**Step by Step Prompt List**

----- SETUP -----

1. **Navigate to Copilot Chat (Microsoft Edge Works Best)**:

Click: <https://m365.cloud.microsoft/chat?auth=2>

1. **Set up environment**:

Prompt: “You are an assistant agent. Work toward fully resolving the user's query but keeping responses clear and concise.

If you're unsure about file contents or code structure, use your tools to inspect them—never guess.

Plan before using tools and reflect on results to guide your next step. Use reasoning, not just function calls, to solve the problem.”

1. **Workshop Overview**

Prompt: “I want to make a machine learning algorithm to predict when someone will have a stroke. I have never built a machine learning algorithm before. Give me a list of steps from start to finish on how I would do this.”

1. **Initial Instruction**

Prompt: "I'm going to provide you with a dataset. Please hold off on processing it until I give you specific instructions to do so. Do not provide any code during this session."

1. **Upload the Dataset**

User Action: Upload the dataset file named healthcare\_dataset.csv.

------ Data ------

1. **Overview of Data** – can spend a little time on this

Prompt: “Can you give an overview of the dataset we’ll be using today?”

1. **Is this even a good idea**

Prompt: “Give me a list of reasons for why this dataset would be a good or bad fit for a machine learning algorithm compared to a simpler model.”

Insight: *You’re better at critical thinking than copilot (hopefully), so let it summarize arguments, not make decisions.*

1. **Describe Age Column (with example best prompts)**

Prompt (choose one):

|  |  |  |
| --- | --- | --- |
| **Good** | **Better** | **Best** |
| "Provide a quantitative summary of the age column from the dataset I gave you.” | "Please generate a statistical summary of the age column in the dataset, including count, mean, standard deviation, minimum, maximum, and quartiles (25th, 50th, 75th percentiles)." | "Please compute a detailed statistical summary of the age column from the uploaded dataset, including count, mean, standard deviation, minimum, maximum, and the 25th, 50th (median), and 75th percentiles. Present the results in a clear and structured format suitable for analysis." |

Insight: *If Copilot is not generating what you want (formatting, crashing, etc.) then providing more context will usually improve the output.*

1. **Create a Scatterplot**

Prompt: "Create a scatterplot that shows the relationship between age and BMI."

Prompt (optional): “What is the correlation between these variables?”

-----Model -----

1. **Choose a Model**

Prompt: “Provide a list of simple models to try and a list of complex models to try on this data set. Do not try to fit any right now.”

1. **Train a Logistic Regression Model**

Prompt: “Train a logistic regression model to predict stroke using all columns except ID."

Insight: *“People will have different model outputs. When making a real model with code you can lock settings in orde to generate consistent outputs.”*

1. **Train a Random Forest Classifier**

Prompt: "Train a Random Forest classifier to predict stroke using all columns except ID.”

1. **Evaluate the Classifier**

Prompt: "Explain the results of the model in more detail”

1. **Show Feature Importances**

Prompt: "Show a bar chart of the model's feature importances using the random forest model and explain which features are most associated with predicting stroke."

1. **Predict Stroke Risk for a Specific Individual** (If time permits)

Prompt: "Use the trained model to predict if someone who is 80, with a BMI of 30. Verify the order and completeness of the features used during training and prediction phases. Do not use the ID column."

**Links to Copilot and AI course Resources:**

1. BD Prompt Engineering resources

[Gen AI at BD Walkaround](https://bd1.sharepoint.com/sites/bdx/it/Intranet%20Documents/Generative%20AI/GenAI-Walkaround-Deck.pdf)

[Crafting Effective Prompts](https://learn.microsoft.com/en-us/training/paths/craft-effective-prompts-copilot-microsoft-365/)

[Generative AI Learning Companion](https://bd1.sharepoint.com/sites/bdx/it/Intranet%20Documents/Generative%20AI/GenAI%20BD%20Learning%20Companion.pdf)

1. AI Course (no-code)

[MIT No-Code AI Course](https://professionalonline2.mit.edu/no-code-artificial-intelligence-machine-learning-online-program?&utm_source=google&utm_medium=performancemax&utm_campaign=NCAI_US_PerformanceMax_Data_Layering_HighDevice&adgroup_id=&campaign_id=21272889974&Keyword=&placement=&gad_source=1&gad_campaignid=21266461842&gclid=CjwKCAjwruXBBhArEiwACBRtHeSlMAkThVmdb3lrUOZP_JnN0g0hlaRg7GZNEomSMyCHBD2QZ0WGbxoCZXgQAvD_BwE)

1. Short courses on machine learning basics

<https://www.deeplearning.ai/>

**Exploration & Deeper Learning Prompt List** that expands on each section.

|  |  |  |
| --- | --- | --- |
| **Section** | **Exploration Prompt** | **Purpose** |
| **Setup** | “What are the ethical considerations when building a health-related ML model?” | Introduce responsible AI principles. |
| **Overview** | “Compare classification vs regression. Which is more appropriate for stroke prediction and why?” | Reinforce problem framing. |
| **Data Quality** | “Check for missing values and explain how you would handle them.” | Teach data cleaning basics. |
| **Data Preparation** | “Normalize or standardize the age and BMI columns. Explain why this is important.” | Teach data preprocessing techniques. |
| **EDA** | “Visualize the distribution of BMI. Are there any outliers?” | Practice univariate analysis. |
| **EDA** | “Create a heatmap of correlations between all numeric features.” | Explore multivariate relationships. |
| **Feature Engineering** | “Suggest new features that could improve model performance.” | Encourage creativity and domain thinking. |
| **Train/Test Split** | “Determine an appropriate train/test split and explain your reasoning.” | Teach model validation techniques. |
| **Model Selection** | “Compare logistic regression and decision trees in terms of interpretability and performance.” | Deepen understanding of model trade-offs. |
| **Model Evaluation** | “Explain precision, recall, and F1-score in the context of stroke prediction.” | Clarify evaluation metrics. |
| **Model Evaluation** | “Plot the ROC curve and explain what it tells you.” | Introduce threshold-based evaluation. |
| **Bias & Fairness** | “Check if the model performs differently across gender or age groups.” | Introduce fairness auditing. |
| **Hyperparameter Tuning** | “How would you tune the Random Forest model to improve accuracy?” | Introduce model optimization. |
| **Explainability** | “Use SHAP or another method to explain individual predictions.” | Teach model interpretability. |
| **Deployment** | “What steps would be needed to deploy this model in a real healthcare setting?” | Connect ML to real-world applications. |
| **Reflection** | “What are the limitations of this model and how could they be addressed?” | Encourage critical reflection. |