World Economic Indicator Project : SatyamKumar

Storyline:

So, the very first step is to understand the business problem and observe the dataset.
So, i was given a dataset of GDP of different regions for different year and i needed to analyse and give the conclusion analysis accordingly

- No such search key was found which is common and unique in all sheets
- So, firstly, we proceeded with the creation of the Unique Key by concatenating region plus year. unique key created.

Now Merging all the sheets

- to make a single merged worksheet with the help of VLOOKUP function.
- Now all the data is in one place.

Now, Handling for missing values:

If the number of missing values is relatively small than total no of datasets then we can remove / ignore those values…. but here we cannot do so.

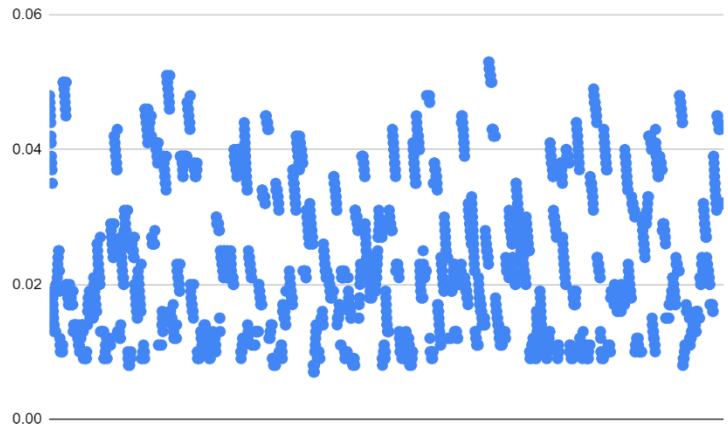
Data cleaning…. some data formatting (heading bold, colour)

Now, Checking for Outliers

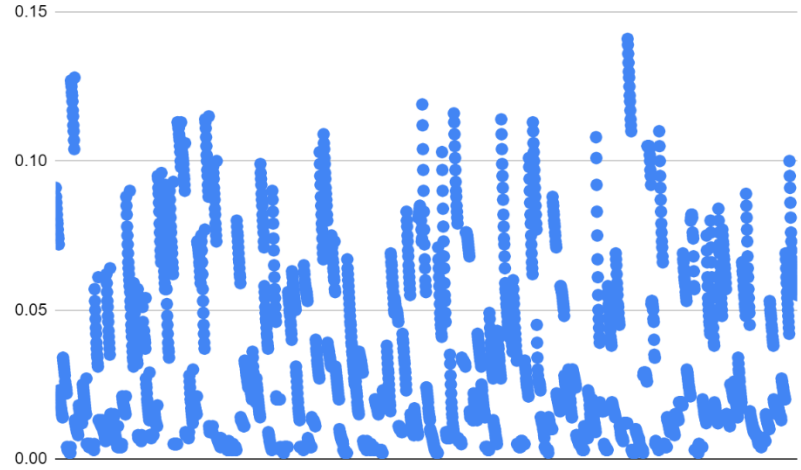
Checked if any outlier is present …through scatter chart

- No outliers in gdp
- No outliers in health expenditure

- This(below) is of birth rate (no outlier detected), because there is no such strange points which is too much out of range very far from other points.

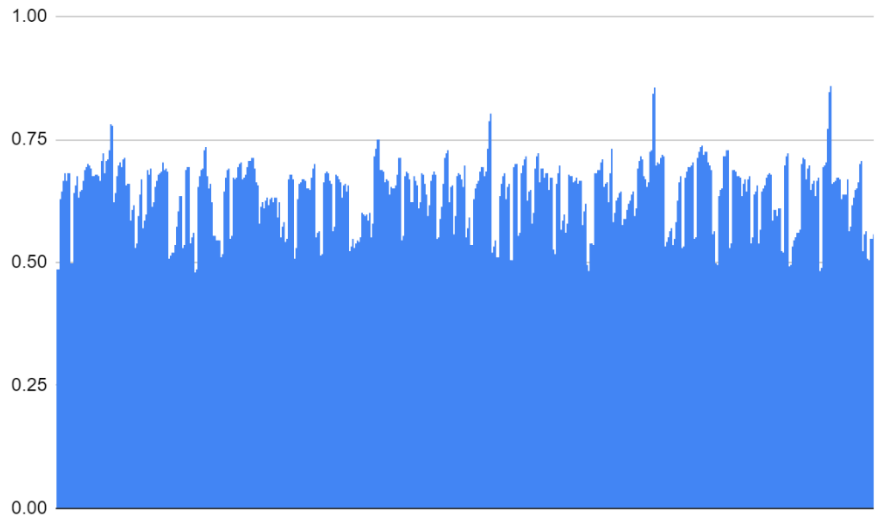


- For infant mortality(no outliers detected)



No outliers in most of the column (detected via charts)

Checking below for outliers in Population via Scatter Chart .



Now, time for Feature Engineering

- Made GDP/capita
- CO2 emission/capita
- Energy consumption/capita
- Tourism outbound/capita
- modified Internet usage to percentage format
- Created tourism outbound and inbound per capita

Now, performed Bivariate analysis of:

- GDP/capita with features like business tax rate
- GDP/capita with features like ease of doing business, urban population, internet usage and a lot of other features as well.

Now, performed Univariate analysis:

- Via average, mean, median, mode and histograms as well

Now the analysis was as follows:

- Correl between GDP of the region is weak negative with the Lending Interest Rate(-0.098)

- The average life is positively correlated (+0.053437) with the health expenditure (Higher the health expenditure, higher the average lifetime of the citizens)

Conclusions:
So, Finally the Insights are:

(Executive Summary)

- GDP/capita is directly proportional to Tourism Inbound/capita
- GDP is directly proportional to Business tax rate
- GDP is directly proportional to ease of doing business.
- More urban population represents more internet usage and thus higher GDP.
- Average Life Expectancy is directly proportional to Expenditure on health.

- The GDP of the Asia region was found to be highest, followed by Europe, then the US and so on.
- GDP of the region is positively correlated to Ease of business.
- GDP of the region is positively correlated to Ease of business
- The region with higher GDP per capita had more percentage of urban population

Also,

- Average Life Expectancy of Male is Highest in: Europe
- Average Life Expectancy of Female is Highest in: Europe
- Average life expectancy of male globally = 69.29266
- Average life expectancy of Female globally = 75.25
- Other such required mean, median, mode, trimmean, etc were also calculated during the work flow of the process. (Univariate analysis).