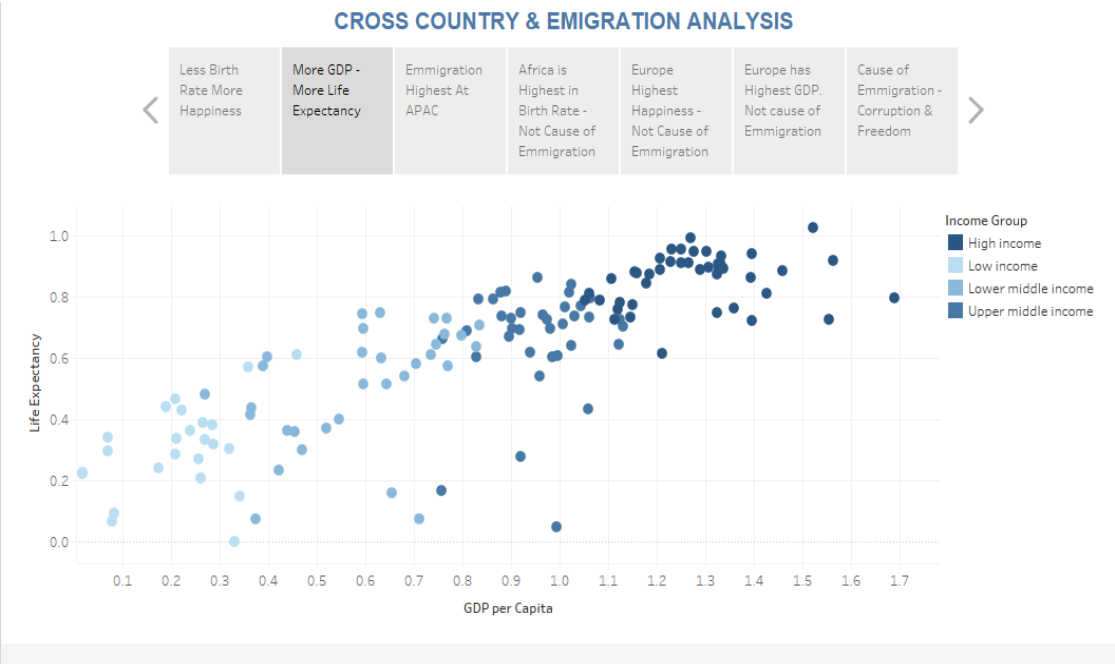
Story 1 - Scatter Plot



Note. The scatter plot shows that the less birth rate results in more happiness. The colors on the scatterplot show the data by different income groups - high income, low income, lower middle income, and upper middle income.

Figure 5

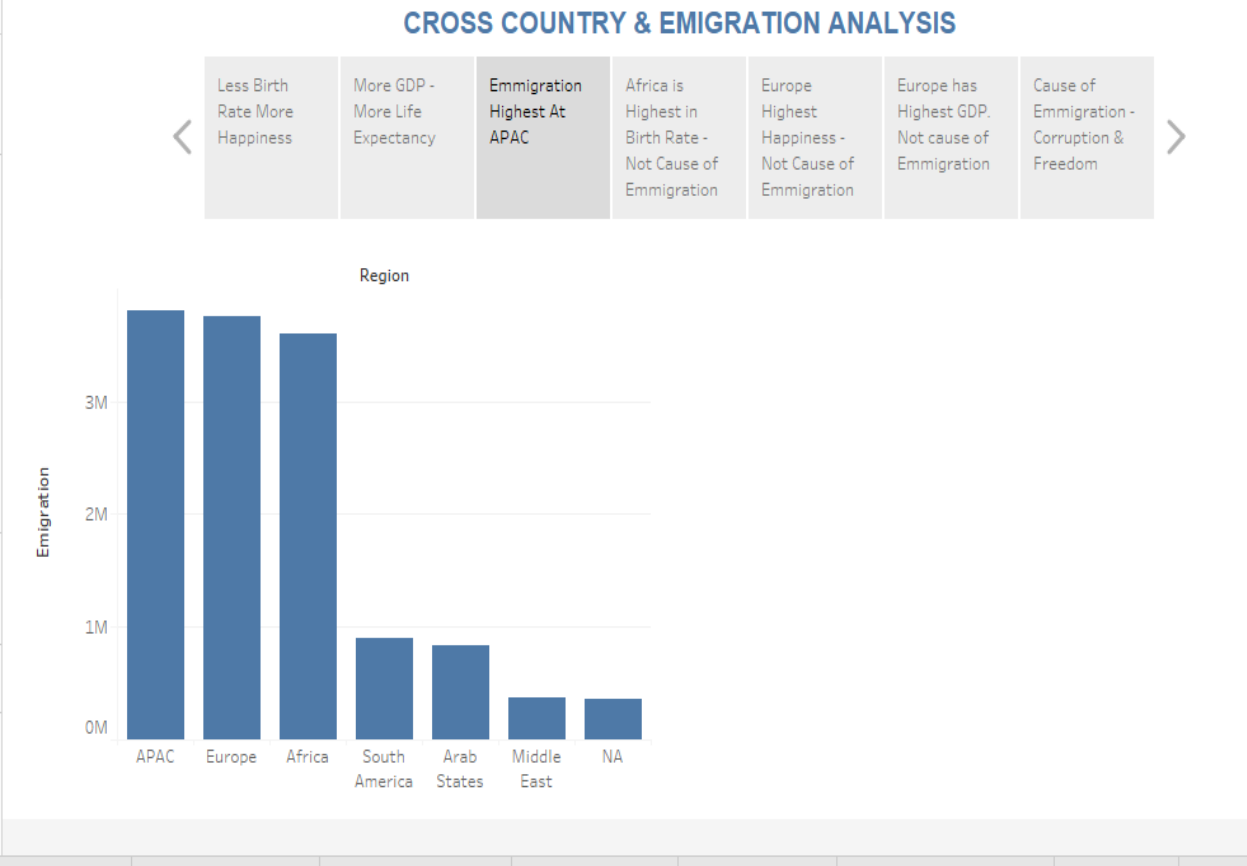
Story 2 - Scatter Plot



Note. The scatter plot shows as GDP increases the life expectancy also increases. The colors on the scatterplot show the data by different income groups - high income, low income, lower middle income, and upper middle income. The high-income group has high life expectancy compared to the low-income group.

Figure 6

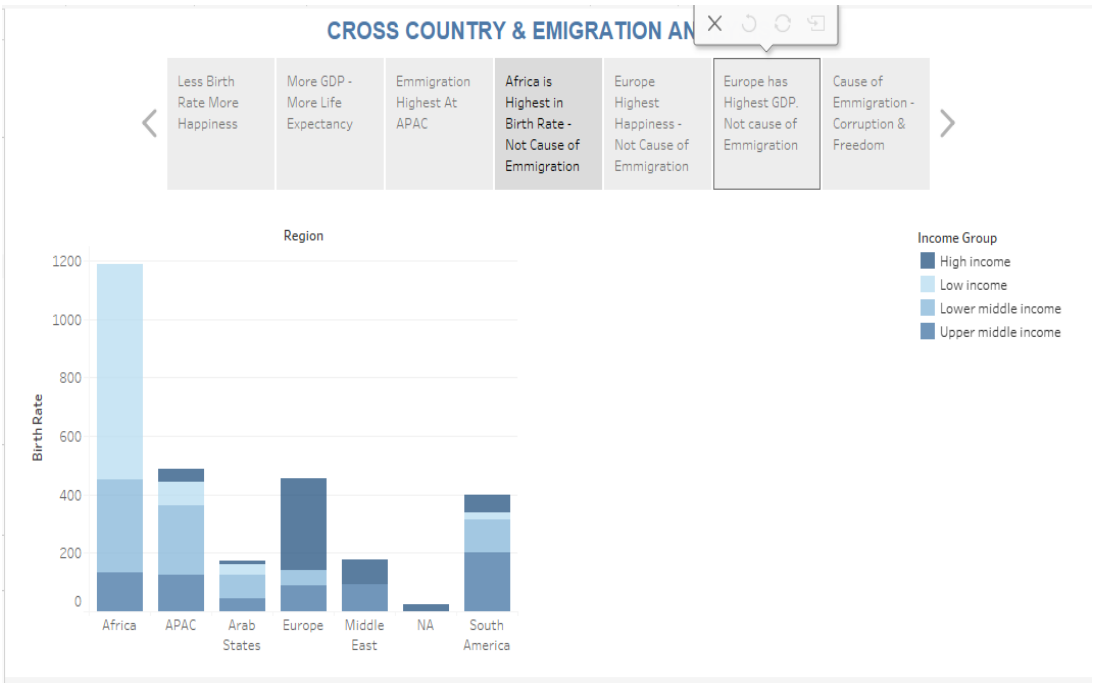
Story 3 - Bar Graph



Note. The bar graph shows the different regions as APAC, Europe, Africa, South America, Arab States, Middle East, and North America. The highest emigration can be seen in the APAC region and the lowest emigration is in North America.

Figure 7

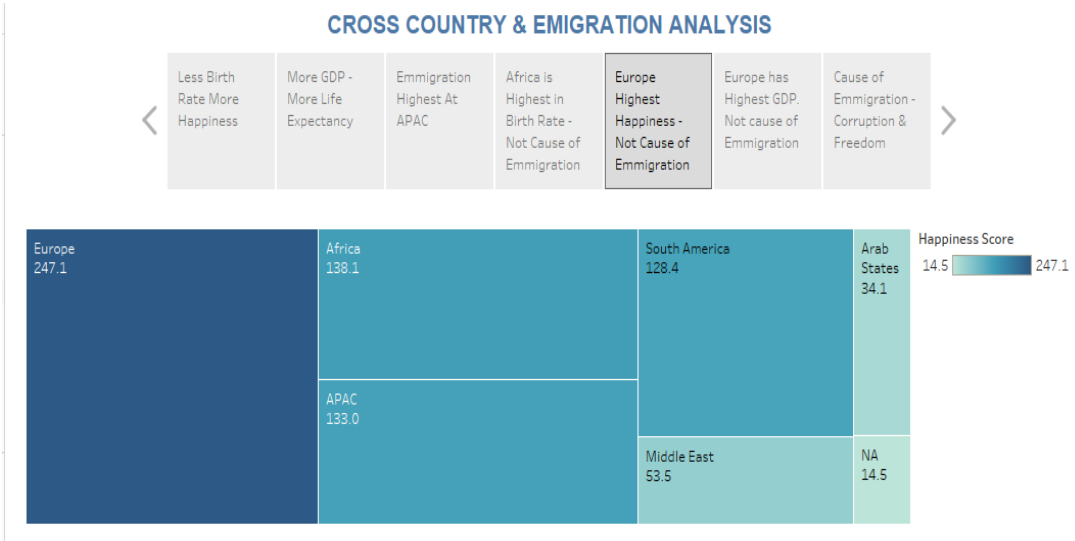
Story 4 - Stacked Bar Graph



Note. The bar graph shows the different regions as APAC, Europe, Africa, South America, Arab States, Middle East, and North America. It also shows the birth rate is highest in Africa with respect to different income groups. The graph does not show us that birth rate is the cause of emigration.

Figure 8

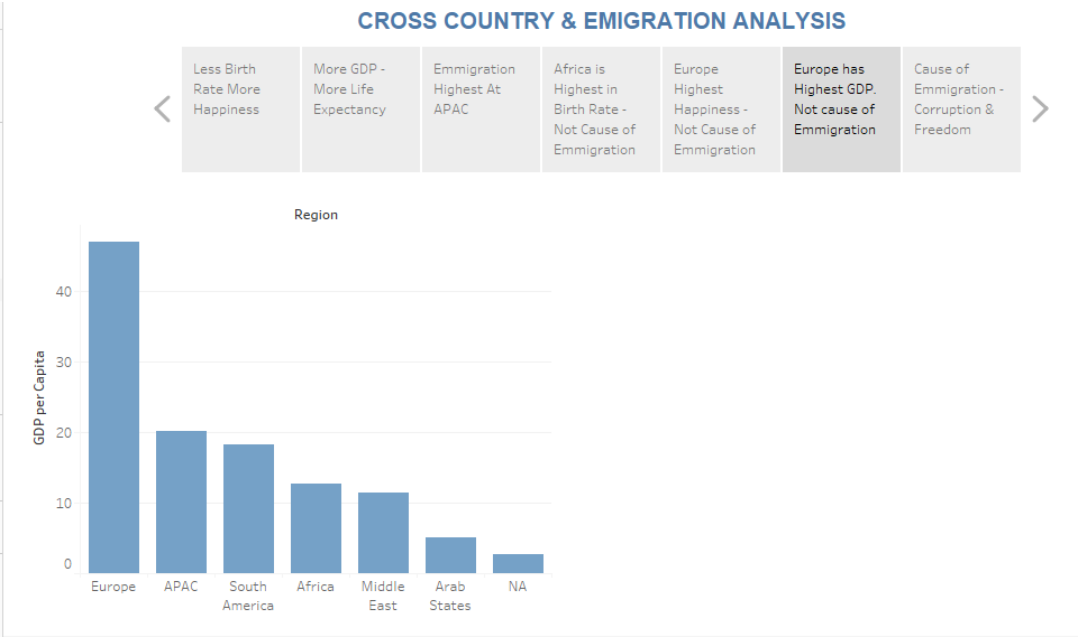
Story 5 - Tree Map



Note. The tree map shows the different regions as APAC, Europe, Africa, South America, Arab States, Middle East, and North America. It shows Europe has the highest happiness score. The tree map does not show us that happiness is the cause of emigration.

Figure 9

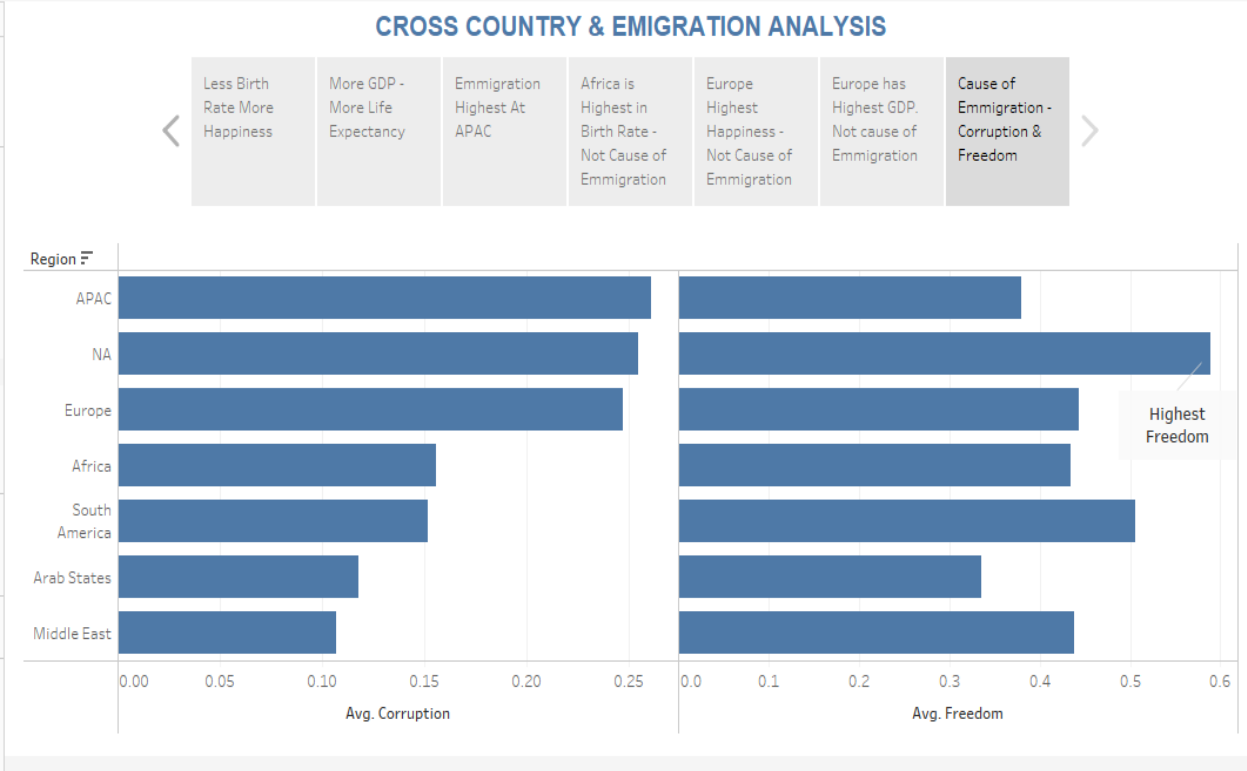
Story 6 - Bar Graph



Note. The bar graph shows the different regions as APAC, Europe, Africa, South America, Arab States, Middle East, and North America. It shows Europe has the highest GDP per capita. The graph does not show us that GDP is the cause of emigration.

Figure 10

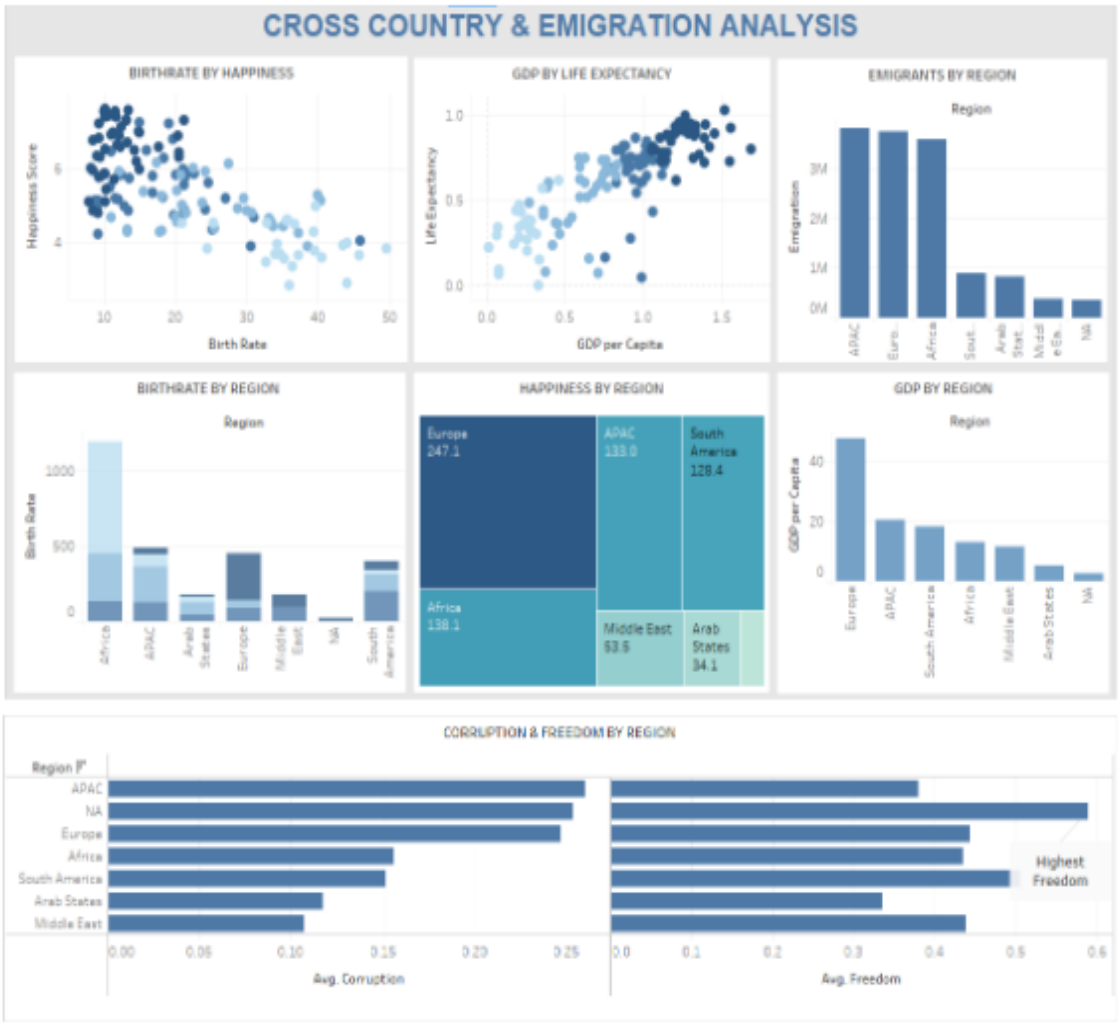
Story 7 - Bar Graph



Note. The bar graph shows the different regions as APAC, Europe, Africa, South America, Arab States, Middle East, and North America. It shows that the APAC has the highest corruption and the order of regions in corruption matches the order of emigration. However, there is one outlier as North America which is lowest in emigration. The second graph for freedom shows that the highest freedom can be seen in North America. Thus, we can conclude that the combination of corruption and freedom contribute to emigration.

Figure 11

Final Dashboard



Notes. The Figure shows all the graphs created for this project on one single dashboard.

6. Deliverables

The project deliverables for this project are as follows:

Table 3

Summary of Project Deliverables

Deliverable	Phase	Explanation
Project proposal document	Initial phase	Includes summary of project, its benefits and plan.
Data Prep Cleaning File	Final phase	CSV file generated after data preparation
Data Prep Flow File	Final phase	.Hyper to show flow chart by tableau prep builder
Tableau visualization	Final phase	Contains all visualizations, dashboard, and storytelling
Final report	Final phase	Detailed documentation project

Note. Table 3 shows the deliverable given in different phases of the project with its explanation.

7. Project risks

The Table 4 shows the risk possibility and its prevention action that can be taken.

Table 4

Project Risks

Possibility	Risk	Preventive action
There could be more factors contributing to emigration	The results are limited to only information collected. But there could be more factors contributing to emigration.	More factors should be involved for not having biased results

8. Communication

The communication involves all the files which will be shared through the canvas. I am also sharing the GitHub details for the same.

8.1 Canvas

Table 5

Details of Person Worked on Project

SJSU ID	NAME	EMAIL ID
015324478	Ritu Bhamrah	ritu.bhamrah@sjsu.edu

Note. The table contains the name, student id, and email address for person worked on this project.

8.2 GitHub

Table 6

GitHub Details of Person Worked on Project

GitHub ID	GitHub Email
ritub5	bhamrah.ritu@gmail.com

Note. The table contains the GitHub ID and email address for person worked on this project.

9. Project plan and Time Schedule

The project plan and time schedule are maintained by dividing the project into various tasks and milestones. For this project we have divided the project into milestones as shown in Figure 12. This also includes the data when each phase will be completed.

Figure 12

Milestones with Date of Delivery

Note.

The figure explains the different phases of the project - project proposal, initial study, data preparation, planning, data visualization, presentation, and report. It also includes dates when each phase will be completed.

The detailed explanation of each phase is as shown below:

- Project proposal - The initial document for project proposal.
- Initial study - In this phase we try to understand the factors which can influence the emigration of different regions. This will help in the data collection process.
- Data preparation and planning - We collect the data from various Kaggle sources. We will do data preparation on raw data and merge all the files.

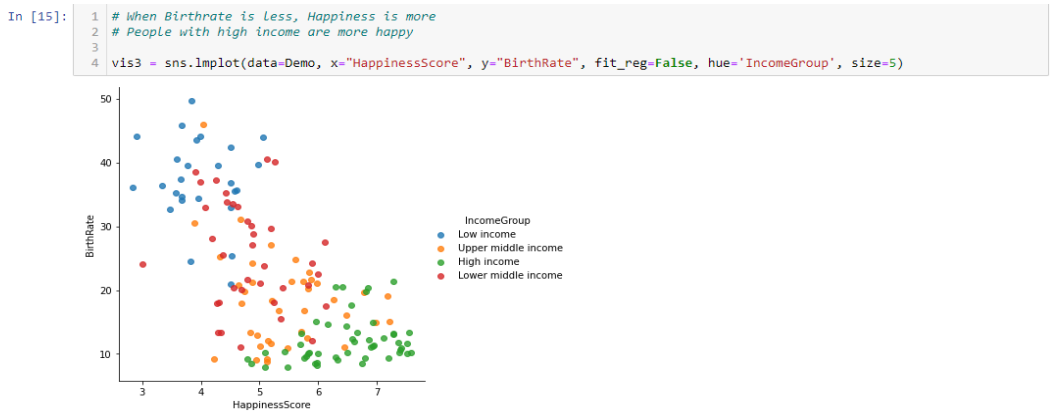
- Execute visualization - We perform visualization using the clean data file generated by tableau data prep builder.
- Presentation and report - We present our work and findings with the class and document the work in detail.

10. Test Plan

For testing the visualizations, we imported the data to a Jupyter notebook and checked if the graphs were correct or not. Some of the visualizations can be seen as below:

Figure 13

Happiness Score and Birth Rate with respect to Income Group



Note. The Figure 13 is the same graph which we also created from tableau. The results obtained in both cases are the same.

References

Causes of Emigration. (n.d.). MayoCoCo. https://www.mayo.ie/library/local-history/historical-events/emigration/causes-of-emigration

Castelli, F. (2018). Drivers of migration: why do people move? *Journal of Travel Medicine*, *25*(1). https://doi.org/10.1093/jtm/tay040

Canache, D., Hayes, M., Mondak, J. J., & Wals, S. C. (2013). Openness, extraversion and the intention to emigrate. *Journal of Research in Personality*, *47*(4), 351–355. https://doi.org/10.1016/j.jrp.2013.02.008

World Demographics. (2020, November 28). Kaggle. https://www.kaggle.com/hbao12/world-demographics?select=P2-Demographic-Data.csv

World Happiness Report up to 2020. (n.d.). Kaggle. https://www.kaggle.com/mathurinache/world-happiness-report

Emigration data up to 2020. (n.d.). OECD. https://stats.oecd.org/Index.aspx?DataSetCode=MIG#

Mlambo, V. (2018). An overview of rural-urban migration in South Africa: its causes and implications. *Archives of Business Research*, *6*(4). https://doi.org/10.14738/abr.64.4407