Machine Learning Canvas

	PREDICTIONS	OBJECTIVES	DATA
IDEA	**Context** Who will use the predictive system / who will be affected by it? Provide some background. Automotive dealerships, financial institutions, and individual sellers who want accurate car price predictions based on market trends and vehicle characteristics. The predictive system aims to provide these users with reliable price estimates to make informed buying, selling, and lending decisions	**Value Proposition** What are we trying to do? E.g. spend less time on X, increase Y... The goal is to develop a predictive model that can accurately forecast car prices, thereby increasing pricing accuracy, enhancing customer satisfaction, reducing the time spent on manual price evaluation, and potentially increasing sales and revenue.	**Data Sources** Where do/can we get data from? (internal database, 3rd party API, etc.) The data will be sourced from the "Vehicle Sales and Market Trends Dataset" available on Kaggle, which includes comprehensive details about vehicle sales transactions, market trends, and various vehicle characteristics.
SPECS	**Problem** Question to predict answers to (in plain English) "What should be the selling price of a given vehicle Input (i.e. question "parameter") Vehicle details (make, model, year, trim, body type, transmission type), condition (condition rating, odometer reading), market data (MMR values), and sale information (state of registration, sale dates). Possible outputs (i.e. "answers") Predicted selling price of the vehicle Type of problem (e.g. classification, regression, recommendation...) Regression **Baseline** What is an alternative way of making predictions (e.g. manual rules based on feature values)? Manual evaluation by sales experts using historical sales data, market trends, and individual judgment.	**Performance evaluation** Domain-specific / bottom-line metrics for monitoring performance in production Increase in revenue from accurate pricing. Reduction in time spent on pricing evaluation. Improvement in customer satisfaction scores. Prediction accuracy metrics (e.g. MSE if regression; % accuracy, #FP for classification) MAE MSE R^2 Offline performance evaluation method (e.g. cross-validation or simple training/test split) Cross-validation and training/test split.	**Dataset** How do we collect data (inputs and outputs)? How many data points? Data is collected from the Kaggle dataset, which contains thousands of vehicle sales transactions with multiple features for each transaction. **Features** Used to represent inputs and extracted from data sources above. Group by types and mention key features if too many to list all. year - int64 body - object transmission - object vin - object condition - float odometr - float color - object sellingprice - target

DEPLOYMENT	
Using predictions	**Learning predictive models**
When do we make predictions and how many? Predictions are made whenever a new vehicle needs to be priced. The number of predictions will depend on the volume of vehicles being processed.	When do we create/update models? With which data / how much? Models are periodically updated, for example, when new data are available
What is the time constraint for making those predictions? Predictions should be made in real-time or within a few seconds to facilitate quick decision-making.	What is the time constraint for creating a model? Model creation should be completed within a few hours to a day, depending on the complexity and amount of data.
How do we use predictions and confidence values? Predictions will be used to set prices for vehicles. Confidence values can be used to determine the reliability of the predictions and to flag cases that may require manual review.	Criteria for deploying model (e.g. minimum performance value — absolute, relative to baseline or to previous model) The model should achieve a minimum performance threshold, such as an MSE below a certain value or an R^2 above a specific threshold, and it should outperform the baseline manual evaluation method.