1. **Geospatial Analysis: Utilize distribution_centers.csv and users.csv for mapping and analysing the geographic distribution of users and distribution centres.**

For first problem firstly I have imported all necessary libraries along with datasets and did some EDA in the data and fetch insights from it following the problem statement

In [107]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from mpl_toolkits.basemap import Basemap
import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style("whitegrid")
from mpl_toolkits.basemap import Basemap
from geopy.geocoders import Nominatim
```

Geospatial Analysis:

In [3]:
```
# Load data
distribution_centers = pd.read_csv(r"C:\Users\Harshit\Downloads\data\data\distribution_centers.csv")
users = pd.read_csv(r"C:\Users\Harshit\Downloads\data\data\users.csv")
```

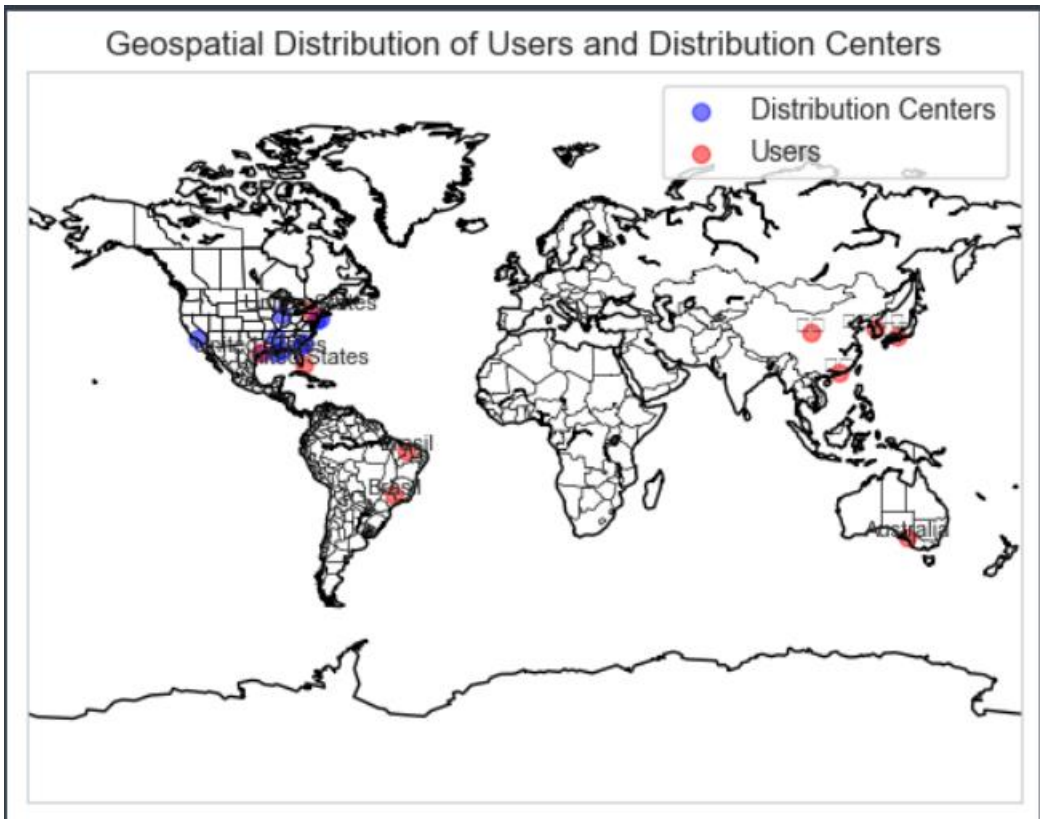
In [4]:
```
distribution_centers.head()
```

	id	name	latitude	longitude
0	1	Memphis TN	35.1174	-89.9711
1	2	Chicago IL	41.8369	-87.6847
2	3	Houston TX	29.7604	-95.3698
3	4	Los Angeles CA	34.0500	-118.2500
4	5	New Orleans LA	29.9500	-90.0667

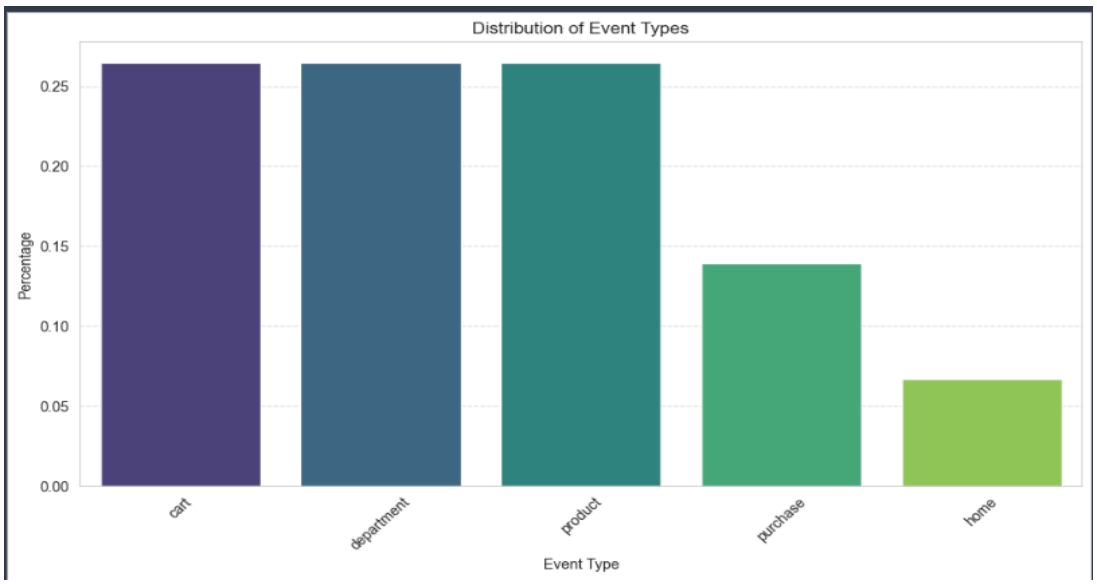
In [5]:
```
users.head()
```

	id	first_name	last_name	email	age	gender	state	street_address	postal_code	city	country
0	457	Timothy	Bush	timothybush@example.net	65	M	Acre	87620 Johnson Hills	69917-400	Rio Branco	Brasil
1	6578	Elizabeth	Martinez	elizabethmartinez@example.com	34	F	Acre	1705 Nielsen Land	69917-400	Rio Branco	Brasil
2	36280	Christopher	Mendoza	christophermendoza@example.net	13	M	Acre	125 Turner Isle Apt. 264	69917-400	Rio Branco	Brasil
3	60193	Jimmy	Conner	jimmyconner@example.com	64	M	Acre	0966 Jose Branch Apt. 008	69917-400	Rio Branco	Brasil
4	64231	Natasha	Wilson	natashawilson@example.net	25	F	Acre	20798 Phillip Trail Apt. 392	69917-400	Rio Branco	Brasil

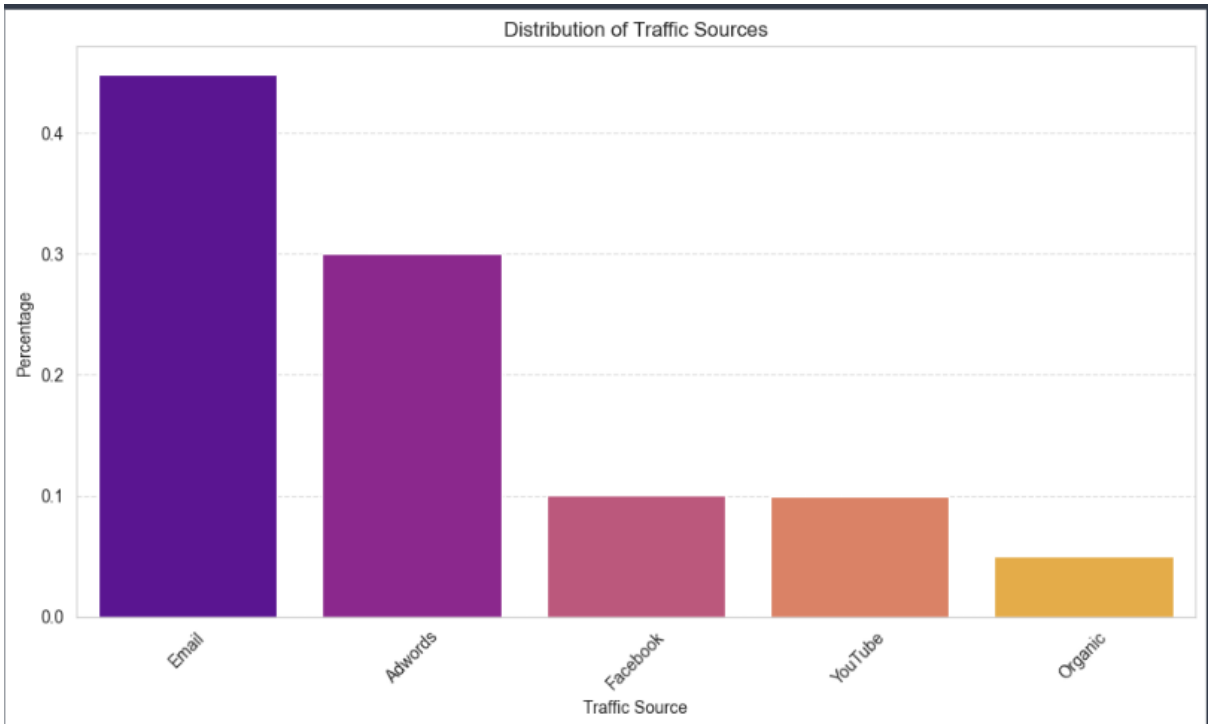
And now as given below visualization we can easily see the geographical distribution and centres.



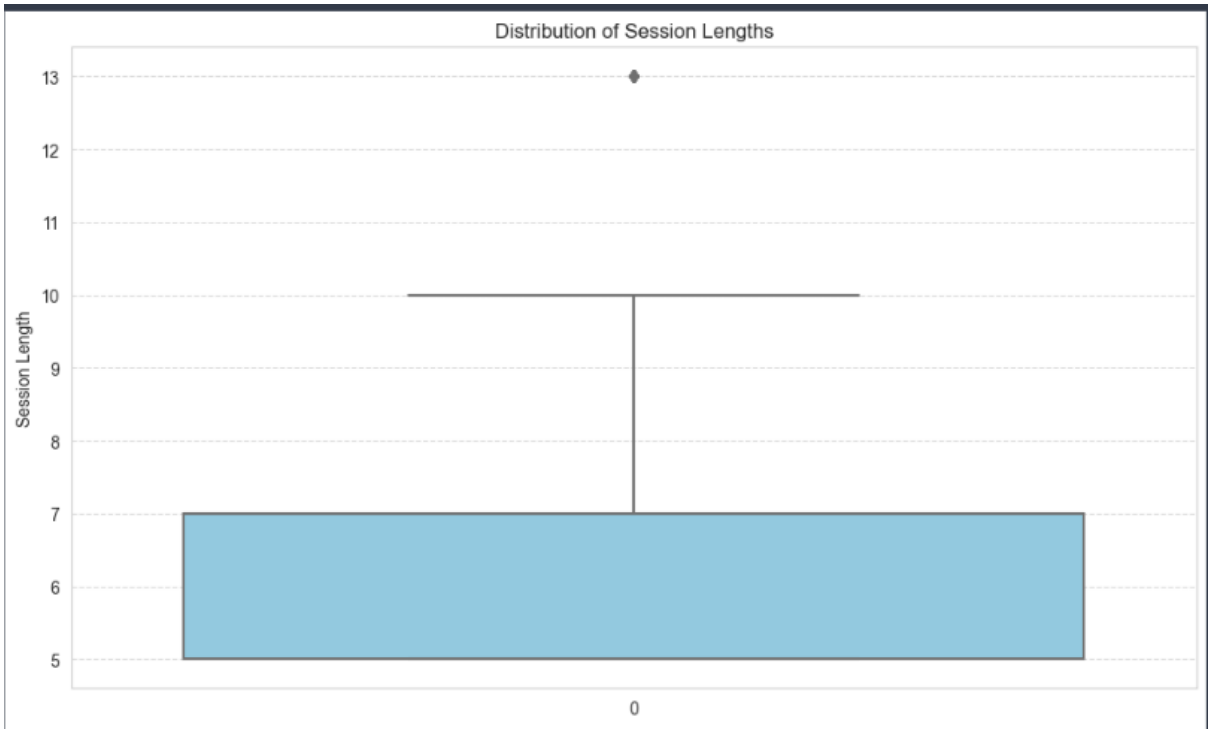
2. User Behaviour Analysis: Use events.csv to analyse user behaviour, including session patterns, traffic sources, and event types.



Event type distribution



Traffic source distribution



Session length distribution using a boxplot

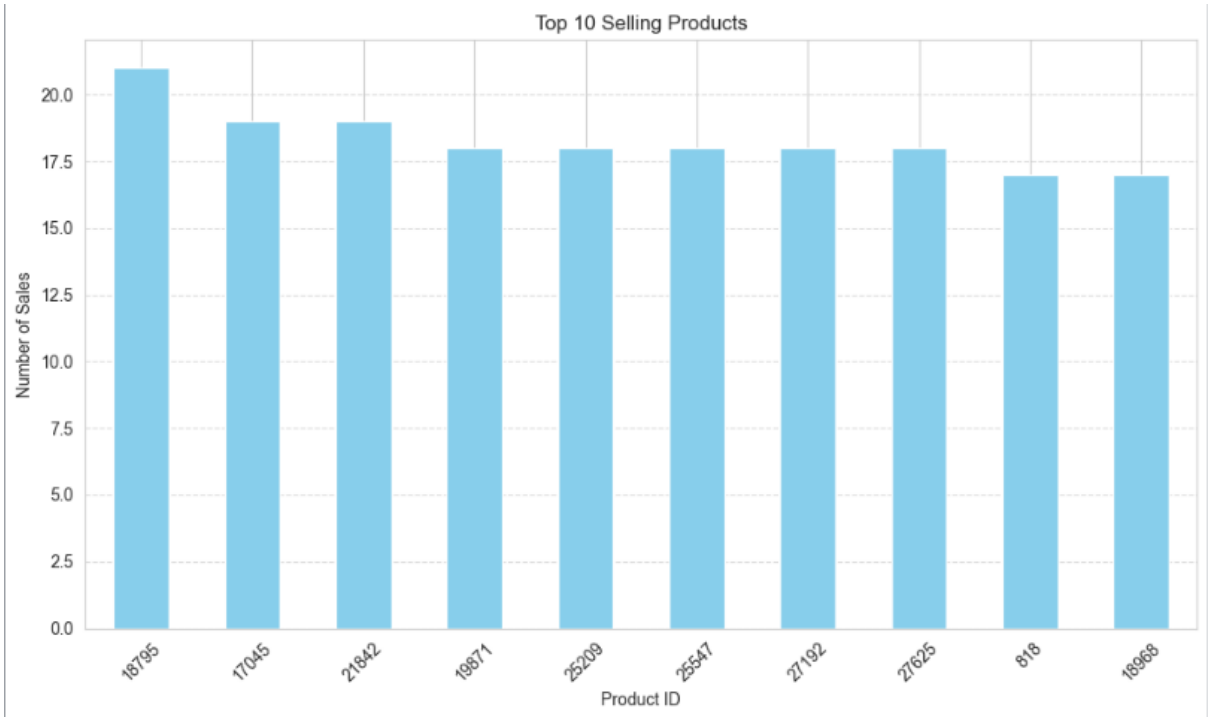
3. Sales and Revenue Analysis: Leverage order_items.csv and inventory_items.csv to analyse product sales, revenue, and profitability.

This is the table of inventory items in which there was missing value at very high percentage in the column name sold at so I have filled that NaN value with “Not Sold”.

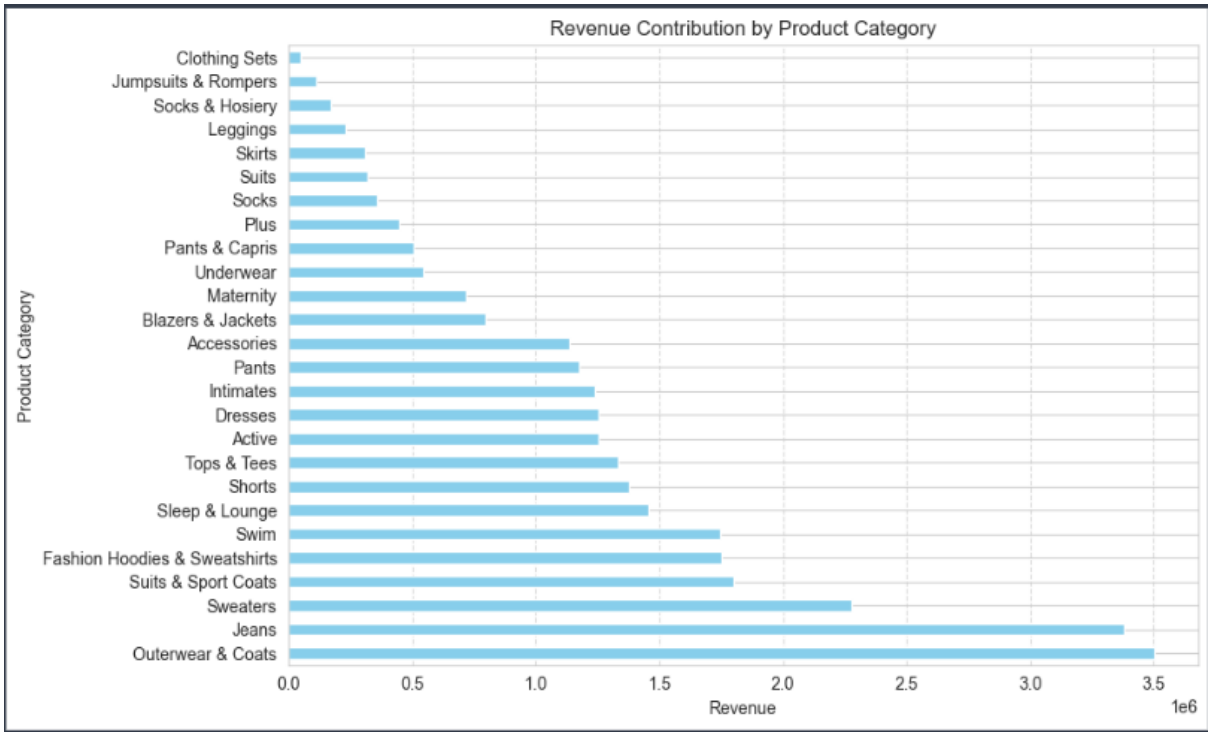
```
In [115]:  1  inventory_items.head()
```

	id	product_id	created_at	sold_at	cost	product_category	product_name	product_brand	product_retail_pric
0	67971	13844	2022-07-02 07:09:20+00:00	2022-07-24 06:33:20+00:00	2.76804	Accessories	(ONE) 1 Satin Headband	Funny Girl Designs	6.99
1	67972	13844	2023-12-20 03:28:00+00:00	Not Sold	2.76804	Accessories	(ONE) 1 Satin Headband	Funny Girl Designs	6.99
2	67973	13844	2023-06-04 02:53:00+00:00	Not Sold	2.76804	Accessories	(ONE) 1 Satin Headband	Funny Girl Designs	6.99
3	72863	13844	2021-10-16 22:58:52+00:00	2021-11-22 02:19:52+00:00	2.76804	Accessories	(ONE) 1 Satin Headband	Funny Girl Designs	6.99
4	72864	13844	2021-08-07 16:33:00+00:00	Not Sold	2.76804	Accessories	(ONE) 1 Satin Headband	Funny Girl Designs	6.99

Bar plot for top-selling products



Revenue for each product category



Total Profit for Each Year

```
Total Profit for Each Year:
year
2018    -5.871489e+03
2019    -5.648322e+05
2020    -1.261111e+07
2021    -1.403636e+07
2022    -1.646481e+07
2023    -2.255754e+07
2024    -6.352898e+05
dtype: float64
```

4. **Product Performance Analysis: Explore products.csv to analyse product performance, including costs, categories, and popularity.**

Here's how data looks like:

	id	cost	category	name	brand	retail_price	department	sku	distribution_center
0	13842	2.51875	Accessories	Low Profile Dyed Cotton Twill Cap - Navy W39S55D	MG	6.25	Women	EBD58B8A3F1D72F4206201DA62FB1204	1
1	13928	2.33835	Accessories	Low Profile Dyed Cotton Twill Cap - Putty W39S55D	MG	5.95	Women	2EAC42424D12436BDD6A5B8A88480CC3	1
2	14115	4.87956	Accessories	Enzyme Regular Solid Army Caps- Black W35S45D	MG	10.99	Women	EE364229B2791D1EF9355708EFF0BA34	1
3	14157	4.64877	Accessories	Enzyme Regular Solid Army Caps- Olive W35S45D (...	MG	10.99	Women	00BD13095D06C20B11A2993CA419D16B	1
4	14273	6.50793	Accessories	Washed Canvas Ivy Cap - Black W11S64C	MG	15.99	Women	F531DC20FDE20B7ADF3A73F52B71D0AF	1

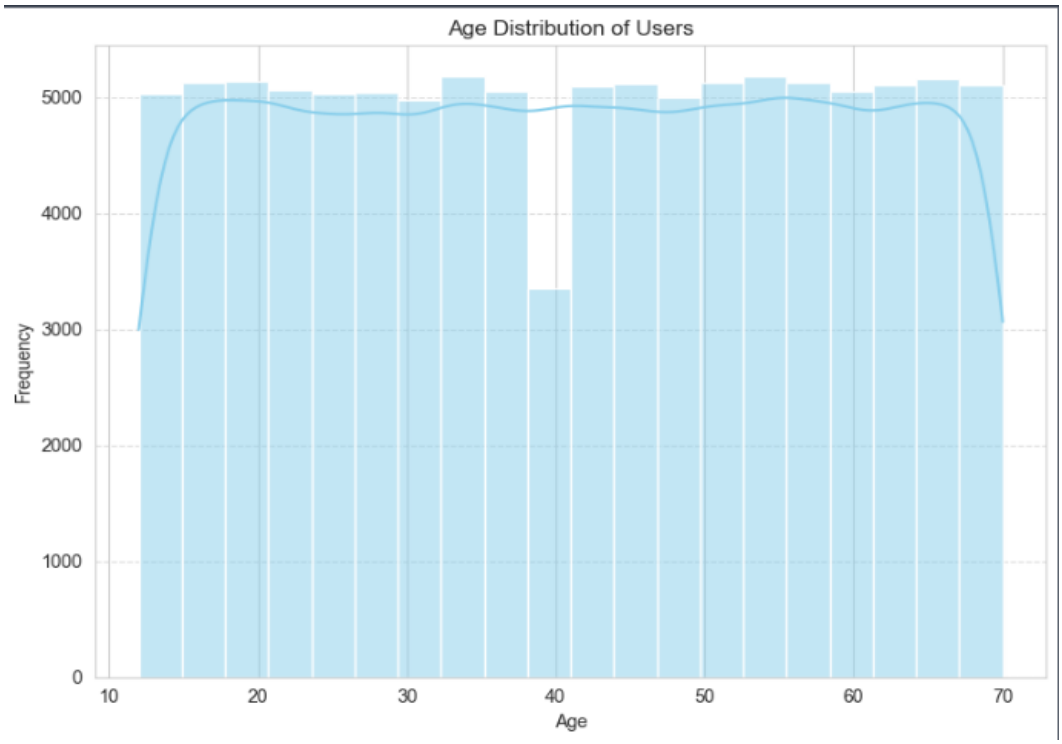
Analysis based on given problem statement

```python
# Display results
print(f"Total cost of all products: ${total_cost:.2f}")
print(f"Average cost per product: ${average_cost:.2f}")
print("\nCategory Analysis:")
print(category_counts)
print(f"\nTotal retail price of all products: ${total_retail_price:.2f}")
print(f"Average retail price per product: ${average_retail_price:.2f}")
print("\nDepartment Analysis:")
print(department_counts)
```

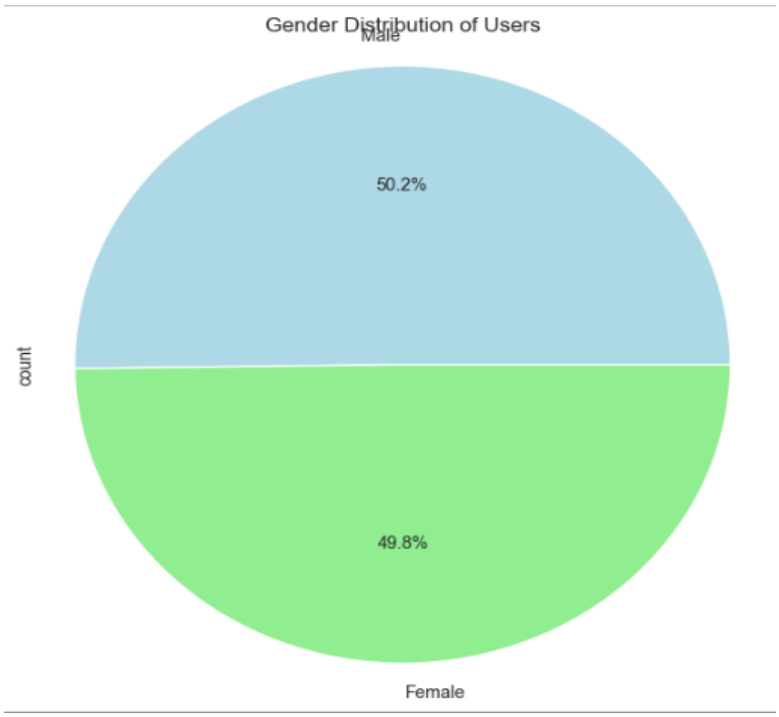
```
Total cost of all products: $829389.27
Average cost per product: $28.48

Category Analysis:
category
Intimates                       2363
Jeans                           1999
Tops & Tees                     1868
Fashion Hoodies & Sweatshirts   1866
Swim                            1798
Sleep & Lounge                  1771
Shorts                          1765
Sweaters                        1737
Accessories                     1559
Active                          1432
Outerwear & Coats               1420
Underwear                       1088
Pants                           1041
Dresses                          955
Socks                            905
Maternity                        898
Plus                             758
Suits & Sport Coats              739
Socks & Hosiery                  666
Pants & Capris                   613
Leggings                         564
Blazers & Jackets                561
Skirts                           367
Suits                            188
Jumpsuits & Rompers              162
Clothing Sets                     37
Name: count, dtype: int64

Total retail price of all products: $1724491.17
Average retail price per product: $59.22

Department Analysis:
department
Women    15989
Men      13131
Name: count, dtype: int64
```

5. **User Demographics Analysis: Use users.csv to analyse user demographics, such as age, gender, and location.**

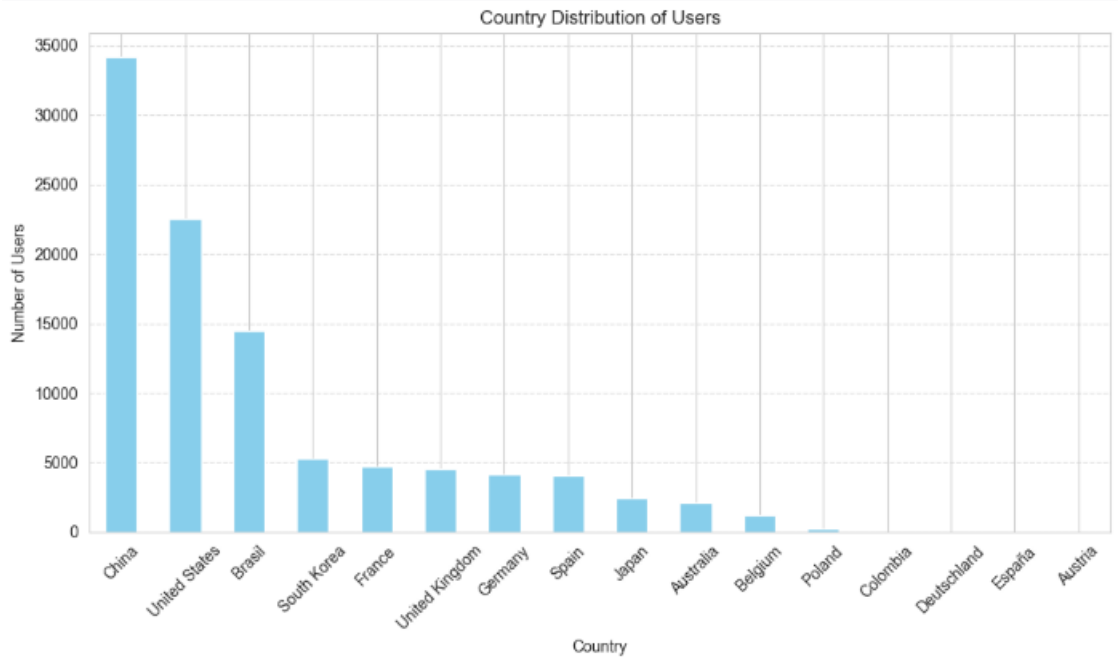
Age Distribution



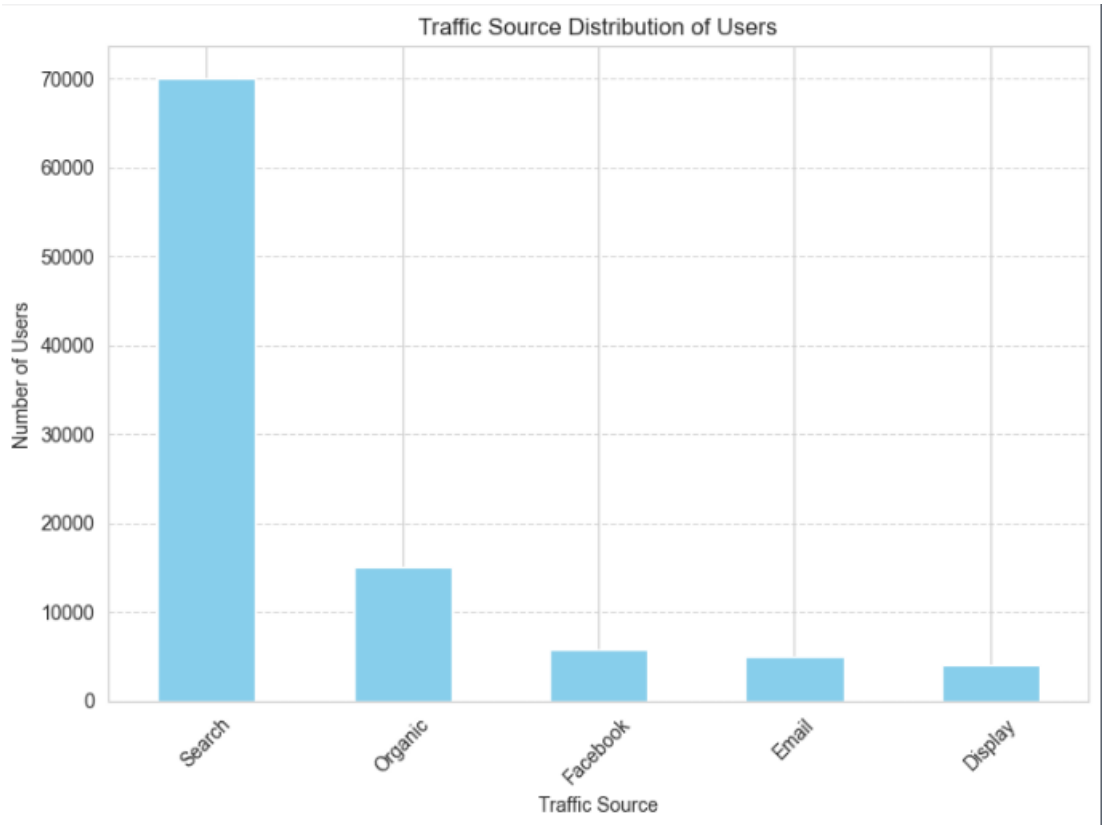
Gender Distribution



Country Distribution



Traffic Source Distribution



Displaying Results

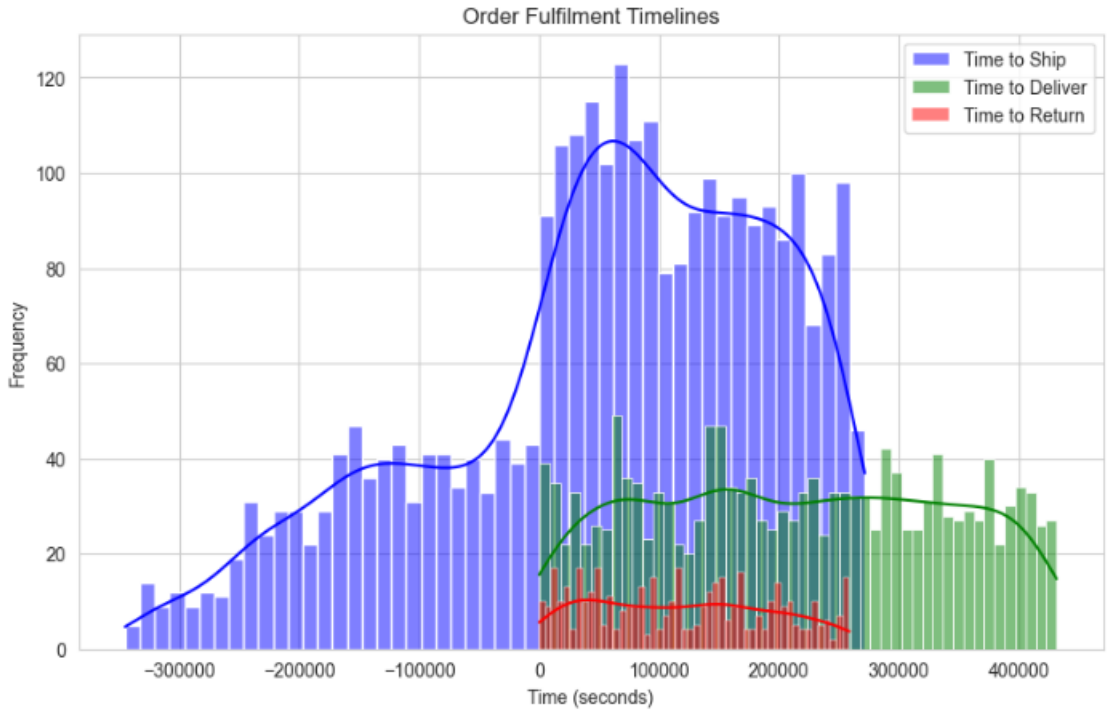
```
Average age of users: 41.05
Median age of users: 41.0
Mode age of users: 33

Gender Analysis:
gender
F     50208
M     49792
Name: count, dtype: int64

Country Analysis:
country
China              34150
United States      22522
Brasil             14507
South Korea         5316
France              4700
United Kingdom      4561
Germany             4155
Spain               4062
Japan               2438
Australia           2146
Belgium             1185
Poland               235
Colombia              17
Deutschland            2
España                 2
Austria                2
Name: count, dtype: int64

Traffic Source Analysis:
traffic_source
Search       70075
Organic      15110
Facebook      5816
Email         4947
Display       4052
Name: count, dtype: int64
```

6. Order Fulfilment Analysis: Analyse order_items.csv and orders.csv to understand order fulfilment timelines and status.

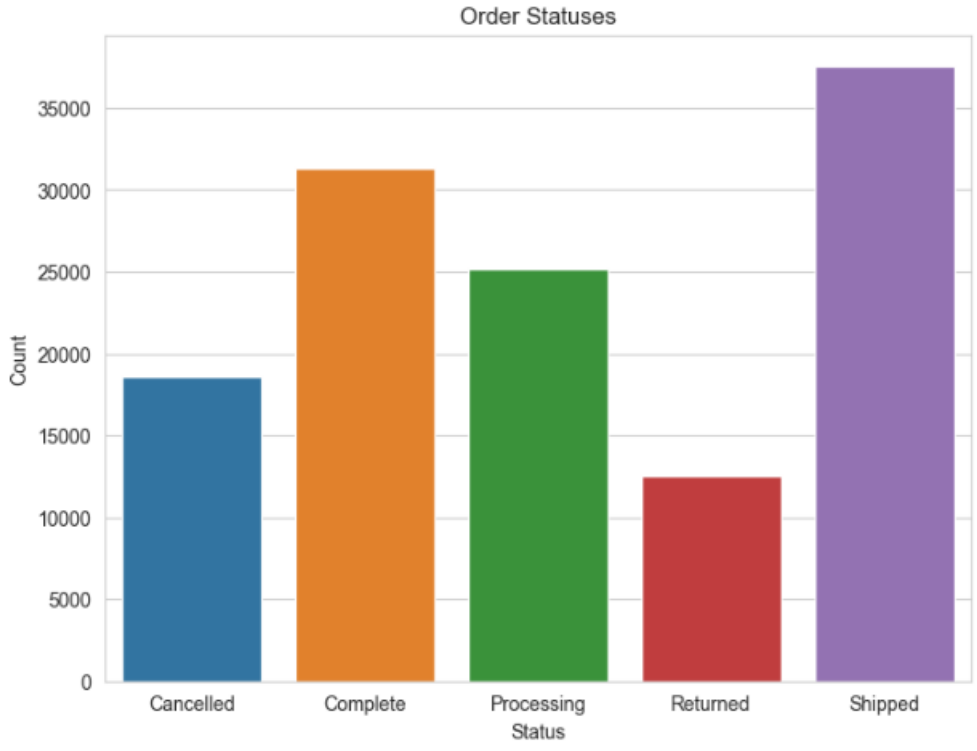
Time to ship

```
Time to Ship:
count      2871.000000
mean      53657.491815
std      142040.469327
min     -345197.000000
25%      -26866.000000
50%       72266.000000
75%      165445.500000
max      271546.000000
Name: time_to_ship_seconds, dtype: float64

Time to Deliver:
count      1555.000000
mean     213992.951768
std      123626.093424
min           0.000000
25%      105630.000000
50%      213000.000000
75%      320400.000000
max      431580.000000
Name: time_to_deliver_seconds, dtype: float64

Time to Return:
count       448.000000
mean     119374.553571
std       74751.930437
min           0.000000
25%       50865.000000
50%      116670.000000
75%      181335.000000
max      258660.000000
Name: time_to_return_seconds, dtype: float64
```

Order fulfilment timelines



Order status



Return status