STEPHAN GRIMM

Predicting Sales

Using linear regression

Goal: Finding most impactful variables

- Sales data of all orders
- Build linear regression that predicts sales numbers
- Determine impactful variables

Raw dataset

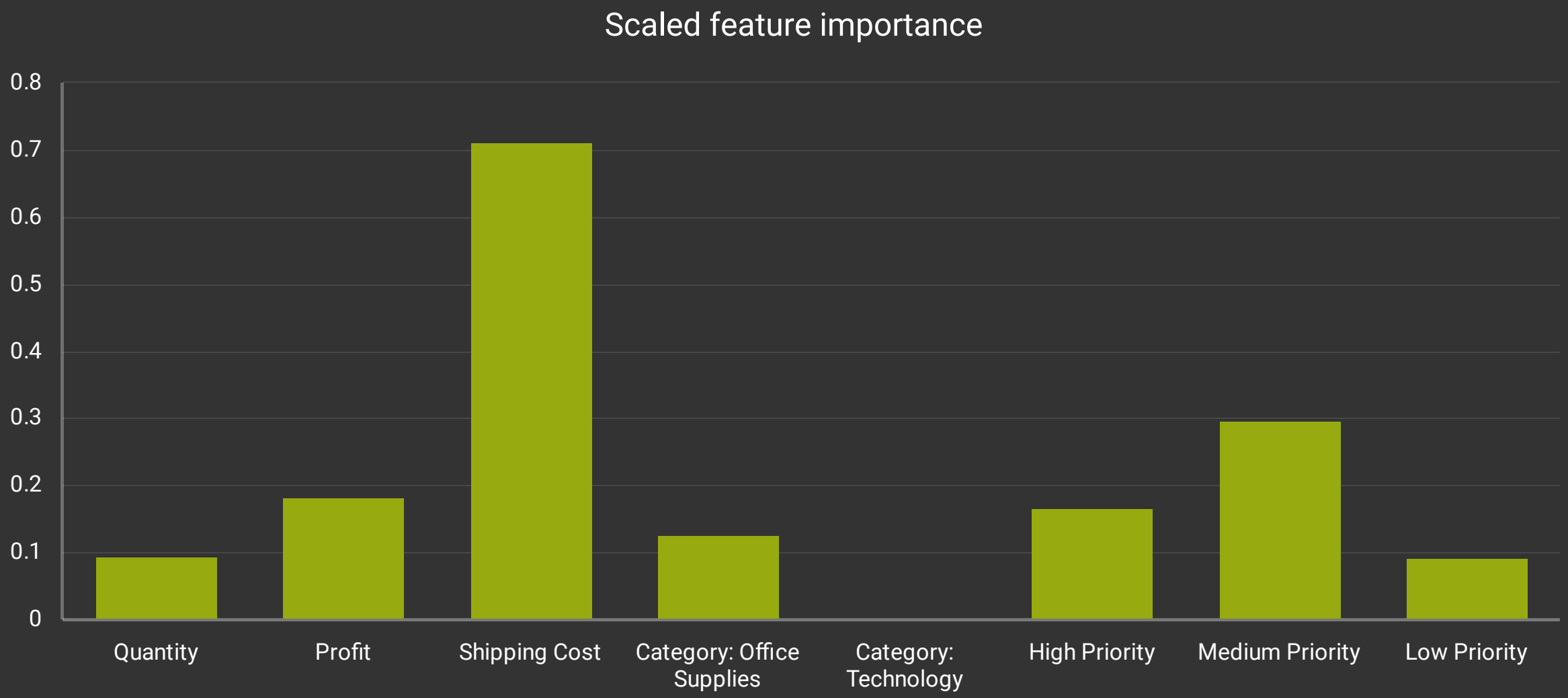
- 24 variables
- 51290 records

- Feature selection criteria:
  - Unique values < 17
  - Correlation with Sales > 0.2

5 variables remain after selecting for statistical significance

Variable	Correlation	# Unique
Quantity	0.31	-
Profit	0.48	-
Shipping Cost	0.76	-
Category	-	3
Order Priority	-	4

Shipping cost determines sales numbers!



Final sales formula

Sales =

0.095*(Quantity) + 0.182*(Profit) + 0.711*(ShippingCost)– 0.126*(Category OfficeSupplies) – 0.003*(Category Technology) + 0.167*(Order Priority High) + 0.093*(Order Priority Low) + 0.297*(Order Priority Medium)