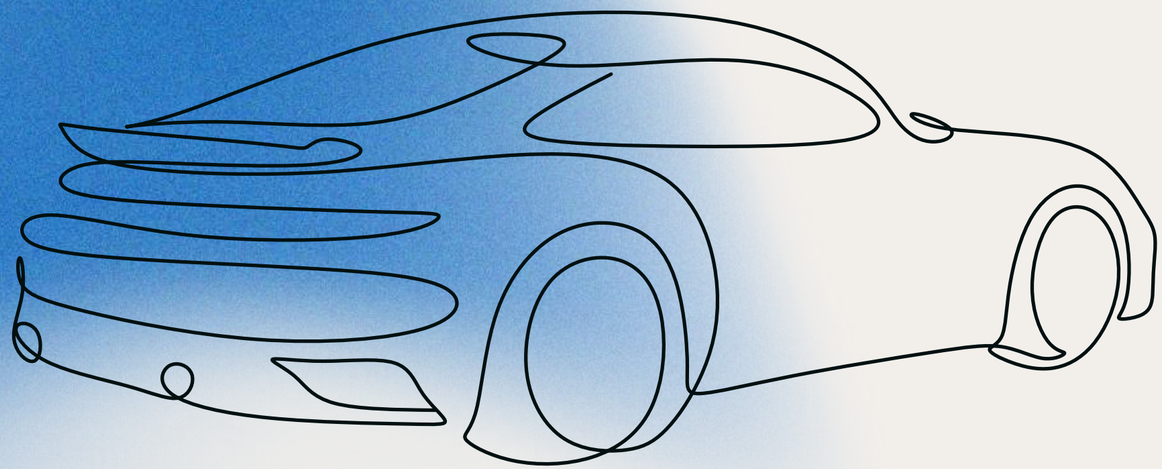
# Multimodal Car Analytics



PRESENTED BY: F20DL GROUP 12

# Objectives & Targets

**Regression (Tabular-only)**

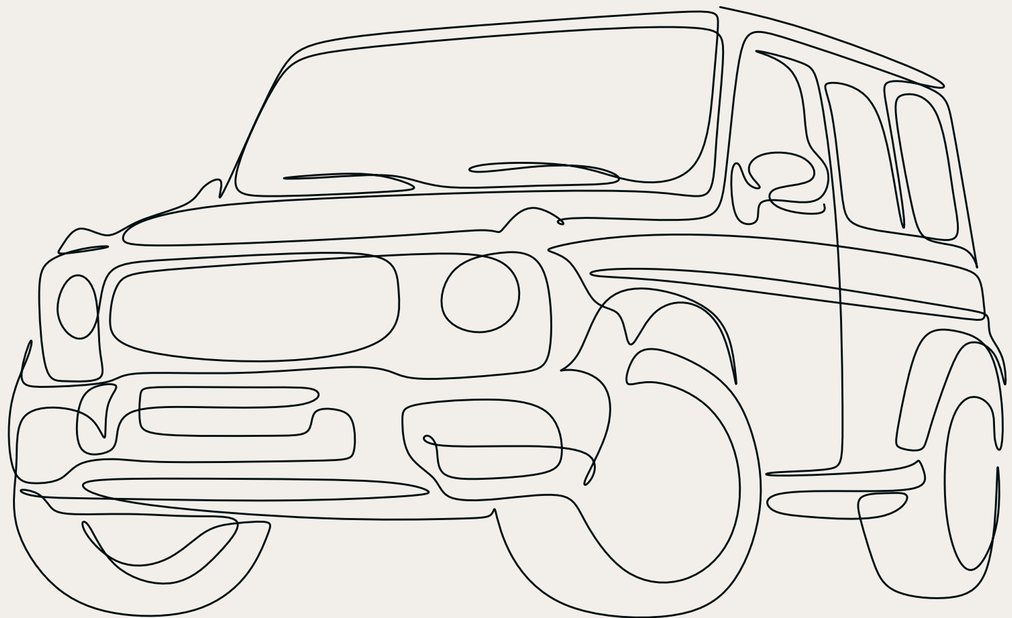
- Predict sale price from listing/spec features
- Predict alternative continuous target: % depreciation vs MSRP = (MSRP − Price) / MSRP

**Classification (Image-only)**

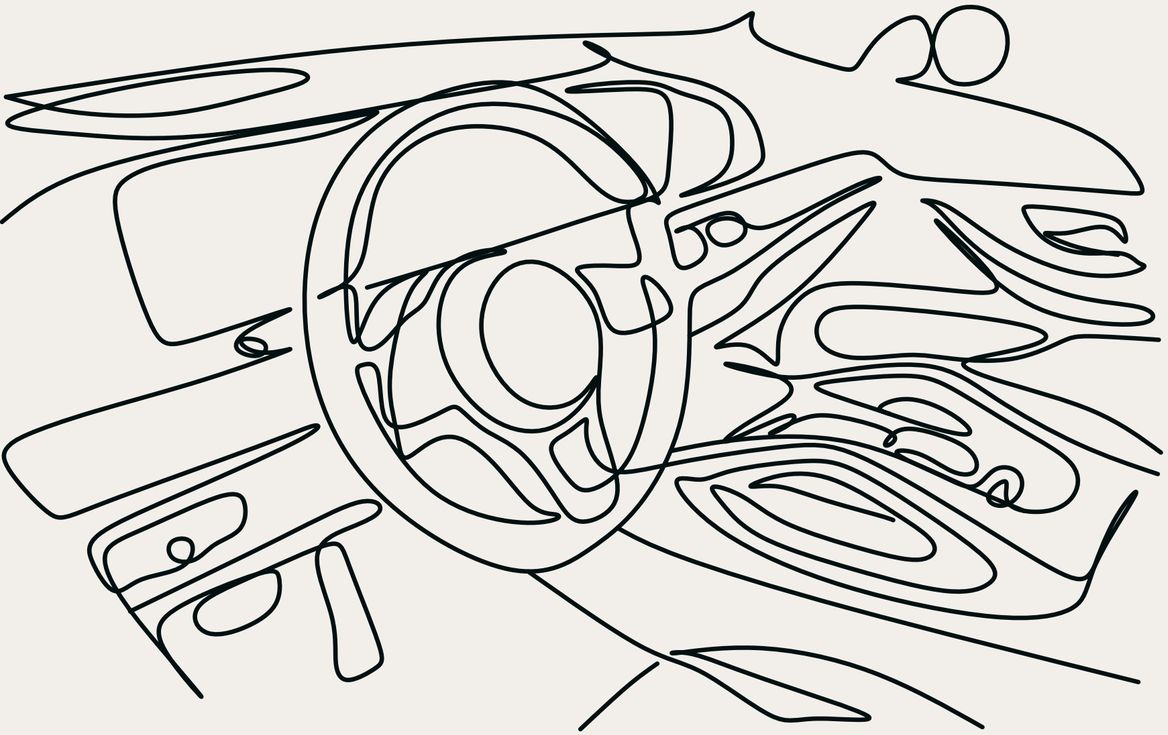
- Predict body type (e.g., sedan/SUV, etc.) and manufacturer from a single listing's images

**Deliverables alignment**

- These objectives set up analysis & modelling for subsequent coursework milestones



# Research Questions (RQ1 & RQ2)



| | |
|---|---|
| **RQ1 (Regression)** | To what extent can listing/features (age, mileage, trim, fuel, etc.) predict:<br>(a) Used Selling Price<br>(b) Percentage depreciation from MSRP (Manufacturer suggested retail price)? |
| **RQ2 (Classification)** | Can body type and manufacturer be accurately inferred from listing images of varying angles? |
| **Success indicators** | RQ1: Mean Absolute Error (MAE) & Error Breakdowns by make and price deciles<br>(↓ MAE = ↑ success)<br><br>RQ2: Accuracy & F1 scores<br>precission + recall<br>(↑ F1 = ↑ recognition) |

# Dataset: DVM-CAR

## What it includes

- 1.45M cleaned car images (background removed, front/side views)
- These images cover 899 car models in the UK market over past two decades.
- The image data is paired with non-visual tabular data, stored in six CSV tables
    1. Basic table (model name, model ID, brand)
    2. Sales table (annual sales over ~10 years)
    3. Price table (entry / new car prices over years)
    4. Trim table (trim-level attributes: engine, engine size, etc.)
    5. Ad table (used car advertisement data)
    6. Image table
- Some of the non-visual features include: brand / automaker, model, trim, engine size / type, fuel type, gear / transmission, sales, listing price, etc.

Dataset Citation (Source of Dataset):

Jingming Huang, Bowei Chen, Lan Luo, Shigang Yue, and Iadh Ounis. (2022). "DVM-CAR: A large-scale automotive dataset for visual marketing research and applications". In Proceedings of IEEE International Conference on Big Data, pp.4130–4137. [PDF_link]

# Dataset: DVM-CAR

## Why this dataset

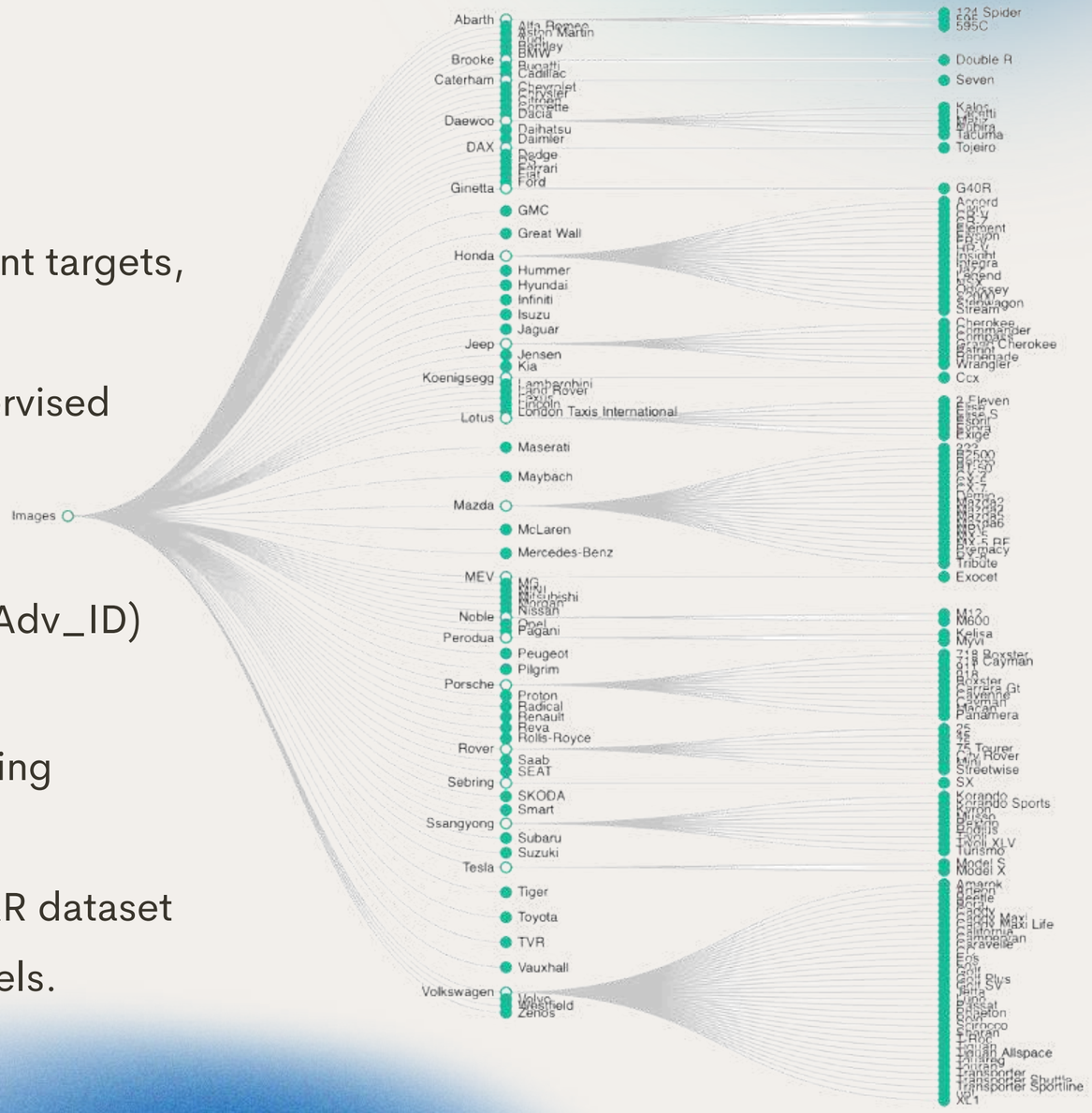
- The dataset has both images and tabular data and consistent targets, making it easy to join and use for the project.
- The dataset is scalable and very large and suitable for supervised learning.

## Join keys & structure

- Link images and tabular via listing ID (e.g.,GenModel_ID, Adv_ID)

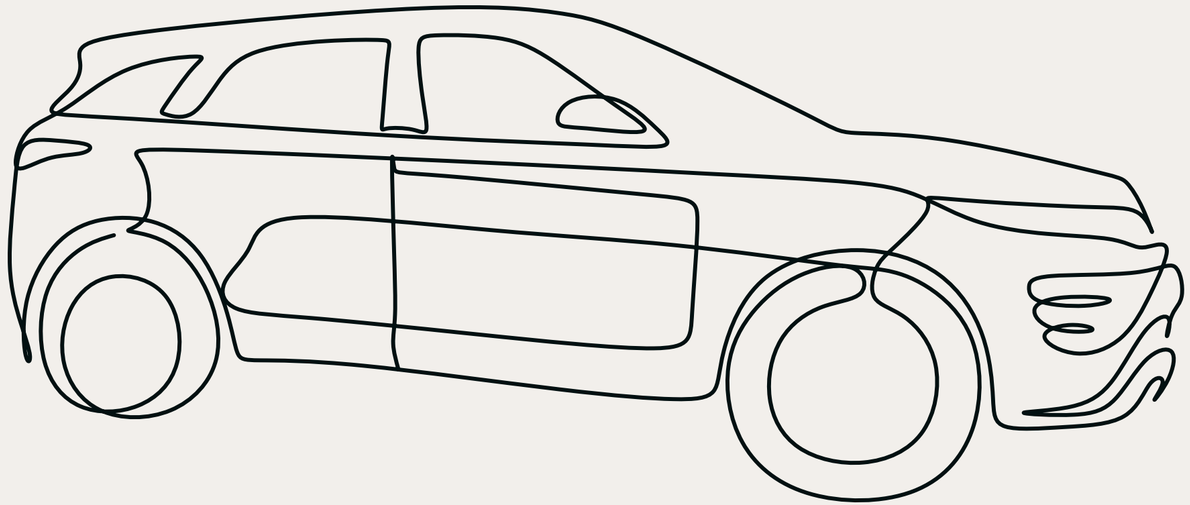
## How it's been used in the ML / research community

- In the original paper, the authors use it to demonstrate linking aesthetic design with sales forecasting.
- In more recent works on multimodal learning, the DVM-CAR dataset is used as a testbed for image and tabular multimodal models.



# Approach for RQ1

- Linear Regression to set a simple benchmark.
- Tree-based ensembles like Random Forest, XGBoost



# Approach for RQ2

- Transfer learning with with a pre-trained CNN.
- Softmax activation function is applied to output class probabilities corresponding to vehicle body types and manufacturers.

# Team Contributions:

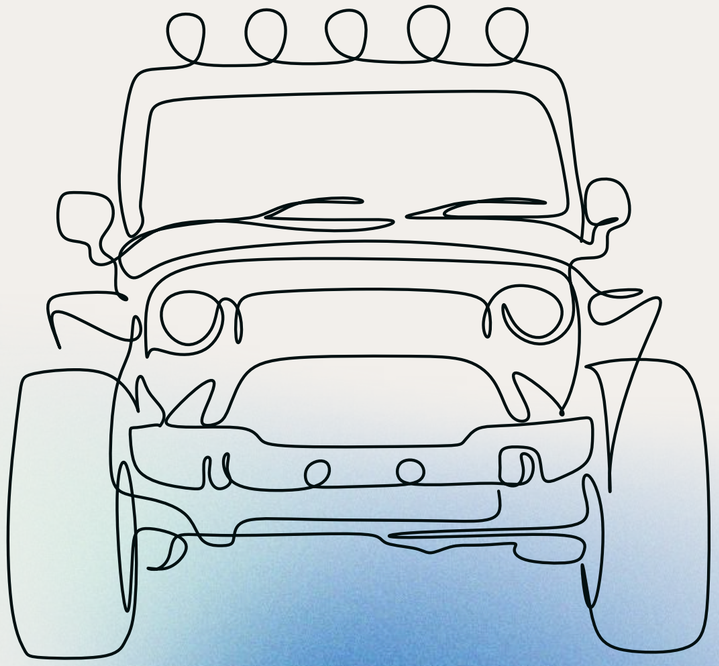
ADITYA:
JOINS TABLES, BUILDS REGRESSION MODELS FOR SELLING PRICE & DEPRECIATION, ENGINEERS KEY FEATURES, LEADS ERROR ANALYSIS

JOSEPH:
FILTERS & PREPROCESSES IMAGES, LINKS IMAGES TO ADVERTS, BUILDS CNN CLASSIFIERS, HANDLES AUGMENTATIONS & CLASSIFICATION METRICS

VIR:
STRUCTURES PIPELINES & EXPERIMENT ORCHESTRATION, ENSURES REPRODUCIBILITY, MANAGES COMPUTE & LOGGING, PRODUCES AGGREGATED RESULTS

LAIBA:
DEFINES ENCODING AND FEATURE STRATEGIES, RUNS ABLATION STUDIES, CONNECTS FINDINGS TO LITERATURE, DRAFTS METHODS & LITERATURE WRITE-UPS

NIBRAS: SETS UP PROJECT STRUCTURE & VERSION CONTROL, INTEGRATES CLASSIFICATION + REGRESSION CODEBASES, LEADS DOCUMENTATION, PREPARES FINAL DELIVERY AND SLIDES



# APPENDIX

**Literature Review:**

**Huang, J., Chen, B., Luo, L., Yue, S., & Ounis, I. (2022). "DVM-CAR: A large-scale automotive dataset for visual marketing research and applications." IEEE Big Data 2022.**

- The original dataset paper describes the DVM-CAR dataset itself, with ~1.45 million images and six relational tables (Basic, Sales, Price, Trim, Ad, and Image). They show example applications such as using CNNs (e.g., fine-tuned VGG/ResNet) to infer "design modernity scores" from images and correlate them with sales trends or model withdrawals.
- We build directly on their dataset, but whereas they only showcase illustrative uses, we will implement a rigorous supervised pipeline aimed at classification (body/manufacturer) and regression (price & depreciation) on identical adverts, with error breakdowns, leakage control, and a dual-task focus they don't carry out.

**Du, Siyi; Zheng, Shaoming; Wang, Yinsong; Bai, Wenjia; O'Regan, Declan P.; Qin, Chen (2024). "TIP: Tabular-Image Pre-training for Multimodal Classification with Incomplete Data." ECCV 2024.**

- They use 176,414 image–tabular pairs sampled from DVM to train a car model classification task with 283 classes. Their method (TIP) combines masked tabular reconstruction, image–tabular contrastive matching, and an interaction module, using CNN + tabular encoder architectures.
- Unlike TIP, which is a multimodal classification work, we separate tasks: image classification (RQ2) and tabular regression (RQ1). That decomposition simplifies attribution of performance and lets us study the independent predictive power of each modality. We will also tackle the regression of price and depreciation, which TIP does not.

**Yaman Abu Ghareebaih (2025). "Car Price Prediction Using Machine Learning: Analyzing the DVM-CAR Dataset." Master's Thesis, East Tennessee State University.**

- This thesis explicitly uses DVM-CAR to build and compare regression models that predict used car prices from features like mileage, engine power, registration year, etc. He performs cleaning and encoding and applies linear regression and random forest models, evaluating with MSE/MSEP metrics.
- His work aligns with our RQ1 regression focus, but he does not integrate image classification or depreciation prediction. We enhance his methodology by also modelling % depreciation, plus combining classification (RQ2) and regression (RQ1) in one unified project, with leakage control and error stratification.

**Roy, K., Krämer, L., et al. (2025). "Multi-Modal Contrastive Pre-training for Enhanced Tabular Data Analysis (MT-CMTM)." Preprint / arXiv.**

- They evaluate MT-CMTM on DVM. Their method fuses contrastive learning and masked tabular modelling with image signals, using a 1D-ResNet + attention tabular backbone plus contrastive alignment. They report a 2.38% absolute accuracy gain on DVM classification tasks over baseline modelling.
- Their work is advanced in multimodal fusion for classification, but they don't target regression of price or depreciation. Our plan differs by assessing classification and regression as separate tasks initially, before optional fusion—thus giving clearer insights and more direct alignment with price prediction goals.

**Du, S., Luo, X., O'Regan, D. P., & Qin, C. (2025). "STiL: Semi-supervised Tabular-Image Learning for Comprehensive Task-Relevant Information Exploration in Multimodal Classification." CVPR 2025.**

- STiL is a semi-supervised multimodal classification framework that is evaluated partly on DVM. It learns from both labelled and unlabelled image-tabular pairs using a disentangled contrastive consistency module, consensus-guided pseudo-labelling, and prototype-guided label smoothing. It uses CNN encoders for images and transformer/tabular encoders for tabular data.
- We don't plan to use semi-supervised or fusion models at first. Instead, we treat classification (RQ2) and regression (RQ1) as separate supervised tasks, which gives clearer attribution of performance to each modality. We also introduce percentage depreciation prediction, strict advert-level data splits, and error breakdowns by make/price decile, which STiL's classification focus does not cover.