Amazon x Columbia Hackathon

Diva Coders

Felicity Bui, Sophia Chan, Ritika Nadi,
Surbhi Panwar

The scenario

A dog food brand (Hanover) currently has only one product and wants to develop a new flavor for their brand. They consult Amazon for product strategy.

Why is it worth solving?

Amazon provides technical solutions for existing clients to strengthen their relationship.

Doing so will help generate more revenue for Amazon Ads when the brand launch its new product through Amazon's ad solutions.

01

Data Source

Dog Food Raw Dataset

1 Shape

8894 rows
21 columns

2 Variable Types

Integers, Strings, Numerical

3 Predictors

Ads experience, sns, gender, prime, etc.

	sale_id	sale_date	ad_exp	sns	product_id	product_brand	product_name	price	qty	customer_id	...	city	st	zip	lat	lng	marital
0	1	2021-10-01	Don't recall seeing an ad	0	3	Alpha	Alpha Natural Sensitive Systems, Skin & Coat S...	48.99	1	370	...	Boise	ID	83711	43.4599	-116.2440	Married
1	2	2021-10-01	Don't recall seeing an ad	1	14	Arf	Arf Soft & Tender American Jerky Dog Treats	11.52	1	625	...	Durham	NC	27710	36.0512	-78.8577	Married
2	3	2021-10-01	Don't recall seeing an ad	0	23	Bezt	Bezt Adult Chicken and Brown Rice Recipe Dry D...	14.22	1	905	...	Phoenix	AZ	85099	33.2765	-112.1872	Married
3	4	2021-10-01	Don't recall seeing an ad	0	33	Alpha	Alpha Probiotics Shredded Blend High Protein, ...	79.31	1	1061	...	Portsmouth	NH	214	43.0059	-71.0132	Married
4	5	2021-10-01	Don't recall seeing an ad	0	26	Alpha	Alpha Natural Adult Lamb & Rice Dry Dog Food	25.16	1	1411	...	Chicago	IL	60624	41.8804	-87.7223	Single
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
8889	8890	2022-09-30	Sponsored Brands	0	26	Alpha	Alpha Natural Adult Lamb & Rice Dry Dog Food	25.16	2	3657	...	Bellevue	WA	98008	47.6115	-122.1162	Single

02

Data Preprocessing

Transforming Columns

- Dropping sale_id, customer_id, product_id using function .drop
- Standardize price using function StandardScaler
- Convert categorical variable into dummy variables using function .get_dummies

Our resulting dataframe now has **8894 rows, 1044 columns**

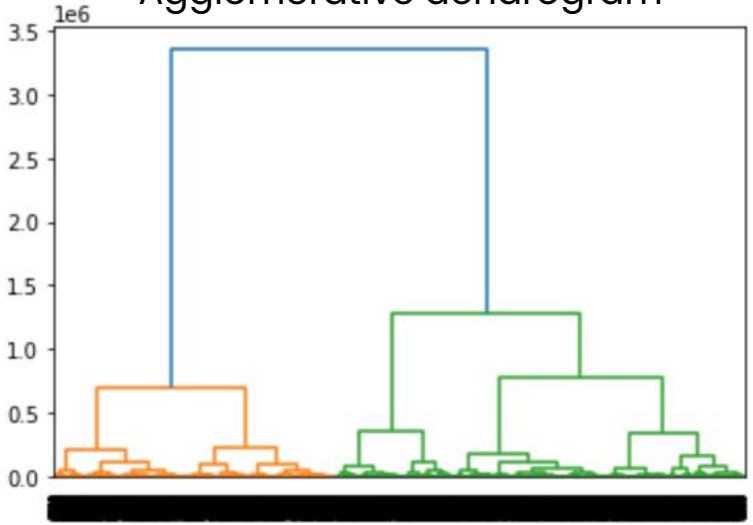
03

Predictive Models

Agglomerative clustering

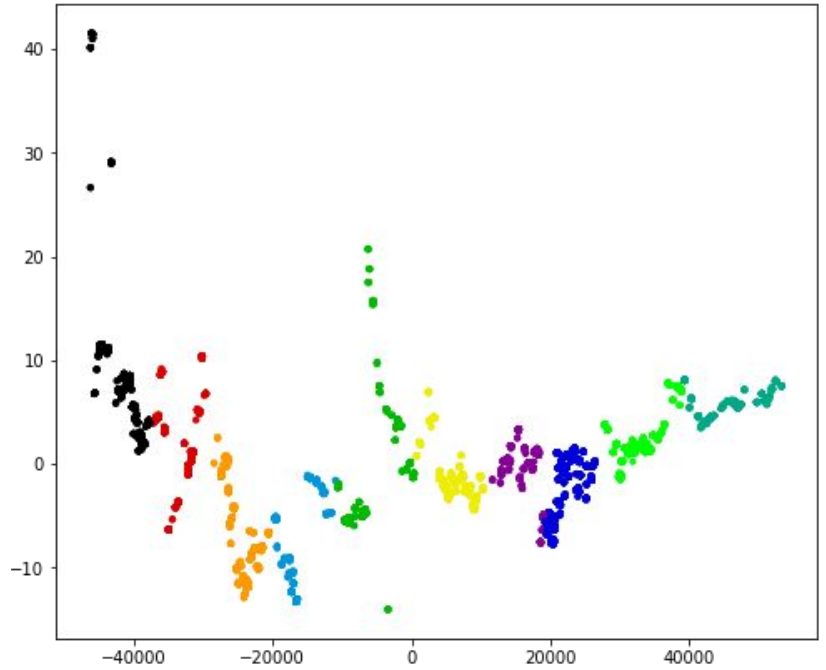
- What:
  - Bottom up approach.
  - Start with 8894 clusters and merge them together to create bigger clusters.
- Why:
  - No assumption of a particular number of clusters
  - May correspond to meaningful taxonomies
- How:
  - Use the euclidean distance as the measure of distance between points and ward linkage to calculate the proximity of cluster
  - Labels_ property returns an array of integers where the values correspond to the distinct categories.

Agglomerative dendrogram



K-means Clustering

- What:
  - Collection of data points aggregated together because of certain similarities.
- Why:
  - Rather than defining groups before looking at the data, clustering allows you to find and analyze the groups that have formed organically.
- How:
  - Define a target number k, which refers to the number of centroids needed in the dataset.
  - Use the euclidean distance as the measure of distance between points and and the centroid.
  - Returns an array of integers where the values correspond to the distinct clusters.



Calculating the Woof Score

- We use the clusters generated using our Unsupervised learning models.
- Since the aim is to maximize profits, we calculate the revenue, i.e. the average of price times quantity for each cluster and rank the clusters based on the average revenue of each cluster.
- We normalize the revenue to between 0 and 1 and call this the woof score.

Woof Function

- The woof function can be defined as -
  woof(X, model) = woof_score

  Where,
  - *X is features of the product like brand, price, whether they subscribe & save (sns) or not not*
  - *model is our trained unsupervised learning model*
  - *woof_score is the assigned score for each cluster that ranges between 0 and 1.*

High woof score

Low woof score

Strategy discussion with the team

- Alpha : best performing brand based on WOOF score
- Brand focus: Hanover
    - based on WOOF score, Hanover has a high occurrence in the most-valuable cluster which generates the highest average revenue
- Alpha has a product in the similar price range as Hanover’s existing product - Hanover should launch a similar dog food flavor

Strategy discussion with Hanover

- Hanover's only product is priced at an average of $58.49
- In this price range, we identified a similar product which has a high potential to be the next flavor Hanover should launch
- Furthermore, the sns percentage for the product we are recommending is 100% - which shows this product is well-received by its customers

The end