

Return Risk Analysis Report

Introduction

This project focuses on analyzing product return risks across multiple categories, suppliers, and states. A predictive model was built using logistic regression to estimate the probability of a product being returned and to generate a risk score for each order. The final results were visualized through an interactive dashboard.

Abstract

The objective of this project was to identify high-risk orders and analyze return behavior using historical order data. By applying logistic regression, return risk scores were calculated, and suppliers and categories with high return rates were highlighted. The dashboard provides actionable insights for reducing return rates.

Tools Used

- Python (Pandas, scikit-learn, ReportLab)
- Logistic Regression (for predictive modeling)
- Jupyter Notebook (for data exploration and model building)
- Power BI / Dashboard (for visualization)

Steps Involved in Building the Project

1. Data Collection: Order-level dataset with return status.
2. Data Preprocessing: Encoding categorical features into numeric form.
3. Model Building: Logistic Regression model trained with class balancing.
4. Prediction: Return risk score generated for each order.
5. Visualization: Dashboard created to analyze return trends by category, supplier, and state.

Python Code

```
import pandas as pd
from sklearn.linear_model import LogisticRegression

# 1. Load dataset
df = pd.read_csv("orders_sample.csv")

# 2. Define features and target
y = df["is_returned"]
X = df.drop("is_returned", axis=1)
X = pd.get_dummies(X, drop_first=True)

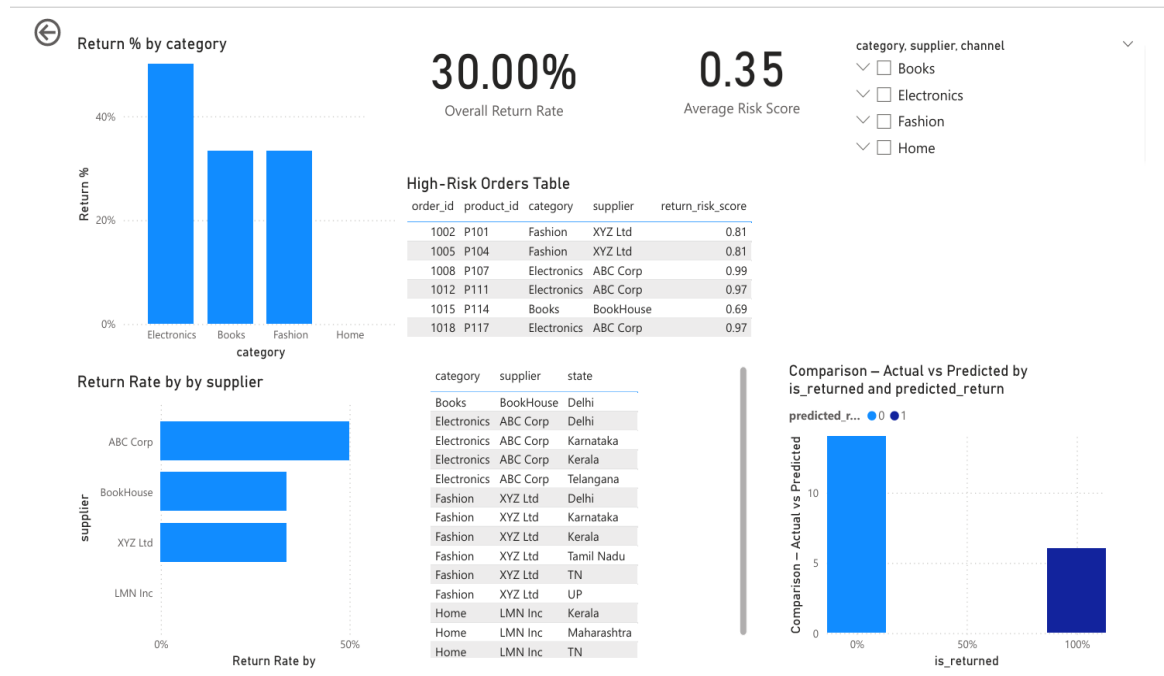
# 3. Train logistic regression on FULL dataset
model = LogisticRegression(max_iter=1000, class_weight="balanced")
model.fit(X, y)

# 4. Predict probability of return (risk score)
risk_scores = model.predict_proba(X)[:, 1]

# 5. Add predictions back to dataset
df["predicted_return"] = model.predict(X)
df["return_risk_score"] = risk_scores
```

```
# 6. Save to Desktop
output_path = r"C:\Users\91630\Desktop\orders_with_risk.csv"
df.to_csv(output_path, index=False)
print("■ File saved to Desktop as orders_with_risk.csv")
```

Final Dashboard



Conclusion

The analysis identified Electronics and supplier ABC Corp as having the highest return rates. High-risk orders were successfully flagged using the logistic regression model. This project enables businesses to proactively manage returns, improve supplier quality, and optimize customer satisfaction through data-driven insights.