

## Questions to ask of every numeric feature (based on Age feature)

Category	Question
<b>Basic Data Profile</b>	What is the data type of the variable?
	How many entries does the variable have?
	Are there any missing values? How many and what percent?
	What is the range (min and max) of the variable?
	What is the mean (average) value?
	What is the median value?
	What is the standard deviation?
	What is the MAD (Median Absolute Deviation)?
	What is the coefficient of variation (CV)? What does it say about variability relative to the mean?
	What is the skewness of the distribution? Is the data symmetric?
<b>Distribution Insights</b>	What is the kurtosis? Are the tails heavier or lighter than normal?
	Is the variable's distribution skewed? If so, in which direction?
	Where does the bulk of the data lie (e.g., most frequent values, mode)?
	Are there long tails, gaps, or notable cutoffs in the distribution?
	Are there extreme values or potential outliers?
	Are the values ordered or monotonic in any way?
	What does the interquartile range (IQR) tell you?
	How evenly or unevenly is the variable distributed across its range?
	Are there key changes in density or frequency in specific ranges?
	Are there key changes in density or frequency in specific ranges?
<b>Business Insights</b>	What groups dominate the dataset?
	What groups are underrepresented, and does this affect targeting?
	Which segment(s) of the population are most relevant to the business goal?
	Are there opportunities for tailored outreach based on this variable?
	Could the variable impact campaign strategy or customer engagement?
<b>Risk and Considerations</b>	Do certain value ranges align with life stages or financial goals?
	Are there limitations or bias in the data?
	Are there segments that may behave differently or require caution?
	Could over-/under-representation skew the analysis or model?
	Are there ethical, legal, or regulatory considerations for this variable?
<b>Recommendations and Next Steps</b>	Should the variable be binned into groups for analysis?
	Should different personas or messaging be created based on this variable?
	Should this variable be combined with another for deeper insights?
	How could this variable be visualized for stakeholders?
	Is this variable likely to be a strong predictor of the target outcome?

## Questions to ask of every categorical feature (based on Job feature)

Category	Question
<b>Basic Data Profile</b>	What is the data type of the variable?
	How many total entries does the variable have?
	Are there any missing values? How many and what percent?
	How many distinct categories (classes) are there?
	What are the most frequent categories, and what percentage of the data do they represent?
	What are the least represented or rare categories?

Category	Question
<b>Distribution Insights</b>	Is the distribution of categories balanced or imbalanced?
	Is the class distribution statistically significant (e.g., via chi-squared test)?
	What is the average or typical length of category names (string length)?
	Are any categories labeled as “unknown” or similar placeholders?
	Is the variable heavily skewed toward certain categories?
	Are some categories disproportionately large?
	Are rare categories meaningful or noisy?
	Are string labels consistent in structure and format?
<b>Business Insights</b>	Are there opportunities to combine similar or redundant categories?
	Do long-tail categories require grouping or special handling?
	What do the categories represent in the context of the business problem?
	Which categories are most likely to align with the business goal (e.g., term deposit subscriptions)?
	Do specific categories imply certain financial behavior or risk tolerance?
	Could job type influence how a client responds to marketing?
	Which customer segments might be most valuable to target?
	Are there underserved or overlooked segments worth focusing on?
<b>Risk and Considerations</b>	Could category imbalance lead to model bias?
	Could the "unknown" category mask important patterns?
	Is there risk of misinterpretation or stereotyping based on this variable?
	Are there ethical or compliance concerns with using this variable for targeting?
	Should certain categories be grouped together for analysis or modeling?
<b>Recommendations</b>	Should this variable be combined with others (e.g., age or education) to derive richer personas?
	Would it help to calculate conversion/response rates by category?
	Should campaign messaging vary by category?
	How should missing or "unknown" values be handled in modeling?
	Is this feature likely to be useful for segmentation, prediction, or personalization?