

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

<div><div>1. Problem Statement/Background</div><div><div><div></div></div><ul style="list-style-type: none">What do we know about?What problem are you trying to solve?<ul style="list-style-type: none">SWU Motors เป็นดีลเลอร์รถมือสองที่กำลังขยายตัวและจ้างพนักงานขายใหม่ (Junior Salespeople) จำนวนมากยอดขายลดลง 18% ในช่วงที่ผ่านมาพนักงานขายใหม่ขาดประสบการณ์ในการตั้งราคา (Pricing) รถ Toyota มือสองที่รับเข้ามา ทำให้เกิดความผิดพลาดในการตั้งราคาขายเรามี dataset (toyota.csv) ที่รวบรวมราคาขายจริงจากผู้ค้าปลีกรายอื่นในตลาดWhat is the business problem?<ul style="list-style-type: none">Revenue Loss: การตั้งราคาผิดพลาดทำให้เสียโอกาสในการขาย (ถ้าราคาสูงเกินไปรถขายไม่ออก) หรือเสียกำไร (ถ้าราคาต่ำเกินไป)Who are the stakeholders?<ul style="list-style-type: none">End-Users: พนักงานขายใหม่ (Junior Salespeople)Management: ผู้จัดการฝ่ายขาย (ต้องการยอดขาย)Data Science, Data Analytics</div></div>	<div><div>2. SMART Objectives/ Value Propositions</div><div><div><div></div></div><div>SMART Objectives:<ul style="list-style-type: none">พัฒนาโมเดลทำนายราคาให้แล้วเสร็จภายใน 2 สัปดาห์ โดยมีค่าความคลาดเคลื่อน (RMSE) ต่ำกว่า 1,250 ปอนด์ เพื่อช่วยให้พนักงานใหม่ตั้งราคาได้อย่างถูกต้องและกู้คืนยอดขายที่ลดลง 18%</div><div>Value Proposition:<ul style="list-style-type: none">สร้างเครื่องมือ "Pricing Tool" เพื่อช่วยให้พนักงานใหม่ตัดสินใจได้เร็วขึ้นแม่นยำขึ้น และกู้ยอดขายกลับคืนมา</div></div></div>	<div><div>3. Questions/Hypothesis</div><div><div><div></div></div><ul style="list-style-type: none">Analytical Questions<ul style="list-style-type: none">ปัจจัยใดส่งผลต่อราคา Toyota มากที่สุด? (ปัจจัยทะเบียน?, เลขไมล์?, ขนาดเครื่องยนต์?)ลักษณะการกระจายตัวของราคาในตลาดเป็นอย่างไร?Predictive Hypothesis<ul style="list-style-type: none">H1: เลขไมล์ (Mileage) น่าจะมีความสัมพันธ์เชิงลบกับราคา (ยิ่งวิ่งเยอะ ราคายิ่งตก)H2: ขนาดเครื่องยนต์ (Engine Size) น่าจะมีความสัมพันธ์เชิงบวกกับราคา (เครื่องใหญ่ = ราคาแพง)<div>What can we predict?<ul style="list-style-type: none">ราคาขายต่อของรถยนต์ Toyota (£)</div></div></div>	<div><div>4. Data Sources/Attributes</div><div><div><div></div></div><ul style="list-style-type: none">Data sources & collection<ul style="list-style-type: none">ไฟล์ toyota.csv (รวบรวมข้อมูลราคาและคุณลักษณะรถจากผู้ค้าปลีกภายนอก)Data cleaning & preprocessing<ul style="list-style-type: none">ตรวจสอบ Missing Values และ Duplicate Rowsตรวจสอบ Outliers (เช่น รถที่มีค่าภาษี tax เป็น 0 หรือค่า mpg ผิดปกติ)Target variables & feature<ul style="list-style-type: none">Target: price.Features: model, year, transmission, mileage, fuelType, tax, mpg, engineSize.Encoding & scaling strategies<ul style="list-style-type: none">Encoding: ใช้ One-Hot Encoding แปลงข้อมูลกลุ่ม (model, transmission, fuelType) เป็นตัวเลขScaling: ใช้ StandardScaler เพื่อปรับมาตรฐานข้อมูลตัวเลข (mileage, tax, mpg)</div></div>
<div><div>5. Analysis/Model Development</div><div><div><div></div></div><ul style="list-style-type: none">Analytics Methodology<ul style="list-style-type: none">Descriptive Analytics: ใช้สถิติพื้นฐาน (Mean, Median, Skewness) และ Visualization เพื่อเข้าใจโครงสร้างตลาดPredictive Analytics: ใช้ Supervised Learning (Regression) สร้างโมเดลทำนายราคาขายModeling Methodology<ul style="list-style-type: none">Pipeline: ใช้ OneHotEncoder (แปลงกลุ่มข้อมูล) และ StandardScaler (ปรับสเกลตัวเลข) เพื่อเตรียมข้อมูลอัตโนมัติModel Selection: เปรียบเทียบ 2 วิธีการ:<ul style="list-style-type: none">Linear Models (Baseline): ทดสอบ Ridge, Lasso, Elastic Net (สมมติฐานเส้นตรง)Random Forest (Challenger): ใช้ Ensemble Learning เพื่อจับรูปแบบ Non-linear ที่ซับซ้อนEvaluation: จูนพารามิเตอร์ด้วย GridSearchCV และวัดผลด้วย 5-Fold Cross-Validation เพื่อความแม่นยำสูงสุด</div></div>	<div><div>6. Findings and Insights</div><div><div><div></div></div><ul style="list-style-type: none">Business Insights<ul style="list-style-type: none">Segmentation: "ขนาดเครื่องยนต์" คือปัจจัยหลักในการแบ่งเกรดและราคารถ (ชัดเจนกว่าปีผลิต)Market Structure: ตลาดเป็นแบบ "เบ้ขวา" (Right Skewed) รถส่วนใหญ่ราคาประหยัด แต่มีรถหรูจำนวนน้อยดึงค่าเฉลี่ยขึ้นValue Retention: รถเครื่องยนต์ดีเซลรักษามูลค่าได้ดีกว่าและเสื่อมราคาช้ากว่ารถเบนซินเมื่อใช้งานหนักPredictive Results<ul style="list-style-type: none">โมเดล Linear Regression (Lasso) ให้ค่า RMSE เท่ากับ £1,680 ซึ่งไม่ผ่านเกณฑ์ที่กำหนดสาเหตุหลักมาจากข้อจำกัดของโมเดลเชิงเส้นในการอธิบายความสัมพันธ์ที่ซับซ้อนของข้อมูลPerformance of Random Forest Model<ul style="list-style-type: none">โมเดล Random Forest แสดงผลลัพธ์ที่ดีกว่าอย่างชัดเจน โดยมีค่า RMSE เท่ากับ £1,102.96 ซึ่งต่ำกว่าเกณฑ์เป้าหมาย (£1,250)โมเดล Random Forest มีค่าความแม่นยำสูง ($R^2 = 97\%$) และมีความคลาดเคลื่อนเฉลี่ย (MAE) ประมาณ $\pm \text{£}750$ ต่อคันแสดงให้เห็นถึงศักยภาพในการนำไปใช้งานจริงในการตั้งราคารถยนต์</div></div>	<div><div>7. Recommendation/Action and Impact</div><div><div><div></div></div><div>Action:<ul style="list-style-type: none">Deploy Tool: ติดตั้งเครื่องมือ "Pricing Tool" ให้ทีมขายใช้เป็น "ราคากลาง" หน่วยงานทันที</div><div>Impact:<ul style="list-style-type: none">Sales Recovery: แก้ปัญหาการตั้งราคาผิดพลาดช่วยกู้คืนยอดขายที่ตกลง 15% ให้กลับมาเติบโตProfit Optimization: ลดโอกาสการ (ขายถูกไป) และลดความเสี่ยงในการ (รับซื้อแพงไป)Standardization: สร้างมาตรฐานราคากลาง (Fair Price) ให้กับองค์กร ลดการพึ่งพาประสบการณ์ส่วนบุคคล (Gut Feeling) ของพนักงาน</div></div></div>	