

## Q1 Report

### Project 2 Group 6

08/15/25

The goal of Q1 was to take a scraped table of country statistics and add extra information from an API so we could find new insights. We started with Worldometers' population projections and used the Rest Countries API to get more details like geography, demographics, and languages. We used Python with pandas and requests for scraping, API calls, and cleaning. HTML scraping was done with BeautifulSoup. We made API calls, worked with the JSON results, and merged everything into CSV files that we could summarize and run stats on with pandas.

From Worldometers, we scraped the "Population by Country" table, which we called df1. It had 233 rows and 11 columns. We matched each country in that list to the Rest Countries API and pulled things like ISO code, region, subregion, capital, area, population, coordinates, and number of languages. That API data became df2. After fixing name mismatches and any missing results, we merged df1 and df2 by country to make df3, which had 21 columns. All of these datasets were saved in the data folder, and we ran descriptions on them for analysis.

The final dataset has 233 countries and dependencies. The average area is about 605,354 km<sup>2</sup>, with the smallest at 0.44 km<sup>2</sup> and the largest over 1.8 million km<sup>2</sup>. The average population from the API is 34.7 million, ranging from 451 people to over 1.3 billion. On average, countries have about 1.7 languages, with some having as many as 8. There are 5 unique regions and 24 subregions. The median age from the scraped data averages 31.9 years, with the youngest country's median age being 14.5 years.

Combining scraped data with API data makes it easier to double-check information and spot errors. We had to fix country name issues for the merge to work. In the future, it would be useful to refresh the data automatically and include more time-based fields.