1. **Brief History of Morden AI and its Applications**

* Introduce
  + AI: any system that can *sense – reason – act – adapt* (includes rule-based systems).
  + ML: a subset of AI, able to learn from data.
  + DL: a subset of ML, using multi-layered neural networks, improving with large amounts of data.
* Major breakthroughs:
  + Image processing: since 2015, machines classify images (dog and cat) more accurately than humans.
  + Machine translation: near-human translation, considering grammar and word order.
  + Drivers of progress: deep learning + big data + computing power.
  + Real-world applications: personalized advertising, supply chain optimization, self-driving cars, smart homes (entertainment, security).
  + Historical significance: AI is considered the *“electricity of the 21st century”* (Andrew Ng) → impact across all industries.
* History
  + 1996: Deep Blue defeats the world chess champion.
  + 2006: Geoffrey Hinton solves the gradient problem, neural networks are rebranded as Deep Learning.
  + 2009: ImageNet is created, with millions of labeled images.
  + 2012: AlexNet wins the ImageNet competition with a significantly lower error rate than other algorithms.
  + 2014: Major progress in NLP and Computer Vision.
  + 2015: TensorFlow is released.
  + 2016: AlphaGo defeats a professional Go player.
  + 2018: Waymo launches self-driving taxi services.
  + 2019: IBM Project Debater debates against humans.
* Applications
  + Transportation: Google Maps, Waze calculate shortest routes and predict traffic.
  + Ride-sharing: Uber, Lyft use dynamic pricing based on supply and demand.
  + Social Media: personalized content recommendation, connection suggestions, targeted advertising.
  + NLP: Siri, Alexa process voice commands.
  + Computer Vision: Facebook face recognition, self-driving cars, unattended baggage detection.

1. **Retrieving and Cleaning Data**

## Data Retrieval

* + **SQL Databases**: Structured, relational data with fixed schema, queried using SQL.
  + **NoSQL Databases**: Unstructured or semi-structured data (JSON, documents, graphs, wide-column), more flexible.
  + **APIs & Cloud Sources**: External/web/cloud data, often delivered in CSV or JSON.
* Data Cleaning
  + Duplicate / unnecessary data.
  + Inconsistent / typos.
  + Missing data.
  + Outliers.
  + Data source issues.
* Solution
  + **Missing data**: remove row / impute / mask category.
  + **Outliers**: remove / impute / transform / robust model.
  + **Residuals**: standardized/ deleted/ studentized

1. **Exploratory Data Analysis and Feature Engineering**

### **Exploratory Data Analysis**

* + **Patterns/ trends/ hypothesis**
  + **Summary statistics:** mean, median, std, min, max

### **EDA with Visualization**

* + Visualization libraries: Matplotlib/Pandas/Seaborn
  + Plot types: scatter/histograms/box plots/pairplot/groupby plots

### **Feature Engineering**

* + **Variable transformation**: log, polynomial features
  + **Feature encoding**: binary, one-hot, ordinal
  + **Feature scaling**: standard, min-max, robust

**IV. Inferential Statistics and Hypothesis Testing**

* **Estimation and Inference**
  + Inferential statistics: Learn population characteristics from sample
    - Population = parameters (mean, std)
    - Sample = statistics
  + Parametric model: uses fixed number of parameters
  + Maximum Likelihood Estimation (MLE): common method to estimate parameters
  + Estimation: determine population parameters from model fit
  + Common distributions: uniform, normal, log-normal, exponential, Poisson
  + Approaches:
    - Frequentist: based on repeated experiments
    - Bayesian: parameters described by probability distributions
* **Hypothesis Testing**
  + Hypothesis: statement about population parameter
    - Null (H0) vs Alternative (H1)
  + Hypothesis test: rule to accept/reject H0 using test statistic
  + Errors:
    - Type 1 (α): false positive
    - Type 2 (β): false negative
  + Significance level (α): usually 0.01 or 0.05
  + p-value: smallest threshold to reject H0
  + Confidence interval: range of statistics where H0 is accepted
* **Correlation & Causation**
  + Correlation predicts but does not imply causation
  + Consider confounding variables
  + Spurious correlations: random correlations in data