

Lab: Guided Practice Project



Estimated Effort: 60 mins

Project Scenario

You have been employed as a data analyst by a Healthcare consultancy firm which has been conducting a survey on the state of global happiness annually. The World Happiness Report offers valuable insights into factors influencing happiness across countries. The firm wants you to produce a report to find out whether there are demographic, regional, and/or economic characteristics that lead to a better life.

The project tasks are data preparation, analysis, visualization, and dashboarding. Based on the data set, you must write prompts to generate the Python codes for performing specific tasks. You can access a JupyterLite-based testing environment to test the generated codes using the Generative AI classroom prompts.

The tasks assigned to you are as follows:

1. Check the correctness of the data types in the dataset.
2. There might be a few missing values in the dataset. Data cleaning will be a part of the assignment.
3. You have to perform exploratory data analysis to draw insights on the data:
 - o Identify the GDP per capita and Healthy Life Expectancy of the top 10 countries and represent it as a bar chart
 - o Find the correlation between the Economy (GDP per Capita), Family, Health (Life Expectancy), Freedom, Trust (Government Corruption), Generosity, and Happiness Score
 - o Create a scatter plot to identify the effect of GDP per Capita on Happiness Score in various Regions
 - o Create a pie chart to present Happiness Score by region
 - o Create a map to display GDP per capita of countries and include Healthy Life Expectancy to be shown as a tooltip
4. Create a dashboard with at least four of the above visualizations
5. Generate the narrative to present the dashboard

You decide to use Generative AI to create python codes that can help you analyze the data, determine the best features and create the visualization as per requirement.

Disclaimer: This is a fictitious scenario created for the purpose of this project. The dataset being used is publicly available.

About the data set

The World Happiness Report is a landmark survey of the state of global happiness. The reports review the state of happiness in the world today and show how the new science of happiness explains personal and national variations in happiness. This is a public dataset available on the [Kaggle](#) website as [World Happiness Report](#) under the [CC0: Public Domain](#) license.

For this guided practice project we will work on the year [2016](#) data, which has been slightly modified for the purpose of this guided practice project. You can download it from here: [2016](#)

Attributes of this dataset have been explained below.

Variable	Description
Country	Name of the country
Region	Region the country belongs to
Happiness Rank	Rank of the country based on the Happiness Score
Happiness Score	A metric measured in 2016 by asking the sampled people the question: "How would you rate your happiness?"
Lower Confidence Interval	Lower Confidence Interval of the Happiness Score
Upper Confidence Interval	Upper Confidence Interval of the Happiness Score
Economy (GDP per Capita)	The extent to which GDP contributes to the calculation of the Happiness Score

Variable	Description
Family	The extent to which Family contributes to the calculation of the Happiness Score
Health (Life Expectancy)	The extent to which Life expectancy contributes to the calculation of the Happiness Score
Freedom	The extent to which Freedom contributes to the calculation of the Happiness Score
Trust (Government Corruption)	The extent to which Trust contributes to the calculation of the Happiness Score
Generosity	The extent to which Generosity contributes to the calculation of the Happiness Score
Dystopia Residual	Dystopia is an imaginary country that has the world's least-happy people. The residuals, or unexplained components, differ for each country, reflecting the extent to which the six variables either over- or under-explain average 2014-2016 life evaluations. These residuals have an average value of approximately zero over the whole set of countries

Code execution environment

To test the prompt-generated code, open the Jupyter Notebook that you have been provided with in the succeeding lab in the course.

The data set for this lab is available in the following URL.

```
URL = "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMSkillsNetwork-AI0272EN-SkillsNetwork/labs/dataset/2016.csv"
```

Complete the setup in the Jupyter Notebook and then proceed further.

Important Note: All prompts that are made available have been hidden and the users are encouraged to first try to write their own prompts to create the solutions. Also, the prompts given as solutions have also been maintained as ones which will create generic code structures which you can modify according to the question at hand.

Importing the Dataset

You can begin by using the Generative AI model to create a python script that can load the dataset to a pandas dataframe. The dataset file already has the headers in the first row.

NOTE: Write the prompt to generate the said code and test it in the JupyterLite environment. For verification of appropriate loading, include a step for printing the first 5 values of the loaded dataframe.

- Click here for the prompt

Data Preparation

Data Cleaning - Checking for correct data types

You need to check the columns in the dataset for the right data type. If the data type is not appropriate, it may lead to misinterpretation of the data. Write a prompt that performs the following tasks:

1. List the data types of the columns and check if there is any column type that is unsuitable.

- Click here for the prompt

Data Cleaning - Change the data types

1. Change the data type to an appropriate type.

If you need specific types, you can specify so.

- Click here for the prompt

Please note as of the future version 3 of pandas it is recommended to use `df.method({col: value}, inplace=True)` instead of `df[col].method(value, inplace=True)`. You may see warning message in this regard, when you generate code and execute it in the notebook.

Data Cleaning - Checking for missing values

At this stage, you need to clean up the data. As has been shared earlier, the data may have missing values. Write a prompt that performs the following tasks:

1. Identify the columns with missing values and fill the blank cells with mean value of the columns.

Please ensure that you have changed all the column to the appropriate data type before you do this.

► Click here for the prompt

Database Querying (optional)

You may choose to query the dataset using [dbsensei](#). Use “Text to SQL” option for generating SQL queries.

1. Generate SQL query to Count the number of rows for each country
2. Generate SQL query to Calculate AVG Happiness score, Region wise for comparison

Please note the dbsensei application portal is still in beta. It may be down and unavailable sometimes.

Data Insights and Visualization

Write prompts that generate codes to perform the following actions:

1. Identify the GDP per capita and Healthy Life Expectancy of the top 10 countries.

► Click here for the prompt

2. Find the correlation between the Economy (GDP per Capita), Family, Health (Life Expectancy), Freedom, Trust (Government Corruption), Generosity and Happiness score. You may like to represent the correlation as a heatmap of a readable, visually appealing size.

► Click here for the prompt

3. Create a scatter plot to identify the effect of GDP per Capita on Happiness Score in various Regions. Use plotly for creating the plot.

► Click here for the prompt

4. Create a pie chart to present Happiness Score by Regions

► Click here for the prompt

5. Create a map to display GDP per capita of countries and include Healthy life expectancy to be shown as a tooltip

► Click here for the prompt

Dashboarding and Storytelling

Write prompts that generate codes to write at least four of the graph plots generated in the previous steps into a HTML page.

► Click here for the prompt

NOTE: Please ensure that the code has the plotly codes generated for creating various visualizations in the previous section.

2. Generate the narrative to present the dashboard

► Click here for the prompt

You can use this narrative to showcase your dashboard in a business meeting!

Conclusion

Congratulations! You have completed this guided project on using Generative AI for different data analytics tasks.

By the end of this project, you are now capable of using Generative AI for the tasks of:

- Data preparation
- Data analysis
- Dashboarding

- Storytelling

Author(s)

[Dr. Pooja](#)

© IBM Corporation. All rights reserved.