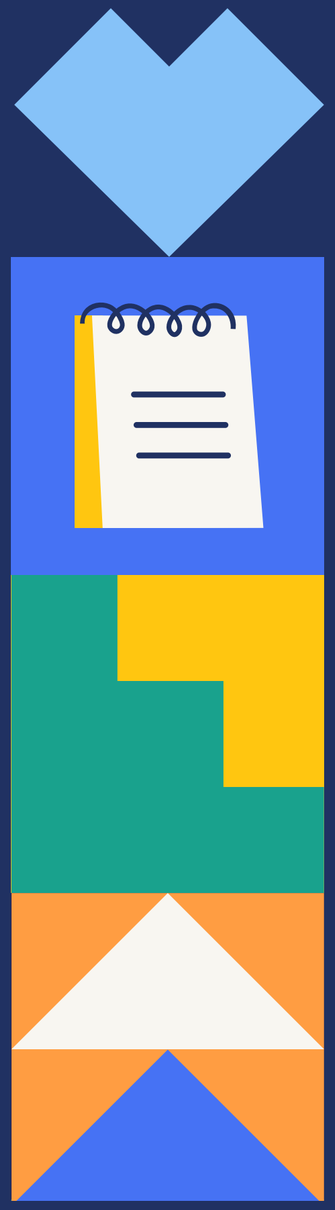
Introduction To Data Science

BY: DAREL OBALLA



Lesson Outline

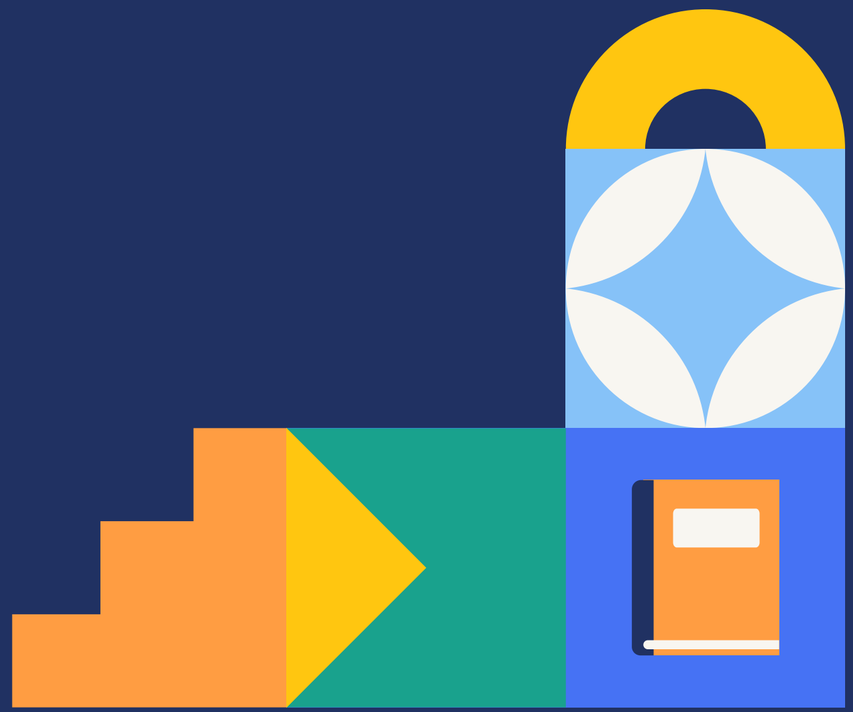
- What is Data Science?
- Why is Data Science Important?
- The Data Science Process
- Key Components of Data Science
- Data Science vs Other Fields
- Tools Used in Data Science
- Data Science Applications
- Career Paths in Data Science
- Skills Needed to Become a Data Scientist

What is Data Science?

Data Science is an interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data.

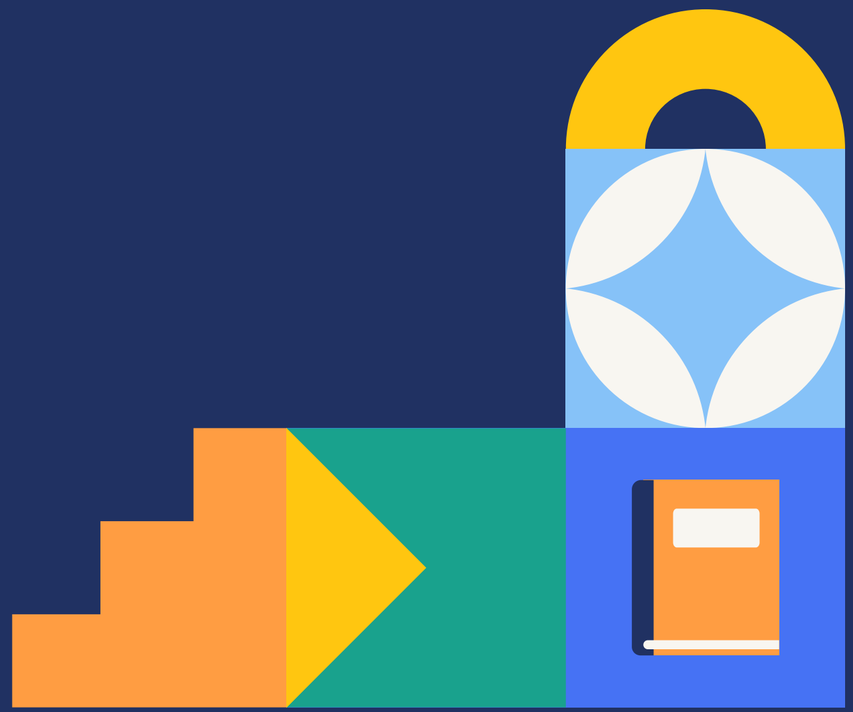
Data is a collection of information. One purpose of Data Science is to structure data, making it interpretable and easy to work with

Goal: Turn raw data into meaningful information for decision-making.



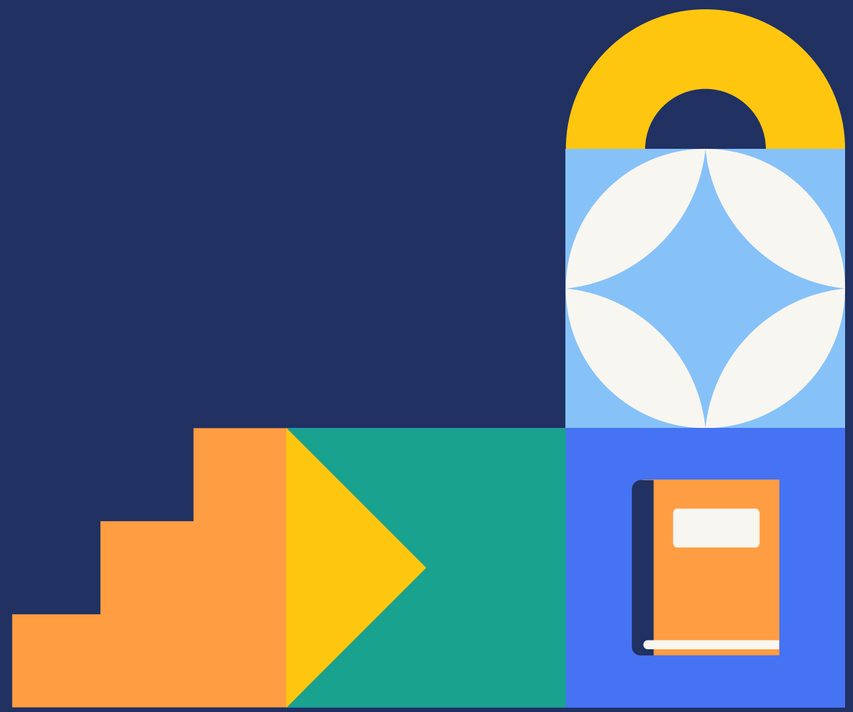
Why is Data Science Important?

- Data Explosion: 90% of the world's data has been created in the last two years.
- Business Decision-Making: Data-driven companies are 23 times more likely to acquire customers.
- Applications: Healthcare, Finance, Retail, Robotics, AI, and more.



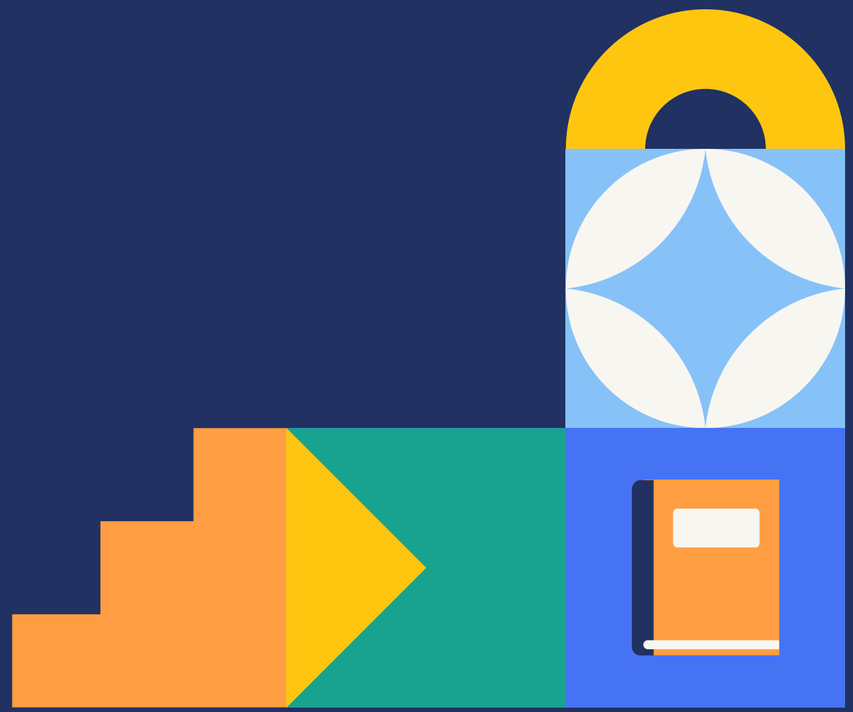
The Data Science Process

- Data Collection: Gathering relevant data from different sources (databases, APIs, web scraping, etc.).
- Data Cleaning: Removing errors, duplicates, and irrelevant data.
- Data Exploration: Visualizing and understanding the patterns.
- Modeling: Applying machine learning or statistical models to find insights.
- Interpretation: Converting insights into actionable business strategies.



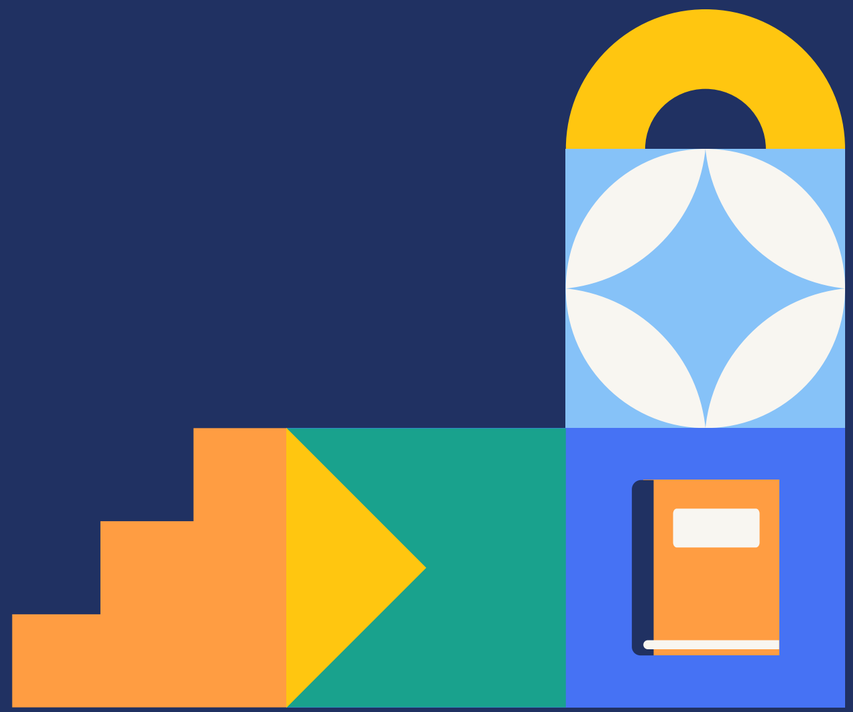
Key Components of Data Science

- Data: The foundation of data science (structured, semi-structured, unstructured).
- Statistics & Probability: Core to understanding and modeling data.
- Programming: Python, R, SQL for data manipulation and analysis.
- Machine Learning: Algorithms that allow computers to learn from and make predictions on data.
- Data Visualization: Communicating findings through graphs, charts, dashboards (using tools like Matplotlib, Tableau).



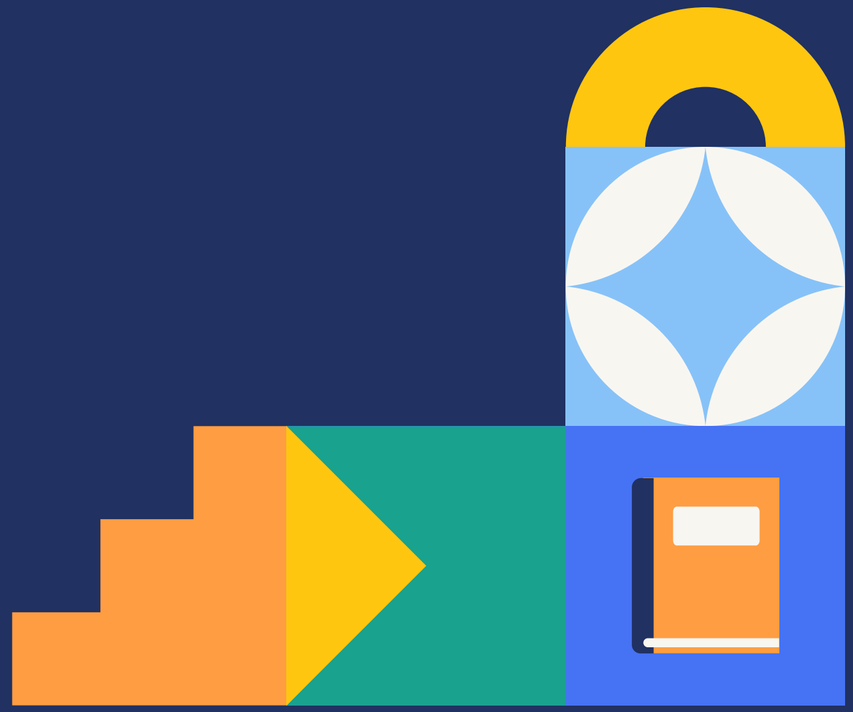
Data Science & Other Fields

- Data Science vs Business Intelligence (BI): BI focuses on historical data, while Data Science predicts future trends.
- Data Science vs Data Engineering: Engineers build data pipelines, while Data Scientists analyze the data.
- Data Science vs Machine Learning: Machine learning is a subset of Data Science focused on creating models that can predict outcomes.



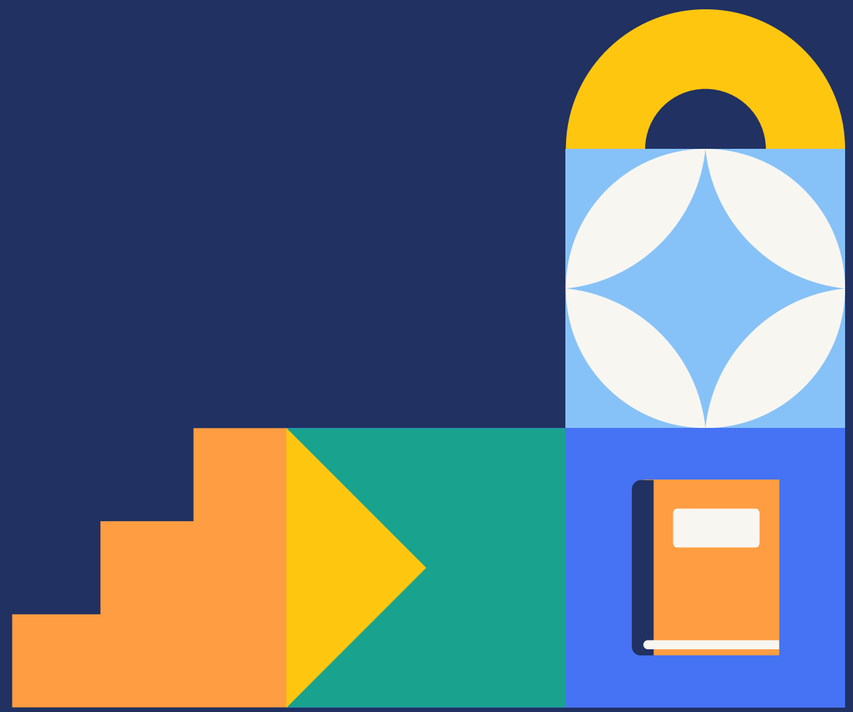
Tools Used in Data Science

- Programming Languages: Python, R, SQL.
- Data Wrangling: Pandas, Numpy.
- Data Visualization: Matplotlib, Seaborn, Tableau.
- Machine Learning: Scikit-learn, TensorFlow, PyTorch.
- Big Data: Hadoop, Spark.
- Databases: MySQL, MongoDB, PostgreSQL



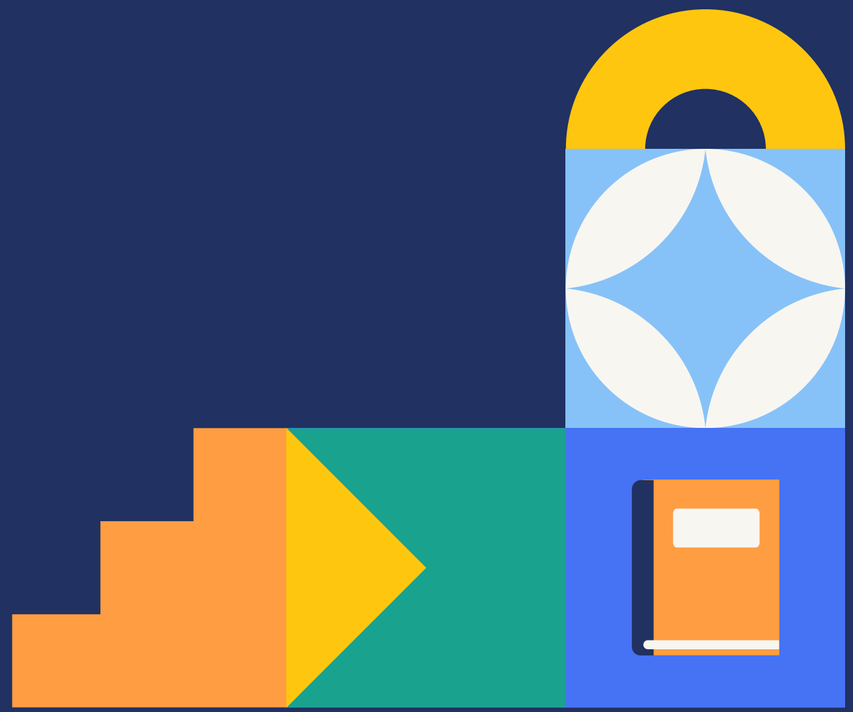
Types of Data Science Projects

- Predictive Modeling: Forecasting sales, weather, stock prices.
- Classification: Categorizing data (e.g., spam vs. non-spam emails).
- Clustering: Grouping customers based on purchasing behavior.
- Recommendation Systems: Suggesting products or content (e.g., Netflix, Amazon).



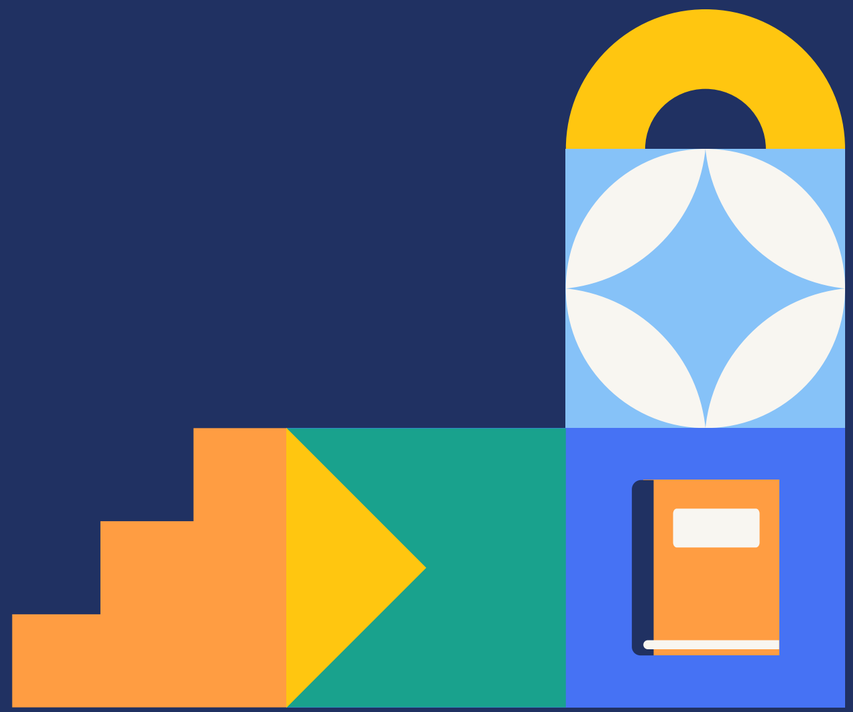
Data Science Applications

- Healthcare: Predicting diseases, personalized treatment plans.
- Finance: Fraud detection, algorithmic trading.
- Retail: Customer behavior analysis, personalized recommendations.
- Robotics and AI: Automation, predictive maintenance



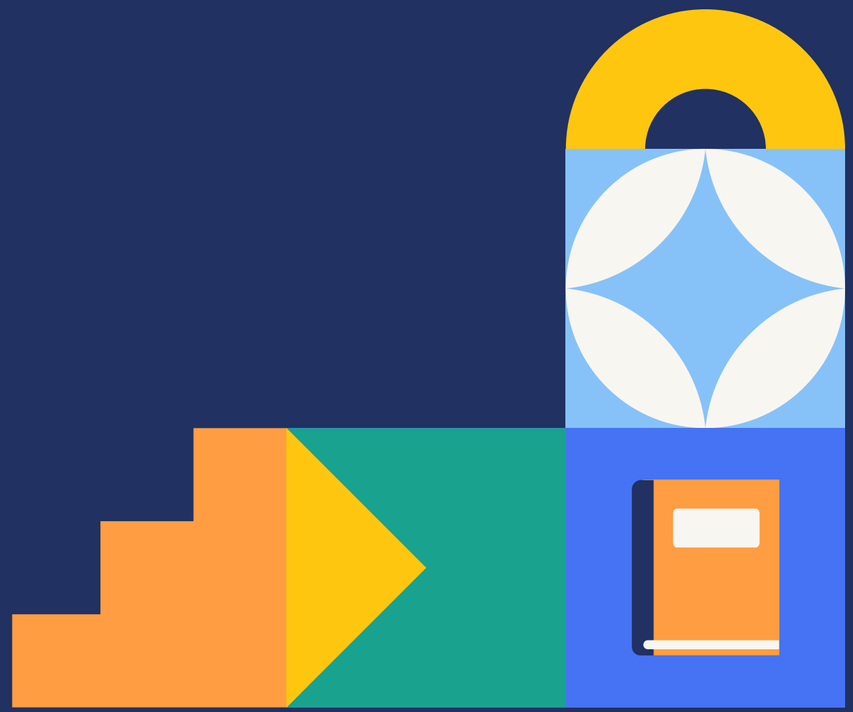
Career Paths in Data Science

- Data Scientist: Uses data to make predictions and insights.
- Data Analyst: Analyzes historical data to inform decisions.
- Machine Learning Engineer: Builds models for predictive tasks.
- Data Engineer: Designs infrastructure for data generation and storage.
- AI Researcher: Develops new algorithms and models.



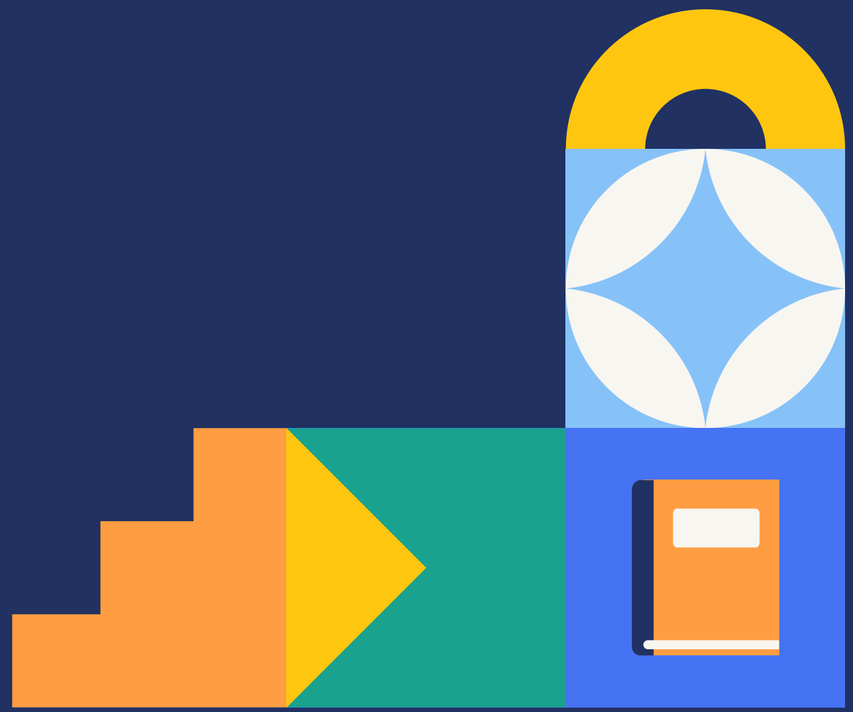
Skills Needed to Become a Data Scientist

- Mathematics and Statistics: Strong analytical foundation.
- Programming: Python, R, SQL.
- Data Visualization: Ability to convey insights clearly.
- Machine Learning: Knowledge of key algorithms and models.
- Critical Thinking: Ability to ask the right questions and interpret results

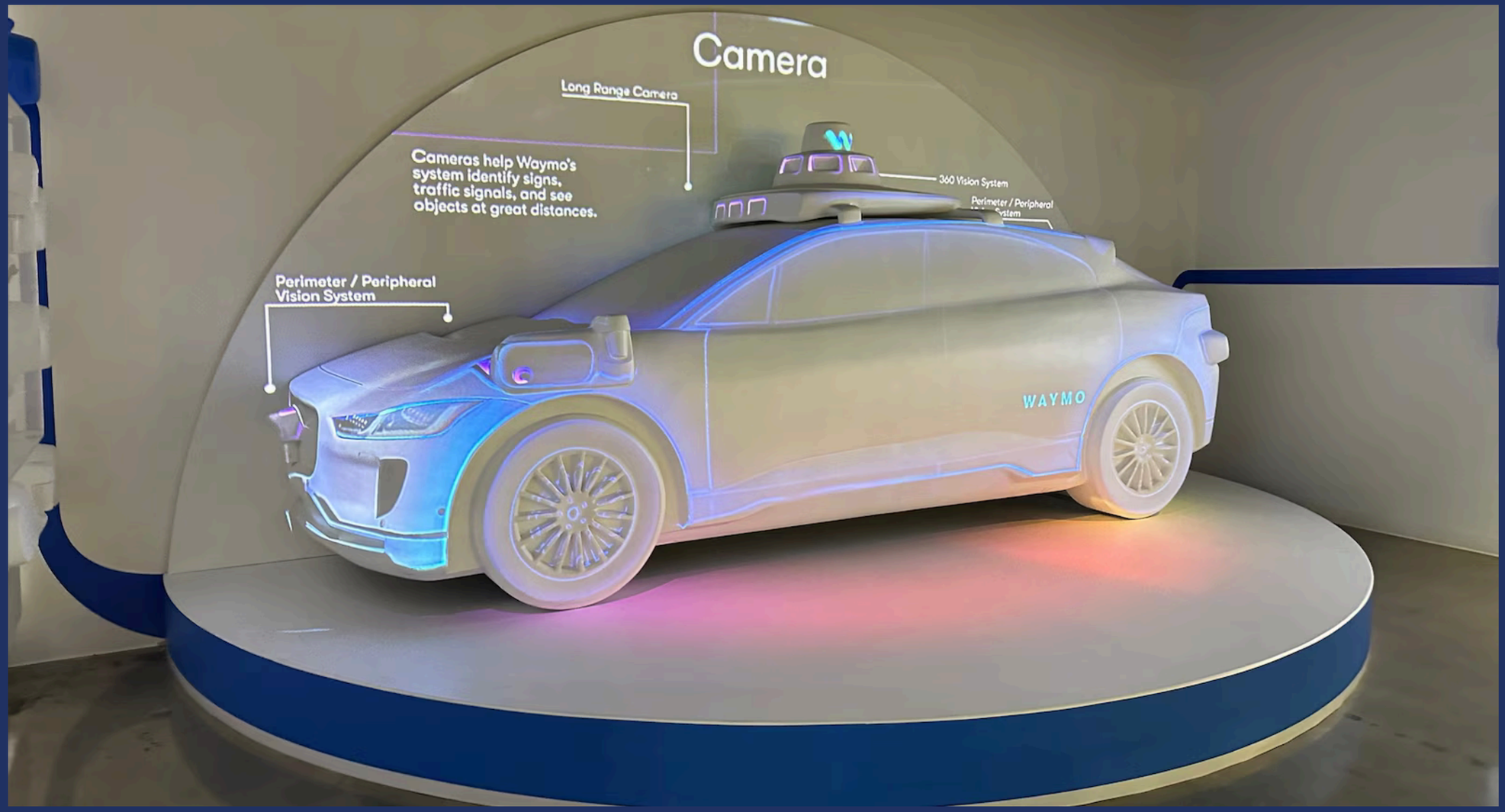


Getting Started ??

1. Ask the right questions - To understand the business problem.
2. Explore and collect data - From database, web logs, customer feedback, etc.
3. Extract the data - Transform the data to a standardized format.
4. Clean the data - Remove erroneous values from the data.
5. Find and replace missing values - Check for missing values and replace them with a suitable value (e.g. an average value).
6. Normalize data - Scale the values in a practical range (e.g. 140 cm is smaller than 1,8 m. However, the number 140 is larger than 1,8. - so scaling is important).
7. Analyze data, find patterns and make future predictions.
8. Represent the result - Present the result with useful insights in a way the "company" can understand.



Waymo


Camera
Long Range Camera
Cameras help Waymo's system identify signs, traffic signals, and see objects at great distances.
360 Vision System
Perimeter / Peripheral Vision System
Perimeter / Peripheral
WAYMO

Conclusion

Data Science: powerful tool for solving complex problems and driving business decisions.

Future: AI, IoT, Robotics, and Big Data are making Data Science more impactful

Start learning: Data Science unlocks endless opportunities!

The End

Introduction to Machine Learning

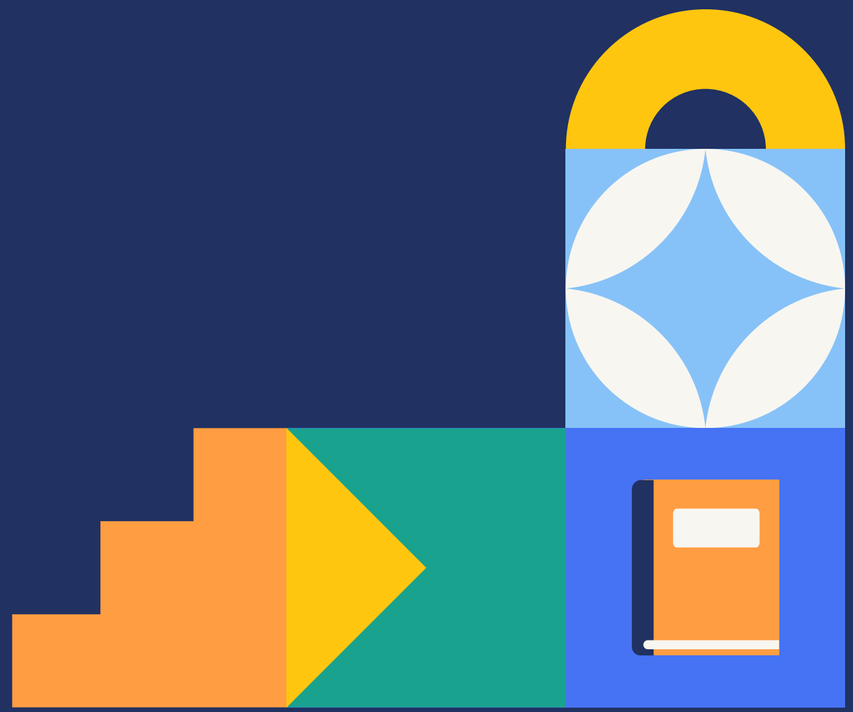
BY: DAREL OBALLA

What is Machine Learning?

Machine learning is a branch of artificial intelligence (AI) focused on building systems that learn from data to make decisions or predictions.

Instead of being explicitly programmed, machines learn patterns from data.

Example: Think about Netflix recommending movies or TV shows. Netflix doesn't know you personally, but it learns from your viewing habits and the habits of similar users to suggest content you'll likely enjoy.



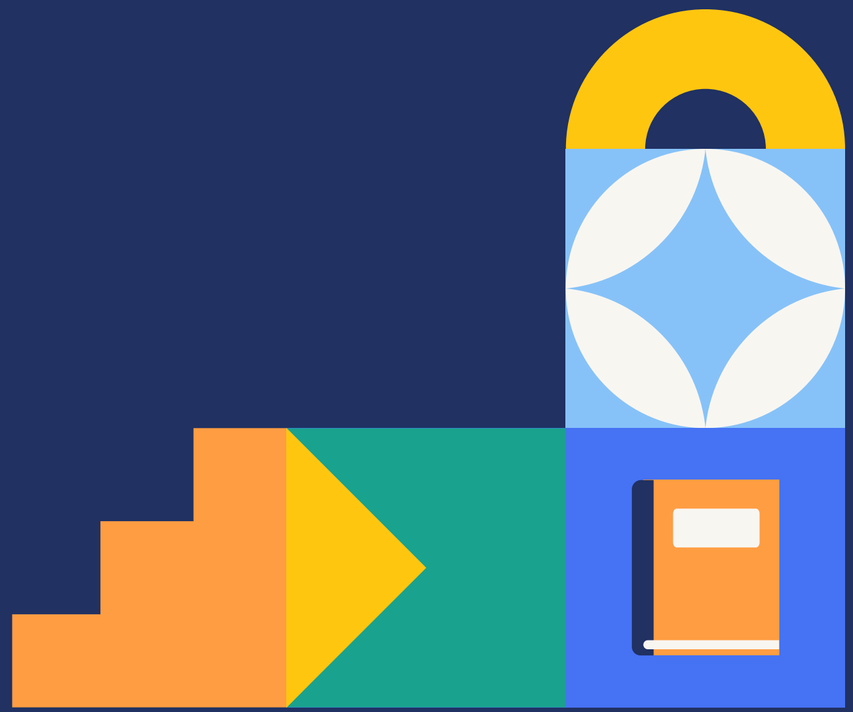
Traditional Programming vs. Machine Learning

Traditional Programming: Programmer gives instructions, and the machine follows

Machine Learning: Machine learns from data and predicts or performs tasks without explicit instructions

Example:

- Traditional Programming: You instruct a computer to add 2 + 2, and it always gives you 4.
- Machine Learning: Spotify learns your music preferences over time and builds personalized playlists based on your listening behavior without you explicitly telling it what you like.



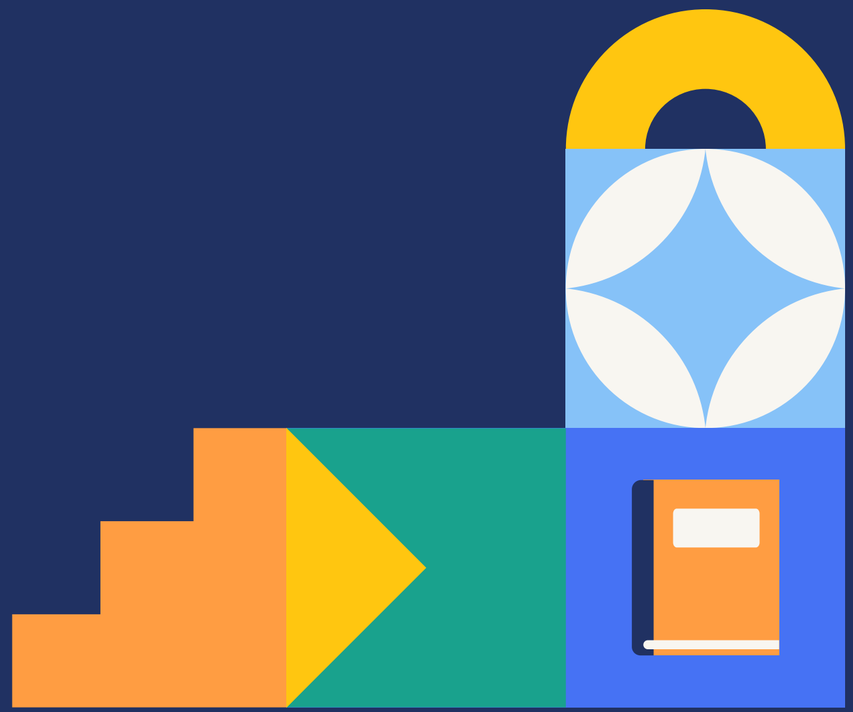
Traditional Programming vs. Machine Learning

Traditional Programming: Programmer gives instructions, and the machine follows

Machine Learning: Machine learns from data and predicts or performs tasks without explicit instructions

Example:

- Traditional Programming: You instruct a computer to add 2 + 2, and it always gives you 4.
- Machine Learning: Spotify learns your music preferences over time and builds personalized playlists based on your listening behavior without you explicitly telling it what you like.



Types of Machine Learning

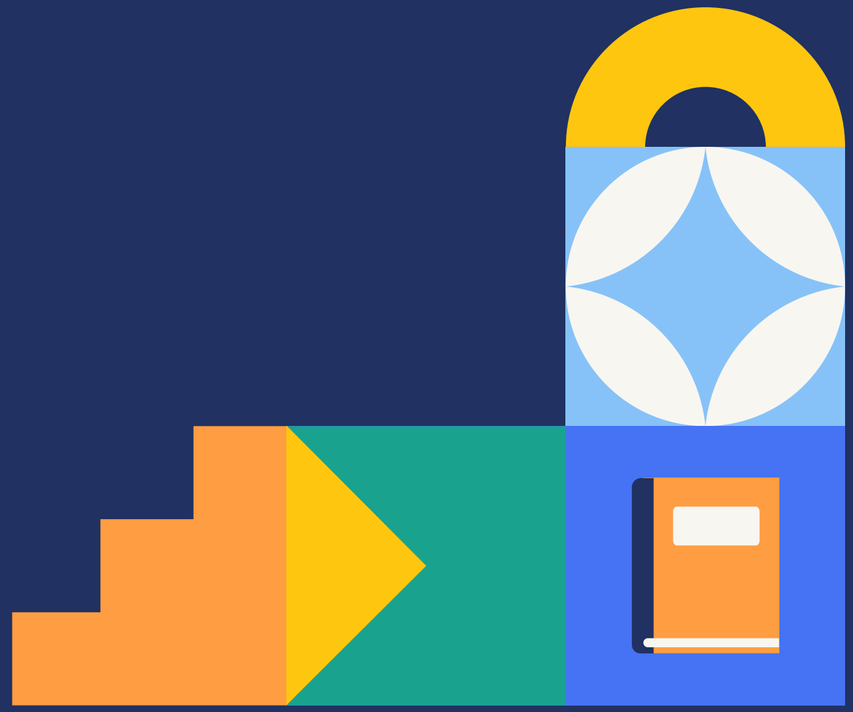
Supervised Learning: Learn from labeled data (e.g., classification, regression).

Unsupervised Learning: Find hidden patterns in data (e.g., clustering, dimensionality reduction).

Reinforcement Learning: Learn by trial and error (e.g., game AI, robotics).

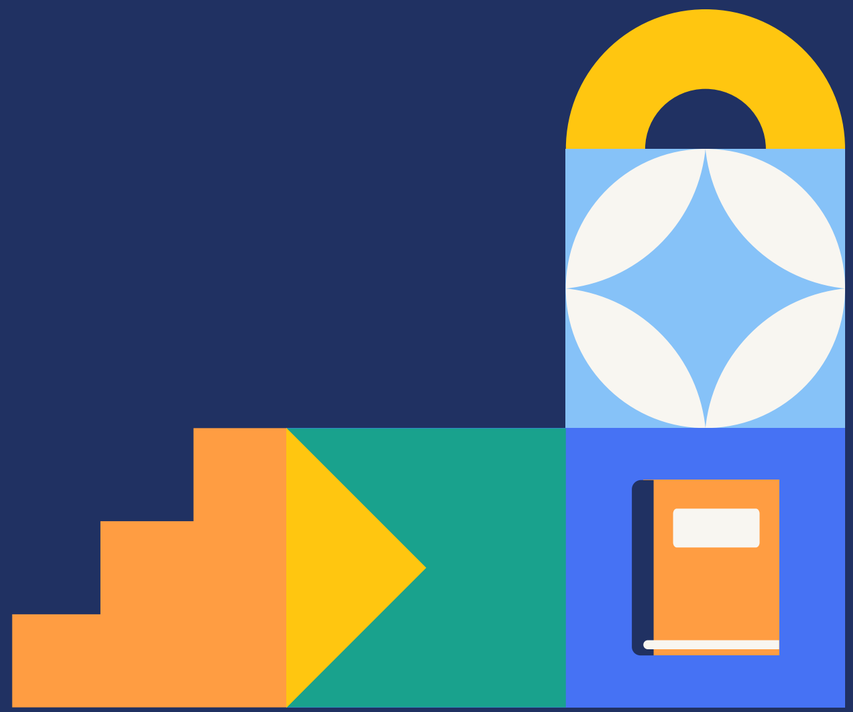
- Email Spam Detection – Gmail or Outlook classifies emails into “spam” and “not spam” based on past labeled data of spam emails.
- Example 2: Predicting House Prices – Apps like Zillow estimate home prices based on features like size, location, and market trends.

- Example 1: Customer Segmentation – Retailers like Amazon group their customers into different segments (e.g., high spenders, frequent shoppers) without prior labeling. This helps them target customers with relevant offers.
- Example 2: Social Media – Facebook clusters users’ posts and forms user communities based on shared interests or topics, which is why you see recommended groups or pages based on your activity.



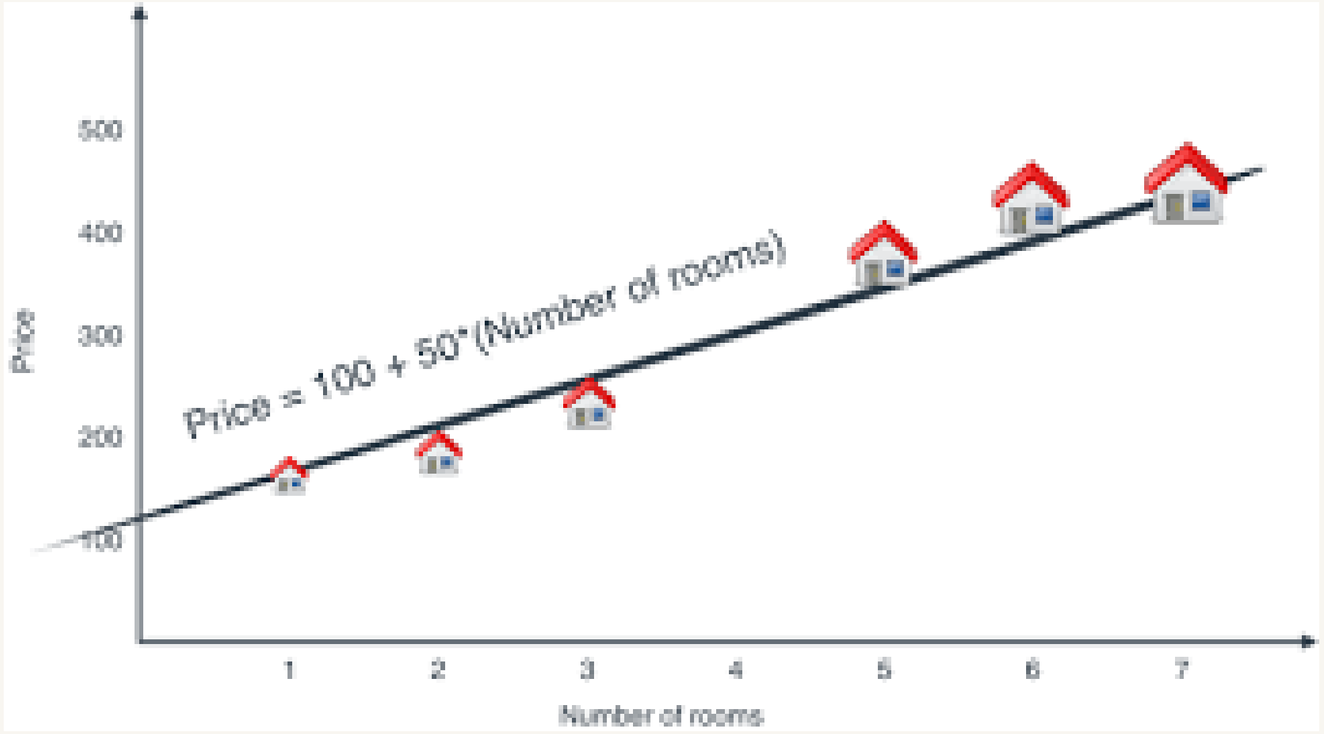
Reinforcement Learning

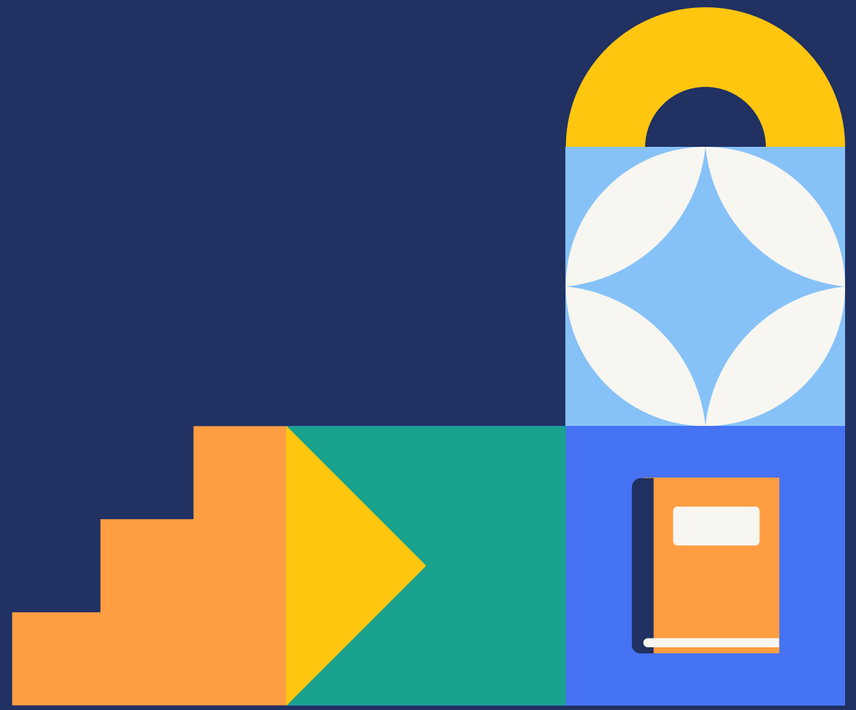
- Example 1: Self-driving cars – Companies like Tesla use reinforcement learning, where the car learns by driving through real-world situations, getting rewarded for safe driving and penalized for mistakes.
- Example 2: Video games – AI opponents in games like FIFA or chess learn strategies over time by playing multiple rounds, improving their gameplay with each trial.



Supervised Learning

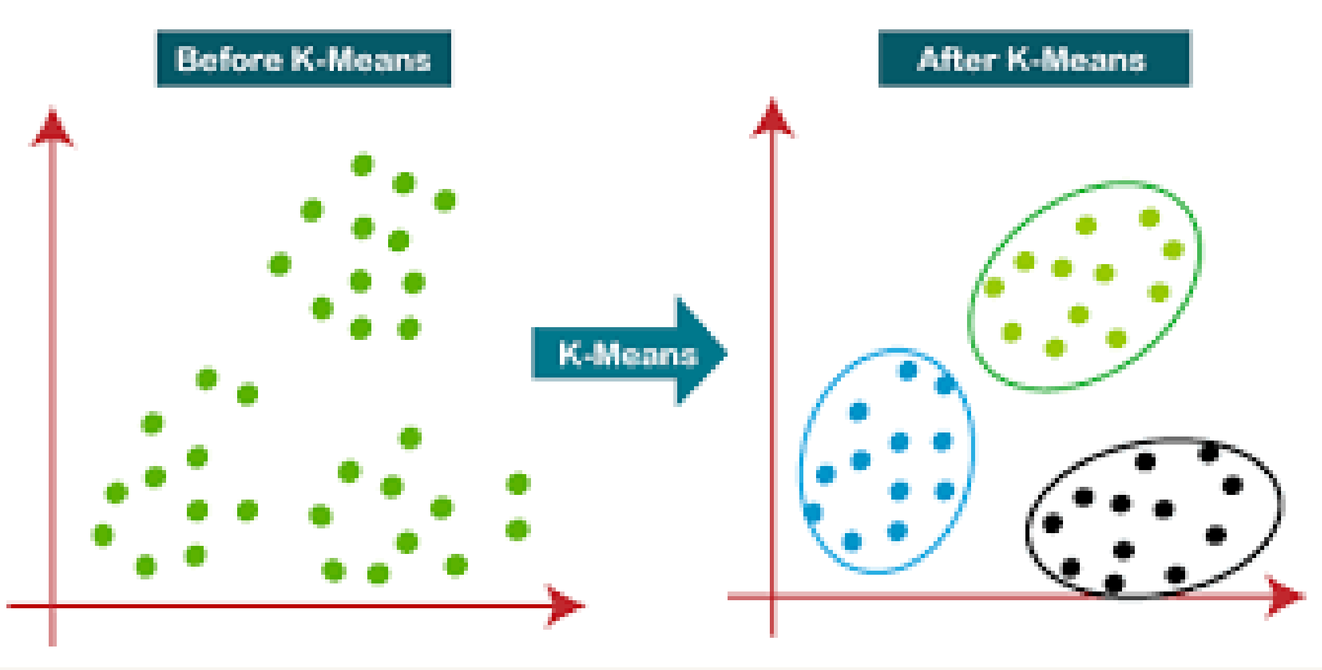
- Key Concept: Models learn from input-output pairs.
- Examples: Predicting house prices (regression), spam detection (classification).
- Algorithms: Linear Regression, Decision Trees, Support Vector Machines (SVM).

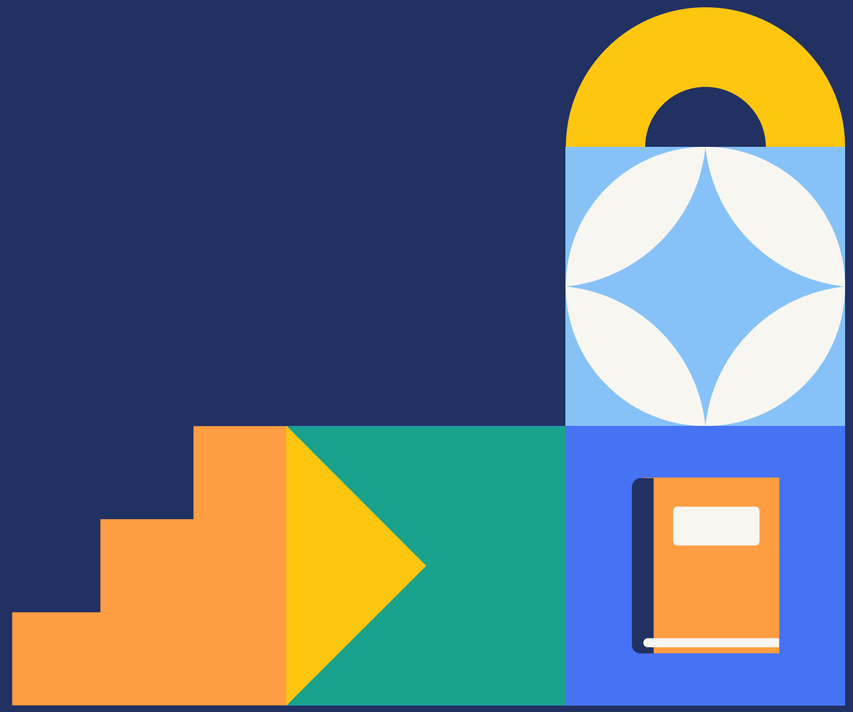




Unsupervised Learning

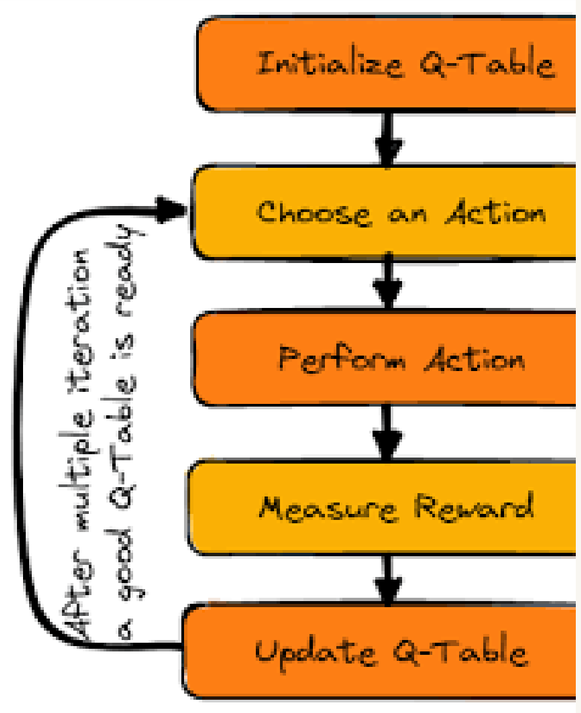
- Key Concept: Models find patterns without labeled data.
- Examples: Customer segmentation, anomaly detection.
- Algorithms: K-Means Clustering, PCA (Principal Component Analysis).

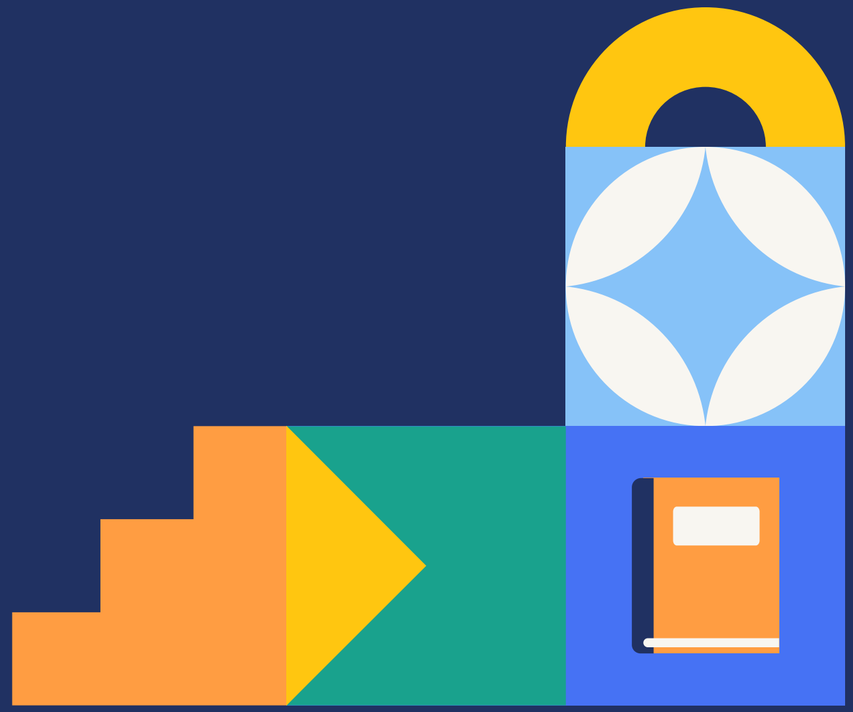




: Reinforcement Learning

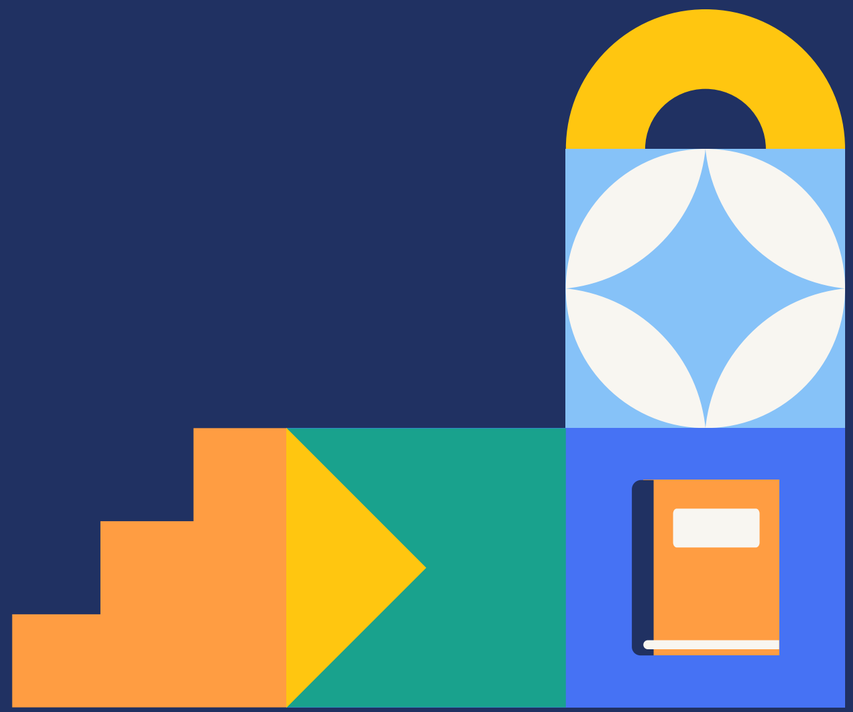
- Key Concept: Learn by interacting with an environment, receiving feedback via rewards or penalties.
- Examples: Self-driving cars, game bots (AlphaGo).
- Algorithms :Q-Learning, Deep Q Networks (DQN).





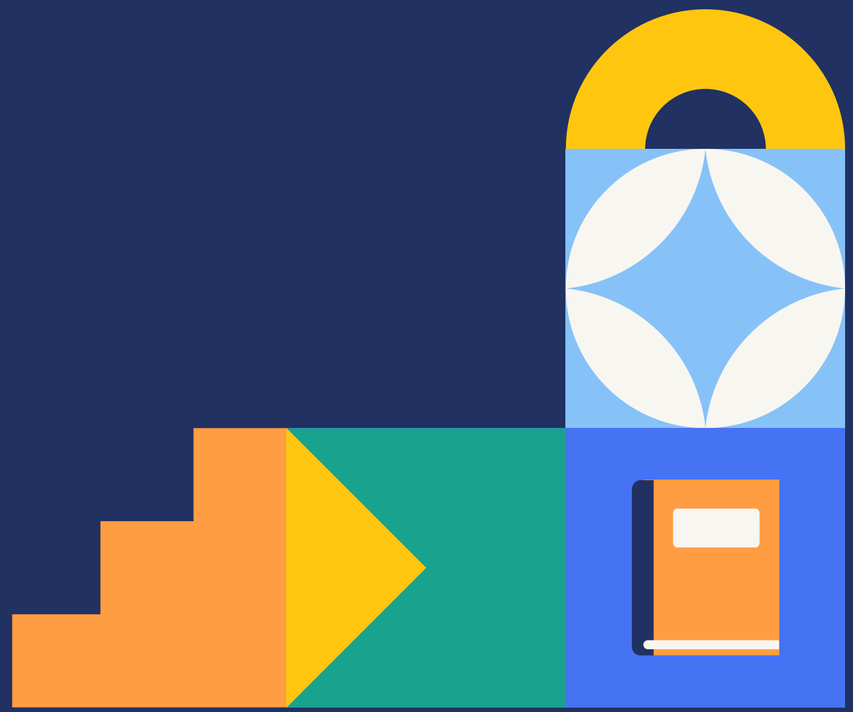
Key Machine Learning Terminologies

- Model: The mathematical representation of the real-world problem.
- Training: The process of learning from data.
- Features: The inputs to the model (e.g., age, income).
- Labels: The outputs for supervised learning.
- Overfitting: When the model performs well on training data but poorly on unseen data.



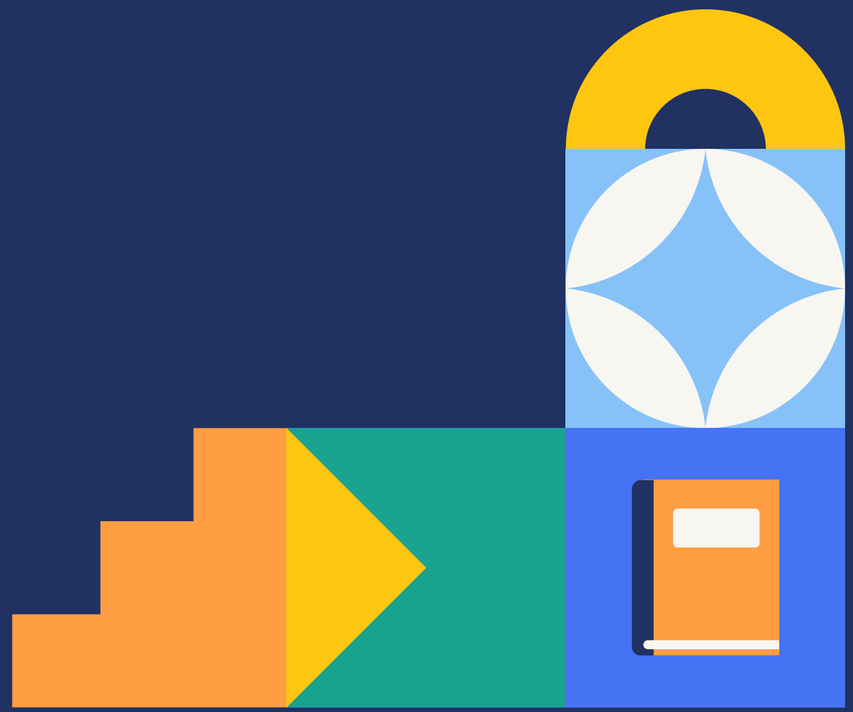
Creating a Machine Learning Model

- Define Problem: Understand what you're trying to predict or discover.
- Collect Data: Gather and preprocess relevant data.
- Choose a Model: Select the appropriate algorithm.
- Train the Model: Fit the model to the training data.
- Evaluate the Model: Test the model's performance on unseen data.
- Deploy: Use the model in real-world scenarios.



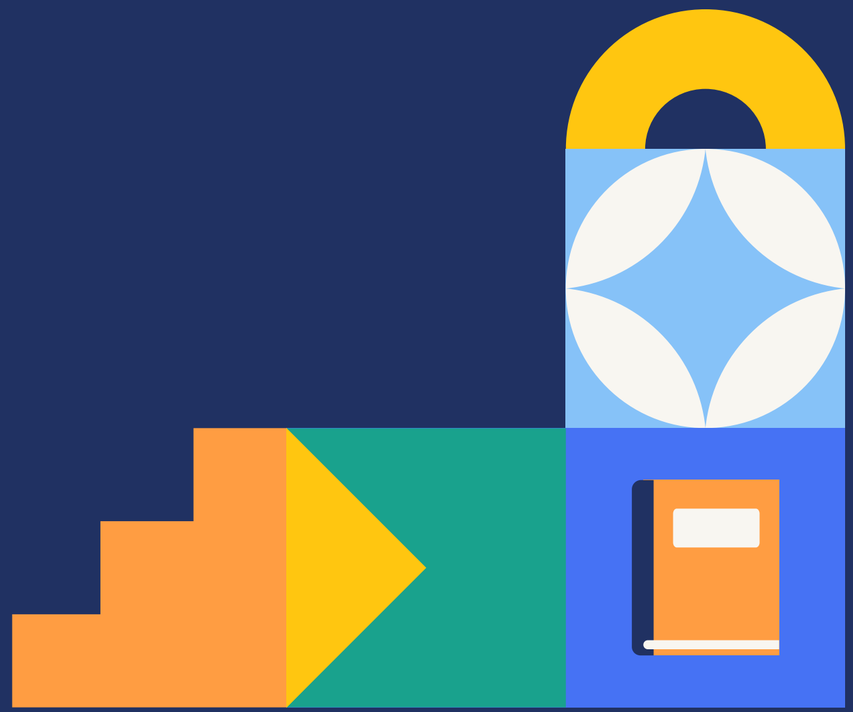
Applications of Machine Learning

- Healthcare: Disease prediction, drug discovery.
- Finance: Fraud detection, stock market analysis.
- E-commerce: Product recommendations, personalized marketing.
- Robotics: Autonomous systems, industrial automation.



Challenges in Machine Learning Model Development

- Data Quality: Garbage in, garbage out.
- Overfitting and Underfitting: Balancing model complexity.
- Ethics: Bias in data, fairness, transparency.



Resource Page

Use these icons and illustrations in your Canva Presentation. Happy designing! Don't forget to delete this page before presenting.

Try this background for online class.

*Please delete this section before downloading.