

Title: Toyota Used Car Price Prediction Project (SWU Motors)

1. Problem Statement/Background



- What do we know about?
- What problem are you trying to solve?
 - SWU Motors เป็นตัวเลือกมือสองที่กำลังขยายตัวและจ้างพนักงานขายใหม่ (Junior Salespeople) จำนวนมาก
 - ยอดขายลดลง 18% ในช่วงที่ผ่านมา
 - พนักงานขายใหม่ขาดประสบการณ์ในการตั้งราคา (Pricing) รถ Toyota มือสองที่รับเข้ามา ทำให้เกิดความผิดพลาดในการตั้งราคายา
 - เรายัง dataset (toyota.csv) ที่รวมราคายาจากผู้ค้าปลีกรายอื่นในตลาด
- What is the business problem?
 - Revenue Loss: การตั้งราคาผิดพลาดทำให้เสียโอกาสในการขาย (ถ้าราคาสูงเกินไปขายไม่ออก) หรือเสียกำไร (ถ้าราคาต่ำเกินไป)
- Who are the stakeholders?
 - End-Users: พนักงานขายใหม่ (Junior Salespeople)
 - Management: ผู้จัดการฝ่ายขาย (ต้องการกู้ยอดขาย)
 - Data Science, Data Analytics

2. SMART Objectives/ Value Propositions



SMART Objectives:

- พัฒนาโมเดลนายาราคาให้แล้วเสร็จภายใน 2 สัปดาห์ โดยมีค่าความคลาดเคลื่อน (RMSE) ต่ำกว่า 1,250 ปอนด์ เพื่อช่วยให้พนักงานใหม่ตั้งราคาได้อย่างถูกต้องและกู้คืนยอดขายที่ลดลง 18%

Value Proposition:

- สร้างเครื่องมือ "Pricing Tool" เพื่อช่วยให้พนักงานใหม่ตัดสินใจได้เร็วขึ้น แม่นยำขึ้น และกู้คืนยอดขายกลับคืนมา

3. Questions/Hypothesis



Analytical Questions

- ปัจจัยใดส่งผลต่อราคา Toyota มากที่สุด? (ปีจดทะเบียน?, เลขไมล์?, ขนาดเครื่องยนต์?)
- ลักษณะการกระจายตัวของราคาในตลาดเป็นอย่างไร?

Predictive Hypothesis

- H1: เลขไมล์ (Mileage) น่าจะมีความสัมพันธ์เชิงลบกับราคา (ยิ่งวิ่งเยอะ ราคาถูก)
- H2: ขนาดเครื่องยนต์ (Engine Size) น่าจะมีความสัมพันธ์เชิงบวกกับราคา (เครื่องใหญ่ = ราคาถูกยิ่ง)

What can we predict?

- ราคายาต่อชิ้นของรถยนต์ Toyota (£)

4. Data Sources/Attributes



Data sources & collection

- ไฟล์ toyota.csv (รวบรวมข้อมูลราคาและคุณลักษณะรถจากผู้ค้าปลีกภายนอก)

Data cleaning & preprocessing

- ตรวจสอบ Missing Values และ Duplicate Rows
- ตรวจสอบ Outliers (เช่น รถที่มีค่าภาษี tax เป็น 0 หรือค่า mpg ผิดปกติ)

Target variables & feature

- **Target:** price.
- **Features:** model, year, transmission, mileage, fuelType, tax, mpg, engineSize.

Encoding & scaling strategies

- **Encoding:** ใช้ One-Hot Encoding แบ่งข้อมูลกลุ่ม (model, transmission, fuelType) เป็นตัวเลข
- **Scaling:** ใช้ StandardScaler เพื่อปรับมาตรฐานข้อมูลตัวเลข (mileage, tax, mpg)

5. Analysis/Model Development

Analytics Methodology

6. Findings and Insights



Business Insights

- Segmentation: "ขนาดเครื่องยนต์" คือปัจจัยหลักในการแบ่งเกรดและราคารถ (ชัดเจนกว่าปีผลิต)
- Market Structure: ตลาดเป็นแบบ "เมี้ยงข้าว" (Right Skewed) รถส่วนใหญ่ราคาประ Eh แต่มีรถหรูจำนวนน้อยตึงค่าเฉลี่ยขึ้น
- Value Retention: รถเครื่องยนต์ดีเซลรักษาภูมิคุ้มได้ดีกว่าและเสื่อมราคชาากว่ารถเบนซินเมื่อใช้งานหนัก

Predictive Results

- โมเดล **Linear Regression (Lasso)** ให้ค่า RMSE เท่ากับ £1,680 ซึ่งไม่ผ่านเกณฑ์ที่กำหนด สาเหตุหลักมาจากการขาดแคลนข้อมูลเชิงเส้นในการอธิบายความสัมพันธ์ที่ซับซ้อนของข้อมูล
- **Performance of Random Forest Model** โมเดล Random Forest แสดงผลลัพธ์ที่ดีกว่าอย่างชัดเจน โดยมีค่า RMSE เท่ากับ £1,102.96 ซึ่งต่ำกว่าเกณฑ์เป้าหมาย (£1,250)
- โมเดล Random Forest มีค่าความแม่นยำสูง ($R^2 = 97\%$) และมีความคลาดเคลื่อนและ MAE ประมาณ ± £750 ต่อคัน แสดงให้เห็นถึงศักยภาพในการนำไปใช้งานจริงในการตั้งราคารถยนต์

7. Recommendation/Action and Impact



Action:

- Deploy Tool: ติดตั้งเครื่องมือ "Pricing Tool" ให้ทีมขายใช้เป็น "ราคากลาง" หน้างานทันที

Impact:

- Sales Recovery: แก้ปัญหาการตั้งราคายาตัวเล็กๆ ที่ตกลง 15% ให้กลับมาเติบโต
- Profit Optimization: ลดโอกาสสูญเสีย (ขายถูกไป) และลดความเสี่ยงในการรับรู้เงินสด
- Standardization: สร้างมาตรฐานราคากลาง (Fair Price) ให้กับองค์กร ลดการพึงพาประสบการณ์ส่วนบุคคล (Gut Feeling) ของพนักงาน