

# Exploration - Now I Know Game Results

## Now I Know

We exploration to discover features that drive the target

You have to explore bivariate and multivariate after splitting

Only only explore on train

Heatmaps only for continuous variables

start by forming your initial questions + hypotheses

1. Ask questions in English about your data (what's driving y)
2. Analyze variables to answer that question
3. Write down your takeaways in plain English (no buzzwords)

hypothesis testing when visualization isn't clear

encode the data before splitting

Recommend exploring un-encoded categorical values

Recommend exploring un-scaled data

- because exploration viz + tests is for people
- People read "red, green, blue" a lot easier than the encoded version
- People understand "degrees and feet" not scaled versions of those
- Approaches:

- for each column in `payment_type`, make an encoded `payment_type` column on that same data