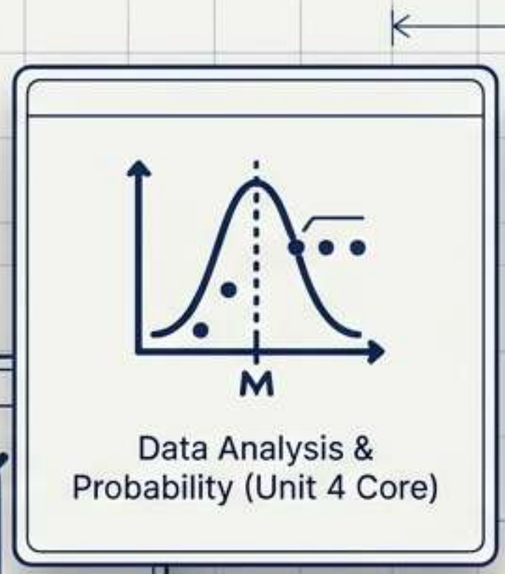
Statistical Data & No-Code AI

The Ultimate Revision Playbook for CBSE Class 10 (Unit 4)









Mastering numbers, predicting the future, and building AI models without typing a single line of code.

Gemini Notebook

CURRICULUM UNIT MAP: THE FOUR PILLARS

DATA SCIENCE BASICS



- Defining statistical data and real-world applications.

CORE STATISTICS



- Underlying mathematical frameworks (Mean, Median, Variance, Probability).

THE NO-CODE REVOLUTION



- High-Code vs. Low-Code vs. No-Code software approaches.

ORANGE DATA MINING



- Practical application and the Palmer Penguins case study.

PHASE 1: FOUNDATION

PHASE 2. CORE CONCEPTS

PHASE 3. TRANSITION

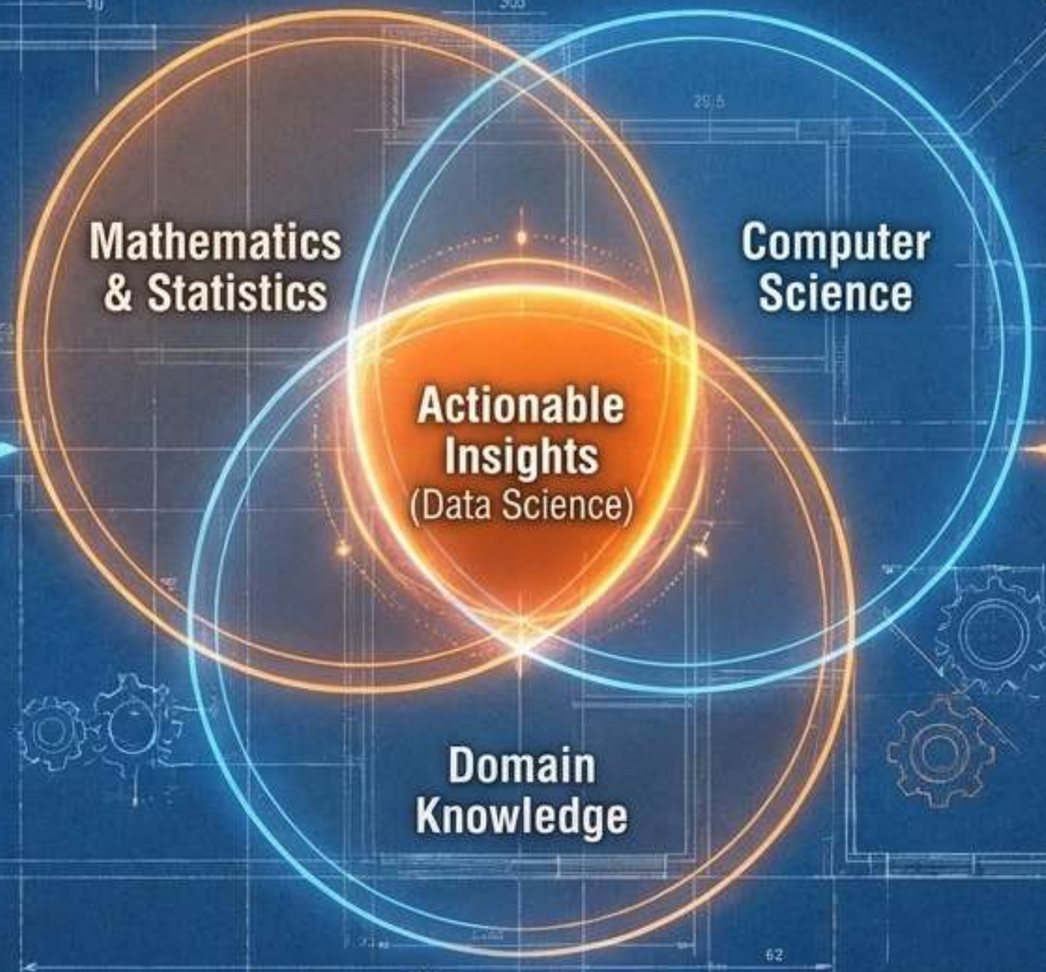
PHASE 4: APPLICATION

Gemini Notebook

WHAT IS STATISTICAL DATA IN AI?

THE DEFINITION

- **What it is:** An AI domain that collects, maintains, and analyzes large amounts of structured **NUMERICAL** and categorical data (numbers, tables, records).
- **What it is NOT:** It does not look at images (Computer Vision) or listen to speech (NLP).



THE GOAL

- Extract useful patterns and trends from massive datasets.
- Support data-driven decision-making (e.g., predicting next month's inventory).

PHASE 1: FOUNDATION | PHASE 2. CORE CONCEPTS | PHASE 3. TRANSITION | PHASE 4. APPLICATION

Gemini Notebook

REAL-WORLD APPLICATIONS MATRIX



INTERNET SEARCH

- **MECHANISM**: Algorithms match queries against 20+ petabytes of daily data.
- **EXAMPLE**: Finding relevant results for 'Buy iPhone without money' instantly.



TARGETED ADVERTISING

- **ANALOGY**: Traditional ads = Rocket Launcher (hits everyone). Targeted ads = Sniper (hits specific audience based on behavior).
- **MECHANISM**: Analyzes past user behavior for high Click-Through Rates (CTR).



WEBSITE RECOMMENDATIONS

- **MECHANISM**: Suggests products based on past searches (Association Models).
- **EXAMPLE**: Amazon, Netflix, Spotify predicting what you want next.



GENETICS & HEALTHCARE

- **MECHANISM**: Combines medical and genomic data for personalized treatment.
- **EXAMPLE**: Predicting diabetes risk based on patient statistics (age, BMI, glucose).

PHASE 1: FOUNDATION | PHASE 2. CORE CONCEPTS | PHASE 3. TRANSITION | PHASE 4. APPLICATION

Gemini Notebook

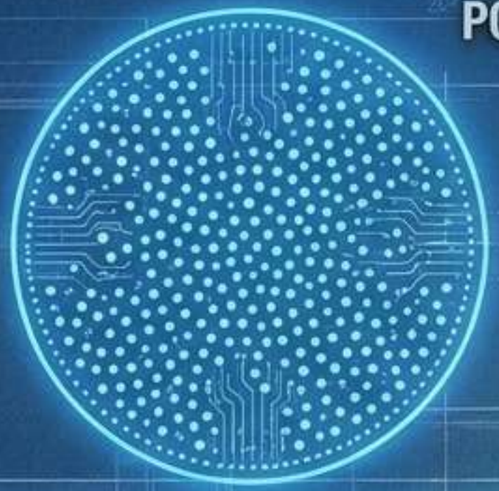
FOUNDATIONAL STATISTICS: THE CORE METRICS

POPULATION VS. SAMPLE

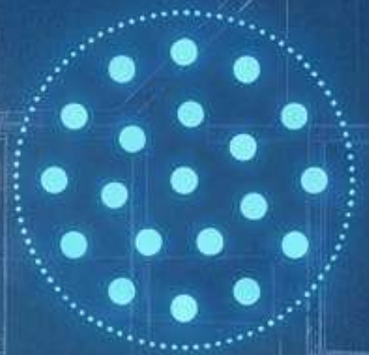
PHASE 3: CORE CONCEPTS

PHASE 3: TRANSITION

POPULATION:
The entire universe of raw data







SAMPLE:
A subset extracted for feasible testing

DESCRIPTIVE STATISTICS (CENTRAL TENDENCY)

MEAN



The central average value. ($\sum x / n$)

MEDIAN



The middle value in a sorted list.

MODE



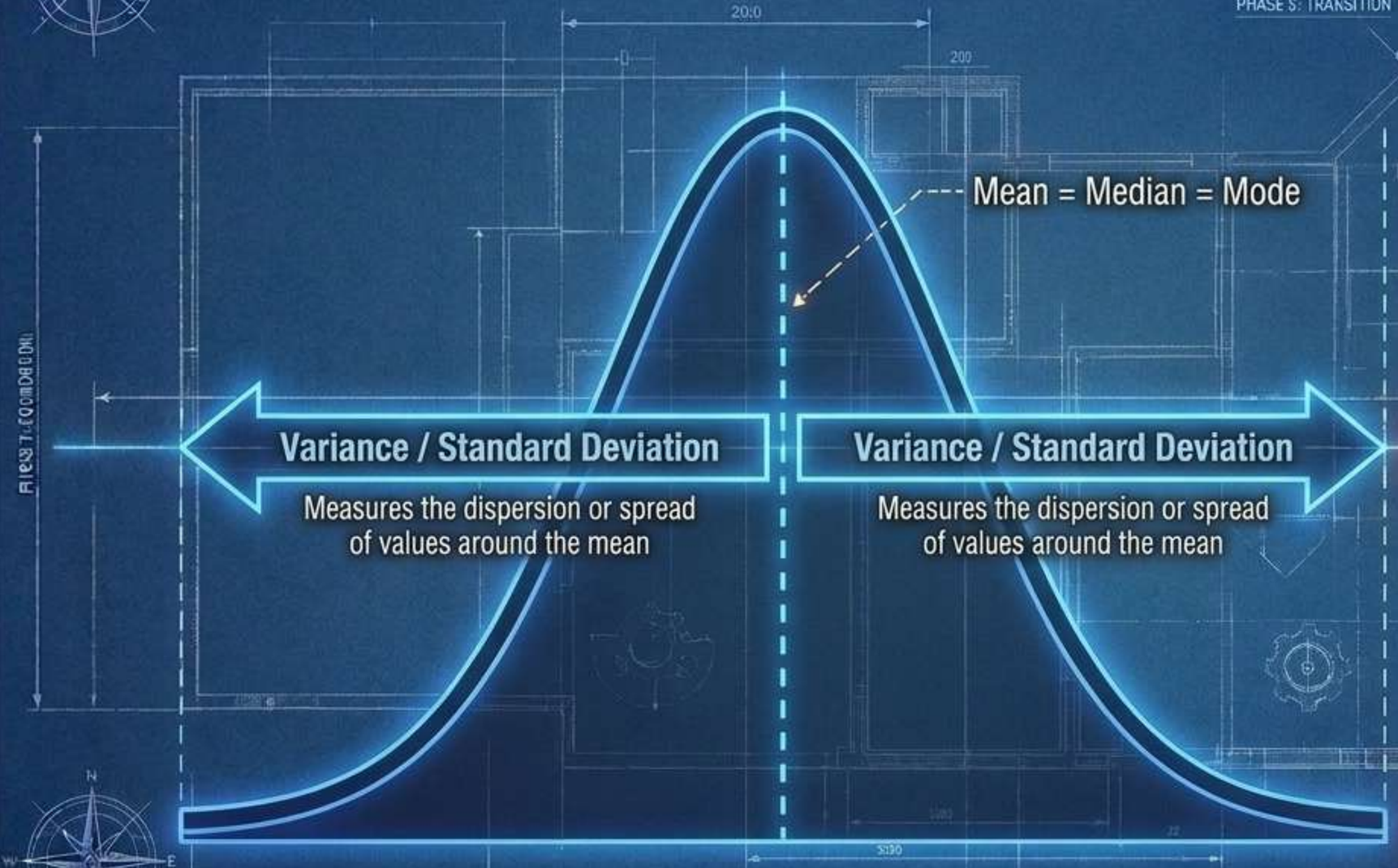
The data point that occurs most frequently.

PHASE 2: CORE CONCEPTS

PHASE 3: TRANSITION

Gemini Notebook

DISTRIBUTION, VARIANCE & PROBABILITY



Probability

- **Definition:** The likelihood (chance) of an event occurring.
- **Independent:** Not affected by other events.
- **Dependent:** Affected by past events.

Outlier:
A data point lying at an abnormal distance from other values

PHASE 1: FOUNDATION
PHASE 2. CORE CONCEPTS
PHASE S. TRANSITION
PHASE 4. APPLICATION
PHASE 2: CORE CONCEPTS
PHASE 3. TRANSITION

Gemini Notebook

THE SOFTWARE DEVELOPMENT SPECTRUM



Feature	High-Code	Low-Code	No-Code
Definition	Write all code manually (Python/Java)	Visual platforms + custom scripting	100% visual drag-and-drop
Target Users	Specialized Software Engineers	Users with basic technical logic	Anyone (Beginners, Doctors, Architects)
Cost & Speed	Most Expensive / Slowest	Moderate Cost / Faster	Least Expensive / Fastest
Customization Level	Infinite flexibility	High flexibility	Limited to pre-built tool widgets

PHASE 1: FOUNDATION
PHASE 2: CORE CONCEPTS
PHASE 3: TRANSITION
PHASE 4: APPLICATION

PHASE 2: CORE CONCEPTS
PHASE 3: TRANSITION

Gemini Notebook

"No-code removes the typing, never the thinking."

The No-Code AI Paradigm

Eliminates Syntax Errors:
No missing semicolons or indentation bugs.

Fast Prototyping:
A working model can be ready on the exact same day.



Doctor



Student

AI Hub



AI Hub



Architect



Wildlife Manager

Cost-Effective:
Companies don't need a team of 20 expensive data scientists.

Real-Time Visualization:
See the blocks and data flows visually rather than staring at a black screen of code.

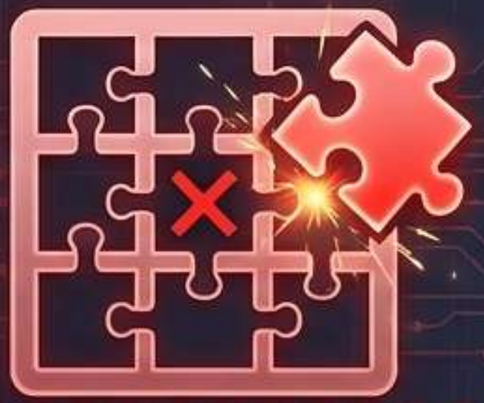
PHASE 1: FOUNDATION
PHASE 2: CORE CONCEPTS
PHASE 3: TRANSITION
PHASE 4: APPLICATION

PHASE 2: CORE CONCEPTS
PHASE 3: TRANSITION

Gemini Notebook

THE HIDDEN RISKS OF NO-CODE AI

LACK OF FLEXIBILITY



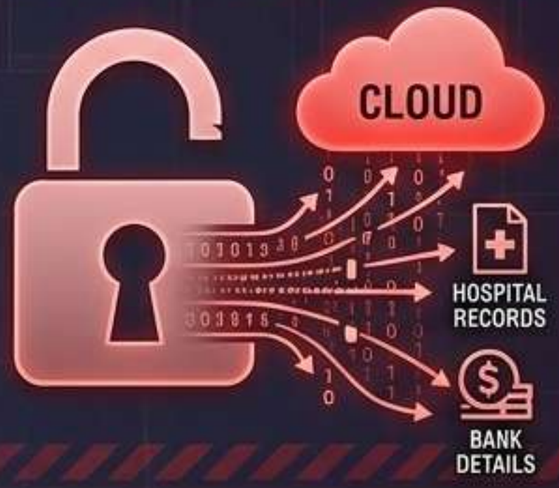
Restricted to the pre-built widgets provided by the tool. Cannot customize advanced algorithms or features.

AUTOMATION BIAS



Over-trusting the machine's results. Blindly trusting an 80% confident model without applying human context or judgment.

SECURITY & PRIVACY ISSUES



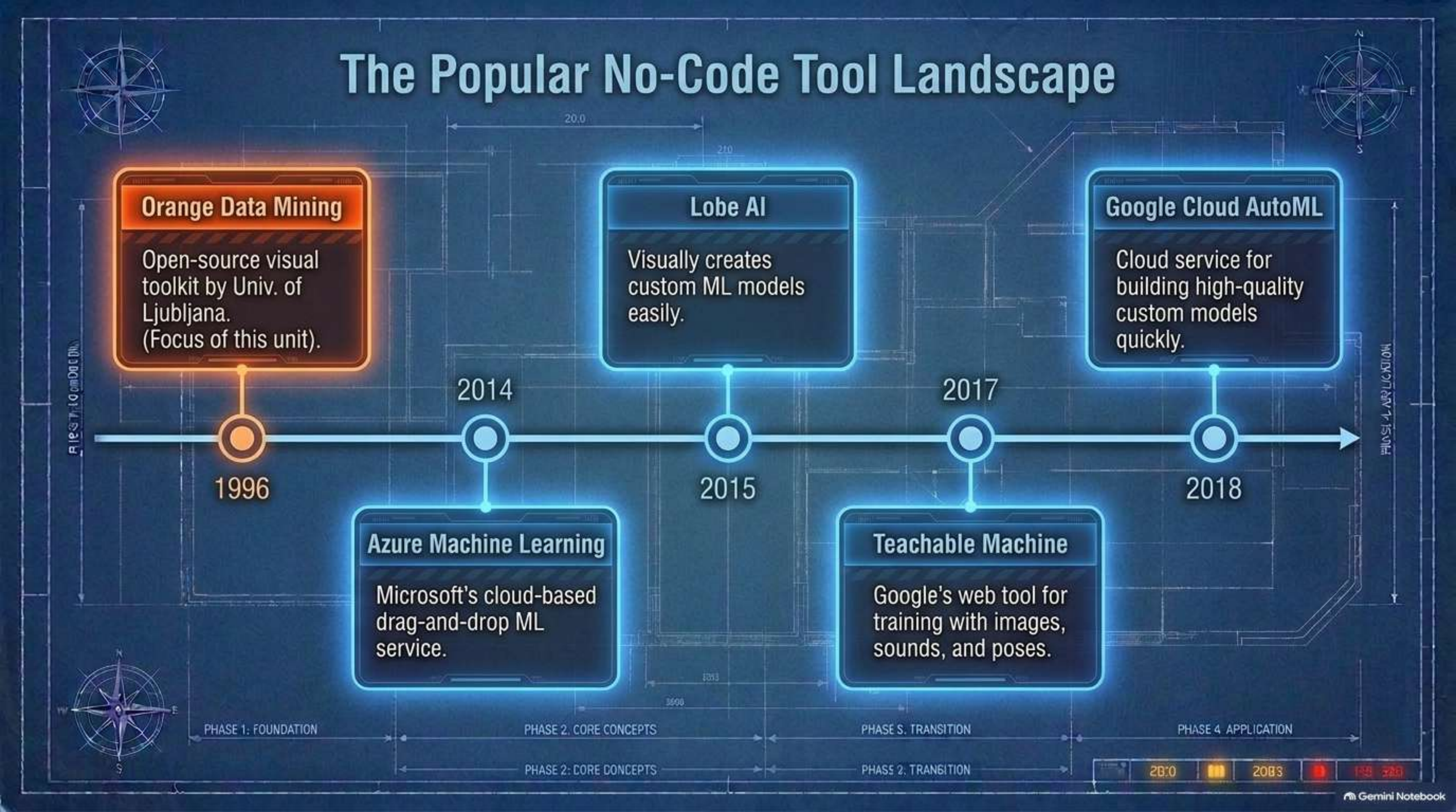
Cloud-based No-Code tools may not not be secure enough for highly sensitive information like hospital records or bank details.

PHASE 1: FOUNDATION
PHASE 2: CORE CONCEPTS
PHASE 3: TRANSITION
PHASE 4: APPLICATION

PHASE 2: CORE CONCEPTS
PHASE 2: TRANSITION

Gemini Notebook

The Popular No-Code Tool Landscape



Gemini Notebook

Introduction to Orange Data Mining

```python
import pandas as pd
from sklearn.ensemble import RandomForestClassifier

# Load data
X, y = load_data()

# Initialize and train model
clf = RandomForestClassifier()
clf.fit(X, y)

# Predict
predictions = clf.predict(X_test)
```

Visual Interface



Key Facts

What it is:
- An open-source, No-Code machine learning and data visualization toolkit.

How it works:
- Users build workflows by placing and connecting modular 'Widgets' on a digital canvas.

The Magic:
- Data analysis is done through a hidden Python backend, completely operated via visual programming.

Gemini Notebook

The Orange Widget Catalog

Data Exploration



Examples: Data Table, Scatter Plot

Purpose: Used to view records and understand data distributions.

Preprocessing



Examples: Impute, Normalize

Purpose: Used to clean data (handle missing N/A values) and scale it to a common range.

Feature Selection



Examples: Select Columns

Purpose: Used to choose the most important variables for analysis.

Modeling



Examples: Random Forest, Support Vector Machine, Logistic Regression

Purpose: The algorithms that learn patterns from the data.

Evaluation



Examples: Test & Score

Purpose: Used to check model accuracy and performance metrics.

PHASE 1: FOUNDATION

PHASE 2: CORE CONCEPTS

PHASE 3: TRANSITION

PHASE 4: APPLICATION

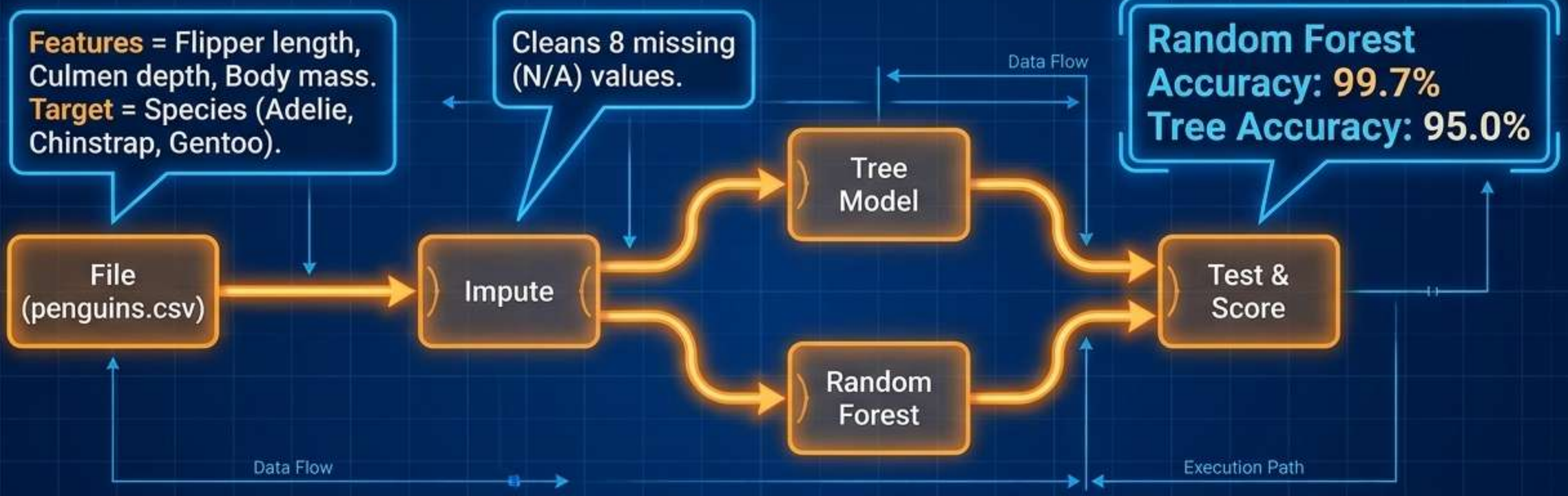
Gemini Notebook

Mapping the AI Project Cycle in Orange

AI Project Cycle Stage	Corresponding Orange Action
1. Problem Scoping	Define the goal: Predict penguin species.
2. Data Acquisition	Use the File Widget to load penguins.csv.
3. Data Exploration	Use Data Table & Scatter Plot to visualize missing values.
4. Data Preprocessing	Use Impute Widget to handle N/A (missing data).
5. Modeling	Connect data to the Random Forest or Tree Widget.
6. Evaluation	Connect to Test & Score to read Classification Accuracy.

Gemini Notebook

Practical Walkthrough: Palmer Penguins



Key Takeaway: Random Forest algorithm significantly outperforms the simple Tree model on this dataset.

Gemini Notebook

Chapter Synthesis & Exam Cheat Sheet

Unit 4: Statistical Data

The Core (Data & Stats)

- Large Numerical Datasets.
- Mean, Median, Mode (Central Tendency).
- Variance & Outliers (Spread).

The Approach (No-Code)

- 100% Visual, Drag & Drop.
- Fast, Accessible, Cheap.
- Risks: Automation Bias, Inflexible, Security.

The Tool (Orange)

- Open-source, Python backend.
- Workflow: File -> Impute -> Model -> Test & Score.



No-code removes the typing, never the thinking. Good luck on your practicals!

Gemini Notebook