## **Case Study: Vehicle Advertisements Analysis**

### **Background**

You are provided with a dataset containing vehicle advertisements. Your task is to analyze this data to address three key areas: Price Prediction, User Segmentation, and Geographical Analysis of Demand and Supply. Each section of this case study will guide you through specific tasks and questions to help you build models and draw insights from the data.

### **Dataset Description**

The dataset contains the following columns:

* ad\_title: Title of the advertisement
* ad\_description: Description of the advertisement
* details: Additional details about the vehicle
* slug: Original URL of the advertisement
* title: Long title of the advertisement
* type: Type of advertisement (e.g., for\_sale)
* price: Price of the vehicle
* timestamp: Timestamp of the posting
* posted\_date: Date when the ad was posted
* deactivation\_date: Date when the ad was deactivated
* category: Category of the vehicle
* parent\_category: Parent category of the vehicle
* location: Location of the vehicle
* geo\_region: Geographical region of the vehicle
* area: Specific area of the vehicle
* is\_delivery\_free: Whether delivery is free
* is\_doorstep\_delivery: Whether doorstep delivery is available
* is\_dsd\_applicable: Whether DSD is applicable
* is\_member: Whether the advertiser is a member
* is\_authorized\_dealer: Whether the advertiser is an authorized dealer
* is\_featured\_member: Whether the advertiser is a featured member
* is\_verified: Whether the advertiser is verified
* membership\_level: Membership level of the advertiser
* member\_since: Date since the advertiser is a member
* properties: Additional properties of the vehicle
* user: User ID of the advertiser

### **Section 1: Price Prediction**

**Problem Statement:** Develop a predictive model to estimate the price of a vehicle based on the provided features.

**Tasks:**

1. **Data Cleaning & Preprocessing:**
   * Handle missing values appropriately.
   * Encode categorical variables.
   * Normalize/standardize numerical features if necessary.
2. **Feature Selection:**
   * Perform correlation analysis to identify relevant features.
   * Apply feature selection techniques such as Recursive Feature Elimination (RFE) or Lasso Regression.
3. **Model Training:**
   * Train at least three different regression models
   * Evaluate the models using cross-validation.
4. **Hyperparameter Tuning:**
   * Use Grid Search or Random Search for hyperparameter optimization.
5. **Model Evaluation:**
   * Compare the models based on RMSE, MAE, and R².
   * Select the best-performing model.

**Deliverables:**

* A detailed report on data preprocessing and feature selection.
* Performance metrics of the trained models.
* A final price prediction model with tuned hyperparameters.
* A brief discussion on the model's strengths and weaknesses.

### **Section 2: User Segmentation**

**Problem Statement:** Segment users based on their advertisement behaviors and characteristics.

**Tasks:**

1. **Data Preprocessing:**
   * Handle missing values.
   * Encode categorical variables.
2. **Feature Selection:**
   * Use PCA to reduce dimensionality if necessary.
   * Select relevant features for clustering.
3. **Clustering:**
   * Apply K-means clustering to segment users.
   * Experiment with different numbers of clusters and evaluate using the Elbow Method and Silhouette Score.
4. **Cluster Analysis:**
   * Interpret and describe the characteristics of each cluster.
   * Identify any patterns or insights about user behavior.

**Deliverables:**

* A detailed report on data preprocessing and feature selection.
* Visualization of the clusters and their characteristics.
* A description of each user segment and potential business implications.

### **Section 3: Geographical Analysis of Demand and Supply**

**Problem Statement:** Analyze the geographical distribution of vehicle advertisements to understand demand and supply patterns.

**Tasks:**

1. **Data Preprocessing:**
   * Handle missing values.
   * Encode geographical variables.
2. **Geographical Analysis:**
   * Visualize the distribution of advertisements across different regions.
   * Identify regions with high demand and supply.
3. **Temporal Analysis:**
   * Analyze how demand and supply vary over time in different regions.
   * Identify any seasonal trends or patterns.
4. **Modeling Demand and Supply:**
   * Develop a regression model to predict demand and supply based on geographical and temporal features.
   * Evaluate the model using appropriate metrics.

**Deliverables:**

* Visualizations of geographical and temporal trends in demand and supply.
* A detailed report on the findings from the geographical analysis.
* Performance metrics of the demand and supply prediction model.
* Insights and recommendations based on the analysis.

### **Submission Guidelines**

* Submit a well-documented Jupyter notebook, your code and analysis.
* Provide a comprehensive report summarizing your findings and insights for each section.
* Ensure that all visualizations are clearly labeled and interpreted.