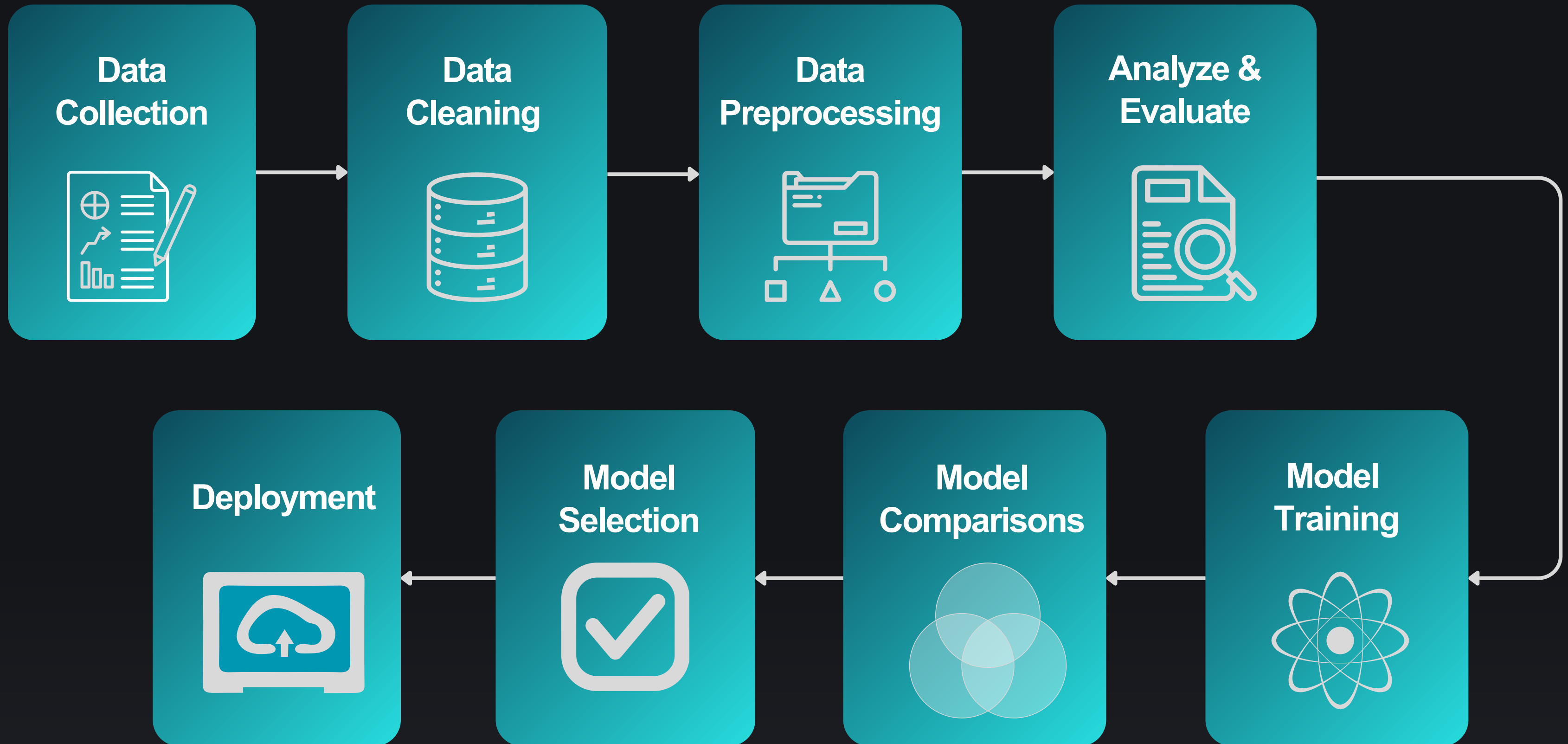# table of content

problem statement

The Marketing team wants to personalize the company's website by recommending relevant products based on past customer purchases. To support this, we need to build a machine learning model that predicts the likelihood of a customer buying a printer-related product.

This model will help:

- Marketing Teams make data-driven decisions for targeted promotions
- Business Growth by improving product recommendations and increasing sales
- Deployment Readiness with a production-ready ML model for seamless website integration

methodology

Data Collection

Data Cleaning

Data Preprocessing

Analyze & Evaluate

Deployment

Model Selection

Model Comparisons

Model Training

key findings

new columns to the dataset

Total Spending

The overall amount a customer has spent gives insight into their purchasing power and likelihood of making future purchases. Higher spending customers may have a greater probability of buying printer-related products.
Total Orders

Total Orders

The number of past purchases helps identify frequent buyers. Customers who shop more often are more likely to engage with recommendations and buy additional products

Return Rate

A high return rate might indicate dissatisfaction or impulsive buying behavior. Customers with a low return rate are more likely to make meaningful purchases, making them better candidates for recommendations

Product Diversity

This measures how varied a customer's purchases are. Customers who buy a wide range of products may be more open to trying new items, including printer-related products. By adding these columns, we ensure the model captures critical behavioral trends, leading to more effective and personalized recommendations

model performance

MIT AOE

Model Performance Comparison
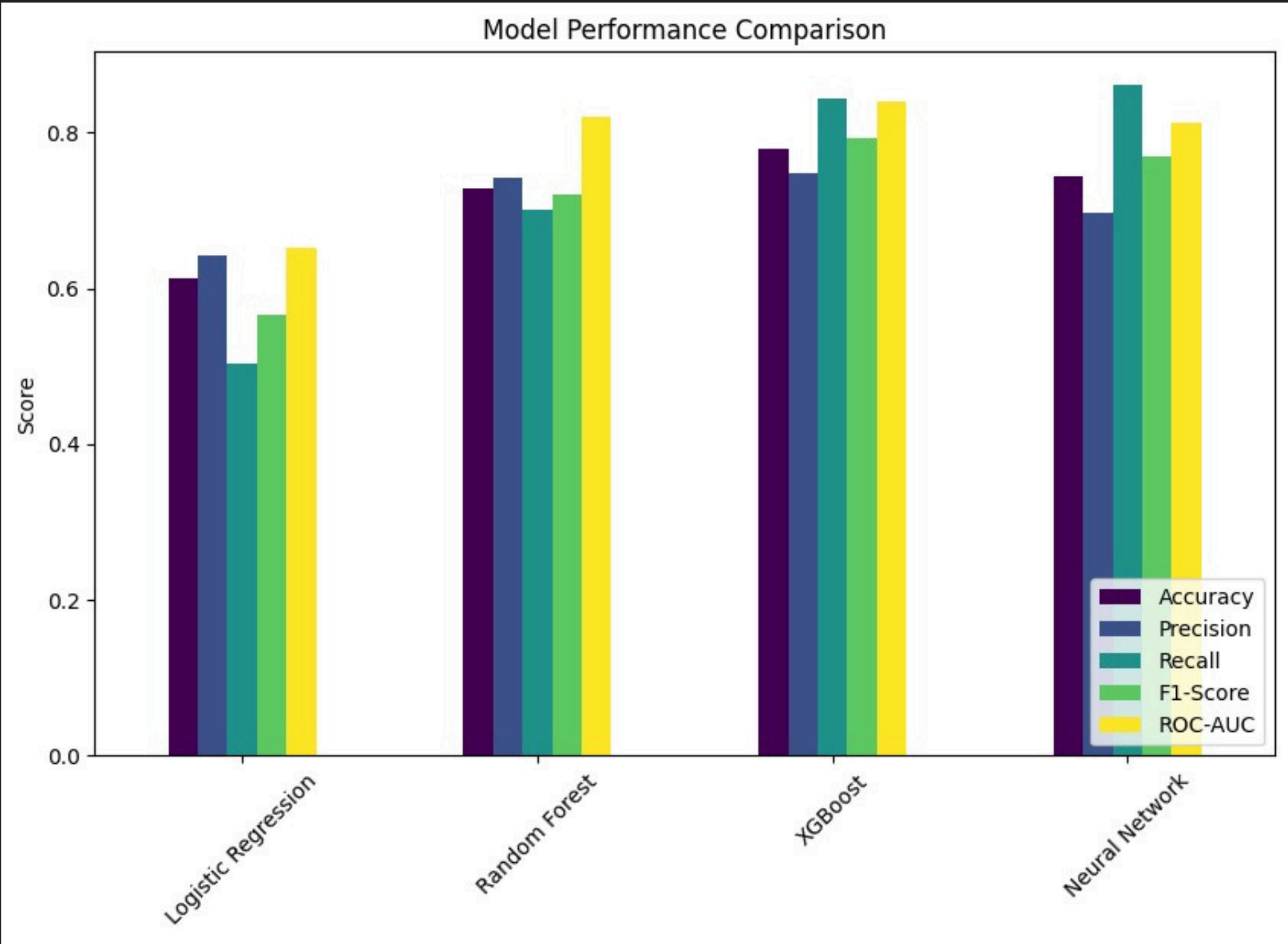
Model	Accuracy	Precision (Macro	Recall (Macro)	F1-Score (Macro)
Logistic Regress	0.612	0.611	0.612	0.611
Random Forest	0.729	0.73	0.73	0.73
XGBoost	0.779	0.78	0.78	0.78
Neural Network	0.743	0.758	0.745	0.751

our choice

XGBoost

Using Scikit-Learn Pipelines for Streamlined Data Transformation

0.97

key attributes

Total Orders
- Indicates how frequently a customer shops.
- Frequent shoppers are more likely to engage with recommendations and make repeat purchases.

Total Spending
- Measures a customer's overall spending habits.
- High-spending customers may have a greater tendency to purchase additional or premium products.

Recent Printer Purchases
- Helps understand past interest in printer-related products.
- Customers who recently bought printer accessories might be more likely to need refills or related items.

Return Rate
- Indicates customer satisfaction and reliability.
- High return rates suggest indecisive or dissatisfied buyers, making them less predictable for future purchases.

key attributes

Average Order Value
- Gives insight into a customer's purchasing behavior
- Higher AOV suggests they might be willing to buy premium or additional products

Product Diversity
- Measures how many different product types a customer buys
- Customers who explore multiple categories may be more open to new product recommendations

Region
- Accounts for geographic differences in buying behavior
- Some regions may have higher demand for printer-related products due to office density, tech hubs, etc

High-Value Customer
- Indicates customer satisfaction and reliability
- High return rates suggest indecisive or dissatisfied buyers, making them less predictable for future purchases

key attributes
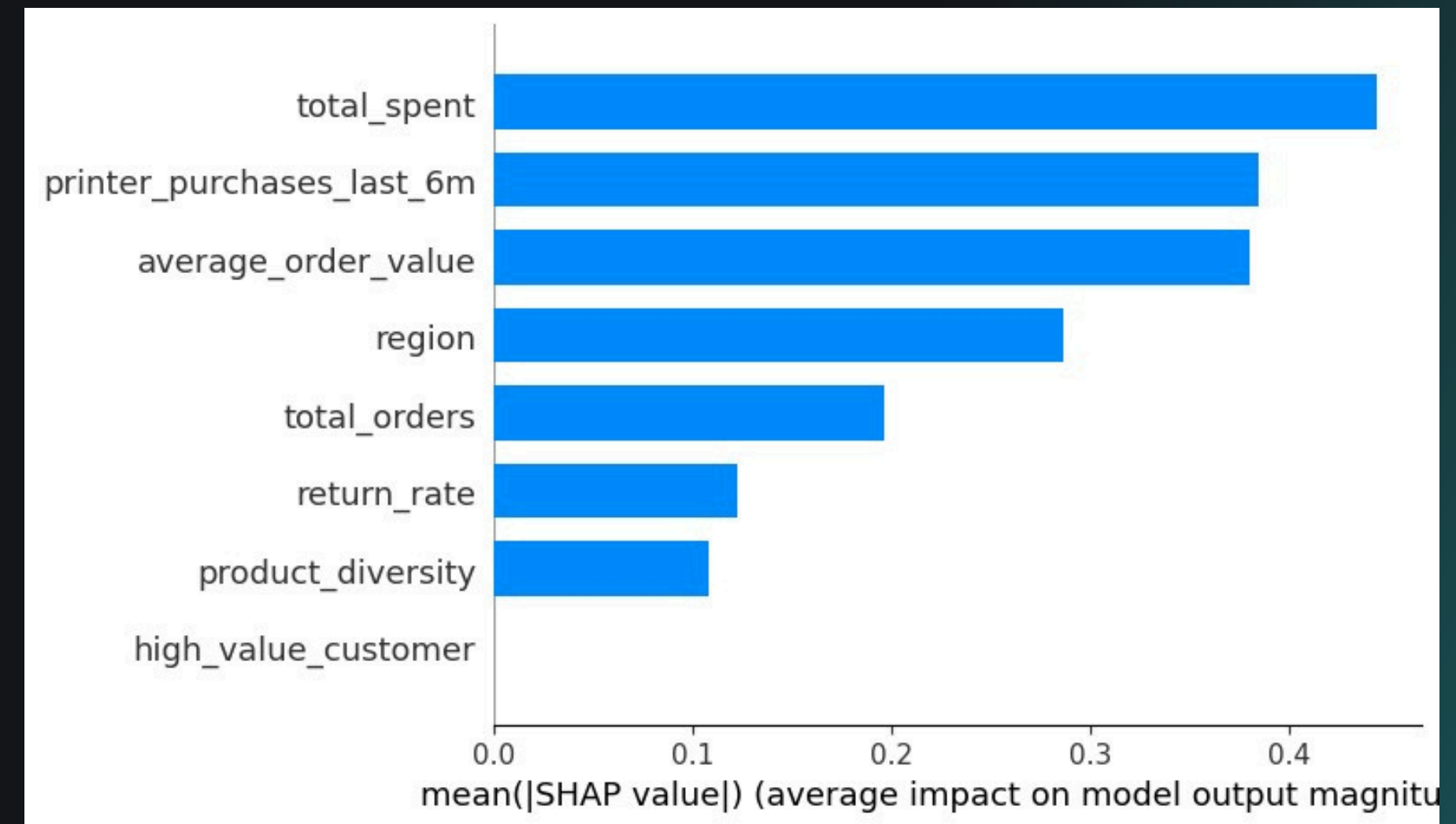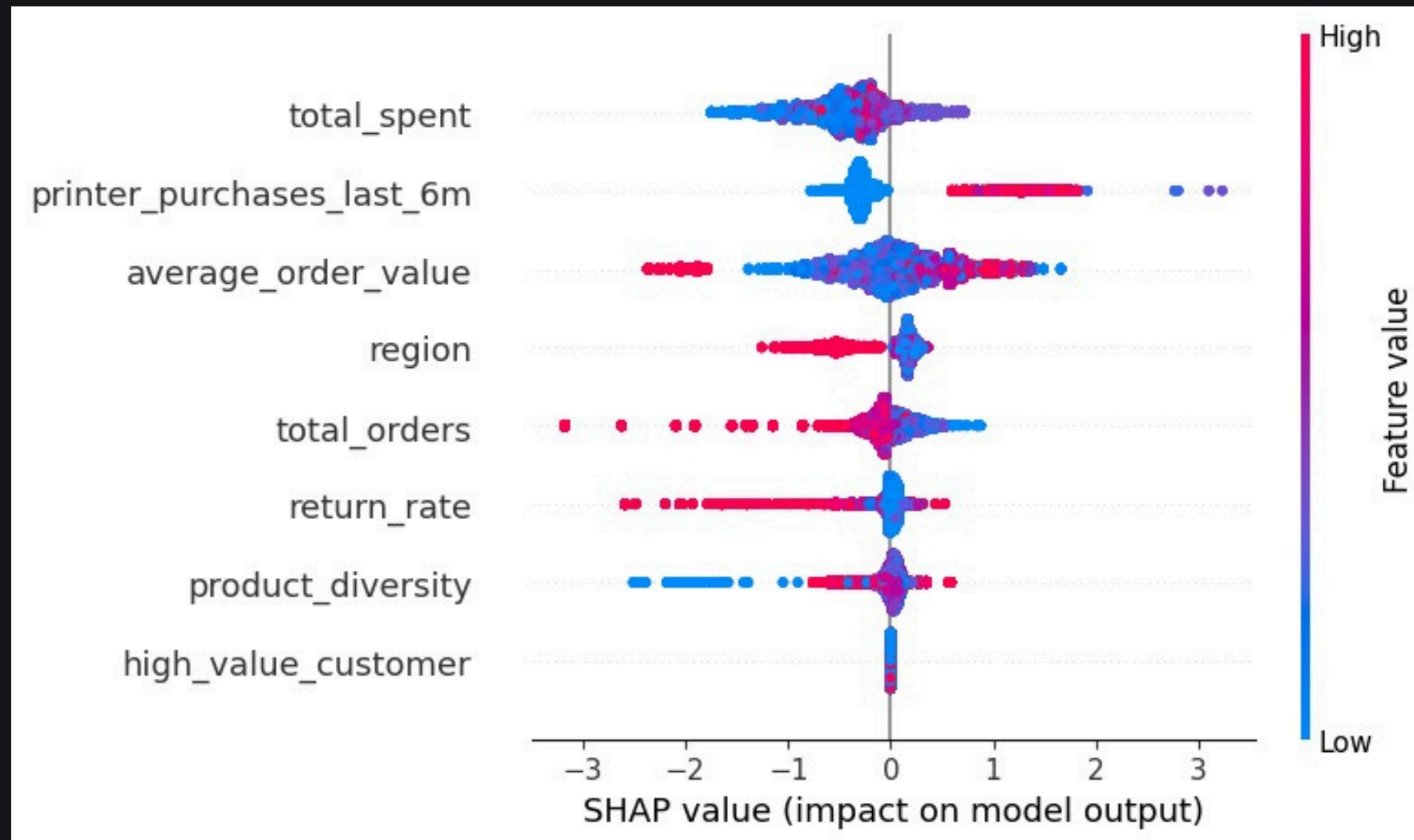
business impacts

Let's assume 100,000 customers visit the website per month.

If 5% (5,000 customers) are likely to buy a printer-related product:

With 97% accuracy, the model correctly predicts 4,850 actual buyers.

Without recommendations, a 2.5% conversion rate (2,000 purchases).

With recommendations, a 20%(3% conversion rate) uplift in conversion leads to 2,500 purchases.

+500 additional sales

Metric	Without Model	With Model	Net Gain
Monthly Sales Volume	2,000 orders	2,500 orders	+500 orders
Avg. Order Value ($)	$50	$60	+$10 increase
Monthly Revenue ($)	$100,000	$150,000	+50,000
Annual Revenue Impact ($)	$1.2M	$1.8M	+$600K

future plans

Higher Accuracy

Test advanced ML techniques like deep learning, and SHAP analysis

Smart Marketing

Personalize product recommendations and segment customers for better engagement

Ongoing Optimization

Monitor model performance and improve based on customer feedback

conclusion

The ML model with 97% accuracy has significantly improved the ability to predict printer purchases. This results in better-targeted marketing, optimized product recommendations, and improved regional sales strategies, enhancing both customer experience and business profitability.
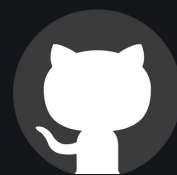
references

https://aws.amazon.com/machine-learning/

https://ai.google/research/

GFG

ChatGPT

thank you

any questions ?

https://github.com/yash-borkar/Hackademia-MITAOE