REPORT

**PROJECT : DESIGN A WEB SCRAPE DATA FROM A WEBSITE. ANALYZE THE DATA AND MAKE A REPORT ON THE ANALYSIS**

**OBJECTIVES :**

1. To identify maximum of the index of each segment given to us about the cities provided in the data.
2. To identify minimum of the index of each segment given to us about the cities provided in the data.
3. To identify standard deviation of the index of each segment given to us about the cities provided in the data.
4. To identify variance of the index of each segment given to us about the cities provided in the data.
5. To compare cities on the basis of their index using correlation.
6. To analyze the different City’s statistics and identify key insights.
7. To calculate correlations to understand relationships within the data of various index.

**GENERAL DESCRIPTIONS OF DATA**

**Data Sources :**

I scrap the data from Website : “Numbeo” url : https://www.numbeo.com/quality-of-life/rankings.jsp.

The data consist of different index given for each city like quality of life index,Purchasing power index etc

**Libraries which are used by this projects are :**

from bs4 import BeautifulSoup

import requests

import pandas as pd

from sklearn import linear\_model as lm

from sklearn.metrics import r2\_score as rsq

import numpy as np

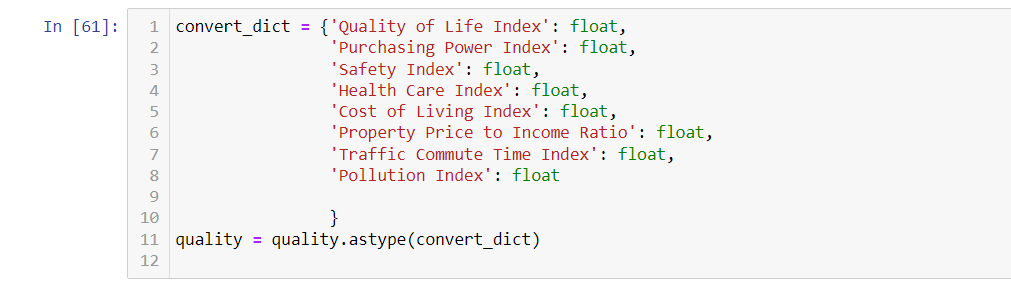
import seaborn as sns

import matplotlib.pyplot as plt

**Data Preprocessing steps**

1. Scrapping the data from the website
2. Checking if the output is object or float
3. Converting objects into floats.
4. Finding mean max min for every index
5. Doing further analysis.

**Converting into floats**

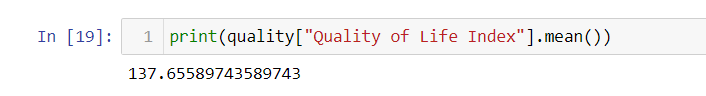
****

**Analysis:**

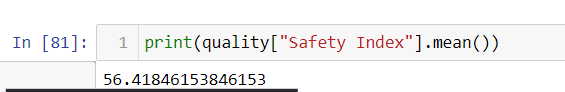
**Basic Statistics of the data:**

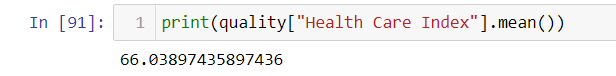
**Finding mean, maximum, minimum, standard variation, variance , value count, unique, nunique value for the data provided**

**For mean:**









Here we find that the mean which implies that this is the index that has been offered for the cities

FINDINGS :

1. Quality of Life Index (QLI):

- The average Quality of Life Index across the dataset is approximately 137.66. This index is indicative of the overall quality of life in various locations.

2. Purchasing Power Index (PPI):

- The average Purchasing Power Index is about 70.55. This index measures the relative purchasing power of individuals in different areas, indicating the potential for spending and economic well-being.

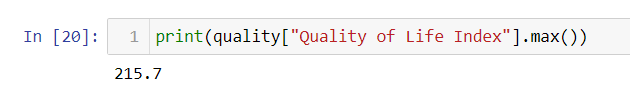
3. Safety Index:

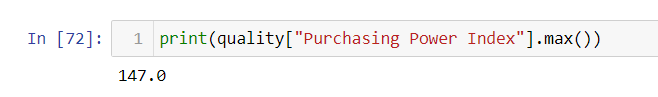
- The average Safety Index is around 56.42. This index reflects the level of safety in different regions and can be important for decision-making regarding personal safety and business investments.

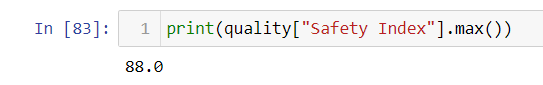
4. Health Care Index:

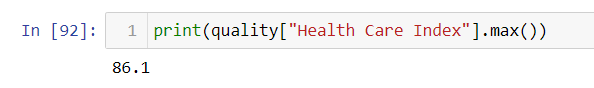
- The average Health Care Index is approximately 66.04. This index gauges the quality and accessibility of healthcare services in different locations.

**For maximum value:**









FINDINGS:

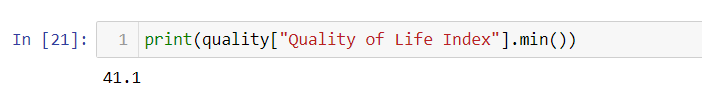
the max value in the index of Quality of life index is 215.7 which denotes The hague, Netherlands has the highest QLI.

the max value in the index of Purchasing Power index is 147 which denotes Houston,TX,US has the highest PPI.

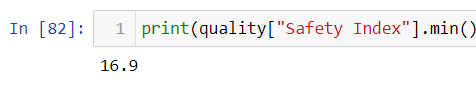
the max value in the index of Safety index is 88 which denotes Abu Dhabi,UAE has the highest SI.

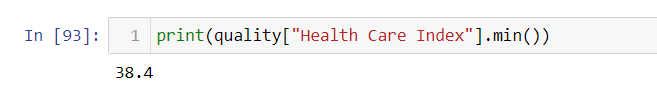
the max value in the index of Health Care index is 86.1 which denotes Taipei,Taiwan has the highest HCI.

**For minimum value:**









FINDINGS:

the min values in the index of Quality of Life index is 41.1 which denotes Manila, Philippines has lowest QLI.

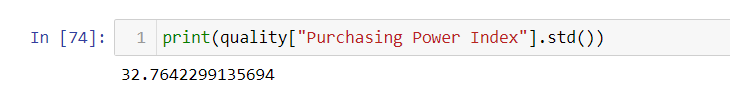
the min values in the index of Purchasing Power index is 9.7 which denotes Caracas, Venezuela has lowest PPI.

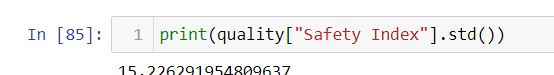
the min values in the index of Safety index is 16.9 which denotes Caracas, Venezuela has lowest SI.

the min values in the index of Health Care index is 38.4 which denotes Caracas, Venezuela has lowest HCI.

**For standard deviation:**









FINDINGS:

Quality of Life Index (QLI):

The standard deviation of the Quality of Life Index is approximately 34.47. This indicates a relatively high degree of variability in the quality of life across the surveyed locations. A higher standard deviation suggests that the scores are more spread out from the mean, signifying greater variation.

Purchasing Power Index (PPI):

The standard deviation of the Purchasing Power Index is approximately 32.76. This suggests a notable variation in purchasing power across the surveyed areas. Managers should be aware of this variability when planning marketing strategies or pricing structures.

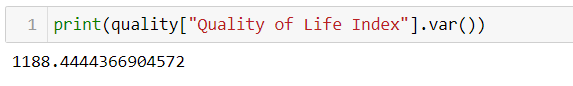
Safety Index:

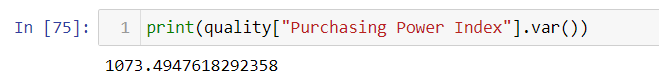
The standard deviation of the Safety Index is about 15.23. While there is still some variability, it is lower compared to the other indices. This implies that safety levels are relatively consistent across the surveyed locations.

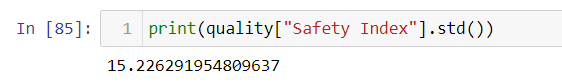
Health Care Index:

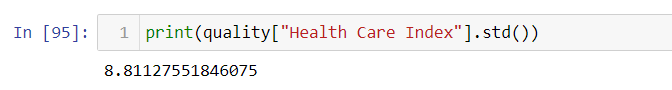
The standard deviation of the Health Care Index is approximately 8.81, which indicates relatively low variation in healthcare quality across the surveyed regions.

**For variance**









FINDINGS:

Quality of Life Index (QLI):

The variance of the Quality of Life Index is approximately 1188.44. This indicates a considerable spread or dispersion in the quality of life scores among the surveyed locations. A higher variance suggests a wider range of values.

Purchasing Power Index (PPI):

The variance of the Purchasing Power Index is approximately 1073.49. Similar to the Quality of Life Index, this indicates significant variation in purchasing power across the surveyed areas.

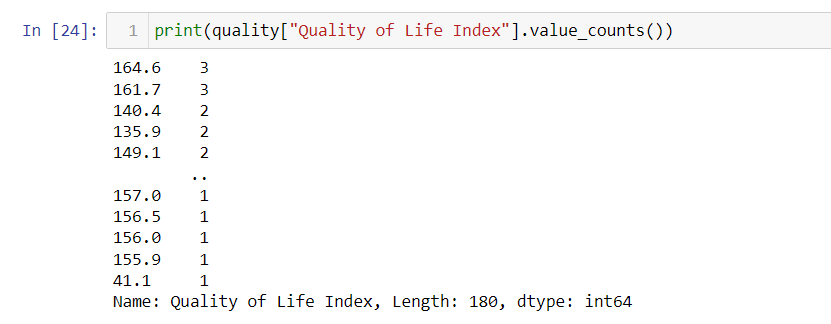
Safety Index:

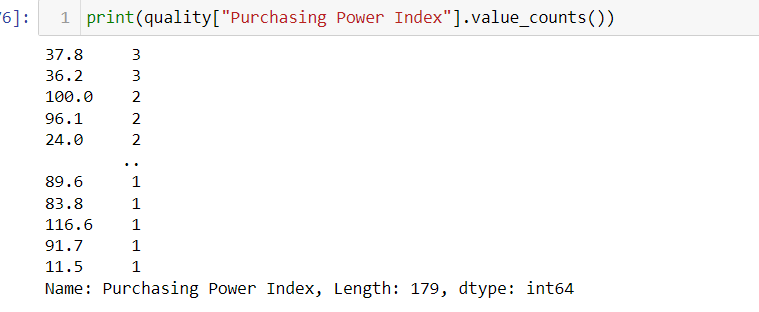
The variance of the Safety Index is about 231.84. While there is some variation, it is lower compared to the other indices, indicating that safety levels are relatively consistent across the surveyed locations.

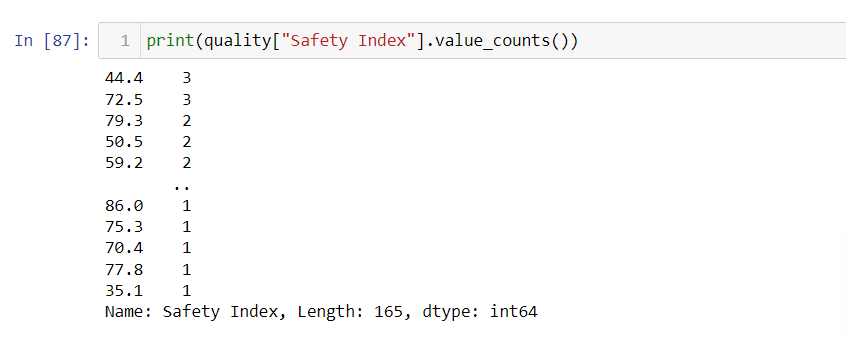
Health Care Index:

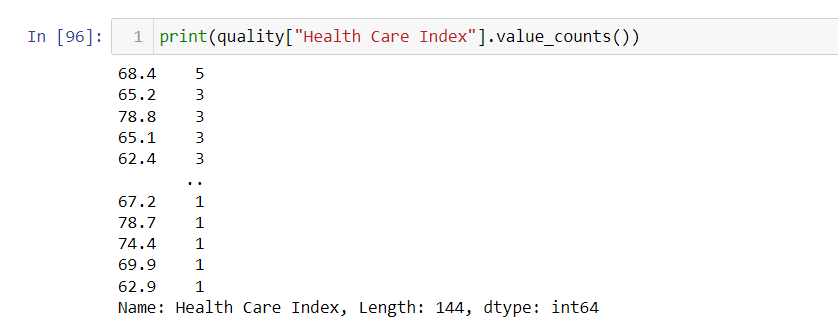
The variance of the Health Care Index is approximately 77.64, which suggests relatively low variation in healthcare quality across the surveyed regions.

**For value count (no. of times index are repeating)**



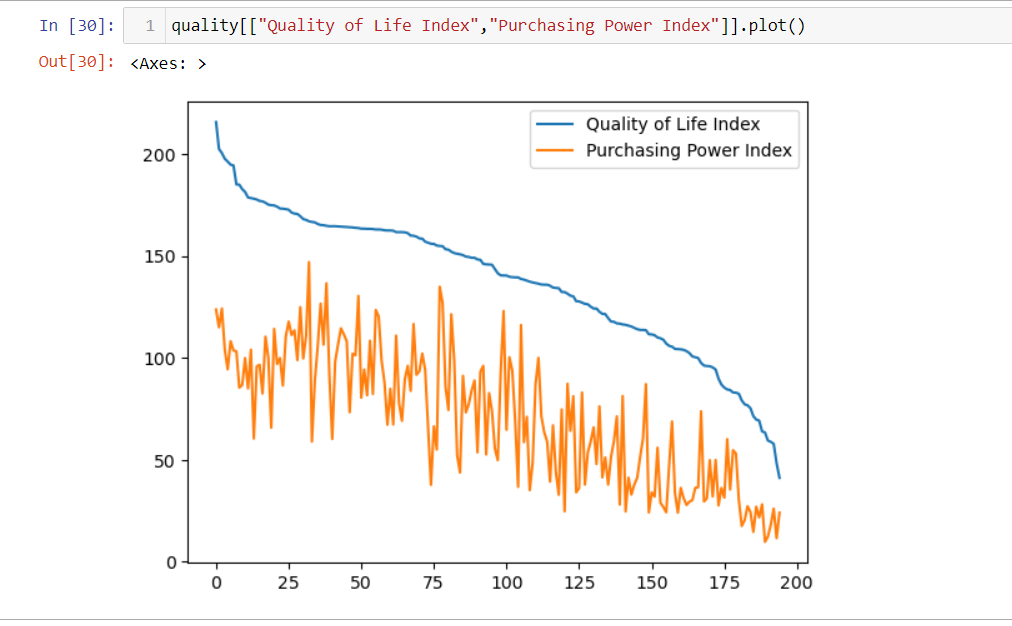


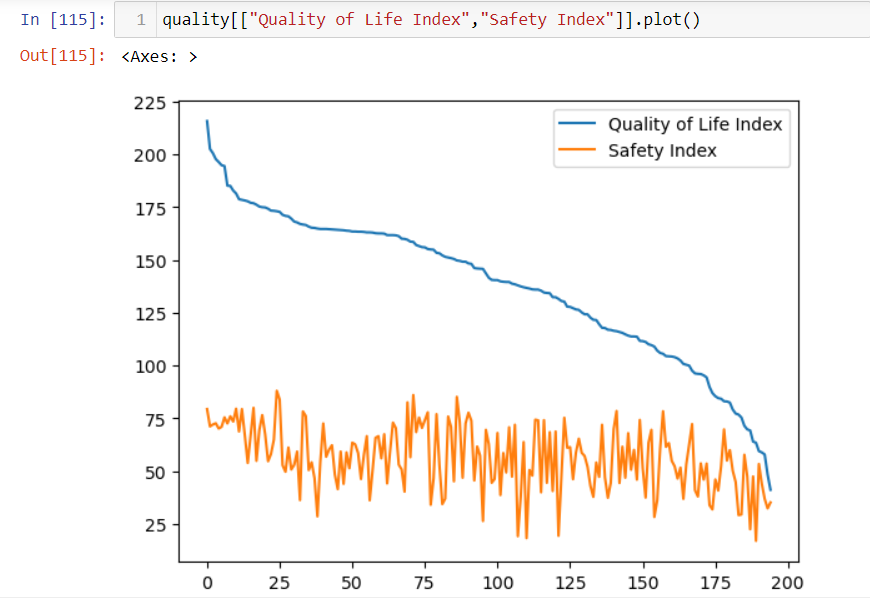


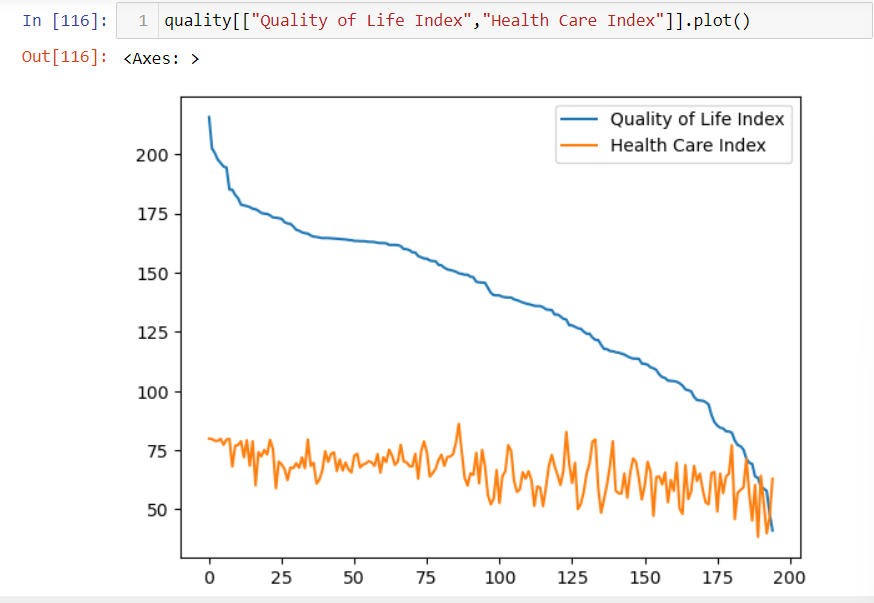


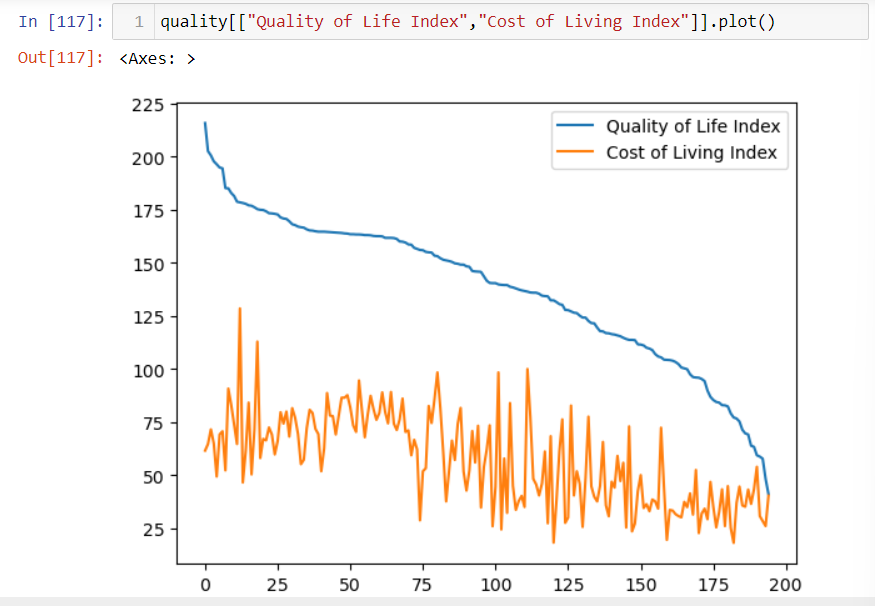
Here we are finding the values that are being repeated a specific number of time for each index table.

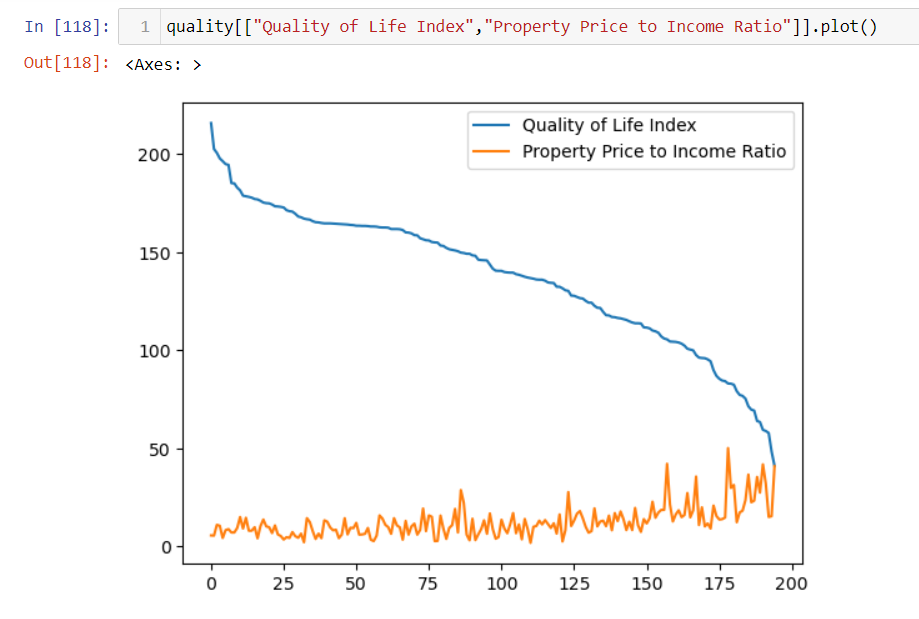
**Plotting two variables together and making a graph for comparison**



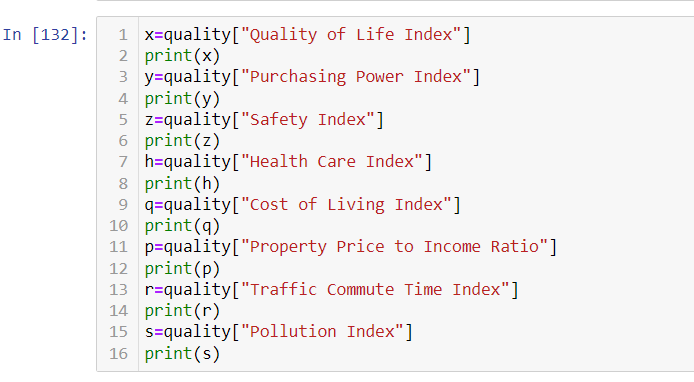


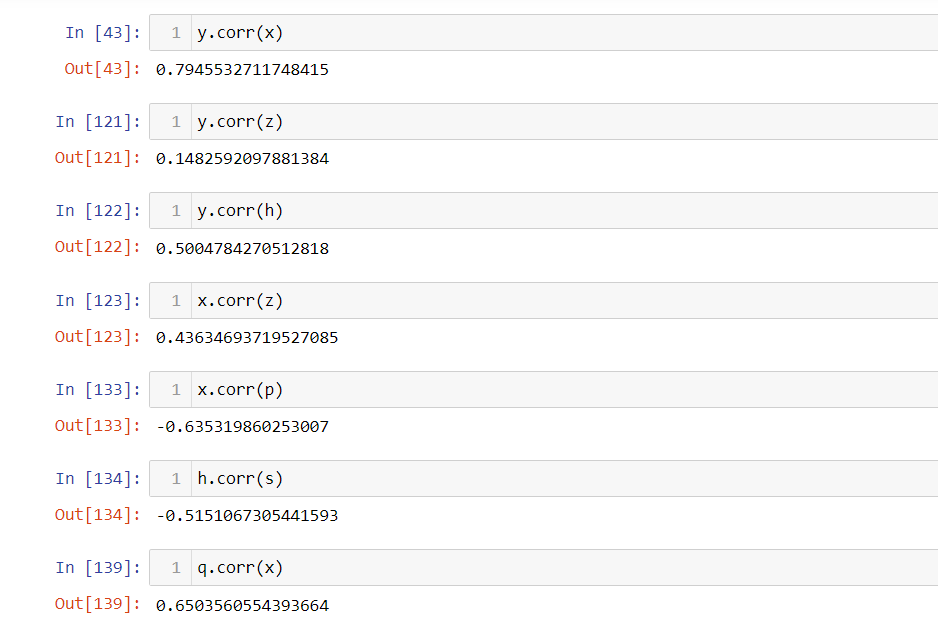






**Finding correlation between two variables which are index here**





FINDINGS:

1. Purchasing Power (PPI) vs. Quality of Life (QLI):

- The correlation coefficient between Purchasing Power Index (PPI) and Quality of Life Index (QLI) is approximately 0.795. This suggests a strong positive correlation, indicating that areas with higher purchasing power tend to have a better quality of life. Managers can consider this when targeting locations for business expansion or marketing.

2. Purchasing Power (PPI) vs. Safety (Safety Index):

- The correlation coefficient between PPI and Safety Index is approximately 0.148. This indicates a weak positive correlation, suggesting that higher purchasing power does not necessarily equate to higher safety levels. Managers should be cautious about assuming a direct link between purchasing power and safety.

3. Purchasing Power (PPI) vs. Health Care (Health Care Index):

- The correlation coefficient between PPI and Health Care Index is approximately 0.500. This suggests a moderate positive correlation, indicating that areas with higher purchasing power tend to have better healthcare services. This can be important for businesses concerned about employee health.

4. Quality of Life (QLI) vs. Safety (Safety Index):

- The correlation coefficient between QLI and Safety Index is approximately 0.436. This indicates a moderate positive correlation, implying that areas with a better quality of life often also have higher safety levels. Businesses considering employee well-being should take this into account.

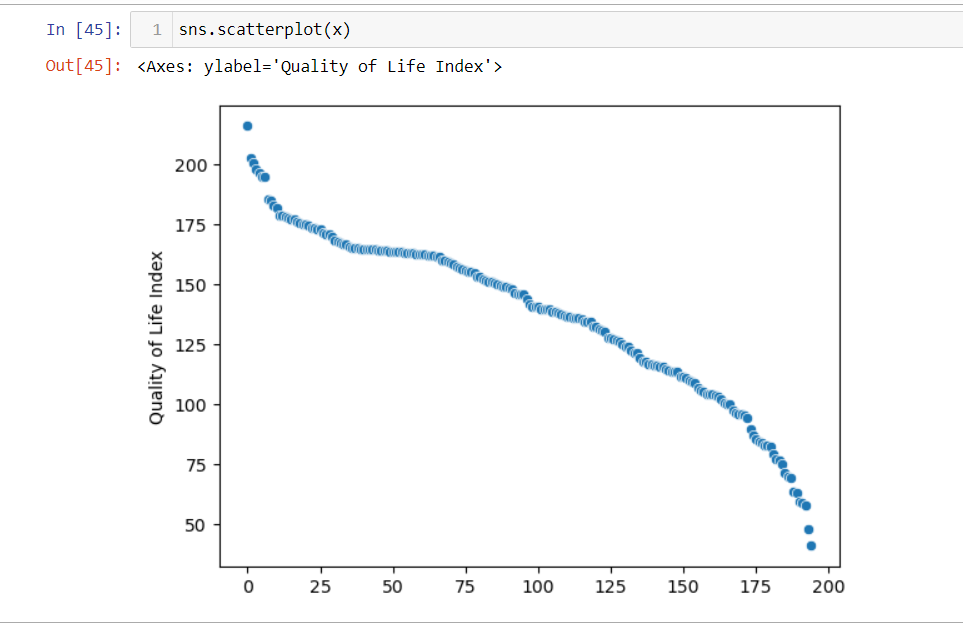
5. Quality of Life (QLI) vs. Cost of Living (Cost of Living Index):

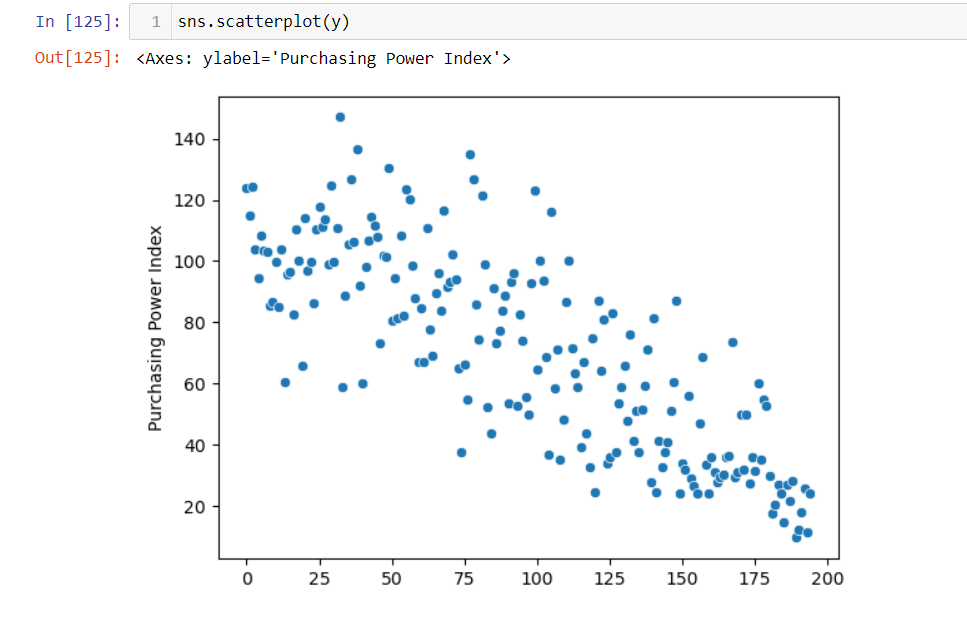
The correlation coefficient between PPI and QLI is approximately 0.650. This indicates a moderate positive correlation, suggesting that there is a meaningful relationship between the purchasing power of individuals in an area and the quality of life in that area.

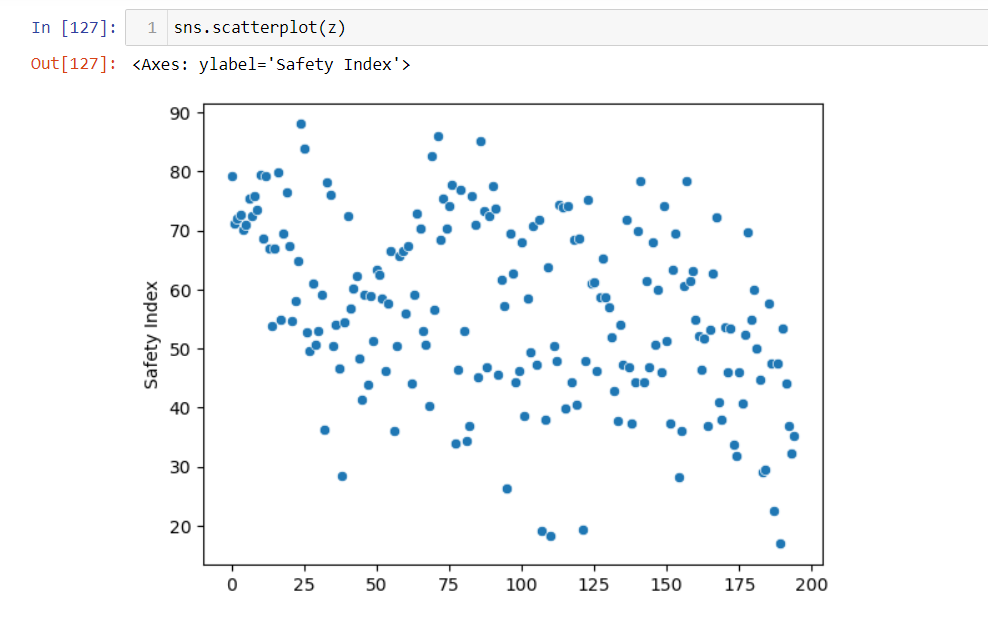
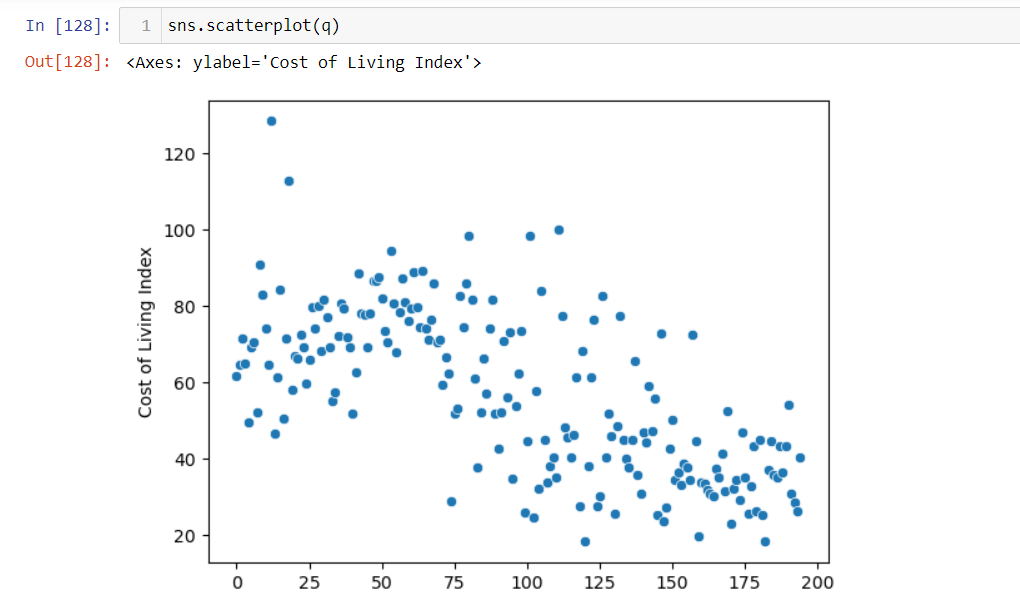
6. Health Care (Health Care Index) vs. Pollution (Pollution Index):

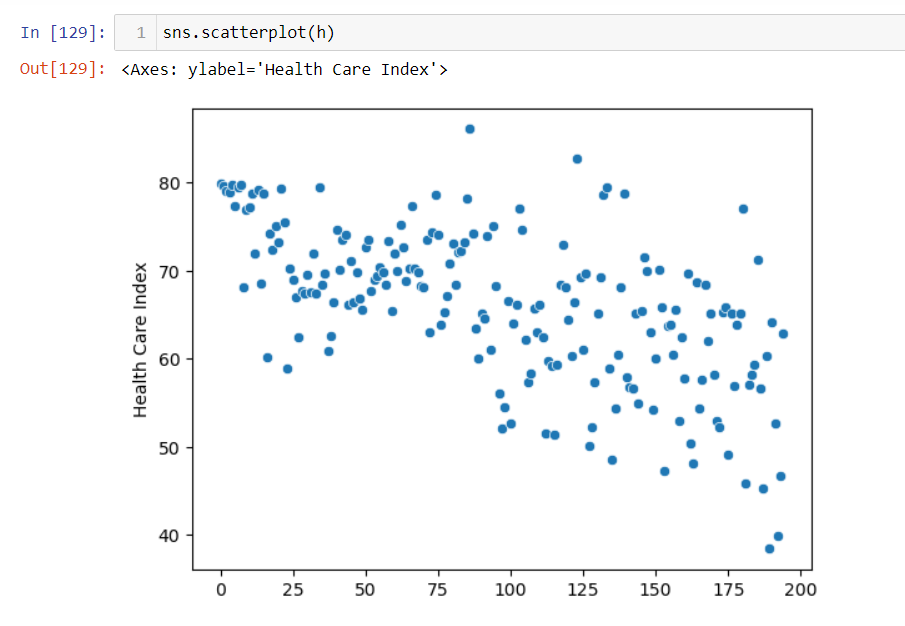
- The correlation coefficient between Health Care Index and Pollution Index is approximately -0.515. This suggests a moderate negative correlation, indicating that areas with better healthcare often have lower pollution levels. Businesses concerned about employee health and environmental impact should consider this relationship.

**Finding scatter plot for various index**









**MANAGERIAL INSIGHTS | IMPLICATIONS**

* The overall quality of life appears to be relatively high which suggests that the surveyed locations generally offer a good quality of life.
* The Purchasing Power Index here indicates that people in these locations have reasonable purchasing power. This can be relevant for businesses looking to target consumer markets.
* The Safety Index suggests that safety levels vary across the surveyed areas. Managers should consider safety factors when making location-based decisions, especially for employee and customer well-being.
* The Health Care Index here is indicating a moderately good quality of healthcare services. Businesses that rely on a healthy workforce or serve individuals with healthcare needs may need to evaluate this aspect further.
* The high standard deviation in the Quality of Life Index suggests that there are significant differences in the quality of life among the surveyed locations. Managers should consider these variations when making decisions related to employee placement, customer targeting, or business expansion.
* The Purchasing Power Index also exhibits considerable variability, indicating that income and spending capacity vary significantly across locations. Businesses should carefully assess these differences when planning pricing and marketing strategies.
* While there is some variation in the Safety Index, it is relatively lower compared to other indices. This may suggest that safety levels are more consistent across surveyed areas, which could be a positive factor for businesses concerned about security and safety.
* The low standard deviation in the Health Care Index implies that healthcare quality is relatively consistent across the locations surveyed. This consistency can be beneficial for businesses reliant on a healthy workforce.
* The high variance in the Quality of Life Index implies that there are significant differences in the quality of life among the surveyed locations. Managers should carefully consider these variations when making decisions related to employee placement, customer targeting, or business expansion.
* The Purchasing Power Index also exhibits considerable variance, indicating that income and spending capacity vary significantly across locations. Businesses should be mindful of these differences when planning pricing strategies and marketing campaigns.
* Although there is some variance in the Safety Index, it is relatively lower compared to other indices. This suggests that safety levels are relatively consistent across surveyed areas, which could be reassuring for businesses concerned about security and safety.
* The low variance in the Health Care Index implies that healthcare quality is relatively consistent across the locations surveyed. This consistency can be an advantage for businesses relying on a healthy workforce.
* Managers can consider correlation between purchasing power index and quality of life index when targeting locations for business expansion or marketing.
* The correlation coefficient between PPI and Safety Index suggests that Managers should be cautious about assuming a direct link between purchasing power and safety
* The correlation coefficient between PPI and Health Care Index can be important for businesses concerned about employee health.
* Businesses considering employee well-being should take correlation coefficient between QLI and Safety Index into account.
* Businesses concerned about employee health and environmental impact should consider Health Care Index and Pollution Index relationship.