

# WHY DOES AMAZON NEED OUR HELP?

by DALMIRO VENTURINI

# TARGET: LEAD-TIME

REDUCING CUSTOMER INSATISFACTION

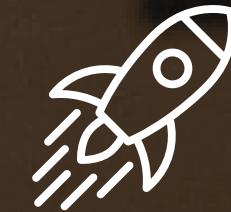
# DATASET OVERVIEW



E.E.U.U  
Puerto Rico



Sales  
2015 - 2017



Product Types  
118



Rows 185.500  
Columns 53

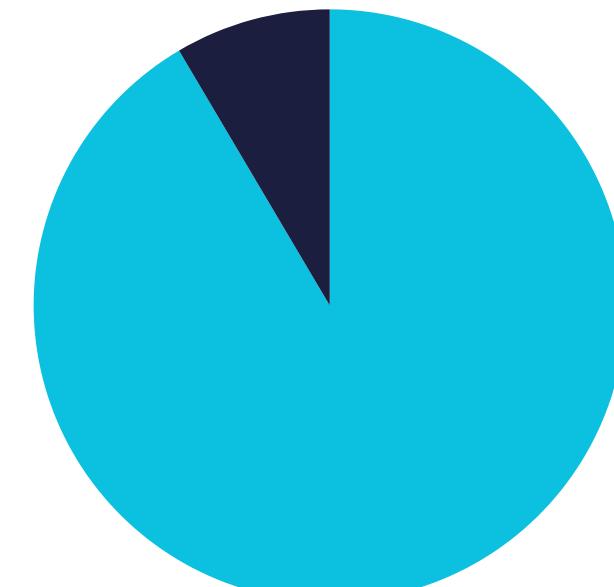
# TOO MANY PRODUCTS?

# CONTRIBUTION MARGIN

PARETTO RULE  $\approx$  80/20

Top 10 Products

8.5%

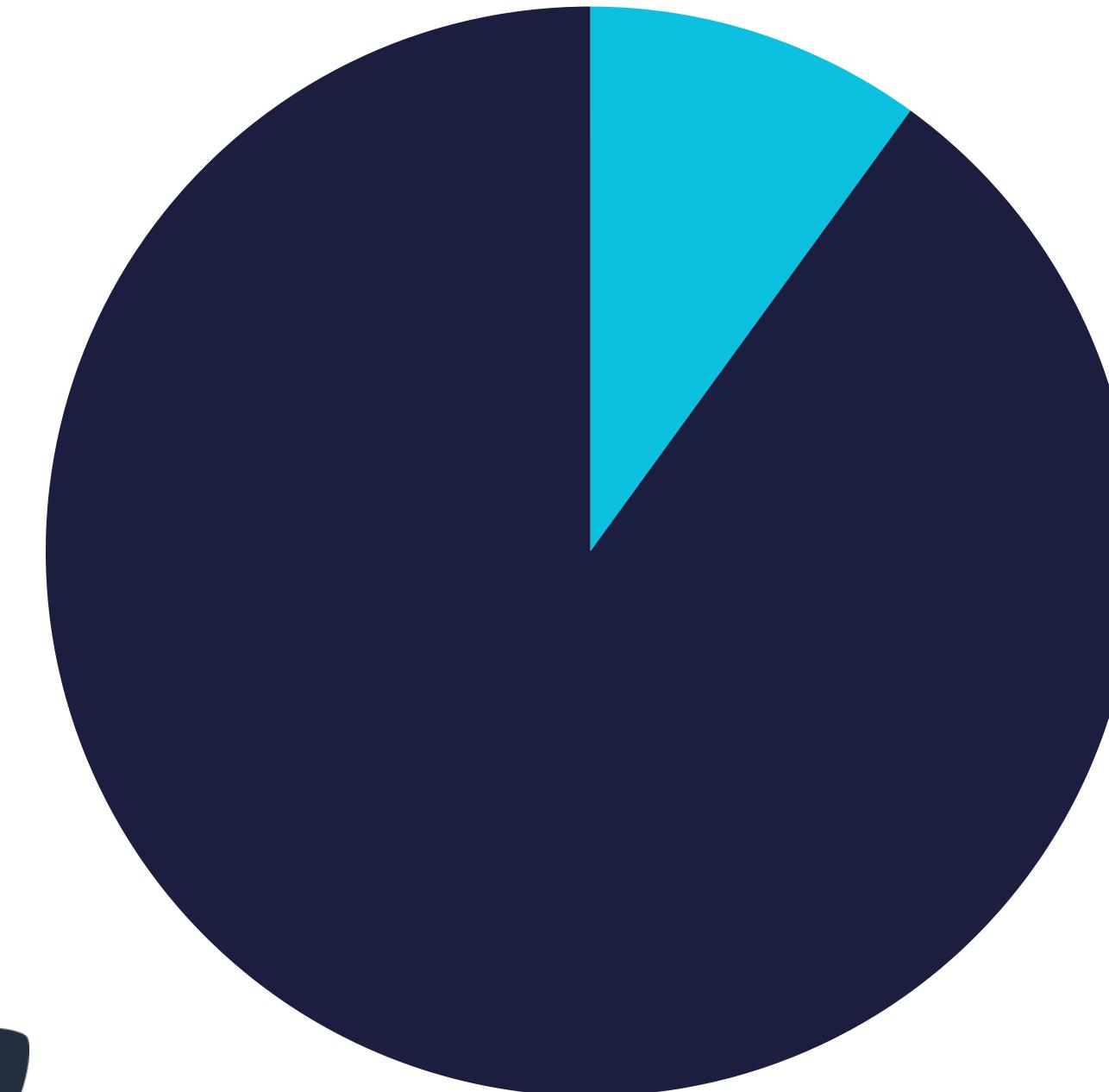


Others  
91.5%



U\$D 3.5M  
90%

U\$D 0.4M  
10%



# Top 10 Products



ID 1004

Field & Stream  
Sportsman 16 Gun  
Fire Safe



ID 365

Perfect Fitness  
Perfect Rip Deck



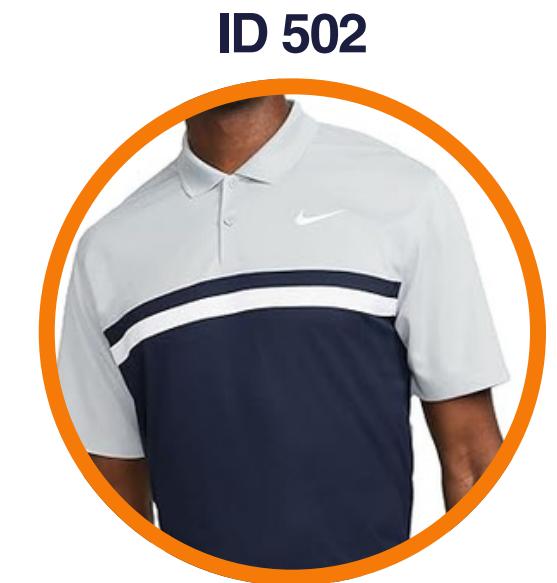
ID 957

Diamondback  
Women's Serene  
Classic Comfort Bi



ID 191

Nike Men's Free 5.0+ Running  
Shoe



ID 502

Nike Men's Dri-FIT Victory  
Golf Polo



ID 1073

Pelican Sunstream  
100 Kayak



ID 1014

O'Brien Men's  
Neoprene Life Vest



ID 403

Nike Men's CJ Elite  
2 TD Football Cleat



ID 627

Under Armour Girls' Toddler  
Spine Surge Runni



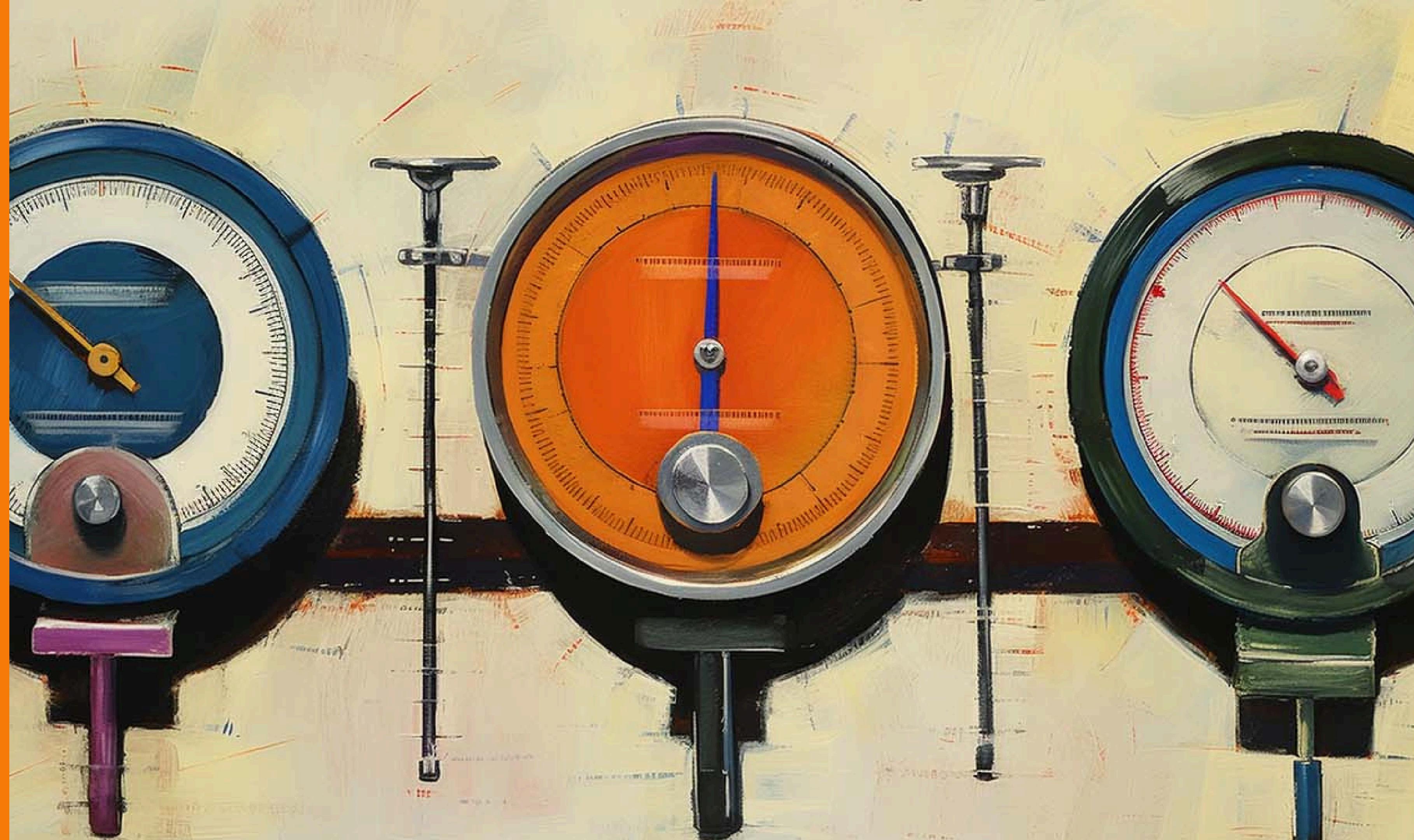
ID 1351

Dell Laptop

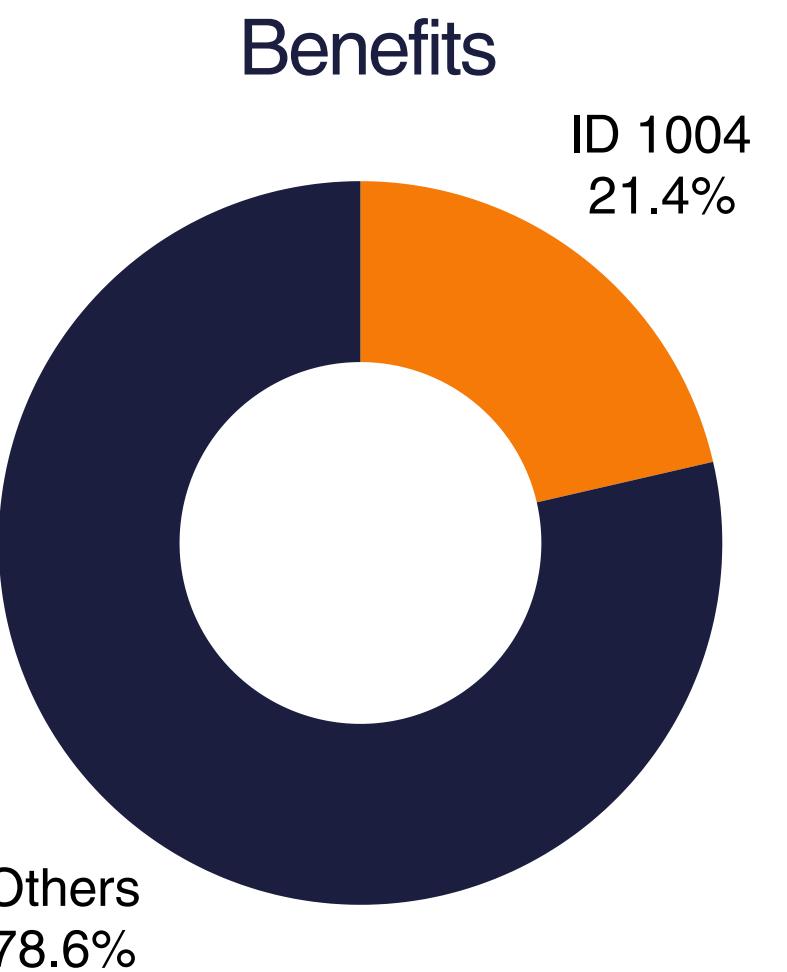
# amazon

# ?

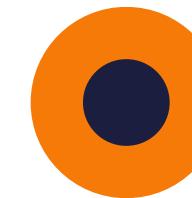




# PREDICTING LEAD TIME

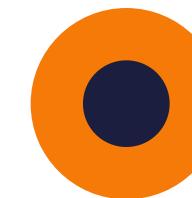


# PRE PROCESSING



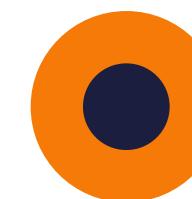
## Used Features

- payment\_type
- lead\_time
- late\_delivery\_risk
- customer\_country
- customer\_segment
- shipping\_mode
- shipping\_year
- shipping\_month
- shipping\_day
- shipping\_weekday



## Get Dummies

For Categorical Features



## X-y Split

y: df["lead\_time"]  
X: df.drop("lead\_time", axis=1)

## Nunique()

- 4 (CASH, DEBIT, TRANSFER, CREDIT)
- 4 (0, 1, 2, 3, 4)
- 2 (0, 1)
- 2 (EEUU, Puerto Rico)
- 3 (Consumer, Corporate, Corporate)
- 4 (Standard Class, First Class, Second Class, Same Day)
- 3 (2015, 2016, 2017)
- 12 (1 - 12)
- 31 (1 - 30)
- 7 (0 - 6)

# MODELING

with REGRESSION

Linear  
Regression

Tree  
Regression

Ridge  
Regression

KNN  
Regression

Lasso  
Regression

XGBoost  
Regression

Regression Models	R2	RMSE	MSE	MAE
Linear	1.00	0.00	0.00	0.00
Ridge	1.00	0.00	0.00	0.00
Lasso	-0.00	1.36	1.84	1.26
Decision Tree	1.00	0.00	0.00	0.00
<u>KNN</u>	<u>0.58</u>	<u>0.88</u>	<u>0.77</u>	<u>0.53</u>
XGBoost	1.00	0.00	0.00	0.00

NORMALIZING with MINMAXSCALER

R2 = 1.00   RMSE = 0.00   MSE = 0.00   MAE = 0.00

# CONCLUSIONS

## KEY TAKE AWAYS

- Shipping Modes significantly affect delivery times.
- Optimize shipping process for top10-performing products.
- Only the 8.5% of products generate the 90% of benefits.

## RECOMMENDATIONS

- Use the model to adjust delivery time promises based on shipping modes.
- KNN is an effective model for predicting delivery lead times with high accuracy..

# THANK YOU

---



ME WAITING FOR MY  
AMAZON PACKAGE TO ARRIVE



(IT'S  
BEEN 7 HOURS)