**HO CHI MINH UNIVERSITY OF SCIENCE**

**FACULTY OF INFORMATION TECHNOLOGY**

**REPORT HW2**

**Subject: Data Visualization Class:** 22KHDL1

**Students:**

Trương Thuận Kiệt – 22127224 Phạm Minh Mẫn - 22127257

**TP. Hồ Chí Minh, tháng 3, năm 2025**

Contents

[**1.** **Group information and task assignments** 3](#_Toc192441760)

[**2.** **External Used Libraries** 3](#_Toc192441761)

[**3.** **Data Crawling** 3](#_Toc192441762)

[**a.** **Spotify API** 3](#_Toc192441763)

[**b.** **Mystreamcount** 6](#_Toc192441764)

[**c.** **Getgenre and Musicbrainz** 8](#_Toc192441765)

[**4.** **Data Overview and Preprocessing** 8](#_Toc192441766)

[**a.** **Shape of data** 8](#_Toc192441767)

[**b.** **Meaning of each row** 8](#_Toc192441768)

[**c.** **Meaning of each column** 8](#_Toc192441769)

[**d.** **Duplicate** 9](#_Toc192441770)

[**e.** **Data Retyping** 9](#_Toc192441771)

[**f.** **Missing Values** 10](#_Toc192441772)

[**g.** **Distribution of Numerical Values** 10](#_Toc192441773)

[**i.** **Songs Dataset** 10](#_Toc192441774)

[**ii.** **Artists Dataset** 11](#_Toc192441775)

[**h.** **Distribution of Categorical Values** 12](#_Toc192441776)

[**i.** **Songs Dataset** 12](#_Toc192441777)

[**ii.** **Artists Dataset** 14](#_Toc192441778)

[**5.** **Questions** 15](#_Toc192441779)

[**a.** **Question 1** 15](#_Toc192441780)

[**b.** **Question 2** 20](#_Toc192441781)

[**c.** **Question 3** 23](#_Toc192441782)

[**d.** **Question 4** 25](#_Toc192441783)

[**e.** **Question 5** 27](#_Toc192441784)

1. **Group information and task assignments**

|  |  |  |  |
| --- | --- | --- | --- |
| **ID** | **Full Name** | **Tasks** | **Rate of completion** |
| **22127224** | **Trương Thuận Kiệt** | * Crawl data from World Bank API. * Provide the overview of dataset (numerical columns). * Pose and answer problems 1 and 2.   Write report. |  |
|  |  |  |  |
|  |  |  |  |

1. External Used Libraries
   1. Requests

* Used to crawl data from API
  1. Geopandas
* To plot the map of the world

1. Data Crawling
2. Data Overview and Preprocessing
3. Questions