Problem Statements – Advanced Apex Project 1 (2025-26, Cohort 1, Trimester 1)

Problem ID	Problem Statement	Brief Description	Expected Machine Learning (ML) Model(s)	Dataset Clues (examples only)
P1	Predicting Rental Demand	Predict daily/weekly/etc rental demand for shared vehicles (bikes, scooters, cars, etc.) using environmental and temporal factors (temperature, humidity, day type, holidays, seasonality, etc) and/or other task relevant features. The aim is to understand which factors most strongly influence demand and to support better resource allocation and planning.	Regression: Simple Linear Regression (SLR), Multiple Linear Regression (MLR)	Seoul Bike Sharing (UCI), CitiBike NYC (Open Data), Divvy Chicago, Car/Scooter rental datasets, etc. *Your assigned Project Supervisor can also help you in choosing the appropriate dataset.*
P2	Customer Segmentation for Retail/E-Commerce/Any relevant Domain	Segment customers into groups based on purchase frequency, recency, and monetary value (RFM), as well as other relevant features such as average spend per visit, basket size, and diversity of products purchased. The goal is to identify actionable customer segments for personalized marketing, retention, and upselling strategies.	Clustering: K-Means, Hierarchical Clustering	Online Retail II Dataset (Kaggle), Supermarket Sales Dataset, or any e-commerce/transaction dataset with customer-level data pertaining to relevant domains, etc. *Your assigned Project Supervisor can also help you in choosing the appropriate dataset.*
P3	Real Estate Price Prediction	Predict property prices based on features such as size, number of rooms, age, location, and neighborhood characteristics, and other relevant features. The goal is to identify key factors influencing pricing and support valuation and investment decision-making.	Regression: Simple Linear Regression, Multiple Linear Regression	Zillow Research (US), NHB RESIDEX (India), Ames Housing (Kaggle), Real Estate Valuation (UCI, Taiwan), etc. *Your assigned Project Supervisor can also help you in choosing the appropriate dataset.*
P4	Forecast Passenger Demand	Forecast the number of passengers airlines, railways, or public transport, etc. systems for upcoming Week/months/Quarter/etc using historical transport data. Incorporate lag features, trends, and seasonality and other task relevant features to model demand patterns. Goal: To aid planning for airlines, railways, or public transport systems.	Time-Series Forecasting: Regression-based forecasting (Lag Features, Moving Average, EWMA, Trend/Seasonality analysis, etc.)	AirPassengers (UCI), US Airline Passengers (Kaggle), Indian Railways Passenger Data, NYC Subway Ridership, TfL London Transport, etc. *Your assigned Project Supervisor can also help you in choosing the appropriate dataset.*
P5	Clustering Biological Specimens by Morphological Features	Group biological specimens (e.g., flowers, seeds, penguins, etc) into clusters based on morphological or physical measurements, without using class labels. The goal is to discover natural groupings where specimens in the same cluster share similar characteristics and potentially similar behaviors or biological functions. You have the privilege to compare the discovered clusters to known species/classes (if available) and analyze clustering quality.	Clustering: K-Means, Hierarchical Clustering	Iris Dataset (UCI), Seeds Dataset (UCI), Palmer Penguins (Kaggle), Plant/Leaf datasets, etc. *Your assigned Project Supervisor can also help you in choosing the appropriate dataset.*