



E-COMMERCE SHOPPERS BEHAVIOR ANALYSIS

DATA MINING FOR CUSTOMER INSIGHTS & REVENUE OPTIMIZATION

Ntwali Bruno Bahongere

Advanced Big Data and Data Mining

University of the Cumberland

INTRODUCTION & RESEARCH PURPOSE



E-commerce generates vast behavioral data.



Project goal: Extract actionable insights to improve revenue and customer experience.



Research Questions:

What drives purchases?

How can we segment customers?

Which patterns predict revenue?

DATASET OVERVIEW

Source: Kaggle Shoppers Behavior and Revenue Dataset

Size: 12,330 sessions × 18 features (10 months)

Features:

- 14 numerical (page visits, durations, rates, values)
- 4 categorical (month, visitor type, weekend, revenue)

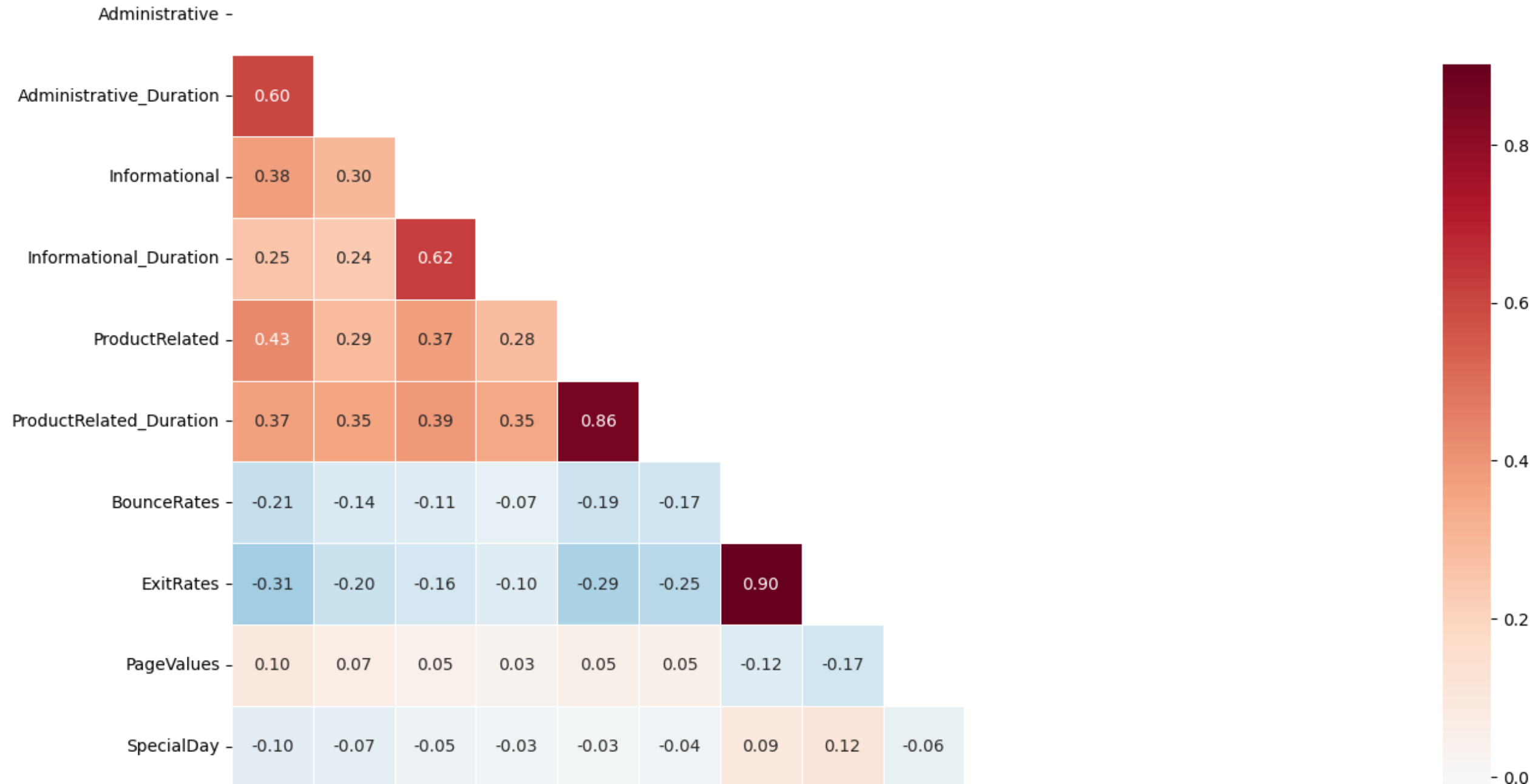
DATA PREPARATION & QUALITY

Actions:

- Checked for missing values (none found)
- Removed 1% duplicates
- Outlier analysis (retained for behavioral insight)

```
Administrative      0
Administrative_Duration  0
Informational       0
Informational_Duration  0
ProductRelated     0
ProductRelated_Duration  0
BounceRates        0
ExitRates          0
PageValues         0
SpecialDay         0
Month              0
OperatingSystems   0
Browser            0
Region             0
TrafficType        0
VisitorType        0
Weekend            0
Revenue            0
dtype: int64
```

Correlation Matrix of Numerical Features





FEATURE ENGINEERING

Created 32 new features:

- Engagement metrics (total pages, duration, score)
- Behavioral patterns (bounce, exit, product focus)
- Temporal/seasonal, technology, composite features

Impact: Improved model performance and business interpretability.



DATA MINING TECHNIQUES USED

Supervised Learning:

- Logistic Regression, SVM, K-NN (with hyperparameter tuning)
- Lasso Regression for value prediction

Unsupervised Learning:

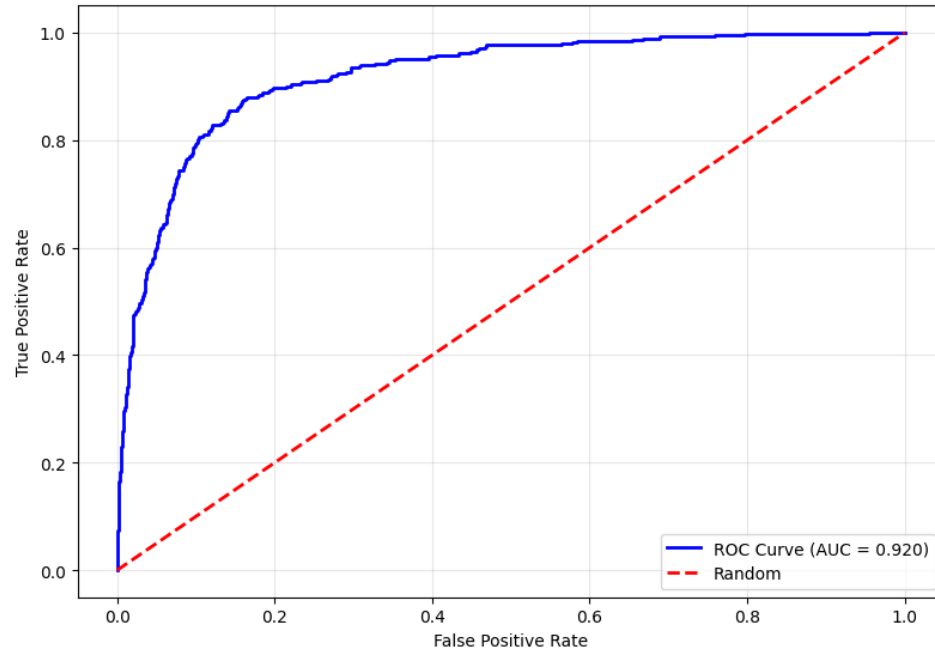
- DBSCAN clustering for customer segmentation
- PCA for visualization

Pattern Discovery:

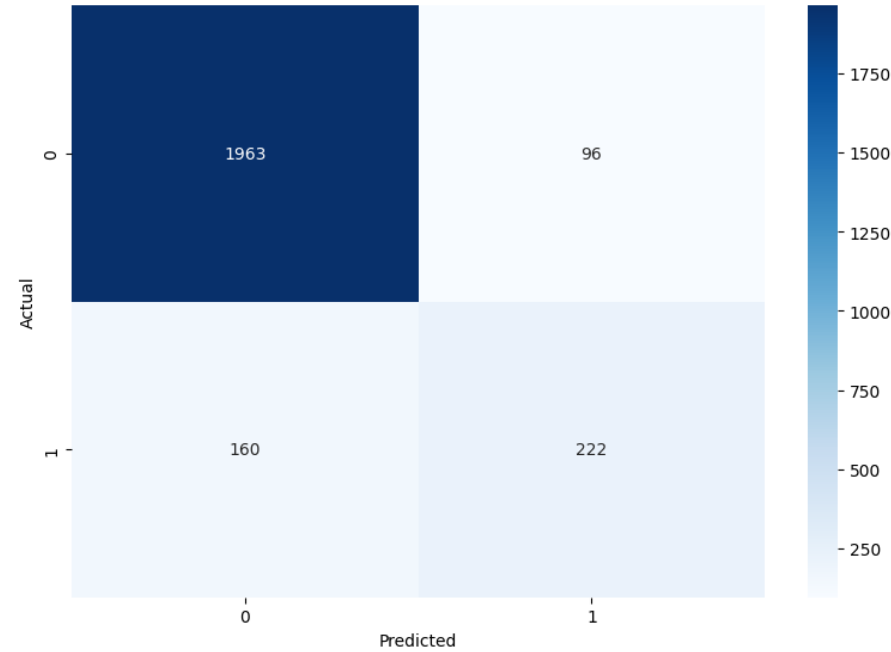
- FP-Growth association rule mining

Model Performance Evaluation

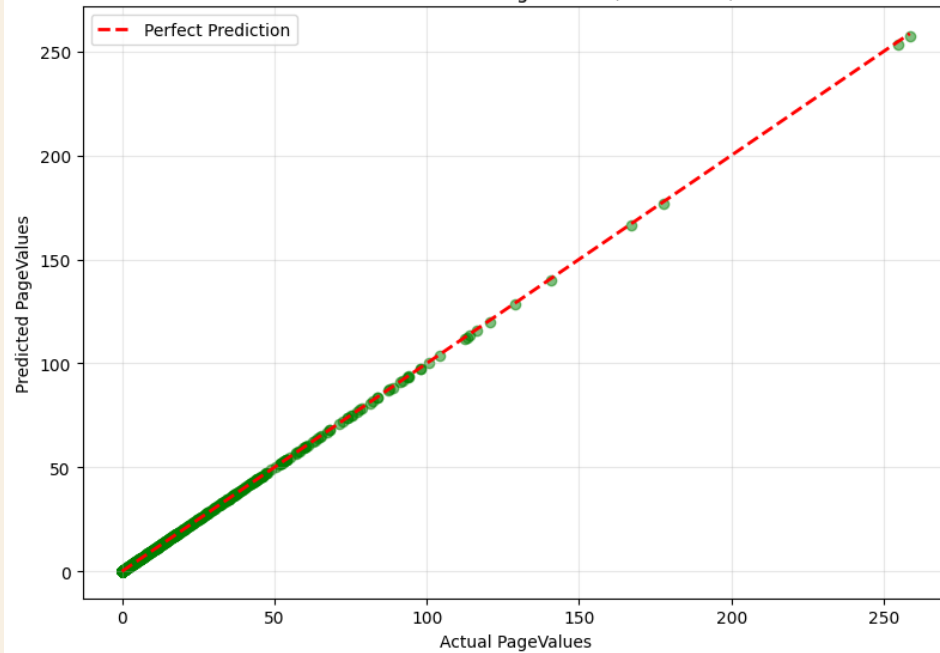
ROC Curve - Revenue Classification



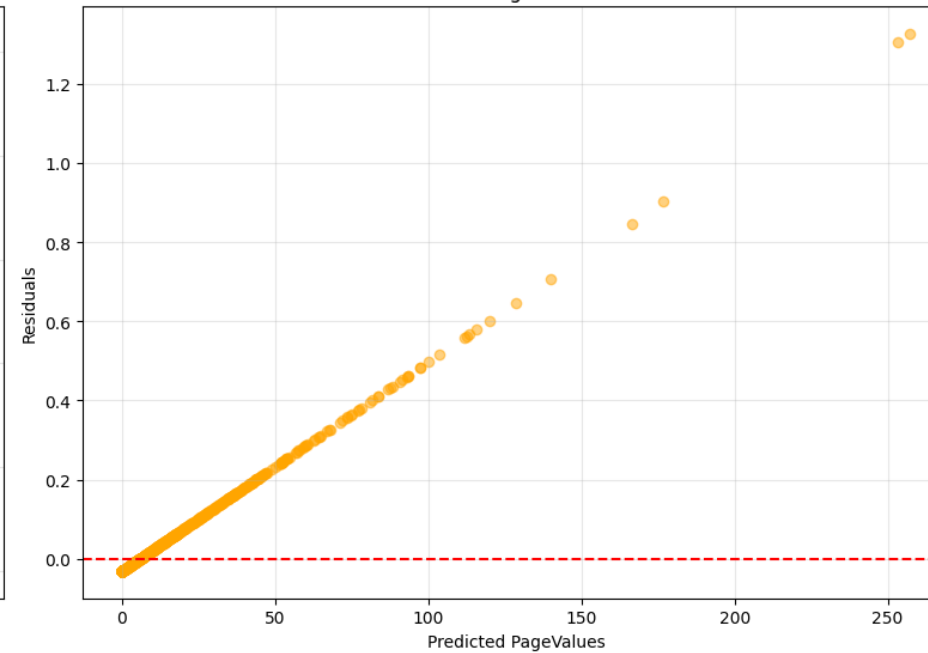
Confusion Matrix - Revenue Classification



Actual vs Predicted PageValues ($R^2 = 1.000$)



Residuals Plot - PageValues Prediction



KEY RESULTS - PREDICTIVE MODELING

- **Classification:**

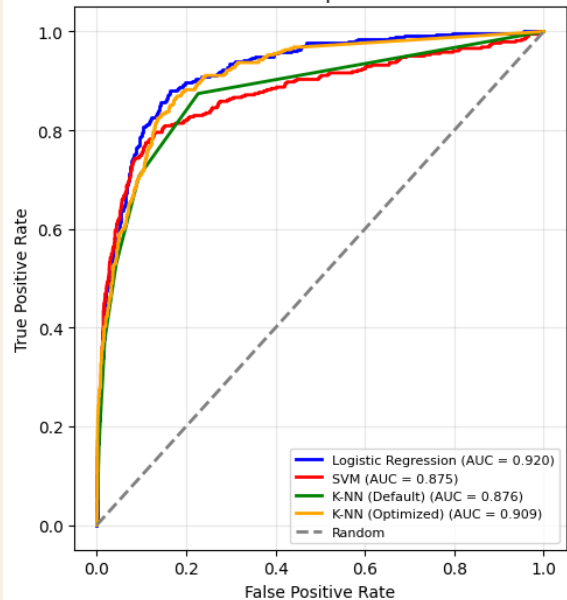
- Logistic Regression: ROC-AUC 0.90+, Accuracy 85%+
- SVM: ROC-AUC 0.87+, Accuracy 87%+
- K-NN (Optimized): ROC-AUC 0.88+, Accuracy 86%+

- **Regression:**

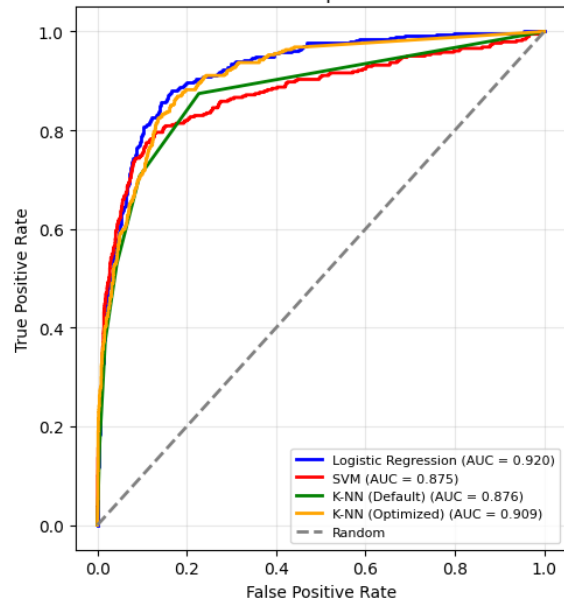
- Lasso: R^2 0.999+, RMSE <1.0

Comprehensive Classification Model Performance Evaluation

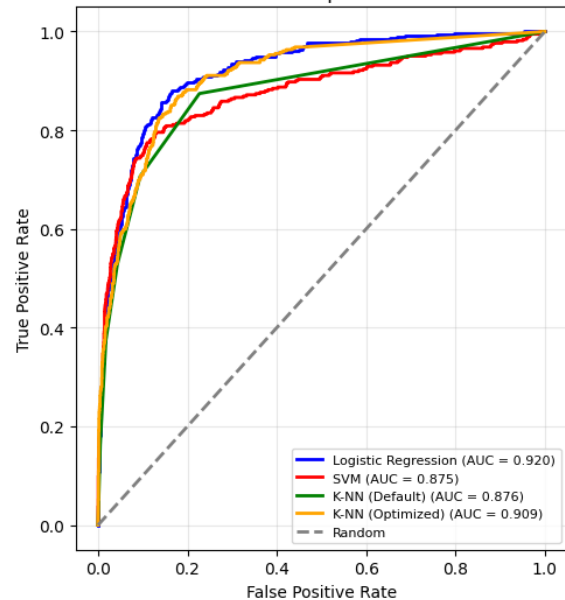
ROC Curves Comparison - View 1



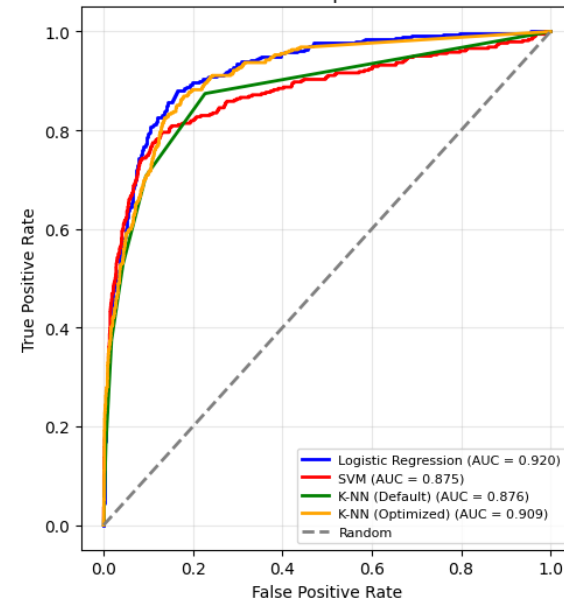
ROC Curves Comparison - View 2



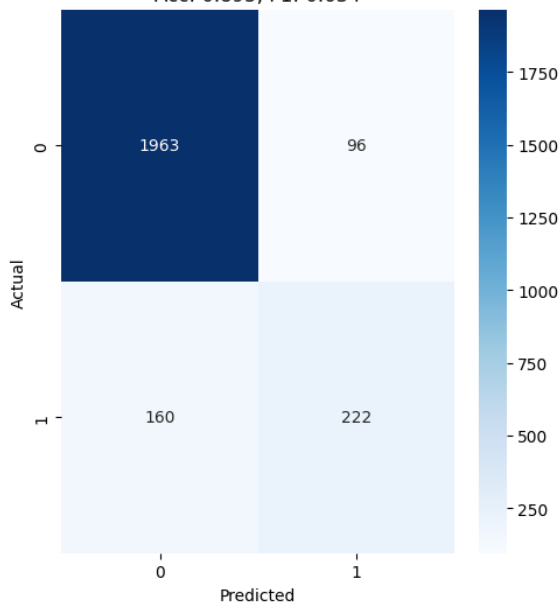
ROC Curves Comparison - View 3



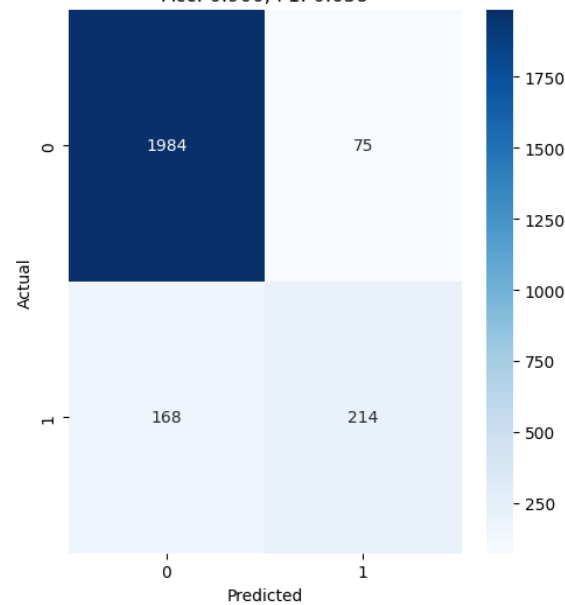
ROC Curves Comparison - View 4



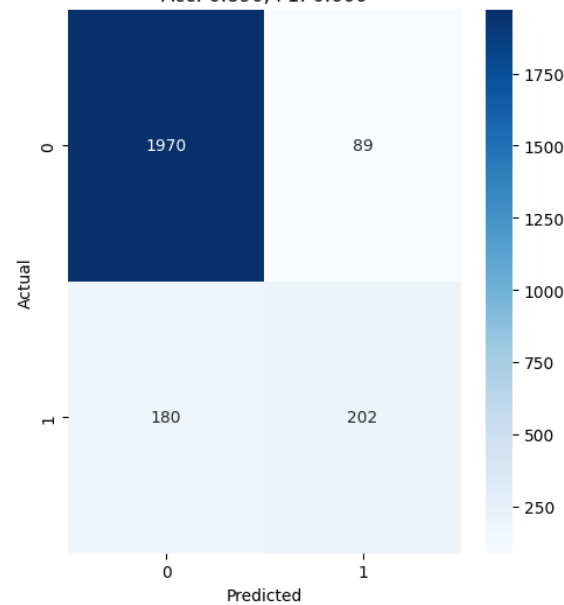
Logistic Regression Acc: 0.895, F1: 0.634



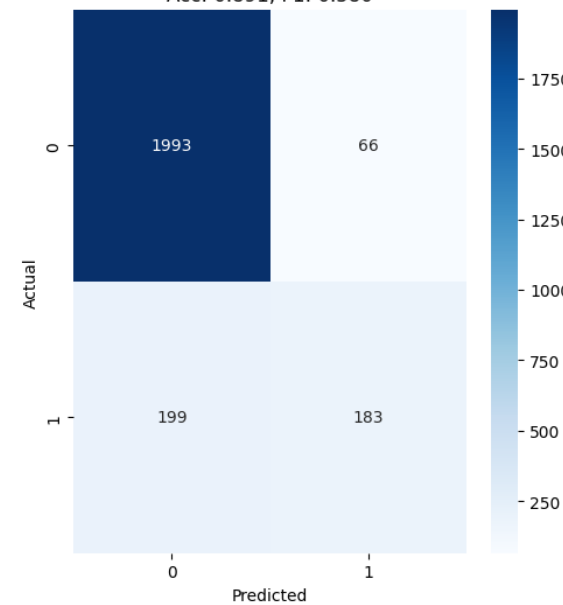
SVM Acc: 0.900, F1: 0.638



K-NN (Default) Acc: 0.890, F1: 0.600



K-NN (Optimized) Acc: 0.891, F1: 0.580



KEY RESULTS - CLUSTERING & PATTERNS

DBSCAN Clustering:

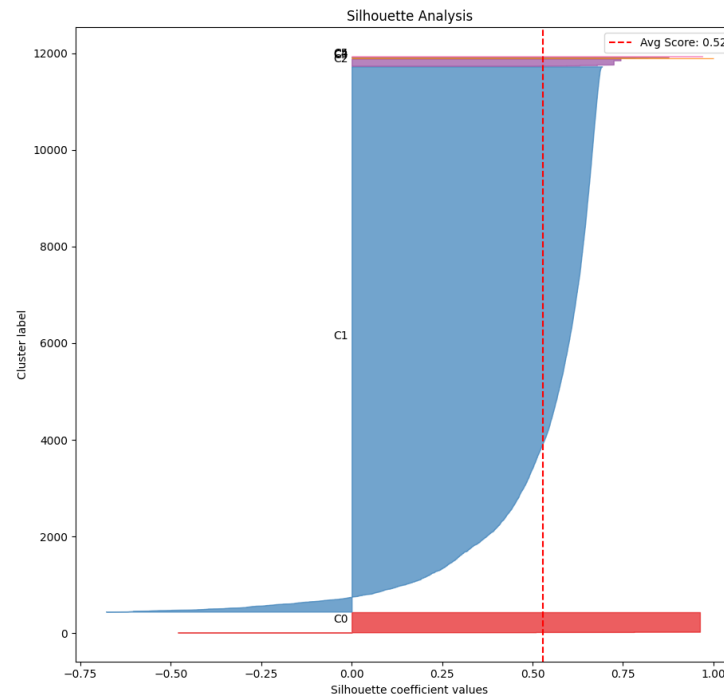
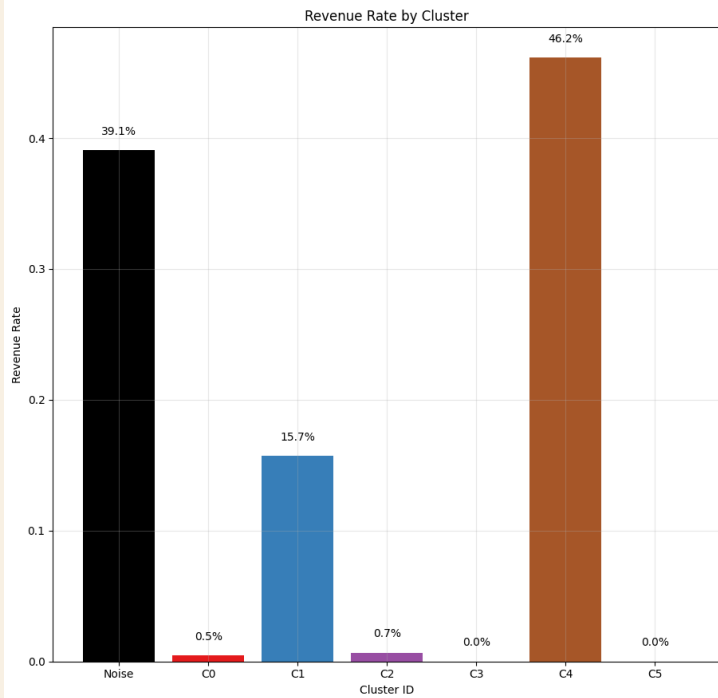
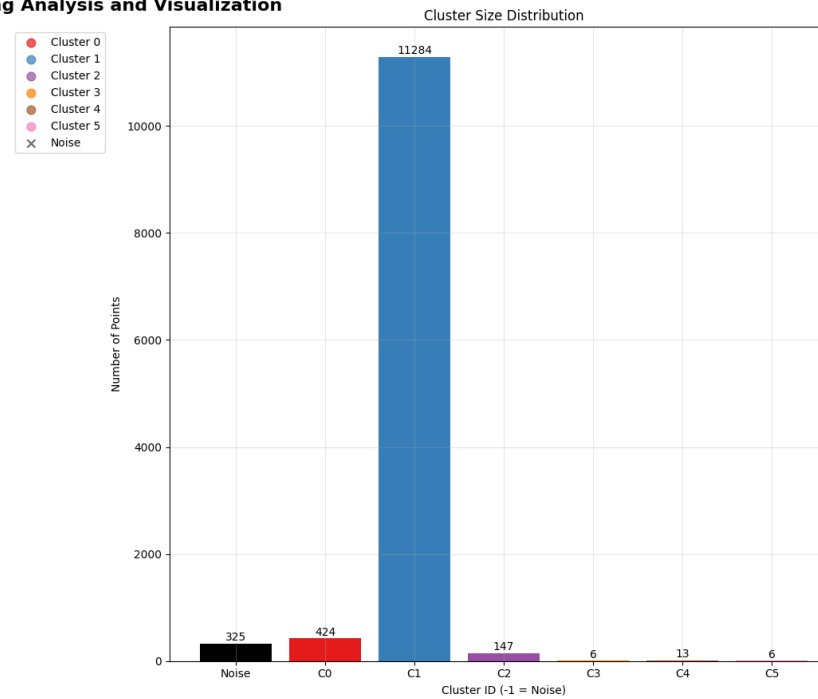
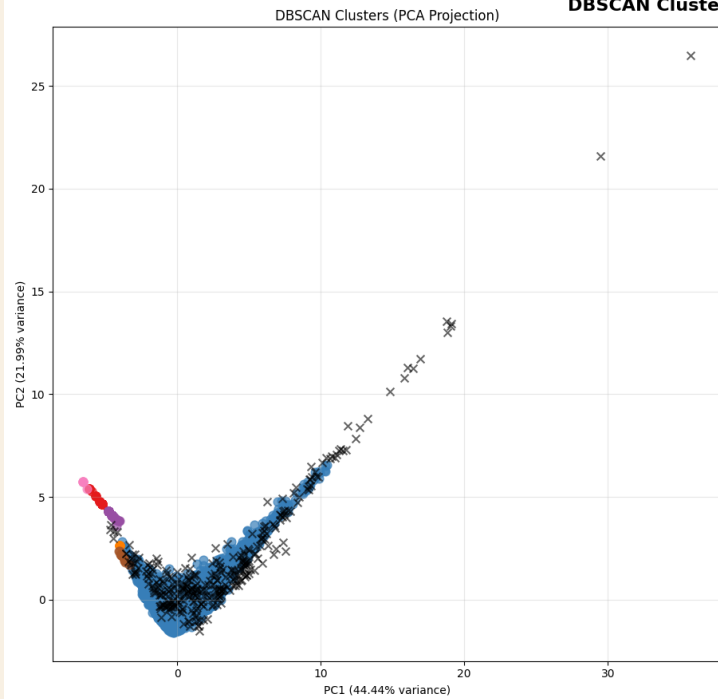
- 6 customer segments (main: 91.1%, high-value: 1.4%)
- Outlier detection (6.4%)

Association Rules:

- 66,564+ patterns, strong predictors (e.g., low bounce + product page → revenue)

Diagram: PCA cluster plot and association rule network.

DBSCAN Clustering Analysis and Visualization



BUSINESS INSIGHTS & RECOMMENDATIONS



Recommended Actions:

- Deploy predictive scoring for real-time offers
- Optimize navigation based on discovered patterns
- Segment-based marketing campaigns



Short/Long-Term:

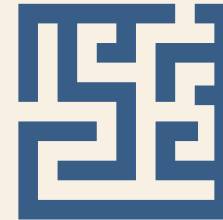
- Personalization engines
- Predictive analytics platforms
- Continuous model retraining and monitoring

CONCLUSION & FUTURE WORK



Achievements:

High-accuracy models, actionable segmentation, and pattern discovery
Projected 15-25% conversion improvement,



Future Recommendations:

Integrate more data sources (search, social)
Explore deep learning with real-time analytics
Continuous improvement and business integration

An abstract geometric design on the left side of the slide. It features a dark blue background with various geometric shapes and patterns. A white circle is positioned near the top left. Below it, a light blue semi-circle is visible. To the right of the semi-circle, there is a pink triangle with diagonal lines. Below the semi-circle, there is a pink square with a pattern of concentric lines. To the right of the square, there is a light blue triangle. Below the square, there is a pink triangle. To the right of the triangle, there is a dark blue triangle. The overall design is modern and minimalist.

THANK YOU