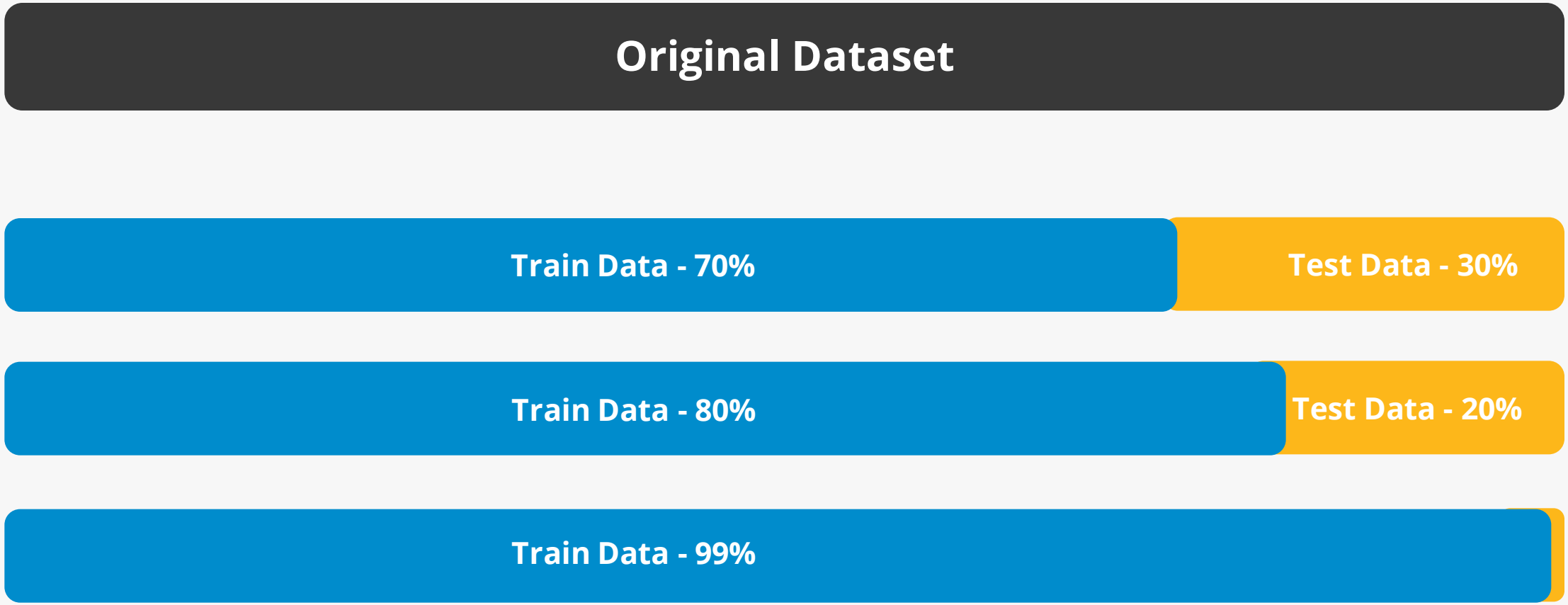
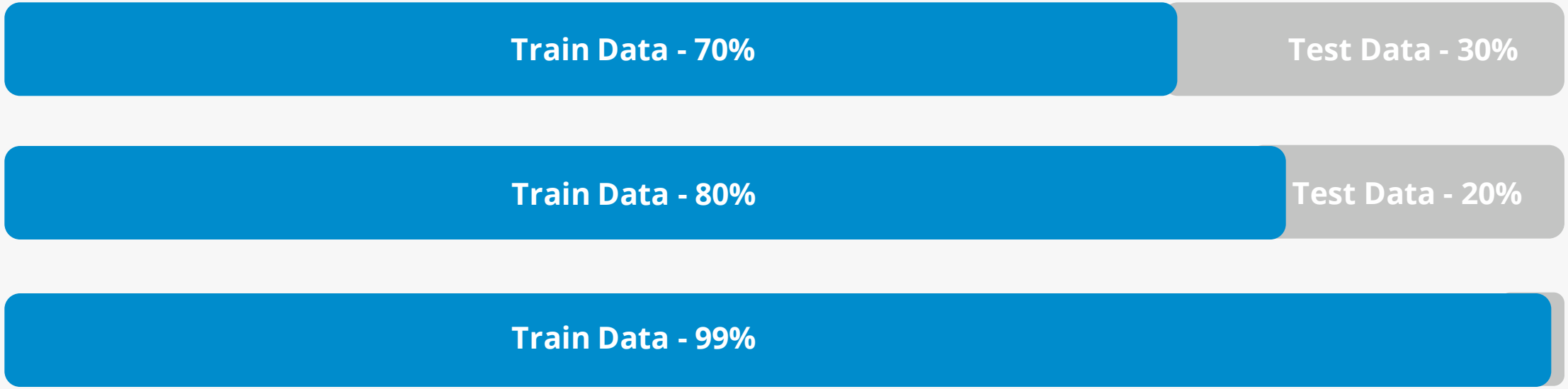
Train-Test Ratio

Train-Test Ratio

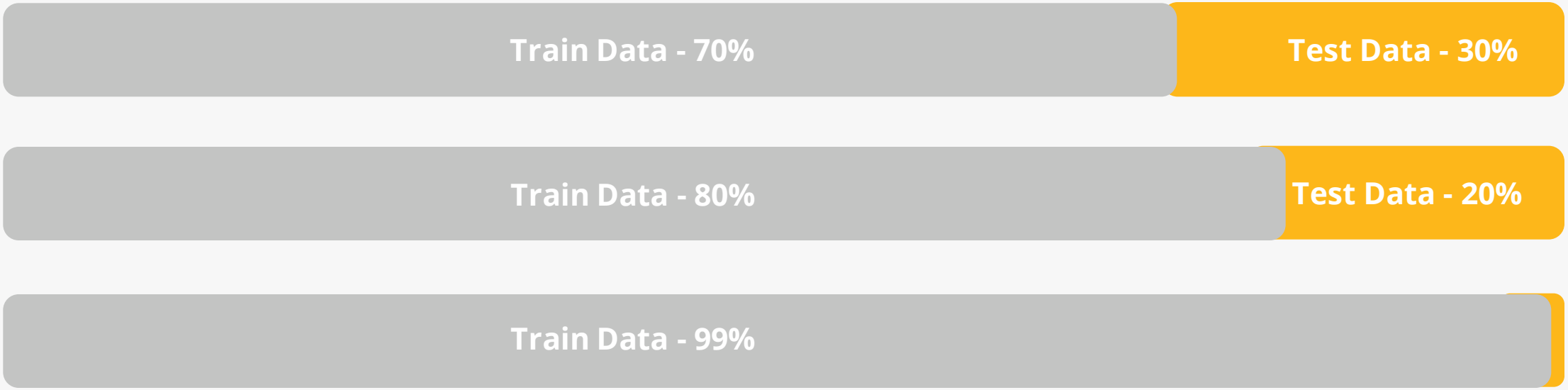


Train-Test Ratio



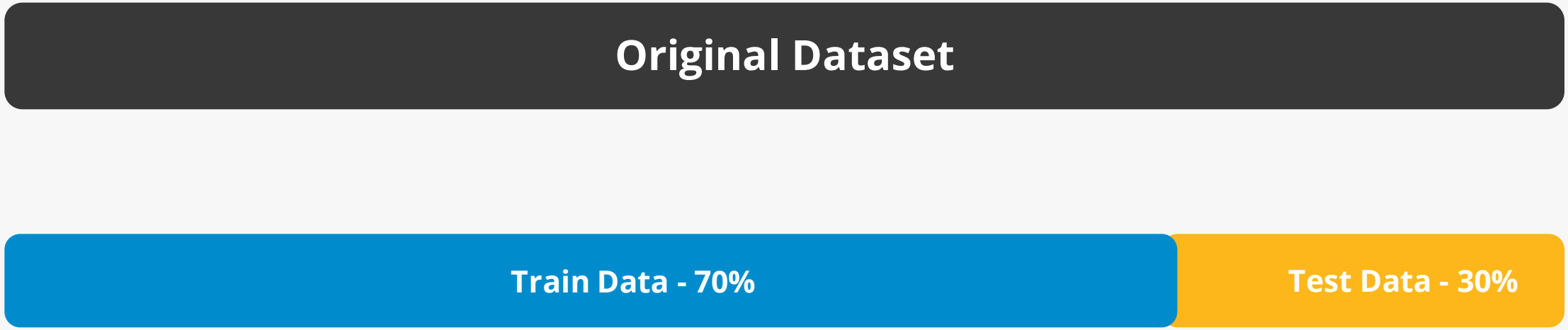
Train data is kept higher to ensure sufficient data is available to make insights and build the prediction system.

Train-Test Ratio

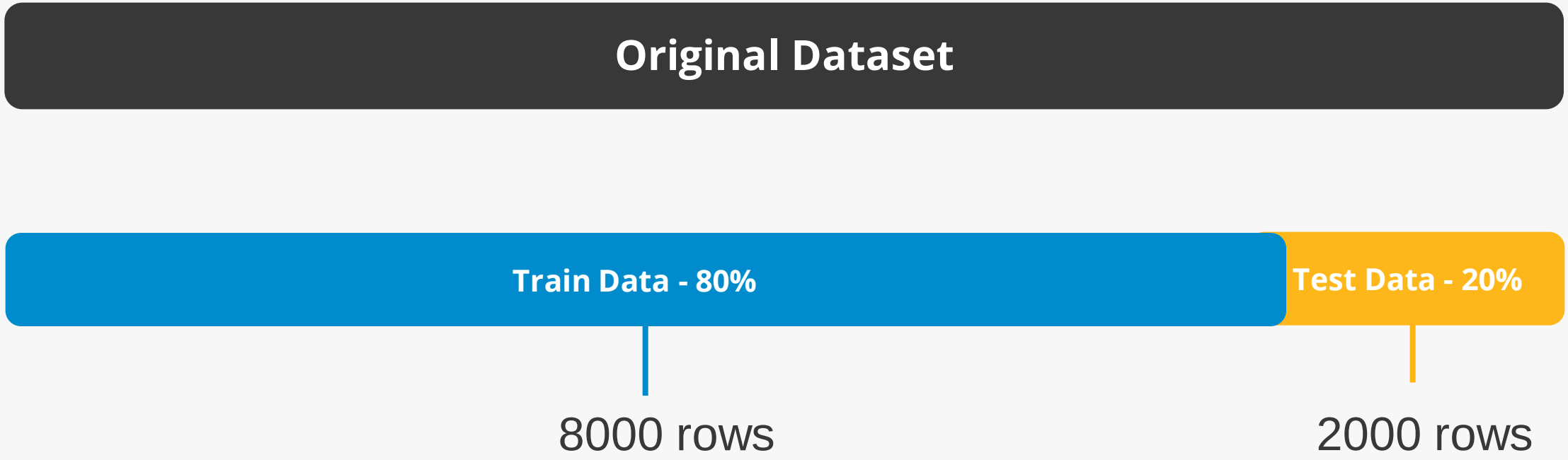


Test data should have sufficient samples to reduce the impact of sample bias.

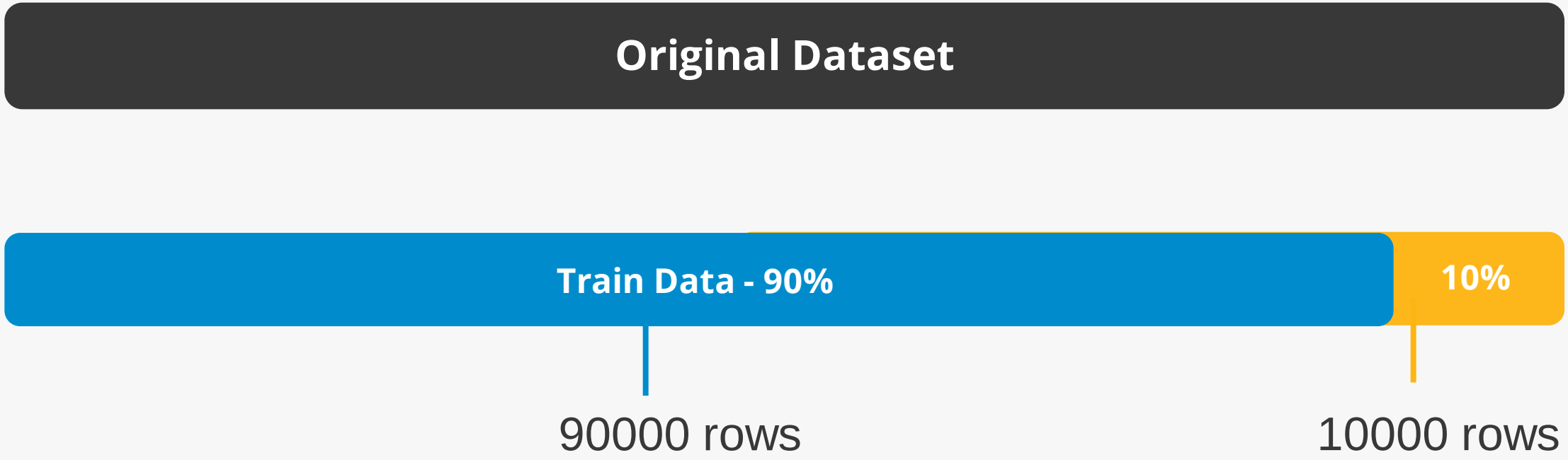
Train-Test Ratio: 500 rows



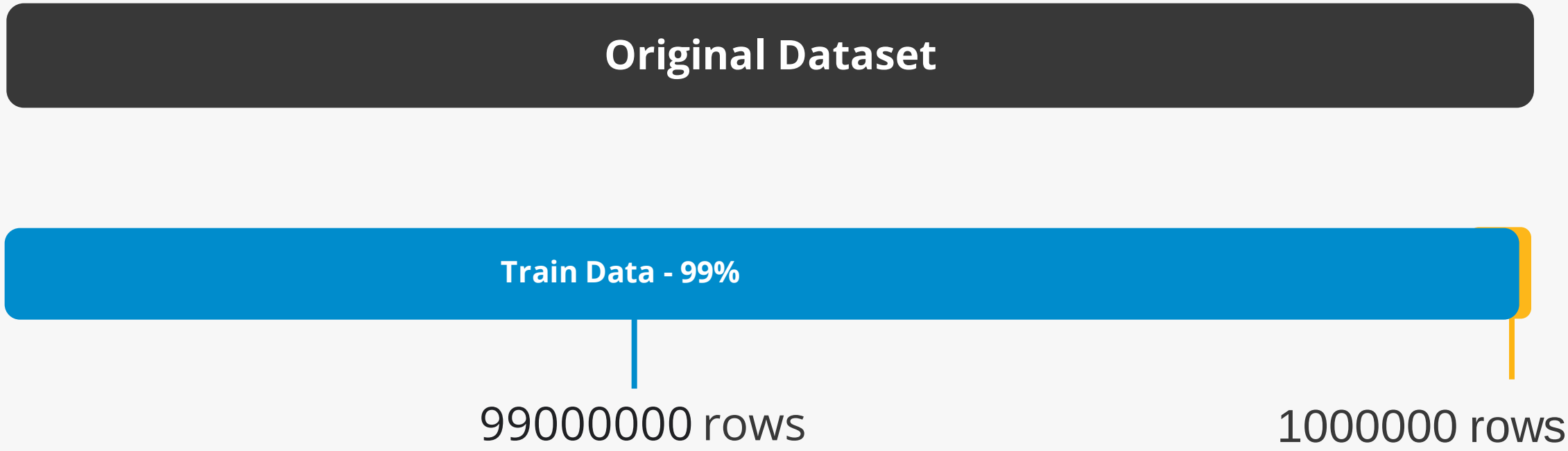
Train-Test Ratio: 10,000 rows



Train-Test Ratio: 100,000 rows



Train-Test Ratio: 100 million rows



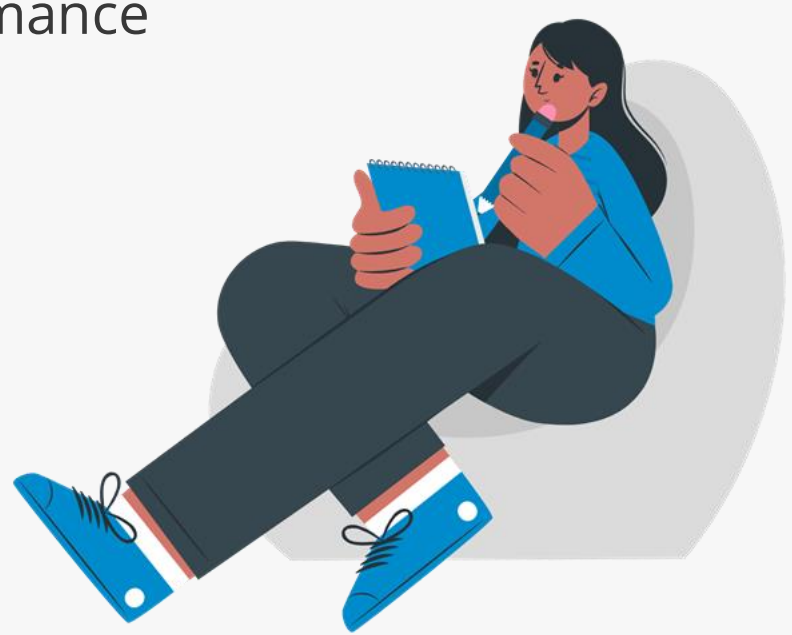
Limitations of Train-Test Split

Limitations of Train-Test Split

- **Difference in Evaluation**

1. Highly dependent on specific random split of data

2. Different splits lead to different evaluation performance



Limitations of Train-Test Split

- **Less Efficient Data Utilization**

1. A portion of data is reserved for testing

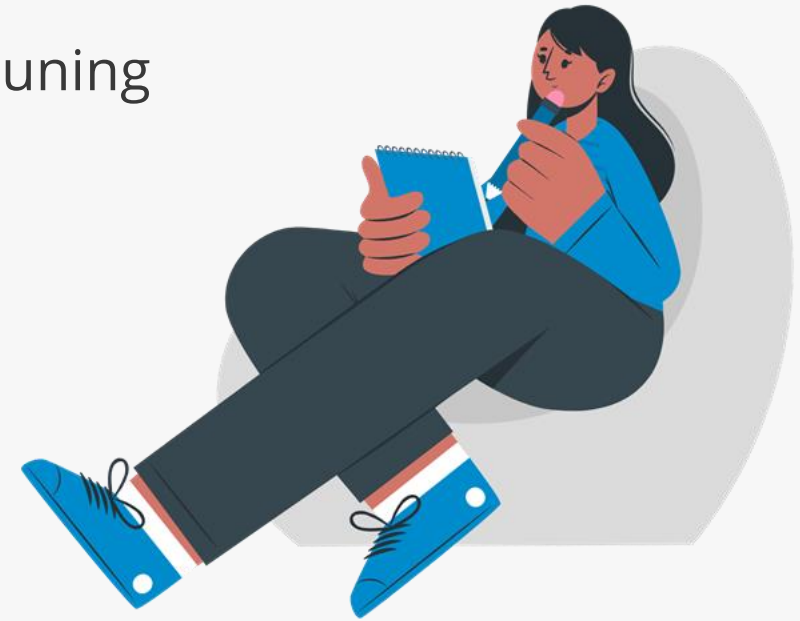
2. Amount of data available for training may not be sufficient

Limitations of Train-Test Split

- **Limited Information for Model Selection**

1. Train-Test Split provides only one evaluation of mode's performance

2. Might not be reliable especially when it comes to tuning hyperparameters



Cross Validation