

A decorative graphic on the left side of the slide. It consists of a blue parallelogram and a light green parallelogram, both tilted at an angle. The blue shape is in the foreground, and the green shape is partially behind it. They are set against a dark blue background with faint, lighter blue diagonal stripes.

# Project 3: Apriori Algorithm



# Background Info

What is apriori?

What is at the core of the apriori algorithm?

What are the 2 basic operations of apriori?

What is the apriori principle?

How can you overcome overfitting?

What are the benefits and challenges of apriori?



# Dataset

- You can find a reliable dataset on github, kaggle, or using this link  
<http://archive.ics.uci.edu/ml/datasets.php>

Pick something you are interested in and one that would make a professional project!



# Problem Statement

Write a “problem statement” and an introductory paragraph that clearly explains your goals, it should include at least the following info:

- Describe your dataset, why you picked it, and write a small paragraph discussing your goal with your dataset, what models you can use to analyze it, and why.



# Import Data

Import the dataset and print the first few rows, and use `info()` and `describe()` to get a better sense of the data.

What features does the dataset contain? Which are categorical vs continuous? Explain.

Any missing values, null values, how many unique values, etc... describe your dataset in detail!



# Data Exploration & Visualization

Create at least 5 plots that help understand your data more. Explain each plot and what conclusions you can draw from each in detail.

Examples: countplot, distribution plots, heatmap, etc..

Explore your data and report your conclusions.



# Scaling

1. Is it necessary to scale the data? What benefits would it provide?
2. Which scaler will you use for this data set? Min Max, Standard, Robust, etc.
3. Are the features or the response variables scaled?

\*Don't forget to split your data into test-train splits before scaling!



# Data Preprocessing

Now that you've really studied your data, are you going to take any preprocessing measures?  
For which columns, and why?

Define any measures you've taken and address why you chose to do so.





# Model

By now, you should have completed any necessary scaling, encoding, preprocessing measures. Next steps would be creating and training your model. Split your data into training and testing sets.

Explain what 2 models you've chosen and why? Explain and define each model, giving background info, its uses, why it's beneficial, why you chose it over another model, and compare both models you've chosen. Is your approach parametric or non parametric? Which features were most important?

How did both models perform? Make sure you give all relevant info. Show the confusion matrix and classification report. Write a conclusion wrapping up your findings.