

Worksheet: AI Data Prep Checklist and Considerations

Use this checklist to plan your data strategy for predictive modeling using Einstein Studio Model Builder. The considerations listed below can vary by use case and business needs.

Goal Use Cases Permissions Considerations Integrations Ethical Use of Data Identity Resolution Requirements Variables Data Model Analyze Test

1 Determine the Goal

Determine the business outcome that you want to improve

Review the Trailhead module, [Machine Learning Predictions: Quick Look](#)

Complete the [Build AI Models in Einstein Studio Trailhead](#) module



2

Define Your Use Case(s)

Explore the [Salesforce AI Use Case Library](#) for practical use cases. Align your use case to the AI capabilities that can help you achieve your goal. Here are some predictive AI examples to consider:

- Prioritize leads based on their likelihood to convert
- Automate case routing to determine the best team to handle incoming cases
- Identify the likelihood of customer churn

USE CASE

DESCRIPTION

FEATURES

STAKEHOLDERS

3

Understand Permissions

Review the standard Data Cloud [permission sets](#)

Review permissions to [manage your AI models](#)



4

Review Data Requirements and Considerations

Understand [data usage types that impact Data Cloud billing](#)

Understand [Data Cloud Limits and Guidelines](#)

5

Identify Data Sources and Integrations

Consider [data types and formats](#), including the structure of date/time data types

Evaluate the use of [connectors](#) to bring in external data

Identify your [data sources](#) and datasets

6

Review Ethical Use of Data

Complete the [Ethical Data Use Best Practices: Quick Look Trailhead module](#)

Learn about the [Ethical Use of Data](#) in the Data Platform trailmix

Understand [ethics, privacy, and consent](#) and its impact on customer data usage

7

Consider Identity Resolution

Evaluate if you need to use identity resolution to [unify customer profiles](#)

Review [data modeling requirements](#) to use individual entity rulesets

Identify what data is shared across sources (examples: email, first name, customer ID, and so on)

8

Understand Data Model Requirements

Complete the [Customer 360 Data Model for Data Cloud Trailhead module](#)

Learn about [data model concepts](#) and what to consider when building a data model

Learn about [data model objects \(DMOs\)](#) in Data Cloud Reference Guide

Understand [data model subject areas](#)

Understand [mapping](#) requirements

9

Identify or Create Model Variables

Define what variables to use, as well as the granularity and scope of the data

Create variables by aggregating engagement or activity data



10 Create a Data Model

Review the [Customer 360 Data Model for Data Cloud Trailhead module](#)

Evaluate which source has the most reliable and highest quality data

11 Analyze Your Data

Review the [Data Quality Trailhead module](#)

Review your data sources and audit your data

- Consider data duplication, missing values, outliers, inconsistencies, and other issues

- Consider issues related to the variables you decide to use

- Determine the scope of your historical data

12 Test Your Data

Resolve data quality issues such as cleanliness, relevance, ethics, bias, and efficiency

Evaluate sandbox requirements and best practices to improve your model quality

Determine if you need to use [packages](#) and/or [data kits](#) to move your objects and models for testing in other orgs