

# Egyptian Used Cars Full Project (From Data Collection to Price Prediction)

## 1. Project Overview

The Context:

The Egyptian car market changed drastically after the 2022 decision to restrict car imports. This decision caused a shortage of new cars, forcing most buyers into the used car market. As a result, prices skyrocketed, and the market became chaotic and hard to understand.

The Goal:

To build a complete system that collects data from the market, analyzes it, and creates a Machine Learning model to help users price their cars accurately based on real market data.

The Value:

This project provides a clear understanding of the Egyptian market, solving the problem of unclear pricing and helping users make data-driven decisions.

## 2. Workflow

The project follows four main steps:

1. **Web Scraping:** Collecting data from the two most important used car websites in Egypt: [Hatla2ee](#) and [ContactCars](#).
2. **Data Cleaning:** Cleaning and formatting the data from both sources to merge them into a single, unified dataset.
3. **EDA & Visualization:** Extracting insights and visualizing market trends using **Tableau**.
4. **Machine Learning:** Building an ML model to predict car prices based on the collected data.

## 3. Detailed Phases

### Phase 1: Web Scraping

To ensure data variety, we scraped data from multiple sources:

- **Over 8,000 rows** from Hatla2ee.
- **Over 3,000 rows** from ContactCars.

### Phase 2: Data Cleaning

The focus was on unifying the format to merge the two datasets successfully. Key steps included:

- **Removing Duplicates:** Deleting repeated ads to ensure accuracy.
- **Standardizing Prices:** Converting text formats (e.g., "1.5 Million") into numeric values.

- **Handling Outliers:** Removing unrealistic values in prices or kilometers.

### Phase 3: Tableau Dashboards

We analyzed the data and visualized the insights to show:

- **Best-selling cars** in Egypt.
- **Price distribution** across different Governorates.
- **Depreciation rates** (Drop in price over time).

### Phase 4: Machine Learning Model

We built a model to predict prices based on inputs like **Make, Model, Transmission, and Kilometers**.

- **Preprocessing:** Applied feature scaling, encoding, and feature engineering. Dropped unnecessary columns.
- **Algorithms:** Used **XGBoost** and **Random Forest Regressor**.
- **Evaluation:** The model achieved high accuracy, evaluated using **R2 Score, MAE, and RMSE**.

## 4. Results & Market Insights

Through our analysis, we discovered significant facts about the market post-2022:

- **Market Share: Mercedes-Benz** holds approximately **10%** of the used car market share.
- **Electric Vehicles (EVs):** Represent **less than 1%** of the market, indicating the difficulty of importing modern EV technology.
- **Price Inflation:** Used car prices are now nearly equal to "Zero" (New) car prices due to the import shortage and high demand.
- **Sales Drop (2022 vs. 2021):**
  - **BMW** sales dropped by over **70%**.
  - **Chevrolet** sales dropped by about **30%**.

Economic Impact of the Import Restriction:

While the decision affected the market, it had positive economic outcomes:

- **Currency Protection:** Saved the country **\$4-5 Billion annually**, securing foreign currency for essential goods.
- **Local Manufacturing:** Forced global companies to rely on local assembly instead of easy imports.
- **Regulation:** Closed customs loopholes (such as the misuse of disability car exemptions) to better control state resources.

## 5. Technologies Used

- **Python** (BeautifulSoup, Selenium, Pandas, NumPy)
- **Visualization:** Tableau Public
- **Machine Learning:** Scikit-Learn, XGBoost

## 6. Team Members

- Abdelrhman Mohamed Yakout (leader)
- Ziad Oun
- Ahmed Mohmed Hussien
- Abdelrhman Islam
- Ahmed Refaay
- Ziad Elgendy