



# Dynamic Pricing

Milestone Presentation on Data Cleaning and Preliminary Modeling

## Presented By

Pomelo Wu  
Jiahua Wang  
Shiqi Liu

## Advised By

David Banks



# Agenda

**03** Milestone Summary

**04** Project Motivation & Background

**05** Dataset Overview

**06** Methodologies

**08** Preliminary Results

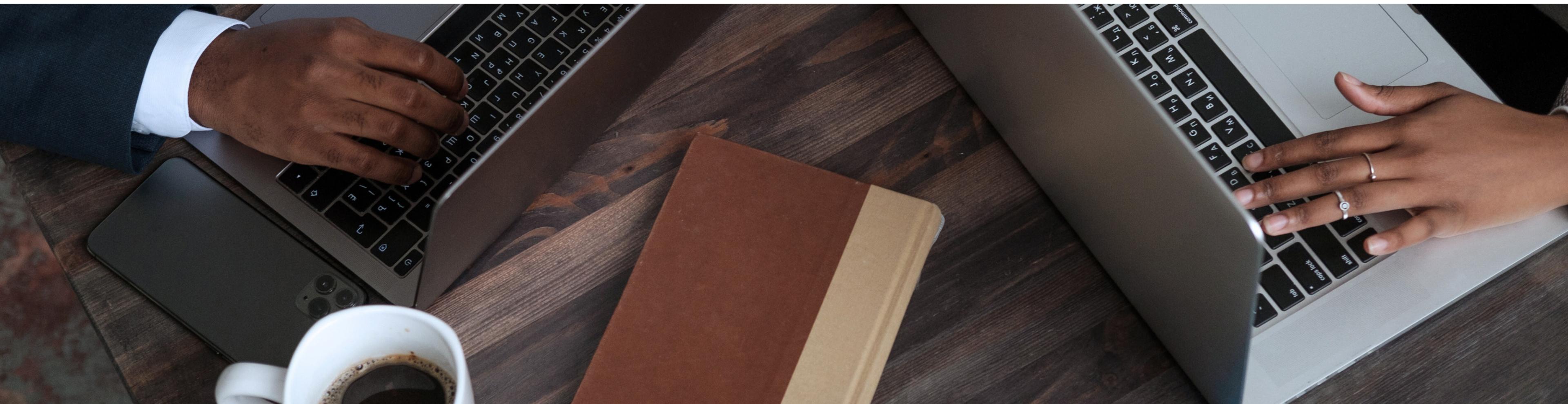
**11** Proposed Timeline

**12** Q&A

**13** Appendix



# Milestone Summary



- Gained familiarity of AMECO's current business tools such as Power Bi and SAP
- Built a clean dataset with augmented information
- Built a linear modeling incorporating time-series effect

»»

[Back to Agenda](#)

# Project Motivation & Background

- This project aims to develop a dynamic pricing model that responds to market fluctuations and customer loyalty information more effectively.
- The goal is to transform AMECO's pricing strategy ultimately enhancing profitability and customer relationships.

[Back to Agenda](#)

# Dataset Overview

[Back to Agenda](#)

Data columns (total 45 columns):			
#	Column	Non-Null Count	Dtype
0	X.	1155850	non-null object
1	Customer.No.	1155850	non-null object
2	Customer.Name	1155835	non-null object
3	DPM.Exempt	572301	non-null object
4	Customer.Category	590461	non-null object
5	Bill.To.Block	424715	non-null object
6	Bill.To.Street.No	13	non-null object
7	Bill.To.Street	569134	non-null object
8	Bill.To.City	571709	non-null object
9	Bill.To.State	571587	non-null object
10	Bill.To.Zip	571471	non-null object
11	Ship.To.Name	571546	non-null object
12	Address.2	260048	non-null object
13	Address.3	178734	non-null object
14	Ship.To.Block	437389	non-null object
15	Ship.To.Street	571564	non-null object
16	Ship.To.City	571754	non-null object
17	Ship.To.State	571691	non-null object
18	Ship.To.Zip	571724	non-null object
19	Invoice.No.	1155850	non-null object
20	Invoice.Date	1155850	non-null object
21	Invoice.Status	572301	non-null object
22	Invoice.Amt	572301	non-null object
23	Freight	572301	non-null object
24	Outside.Salesman	572301	non-null object
25	Inside.Salesman	491743	non-null object
26	Current.BP.Supply.Inside.Salesperson	564286	non-null object
27	Item.Group	572291	non-null object
28	Invoice.Profit.Center	572301	non-null float64
29	Profit.Center.Name	572301	non-null object
30	Product.Classification	572290	non-null object
31	Product.Classification.Description	572054	non-null object
32	Item.No.	1155840	non-null object
33	Item.Description	1155835	non-null object
34	UoM.Code	1155835	non-null object
35	Quantity	1155850	non-null object
36	Warehouse	572291	non-null float64
37	Stock.Status	691242	non-null object
38	Inventory.Cost	1155847	non-null object
39	Replacement.Cost	572301	non-null object
40	Price	1155849	non-null object
41	DPM.Factor	572301	non-null float64
42	DPM.Price	868869	non-null object
43	Vendor.No	562518	non-null object
44	Strikeforce.Flag	572301	non-null object

# Methodologies

[Back to Agenda](#)

## Data Cleaning

### Challenges

- Data System Transition  
PRE -> SAP
- Superfluous Information  
45 variables
- Missing Data  
50% missing
- Insufficient Information  
Inflation rate and competitor info

### Solutions

- Data Harmonization
- Variable Selection
- Handling Missing Data  
construction supplies  
welding equipment
- Data Augmentation

# Methodologies

[Back to Agenda](#)

Modeling Strategy

Train: 2016 - 2022 Test: 2023

Stage 1: Preliminary Linear Model

Stage 2: Iterative Refinement

# Results

## Data Cleaning

455978 rows with 12 Distinct Variables as Independent Variables

	Name	Category	State	Item	Group	Classification	UoM	Quantity	Stock	Cost	Price	Factor	DPMPrice	Date	Year	Month	Quarter	Inflation	Prev_Quarter	prev_quar_price
755043	WESCO - TVA INTEGRATED SUPPLY	Good	Gulf_Coast	GEC 647-DH326-1	Const Supplies CS/SD	CHECLE	EA	10.0	Y	3.980	5.32	0.0	0	2022-02-18	2022	2	2022Q1	6.4	2021Q1	5.404706
869456	KIEWIT POWER WHS 146	Good	Midwest	WMK 8SC24	Welding Equip WS/SD	SAFLEN	EA	16.0	Y	0.200	0.27	0.0	0	2020-04-09	2020	4	2020Q2	1.4	2019Q1	0.285407
859131	WESCO-DUKE INTEGRATED SUPPLY	Good	Mid-Atlantic	CEP 932-146-250	Welding Equip WS/SD	SAFLEN	EA	24.0	N	2.880	2.87	0.0	0	2020-04-06	2020	4	2020Q2	1.4	2019Q1	2.778842
849914	INDUSTRIAL RELIABILITY & REP.	Good	Midwest	LFA EN60X20	Const Supplies CS/SD	LIFRSL	EA	4.0	Y	44.600	72.95	0.0	0	2023-04-21	2023	4	2023Q2	5.5	2022Q2	60.645000
349946	UNITED GROUP SERVICES	Good	NaN	BWS 901-932-107-10	Welding Equip WS/SD	SAFLEN	EA	6.0	Y	0.872	1.50	NaN	1.25	2018-01-01	2018	1	2018Q1	1.8	2017Q1	1.973428
722295	WESCO - TVA INTEGRATED SUPPLY	Good	Midwest	LFA EE2801NFX3	Const Supplies CS/SD	LIFNSL	EA	5.0	N	6.850	7.63	0.0	0	2021-11-30	2021	11	2021Q4	4.9	2020Q4	7.503333
1103175	WOOD-MIZER LLC WHSE 500	Good	Midwest	LIN KP2742-1-62R-B25VEND	Welding Equip WS/SD	MIGNOZ	EA	1.0	Y	7.830	8.35	0.0	0	2021-04-01	2021	4	2021Q2	3.0	2020Q2	8.350000
275540	AGC MATERIALS	Good	NaN	GBU 202208	Const Supplies CS/SD	LADSTF	EA	1.0	Y	111.725	140.00	NaN	128.19	2017-01-01	2017	1	2017Q1	2.3	2016Q1	123.717194
508845	TQ CONSTRUCTORS INC.	Good	NaN	FTP 4511	Welding Equip WS/SD	SAFLEN	EA	12.0	Y	1.286	1.72	NaN	0	2019-01-01	2019	1	2019Q1	2.2	2018Q1	1.808554
546675	CENTRAL FABRICATORS	Good	NaN	CGW 45086	Welding Equip WS/SD	ABRCUT	EA	100.0	Y	0.700	1.00	NaN	NaN	2019-01-01	2019	1	2019Q1	2.2	2018Q1	0.954000

In this dataset we only chose customers in the ‘good’ category and unit of measure as ‘EA’

[Back to Agenda](#)

# Results

## Preliminary Modeling Results

### Linear Model

MSE: 0.052

R\_square: 0.98

OLS Regression Results						
Dep. Variable:	log_Price	R-squared:	0.980			
Model:	OLS	Adj. R-squared:	0.980			
Method:	Least Squares	F-statistic:	1.370e+06			
Date:	Mon, 11 Dec 2023	Prob (F-statistic):	0.00			
Time:	14:12:35	Log-Likelihood:	20677.			
No. Observations:	341868	AIC:	-4.133e+04			
Df Residuals:	341855	BIC:	-4.119e+04			
Df Model:	12					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.0602	0.003	20.945	0.000	0.055	0.066
C(Group) [T.Welding Equip WS/SD]	-0.0175	0.001	-20.119	0.000	-0.019	-0.016
C(Stock) [T.Y]	-0.0271	0.001	-27.353	0.000	-0.029	-0.025
C(State) [T.Central]	0.2624	0.003	80.631	0.000	0.256	0.269
C(State) [T.Gulf_Coast]	0.3000	0.003	108.850	0.000	0.295	0.305
C(State) [T.Mid-Atlantic]	0.2732	0.003	84.872	0.000	0.267	0.279
C(State) [T.Midwest]	0.3053	0.003	112.976	0.000	0.300	0.311
C(State) [T.Northeast]	0.3223	0.014	22.643	0.000	0.294	0.350
C(State) [T.Pacific]	0.2121	0.004	48.410	0.000	0.204	0.221
C(State) [T.VI]	-0.0128	0.006	-2.018	0.044	-0.025	-0.000
C(State) [T.West]	0.3183	0.004	85.824	0.000	0.311	0.326
log_Cost	0.9789	0.000	3962.445	0.000	0.978	0.979
Inflation	0.0008	0.000	3.357	0.001	0.000	0.001
Omnibus:	379557.011	Durbin-Watson:	1.803			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	78764127.061			
Skew:	5.419	Prob(JB):	0.00			
Kurtosis:	76.566	Cond. No.	155.			

[Back to Agenda](#)

# Results

Experiment with Random Forest

## Random Forest (Black box)

Group: 0.0006223548979141022

Stock: 0.0005569428739923778

log\_Cost: 0.00043116736752880084

State: 0.0004727349180723709

Inflation: 0.0007253158563097514

MSE: 0.0228

# Proposed Timeline

Plans for next semester



**Jan - Feb**

Incorporate other factors such as competitor information;  
Iterative Approach on Refining Pricing Model

**Feb - Mar**

Allow more flexible inputs, i.e., customer category, to set the prices on dashboard

**Mar - Apr**

TBA

**Apr - May**

Wrap up Work



[Back to Agenda](#)

# Q&A

Thank you



# Appendix

This is pre and post cleaning dataset overview.

Data columns (total 45 columns):			
#	Column	Non-Null Count	Dtype
0	X.	1155850	non-null object
1	Customer.No.	1155850	non-null object
2	Customer.Name	1155835	non-null object
3	DPM.Exempt	572301	non-null object
4	Customer.Category	590461	non-null object
5	Bill.To.Block	424715	non-null object
6	Bill.To.Street.No	13	non-null object
7	Bill.To.Street	569134	non-null object
8	Bill.To.City	571709	non-null object
9	Bill.To.State	571587	non-null object
10	Bill.To.Zip	571471	non-null object
11	Ship.To.Name	571546	non-null object
12	Address.2	260048	non-null object
13	Address.3	178734	non-null object
14	Ship.To.Block	437389	non-null object
15	Ship.To.Street	571564	non-null object
16	Ship.To.City	571754	non-null object
17	Ship.To.State	571691	non-null object
18	Ship.To.Zip	571724	non-null object
19	Invoice.No.	1155850	non-null object
20	Invoice.Date	1155850	non-null object
21	Invoice.Status	572301	non-null object
22	Invoice.Amt	572301	non-null object
23	Freight	572301	non-null object
24	Outside.Salesman	572301	non-null object
25	Inside.Salesman	491743	non-null object
26	Current.BP.Supply.Inside.Salesperson	564286	non-null object
27	Item.Group	572291	non-null object
28	Invoice.Profit.Center	572301	non-null float64
29	Profit.Center.Name	572301	non-null object
30	Product.Classification	572290	non-null object
31	Product.Classification.Description	572054	non-null object
32	Item.No.	1155840	non-null object
33	Item.Description	1155835	non-null object
34	UoM.Code	1155835	non-null object
35	Quantity	1155850	non-null object
36	Warehouse	572291	non-null float64
37	Stock.Status	691242	non-null object
38	Inventory.Cost	1155847	non-null object
39	Replacement.Cost	572301	non-null object
40	Price	1155849	non-null object
41	DPM.Factor	572301	non-null float64
42	DPM.Price	868869	non-null object
43	Vendor.No	562518	non-null object
44	Strikeforce.Flag	572301	non-null object



Data columns (total 20 columns):			
#	Column	Non-Null Count	Dtype
0	Name	365534	non-null object
1	Category	365534	non-null object
2	State	365534	non-null object
3	Item	365534	non-null object
4	Group	365534	non-null object
5	Classification	365534	non-null object
6	UoM	365534	non-null object
7	Quantity	365534	non-null float64
8	Stock	365534	non-null object
9	Cost	365534	non-null float64
10	Price	365534	non-null float64
11	Factor	365534	non-null float64
12	DPMPrice	365534	non-null object
13	Date	365534	non-null datetime64[ns]
14	Year	365534	non-null int64
15	Month	365534	non-null object
16	Quarter	365534	non-null period[Q-DEC]
17	Inflation	365534	non-null float64
18	Prev_Quarter	365534	non-null period[Q-DEC]
19	prev_quar_price	365534	non-null float64