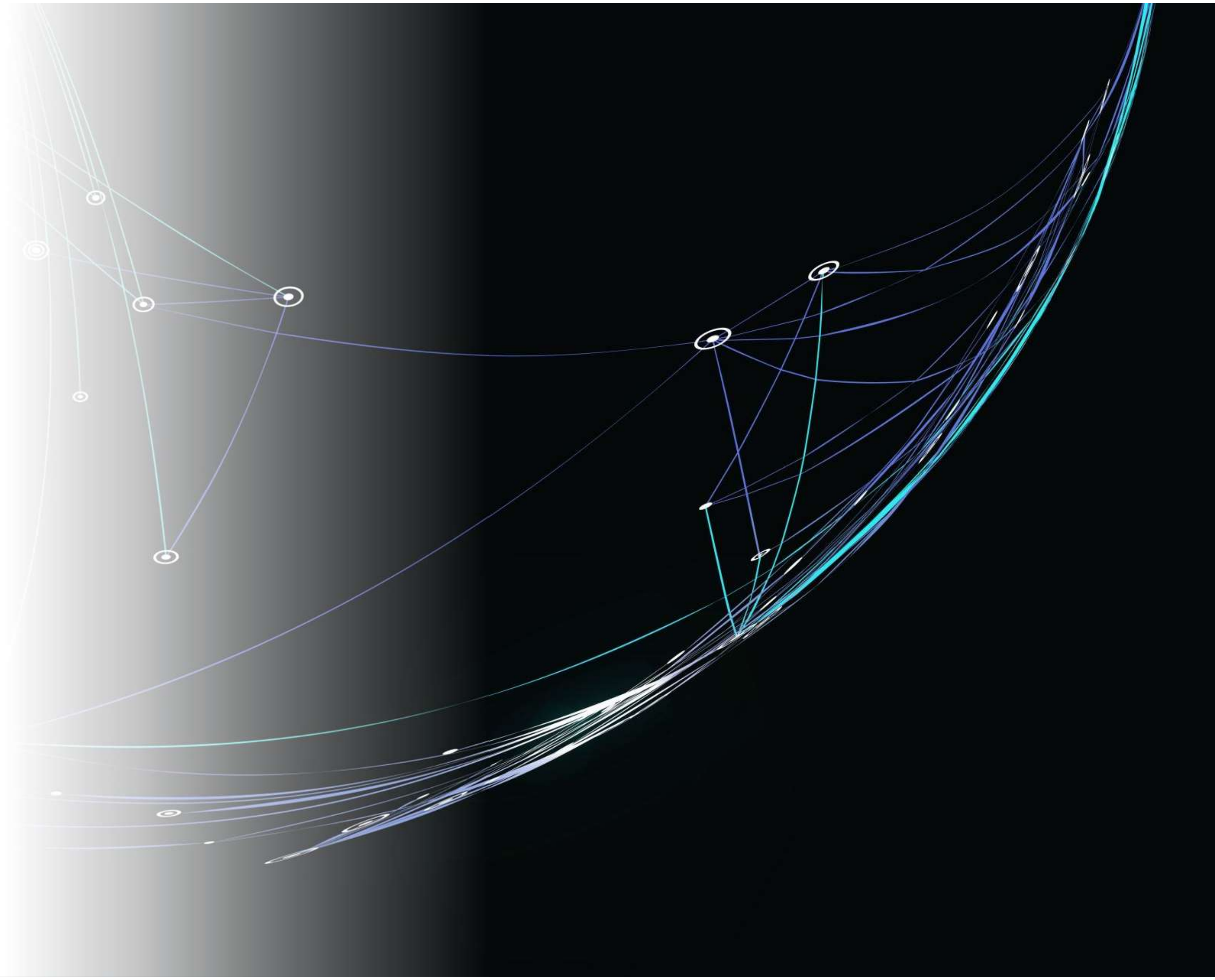
Product Demand Predictor Tool

Project 4

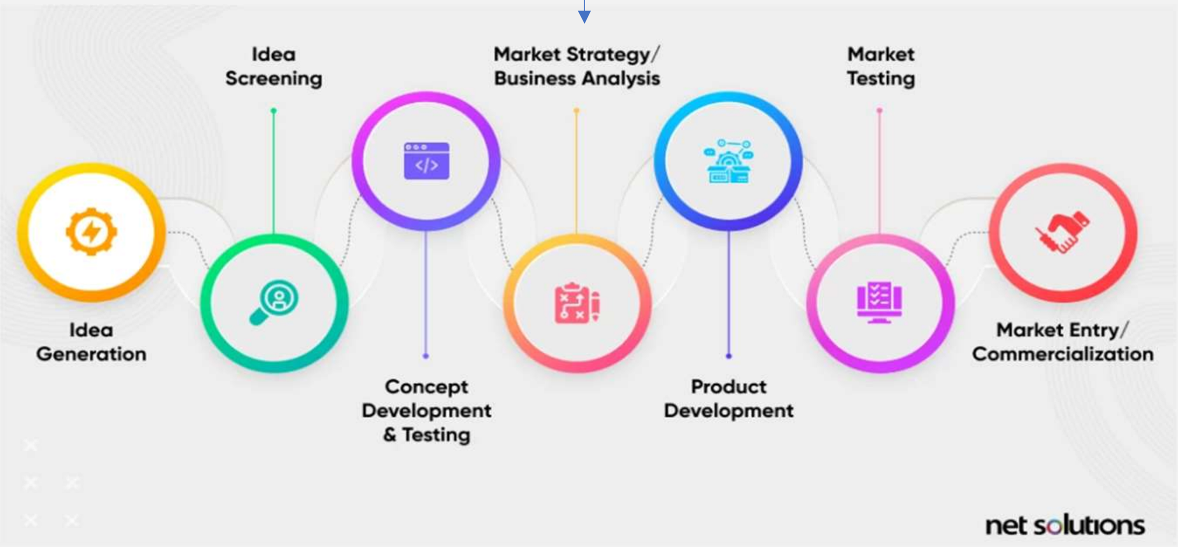
Lina Orjuela



Analysis question:
How many units could a product sell on Amazon Canada?

- Predicting the demand that a product will have in the market is a complex task for a person who wants to sell their product online.
- Also, is a critical point in the development of a new product for a company. This tool is intended to help with this task.

product demand prediction

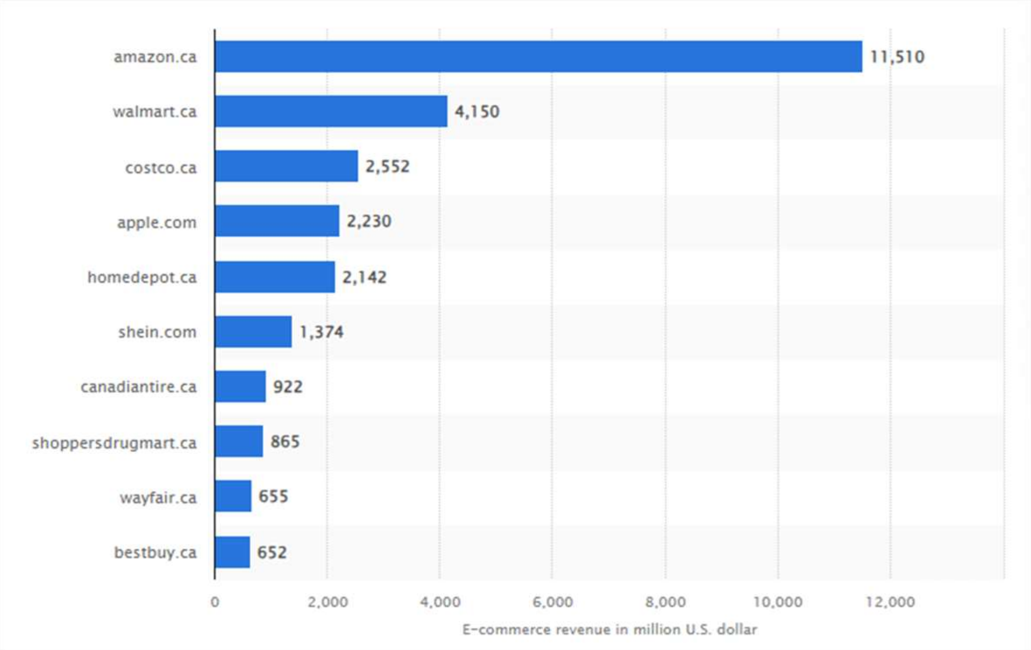

Idea Screening
Market Strategy/ Business Analysis
Market Testing
Idea Generation
Market Entry/ Commercialization
Concept Development & Testing
Product Development
net solutions

Source: https://www.netsolutions.com/insights/everything-about-new-product-development/

Why Amazon Canada ?

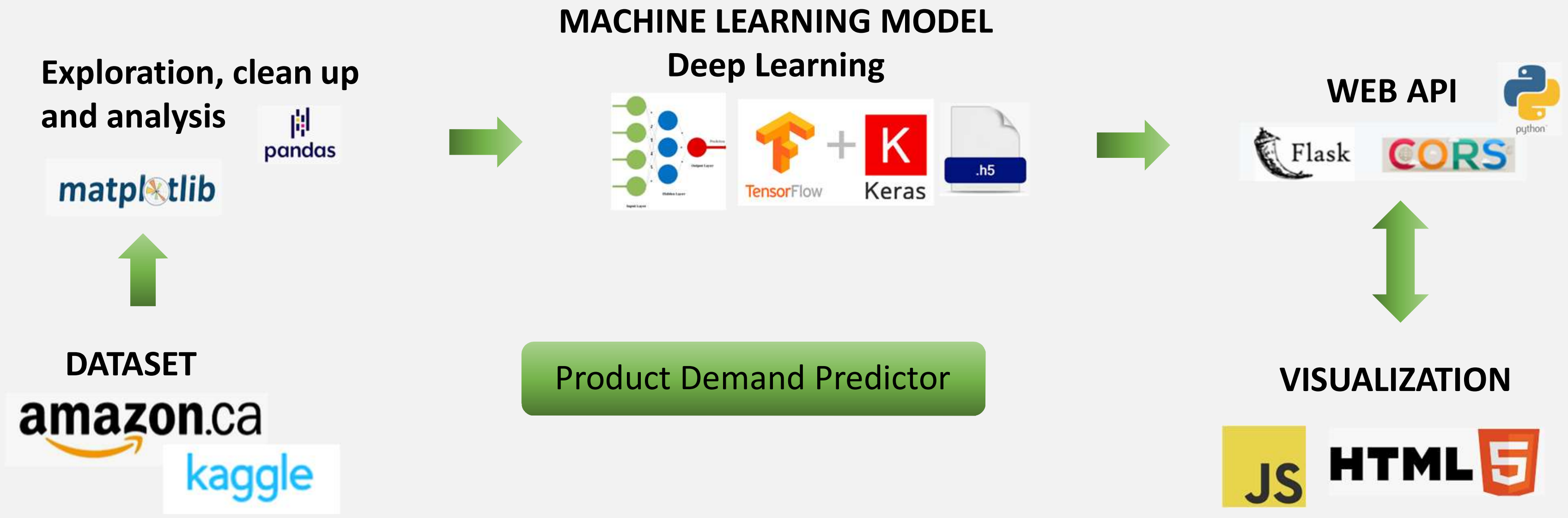
Top online stores in Canada in 2022, by e-commerce net sales *(in million U.S. dollar)*

- In February 2023, Amazon Canada received close to 160 million visits
- Over 50% of Canadians selling on Amazon make over $50,000 per year.



Source: https://www.statista.com/forecasts/871090/canada-top-online-stores-canada-ecommercedb
https://madeinca.ca/amazon-statistics-canada/

How is the tool built?



Exploration, clean up and analysis

DATASET

MACHINE LEARNING MODEL
Deep Learning

Product Demand Predictor

WEB API

VISUALIZATION

pandas
matplotlib
amazon.ca
kaggle
TensorFlow
Keras
.h5
python
Flask
CORS
JS
HTML

Source: https://www.kaggle.com/datasets/asaniczka/amazon-canada-products-2023-2-1m-products/

DATA cleanup

Initial Dataset

2.1MProducts
11 columns
246 Category name

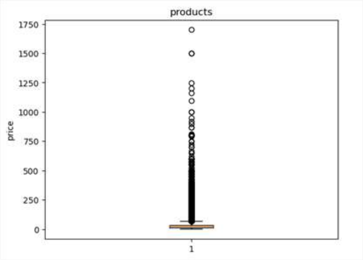
Data Cleanup

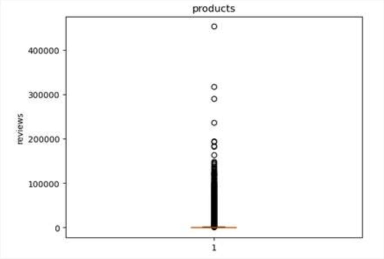
- Delete: null values, products with price $0, outliers in price and reviews and 5 columns
- Creation of 2 new columns: word count and discount
- Filter: top 10 product categories with bought in las month.

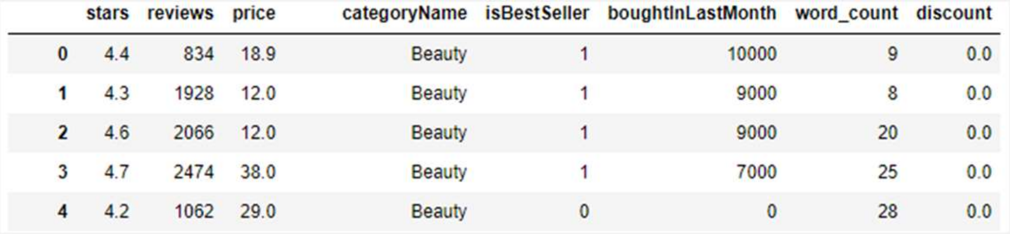
Clean Dataset

78 K Products
8 columns
10 Category name

Outliers

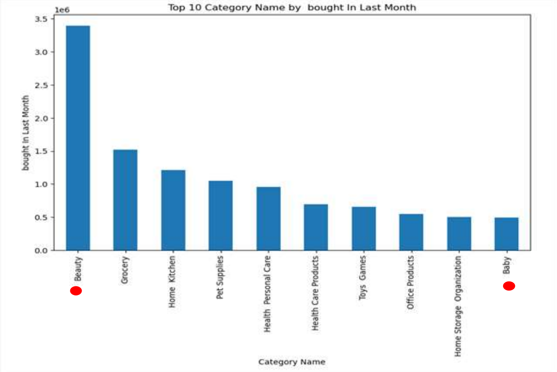






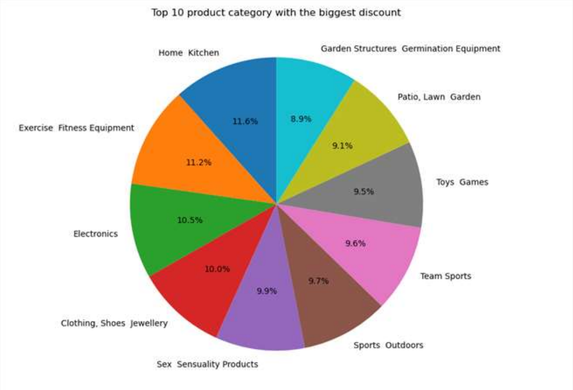
	stars	reviews	price	categoryName	isBestSeller	boughtInLastMonth	word_count	discount
0	4.4	834	18.9	Beauty	1	10000	9	0.0
1	4.3	1928	12.0	Beauty	1	9000	8	0.0
2	4.6	2066	12.0	Beauty	1	9000	20	0.0
3	4.7	2474	38.0	Beauty	1	7000	25	0.0
4	4.2	1062	29.0	Beauty	0	0	28	0.0

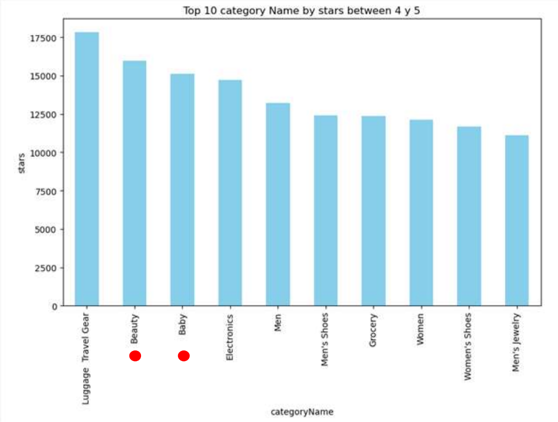
Exploration and Analysis

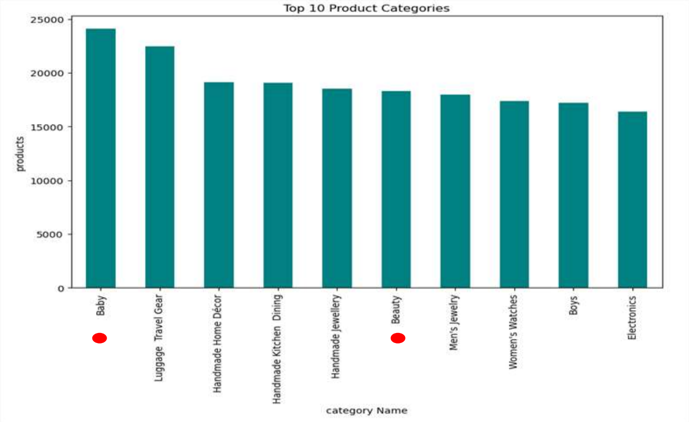


* **Baby** and **Beauty** categories are part of the top 10 in:

- Product quantity
- Products bought in las month
- highest number of stars between 4 and 5.





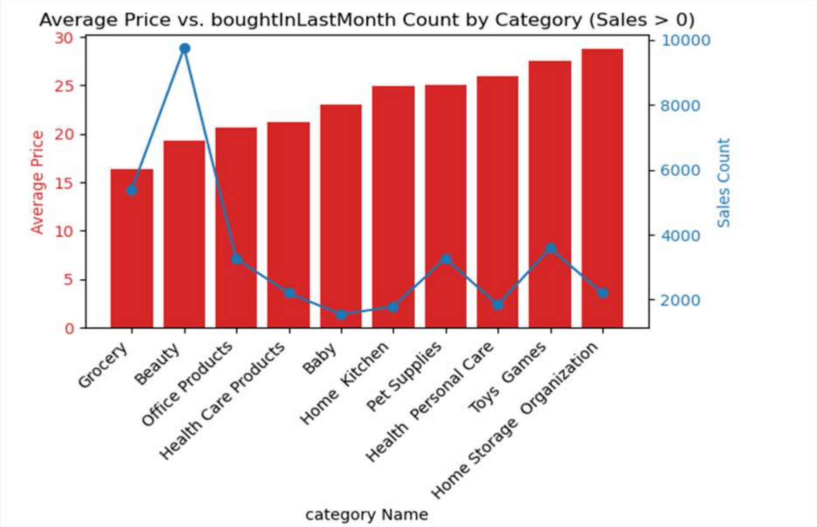


* The top 10 product categories that offer the most discounts are not the same ones that sell the most and have the most products.

* Initial Dataset

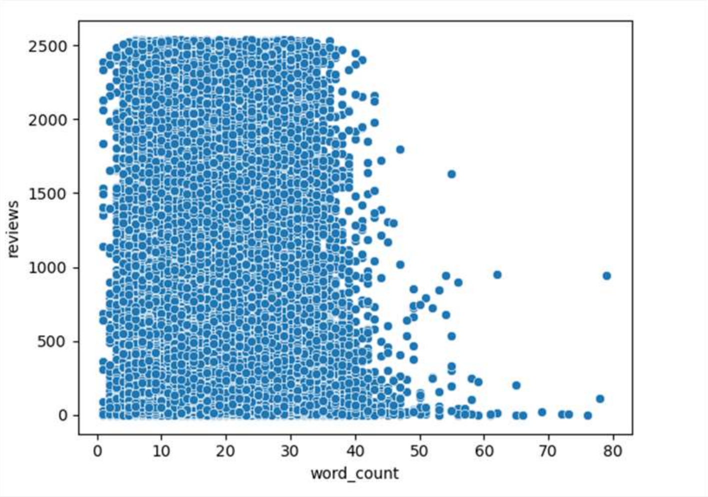
Exploration and Analysis

Average Price vs Bought in Last month



*****Beauty** category with an average price of 20, achieves the highest number of sales in 1 month

Comparing reviews and number of words in the product description

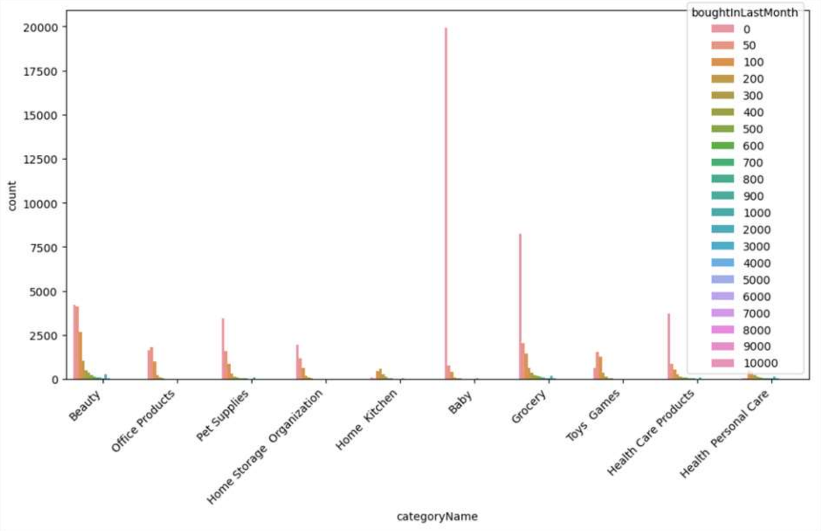


*The relationship between the number of words in a product description and the number of reviews shows that descriptions with less than 40 words are the ones that receive the most reviews.

*clean Dataset

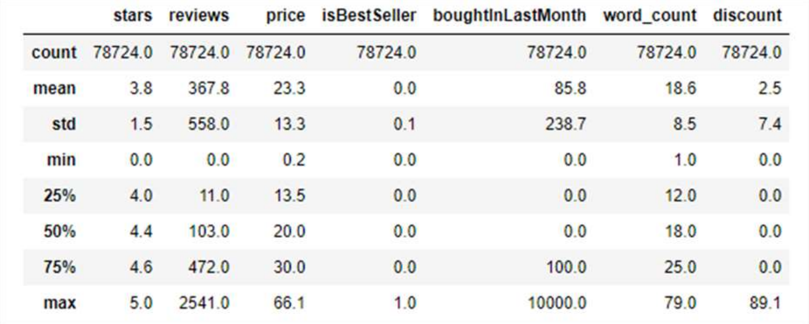
Exploration and Analysis

Quantity of products purchased in the last month by product category



****Baby** and **Grocery** categories were the ones with the most unsold products in the last month

summary statistics



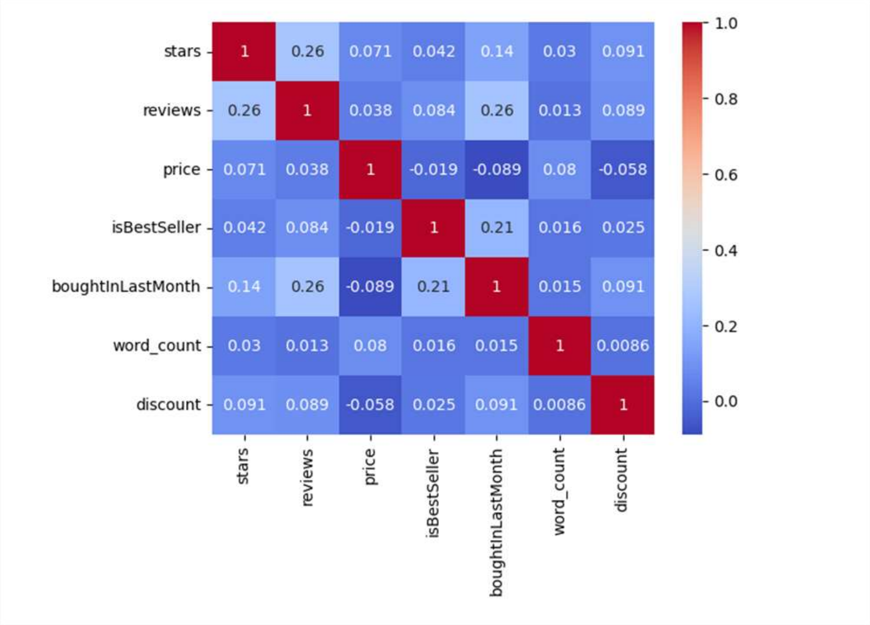
	stars	reviews	price	isBestSeller	boughtInLastMonth	word_count	discount
count	78724.0	78724.0	78724.0	78724.0	78724.0	78724.0	78724.0
mean	3.8	367.8	23.3	0.0	85.8	18.6	2.5
std	1.5	558.0	13.3	0.1	238.7	8.5	7.4
min	0.0	0.0	0.2	0.0	0.0	1.0	0.0
25%	4.0	11.0	13.5	0.0	0.0	12.0	0.0
50%	4.4	103.0	20.0	0.0	0.0	18.0	0.0
75%	4.6	472.0	30.0	0.0	100.0	25.0	0.0
max	5.0	2541.0	66.1	1.0	10000.0	79.0	89.1

*The average sales for 1 month is 85 units of product

*clean Dataset

Exploration and Analysis

Correlation Matrix



	stars	reviews	price	isBestSeller	boughtInLastMonth	word_count	discount
stars	1	0.26	0.071	0.042	0.14	0.03	0.091
reviews	0.26	1	0.038	0.084	0.26	0.013	0.089
price	0.071	0.038	1	-0.019	-0.089	0.08	-0.058
isBestSeller	0.042	0.084	-0.019	1	0.21	0.016	0.025
boughtInLastMonth	0.14	0.26	-0.089	0.21	1	0.015	0.091
word_count	0.03	0.013	0.08	0.016	0.015	1	0.0086
discount	0.091	0.089	-0.058	0.025	0.091	0.0086	1

The correlation between price and discount is -0.058, indicating a weak negative correlation. The price increases, the discount tends to decrease.

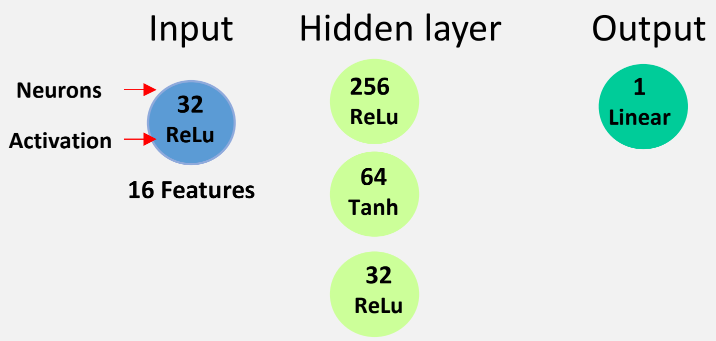
The correlation between isBestSeller and boughtInLastMonth is 0.21, moderate positive correlation. Products that are best sellers tend to have been purchased in the last month.

Machine Learning Model

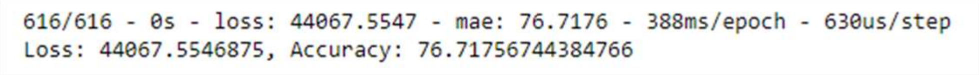
Model: Deep Learning
Problem type: Regression - goal is to perform a regression to predict a continuous value.

Final Model

Epoch:100

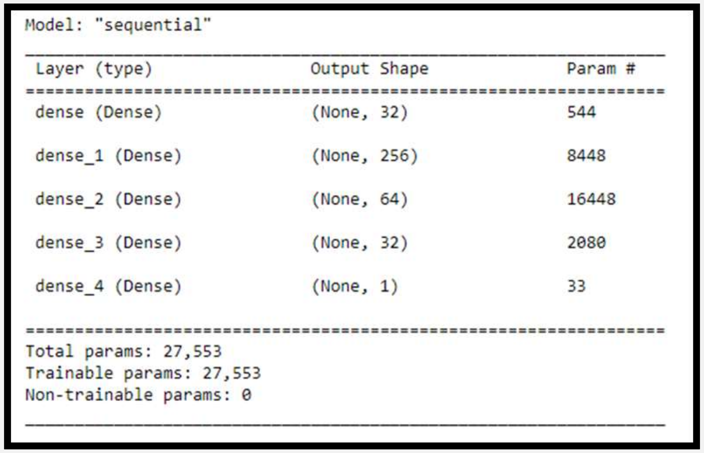

Input
Hidden layer
Output
Neurons
Activation
32
ReLu
16 Features
256
ReLu
64
Tanh
32
ReLu
1
Linear

MAE: 76.7176
Loss: 44067.55547

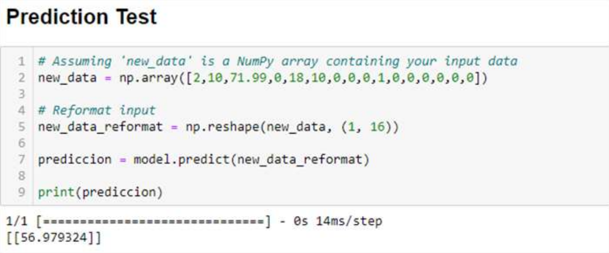


```
616/616 - 0s - loss: 44067.5547 - mae: 76.7176 - 388ms/epoch - 630us/step
Loss: 44067.5546875, Accuracy: 76.71756744384766
```

Neural Network built with Keras using the Sequential interface



```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 32)                544

 dense_1 (Dense)             (None, 256)               8448

 dense_2 (Dense)             (None, 64)                16448

 dense_3 (Dense)             (None, 32)                2080

 dense_4 (Dense)             (None, 1)                 33

=================================================================
Total params: 27,553
Trainable params: 27,553
Non-trainable params: 0
_________________________________________________________________
```



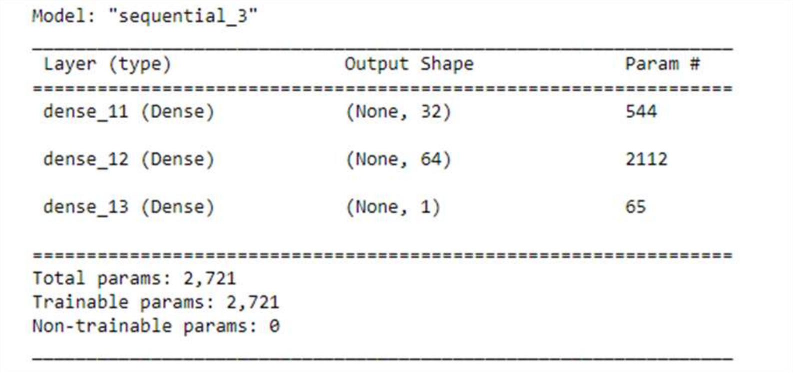
Prediction Test

```
# Assuming 'new_data' is a NumPy array containing your input data
new_data = np.array([2,10,71.99,0,18,10,0,0,0,1,0,0,0,0,0,0])

# Reformat input
new_data_reformat = np.reshape(new_data, (1, 16))

prediccion = model.predict(new_data_reformat)

print(prediccion)
```

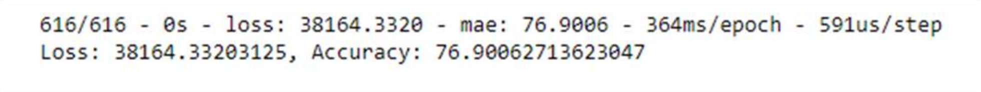
```
1/1 [==============================] - 0s 14ms/step
[[56.979324]]
```

Machine Learning Model

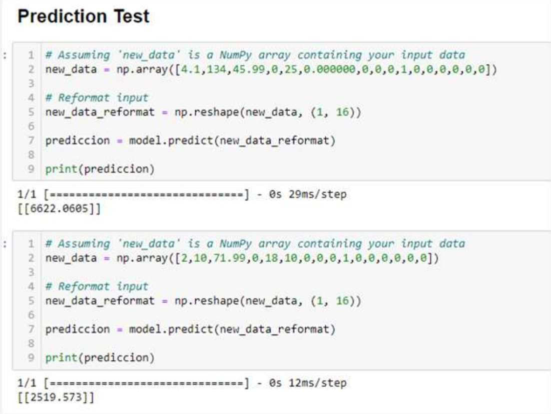
First approximation to the final model



```
Model: "sequential_3"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense_11 (Dense)            (None, 32)                544

 dense_12 (Dense)            (None, 64)                2112

 dense_13 (Dense)            (None, 1)                 65

=================================================================
Total params: 2,721
Trainable params: 2,721
Non-trainable params: 0
_________________________________________________________________
```



```
616/616 - 0s - loss: 38164.3320 - mae: 76.9006 - 364ms/epoch - 591us/step
Loss: 38164.33203125, Accuracy: 76.90062713623047
```



Prediction Test

```python
# Assuming 'new_data' is a NumPy array containing your input data
new_data = np.array([4.1,134,45.99,0,25,0.000000,0,0,0,1,0,0,0,0,0,0])

# Reformat input
new_data_reformat = np.reshape(new_data, (1, 16))

prediccion = model.predict(new_data_reformat)

print(prediccion)
```

```
1/1 [==============================] - 0s 29ms/step
[[6622.0605]]
```

```python
# Assuming 'new_data' is a NumPy array containing your input data
new_data = np.array([2,10,71.99,0,18,10,0,0,0,1,0,0,0,0,0,0])

# Reformat input
new_data_reformat = np.reshape(new_data, (1, 16))

prediccion = model.predict(new_data_reformat)

print(prediccion)
```

```
1/1 [==============================] - 0s 12ms/step
[[2519.573]]
```

After this model, more hidden layers were added and the activations were changed to reach the final model.

Machine Learning Model

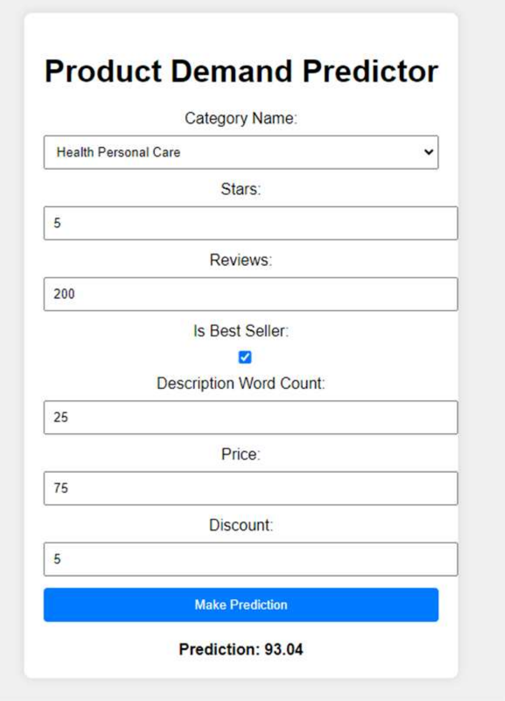
Comparison between the two models

	Input	Hidden layer	Neurons	Activation	Output	MAE	Loss	Prediction Test
Test Model	32 Relu 16 Features	1	64	Relu	1 Linear	79.9006	38164.33	6,622 2,519.5
Final Model	32 Relu 16 Features	1	256	Relu	1 Linear	76.7176	44067.55	79.2 83.9
		2	64	Tanh				
		3	32	Relu				

→ Prediction outside the average bought in last month

The two models present proximate MAE and LOSS indicators; however, the prediction test is closer to reality in the final model.

Product Demand Predictor Tool



Product Demand Predictor

Category Name: Health Personal Care

Stars: 5

Reviews: 200

Is Best Seller: ☑

Description Word Count: 25

Price: 75

Discount: 5

Prediction: 93.04

The user must fill out the form and at the end click on Make Prediction and the tool will provide the number of products that can be sold in a month according to the selected variables.

This will allow the user to determine their marketing strategy, if they need stars, how many stars, how much discount, depending on the number of units they want to sell.

CONCLUSIONS

- Big discounts do not ensure the sale of the product
- The price should not exceed $30
- The product description should be between 10 and 30 words
- The Beauty Category presents the highest number of units sold in a month.