A **Data Readiness Checklist** is a tool to ensure that your data is prepared, clean, and suitable for analysis, modelling, or other data-driven tasks. Below is a comprehensive checklist to assess data readiness:

1. Data Availability

- ☐ **Data Exists**: Ensure the required data is available and accessible.

- ☐ **Data Sources Identified**: All relevant data sources (databases, APIs, files, etc.) are identified.

- ☐ **Access Permissions**: Access to the data is granted and permissions are verified.

- ☐ **Data Volume**: The volume of data is sufficient for the intended analysis or modelling.

2. Data Quality

- ☐ **Completeness**: No missing values or, if present, they are handled appropriately.

- ☐ **Accuracy**: Data is free from errors and reflects real-world values.

- ☐ **Consistency**: Data is consistent across sources and over time.

- ☐ **Timeliness**: Data is up-to-date and relevant for the analysis.

- ☐ **Uniqueness**: No duplicate records exist.

- ☐ **Validity**: Data conforms to the expected format, type, and range.

3. Data Structure and Format

- ☐ **Structured Data**: Data is organized in a structured format (e.g., tables, JSON, CSV).

- ☐ **File Formats**: Data is in a usable format (e.g., CSV, Excel, Parquet, etc.).

- ☐ **Schema**: A clear schema or data dictionary is available.

- ☐ **Normalization**: Data is normalized or denormalized as required for the task.

4. Data Integration

☐ **Data Combined**: Data from multiple sources is integrated (if needed).

☐ **Joins/Merges**: Relationships between datasets are defined and implemented.

☐ **ETL/ELT Processes**: Data pipelines are in place for extraction, transformation, and loading.

5. Data Cleaning

☐ **Outliers Handled**: Outliers are identified and addressed.

☐ **Missing Values**: Missing values are imputed or removed.

☐ **Data Transformation**: Data is transformed (e.g., scaling, encoding, aggregation) as needed.

☐ **Noise Reduction**: Irrelevant or noisy data is removed.

6. Data Documentation

☐ **Metadata Available**: Metadata (e.g., column descriptions, units, etc.)

☐ **Data Dictionary**: Explain variables, types, and meanings.

☐ **Source Documentation**: Sources of data and any transformations.

7. Data Security and Compliance

☐ **Data Privacy**: Sensitive data is anonymized or pseudonymized.

☐ **Compliance**: Data complies with relevant regulations (e.g., GDPR, HIPAA).

☐ **Encryption**: Data is encrypted during storage and transmission.

8. Data Usability

☐ **Data is Queryable**: Data can be easily queried or filtered.

☐ **Data is Shareable**: Data can be shared with stakeholders in a usable format.

☐ **Data is Interpretable**: Data is understandable and ready for analysis.

9. Data Testing

☐ **Data Validation**: Data has been validated against business rules or constraints.

☐ **Sample Testing**: A sample of the data has been tested for quality and usability.

☐ **Edge Cases**: Edge cases and anomalies have been identified and addressed.

10. Stakeholder Alignment

☐ **Business Requirements Met**: Data aligns with the business objectives and requirements.

☐ **Stakeholder Approval**: Key stakeholders have reviewed and approved the data.

☐ **Use Case Alignment**: Data is suitable for the intended use case (e.g., reporting, ML, analytics).

11. Technical Readiness

☐ **Infrastructure Ready**: Systems and tools for data storage and processing are in place.

☐ **Scalability**: Infrastructure can handle the volume and velocity of data.

☐ **Backup and Recovery**: Data is backed up and recovery processes are in place.

12. Final Checks

☐ **Data is Versioned**: Changes to the data are tracked and versioned.

☐ **Data is Monitored**: Ongoing monitoring for data quality and integrity is established.

☐ **Data is Ready for Use**: All checks are complete, and data is ready for analysis or modelling.

This checklist can be customized based on the specific needs of your project or organization. Regularly reviewing and updating the checklist ensures that your data remains reliable and actionable.

Contact ravi@iai.cx for more information.