Customer Segmentation
Individual Final Readout

TEAM
Adam Cespedes-Zaman
Tom Cerny
Jose Garcia
Alvira Khan
Rohit Kundurthi
Brandon Morgeson
Jahnavi Rati
Mary Elizabeth Ward

Fall 2022

Planta Cartagena

Contents

ARGOS

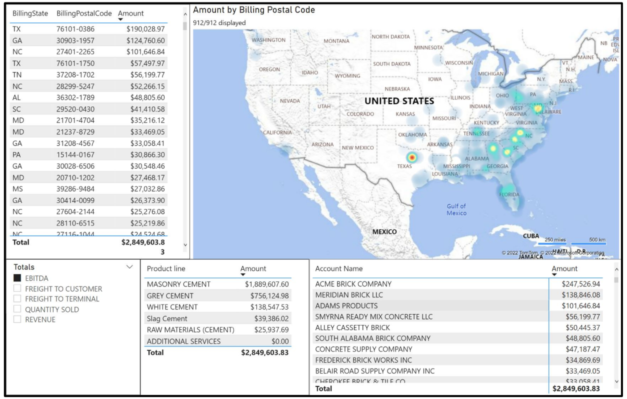
Executive Summary

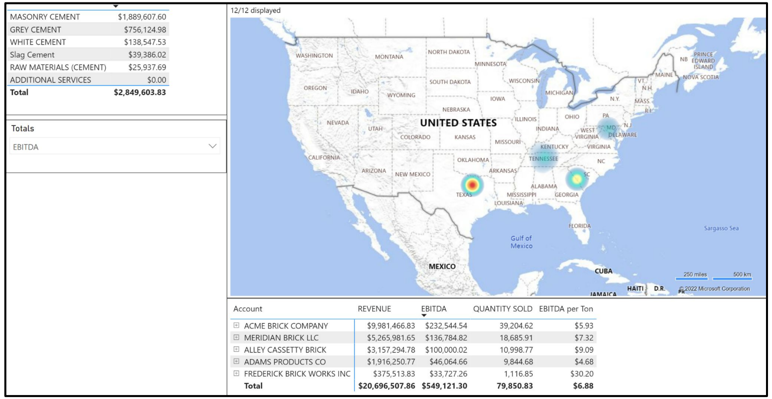
- During the 10/16 and 11/9 updates, Argos reiterated their desire to segment customers via a **qualitative, behavioral-focused approach.**
    - Specific customer attributes, such as EBITDA, Shipping Conditions, and Incoterms were identified by Argos as key data points to prioritize in segmentation.
- A new dataset was provided on 11/17, **requiring a new segmentation methodology** given material changes in the dataset. The new dataset was best segmented via **K-Modes Clustering Analysis.**
    - In this approach, key customer attributes (e.g., Incoterms, Shipping Conditions, Product, etc.) are analyzed through an algorithmic approach.
    - Algorithm creates Clusters based on shared characteristics across the dataset; specifically, the ***distance*** of each datapoint from the centroid of created clusters. A datapoint close to the centroid of a particular cluster therefore shares the same characteristics of other data points close to the centroid.
- Clustering Analysis identified 2 distinct customer segments.
    - Both segments have roughly similar number of transactions and monthly profit in Total EBITDA, and cover similar states, products, and materials.
    - However, the segments **differ** significantly in ship-to cities, shipping conditions and incoterms.
    - This implies there is a distinct set of attributes that provide Argos higher value (Customers who ship to primarily the Southeast vs. Atlantic and Midwest whose preferred shipping conditions are Delivery via Road and Rail).
    - These customers would benefit from targeted promotional pricing strategies and direct sales strategies based on shipping modality. Additionally, Argos' transportation contracts can be adjusted to align with the locations where delivery is preferred.

Project Focus

Following the October 16 meeting with Argos, the team began to focus on:

- **EBITDA over Revenue**
  - Began to assess profitability by measuring Total EBITDA by customer and product line
- **Pickup location** with regard to terminal setup
- Contextualized whether certain customer groups exist in certain geographic locations that can be segmented in the short term
  - By analyzing these geographic hotspots, the team can determine if they reflect specific product types or Incoterms
- Began to map and observe key customer hotspots using Power BI





ARGOS

Project Schedule

Week 1 & 2

Project Overview & Sponsor Research

Team formed structure and key roles for successful 2022 Capstone Project.

Week 3&4

Data Exploration & Storyboard Development

Team began to build structure to data set supplied from Argos in September 2022.

Key Programs: Excel, R, Python

Week 5&6

Client Presentation & Recap

First Client Presentation, asked to focus on EBITDA per Ton per perspective of client.

Key Programs: Excel, R, Python

Week 7&8

Continued to Build on Data & Sent Update to Client

Proof of Concept sent to Argos for signature, building on final product within Power BI

Key Programs: Excel, R, Python, Power BI

Week 9&10

Client Meeting w/ Pivots to New Data Set

Client sent new data set with introduction of TOTAL EBITDA and updated figures that removed duplicates.

Key Programs: Excel, R, Python, Power BI

Week 11

Data Analysis/Final Presentation Prep

Due to data pivot and timing, team updated Python and will deliver discoveries as planned on Nov. 30.

Key Programs: Excel, R, Python, Power BI

Analysis

Data Exploration

- Significant outliers in the data mostly under no assigned State and Ship to being Otros
- The boxplot gives a visual representation of the outliers
- The table shows the top upper limit outlier of $5,348,380
- The df.describe() function shows the mean of $5,182 while the df_noOutliers.describe() shows a mean of $4,825

```
df.describe()
```

	Cluster	Monthly Profit
count	89939.000	89939.000
mean	1.072	5182.141
std	0.894	43797.181
min	0.000	-1581596.290
25%	0.000	-3.850
50%	1.000	687.330
75%	2.000	3479.435
max	2.000	5348380.090

```
df_noOutliers.describe()
```

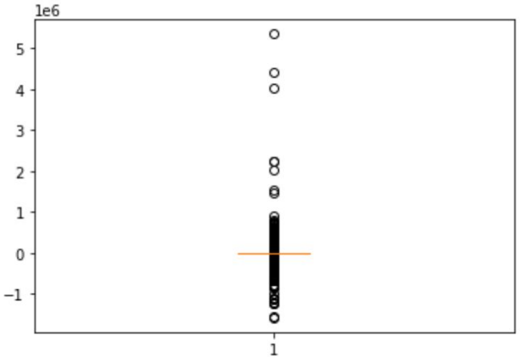
	Cluster	Monthly Profit
count	89939.000	89939.000
mean	0.456	4825.579
std	0.498	18524.969
min	0.000	-126209.402
25%	0.000	-3.850
50%	0.000	687.330
75%	1.000	3479.435
max	1.000	136573.683

```
df[["State","Ship to name", "Monthly Profit"]].sort_values(by=["Monthly Profit"]
```

	State	Ship to name	Monthly Profit
126214	Not assigned	OTROS	5348380.090
126209	Not assigned	OTROS	4406280.590
126373	Not assigned	OTROS	4032189.540

```
plt.boxplot(y)
```

```
{'whiskers': [<matplotlib.lines.Line2D at 0x7f82e01322e0>,
  <matplotlib.lines.Line2D at 0x7f82e0132670>],
 'caps': [<matplotlib.lines.Line2D at 0x7f82e0132a00>,
  <matplotlib.lines.Line2D at 0x7f82e0132d90>],
 'boxes': [<matplotlib.lines.Line2D at 0x7f82e011af10>],
 'medians': [<matplotlib.lines.Line2D at 0x7f82e012b160>],
 'fliers': [<matplotlib.lines.Line2D at 0x7f82e012b4f0>],
 'means': []}
```



Data Exploration

- Upon receiving the new data and based on discussion with Argos, we filtered for EBITDA values only to explore the revenue minus operating costs
- Once we filtered the new data, we ran simple queries and data exploration to understand top profiting states and top profiting product line

Run Queries

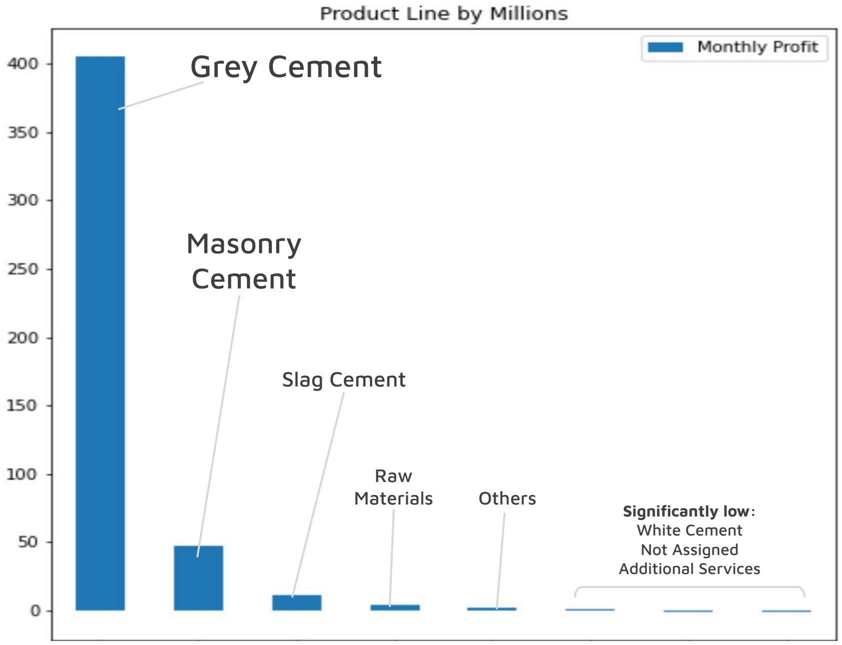
Top 5 Profiting States

```
pd.set_option('display.float_format', lambda x: '%.3f' % x)
df[["State", "Monthly Profit"]].groupby(["State"]).sum().sort_values(by=["Monthl
```

	Monthly Profit
State	
Florida	80090052.970
North Carolina	79554730.170
Georgia	76606967.970
South Carolina	57457310.160
Alabama	31412198.370

```
_df_ebitda.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 89939 entries, 524 to 462656
Data columns (total 9 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   State           89939 non-null  object
 1   Product line    89939 non-null  object
 2   Material        89939 non-null  object
```

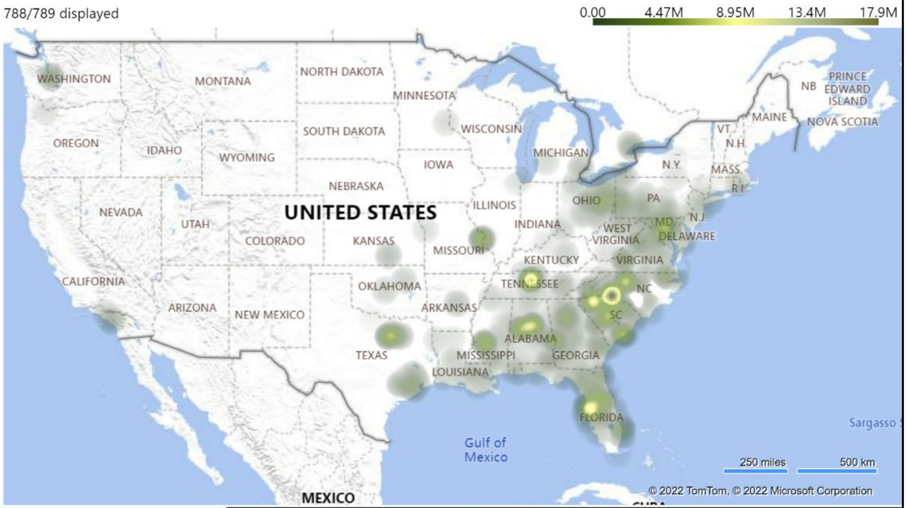


Data Exploration - "EBITDA by State Heat Map"

BillingState	TOTAL EBITDA
NC	$29,912,524.06
FL	$27,026,978.73
SC	$27,013,134.59
AL	$15,810,685.66
GA	$12,986,021.96
TN	$12,152,235.83
TX	$11,592,197.66
MD	$8,645,750.00
OH	$7,707,771.00
VA	$7,647,654.85
	$7,243,674.48
MS	$6,247,193.50
PA	$5,289,626.98
MO	$4,847,040.78
WA	$1,302,753.82
LA	$1,149,985.38
WV	$974,698.77
ON	$916,732.47
CA	$676,533.48
Total	**$191,111,313.37**

Amount by Billing Postal Code

788/789 displayed



Revised Fiscal Pe...

1/1/2021

12/1/2021

Transaction Type

- F
- H

Product line	TOTAL EBITDA
GREY CEMENT	$151,060,790.57
MASONRY CEMENT	$23,989,466.01
Others	$12,049,577.92
Slag Cement	$3,513,708.51
RAW MATERIALS (CEMEN	$1,963,278.46
WHITE CEMENT	$463,434.61
ADDITIONAL SERVICES	($935,508.42)
Not assigned	($993,434.29)
Total	**$191,111,313.37**

Account Name	TOTAL EBITDA
SMYRNA READY MIX CONCRETE LLC	$10,591,207.58
CONCRETE SUPPLY COMPANY	$8,134,910.76
	$7,243,674.48
CONCRETE SUPPLY COMPANY - RAIL ONLY	$7,175,691.28
CEMEX INC	$5,625,599.18
BAMA CONCRETE PRODUCTS CO INC	$5,448,443.70
MISSISSIPPI LIME CO	$4,819,812.43
DELTA INDUSTRIES, INC.	$4,407,906.83
ACME BRICK COMPANY	$4,388,625.28
ADAMS PRODUCTS (GREENSBORO)	$3,587,689.84
METROMONT LLC	$3,133,304.22
Total	**$191,111,313.37**

Key Takeaways

- Using *latest* Profitability 2021 data, we measured "EBITDA" across different regions to determine where most cement sales derive from
- Grey cement generates most EBITDA amongst all product lines throughout Southern Atlantic states
- Top customer/location drivers:
  - Carolinas
  - Florida
  - Alabama
- Top customer/location drivers by billing state:
  - Smyrna
  - Concrete Supply Company
  - Cemex

ARGOS

Correlation Matrix

- **Strategy**: Use a correlation matrix to identify significant features influencing "Monthly Profit" or unit of measure Total EBITDA and to proceed with clustering based on correlations
- Correlation matrix below with **most appropriate and desired qualitative features** from the Profitability dataset

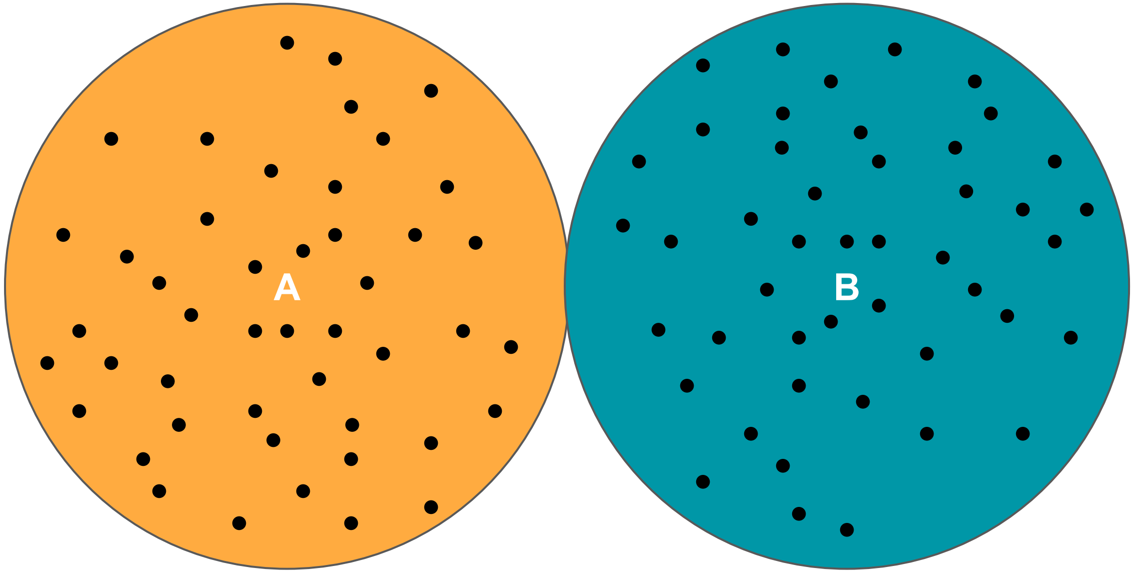
- **Most influential**: "Ship to name" and "Customer name"
- **Moderately influential**: "Material", "City Ship to", "Shipping Conditions" and "Product Line"
- **Least influential**: "State" and "Incoterms"

	State	Product line	Material	Ship to name	City ship to	Incoterms	Shipping Conditions	Customer name	Monthly Profit
State	1.00	0.40	0.33	0.98	0.74	0.28	0.33	0.99	0.15
Product line	0.40	1.00	1.00	0.72	0.54	0.25	0.29	0.71	0.24
Material	0.33	1.00	1.00	0.30	0.24	0.63	0.48	0.30	0.36
Ship to name	0.98	0.72	0.30	1.00	0.67	0.74	0.72	0.99	0.48
City ship to	0.74	0.54	0.24	0.67	1.00	0.66	0.67	0.68	0.35
Incoterms	0.28	0.25	0.63	0.74	0.66	1.00	0.75	0.74	0.15
Shipping Conditions	0.33	0.29	0.48	0.72	0.67	0.75	1.00	0.69	0.25
Customer name	0.99	0.71	0.30	0.99	0.68	0.74	0.69	1.00	0.47
Monthly Profit	0.15	0.24	0.36	0.48	0.35	0.15	0.25	0.47	1.00

Segmentation Methodology

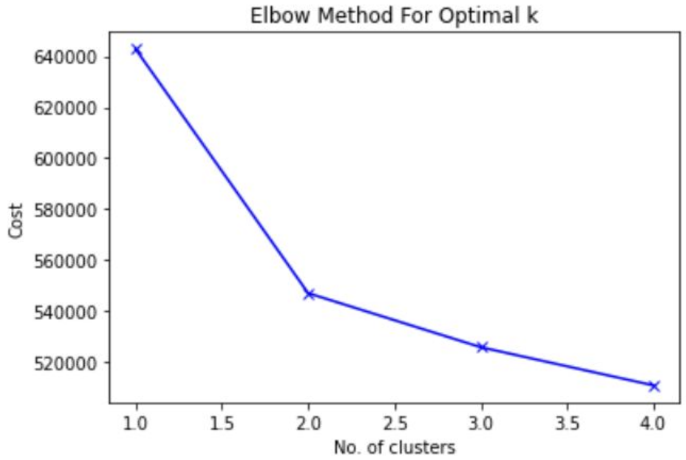
Define key customers characteristics & attributes:

- EBITDA
- Incoterms
- Shipping Conditions
- Product
- Material
- State
- Ship-to City
- Ship-to Name
- Customer Name



Steps to Output

- Using Kmodes clustering algorithm, we identified two optimal number of clusters
- It runs multiple iterations placing centroids in the data and calculates how many moves it takes to get to the centroid
- Using the elbow curve we look for the largest break - in this case two
- Finally, we built the cluster model and assigned each row to its own cluster




```
# #Elbow curve to find optimal K
from kmodes.kmodes import KModes

cost = []
K = range(1,5)
for num_clusters in list(K):
    kmode = KModes(n_clusters=num_clusters, init = "random", n_init = 5, verbose=1)
    kmode.fit_predict(df_noOutliers)
    cost.append(kmode.cost_)

plt.plot(K, cost, 'bx-')
plt.xlabel('No. of clusters')
plt.ylabel('Cost')
plt.title('Elbow Method For Optimal k')
plt.show()
```

```
Init: initializing centroids
Init: initializing clusters
Starting iterations...
Run 1, iteration: 1/100, moves: 0, cost: 642853.0
Init: initiali
Init: initiali
Starting itera
Run 2, iterat
```

```
#Building the model with 2 clusters
from kmodes.kmodes import KModes
kmode = KModes(n_clusters=2, init = "random", n_init = 5, verbose=1)
clusters = kmode.fit_predict(df_noOutliers)
clusters
```

```
Init: initializing centroids
Init: initializing clusters
Starting iterations...
Run 1, iteration: 1/100, moves: 8520, cost: 546935.0
Run 1, iteration: 2/100, moves: 1781, cost: 546935.0
Init: initializing centroids
Init: initializing clusters
Starting iterations...
Run 2, iteration: 1/100, moves: 5106, cost: 546935.0
Run 2, iteration: 2/100, moves: 722, cost: 546935.0
```

Application

Current Segmentation

Customer Understanding

- Argos has 3 types of customers, which all have impact on the **overall pricing strategy**
- To better understand the customers Argos services and what generates the most **EBITDA and largest projects**, it is critical to understand these segments and include these aspects in the segmentation model
- Largest pool of Argos customers are price-driven, so the need for a new segmentation model is critical to pivot to a new segmentation strategy that will be impactful in driving profitability and enticing more prospective customers

Argos Customer Types

Performance Driven	Relationally Driven	Price Driven

ARGOS

Segmentation Output and Results - Cluster Overview

	Cluster A : Pickup	Cluster B : Delivery
No. of Transactions	48,024 *(53%)*	41,915 *(47%)*
Total Monthly EBITDA	$184M *(42%)*	$250M *(58%)*
Top 5 Profitable States	1. South Carolina *(15%)* 2. Florida *(11%)* 3. Alabama *(11%)* 4. Georgia *(11%)* 5. North Carolina *(10%)*	1. Florida *(24%)* 2. North Carolina *(21%)* 3. Georgia *(19%)* 4. South Carolina *(11%)* 5. Tennessee *(6%)*
Top 3 Materials Bought	1. Type I/II Gray Bulk *(35%)* 2. Type I Gray Bulk *(30%)* 3. Type III Gray Bulk *(7%)*	1. Type I Gray Bulk *(54%)* 2. Type I/II Gray Bulk *(17%)* 3. EcoStrong PLC - Type IL Gray Bulk *(12%)*
Top 5 Ship-To City	1. Charlotte *(3%)* 2. Charleston *(2%)* 3. Ocala *(2%)* 4. Canton *(2%)* 5. Lexington *(1%)*	1. Charlotte *(4%)* 2. Atlanta *(3%)* 3. Jacksonville *(2%)* 4. Orlando *(2%)* 5. Doraville *(2%)*
Top 3 Incoterms	1. Pickup *(94%)* 2. Not Assigned 3. Delivery	1. Delivery *(60%)* 2. Not Assigned 3. Pickup
Top 3 Shipping Conditions	1. Pickup Road *(98%)* 2. Delivery Road 3. Delivery Rail	1. Delivery Road *(89%)* 2. Delivery Rail 3. Pickup Road

Roughly similar characteristics

Different characteristics

Source: prof_2021_2020 SR, provided 11/17

ARGOS

Implications of New Segmentation

EBITDA generation by clusters: Customers who prefer pickup generate less monthly EBITDA as 42% of the TOTAL EBITDA, although the EBITDA in each state and city is slightly more spread out.

Incoterms for Marketing: The breakdown of the cluster allows for Argos to target their customers using the incoterms. Cluster A has 94% of orders that are picked-up. This gives Argos a chance to target Cluster A based on Pick-up messaging.

Specialized Transportation Agreements: Transportation costs fluctuate significantly based on demand, which is difficult to predict. The best way to optimize fluctuations in transportation costs is to commit to a minimum weekly load in locations where consistent delivery is known - this can drive down freight costs and build relationships with FTL (full truckload) / LTL (less than truckload) truckers who can then better accommodate spikes in demand. There are clearly defined areas where Argos customers prefer delivery where transportation agreements will be useful.

Data Suggestions

- **Data Intake**: By streamlining the customer intake form, you can eliminate the “Not Assigned” or blanks throughout the data.
    - Ensure data is entered into the system accurately and holistically, reducing human error.
    - Key Areas to Improve: Incoterms, States, City, Zip, Shipping Conditions, Ship to Name

- **Naming Conventions**:
    - Standardize the naming conventions across the different fields throughout data sets.
    - Formalize field integration to solidify correct categories are not misrepresented in different areas of the data.

ARGOS

Implementation & KPIs

Implementation - Sales Process Design

Argos has requested segmentation to be done using “qualitative” descriptors. However, currently these descriptors are currently not captured in the account/sales data.

In order to capture this - implementing sales process changes will be necessary.

Stage 1	Stage 2	Stage 3
Sales Opportunity created	**Negotiation & Terms**	**Opportunity Close**
Seller qualifies opportunity. Non-qualitative data asked for stage advancement. Connecting with all the relevant stakeholders internal and externally.	Commercial information Size of deal (quantity/pricing) Align on close dates	Contracts sent out Final negotiations Signatures Deal Close

Implementation of Insights

The defined **customer segmentation** will then need to be implemented across internal systems at Argos, as well as communicated company wide.

Customer Segmentation

CRM implementation

- Houses all the sales opportunity data
- Segments customer / prospect accounts
- Create sales territories
- Create sales reports/dashboards

Finance / FP&A Team

- Using defined segmentation data & historical analysis, reconcile historical data to fit within new segmentation.
- Break down company growth/sales targets into segment level goals.
- Allocate financial resources towards high success vs. high growth potential segments.

Sales Team

- Redefine sales territories/patches
- Understand sales targets vs. sales headcount/quota capacity.
- Hire/reallocate sales resources based on redefined territories.

Marketing Team

- Create new campaigns targeting newly defined "ICP" (Ideal customer profiles).
- Track new open/click-through/conversion rates based on new segmentation.
- Reallocate marketing resources to digital/physical marketing & region based.

KPIs

Revenue Systems team (CRM)

- Automated segmentation of accounts
- 100% of existing accounts segmented based on new definition
- Clear sales territories (potentially defining an RoE)

Sales Team

- $ bookings
- Conversion rates on new segmentation prospects
- Average contract value
- Average sales cycle
- Segment penetration

Finance / FP&A Team

- Restated historical financials
- Sales/revenue targets based on segmentation
- Reallocation of financial resources

Marketing Team

-Open/click-through/conversion rates based on new segmentation.

- CAC / LTV measurements
- Cost per opportunity

Thank you!