Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Our primary goal is to understand the market of this particular region and try to use our data analysis techniques to draw out key features of the market. Our secondary goal is to draw out actionable insights from our analysis and give conclusions about key aspects of the market such as cancellation rate, distribution channels, and alike Our approach towards the problem statement was well decided by two of us. As our approach was clear and we moved in a sequence. We used 5-6 visualization plots to get useful insights.

As a first step we saw the dataset and divided the complete project into sub categories that is Cleaning the dataset, Handling the null values, Handling the Outliers, Univariates, Bivariate and the Hypothesis questions to get some useful predictions.

We started with Importing some of the useful python libraries, needed for the project. After the basic operations of importing the dataset, seeing at the overview of the dataset we started with cleaning the dataset. In cleaning the dataset, we mostly focused on the null values, duplicates and dropping unnecessary columns.

As we proceed, we looked into the outliers' part in the numeric type of columns, we handled the outliers with the help of bar plots. With the help of inbuilt seaborn and matplotlib libraries, replaced the outliers with some mean values. As we analyzed all year's data to know the high percent of data present in which years and used them for further analysis.
Then we extracted out all the column lists to get a univariant columns name for using these years further in all analysis, years from 1990 to 2014 have most percentage of data.
Then we analyzed all country detail to get into it and know every detail of countries, their columns name, Indicator names, and codes for further analysis and to understand the similarities and differences in them.

When we analysed many countries, we concluded that every country has an almost similar number of rows and has the same indicators in them

Then we took one indicator name and analysed it for different countries by taking the mean values to get the insight of it and analysed that although every country has a similar indicator name, the data value is so different even for the same years
For BAR.NOED. 1519. FE. ZS India has a mean value range between 20 to 60, Nepal has 5-90, whereas the Russian Federation has 0.2 to 1.4 only in the same years. On the other hand, for SE.PRM.TENR Nepal, India, and the Russian federation all have a high mean value range data.

Then we looked into bivariant and tried to compare one indicator with the top 7 countries that have high data availability for the years that have high percentages of data with the help of Python implementation of the grammar of graphics, we came up with some interesting conclusions,

we concluded that every indicator, a key detail of education has a different percentage of data in all countries with the same distribution.

After that, we framed out the different percentages of data for many countries by using the bar plot.
from the above hypothesis, we concluded percentage of present data from 2004 to 2015 for different countries India ranged between 75 to 85%, The United Stated ranged between 80 to 90%, whereas The Arab World ranges between 80 to 95%.

Then we analyzed the median value graph for the same indicator by comparing two countries together using a bar plot.
First, we took India and Zimbabwe and see India has a high value as compared to Zimbabwe, same we did with Bahrain and Belize and get to know that Bahrain has maximum counts, and at the end, we did with Belgium and Algeria and see they have same ranges as Belgium was Algeria was higher in starting then Belgium and at Belgium has high counts then Algeria.

In the end, we mentioned the indicators having PISA and both the genders in them and also concluded all 242 countries' data percent and analyzed the lower, middle, and higher percent data countries
and then give suggestions based upon insights to build a sound marketing strategy.
Country's that has fewer education data need to take loans from the world bank to improve their literacy rate and development and the countries that have less female education as compared to men's, bring some highlights so females get more education, in some countries children with lower age groups have less enrollment rate, make a rule of compulsory education for these age groups.

Contributor's Role:

1. **Kanika Singh**
    - Data Wrangling
    - Handling outliers by mean
    - Bivariant check
        a. Analysis of one Indicator for 7 countries
            Duration of compulsory education
            Adjusted net enrolment rate, primary, both sexes (%)
            Percentage of female population age 15-19 with no education

(India, Israel, United States, Russian Federation, Denmark, China, and World)

- Univariant Check
    a. Country of Data percentages
        India
        The United States
        Arab World
    b. Median Value analysis
        India and Zimbabwe
        Bahrain and Belize
        Belgium and Algeria
- Country's Name
- Country's percentages
- **Conclusion**

1. **Tanjul Gohar**
    - Data Wrangling
    - Handling outliers by mean
    - Bivariant check
        a. Analyzed all year's data
        b. Area-wise population analysis
    - Univariant check
        a. Each Country Analysis
            India
            China
            The United States
        b. Mean of Indicators
            BAR.NOED. 1519.FE. ZS for India, Nepal, and Russian Federation
            SE.PRM.TENR for same countries
    - **Conclusion**

Please paste the GitHub Repo link.

GitHub Link: - https://github.com/Kanika211/World-Bank-Global-Education_EDA_capstone1/tree/732ba9186523a71234e9781ed50c3d6c6a8ea4f0

Please paste the drive link to your deliverables folder, ensure that this folder consist of the project Colab notebook, project presentation and video

Drive Link -
https://drive.google.com/drive/folders/1Ut4wTAzeg_N3etdN7mV1blPamhy2_JAR?usp=sharing