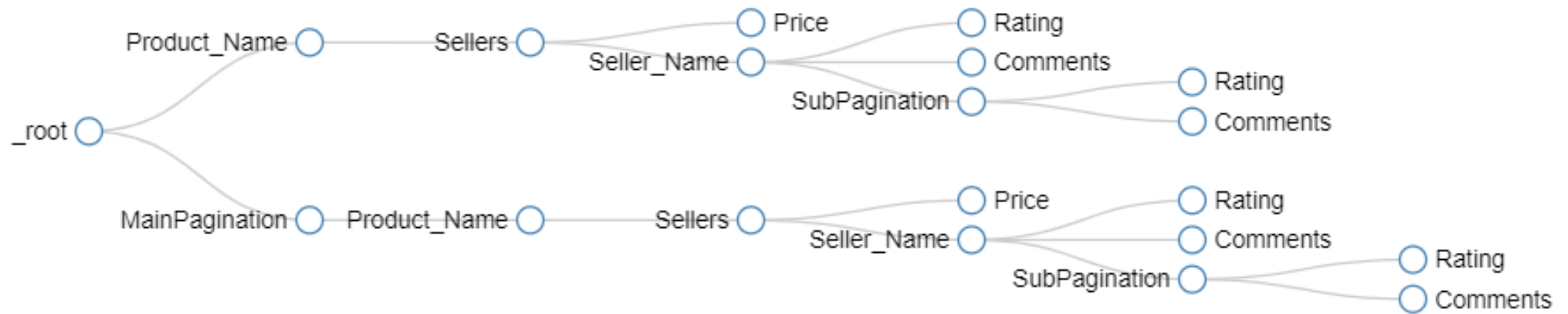# Seller Analysis Report on Amazon.in

Dataset Collection

- For data set collection, I used web scraper. Below is the selector graph of the crawler which I used to collect the dataset.

- In root, I placed a selected URL of the given page and the selector made its way through the crawler

- After Pre-Processing:

|	MultiLink	Product	Brand	Seller	Price	FeedBack_Rating	Product_Rating	Packaging_Rating	Courier_Rating
0	Video Games Deals	Doom (PS4)	Bethesda	Game Addicts	999.00	1	4.4	4.400000	4.400000
1	Headphones	Boat BassHeads 100 Hawk Inspired Earphones wit...	Boat	Techretailer	589.00	5	4.0	5.000000	5.000000
2	Drives & Storage	Sandisk 16GB Ultra Microsdhc (Microsd) Memory ...	SanDisk	Cell Express	280.00	5	4.3	5.000000	5.000000
3	Components	Generic Uno R3 ATmega328P with USB Cable lengt...	Generic	Cloudtail India	500.00	5	4.1	5.000000	5.000000
4	Janitorial & Sanitation Supplies	Sterillium Hand Sanitizer - 500 ml (Blue)	Sterillium	city medical	340.00	5	4.4	5.000000	5.000000
5	Monitors	Samsung 27 inch (68.6 cm) Curved Bezel Less LE...	Samsung	The-EStore	18921.00	5	4.5	5.000000	5.000000

Exploratory Data Analysis

- Total Sellers: 2903
- Total Products: 3224
- Total Categories: 88
- Total Brands: 967
- Generating Reports Like
  - Total Brands per Seller and Vice-Versa
  - Total Products per Seller and Vice-Versa etc.
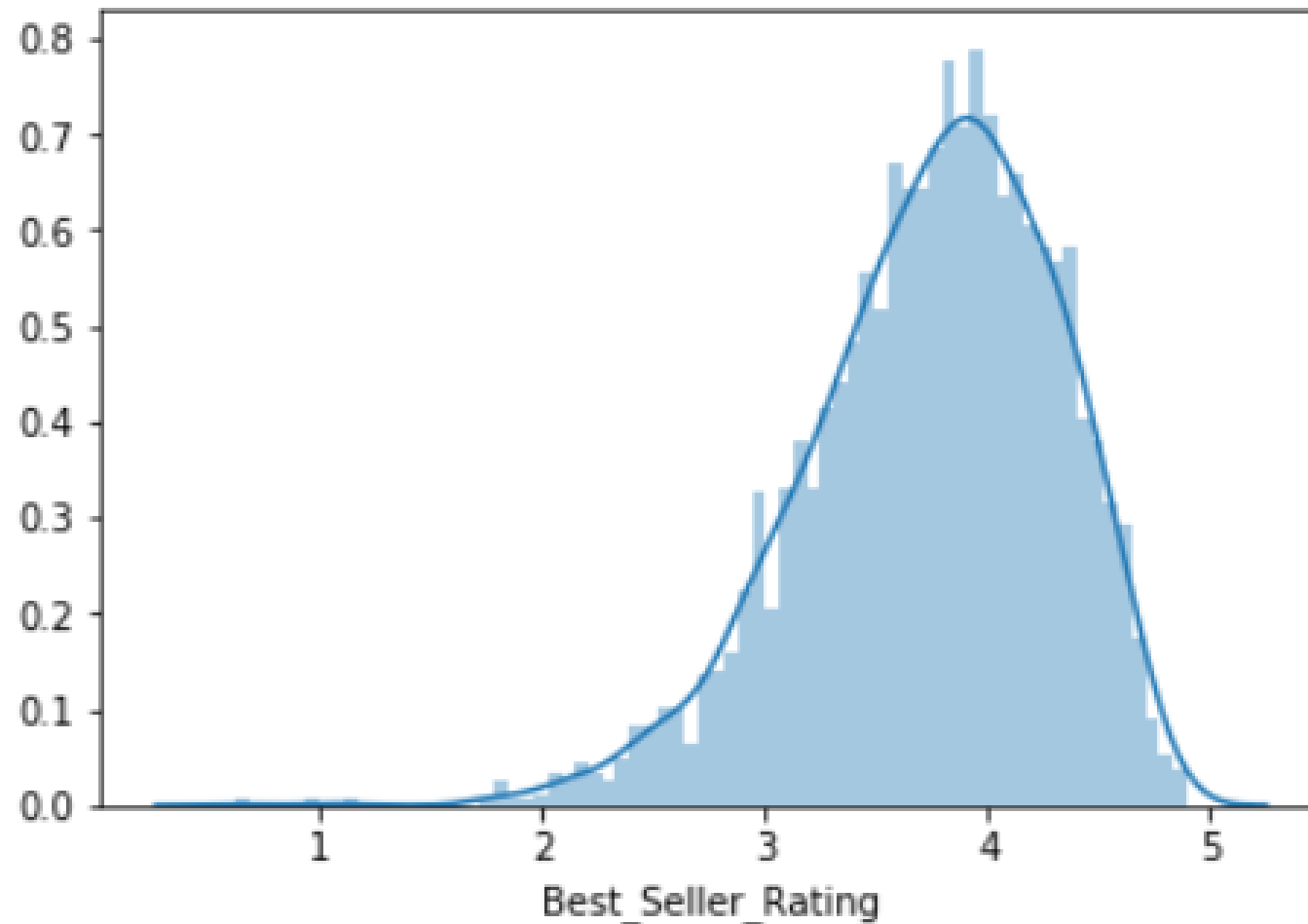
Best Seller Ratings (Category)
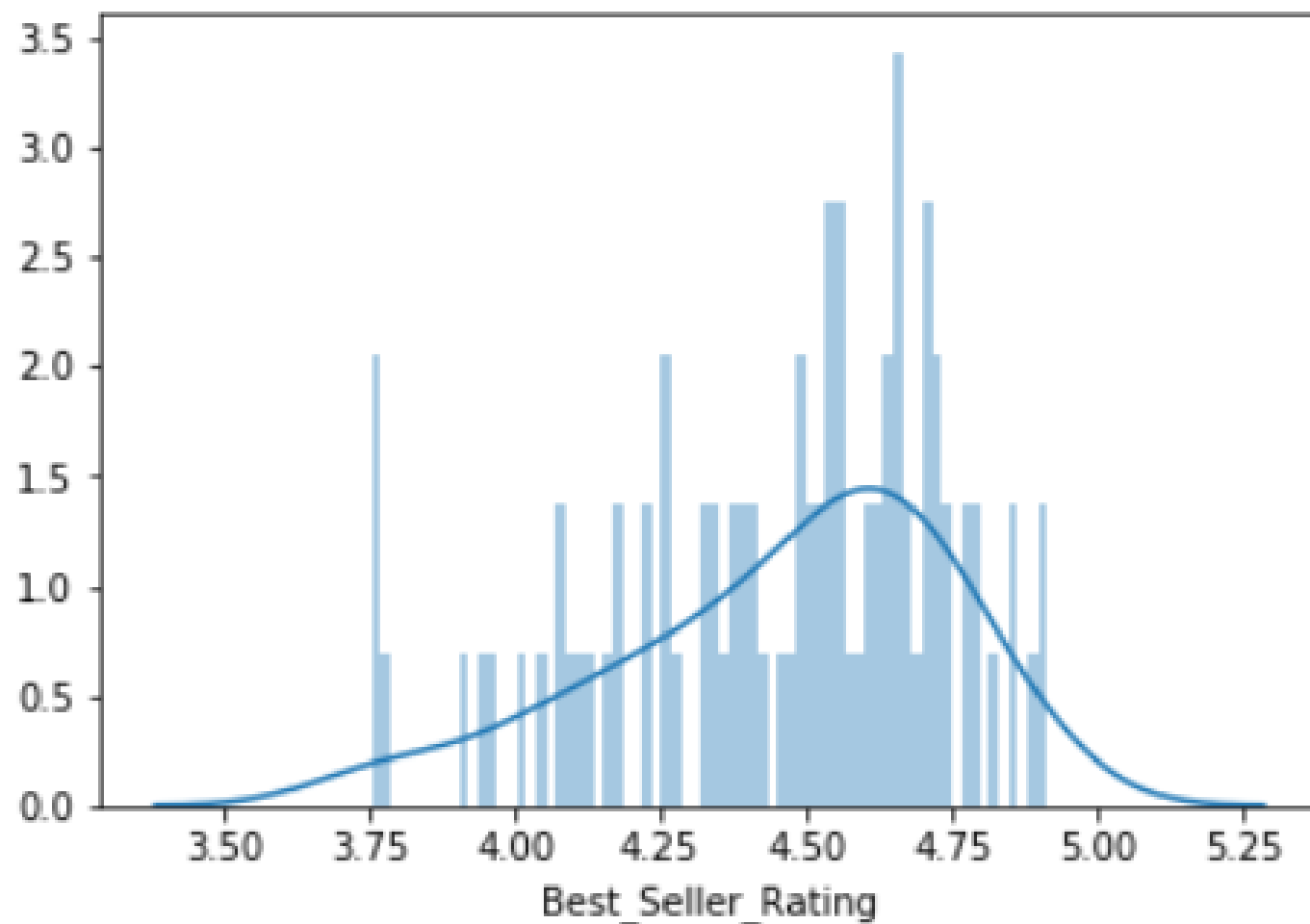
Best Seller Ratings (Brand)
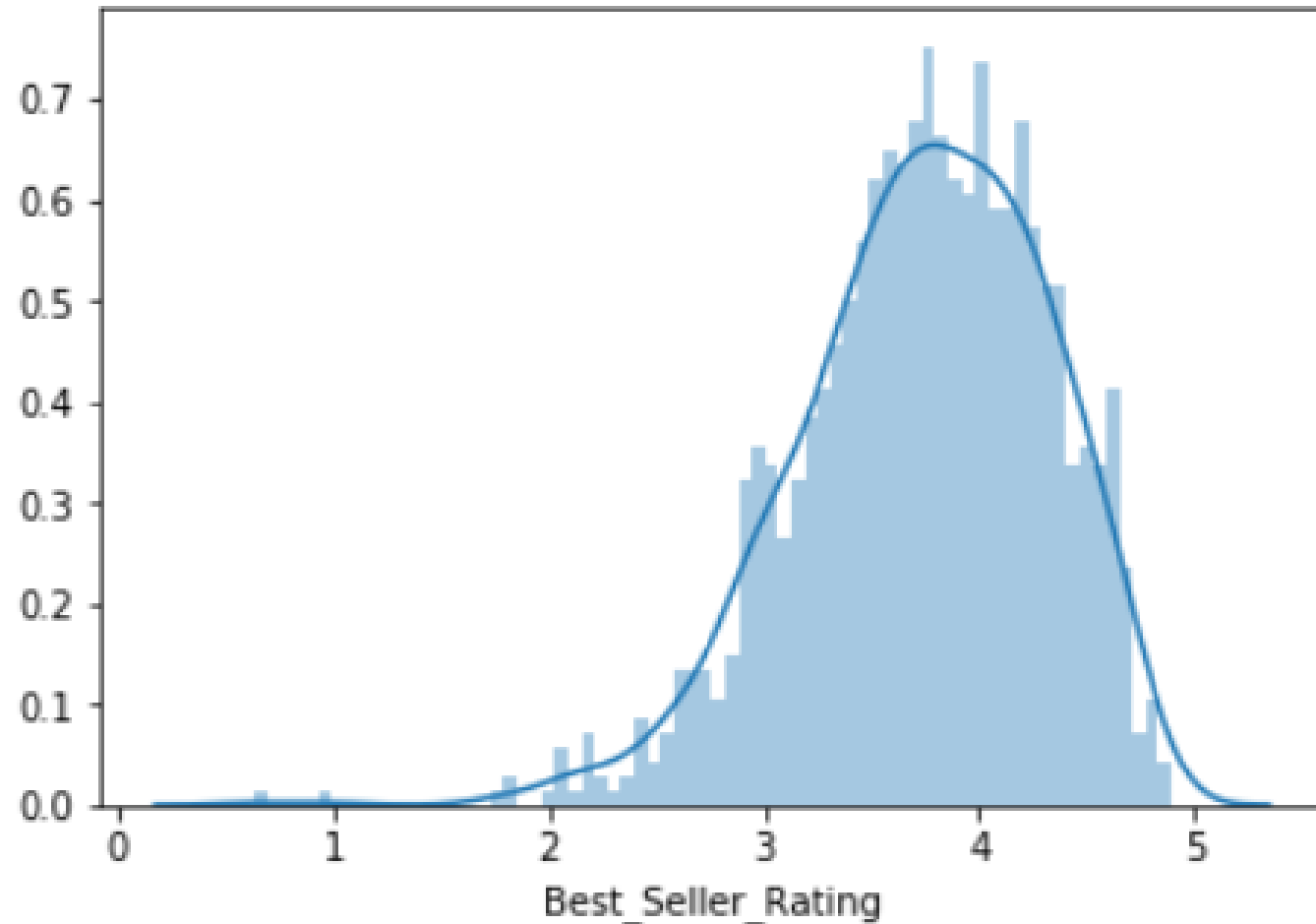
Best Seller Ratings (Product)

Analysis & Model Making

- As mentioned in database there were 5 criteria based on which a particular seller is to be scored:
  - **Price:** Lesser the price, more the score
  - **Count:** Popularity of the seller among the buyers
  - **Buyer Feedback:** Individual buyer rating to the seller's attitude
  - **Courier Rating:** Medium used by the seller to ship the product
  - **Product Rating:** Rating of the product under that seller
  - **Packaging Rating:** Handling the product by seller while shipping

Analysis & Model Making

- Each of the above mentioned criteria has individual value as well as whole value.

- Since every buyer has a perception of what he/she prefers the most out the criteria determines what kind of seller would he/she will prefer.

- To ensure the above, some kind of correlation is to be established on the basis of given dataset.

- But before establishing correlation, it was to be ensured that columns were to be in sync and scored.

- Correlation between features:
  - **# Price Correlation   : corr[0]**
  - **# Count Correlation   : corr[1]**
  - **# Buyer_FeedBack Correlation   : corr[2]**
  - **# Courier_Rating Correlation   : corr[3]**
  - **# Product_Rating Correlation   : corr[4]**
  - **# Packaging_Rating Correlation : corr[5]**

`<matplotlib.axes._subplots.AxesSubplot at 0x27a15d86c50>`
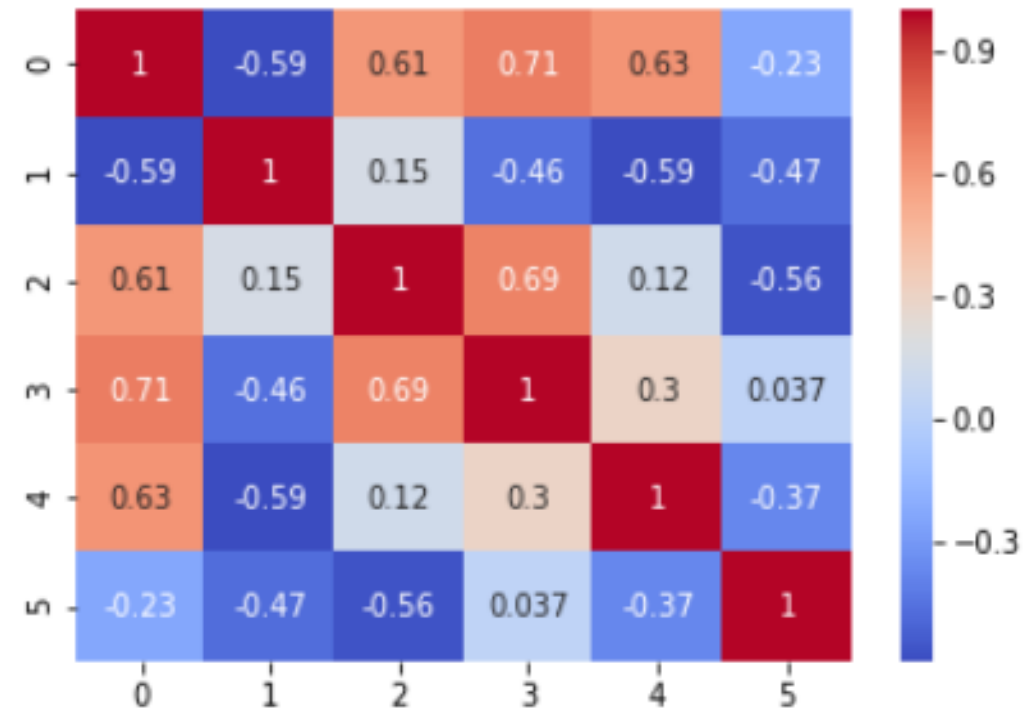
Generating results

- Based on the final dataset, three separate datasets are created to address each categorical feature individually

Case I

Product **X** of Brand **Y**

Best Seller for Product **X**

Best Seller for Brand **Y**

Product DataSet

Brand DataSet

On Criteria: **C**

On Criteria: **C**

SELLER: **A**

SELLER: **B**

Case II

Product **M** of Brand **Y**

Best Seller for Product **X**

Best Seller for Brand **Y**

Product DataSet

Brand DataSet

On Criteria: **C**

On Criteria: **C**
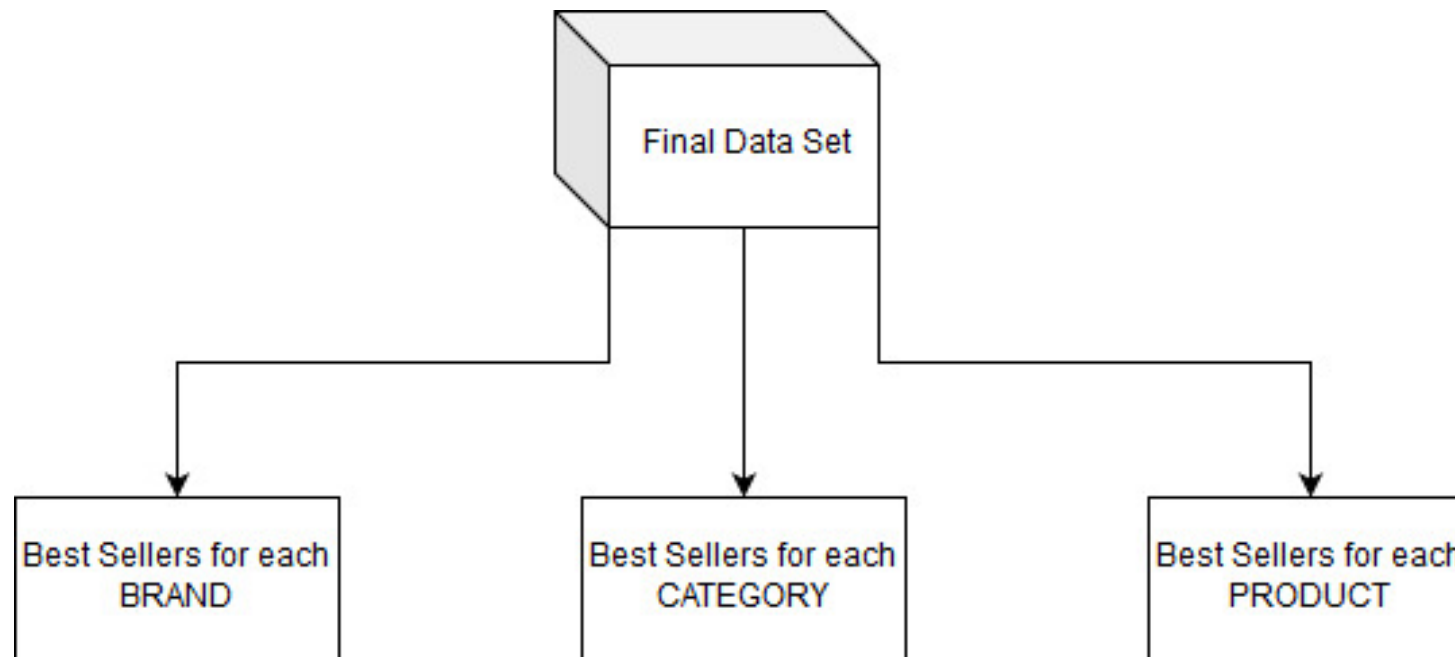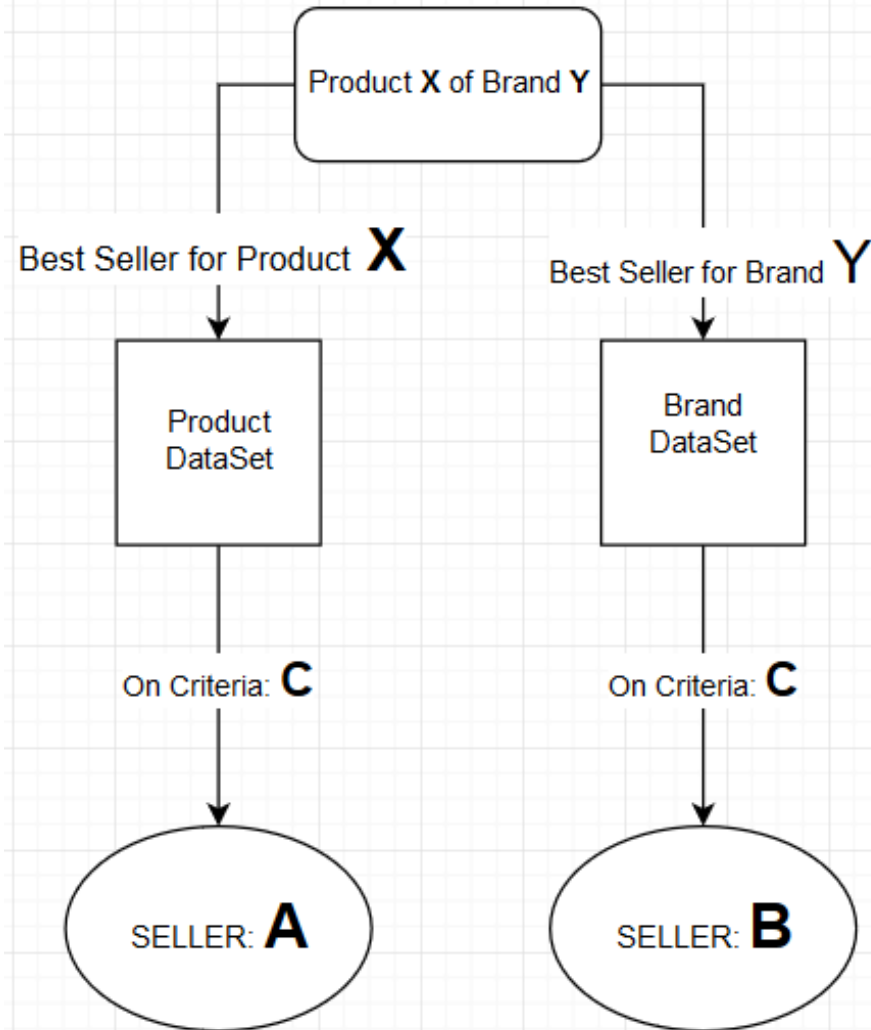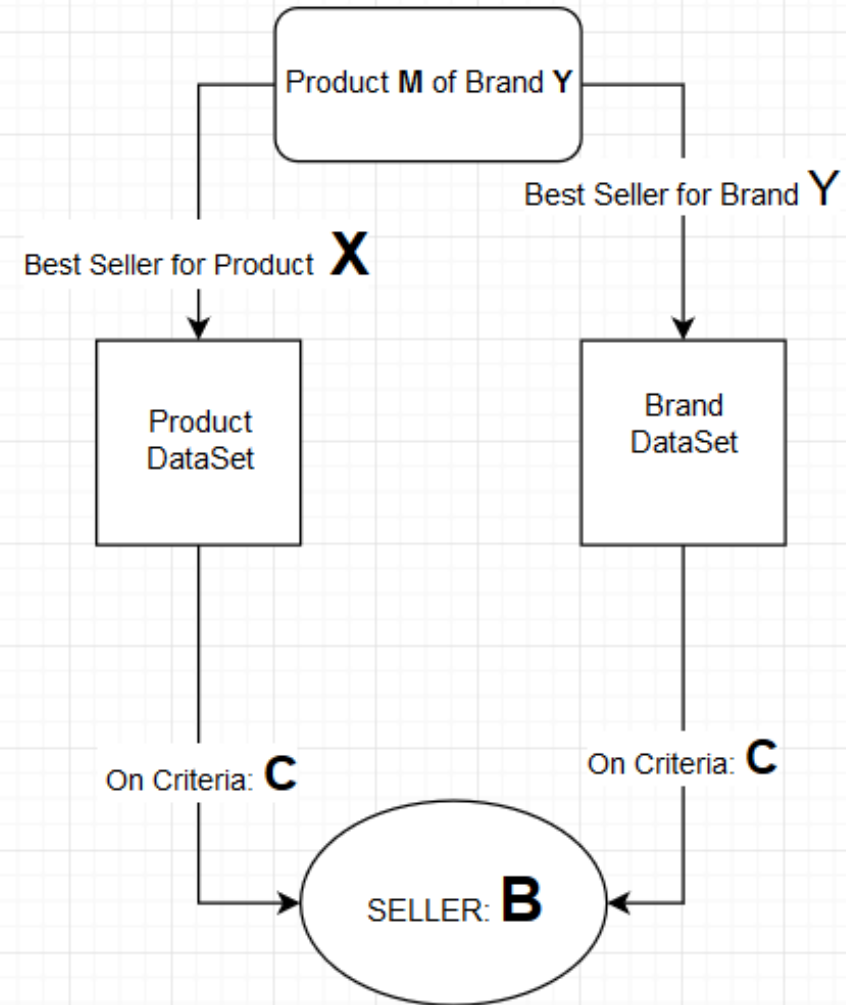
SELLER: **B**

Consumer Point-of-View

```
PRODUCT = 'Canon EOS 6D Mark II 26.2MP Digital SLR Camera Body'
CATEGORY = 'DSLR Cameras'
BRAND = 'Canon'
CITY = 'Ahmedabad'
```

Nearest Seller : MASTER ACCESSORIES { City: Sanand(29 KM)}
Best Rated : Original cartridge store (Rating = 4.893068200426682)

Based On:
Category : DSLR Cameras
Brand : Canon

	Index	INDEX_CATEGORY	Seller	Rating	City	Distance
0	Category	DSLR Cameras	MASTER ACCESSORIES	4.724764	Sanand	29
1	Brand	Canon	Original cartridge store	4.893068	Rajkot	215
2	Product	Canon EOS 6D Mark II 26.2MP Digital SLR Camera...	Meera-Enterprises	3.252713	Bhuj	331

Combined Truth Dataset:

TRUE -> Case II

FALSE -> Case I

Since the effect of both cases with respect to the features reflects upon the ALL_OVER rating, we consider that as our output.

	Product	Price	Count	Buyer_FeedBack	Courier_Rating	Product_Rating	Packaging_Rating	ALL OVER
0	100yellow Game of Thrones Quote Print Designer...	True	True	True	True	True	True	True
1	2-OYSS 10 Emoji and 10 Motivation Plastic Stam...	True	True	True	True	True	True	True
2	2-oyss Kid's Emoji Design Stamp Craft School S...	True	True	True	True	True	True	True
3	A & T Hidden Micro Mini Secret Spy Pen Camera ...	True	True	True	True	True	True	True
4	ADISA BP004 Light Weight 31 Ltrs Casual Laptop...	True	True	True	True	True	True	True

Evaluation of the Analysis (Product-Category)

	precision	recall	f1-score	support
0	0.98	1.00	0.99	617
1	0.77	0.40	0.53	25
micro avg	0.97	0.97	0.97	642
macro avg	0.87	0.70	0.76	642
weighted avg	0.97	0.97	0.97	642

```
[[614    3]
 [ 15   10]]
```


Receiver operating characteristic

ROC curve (area = 0.87)
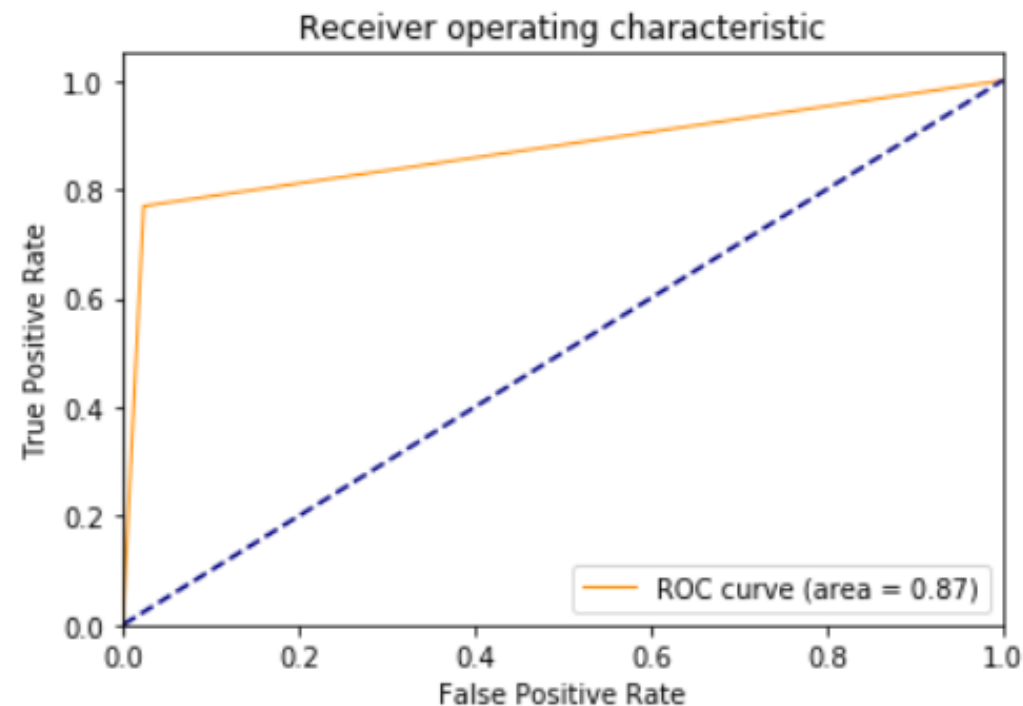
Evaluation of the Analysis (Brand-Category)

	precision	recall	f1-score	support
0	0.97	1.00	0.98	612
1	0.88	0.42	0.57	33
micro avg	0.97	0.97	0.97	645
macro avg	0.92	0.71	0.78	645
weighted avg	0.96	0.97	0.96	645

```
[[610    2]
 [ 19   14]]
```


Receiver operating characteristic — ROC curve (area = 0.92)
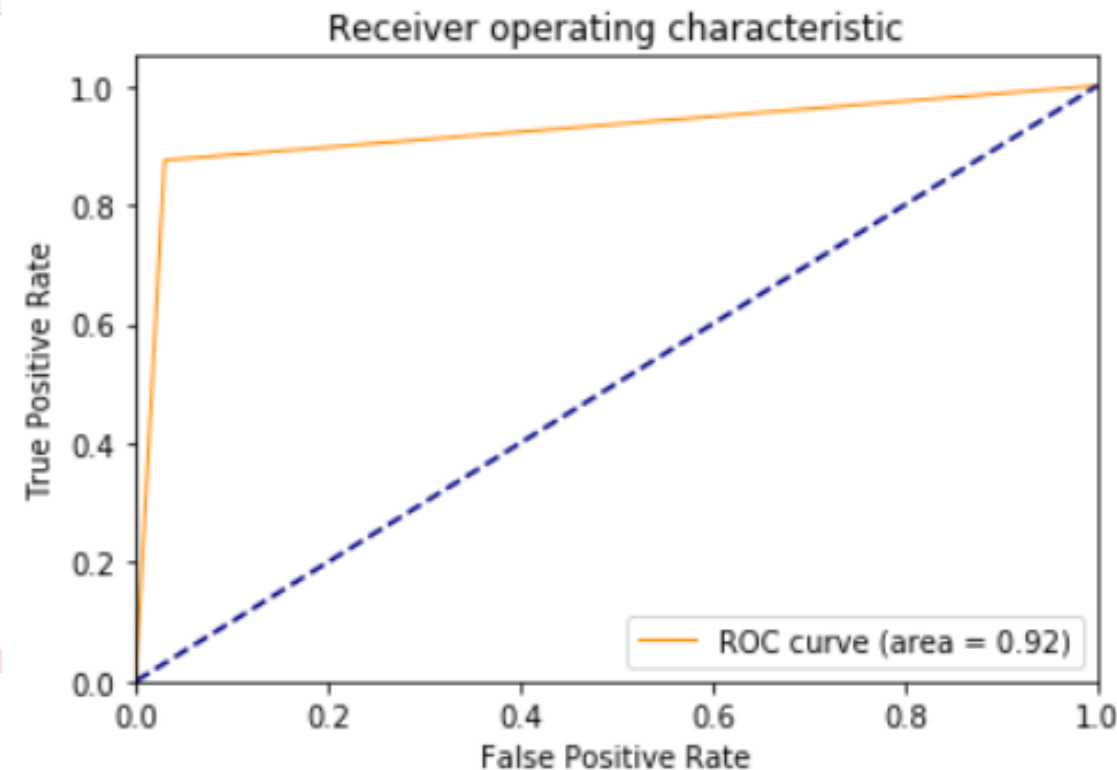
Evaluation of the Analysis (Brand-Product)