Predictive Analytics for Car Prices Project

Project Overview

The goal of this project is to develop a predictive model to estimate car prices based on various features. This model will help in understanding the factors affecting car prices and provide actionable insights for decision-making.

Objectives

1. **Data Collection**: Gather and load the 2023 Cars Dataset.
2. **Data Cleaning and Pre-processing**: Prepare the data for analysis by addressing missing values and ensuring correct data types.
3. **Exploratory Data Analysis (EDA)**: Analyse data patterns and relationships.
4. **Feature Engineering**: Create relevant features for improved model accuracy.
5. **Model Building**: Develop and evaluate predictive models.
6. **Model Deployment**: Deploy the model and integrate it into a decision-support system.
7. **Visualization and Reporting**: Present insights through interactive dashboards.

Phase	Task	Status	Completion	Notes
Week 1: Data Collection and Preparation	Download and explore the dataset	Not Started	Done	Review dataset structure and attributes
	Load data into Python/SQL Server	Not Started	Done	Ensure data is properly loaded
	Inspect data for missing values and data types	Not Started	Done	Identify and document any data quality issues
Week 2: Data Cleaning and Pre-processing	Handle missing values and adjust data types	Not Started	Done	Use Python/SQL for data cleaning
	Normalize and standardize numerical features	Not Started	Done	Apply transformations as needed
	Develop additional features (e.g., car age)	Not Started	Done	Create and document new features
Week 3: Exploratory Data Analysis (EDA)	Perform univariate and bivariate analysis	Not Started	Done	Analyse data distributions and relationships
	Visualize distributions and relationships	Not Started	Done	Use Power BI or Python libraries for visuals
	Document key insights and findings	Not Started	Done	Summarize major discoveries
Week 4: Feature Engineering and Model Building	Create and select relevant features	Not Started	Done	Focus on features impacting car prices

Phase	Task	Status	Completion	Notes
	Split data into training and test sets	Not Started	Done	Prepare data for modelling
	Build and evaluate initial predictive models	Not Started	Done	Test various algorithms
Week 5: Model Refinement and Evaluation	Tune model parameters and enhance accuracy	Not Started	Done	Optimize model performance
	Validate model performance using metrics	Not Started	Done	Evaluate with MAE, RMSE, R^2
Week 6: Deployment and Integration	Prepare model for deployment	Not Started	Done	Ensure model is ready for production
	Integrate model with system	Not Started	Done	Test in a real-world or simulated environment
Week 7: Visualization and Reporting	Develop interactive dashboards in Power BI	Not Started		Create visualizations for insights
	Document analysis results and model performance	Not Started		Prepare final report and recommendations
	Prepare final report or presentation	Not Started		Summarize the project and insights

Tools and Technologies

- **Python**: For data cleaning, pre-processing, and model building.
- **SQL Server**: For data storage and querying.
- **Power BI**: For interactive visualizations and dashboards.

- **Jupyter Notebook/IDE**: For coding and documentation.

Deliverables

1. **Data Preparation Scripts**: Python or SQL scripts for data cleaning and pre-processing.
2. **EDA Report**: Insights and visualizations from exploratory analysis.
3. **Predictive Model**: Trained and evaluated predictive model.
4. **Deployment Package**: Model integrated into a production environment.
5. **Visualization Dashboard**: Interactive dashboards in Power BI.
6. **Final Report**: Comprehensive project documentation and results.

Success Criteria

- **Model Accuracy**: Achieve high performance metrics.
- **Effective Deployment**: Successful integration and functionality in production.
- **Insightful Visualizations**: Clear and actionable dashboards.

Risks and Mitigations

- **Data Quality Issues**: Address through thorough cleaning and validation.
- **Model Performance**: Continuously refine and validate model performance.
- **Deployment Challenges**: Extensive testing to ensure smooth integration.

Initial Report

Final Comprehensive Report