Life Expectancy Analysis

CRISP-DM

Understanding Business

Assessing Current Situation

Inventory of Resources

- Data Analyst , Python Programmer (Personnel)
- LE-File Description.docs, education.csv, crime.csv, income.xlsx, area.csv, lifeExpectancy.csv, region.txt (Data)
- MacOS Big Sur , 8GB RAM, Mac M1 chip (Computing Resources)
- Anaconda Distribution, Python 3, Numpy, Pandas, Matplotlib, Seaborn, Statsmodels, Terminal, Git, Jupyter Notebooks, Keynote (Software)

Assessing Current Situation

Schedule of Completion of Project (26 Days)

Tasks Schedule

Work	Start Date	End Date
CRISP-DM Start	25.9.2021	20.10.2021
Data Collection & Cleaning	27.9.2021	30.9.2021
Data Exploration	1.10.2021	6.10.2021
Unusual Observations	7.10.2021	9.10.2021
Regression Analysis	10.10.2021	15.10.2021
Results & Summary	16.10.2021	20.10.2021

Assessing Current Situation

Project Risks & Potential Solutions

- Over generalisation to other population outside America
- Lack of clarity
- Solution would be to restrict our understanding to limited area and Planning effectively

Business & Data Mining Goals

Primary Business Objective

Our Primary Objective would be to find factors related to life expectancy America (Business) by running Regression Analysis (Data Mining) on the data to finds factors related to life expectancy by finding out positively correlated factors (Measure of Success).

Business & Data Mining Goals

Secondary Business Objective

Our Secondary Objective would be to know more about distribution of certain demographic variables in US population (Business) by analysing data to answer a series of questions (Data Mining) and finding out and plotting results (Measure of Success).

Project Plan & Tools Required

Project Steps	Duration (in Days)	Start Date	End Date	Tools & Techniques
CRISP-DM (Ongoing)	26	25.9.2021	20.10.2021	Understanding Analysis Life
Data Collection & Cleaning	4	27.9.2021	30.9.2021	Numpy, Pandas, Data Preparation
Data Exploration	6	1.10.2021	6.10.2021	Data Visualization, Seaborn, Plotting
Unusual Observations	3	7.10.2021	9.10.2021	Outlier Detection
Regression Analysis	6	10.10.2021	15.10.2021	Regression, Statsmodels
Results & Summary	5	16.10.2021	20.10.2021	Plots, Summary, Report