**PART 1: PYTHON & PANDAS SUMMARY**

**ESSENTIAL COMMANDS**

read\_csv() – Load CSV files into DataFrame  
head(), shape, dtypes, columns – Inspect structure and schema  
loc[], iloc[] – Access data by label or index

**ANALYSIS & CLEANING**

value\_counts(), unique(), mean() – Summary metrics  
describe(), quantile() – Statistical overview  
fillna(), drop(), replace() – Handle missing, unneeded, or erroneous data

**TRANSFORMATIONS & EXPORT**

astype() – Type conversion  
str.lower(), apply() – Normalize and customize  
to\_csv() – Export results for sharing or storage

**GROUPING & MERGING**

groupby(), pivot\_table() – Categorize and summarize  
merge(), concat() – Join datasets by keys or stacking

**FUNCTIONS**

lambda, def, apply, map – Create and apply functions for data transformation

**COMMON PATTERNS**

isnull().sum() – Missing value count  
get\_dummies() – Encode categorical data  
df[df["Age"] > 30] – Row filtering

**PART 2: DATA GOVERNANCE, SECURITY & LIFECYCLE (WEEKS 7–9)**

**WEEK 7: DATA GOVERNANCE**

**Definition:** Ensures data is secure, accurate, and usable  
**Key Principles:** Access control, auditability, compliance (GDPR, HIPAA, PIPEDA)  
**Why Important:** Rising data volumes (175 ZB by 2025), increasing global interaction with data  
**Key Components:**

* **Tools:** Automation, ML for policy enforcement
* **Enterprise Dictionary:** Classifies data, defines policies
* **Roles:** Stewards, custodians, compliance officers
* **Business Value:** Risk management, better decisions, productivity

**WEEK 8: GOVERNANCE OVER DATA LIFE CYCLE**

**Phases:**

* **Creation:** Data entry, ingestion (structured/unstructured)
* **Processing:** ETL, lineage tracking, cleaning, classification
* **Storage:** Data warehouses/lakes, encrypted backups
* **Usage:** Analytics, ML, BI – needs strict access & audits
* **Archiving:** Cold storage with retention rules
* **Destruction:** Policy-driven purging

**Data Quality Factors:** Accuracy, completeness, timeliness  
**Techniques:** Scorecards, profiling, deduplication, lineage tracking  
**Transformations:**

* Extract → Transform → Load into platforms like BigQuery, Redshift

**WEEK 9: DATA PRIVACY & SECURITY**

**Security Layers:**

* Physical: Facilities, disaster recovery
* Network: Firewalls, IDS
* OS: Authentication methods
* App: RBAC
* Storage: Encryption, backup integrity

**Cloud Security:**

* Multi-tenant risks require strict isolation
* IAM: RBAC, OAuth, API keys
* Zero Trust: Least privilege + continuous authentication

**Monitoring:**

* Why: Anomaly detection, compliance, optimization
* What: Data quality, lineage, security logs

**Building Data Culture:**

* Transparency & Trust
* Clear role definitions
* Training & updated protocols

**PART 3: ETHICS IN TECH & AI**

**ETHICAL PRINCIPLES & FRAMEWORKS**

**Codes Studied:** ACM, BCS, CIPS, EU AI Ethics, GenAI Code  
**Values:** Integrity, privacy, public benefit, explainability  
**Frameworks:**

* Utilitarianism – Outcomes-focused
* Deontology – Duty/rights-focused

**BEST PRACTICES**

* Human-in-loop decision-making
* Use model cards and bias audits
* Ensure fairness in outcomes

**CASE STUDIES**

* **Cambridge Analytica** – Unauthorized data use
* **COMPAS** – Discriminatory model bias
* **Amazon Rekognition Pause** – Ethical restraint

**PART 4: MIDTERM 2 THEORY RECAP**

**DATA TYPES**

**Transactional:** Dynamic, real-time (e.g., sales logs)  
**Analytical:** Static summaries (e.g., dashboards)

**CLOUD SECURITY**

**Risks:** Data theft, vendor lock-in, loss of control  
**Controls:** Encryption, IAM, logs, GDPR compliance

**ZERO TRUST & IBAC**

**Zero Trust:** All users/devices verified before access  
**IBAC:** Permissions based on identity/role

**LINEAGE & QUALITY**

**Lineage:** Tracks origin and data changes **Quality:** Ensure completeness, accuracy, consistency

**MONITORING & ROLES**

**What to Monitor:** Usage, compliance, anomalies **Roles:**

* Owner: Sets access & policies
* Steward: Maintains integrity
* Analyst: Extracts value

**ETHICAL AI**

* Needs new rules for fairness, bias, transparency
* Encourage proactive audits, stakeholder consultation