Assignment 4 – Loblaws Digital Customer Churn Prediction
CSCN8030 – Artificial Intelligence for Business Decisions and Transformation

Team Members
- Manu Mathew (8990691)
- Yogesh Kumar Gopal (8996403)
- Jahnavi Pakanati (9013742)
- Kapil (9064347)
- Parag (8877535)

1. Introduction
The customer churn is increasing, and losing these customers means lost revenue and higher costs for re-acquiring new customers. This report presents an integration plan, the development strategy of the AI model, and performance management to support proactive customer retention.
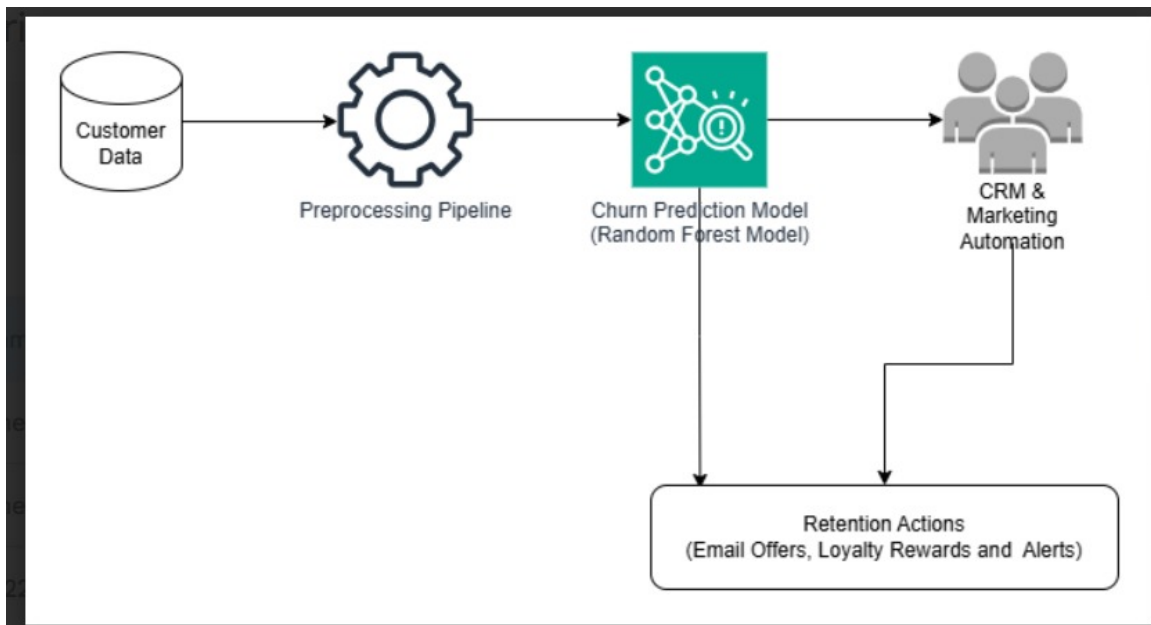
2. Business Context & Stakeholders
The system helps in identifying customers who are likely to exit the platform.

Key stakeholders include

• Product Managers

• CRM Teams

• Data Scientists

• Data Engineers

• IT Security, and Executives.

3. Integration Architecture
The AI model integrates with Loblaws existing ecosystem using ETL workflows, feature storage systems, model scoring pipelines, and CRM integration through APIs or batch scoring mechanisms.

Diagram: Customer Data → Preprocessing Pipeline → Churn Prediction Model (Random Forest Model) → CRM & Marketing Automation → Retention Actions (Email Offers, Loyalty Rewards and Alerts)

4. Data Flow & Security Plan

The system follows a structured pipeline:

data ingestion → preprocessing → model inference → secure result storage.

Security Plan includes

* AES-256 encryption

* RBAC access control

* audit logs

5. Deployment & Scaling Strategy

The churn model is deployed using cloud infrastructure like AWS, containerized with Docker, orchestrated through Kubernetes and monitored using CI/CD .
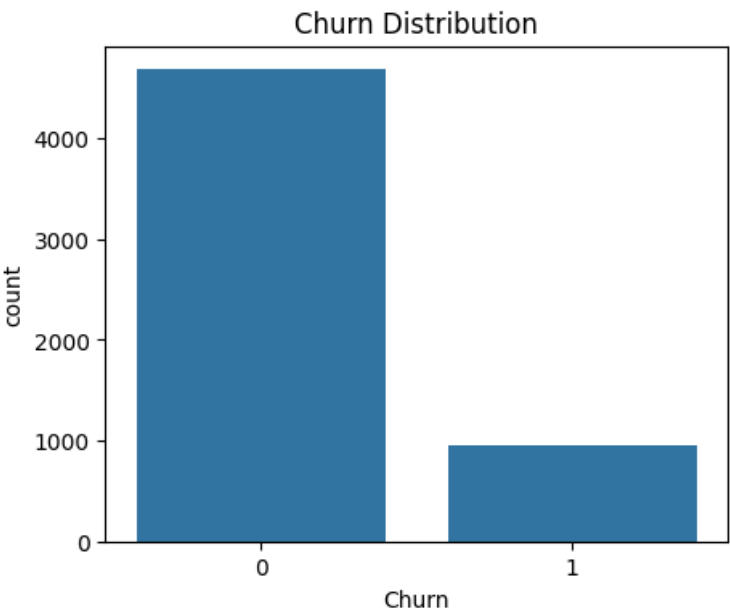
6. Risk Assessment & Mitigation

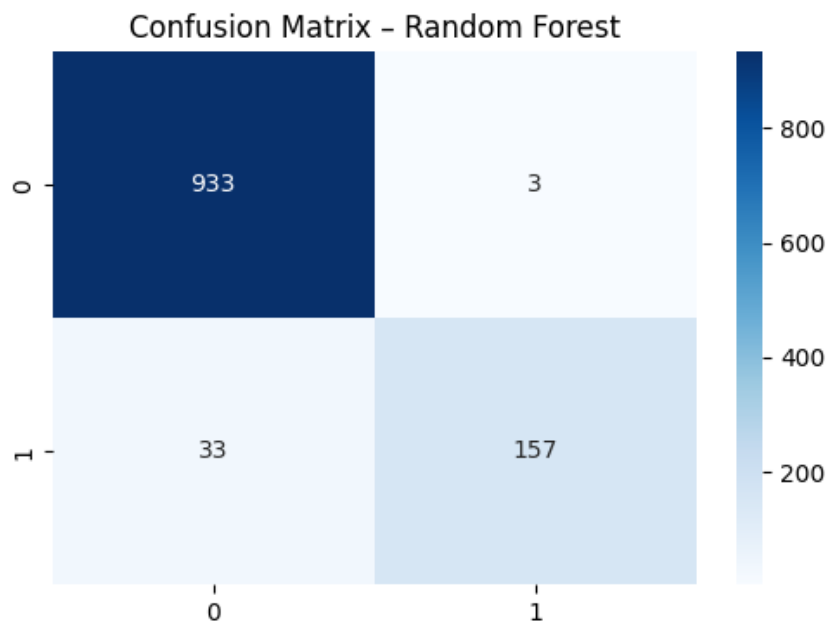| Risk	Description	Mitigation Strategy
Data Drift	Customer behavior changes over time	Monitor weekly, retrain monthly
Bias Risk	Model may disadvantage groups	Fairness metrics, demographic parity checks
Downtime	Model/API failure impacts CRM workflows	Redundant deployment, fallback systems
Overfitting	Poor generalization	Cross-validation, regularization
Security Breach	Exposure of PII	Encryption, IAM, audit logging

7. Model Development & Performance
The Random Forest classifier achieved the best performance across evaluation metrics
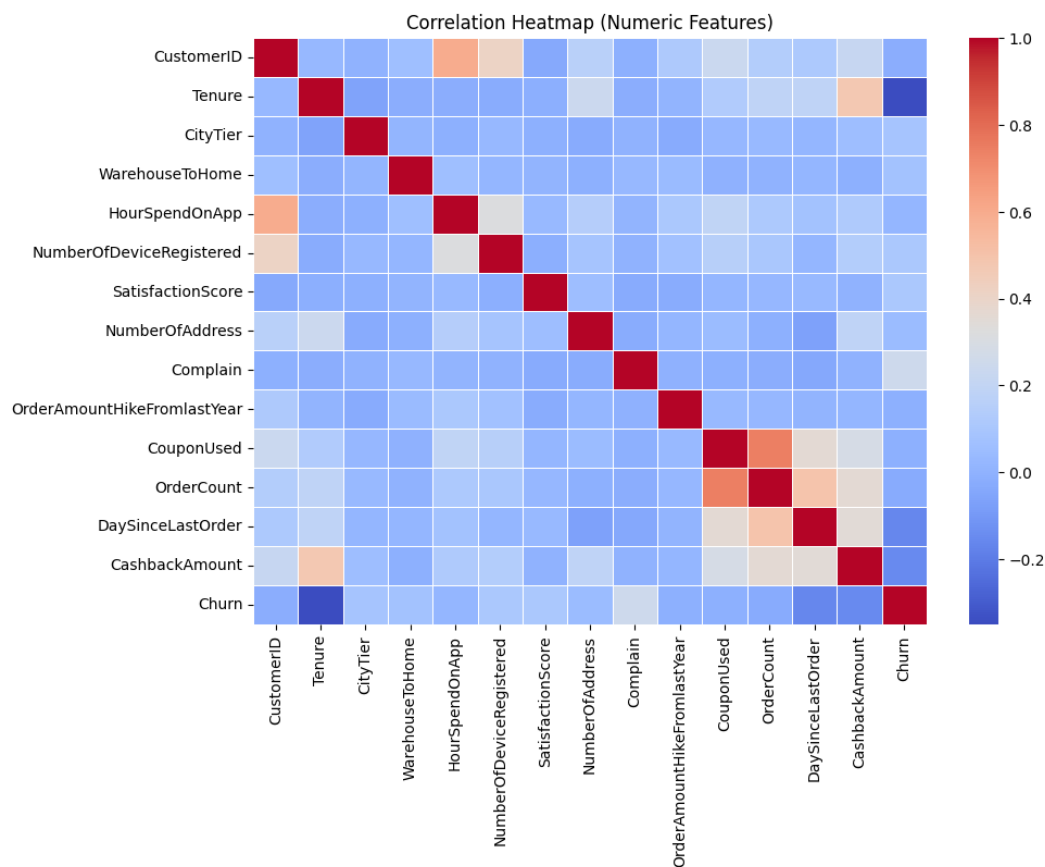The following visual outputs below are summarizing the data exploration and model evaluation.
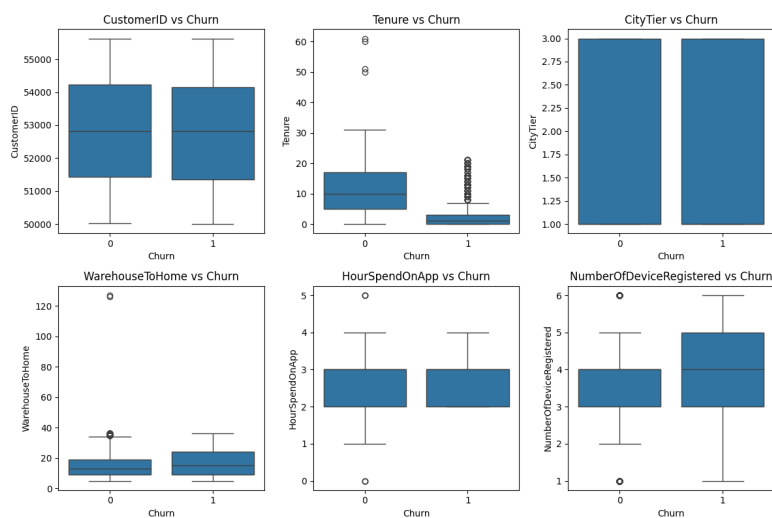
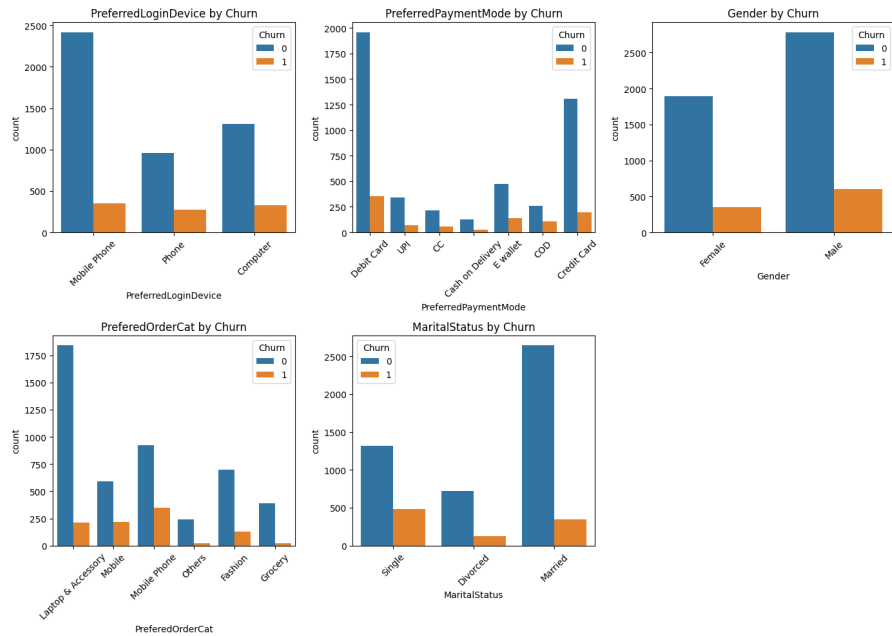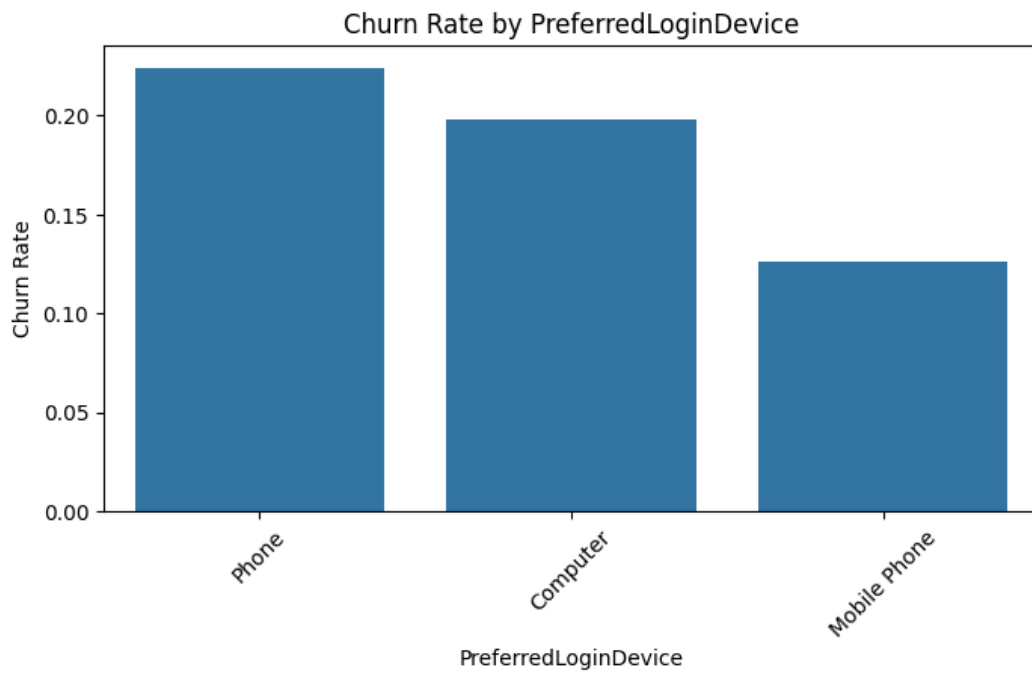## 8. Visual Outputs

Churn Distribution

Confusion Matrix – Random Forest
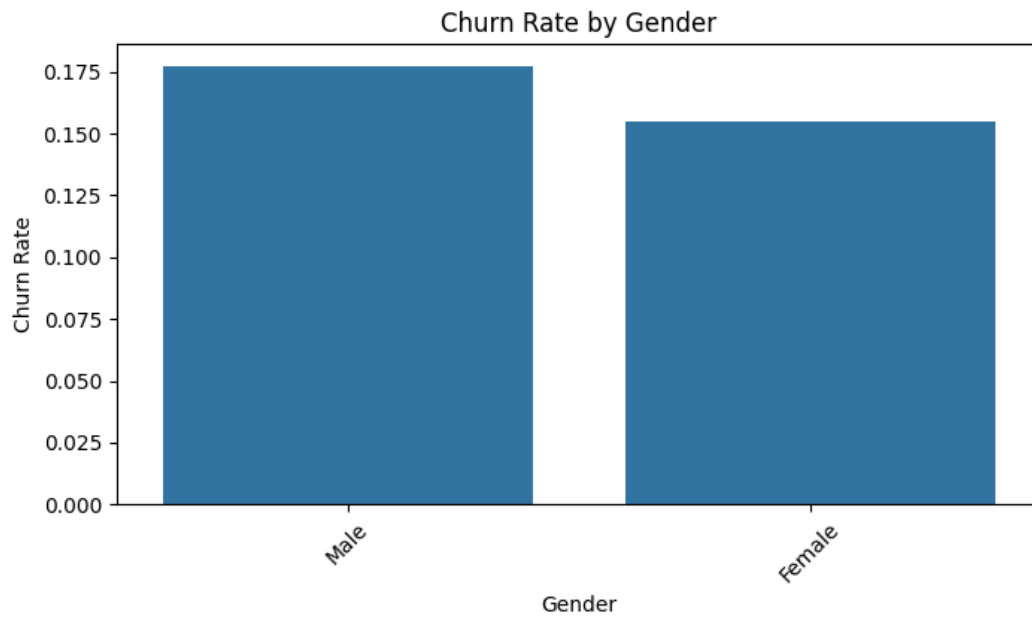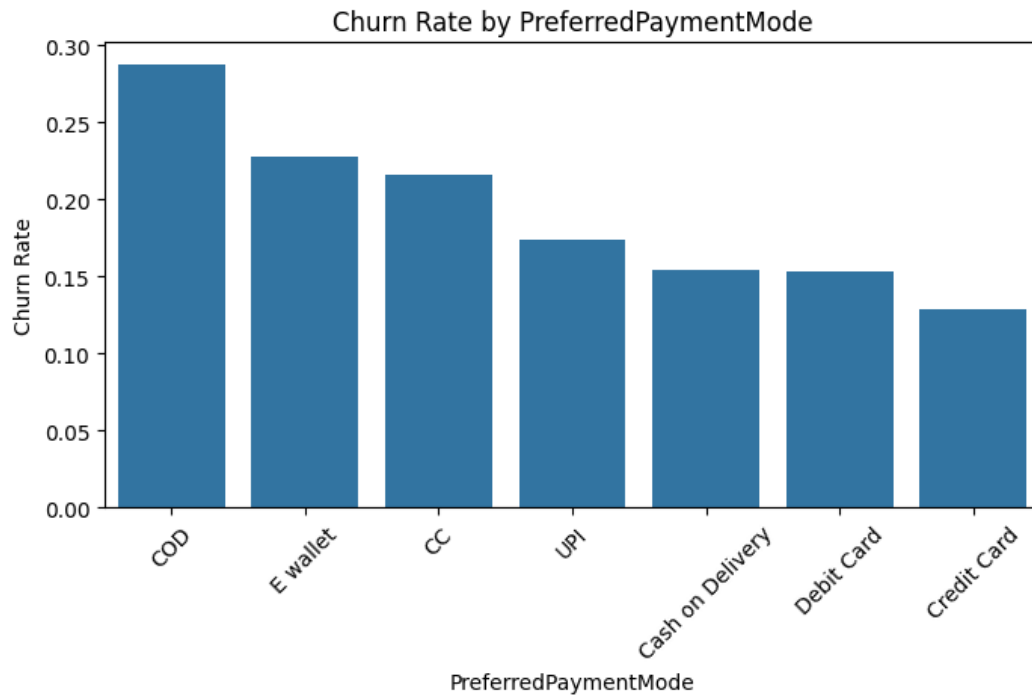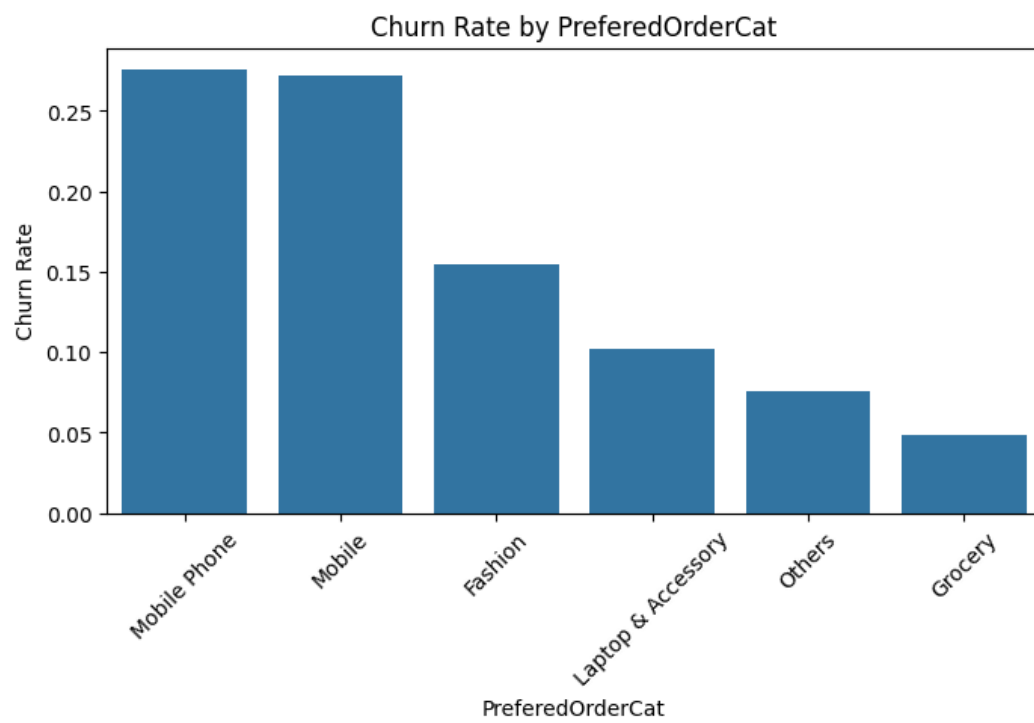
Confusion Matrix

Correlation Heatmap of Numeric Features

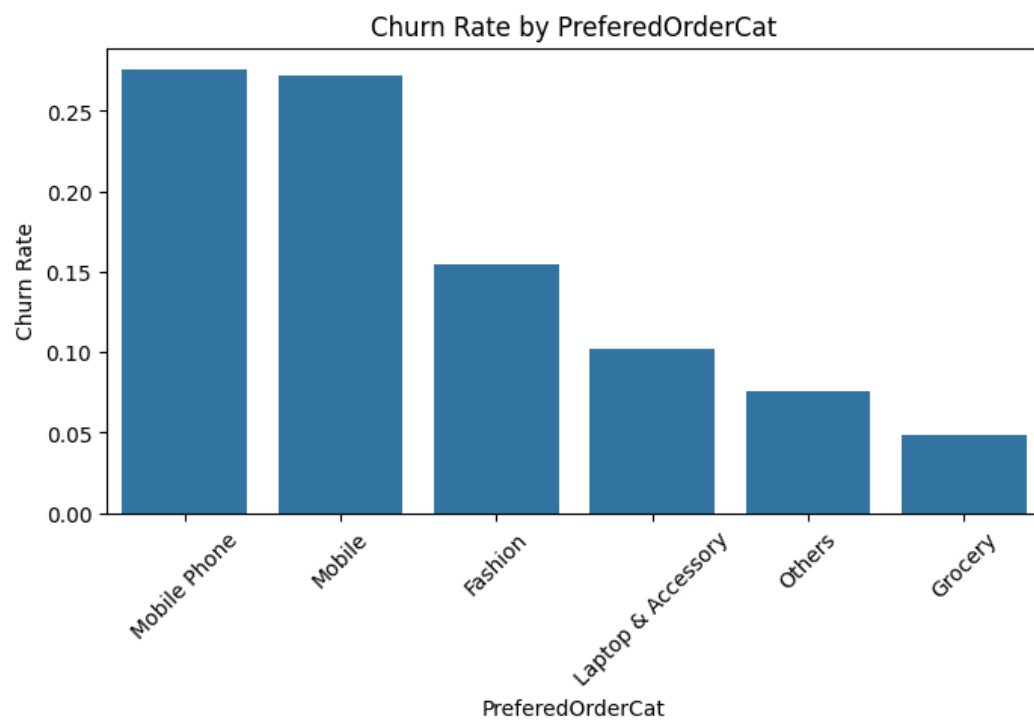

Boxplots: Numeric Features vs Churn

Categorical Feature Distribution by Churn

Churn Rate by PreferredPaymentMode

Churn Rate by Gender

Churn Rate by PreferedOrderCat

Churn Rate by PreferedOrderCat

Distribution of Numeric Features

ROC Curve – Random Forest

AUC = 0.997

Top 15 Feature Importances – Random Forest
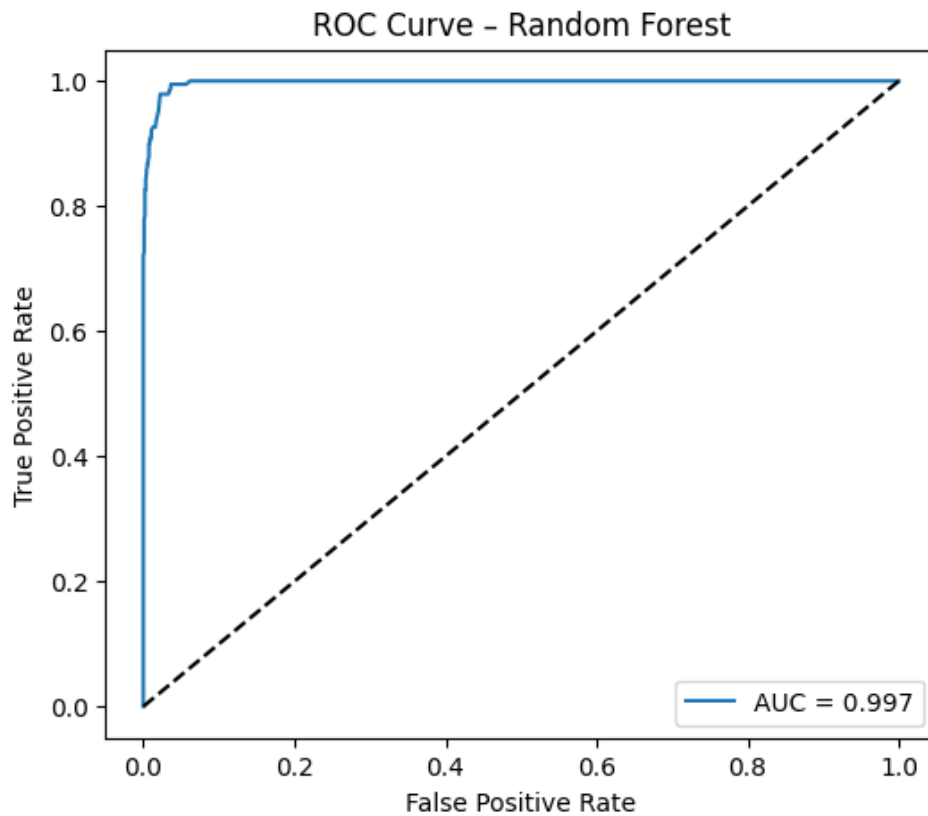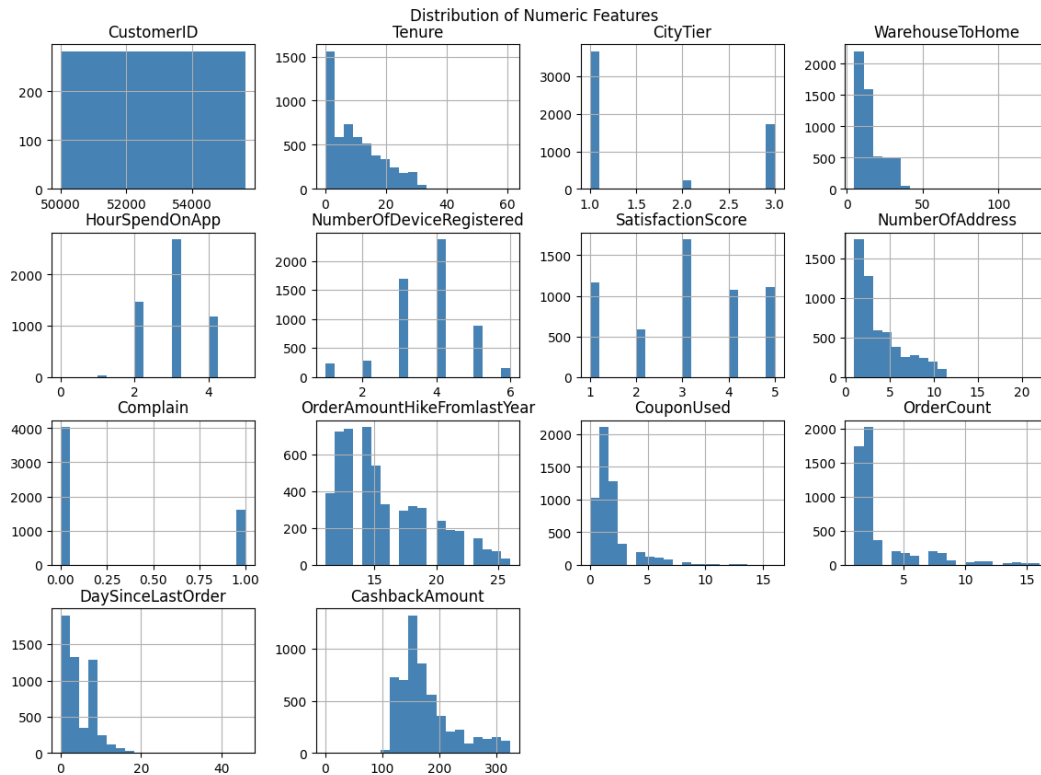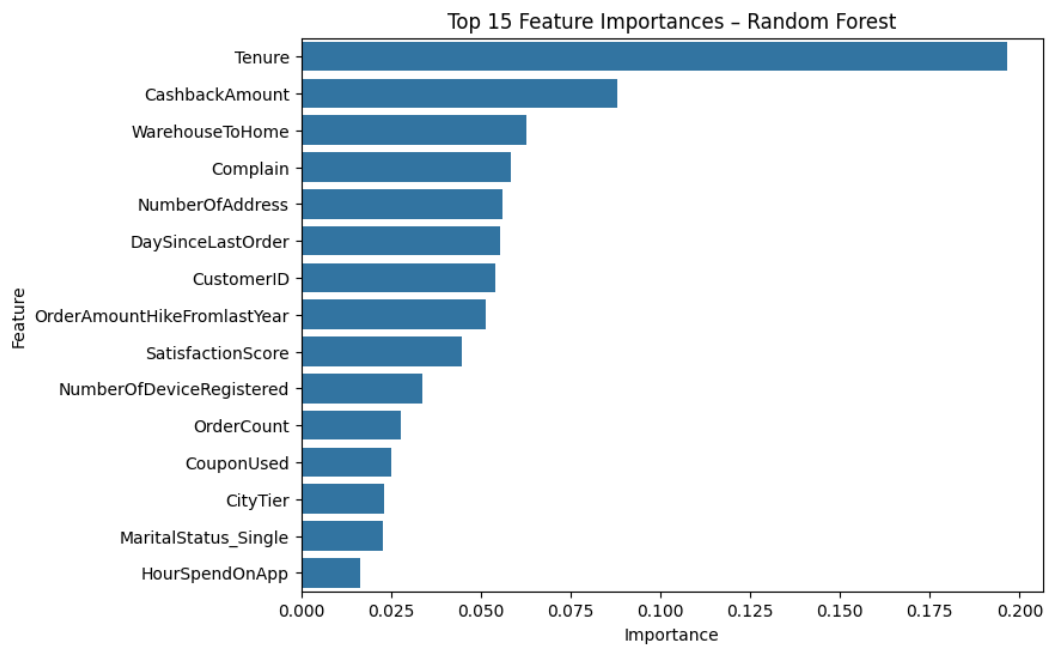

9. Performance KPIs

- Accuracy > 90%

- Precision > 90%

- Recall > 80%

- ROC-AUC > 0.90

- API Latency < 200 ms

- Batch Scoring < 30 minutes

10. Validation Strategy

Technical validation includes

* cross-validation

* ROC-AUC analysis

* confusion matrix review

* class balance analysis.

Business validation includes A/B testing of retention strategies and uplift modeling.

11. Conclusion

The model offers business value by enabling targeted retention interventions by aligning the solution with the Loblaws Digital's strategic goals, ensuring scalability, security, and long-term impact.