



You:

i will be submitting a data set and asking it to perform various data science tasks such as EDA etc etc .
what prompt can i give to you to make you do that?



ChatGPT:

When you're looking to have various data science tasks performed on a dataset, it's essential to be specific about what you want to achieve. Here's a general structure of how you can prompt me:

1. **Introduction:** Briefly describe the dataset and what you know about it.

2. **Tasks:**

- **Exploratory Data Analysis (EDA):** Indicate specific aspects you want to explore. For example:
 - Descriptive statistics (mean, median, standard deviation, etc.)
 - Distribution of specific variables
 - Correlation between variables
 - Missing value analysis
 - Outlier detection
- **Data Cleaning:**
 - Handling of missing values
 - Outlier treatment
 - Feature engineering (if required)
- **Data Visualization:**
 - Histograms, box plots, scatter plots for specific variables
 - Heatmaps for correlation or other analyses
- **Modeling:** If you want to build predictive models, specify:
 - Type of model (regression, classification, clustering, etc.)
 - Specific algorithms you're interested in (linear regression, decision tree, neural network, etc.)
 - Evaluation metrics of interest (MSE, accuracy, precision, recall, etc.)
- **Other Specific Tasks:** Any other data science tasks you're interested in.

3. **Output:**

- Specify how you want the results. For example:
 - Visualizations as plots
 - Tables of statistics
 - Recommendations or insights based on the analysis

Here's a sample prompt you can use:

Introduction: I have submitted a dataset about housing prices in a city. The dataset contains variables like number of bedrooms, area in square feet, location, and the sale price.

Tasks:

- **EDA:**
 - Descriptive statistics for all variables
 - Distribution of sale price
 - Correlation between area in square feet and sale price
 - Check for missing values
- **Data Cleaning:**
 - Handle any missing values found
- **Data Visualization:**
 - Histogram for sale price
 - Scatter plot between area in square feet and sale price
- **Modeling:**
 - Predict sale price using a linear regression model. Use area in square feet as the predictor.
 - Evaluate the model using Mean Squared Error (MSE)

Output:

- Visualizations for EDA and data visualization tasks
- Tables for descriptive statistics
- MSE for the linear regression model

Once you provide the dataset and the specific tasks you want, I can proceed with the analysis.



good! i am using Global YouTube Statistics.csv. Can you give prompt for this ? About this file A collection of YouTube giants, this dataset offers a perfect avenue to analyze and gain valuable insights from the luminaries of the platform. With comprehensive details on top creators' subscriber counts, video views, upload frequency, country of origin, earnings, and more, this treasure trove of information is a must-explore for aspiring content creators, data enthusiasts, and anyone intrigued by the ever-evolving online content landscape. Immerse yourself in the world of YouTube success and unlock a wealth of knowledge with this extraordinary dataset.



Certainly! Given the description of the "Global YouTube Statistics.csv" dataset, here's a structured prompt that you can use to guide the analysis:

Introduction: The dataset "Global YouTube Statistics.csv" provides comprehensive insights into top YouTube creators, detailing metrics such as subscriber counts, video views, upload frequency, country of origin, earnings, and more. This dataset is an invaluable resource for understanding the patterns of success on the YouTube platform.

Tasks:

- **Exploratory Data Analysis (EDA):**

- Descriptive statistics for key variables like subscriber counts, video views, earnings, etc.
- Distribution of top creators by country of origin.
- Correlation between subscriber counts and earnings.
- Analysis of upload frequency and its relation to video views.
- Check for missing values in essential columns.
- **Data Cleaning:**
 - Handle any missing values detected.
 - Identify and treat outliers if they exist in key metrics like subscribers, views, and earnings.
- **Data Visualization:**
 - Histogram or bar plot showcasing distribution of top creators across different countries.
 - Scatter plot between subscriber counts and earnings.
 - Box plot for earnings to visualize its distribution and outliers.
 - Heatmap showcasing correlation between key metrics.
- **Modeling:**
 - Predict earnings of a creator based on metrics like subscriber counts, video views, and upload frequency using a regression model.
 - Evaluate the model using metrics like Mean Squared Error (MSE) and R-squared.
- **Insights and Recommendations:**
 - Based on the EDA and modeling, provide insights into what drives earnings on YouTube.
 - Recommendations for aspiring content creators on how to maximize their success on the platform.

Output:

- Visualizations for EDA and data visualization tasks.
- Tables showcasing descriptive statistics.
- Model performance metrics.
- A summarized report on insights and recommendations based on the analysis.

Once you upload the dataset, I can proceed with the tasks as outlined in the prompt.



You:

give me the prompt.



ChatGPT:

Of course, here's a structured prompt for the analysis of the "Global YouTube Statistics.csv" dataset:

Introduction: The dataset, titled "Global YouTube Statistics.csv," captures comprehensive metrics related to the world's top YouTube creators. These metrics encompass subscriber counts, video views, upload frequency, country of origin, earnings, and several other attributes. By delving into this dataset, one can uncover patterns, trends, and correlations that highlight the dynamics of success on the YouTube platform, serving as an invaluable guide for content creators, data analysts, and digital content enthusiasts.

Tasks:**1. Exploratory Data Analysis (EDA):**

- Compute descriptive statistics for pivotal variables such as subscriber counts, video views, earnings, etc.
- Assess the distribution of top YouTube creators based on their country of origin.
- Investigate the correlation between metrics like subscriber counts and earnings.
- Examine the relationship between upload frequency and its potential influence on video views.
- Scrutinize the dataset for missing values, especially in vital columns.

2.



You:

my prompt should be a paragraph



ChatGPT:

Certainly! Here's a condensed paragraph-style prompt for the analysis of the "Global YouTube Statistics.csv" dataset:

We're delving into the "Global YouTube Statistics.csv" dataset, an extensive collection that captures the pulse of YouTube's top creators. Essential metrics like subscriber counts, video views, earnings, upload frequencies, and countries of origin are laid out for our analysis. Commence with an exploratory data analysis, focusing on descriptive statistics of pivotal variables and the distribution of creators by country. It's crucial to discern the relationship between subscriber counts, earnings, and upload frequencies. While at it, keep an eye out for missing values and ensure the data's cleanliness. As we progress, visualize the data to capture trends vividly, and consider a regression model to predict a creator's earnings based on influential metrics. The end goal is to distill insights and recommendations that can guide content creators to success on this dynamic platform.



You:

give a prompt that i can ask to do simple deep learning



ChatGPT:

Certainly! If you're looking to perform a simple deep learning analysis on the "Global YouTube Statistics.csv" dataset, here's a paragraph-style prompt:

With the "Global YouTube Statistics.csv" in hand, a treasure trove detailing YouTube's top creators, let's harness the power of deep learning to unearth hidden patterns. Start with data preprocessing: normalize the numerical variables and handle any missing values. Split the dataset into training and testing sets. Then, architect a basic neural network model to predict a creator's earnings based on features like subscriber counts, video views, and upload frequency. Use appropriate loss functions and metrics, considering this is a regression task. Train the model, validate its performance on the test set, and visualize the training progress. Conclude with insights drawn from the model's performance and potential improvements or further experiments that can be conducted.

This prompt provides a direction for building a simple deep learning model for the dataset. If you share the dataset, I can help with the actual implementation based on this guidance.



Generate a nice report - in medium.com. use savegpt extension to export your work to pdf, import and insert any images from chatgpt (screenshot extension) how to do this?



Creating a report on Medium.com and incorporating work from ChatGPT requires a few steps. Here's a step-by-step guide:

1. Work in ChatGPT:

- Perform the desired analysis, create visualizations, and generate insights within ChatGPT.

2. Use the SaveGPT Extension:

- If you have the SaveGPT extension (or a similar tool) installed, you can use it to save the content produced in ChatGPT to a PDF.

3. Use the Screenshot Extension:

- If you need to capture specific visualizations or any section of the ChatGPT interface, use the Screenshot extension or any other screenshot tool to capture the content.

4. Create a Medium Article:

1. Go to Medium.com and log in to your account.