Milestone 1

10% of the final grade. This is a preliminary milestone to let you set up goals for your final project and assess the feasibility of your ideas. Please, fill the following sections about your project. (max. 2000 characters per section)

Dataset

Find a dataset (or multiple) that you will explore. Assess the quality of the data it contains and how much preprocessing / data-cleaning it will require before tackling visualization. We recommend using a standard dataset as this course is not about scraping nor data processing.
Hint: some good pointers for finding quality publicly available datasets (Google dataset search, Kaggle, OpenSwissData, SNAP and FiveThirtyEight)

In this project, we intend to utilize the Numbeo dataset, which is the largest global database of cost of living information. Numbeo is a crowd-sourced repository encompassing data on various aspects of quality of life, such as housing indicators, perceived crime rates, healthcare quality, transportation quality, and other relevant statistics.

The dataset comprises cost of living index data spanning from 2009 to 2023, covering 540 cities across 140 countries. This extensive dataset offers a valuable resource for analyzing the variations in living costs across time and geographical locations. Additionally, it encompasses a diverse range of 55 price categories, including food, housing, healthcare, and more. Consequently, Numbeo provides comprehensive insights into the living expenses in cities and countries worldwide.

Problematic

Frame the general topic of your visualization and the main axis that you want to develop.
- *What am I trying to show with my visualization?*
- *Think of an overview for the project, your motivation, and the target audience.*

Title: Global Living Costs Unveiled: A Comparative Analysis Across Time and Space

General Topic: Visualizing the evolution of living costs around the world, highlighting the most significant trends and disparities.

Main Axis: Comparison of living costs across cities and countries, with a focus on cost of living indices, rent, groceries, restaurant prices, and local purchasing power.

Overview: This visualization aims to provide a comprehensive and engaging perspective on global living costs, revealing key patterns and differences across various geographical locations and over time. By presenting a data-driven analysis, our goal is to offer valuable insights for individuals, policymakers, and businesses interested in understanding the dynamics of living expenses worldwide.

Motivation: The motivation behind this project is to raise awareness of the factors affecting living costs and to facilitate better decision-making for people considering relocation, investments, or policy changes.

Target Audience: Our target audience includes individuals considering relocation, policymakers involved in urban planning and development, businesses seeking to expand or invest in new locations, and researchers interested in the study of living costs and their socio-economic implications.

Exploratory Data Analysis

Pre-processing of the data set you chose
• Show some basic statistics and get insights about the data

Web scraping and data pre-processing: As the first step, we scrapped the dataset from Numbeo website. Under the most general Quality of Life Index, there are eight sub-indices calculated from a wide range of factors listed as follows.

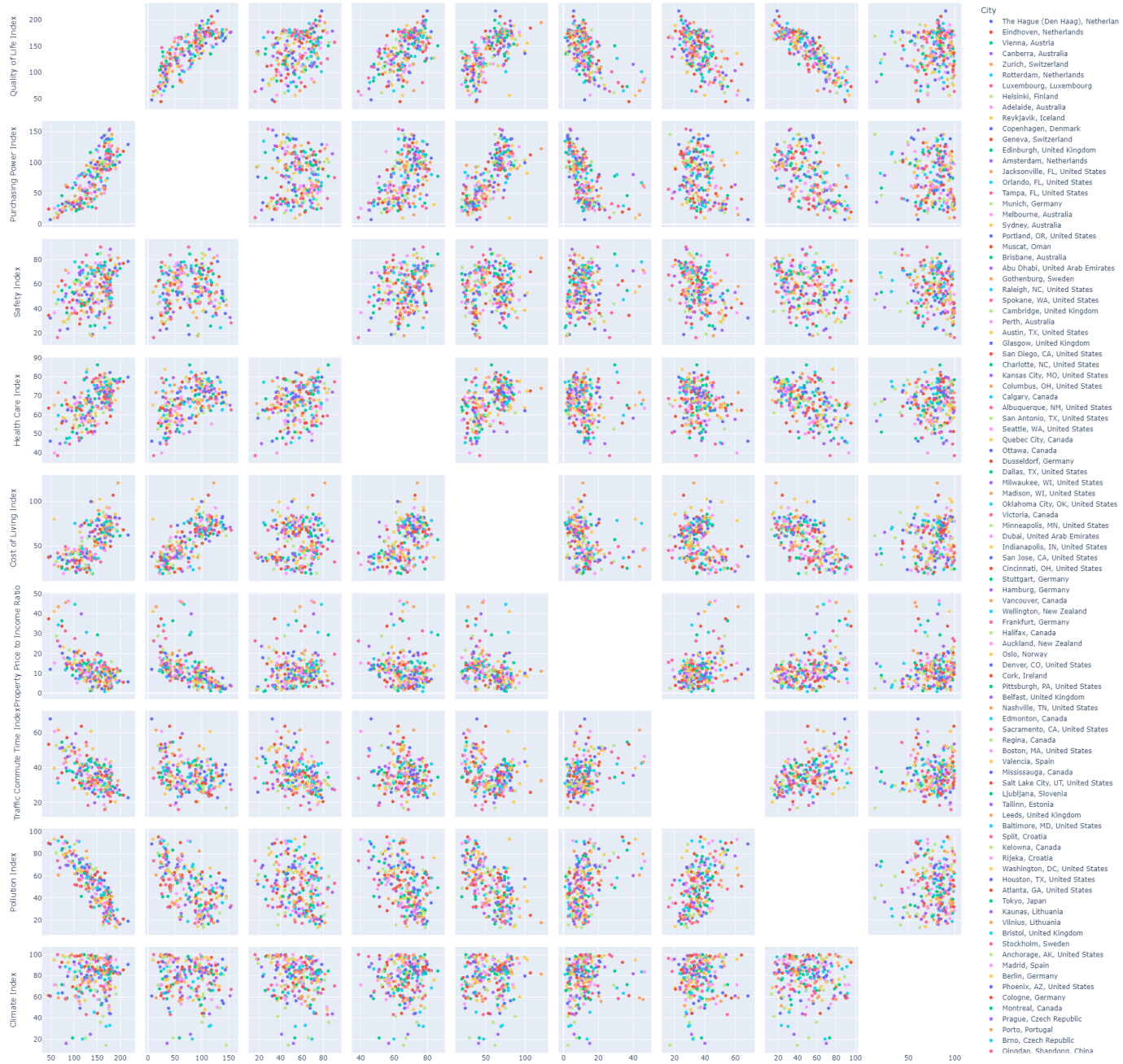
Quality of Life Index	● Quality of Life Index ● Purchasing Power Index ● Safety Index ● Health Care Index ● Cost of Living Index ● Property Price to Income Ratio ● Traffic Commute Time Index ● Pollution Index ● Climate Index
Cost of Living Index by City	● Cost of Lixing Index ● Rent Index ● Cost of Living Plus Rent Index ● Groceries Index ● Restaurant Price Index ● Local Purchasing Power Index

Property Prices Index by City	• Price To Income Ratio • Gross Rental Yield City Centre • Gross Rental Yield Outside of Centre • Price To Rent Ratio City Centre • Price To Rent Ratio Outside Of City Centre • Mortage As A Percentage Of Income • Affordability Index
Crime Index by City	• Crime Index • Safety Index
Health Care Index by City	• Health Care Index • Health Care Exp. Index
Pollution Index by City	• Pollution Index • Exp. Pollution Index
Traffic Index by City	• Traffic Index • Time Index (in minutes) • Time Exp. Index • Inefficiency Index • CO2 Emission Index

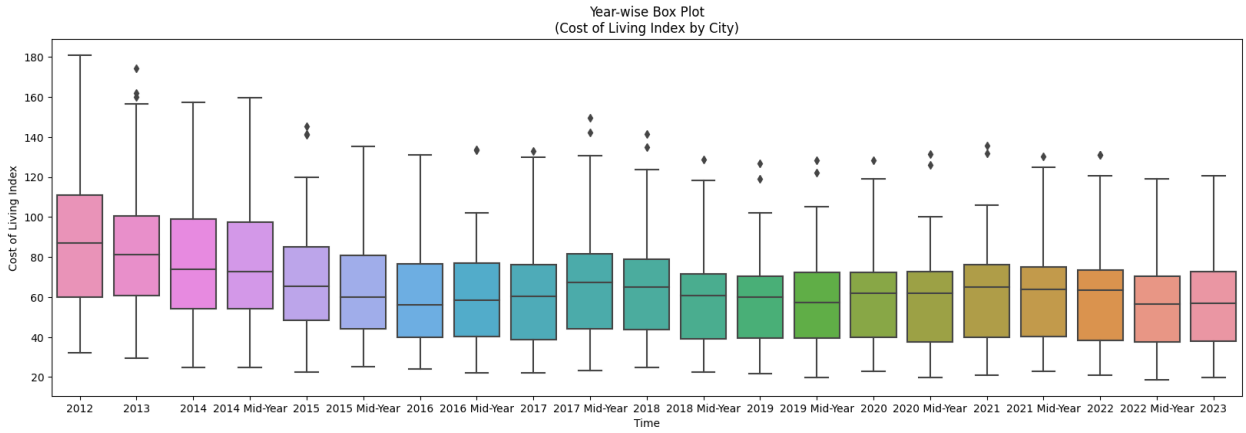
Table 1. Categories of Indices and sub-indices

Statistical analysis: By visualizing the scatter plots among Quality of Life Index and its sub-indices, we could observe its significant positive relation with Purchasing Power Index and negative relation with Pollution Index and Traffic Commute Time Index. Local Purchasing Power shows relative purchasing power in buying goods and services in a given city for the average net salary in that city, which is calculated from the estimated cost of living and salary. It explained the trends shown in the subplots of Cost of Living Index. And Property Price to Income Ratio.

Therefore, the cost of living, property price, and pollution level could be regarded as the most dominant factors for the quality of our life. Since detailed information is not provided on this website, we decided to put our focus on the cost of living.



By plotting the time-series data of the Quality of Life Index and the Cost of Living Index, we could find the worldwide price decreased between 2012 to 2016, meanwhile, the quality of life was improved. However, further investigation is needed to explain the steady trend in the recent 7 years, which requires our visualizations among the categories under the Cost of Living Index and the variation of cities/countries.



name	category	cost_of_living_factor	rent_factor
1 Pair of Jeans (Levis 501 Or Similar)	Clothing And Shoes	0.35	0
1 Summer Dress in a Chain Store (Zara, H&M, ...)	Clothing And Shoes	0.35	0
1 Pair of Nike Running Shoes (Mid-Range)	Clothing And Shoes	0.35	0
1 Pair of Men Leather Business Shoes	Clothing And Shoes	0.35	0
Milk (regular), (1 liter)	Markets	25	0
Loaf of Fresh White Bread (500g)	Markets	31	0
Rice (white), (1kg)	Markets	14	0

Eggs (regular) (12)	Markets	20	0
Local Cheese (1kg)	Markets	12	0
Chicken Breasts (Boneless, Skinless), (1kg)	Markets	15	0
Beef Round (1kg) (or Equivalent Back Leg Red Meat)	Markets	15	0
Apples (1kg)	Markets	31	0
Banana (1kg)	Markets	25	0
Oranges (1kg)	Markets	30	0
Tomato (1kg)	Markets	22	0
Potato (1kg)	Markets	24	0
Onion (1kg)	Markets	10	0
Lettuce (1 head)	Markets	18	0
Water (1.5 liter bottle)	Markets	30	0
Bottle of Wine (Mid-Range)	Markets	4	0
Domestic Beer (0.5 liter bottle)	Markets	6	0
Imported Beer (0.33 liter bottle)	Markets	6	0
Cigarettes 20 Pack (Marlboro)	Markets	15	0
Apartment (1 bedroom) in City Centre	Rent Per Month	0	0.25
Apartment (1 bedroom) Outside of Centre	Rent Per Month	0	0.25
Apartment (3 bedrooms) in City Centre	Rent Per Month	0	0.25
Apartment (3 bedrooms) Outside of Centre	Rent Per Month	0	0.25
Meal, Inexpensive Restaurant	Restaurants	16	0
Meal for 2 People, Mid-range Restaurant, Three-course	Restaurants	3.5	0
McMeal at McDonalds (or	Restaurants	6	0

Equivalent Combo Meal)			
Domestic Beer (0.5 liter draught)	Restaurants	5	0
Imported Beer (0.33 liter bottle)	Restaurants	5	0
Cappuccino (regular)	Restaurants	15	0
Coke/Pepsi (0.33 liter bottle)	Restaurants	6	0
Water (0.33 liter bottle)	Restaurants	6	0
Fitness Club, Monthly Fee for 1 Adult	Sports And Leisure	2.3	0
Tennis Court Rent (1 Hour on Weekend)	Sports And Leisure	3	0
Cinema, International Release, 1 Seat	Sports And Leisure	6	0
One-way Ticket (Local Transport)	Transportation	20	0
Monthly Pass (Regular Price)	Transportation	1.5	0
Taxi Start (Normal Tariff)	Transportation	5	0
Taxi 1km (Normal Tariff)	Transportation	20	0
Taxi 1hour Waiting (Normal Tariff)	Transportation	0.7	0
Gasoline (1 liter)	Transportation	60	0
Volkswagen Golf 1.4 90 KW Trendline (Or Equivalent New Car)	Transportation	0.0035	0
Toyota Corolla 1.6l 97kW Comfort (Or Equivalent New Car)	Transportation	0.0035	0
Basic (Electricity, Heating, Cooling, Water, Garbage) for 85m2 Apartment	Utilities (Monthly)	1	0
1 min. of Prepaid Mobile Tariff Local (No Discounts or Plans)	Utilities (Monthly)	320	0
Internet (60 Mbps or More, Unlimited Data, Cable/ADSL)	Utilities (Monthly)	1	0

Table 2. Calculation Indices of Cost of Living

Related work

What others have already done with the data?

The source website, Numbeo, offers various views for exploring the data, with sections dedicated to comparing living costs between cities or countries and focusing on specific goods and products. Users can perform direct comparisons between locations and access detailed information, such as property prices and apartment locations. Some pages also provide color-coded world maps for an easy overview of certain statistics.

The data is referenced on Kaggle, where users have shared analyses and visualizations through notebooks, though some datasets may be outdated.

Why is your approach original?

Our project aims to enhance data exploration by retaining the website's versatility while increasing the information users can glean from the tool. We intend to create visualizations that quickly reveal the city with the lowest living cost and the size of the gap with other cities, accommodating a large number of locations. The tool will also be applicable to the entire range of columns, necessitating multiple visualization types for optimal readability. Furthermore, we plan to include a time series analysis for insights on the evolution of living costs globally and comparisons between regions.

What source of inspiration do you take?

Kaggle notebooks provide insights on potential analyses for the data, while the MakeoverMonday offers a variety of ideas for dynamic data presentation across numerous topics, with new content added weekly.