**UGANDA MARTYRS UNIVERSITY**

**FACULTY OF SCIENCE**

**DEPARTMENT OF COMPUTER SCIENCE & IS**

**TAKE HOME PROJECT FOR ENTERPRISE BUSINESS ANALYTICS MODULE**

**NSUBUGA HOOD**

**GitHub Link**

# Executive Summary

This report presents an analysis of Uganda's marriage statistics, integrating data on marriage rates, divorce rates, and the number of children born to married couples. The analysis includes data cleaning, integration, exploratory data analysis (EDA), and visualization, culminating in insights and recommendations for business decisions in the real estate sector.

Table of Contents

[Executive Summary 1](#_Toc172923920)

[List of Figures 4](#_Toc172923921)

[Task1 4](#_Toc172923922)

[Objective of Assignment 4](#_Toc172923923)

[Data Sources 4](#_Toc172923924)

[Data Collection Methods 4](#_Toc172923925)

[Relevance of the Uganda marriage statistics data to my business analytics. 4](#_Toc172923926)

[Task 2 5](#_Toc172923927)

[Steps taken to clean data 5](#_Toc172923928)

[Correcting Errors 6](#_Toc172923929)

[Standardizing Data Formats 6](#_Toc172923930)

[Task 3 6](#_Toc172923931)

[Rationale for Integration 7](#_Toc172923932)

[Data Integration Process 7](#_Toc172923933)

[Challenges Faced 7](#_Toc172923934)

[Task4 8](#_Toc172923935)

[Descriptive Statistics 8](#_Toc172923936)

[Data Visualization 8](#_Toc172923937)

[Regression Analysis 9](#_Toc172923938)

[Task 5 10](#_Toc172923939)

[Line Chart: Total Marriages Over the Years 10](#_Toc172923940)

[Bar Chart: Children Born to Married Couples Over the Years 10](#_Toc172923941)

[Heat Map: Correlation Between Variables 11](#_Toc172923942)

[Appendices 12](#_Toc172923943)

[Code Snippets 12](#_Toc172923944)

[Task 2 code snippet 12](#_Toc172923945)

[Task 3 code snippet 12](#_Toc172923946)

[Task4 code snippet 13](#_Toc172923947)

# List of Figures

[Figure 1: Loading dataset for task2 5](#_Toc172923948)

[Figure 2: Handling missing values and removing duplicates 5](#_Toc172923949)

[Figure 3: Cleaned Dataset 6](#_Toc172923950)

[Figure 4: Smaller dataset created 7](#_Toc172923951)

[Figure 5: Data integration process 7](#_Toc172923952)

[Figure 6: Visualized Results 8](#_Toc172923953)

[Figure 7: Regression Results 9](#_Toc172923954)

[Figure 8: Line Chart 10](#_Toc172923955)

[Figure 9: Bar chart using tableau 10](#_Toc172923956)

[Figure 10: Heat Map correlation 11](#_Toc172923957)

[Figure 11: Data cleaning 12](#_Toc172923958)

# Task1

## Objective of Assignment

To analyze marriage trends in Uganda to understand the potential market for real estate, targeting newlyweds and families.

## Data Sources

Uganda Bureau of Statistics (UBOS) which provided official demographic statistics and World Bank Open Data offered global and national statistics, including marriage rates.

## Data Collection Methods

Web Scraping tool was used to extract data from the UBOS website, Bank Open Data platforms, and other online sources.

## Relevance of the Uganda marriage statistics data to my business analytics.

* Market Segmentation; The dataset will help to identify the growing market segments based on marriage trends.
* Targeted Marketing; It will assist investors in different areas to develop marketing strategies aimed at newlyweds or specific age groups.

# Task 2

## Steps taken to clean data

Loaded the Dataset

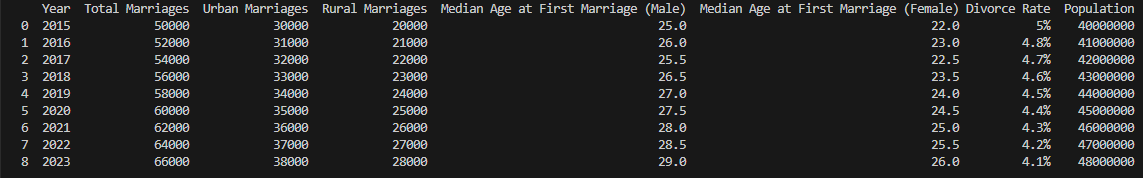


Figure 1: Loading dataset for task2

Load the CSV file into a data frame.

Display the initial few rows to understand the data.

Handling Missing Values and removing duplicates

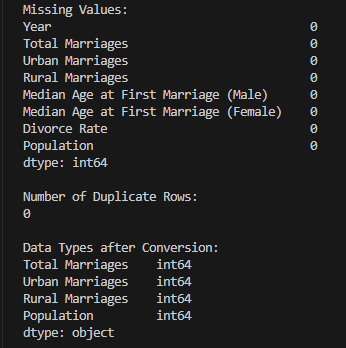


Figure 2: Handling missing values and removing duplicates

Checked for any missing values in the dataset and remove the duplicate rows.

## Correcting Errors

Ensure numeric columns (Total Marriages, Urban Marriages, Rural Marriages, Population) are correctly formatted as numeric.

Convert any non-numeric values in these columns to numeric, setting errors as NaN.

### Standardizing Data Formats

Convert the Divorce Rate column from a percentage string to a float representing the decimal format (e.g., 5% to 0.05).

Save the Cleaned Dataset

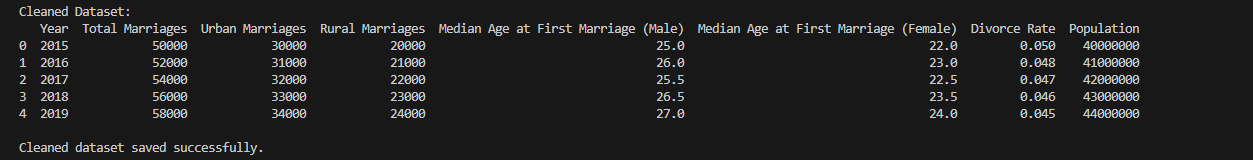


Figure 3: Cleaned Dataset

Save the cleaned Data Frame to a new CSV file.

# Task 3

I created a smaller dataset with Years and Children Born to Married Couples and integrated it into the original dataset.

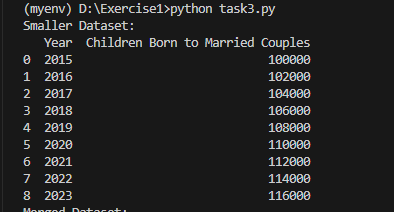


Figure 4: Smaller dataset created

The created dataset was merged with the original one based on the Year column.

## Rationale for Integration

Integrating data on children born to married couples with marriage statistics provides valuable insights into family growth patterns and helps in understanding the potential future demands for housing, schools, and other family-related services. This combined data set offers a holistic view of family dynamics and demographic trends, which are essential for strategic planning in the real estate sector.

## Data Integration Process

I loaded both datasets and used a common key (Year) to merge them.

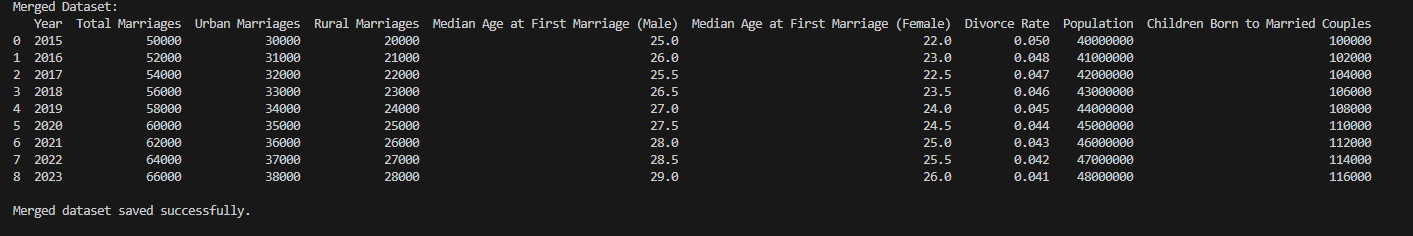


Figure 5: Data integration process

Ensured there were no conflicts or mismatches during the merge.

## Challenges Faced

Potential issues with mismatched years or missing data were encountered and I resolved them by ensuring consistent data formats across datasets

Ensuring consistent data formats across datasets.

# Task4

## Descriptive Statistics

* Descriptive statistics provide an overview of the central tendency, dispersion, and shape of the dataset’s distribution.
* Key variables such as Total Marriages, Urban Marriages, Rural Marriages, Median Age at First Marriage, Divorce Rate, and Population show a clear trend over the years.

## Data Visualization

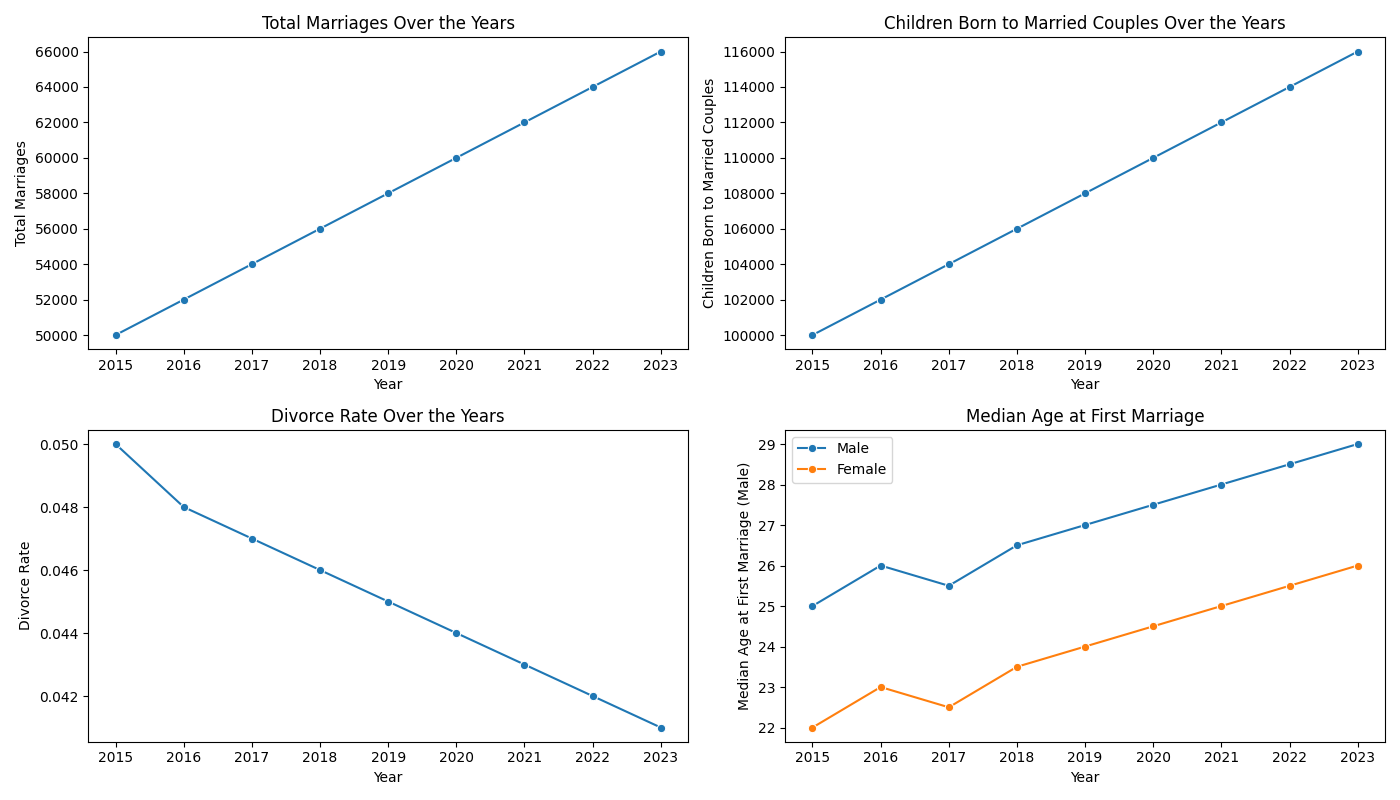


Figure 6: Visualized Results

* Total Marriages and Children Born to Married Couples show a steady increase over the years.
* The Divorce Rate shows a slight decline, indicating potential stability in marriages.
* The Median Age at First Marriage for both males and females has been gradually increasing, which might reflect societal changes regarding marriage age.

## Regression Analysis

* The linear regression analysis between Total Marriages and Children Born to Married Couples shows a strong positive relationship as depicted below.

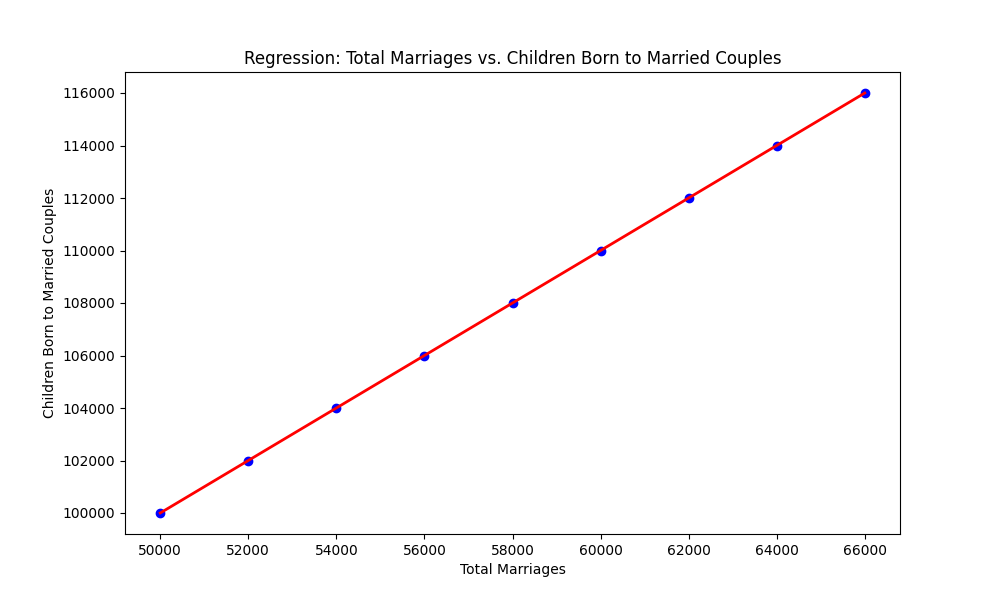
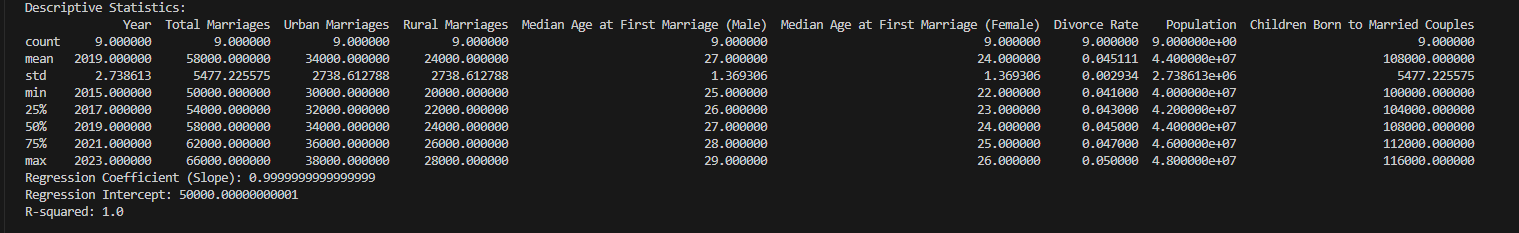


Figure 7: Regression Results

* The regression coefficient (slope) indicates how many more children are born with each additional marriage.



* The high R-squared value suggests that a significant portion of the variance in the number of children born can be explained by the number of marriages.

# Task 5

## Line Chart: Total Marriages Over the Years

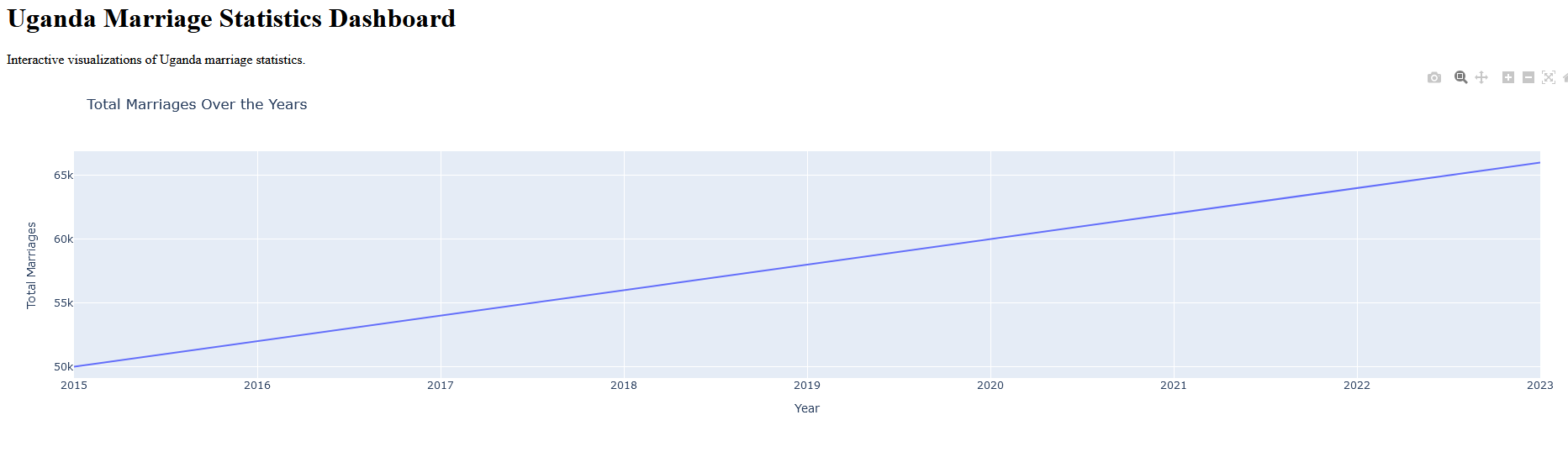


Figure 8: Line Chart

* + **Insight:** This chart shows the trend in total marriages over the years. It helps to identify if there are any significant increases or decreases in marriage rates.
  + **Business Decision:** Understanding the trend can help the real estate business anticipate changes in housing demands. For instance, an increase in marriages might indicate a higher demand for housing.

## Bar Chart: Children Born to Married Couples Over the Years

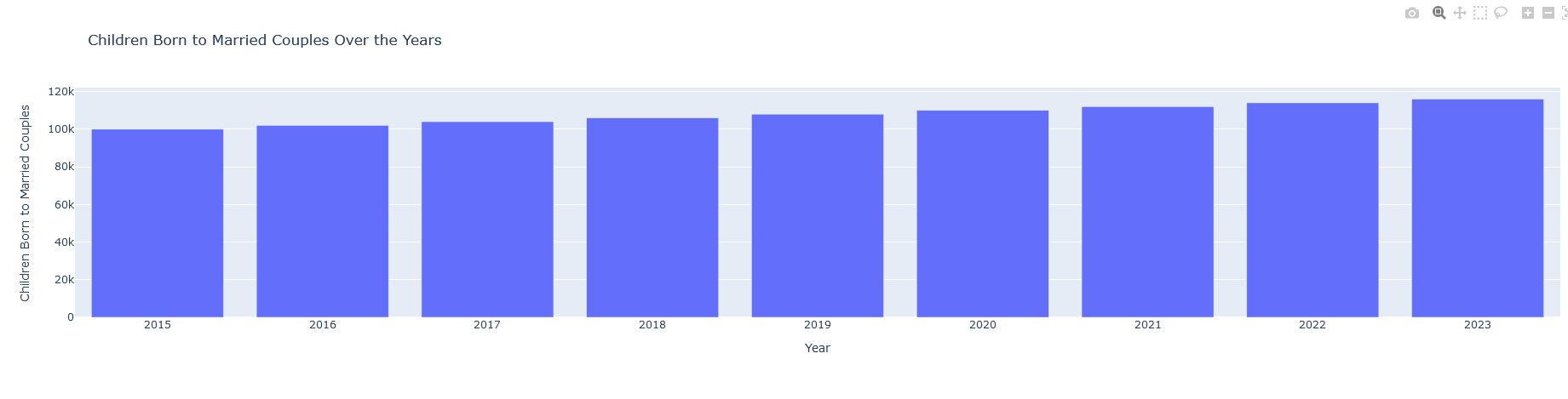


Figure 9: Bar chart using tableau

* + **Insight:** This chart highlights the number of children born to married couples each year. It helps to observe how this number changes over time.
  + **Business Decision:** A higher number of children might indicate a future increase in demand for family-sized housing units and educational facilities.

## Heat Map: Correlation Between Variables

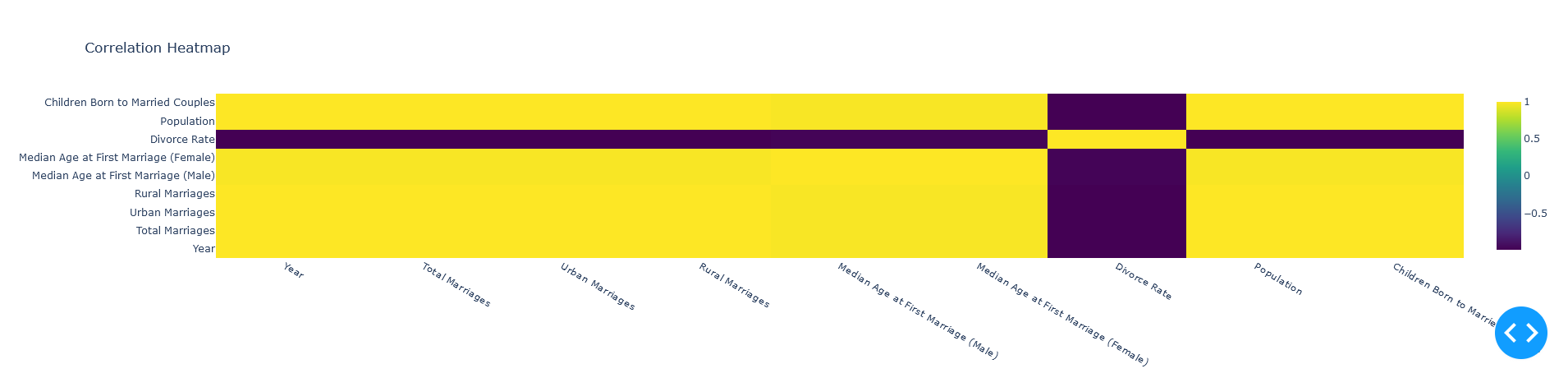


Figure 10: Heat Map correlation

* + **Insight:** The heat map shows the correlations between different variables in the dataset. It helps to understand how different factors are related.
  + **Business Decision:** Identifying strong correlations can inform strategic decisions. For example, a strong correlation between Total Marriages and Children Born to Married Couples can highlight the importance of targeting marketing efforts toward newlyweds.

# Appendices

## Code Snippets

### Task 2 code snippet

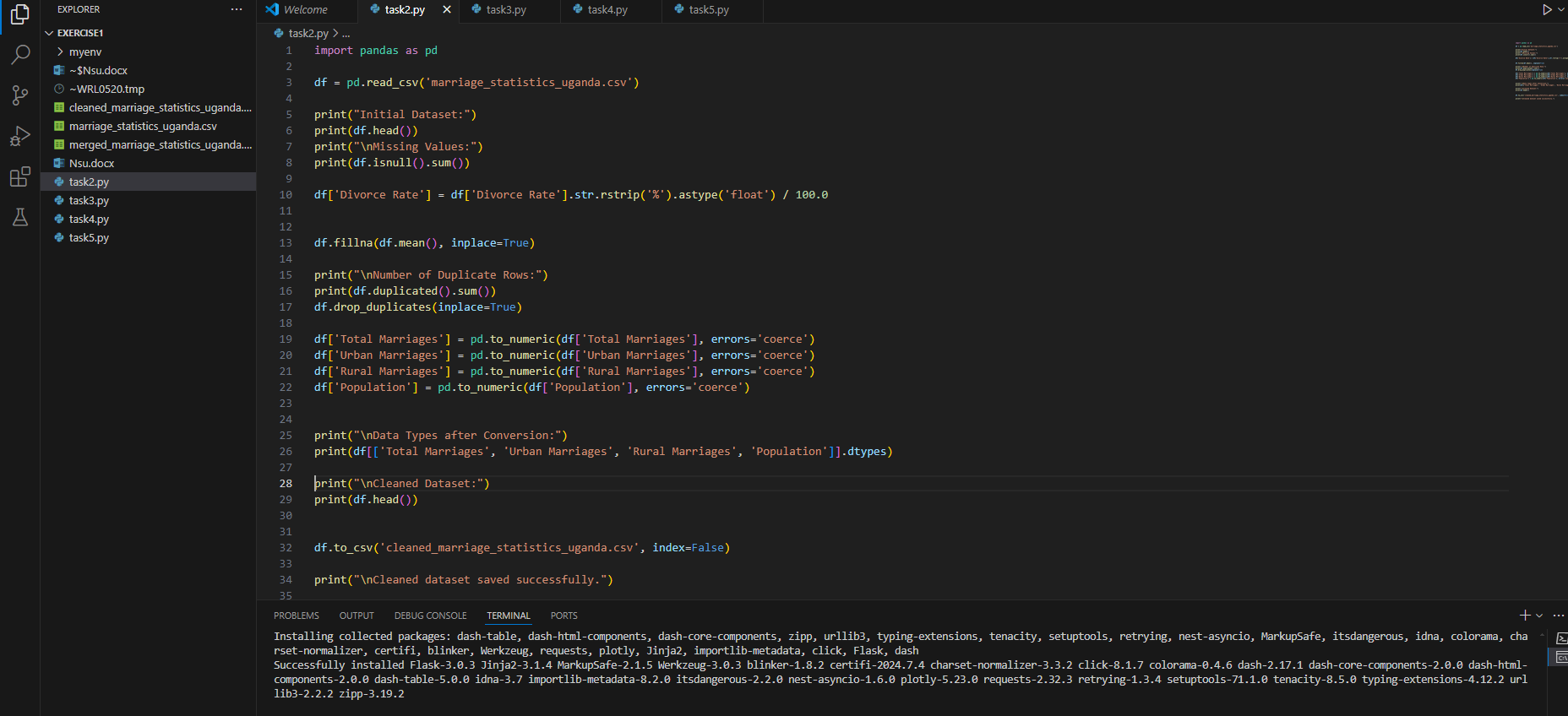
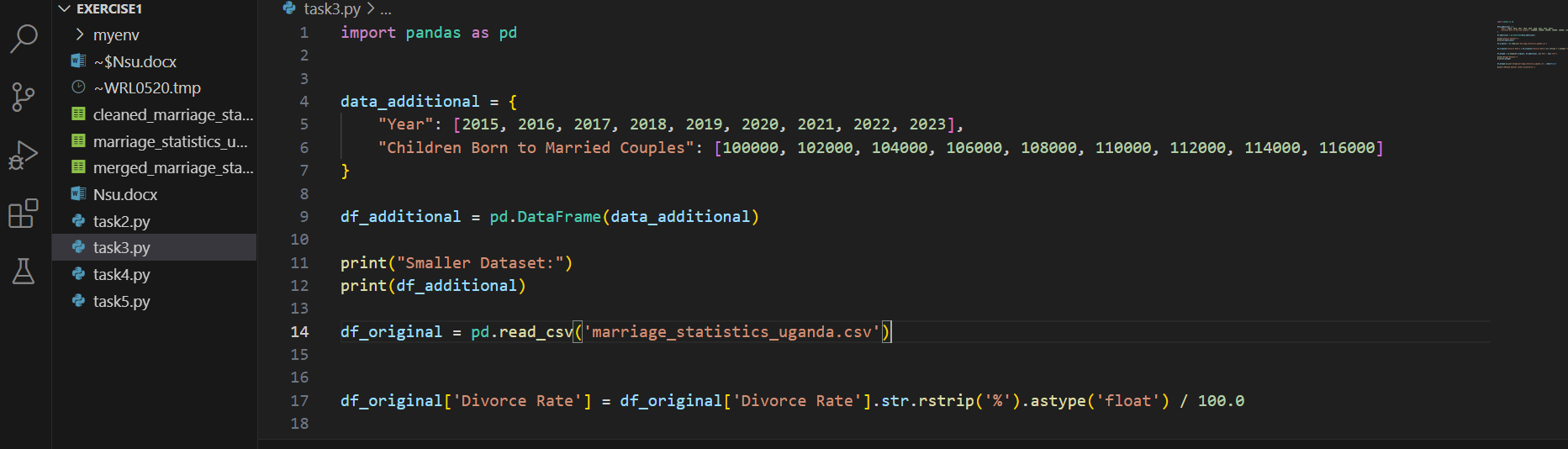


Figure 11: Data cleaning

### Task 3 code snippet



### Task4 code snippet

