Lab 1

Datasets

1 Electric vehicle population Dataset

problem Statement: The dataset contains information about Battery Electric vehicles and plug-in Hybrid Electric vehicles registered in washington state. it Includes details like vehicle type, make, model, Electric range and location. The dataset has 223,995 rows and 17 columns with missing values. possible problem statements include predicting EV adoption based on vehicle features and geographic location, analyzing regional factors influencing EV popularity

Number of rows = 223995
Number of columns = 17

Columns

1. VIN
2. county
3. City
4. State
5. postal Code
6. Model Year
7. Make
8. Model
9. Electric Vehicle Type
10. Clean Alternative fuel vehicle Eligible
11. Elettric Range
12. Base MSRP
13. Legislative District
14. DOL vehicle ID
15. vehicle location
16. Electric Utility
17. 2020 Census Tract

2 Laptop price Dataset

problem statement: This dataset contains 11,768 records of laptops with various hardware specifications and prices. it is designed for predictive modeling, price estimation and exploratory data analysis. the dataset includes real world values to reflect the factors that influence laptop prices. This dataset can be used for ~~pr~~, Building machine learning models to predict laptop prices based on specification, Market Analysis & Feauture importance study.

number of rows = 11769
number of columns = 11

coloumns

① Brand
② processor
③ RAM
④ Storage
⑤ GPU
⑥ Screen Size
⑦ Resolution
⑧ Battery life
⑨ weight
⑩ operating System
⑪ price.

3 Indian Bike Sales Dataset

problem statement: This dataset contains records of motorcycle sales across various indian state, covering top brands like Honda, Royal Enfield, TVS, Yamaha, Hero, Bajaj, KTM and Kawasaki. The dataset includes key attributes such as average daily distance traveled, engine capacity, fuel type etc. it provides insights into bike sales trends, market demand, and resale value across different city tiers

number of rows = 10001
number of columns = 15

Coloums

(1) State
(2) Avg Daily Distance
(3) Brand
(4) Model
(5) Price
(6) Year of Manufature
(7) Engine Capacity
(8) Fuel Type
(9) Mileage
(10) owner Typ
(11) Registration
(12) insurence
(13) Seller Type
(14) Resale price
(15) City Tier

(4) Road Accident Survival Dataset

problem Statement:
This dataset contains detailed records of simulated road accident data, focusing on factors influencing survival outcomes. The dataset includes demographic, behavioral and situational attributes, providing valuable insight into how various factor impact the survival probability during road accidents.

number of rows = 201
number of columns = 6

columns
① Age
② Gender
③ Speed_of_impact
④ Helmet_Used
⑤ Seatbelt_Used
⑥ Survived

⑤ phone usage in India

problem Statement:
The dataset ~~pr~~ represents simulated phone usage data for Indian users, It contains 17,686 rows and 16 columns, Each row representing an individual's phone usage details. The data reflects various aspects of mobile phone behaviour including demographics, phone brand and app usage patterns

number of rows = 14687
number of columns = 16

columns

1. User ID
2. Age
3. Gender
4. Location
5. phone Brand
6. OS
7. Screen Time
8. Data Usage
9. calls Duration
10. Number of Apps install
11. Social Media Time
12. E-commerce Spend
13. Streaming Time
14. Gaming Time
15. Monthly Recharge
16. primary usage

3/3/2025