Data Visualization Project

Name: Swastik Shrey & Preetam Bolla
Section: S11
Roll no's: 211EE148 & 211EE132 respectively

***Problem Statement* — This project aims to create an infographic/visual graph to find a correlation between *a family's household income* and their *inclination towards switching to renewable sources of energy*.**

***Python libraries used*[1]:**

1. ***NumPy* – mathematical operations**
2. ***Pandas* – extracting and organizing data**
3. ***Matplotlib* – plotting data**
4. ***Seaborn.* – building over matplotlib graphs**

DESIGN, PLAN OF THE PROJECT

- Making a **hypothesis** that the inclination of people towards **choosing sustainable sources of energy increases with increase in average household income**.
- Finding data for the following **sub-hypothesis** to increase granularity of the analysis:
    1. Continents like *North America* with higher avg household income than *Africa* has a greater adoption of sustainable energy.
    2. Similar pattern between a low earning region/country and a high earning region/country within a continent.
    3. Similar pattern between a low earning state and a high earning state within a country.
- **Extracting and organizing** the data using ***NumPy*** and ***Pandas*** libraries. Conducting **EDA (Exploratory Data Analysis)** on the procured data.
- Extracting **the main characteristics of the data** using EDA to prove our hypothesis using suitable co-relations.
- Choosing a **suitable data visualization graph/plot** for the data according to the variables.
- **Plotting** the data using ***Matplotlib*** and ***Seaborn*** libraries to visualize the data and conclude.

I. Current Status of Project

A. *Current Status of the Project*

1. **Hypothesis and sub-hypothesis** have been ideated.
2. Detailed **learning of the required libraries** has been done using the resources referenced at the end of the report.
3. **Data** necessary to prove the hypothesis has been **procured** in the form of APIs, CSVs and excel file formats from websites linked at the end of the report.
4. **EDA of the data** has been done using NumPy and Pandas and main characteristics of data have been stored.

The current and future progress of the project can be viewed on our GitHub repo:
S-Shrey-09/DataVizProject (github.com)

B. *How much of the design is converted into code?*

~50% of the data required for plotting has been procured, extracted and organized for plotting.

II. Expected Final Design

A. *Features*

The final output will be able establish a co-relation between a family's household income and their inclination towards switching to renewable sources of energy for household and transportation needs.

B. *Data Visualizers*

The following types of data visualizers will be used:
---

p.s. other types of data visualizers might be added by the end of the project for better understanding.

III. Uniqueness of our Project

The hypothesis chosen is a unique one and none of the tutorials or references used, provide roadmaps to extracting or plotting of the dataset that we have chosen.
The tutorials were simply used to learn the basics of libraries. The co-relation if proved would be a unique one.

References

[1] Documentations and tutorials used to learn libraries:
a) NumPy - *NumPy user guide — NumPy v1.22 Manual*
Complete Python NumPy Tutorial - YouTube
b) Pandas - *User Guide — pandas 1.4.1 documentation (pydata.org)*
Complete Python Pandas Data Science Tutorial -YouTube
c) Matplotlib - *Users guide — Matplotlib 3.5.1 documentation*
Python Plotting Tutorial w/ Matplotlib & Pandas - YT
d) Seaborn - User guide and tutorial — seaborn 0.11.2 (pydata.org)
Seaborn Tutorial 2021 - YouTube

[2] Books used:
1. Python Data Science Handbook – Jake VanderPlas
2. Data visualization with Python – Mario Dobler and Tim Großmann

[3] For choosing data visualizers:
From data to Viz | Find the graphic you need (data-to-viz.com)
44 Types of Graphs & Charts [& How to Choose the Best One]

[4] Data collected from:
UNdata
DataVizGNI/data at main · jiexiangfan/DataVizGNI (github.com)
Open Government Data (OGD) Platform India
public-apis/public-apis: A collective list of free APIs (github.com)
Wikipedia, the free encyclopedia

if any new source of data is used for the project later it will be updated here and uploaded on our GitHub.