

---

---

# Apsey Farms: Historical Sales Order Analysis

— Josh Singer —

Created for: Apsey Farms

---

---

# Covered in this deck

- Background, Objectives & Methodology
- Analysis of Customers & Product Orders

# Background & Methodology

# Background

The dataset provided contained 3,787 line items, where each row corresponded to one product ordered (note that multiple lines (products) may make up one order). Some of the relevant features (columns) in the dataset are included below:

Feature	Description
Name	Unique identifier for the order (order number)
Email	Customer email address & customer identifier
Subtotal	Total cost of the order (sum of the line item prices for the order minus any discounts)
Created at	Date and time the order was processed
Lineitem quantity	Number of items (products) ordered. Depending on the product, the quantity value may be defined by weight (lbs) (e.g. for bulk items) or number of items.
Lineitem name	Name of the product ordered
Lineitem price	Price of the line item (product)
Shipping City	Customer's delivery city
Shipping Zip	Customer's delivery zip code
Shipping Province	Customer's delivery state

# Objectives & Methodology

## Purpose & Objectives:

- The goal of this project was to clean and analyze sales order data to help Apsey Farms better understand their customer base and sales trends between different product offerings.

## Methodology:

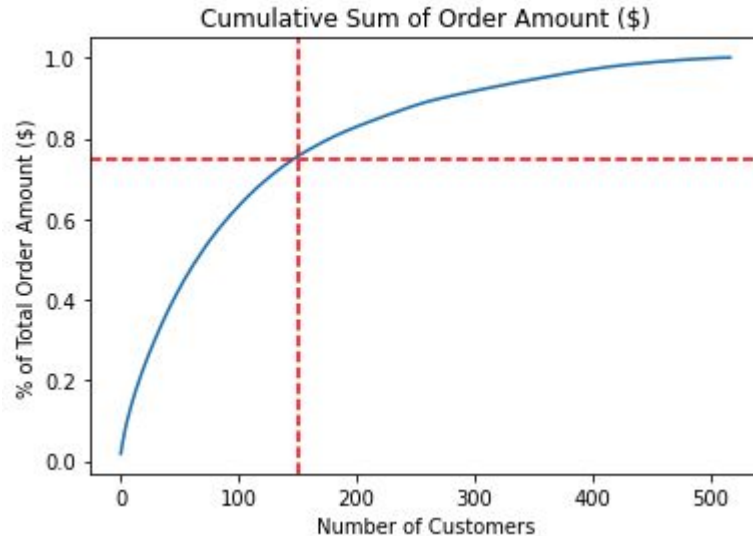
- Dataset contained 3,787 line items, where each row corresponded to one product ordered. Note that multiple lines (products) may make up one order
- Data cleaning/prep activities included:
  - Standardized products and dimensions (enterprise, product type, quantity/weight) into a single, comprehensive crosswalk/lookup table (dataset included 350 different products) using techniques including:
    - Web scraping to collect standard product names and quantities from apseyfarms.com
    - Fuzzy string matching to impute values for ill-defined/unknown products
    - Character extraction with regular expressions (regex)
  - Removed “non-legitimate” orders (noise) such as non-product sales (e.g. farm tours) and refunded/canceled orders
- Used the standardized product crosswalk to analyze and identify trends with customers and products

# Customer Analysis

# Understanding the Customer Base

Customer Statistics*	
Total Number of Unique Customers	Removed for sensitive info
Total Number of Recurring Customers	
Recurring Customers % of Total	74.2%

Order Statistics*	
Total Number of Orders	Removed for sensitive info
Total Number of Recurring Orders	
Recurring Orders % of Total	53.7%

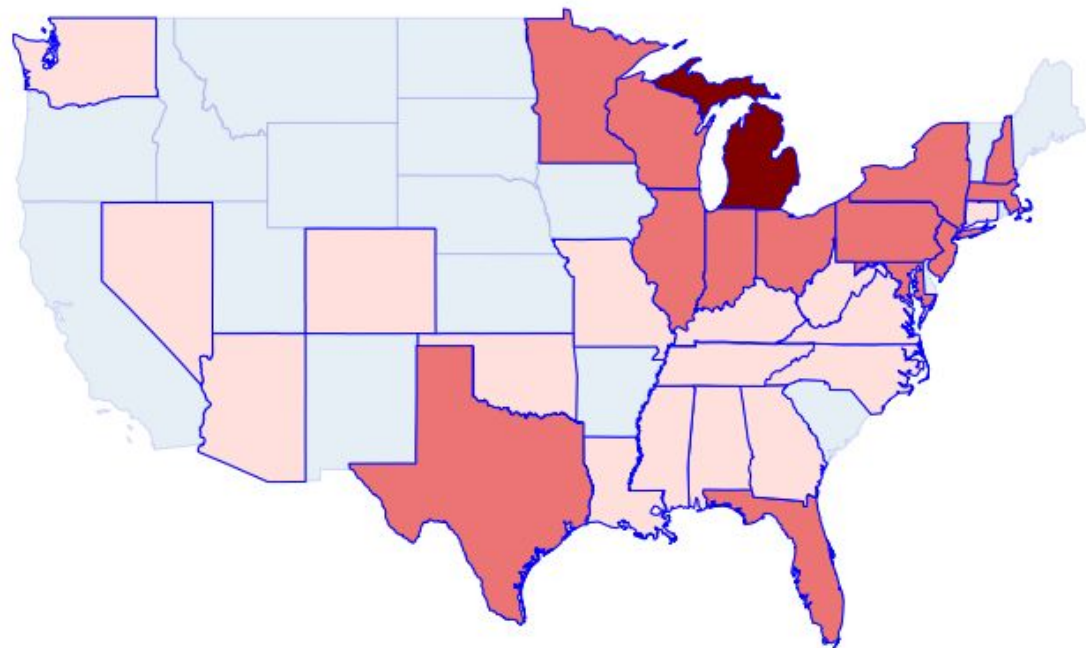


The top 150 paying customers (or the top ~29% of customers) account for ~75% of the total revenue.

\* Customer and Order Statistics are estimates given the data available. For example:

- Customers are identified by email address in the data, which made it difficult to de-dupe repeat customers (e.g. one customer may use multiple emails). Further, some customer emails were missing.
- Recurring orders were identified using the 'Tag' feature, and missing tags were assumed to represent ad-hoc/one-time orders.

# Order Amount (\$) by State (Jan 2018 - July 2021)



State Name	Order Amount
Michigan	Removed for sensitive info
Illinois	
Wisconsin	
New York	
Ohio	
Minnesota	
Pennsylvania	
Texas	
Massachusetts	
New Hampshire	

~86% of the total customer spend comes from Michigan\*

\* Note that due to some orders with missing 'state' values, the total order dollars across all states is slightly less than total order dollars for all orders.



# Product Analysis

# Product Overview

Products were categorized based on two dimensions (note: enterprise names blinded for sensitive information):

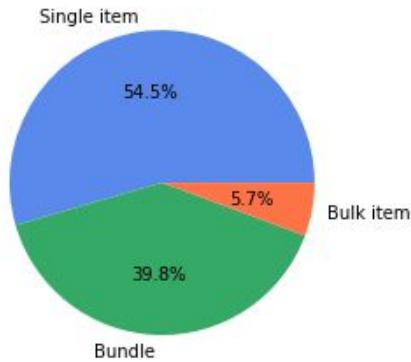
- Enterprise (1, 2, 3, 4, 5, or some combination of these groups)
- Product Type (Single item, Bundle, Bulk item)

From Jan 2018 to July 2021, product orders fell into one of 13 Enterprise-Product Type combinations, as shown on the right.

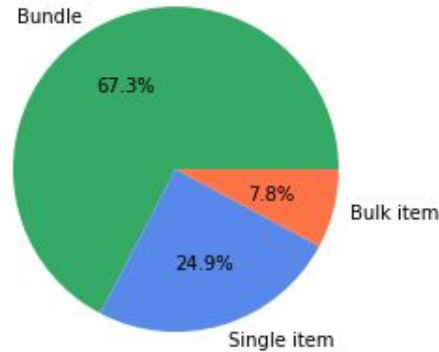
Product Type	Enterprise
Single item	Enterprise 1
	Enterprise 2
	Enterprise 3
	Enterprise 4
	Enterprise 5
Bulk item	Enterprise 1
	Enterprise 3
	Enterprise 4
Bundle	Enterprise 1
	Enterprise 3
	Enterprise 1, 2
	Enterprise 1, 3
	Enterprise 1, 2, 3

# Orders by Product Type (Jan 2018 - July 2021)

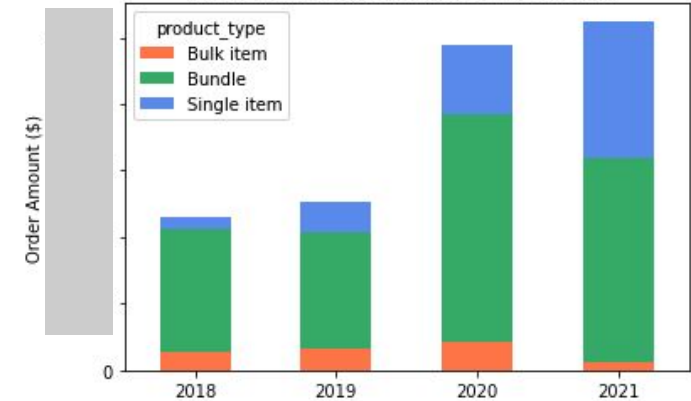
Product Type % of  
Total Number of Line Items Ordered



Product Type % of Total Order Amount (\$)



Annual Order Amount (\$) by Product Type



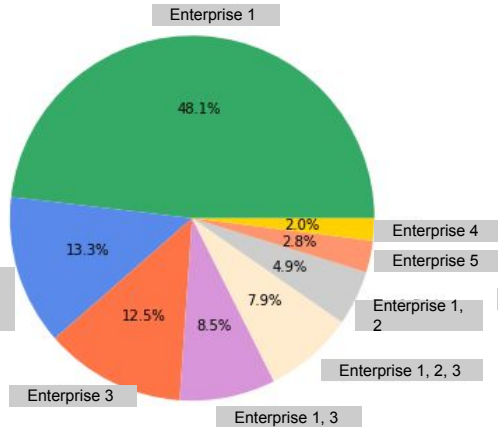
Single items made up the majority (54.5%) of the total number of line items ordered, but Bundles accounted for the largest portion of total order amount (67.3%), or total amount spent by customers.

The contribution of each product type to the total order amount has remained relatively consistent year over year.

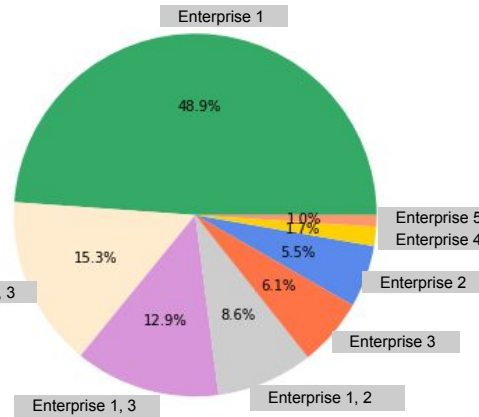
Interestingly, so far in 2021, customers appear to be spending more than typical on single items.

# Orders by Enterprise (Jan 2018 - July 2021)

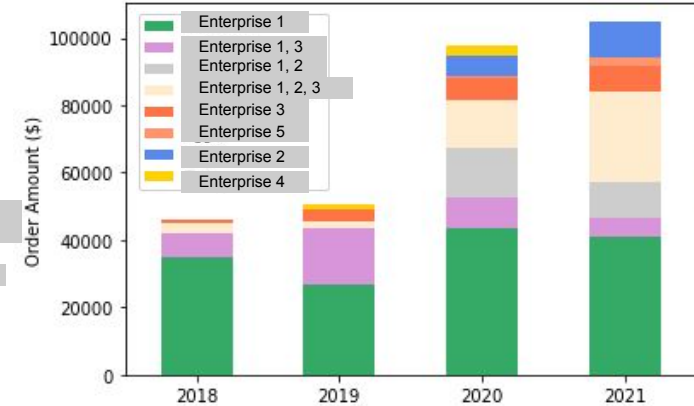
Enterprise % of  
Total Number of Line Items Ordered



Enterprise % of Total Order Amount (\$)



Annual Order Amount (\$) by Enterprise



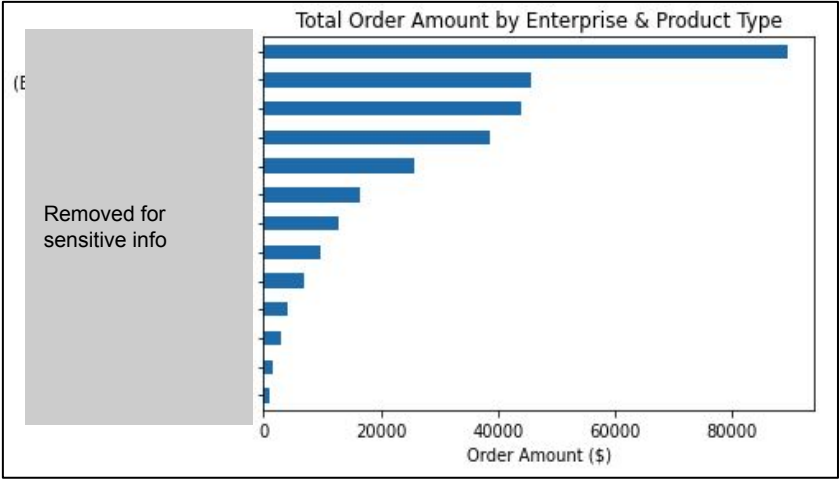
Enterprise 1 made up the majority of the total number of line items ordered (48.1%) as well as the largest portion of total order amount (48.9%). Aside from Enterprise 1 it appears that customers spend the most money on combination orders (e.g. Bundles) that include Enterprise 1 products.

The contribution of each enterprise to the total order amount has remained relatively consistent year over year, with Enterprise 1 or combination products including Enterprise 1 leading the customer spend.

Note that Enterprise 2 and 5 were only introduced recently (as of 2020).

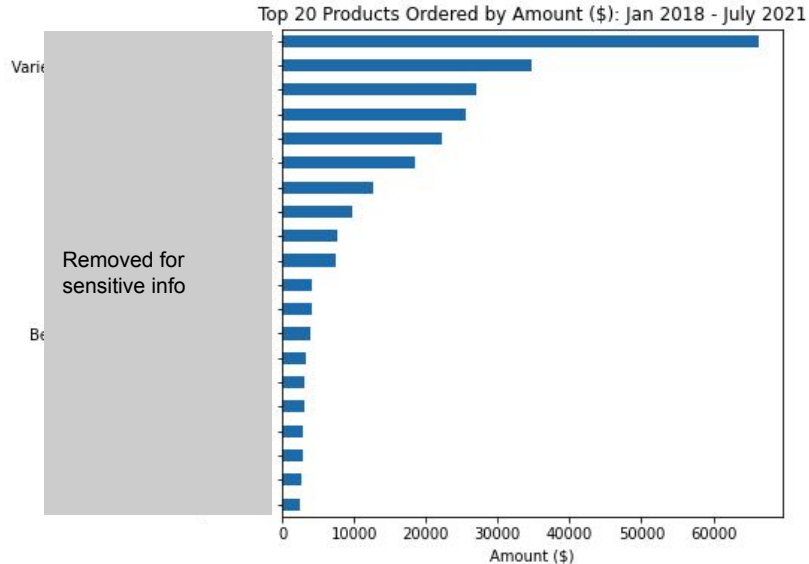
# Orders by Enterprise-Product Type Combination (Jan 2018 - July 2021)

Order Combination (Enterprise + Product Type)	Total # of Order Line Items	# as % of Total	Total Order Amount (\$)	\$ as % of Total
Removed for sensitive info	640	17.62%	Removed for sensitive info	29.94%
	288	7.93%		15.25%
	1,079	29.70%		14.68%
	309	8.51%		12.95%
	177	4.87%		8.59%
	483	13.29%		5.53%
	30	0.83%		4.24%
	158	4.35%		3.24%
	262	7.21%		2.27%
	54	1.49%		1.35%
	102	2.81%		1.03%
	33	0.91%		0.57%
	18	0.50%		0.37%
Total	3,633	100%		100%



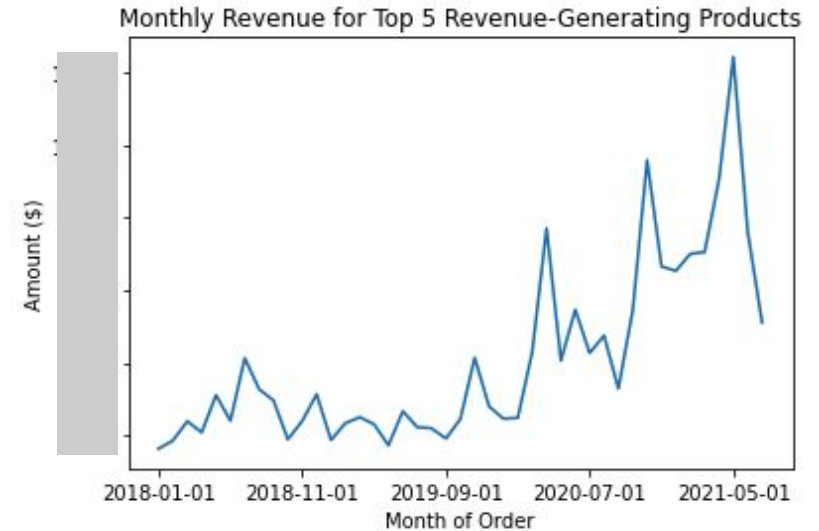
While [redacted] Bundles accounted for 17.6% of the total number of order line items, it accounted for almost 30% of the total cost of orders, which was nearly twice as much as any other Enterprise-Product Type category.

# Drill Down: Products Ordered - All Products



Customers appear to spend the most money on variety bundles, as the top four products orders by dollar amount are all variety boxes.

Interestingly, the [REDACTED] generated almost twice as much revenue as any other product.



If we look at the total monthly revenue for the top 5 products, we see an upward trend since Jan 2018, with some spikes and dips along the way.

Across 2020 and 2021, the three highest revenue-generating months for these products were April 2020 (\$ [REDACTED]), November 2020 (\$ [REDACTED]) and May 2021 (\$ [REDACTED]). Note that July 2021 is only a partial month.

# Drill Down: Products Ordered - All Products



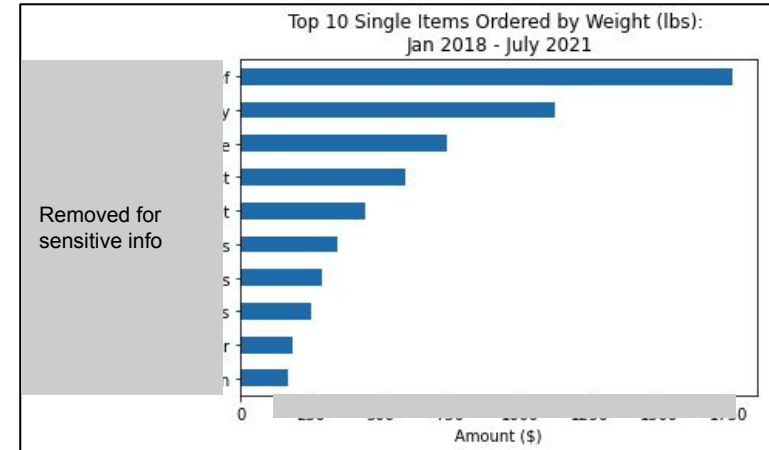
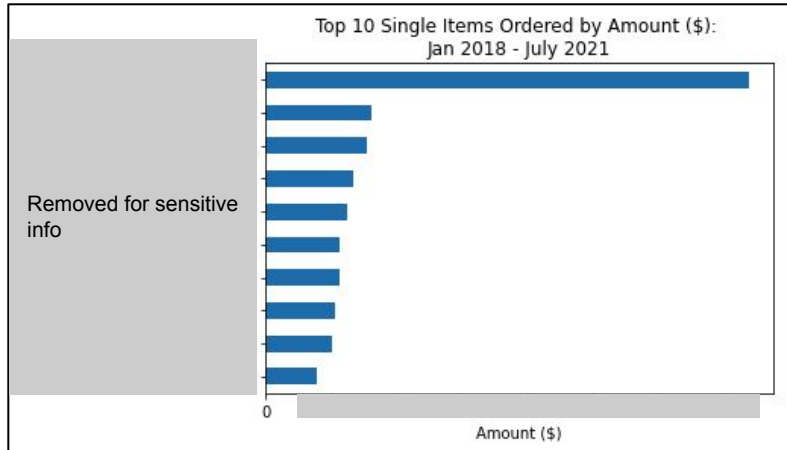
On average, order revenue is:

- Relatively consistent from December through July
- Lowest between August and September
- Highest during the month of November

The higher monthly revenue during November and December is likely due to an increase in demand during the holidays (e.g. bulk Turkey orders, special Holiday Gift Bundles and promotions).

Note: July average revenue may be a slight underestimate as the order data only includes a partial month of July 2021; however, the lack of data for Aug-Dec 2021 was factored into this analysis, thus all months besides July should represent the true averages.

# Drill Down: Products Ordered - Single Items (Jan 2018 - July 2021)



## Notes:

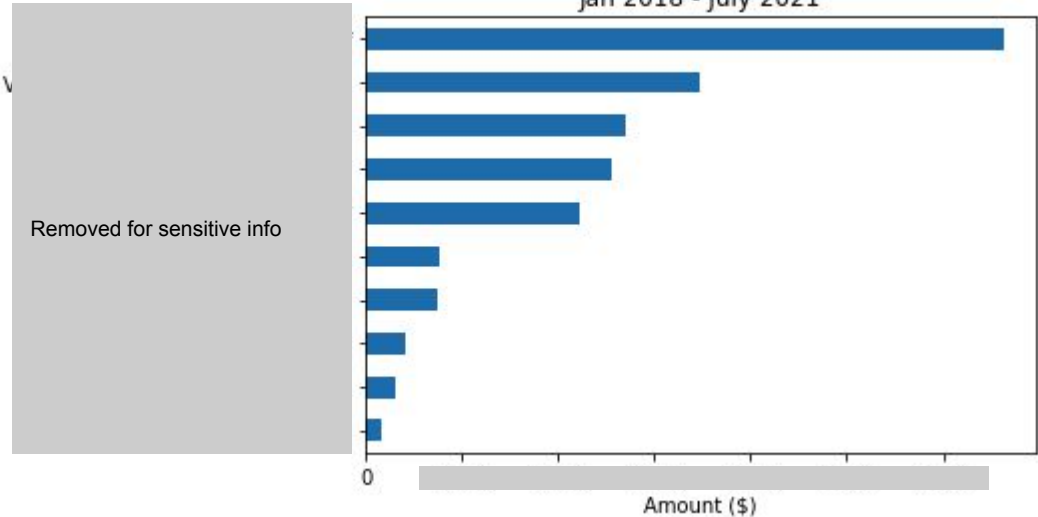
- Single items ordered as part of bundles are not included in this analysis
- [redacted] are not included in the top orders by weight category

- [redacted] is by far the most popular single item ordered by dollar amount and weight
- Some of the top products ordered by weight do not appear to be among the highest revenue-generating products, at least when ordered individually (not in a bundle) e.g. [redacted]



# Drill Down: Products Ordered - Bundles (Jan 2018 - July 2021)

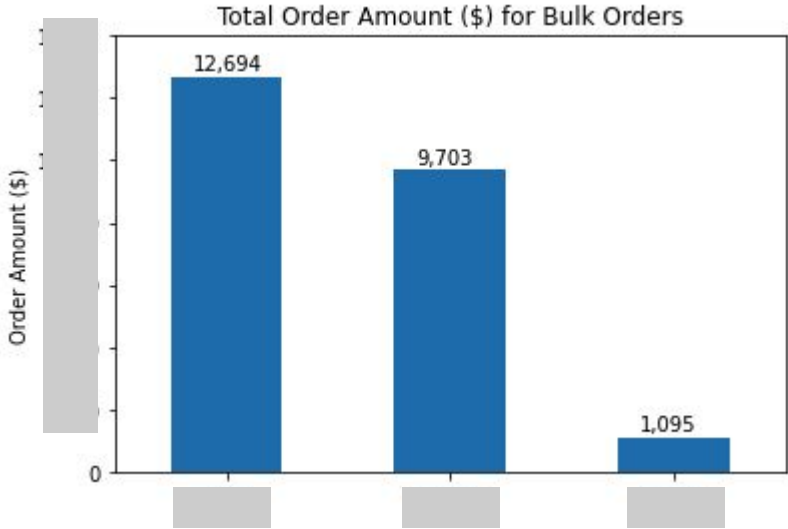
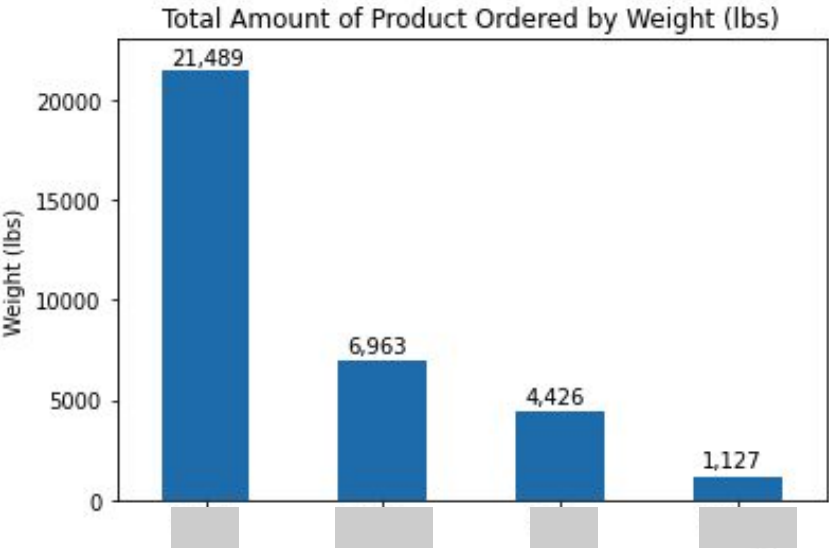
Top 10 Bundles Ordered by Amount (\$):  
Jan 2018 - July 2021



As we saw previously when looking at revenue across all products, variety boxes appear to generate the most customer spend out of all the different bundles.

# Appendix

# Additional Analyses



Note: [redacted] and bulk items excluded (see appendix for full list of bulk items)