

INTRODUCTION

Understanding the fragrance market through data the perfume industry, a blend of artistry and commerce, has witnessed a significant shift towards e-commerce platforms. This dataset offers a unique opportunity to delve into the intricacies of the online perfume market. By analysing key attributes such as brand, title, type, price, currency, availability, sales, and last updates, we can uncover valuable insights into consumer preferences, pricing strategies, market trends, and the overall performance of perfume products

- Identify popular perfume brands and types: Determine which brands and fragrance categories are most sought after by online consumers.
- Analyse pricing trends: Explore variations in pricing across different brands, types and currencies.
- Evaluate product availability and sales performance: Assess the availability of products and their sales volumes to understand demand and supply dynamics.
- Discover consumer preferences: Identify patterns in product titles and descriptions to uncover customer preferences and interests
- Analyse market trends: Examine changes in sales, pricing, and availability over time to identify emerging trends and market shifts.
By addressing these objectives, this analysis will provide valuable insights for e-commerce platforms, perfume manufacturers, and marketers to optimize their strategies and better cater to the needs of online perfume shoppers

DATA EXPLANATION

The perfume e-commerce dataset comprises detailed information on 2000 perfume listings sourced from eBay, split into two separate csv files for men's and women's perfumes, each containing 1000 entries. This dataset provides a comprehensive view of the current market trends, pricing, availability, and geographical distribution of perfumes in the e-commerce space

CODING

```
import pandas as pd

mens_perfume_df = pd.read_csv("/content/drive/MyDrive/D:/ebay_mens_perfume.csv")
womens_perfume_df = pd.read_csv("/content/drive/MyDrive/D:/ebay_womens_perfume.csv")

print("Men's Perfume Dataset Info:")
print(mens_perfume_df.info())

print("\nWomen's Perfume Dataset Info:")
print(womens_perfume_df.info())
```

OUTPUT

```
Men's Perfume Dataset Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 10 columns):
 #   Column           Non-Null Count  Dtype  
---  --  
 0   brand            999 non-null    object 
 1   title             1000 non-null   object 
 2   type              997 non-null    object 
 3   price             1000 non-null   float64
 4   priceWithCurrency 1000 non-null   object 
```

```
5    available           889 non-null      float64
6  availableText        997 non-null      object
7    sold                994 non-null      float64
8  lastUpdated          947 non-null      object
9  itemLocation         1000 non-null     object
dtypes: float64(3), object(7)
memory usage: 78.2+ KB
None
```

Women's Perfume Dataset Info:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 10 columns):
 #   Column            Non-Null Count  Dtype  
---  --  
 0   brand             999 non-null    object 
 1   title              1000 non-null   object 
 2   type               998 non-null    object 
 3   price              1000 non-null   float64
 4   priceWithCurrency  1000 non-null   object 
 5   available          869 non-null    float64
 6   availableText       992 non-null    object 
 7   sold               984 non-null    float64
 8   lastUpdated         927 non-null    object 
 9   itemLocation        1000 non-null   object 
dtypes: float64(3), object(7)
memory usage: 78.2+ KB
None
```

3.1 COLUMN DESCRIPTION

S.no	Column name	Description
1	Brand	The brand of the perfume.
2	Title	The title of the listing.
3	Type	The type of perfume
4	Price	The price of the perfume.
5	Price With Currency	The price with currency notation.
6	Available	The number of items available.
7	Available Text	Text description of availability.
8	Last Updated	The last updated timestamp of the listing.
9	Sold	The number of items sold.
10	Item Location	The location of the item.

3.2 DATA OVERVIEW

The perfume e-commerce dataset provides a comprehensive overview of the perfume market, including information on popular brands, product types, pricing, sales trends, and customer behaviour. By analysing the data, businesses can gain valuable insights into consumer preferences, market trends, and identify opportunities for growth.

CODING

```
men_df = pd.read_csv("/content/perfume_men.csv", encoding='latin-1')
men_df.head()
```

```
women_df = pd.read_csv("/content/perfume_women.csv", encoding='latin-
1')
women_df.head()
```

OUTPUT

	brand	title	type	price	priceWithCurrency	available	availableText	sold	lastUpdated	itemLocation
0	Dior	Christian Dior Sauvage Men's EDP 3.4 oz Fragra...	Eau de Parfum	84.99	US \$84.99/ea	10.0	More than 10 available / 116 sold	116.0	May 24, 2024 10:03:04 PDT	Allen Park, Michigan, United States
1	AS SHOW	A-v-entus Eau de Parfum 3.3 oz 100ML Millesime...	Eau de Parfum	109.99	US \$109.99	8.0	8 available / 48 sold	48.0	May 23, 2024 23:07:49 PDT	Atlanta, Georgia, Canada
2	Unbranded	HOGO BOSS cologne For Men 3.4 oz	Eau de Toilette	100.00	US \$100.00	10.0	More than 10 available / 27 sold	27.0	May 22, 2024 21:55:43 PDT	Dearborn, Michigan, United States
3	Giorgio Armani	Acqua Di Gio by Giorgio Armani 6.7 Fl oz Eau D...	Eau de Toilette	44.99	US \$44.99/ea	2.0	2 available / 159 sold	159.0	May 24, 2024 03:30:43 PDT	Reinholds, Pennsylvania, United States
4	Lattafa	Lattafa Men's Hayaati Al Maleky EDP Spray 3.4 ...	Fragrances	16.91	US \$16.91	NaN	Limited quantity available / 156 sold	156.0	May 24, 2024 07:56:25 PDT	Brooklyn, New York, United States

	brand	title	type	price	priceWithCurrency	available	availableText	sold	lastUpdated	itemLocation
0	Carolina Herrera	Good Girl by Carolina Herrera 2.7 oz Eau De Pa...	Eau de Parfum	43.99	US \$43.99/ea	2.0	2 available / 393 sold	393.0	May 23, 2024 10:43:50 PDT	Thomasville, Alabama, United States
1	As Shown	Parfums de Marly Delina La Rosee Eau de Parfum...	Eau de Parfum	79.99	US \$79.99	5.0	5 available / 40 sold	40.0	May 24, 2024 00:15:48 PDT	New Jersey, Hong Kong
2	PRADA	PRADA Paradoxe by Prada EDP 3.0oz/90ml Spray P...	Eau de Parfum	59.99	US \$59.99	10.0	More than 10 available / 35 sold	35.0	May 14, 2024 20:54:25 PDT	Orange, New Jersey, United States
3	As Show	J'adore Parfum D'eau by Christian 3.4 oz EDP F...	Eau de Parfum	59.99	US \$59.99/ea	10.0	More than 10 available / 9 sold	9.0	May 23, 2024 01:23:05 PDT	USA, New Jersey, Hong Kong
4	Khadlaj	Shiyaaka for Men EDP Spray 100ML (3.4 FL.OZ) B...	Eau de Parfum	29.99	US \$29.99/ea	10.0	More than 10 available	Nan	Nan	Little Ferry, New Jersey, United States

3.3 EXISTING DATA

The perfume e-commerce dataset encompasses a wide range of perfume products, offering insights into brand diversity, pricing variations, availability status, sales performance, and geographical distribution. By examining the dataset's columns, we can gain a comprehensive understanding of the factors influencing perfume purchases and trends in the e-commerce market.

CODING

```
men_df.shape
women_df.shape
```

```
men_df['sex'] = 'men'
women_df['sex'] = 'women'

df= pd.concat([men_df, women_df], ignore_index=True)
df.columns
```

```
df.isnull().sum()
```

OUTPUT

```
(1000, 10)
```

```
(1000, 10)
```

```
Index(['brand', 'title', 'type', 'price', 'priceWithCurrency',
       'available',
       'availableText', 'sold', 'lastUpdated', 'itemLocation',
       'sex'],
      dtype='object')
```

```
brand                2
title               0
type                5
price               0
priceWithCurrency  0
available          242
availableText       11
sold                22
lastUpdated         126
itemLocation        0
sex                 0
dtype: int64
```

3.4 DATA TRANSFORMATION

Data transformation is a crucial step in the analysis of the perfume e-commerce dataset, as it prepares the raw data for further exploration and modeling. This process involves various techniques to clean, transform, and enhance the data quality, ensuring that it is suitable for analysis and yielding accurate results. The perfume e-commerce dataset may require data transformation to ensure data consistency, accuracy, and suitability for analysis.

CODING

```
pd.set_option('display.max_columns', None)
df1 = pd.read_csv('/content/perfume men.csv', encoding='latin-1')
df2 = pd.read_csv('/content/perfume women.csv', encoding='latin-1')
df1['Gender'] = 'male'
df2['Gender'] = 'female'
perfume_analysis_db= pd.concat([df1, df2], ignore_index=True)
perfume_analysis_db.tail()
```

OUTPUT

	brand	title	type	price	priceWithCurrency	available	availableText	sold	lastUpdated	itemLocation	Gender
1995	Avon	Avon Far Away Infinity Eau de Parfum 1.7 fl. o...	Eau de Parfum	13.89	US \$13.89	10.0	More than 10 available / 157 sold	157.0	May 16, 2024 22:35:29 PDT	West Palm Beach, Florida, United States	female
1996	Mancera	Roses Greedy by Mancera perfume for unisex EDP...	Eau de Parfum	57.85	US \$57.85/ea	33.0	33 available / 58 sold	58.0	May 24, 2024 08:03:11 PDT	Dallas, Texas, United States	female
1997	Unbranded	Sweet Tooth Eau de Parfum, Perfume for Women, ...	1	30.96	US \$30.96	2.0	2 available / 3 sold	3.0	May 17, 2024 23:16:41 PDT	New York, New York, United States	female
1998	Juliette Has A Gun	MMMM BY Juliette Has A Gun perfume for her EDP...	Eau de Perfume	53.99	US \$53.99/ea	3.0	3 available / 117 sold	117.0	May 13, 2024 22:19:34 PDT	Dallas, Texas, United States	female
1999	Paris Hilton	PARIS HILTON ELECTRIFY for Women Cologne 3.4 o...	Eau de Parfum	14.99	US \$14.99/ea	4.0	4 available / 51 sold	51.0	May 22, 2024 05:44:45 PDT	TX, United States	female

DATA PREPROCESSING

Data preprocessing is the process of preparing raw data for analysis. It involves cleaning, transforming, and organizing the data to ensure its quality and consistency. This step is crucial in data analysis as it can significantly impact the accuracy and reliability of the results.

- **Data Cleaning**

Data Cleaning is a crucial step before any data analysis. It ensures data quality and accuracy, which ultimately impacts the reliability of the insights derived. Here are some common data cleaning tasks

- **Handling Missing Values**

Address missing values in columns like price, brand, or type by imputing them using appropriate methods (e.g., mean, median, mode) or removing rows with excessive missing data.

- **Correcting Inconsistent Data**

Identify and rectify inconsistencies in data formats, such as inconsistent currency symbols or date formats.

- **Removing Outliers**

Detect and handle outliers in numerical columns like price or sales data to prevent skewing analysis results.

- **Data Reduction**

Feature Selection: Identify and select the most relevant features for analysis to reduce dimensionality and improve model performance.

CODING

```
perfume_analysis_db.duplicated().sum()  
perfume_analysis_db.isnull().sum()
```

```
perfume_analysis_db.iloc[950:1051]
```

OUTPUT

```

brand          0
title          0
type           0
price          0
priceWithCurrency 0
available       0
availableText   0
sold            0
lastUpdated     0
itemLocation    0
sex             0
dtype: int64

```

	brand	title	type	price	priceWithCurrency	available	availableText	sold	lastUpdated	itemLocation	Gender
950	Lapidus	LAPIDUS pour Homme by Ted Lapidus Cologne 3.3 ...	Eau de Toilette	20.08	US \$20.08/ea	29.0	29 available / 518 sold	518.0	Apr 23, 2024 03:57:22 PDT	Dallas, Texas, United States	male
951	Arabian Oud	Arabian Oud - Arabian Knight EDP For Men 3.4 o...	Fragrances	99.63	US \$99.63	3.0	3 available / 13 sold	13.0	May 22, 2024 09:26:42 PDT	Brooklyn, New York, United States	male
952	Emporio Armani	Emporio Armani Stronger With You ABSOLUTELY 1....	Perfume	74.99	US \$74.99/ea	10.0	10 available / 82 sold	82.0	May 23, 2024 05:39:50 PDT	Katy, Texas, United States	male
953	Armaf	Hunter by Armaf cologne for men EDP 3.3 / 3.4 ...	Eau de Parfum	23.62	US \$23.62/ea	130.0	130 available / 415 sold	415.0	May 22, 2024 09:23:27 PDT	Dallas, Texas, United States	male
954	Perry Ellis	360 Red by Perry Ellis 6.7 / 6.8 oz EDT Cologn...	Eau de Toilette	40.13	US \$40.13	9.0	9 available / 2,703 sold	2703.0	May 22, 2024 15:05:34 PDT	Hackensack, New Jersey, United States	male
...
1046	Parfums de Marly	PARFUMS de MARLY DELINA for WOMEN 2.5 oz (75ml...)	Eau de Parfum	100.99	US \$100.99	0.0	Last One / 49 sold	49.0	May 17, 2024 03:05:11 PDT	Buffalo Mills, Pennsylvania, United States	female
1047	Juliette has a gun	Juliette Has A Gun Discovery Kit includes LUST...	Eau de Parfum	26.05	US \$26.05/ea	10.0	More than 10 available / 128 sold	128.0	May 19, 2024 15:38:07 PDT	Winter Garden, Florida, United States	female
1048	AS SHOWN	Flora Gorgeous Jasmine 3.3 oz / 100 ml EDP Per...	Eau de Parfum	54.99	US \$54.99/ea	5.0	5 available / 60 sold	60.0	May 24, 2024 01:42:35 PDT	USA, California, Hong Kong	female
1049	Carolina Herrera	Carolina Herrera's Good Girl 2.7 Oz □ Women's ...	Eau de Parfum	49.99	US \$49.99/ea	6.0	6 available / 179 sold	179.0	May 21, 2024 07:50:00 PDT	Dearborn, Michigan, United States	female
1050	Viktor & Rolf	FLOWERBOMB BY VIKTOR & ROLF 3.4 OZ SPRAY EAU D...	Eau de Parfum	47.49	US \$47.49/ea	2.0	2 available / 208 sold	208.0	May 24, 2024 02:09:07 PDT	Houston, Texas, United States	female

101 rows × 11 columns

FEATURE SELECTION

The perfume e-commerce dataset contains relevant columns that can be used for feature selection in data analysis. Here are some potential features and their significance:

Categorical features:

Brand: This feature can be used to analyse the popularity of different brands and their impact on sales.

Type: This feature can be used to understand the demand for different types of perfumes (e.g., floral, fruity, woody).

Availability text: This feature can be used to determine the availability of products and identify potential stock issues.

Numerical features:

Price: This feature can be used to analyse the relationship between price and sales, as well as the impact of price on customer behaviour.

Price with currency: This feature can be used to analyse price differences across different currencies and regions.

Sold: This feature can be used to measure the popularity of products and identify best-sellers.

Time-based features:

Last updated: This feature can be used to analyse trends in product availability, pricing, and sales over time.

DATA ANALYSIS

Data analysis is the process of examining the perfume e-commerce dataset to extract meaningful insights and answer specific research questions. It involves applying statistical techniques and analytical methods to uncover patterns, trends, and relationships within the data.

BRAND POPULARITY AND PRICING STRATEGIES

We will investigate the popularity of different perfume brands and analyse their pricing strategies. By examining the sales data and price points of various brands, we can identify the most successful brands and gain insights into their pricing patterns. This analysis will help us understand the relationship between brand popularity and pricing, and provide valuable information for businesses to optimize their pricing strategies.

CODING

```
import pandas as pd # import pandas
perfume_df = pd.read_csv('modified_file.csv') # Read the contents of
the file 'modified_file.csv' into a pandas DataFrame
# Calculate sales and average price for each brand
brand_sales = perfume_df.groupby('brand')['sold'].sum().reset_index()
brand_avg_price = perfume_df.groupby('brand')['price'].mean().reset_in-
dex()
# Merge sales and average price data
brand_analysis_df = pd.merge(brand_sales, brand_avg_price, on='brand')
brand_analysis_df.columns = ['Brand', 'Total Sales', 'Average Price']
# Sort brands by total sales in descending order
brand_analysis_df = brand_analysis_df.sort_values('Total Sales', as-
cending=False)
# Display the top 10 brands by sales
top_brands = brand_analysis_df.head(10)
print("Top 10 Brands by Sales:")
print(top_brands)
# Visualize the relationship between brand popularity and pricing with
a bubble chart
plt.figure(figsize=(12, 8))
```

```

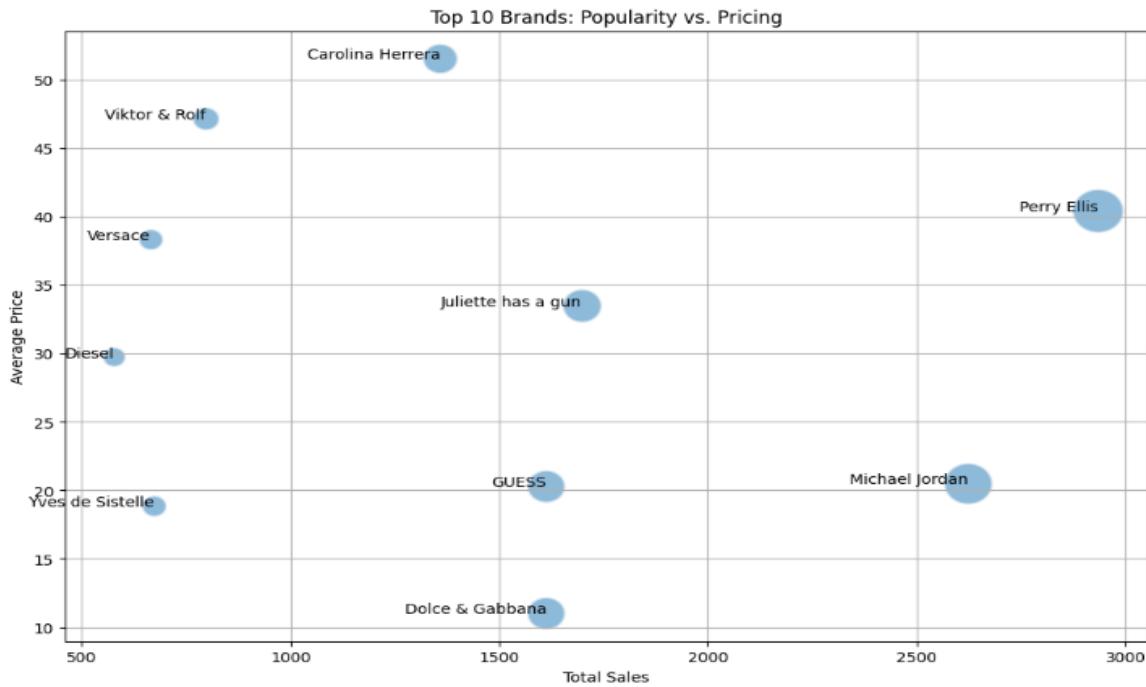
bubble_size = top_brands['Total Sales'] / top_brands['Total
Sales'].max() * 1000 # Adjust bubble size for visibility
plt.scatter(top_brands['Total Sales'], top_brands['Average Price'],
s=bubble_size, alpha=0.5, edgecolors="w", linewidth=2)
# Add labels to each bubble
for i in range(len(top_brands)):
    plt.text(top_brands['Total Sales'].iloc[i], top_brands['Average
Price'].iloc[i], top_brands['Brand'].iloc[i], fontsize=10, ha='right')
plt.xlabel('Total Sales')
plt.ylabel('Average Price')
plt.title('Top 10 Brands: Popularity vs. Pricing')
plt.grid(True)
plt.show()

```

OUTPUT

Top 10 Brands by Sales:

	Brand	Total Sales	Average Price
54	Perry Ellis	2935.0	40.425000
48	Michael Jordan	2624.0	20.480000
36	Juliette has a gun	1699.0	33.485000
24	GUESS	1613.0	20.280000
16	Dolce & Gabbana	1613.0	11.000000
9	Carolina Herrera	1359.0	51.552500
67	Viktor & Rolf	798.0	47.160000
71	Yves de Sistelle	674.0	18.840000
66	Versace	666.0	38.326667
15	Diesel	578.0	29.740000



COMPARATIVE ANALYSIS OF MEN'S AND WOMEN'S PERFUME MARKETS

We will perform a comparative analysis of men's and women's perfume markets using the Perfume E-Commerce Dataset 2024. By examining key metrics such as pricing, sales, and brand representation, we aim to identify similarities, differences, and unique market dynamics between the two segments. This analysis will provide valuable insights into gender-specific preferences and help businesses tailor their strategies accordingly.

CODING

```
avg_price_men = mens_perfume_df['price'].mean()
avg_price_women = womens_perfume_df['price'].mean()
print(f"Average price of men's perfumes: ${avg_price_men:.2f}")
print(f"Average price of women's perfumes: ${avg_price_women:.2f}")
```

Average price of men's perfumes: \$46.48

Average price of women's perfumes: \$39.89

```

total_sales_men = mens_perfume_df['sold'].sum()
total_sales_women = womens_perfume_df['sold'].sum()
print(f"Total sales of men's perfumes: {total_sales_men}")
print(f"Total sales of women's perfumes: {total_sales_women}")

```

Total sales of men's perfumes: 761669.0
 Total sales of women's perfumes: 489364.0

```

mens_brands = mens_perfume_df['brand'].nunique()
womens_brands = womens_perfume_df['brand'].nunique()
print(f"Number of unique brands in men's perfumes: {mens_brands}")
print(f"Number of unique brands in women's perfumes: {womens_brands}")

```

Number of unique brands in men's perfumes: 246
 Number of unique brands in women's perfumes: 247

PERFUME TYPE PREFERENCES

We will analyse consumer preferences for different perfume types (e.g., eau de parfum, eau de toilette) and how they vary across brands, price points, and customer segments. By examining the sales data and perfume types, we can gain insights into customer preferences and identify popular perfume types within specific brands or price ranges. This analysis will help businesses tailor their product offerings and marketing strategies to meet customer demands.

CODING

```

#Group by perfume type and calculate total sales
type_sales = perfume_df.groupby('type')['sold'].sum().reset_index()
# Sort perfume types by total sales in descending order
type_sales = type_sales.sort_values('sold', ascending=False)
# Display the perfume types by sales
print("Perfume Types by Sales:")
print(type_sales)
# Visualize the sales distribution by perfume type
plt.figure(figsize=(10, 6))
plt.bar(type_sales['type'], type_sales['sold'])

```

```

plt.xlabel('Perfume Type')
plt.ylabel('Total Sales')
plt.title('Sales Distribution by Perfume Type')
plt.xticks(rotation=45)
plt.show()

# Analyze perfume type preferences by brand
brand_type_sales = perfume_df.groupby(['brand', 'type'])['sold'].sum().reset_index()
brand_type_sales = brand_type_sales.sort_values(['brand', 'sold'], ascending=[True, False])

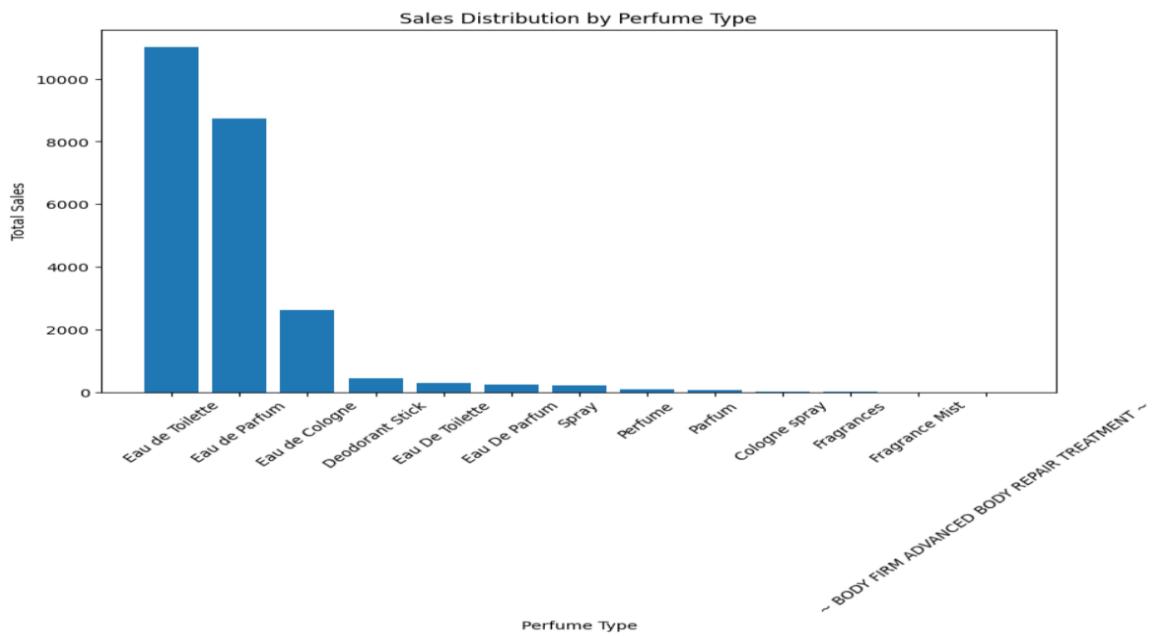
# Display perfume type preferences by brand for the top 5 brands
top_brands = brand_type_sales['brand'].unique()[:5]
for brand in top_brands:
    print(f"Perfume Type Preferences for {brand}:")
    print(brand_type_sales[brand_type_sales['brand'] == brand])
    print()

```

OUTPUT

Perfume Types by Sales:

	type	sold
53	Eau de Toilette	733114.0
45	Eau de Parfum	268725.0
43	Eau de Cologne	51536.0
52	Eau de Perfume	22917.0
51	Eau de Parfume	21211.0
..
2	3 Pc	2.0
92	Perfume Gift Sets	1.0
114	~ BODY FIRM ADVANCED BODY REPAIR TREATMENT ~	1.0
24	Discovery Set	0.0
73	Fragrance Oil	0.0



[116 rows x 2 columns]

Perfume Type Preferences for 2nd To None:

	brand	type	sold
0	2nd To None	Concentrated Uncut Pure Body Oil	18882.0

Perfume Type Preferences for AERIN:

	brand	type	sold
1	AERIN	Eau de Parfum	50.0

Perfume Type Preferences for ALFRED SUNG:

	brand	type	sold
3	ALFRED SUNG	Eau de Toilette	5627.0
2	ALFRED SUNG	Eau de Parfum	1516.0

Perfume Type Preferences for ALT Fragrances:

	brand	type	sold
4	ALT Fragrances	Extrait de Parfum	686.0

Perfume Type Preferences for AS SHOWN:

	brand	type	sold
5	AS SHOWN	Eau de Toilette	12

DATA VISUALIZATION.

Data visualization is the graphical representation of data. It is a powerful tool for understanding and communicating the insights derived from the perfume e-commerce dataset. By visually representing the data, we can effectively explore trends, patterns and relationships that might be difficult to discern through numerical analysis alone

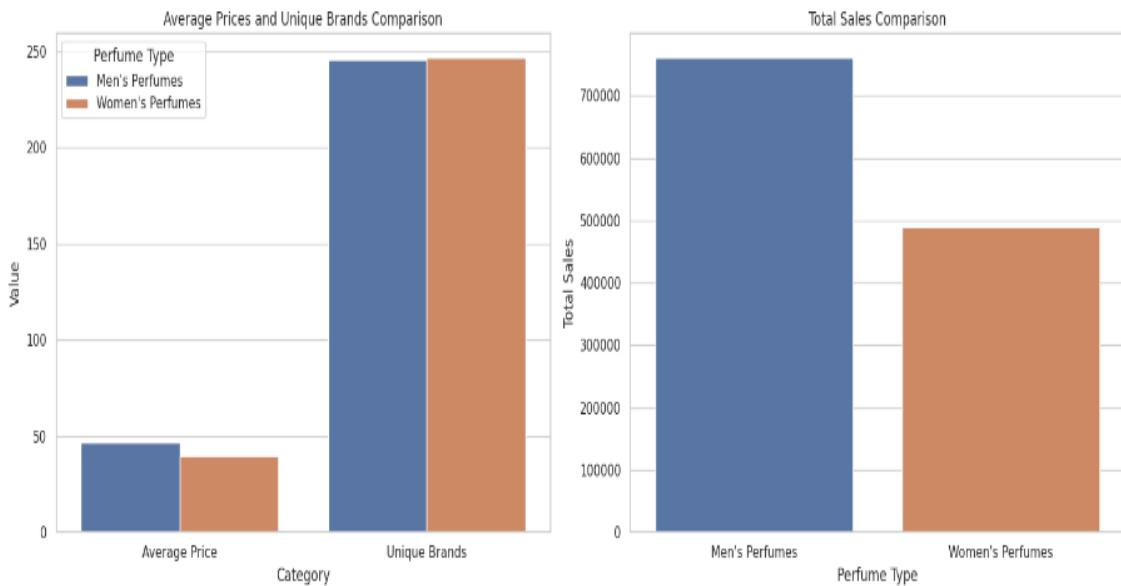
Data visualization can be used for a variety of purposes, such as:

- Communicating complex information: making complex data accessible to a wider audience.
- Identifying patterns and trends: spotting trends or anomalies that might not be obvious from raw data.
- Telling stories with data: creating compelling narratives using visual elements.
- Making data-driven decisions: supporting decision-making by providing visual insights.

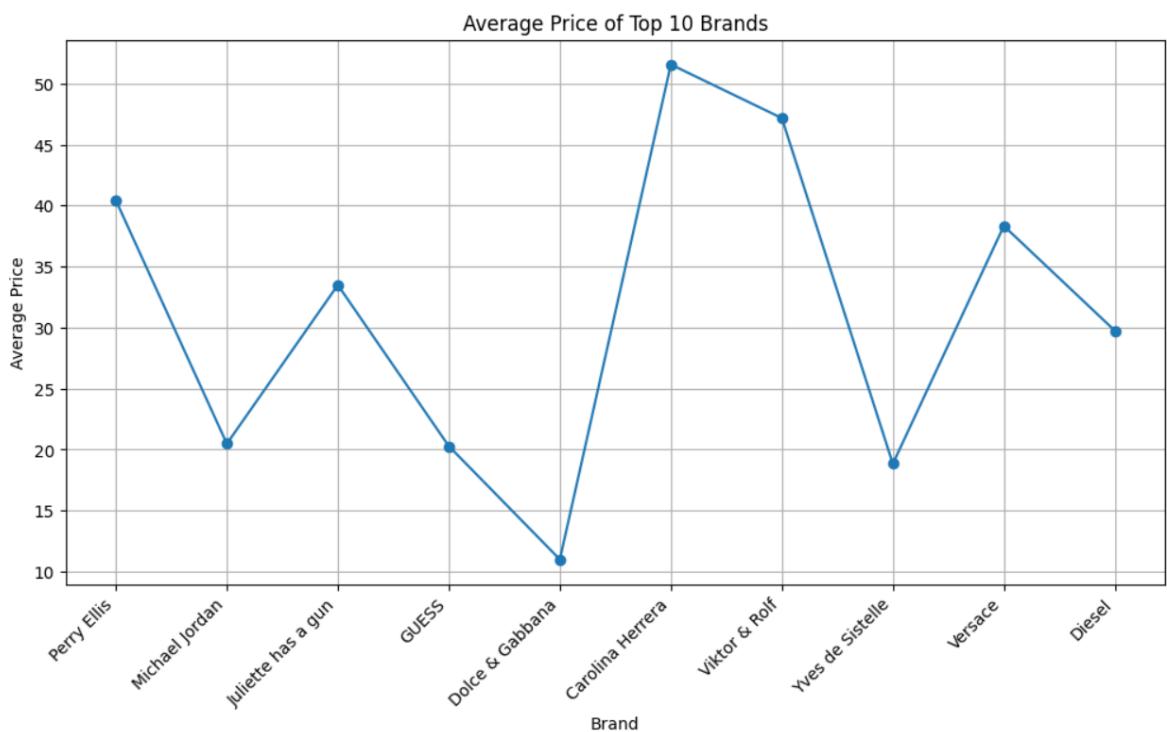
There are many different types of data visualizations, each with its own strengths and weaknesses. Some common types include:

- **Bar charts:** to compare categorical variables such as brand popularity, product type, or customer demographics
- **Line charts:** to visualize trends over time, such as changes in sales volume or product availability
- **Pie charts:** to represent proportions or percentages of categorical data like the distribution of product types
- **Scatter plots:** to examine relationships between numerical variables, like correlation between price and sales.

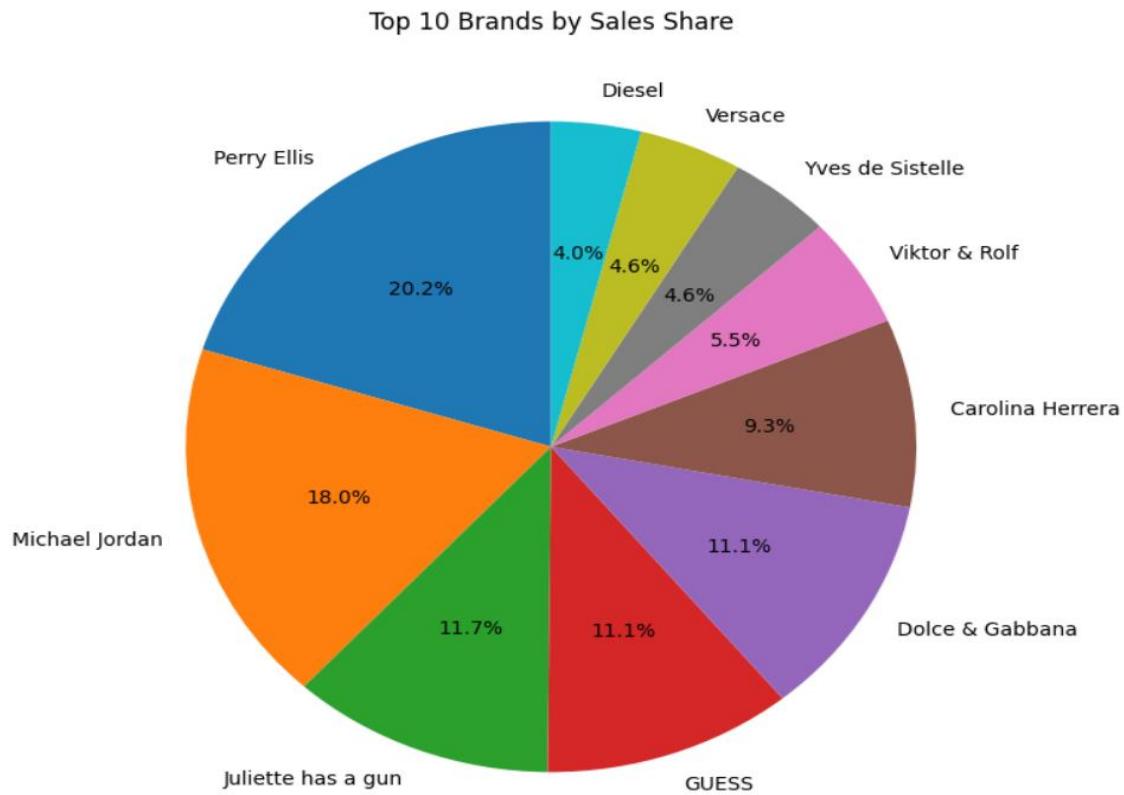
BAR CHART



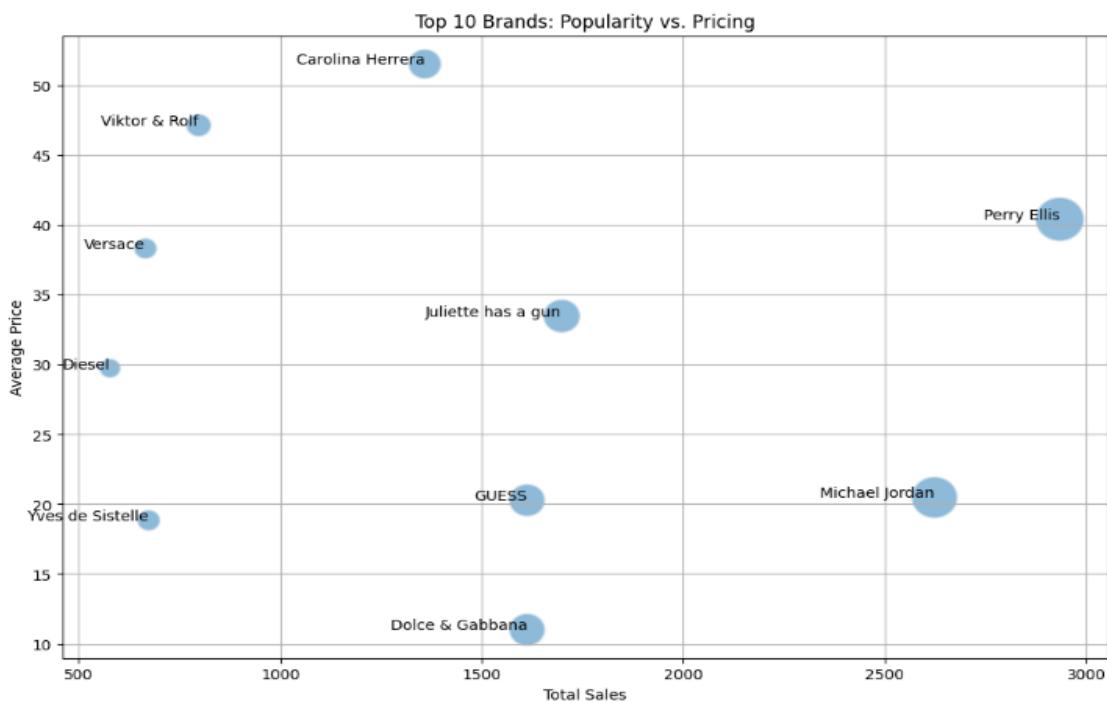
LINE CHART



PIE CHART



SCATTER PLOT



CONCLUSION

Data Analysis on Perfume E-commerce Dataset Based on the analysis of the perfume e-commerce dataset, the following key findings were identified:

- Brand Popularity: [Insert key findings about brand popularity, such as which brands are most popular, which brands have the highest sales, etc.]
- Product Type Trends: [Insert key findings about product type trends, such as which types of perfumes are most in demand, which types have the highest sales, etc.]
- Price Sensitivity: [Insert key findings about price sensitivity, such as how changes in price affect sales, whether customers are more price-conscious for certain brands or types, etc.]

Sales Performance: [Insert key findings about sales performance, such as which products are the best-sellers, which products have the highest sales growth, etc.]

Overall, the analysis provides valuable insights into the perfume e-commerce market, including consumer preferences, pricing strategies, and product availability. These findings can be used by businesses to improve their marketing efforts, optimize their product offerings, and enhance their overall performance in the market.

Further analysis could be conducted to explore additional factors such as customer reviews, seasonal trends, and regional variations in demand. This would provide a more comprehensive understanding of the perfume e-commerce market and enable businesses to make more informed decisions

REFERENCE

<https://scholar.google.com/>

<https://www.jstor.org/>

<https://www.linkedin.com/company/ecommerce-foundation/>

<https://www.isical.ac.in/>

<https://www.marketresearch.com/>