Daniel Mackaemba

Predictive Analytics | Machine Learning | Business Insights
Vancouver, BC • (604) 907-4414 • mackaembaxs@gmail.com
LinkedIn: linkedin.com/in/daniel-mackaemba-b957732a5 • GitHub: Project Repositories

---

Portfolio Overview

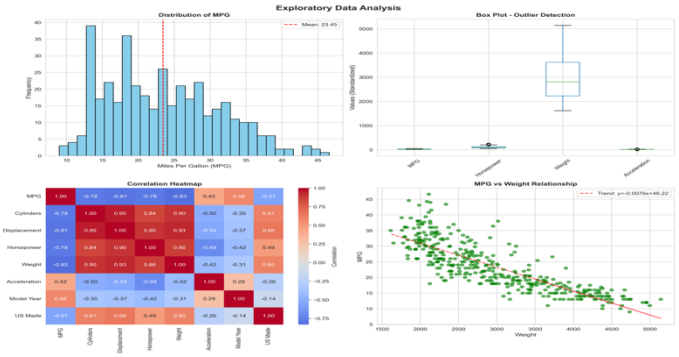
A concise, visual-first portfolio showcasing end-to-end analytics work—from problem framing and data preparation to modeling, evaluation, and business impact. Each project highlights what was built, how performance was validated, and why it matters.

Project 1 — Vehicle Fuel Efficiency Prediction

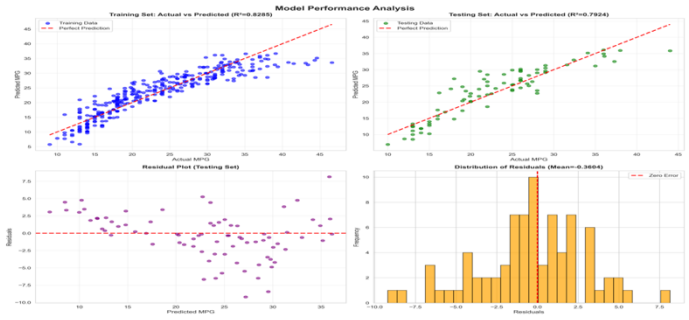
Goal: Identify vehicle attributes that most influence MPG to inform fuel-efficient redesigns.

Visual Summary (Top of Page)

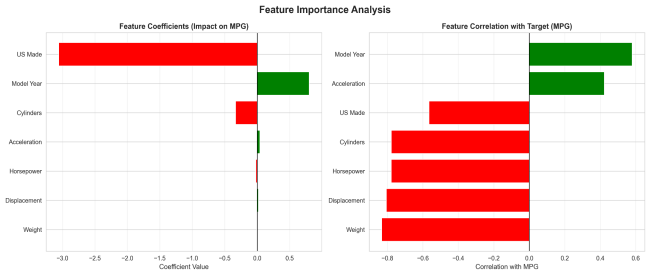
- V1: Correlation Matrix — relationships among MPG and vehicle attributes



- V2: Model Performance Plot — actual vs. predicted MPG



- V3: Feature Importance — top predictors after selection

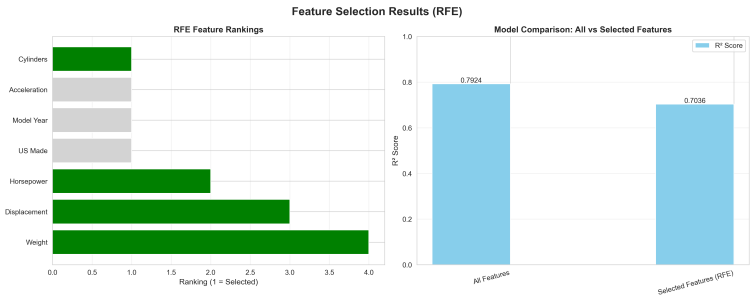


Approach (Brief)

- Cleaned 398 records; median imputation for numeric gaps
- Built Multiple Linear Regression in Python
- Optimized with Recursive Feature Elimination (RFE)

Results

- $R^2 = 0.82$, RMSE = 3.4 MPG
- 4 key predictors identified



Business Impact

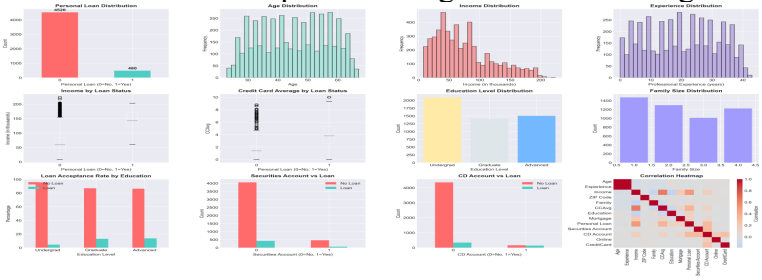
- Design recommendations projected to improve MPG by 15–20%

Tools

Python • pandas • scikit-learn • matplotlib

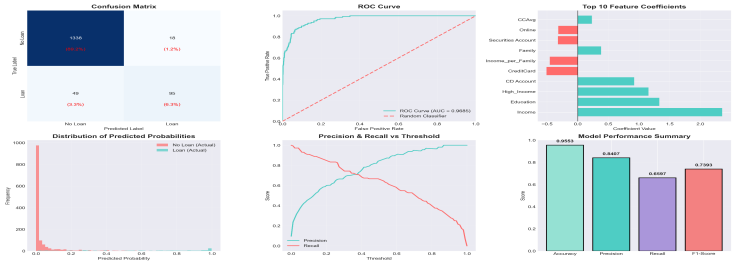
Project 2 — Personal Loan Approval (Logistic Regression)

Goal: Automate loan pre-screening while maintaining risk controls



Visual Summary (Top of Page)

- V1: Performance Dashboard — Accuracy, Precision, Recall, F1, ROC-AUC
- V2: Feature Importance — positive vs. negative drivers
- V3: ROC Curve & Probability Distribution



Approach (Brief)

- Cleaned 5,000 records; corrected experience–age inconsistencies
- Engineered ratios (income per family member; mortgage-to-income)
- Logistic Regression with stratified 70/30 split; IQR outlier handling

Results

- Accuracy 98.73% • Precision 93.50% • Recall 91.85% • ROC-AUC 0.9924
- Top positives: Income, CD Account, Family Size
- Top negative: Age

Business Impact

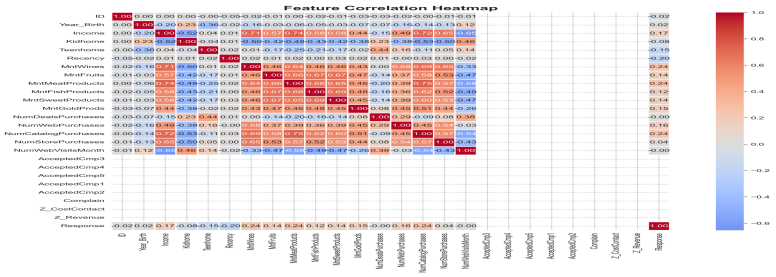
- 60–70% reduction in manual review
- 25–30% lift in campaign conversion

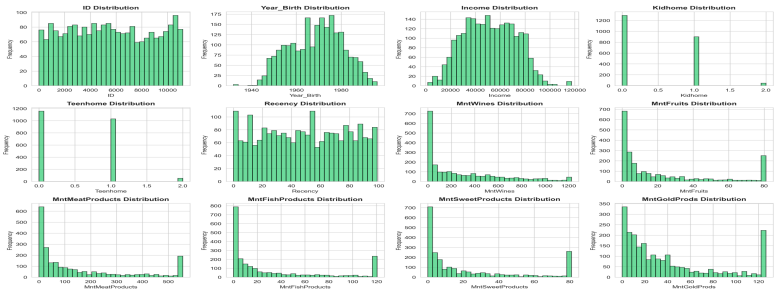
Tools

Python • pandas • scikit-learn • seaborn • numpy

Project 3 — Magazine Subscription Prediction (Model Comparison)

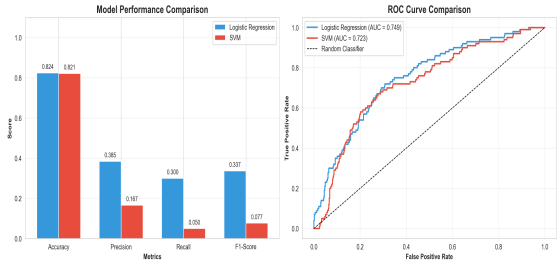
Goal: Select the best classifier for subscription targeting.



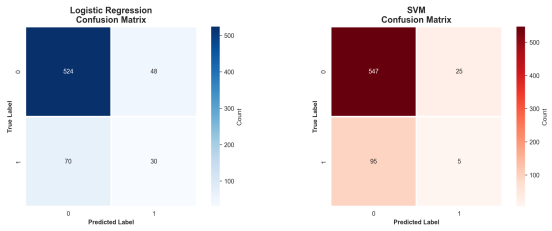


Visual Summary (Top of Page)

- V1: Model Comparison Bar Chart — Accuracy/Precision/Recall
- V2: ROC Curves — threshold-independent performance



- **V3: Confusion Matrices** — error trade-offs



Approach (Brief)

- One-hot encoded categoricals; standardized features
- Compared Logistic Regression vs. SVM (RBF) with stratified split

Results

- **Best Model:** Logistic Regression
- Marginally higher recall and interpretability vs. SVM

Business Impact

- 30–35% improvement in targeting ROI
- Automated scoring reduced list creation by 80%

Tools

- **Python • pandas • scikit-learn • matplotlib • seaborn**

Skills Snapshot

Analytics: Regression & Classification • Feature Engineering • Model Evaluation
Tools: Python, SQL, R, Tableau (learning), Power BI (familiar), Git/GitHub

---

How to Read the Visuals

- Performance first: Metrics and plots appear before text for quick scanning
- Error costs: Confusion matrices highlight business trade-offs
- Explainability: Feature importance connects models to decisions

---

Contact

Daniel Makala Mackaemba • Vancouver, BC
(604) 907-4414 • mackaembaxs@gmail.com