Per i dataset strutturati ho scelto:

* Dataset Bike Sharing:

https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset

Gli attributi sono:

* codice bici
* data
* stagione
* anno
* mese
* ora
* giorno festivo
* weekend
* giorno feriale
* meteo
* temperatura
* temperatura percepita
* umidità
* velocità vento
* utenti occasionali
* utenti registrati
* totale utenti

Gli identificativi possono essere le varie tipologie di bici a noleggio, per esempio:

* mountain bike
* city bike
* e-bike
* bici da corsa

* Dataset valore automobili:

https://archive.ics.uci.edu/ml/datasets/car+evaluation

Gli attributi sono:

* prezzo
* dimensioni
* porte
* posti auto
* dimensione bagagliaio
* sicurezza

Gli identificativi sono classificazioni di auto:

* inaccettabile
* accettabile
* buona
* molto buona

* Dataset consumo di frutta negli USA nel 2021

https://www.statista.com/statistics/477475/us-most-consumed-fruit-and-fruit-products-by-type/

L'attributo è:

* quantità consumata

Gli identificativi sono:

* banane
* fragole
* uva
* mele
* angurie
* arance
* mirtilli
* limoni
* pesche
* avocado

Per i dataset non strutturati ho scelto:

* Dataset travel blogger Instagram:

in un dataset che comprende foto, video, posizione e testo di un profilo Instagram di un travel blogger, dove vorrei capire quali sono le migliori destinazioni di viaggio, cercherei di raggruppare le foto e video per posizione geografica.

Da qui analizzerei le foto e video che sono stati visti, piaciuti e commentati maggiormente per estrapolare le migliori mete di viaggio.

* Dataset e-mail

in una casella di posta in arrivo di un determinato indirizzo e-mail cercherei di filtrarla per i vari mittenti per cui mi interessa ottenere informazioni e le sposterei in varie cartelle dedicate, utilizzando un sistema di regole automatizzato.

Dopodichè creerei delle ulteriori sottocartelle suddivise per argomento trattato utilizzando sempre un sistema di regole automatizzato attraverso le parole chiave usate nell'oggetto delle mail.