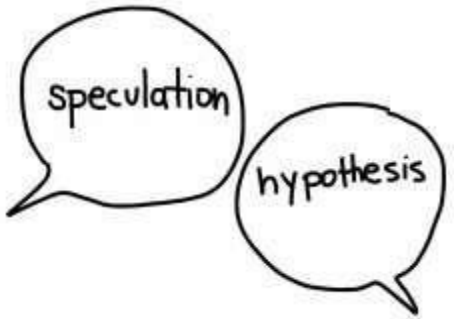
Amazon Sales Report - Predicting Cancellations

Vami Krishna (SM22UBBD154) - Project manager
Masthan reddy (SM22UBBD005) - Visualization specialist (BI)
Ragam ronith patel(SM22UBBD150),
Sahithi (SM22UBBD048) - Data analyst

Rajesh naik (SM22UBBD095) - Project coodinator (PPT)
Rahul konakandla (SM22UBBD165) , chandana(SM22UBBD025),
Challa saketh (SM22UBBD175) - Data collector
Geethvika(SM22UBBD073)

Objective



Problem statement -

Predicting cancellations in advance so that high risk of cancellation can be incorporated in pricing of the product

Research question -

What factors contribute most significantly to order cancellations?

Hypothesis -

Null Hypothesis (H0):-Independent variables (size, category, promotion id's , amount) does not affect the order cancellation rate.

Alternate Hypothesis (H1): At Least one of them effect cancellation rate

Market and industry context

Industry background -

The industry context for predicting order cancellations is crucial, especially if you are dealing with e-commerce, retail, or logistics

Current trends in Business analytics

Predictive modeling

Operational efficiency

Literature review

Key findings -

Order cancellations based on promotion id's , fulfillment, category, size and price

Methodology

Data collection - Source – Kaggle

Data cleaning - Removed blank spaces and missing values and removed unwanted variables

AI model - Logistic regression

Validation technique - Split validation

Used AI studio for prediction , BI and excel for visualization

PerformanceVector (Performance) | Logistic Regression Model (Logistic Regression)

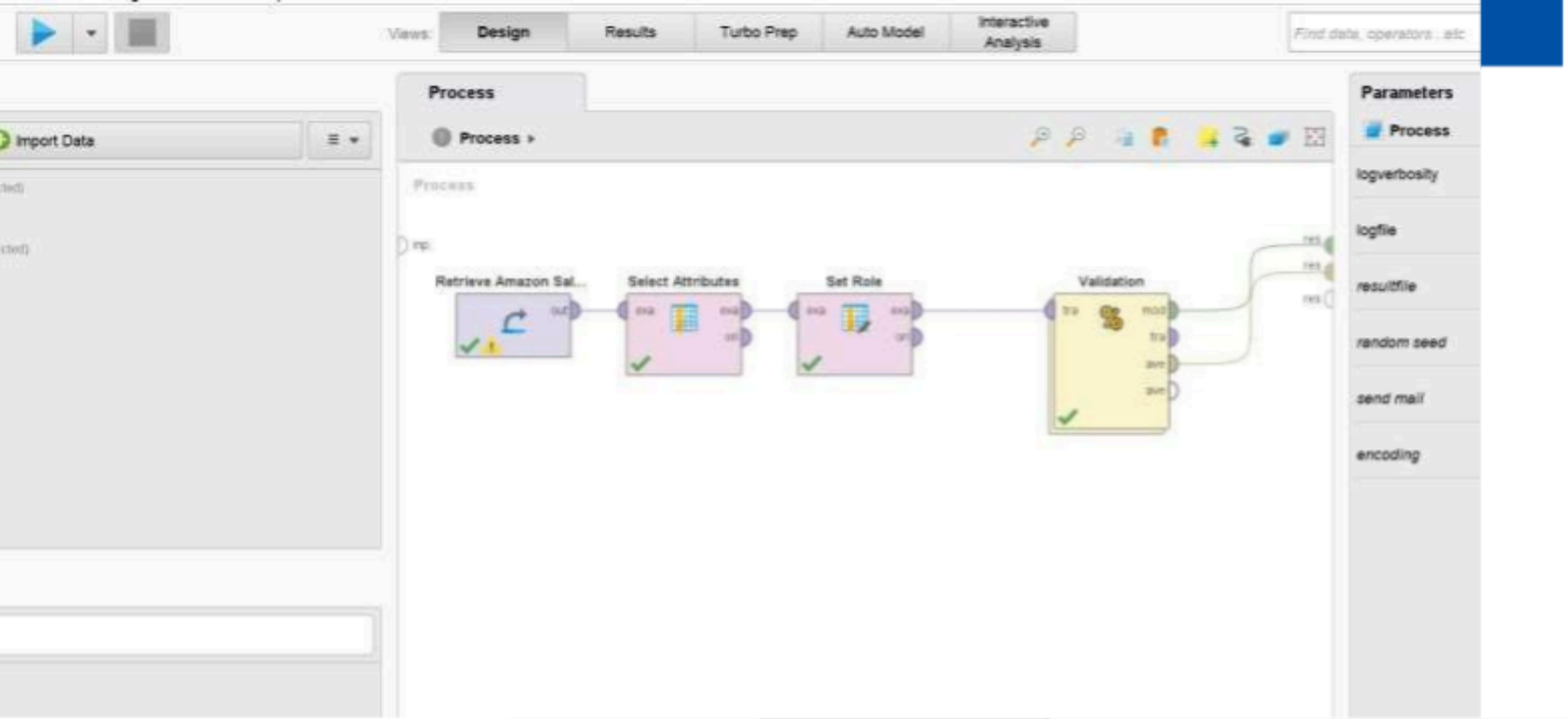
on error

Table View | Plot View

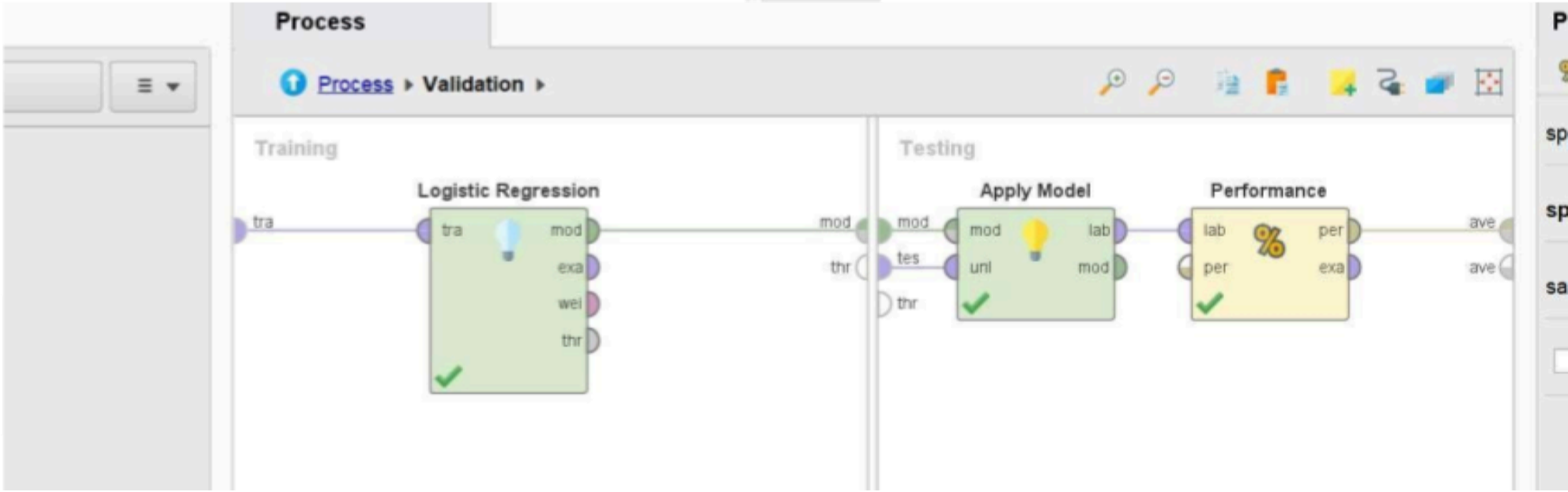
accuracy:

	true Cancelled	true Shipped	class precision
pred. Cancelled	5491	88	98.42%
pred. Shipped	9	33105	99.97%
class recall	99.84%	99.73%	

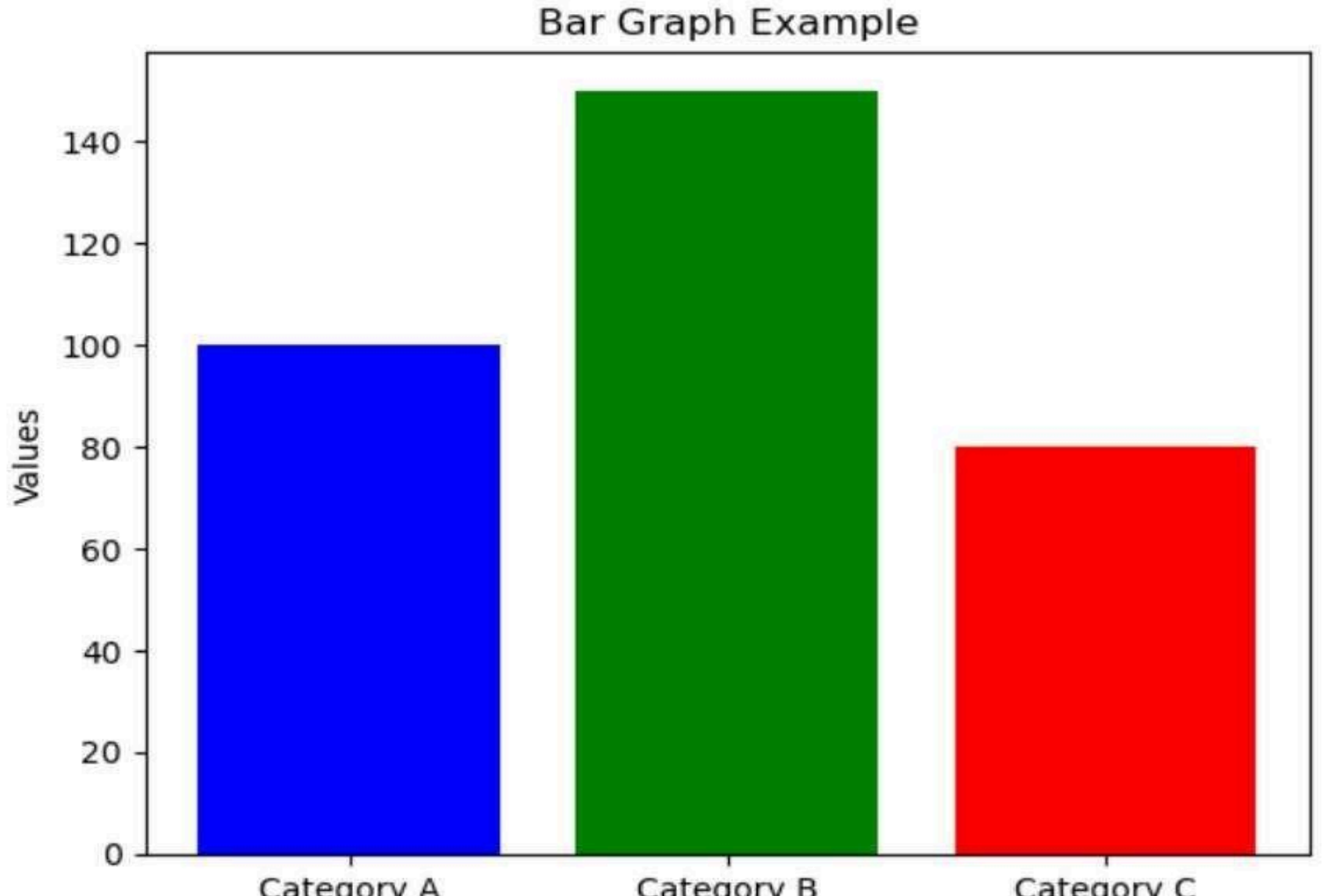
```
Accuracy: 0.8216649694501018
Classification Report:
              precision    recall  f1-score   support

           0       0.82      1.00      0.90      6455
           1       0.00      0.00      0.00      1401

    accuracy                           0.82      7856
   macro avg       0.41      0.50      0.45      7856
weighted avg       0.68      0.82      0.74      7856
```


Views:
Design
Results
Turbo Prep
Auto Model
Interactive Analysis
Import Data
Process
Process
Retrieve Amazon Sal...
Select Attributes
Set Role
Validation
Parameters
Process
logverbosity
logfile
resultfile
random seed
send mail
encoding

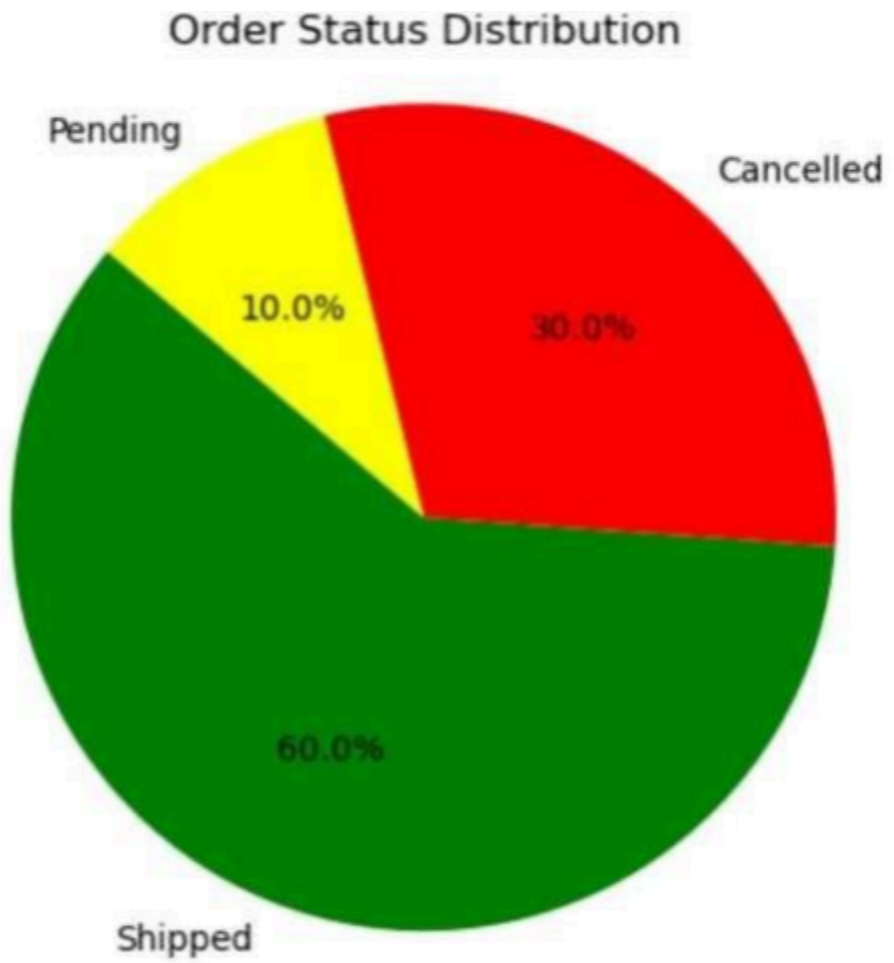
AI studio

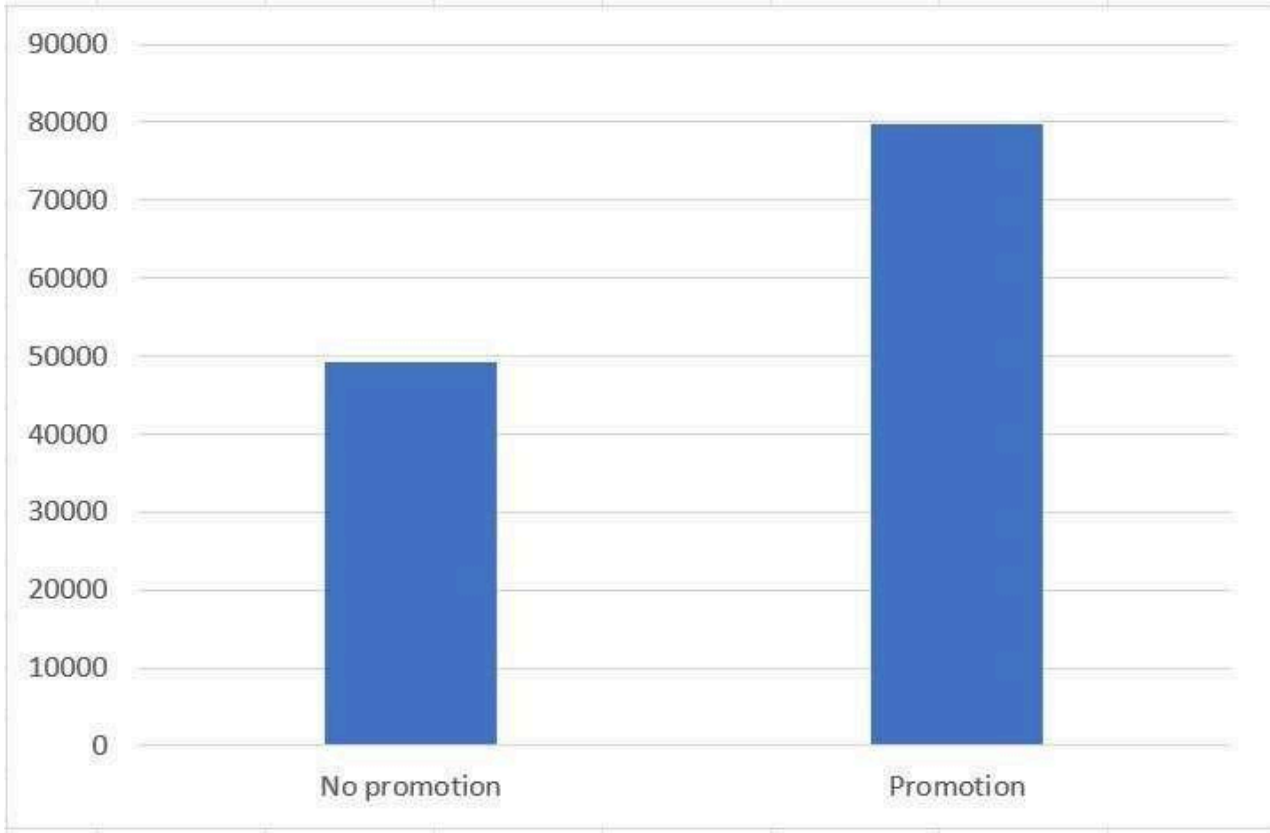

Process
Process
Validation
Training
Testing
Logistic Regression
Apply Model
Performance
tra
mod
exa
wei
thr
tes
unl
lab
per
ave

Visualizations

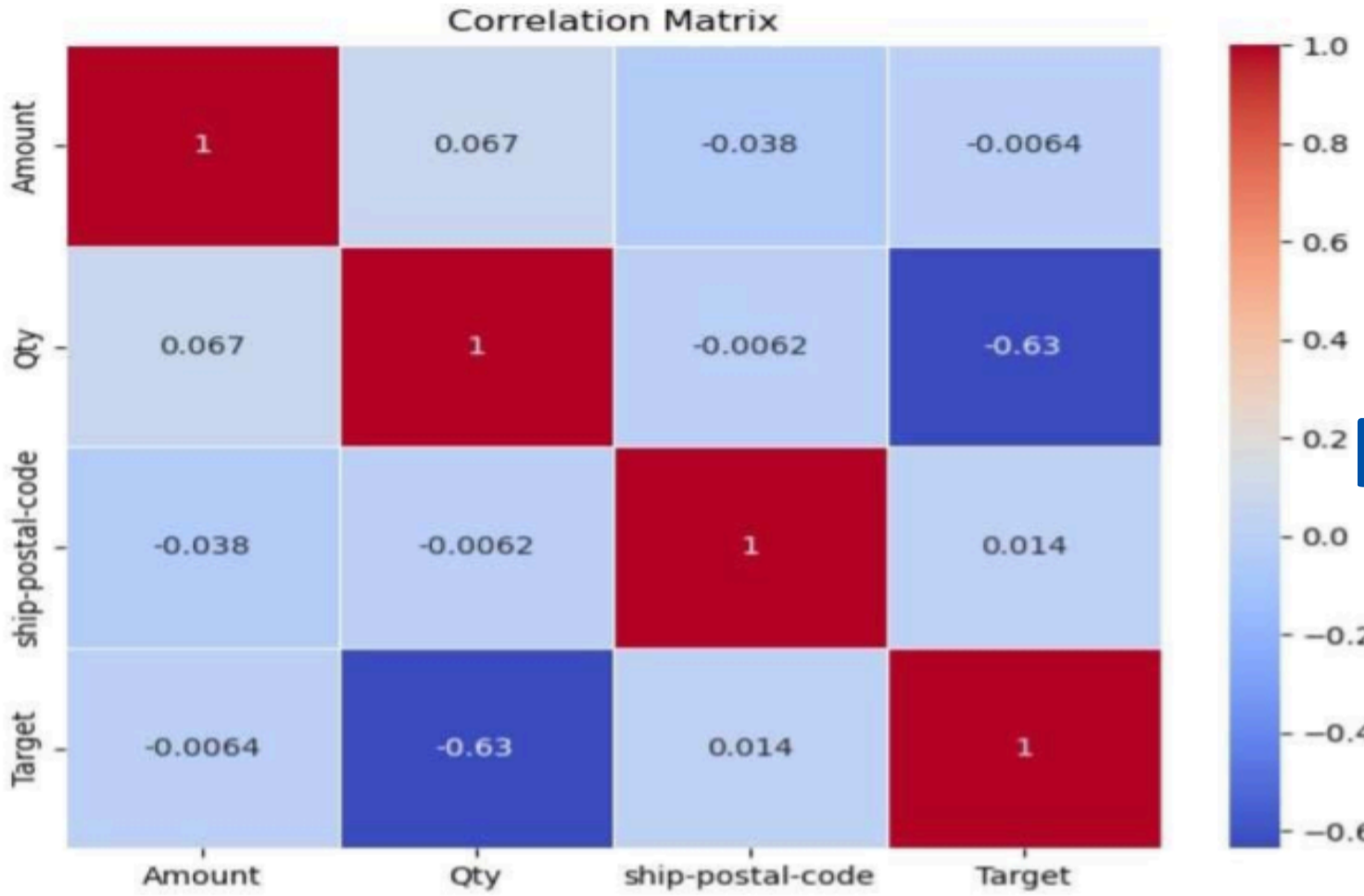


Visualizations





Correlation Analysis



Correlation Matrix:

	Amount	Qty	ship-postal-code	Target
Amount	1.000000	0.066820	-0.038195	-0.006401
Qty	0.066820	1.000000	-0.006157	-0.634904
ship-postal-code	-0.038195	-0.006157	1.000000	0.013871
Target	-0.006401	-0.634904	0.013871	1.000000

Regression Analysis:

	Feature	Coefficient
0	Amount	3.622917e-05
1	Qty	-8.418095e-01
2	ship-postal-code	1.885431e-08
Intercept	Intercept	8.661906e-01
Mean Squared Error	MSE	4.614133e-02
R-squared	R-squared	4.387695e-01

Challenges and Solutions

Interpretability of Predictions

- **Challenge:** Making the predictions understandable and actionable for merchants who may not have a technical background.
- **Solution:** Develop user-friendly dashboards and visualization tools that present the predictions and associated recommendations clearly.

Integration with Existing Systems

- **Challenge:** Seamlessly integrating predictive analytics into existing e-commerce platforms and workflows.
- **Solution:** Ensure compatibility with common e-commerce platforms like Amazon's seller central.

Implementation

Our analysis is useful for merchants who are selling products in e commerce platform amazon as it would help them in pricing the products

Best model and conclusion

- Regression, Correlation Analysis was the best model for prediction because the error rate and standard deviation is low with best accuracy

Future work

- Investigate using deep learning techniques
- Analyze customer behavior
- Cost sensitive learning

Thank

you