

```
In [13]: import pandas as pd
import os
from sqlalchemy import create_engine

In [14]: # Replace the values below with your actual credentials
engine = create_engine('mysql+pymysql://root:your_password@127.0.0.1/inventory.db')

In [15]: def ingest_db(df,table_name, engine):
    df.to_sql(table_name,con=engine,if_exists='replace',index=False)

In [ ]: for file in os.listdir('data'):
    if '.csv' in file:
        df=pd.read_csv('data/'+file)
        print(df.shape)
        ingest_db(df,file[:-4],engine)

In [18]: import pandas as pd

In [20]: purchase=pd.read_csv("C:\\\\Users\\\\Jagveer singh\\\\Desktop\\\\inventory\\\\purchases.csv")

In [21]: purchase.head()
```

Out[21]:

| | InventoryId | Store | Brand | Description | Size | VendorNumber | VendorName |
|----------|---------------------|-------|-------|------------------------------------|-------|--------------|----------------------------|
| 0 | 69_MOUNTMEND_8412 | 69 | 8412 | Tequila Ocho Plata Fresno | 750mL | 105 | ALTAMAR BRANDS LLC |
| 1 | 30_CULCHETH_5255 | 30 | 5255 | TGI Fridays Ultimte Mudslide | 1.75L | 4466 | AMERICAN VINTAGE BEVERAGE |
| 2 | 34_PITMERDEN_5215 | 34 | 5215 | TGI Fridays Long Island Iced | 1.75L | 4466 | AMERICAN VINTAGE BEVERAGE |
| 3 | 1_HARDERSFIELD_5255 | 1 | 5255 | TGI Fridays Ultimte Mudslide | 1.75L | 4466 | AMERICAN VINTAGE BEVERAGE |
| 4 | 76_DONCASTER_2034 | 76 | 2034 | Glendalough Double Barrel | 750mL | 388 | ATLANTIC IMPORTING COMPANY |

```
In [23]: purchase[purchase['VendorNumber']==4466]
```

Out[23]:

| | | InventoryId | Store | Brand | Description | Size | VendorNumber | VendorType |
|---------|---------------------|-------------|-------|------------------------------------|-------------|------|---------------------|------------|
| 1 | 30_CULCHETH_5255 | 30 | 5255 | TGI Fridays Ultimte Mudslide | 1.75L | 4466 | AMER VIN BEVE | |
| 2 | 34_PITMERDEN_5215 | 34 | 5215 | TGI Fridays Long Island Iced | 1.75L | 4466 | AMER VIN BEVE | |
| 3 | 1_HARDERSFIELD_5255 | 1 | 5255 | TGI Fridays Ultimte Mudslide | 1.75L | 4466 | AMER VIN BEVE | |
| 8091 | 38_GOULCREST_5215 | 38 | 5215 | TGI Fridays Long Island Iced | 1.75L | 4466 | AMER VIN BEVE | |
| 8092 | 59_CLAETHORPES_5215 | 59 | 5215 | TGI Fridays Long Island Iced | 1.75L | 4466 | AMER VIN BEVE | |
| ... | ... | ... | ... | ... | ... | ... | ... | |
| 2314214 | 81 PEMBROKE_5215 | 81 | 5215 | TGI Fridays Long Island Iced | 1.75L | 4466 | AMER VIN BEVE | |
| 2314215 | 62_KILMARNOCK_5255 | 62 | 5255 | TGI Fridays Ultimte Mudslide | 1.75L | 4466 | AMER VIN BEVE | |
| 2314216 | 34_PITMERDEN_5215 | 34 | 5215 | TGI Fridays Long Island Iced | 1.75L | 4466 | AMER VIN BEVE | |
| 2314217 | 6_GOULCREST_5215 | 6 | 5215 | TGI Fridays Long Island Iced | 1.75L | 4466 | AMER VIN BEVE | |
| 2314218 | 35_HALIVAARA_5255 | 35 | 5255 | TGI Fridays Ultimte Mudslide | 1.75L | 4466 | AMER VIN BEVE | |

2192 rows × 16 columns

In [24]: `purchase_price=pd.read_csv("C:\\Users\\Jagveer singh\\Desktop\\inventory\\purchase_`In [25]: `purchase_price.head(5)`

Out[25]:

| | Brand | Description | Price | Size | Volume | Classification | PurchasePrice | VendorNumber |
|---|-------|-----------------------------|-------|-------|--------|----------------|---------------|--------------|
| 0 | 58 | Gekkeikan Black & Gold Sake | 12.99 | 750mL | 750 | 1 | 9.28 | 8320 |
| 1 | 62 | Herradura Silver Tequila | 36.99 | 750mL | 750 | 1 | 28.67 | 1128 |
| 2 | 63 | Herradura Reposado Tequila | 38.99 | 750mL | 750 | 1 | 30.46 | 1128 |
| 3 | 72 | No. 3 London Dry Gin | 34.99 | 750mL | 750 | 1 | 26.11 | 9165 |
| 4 | 75 | Three Olives Tomato Vodka | 14.99 | 750mL | 750 | 1 | 10.94 | 7245 |



In [27]: `purchase_price[purchase_price["VendorNumber"] == 4466]`

Out[27]:

| | Brand | Description | Price | Size | Volume | Classification | PurchasePrice | VendorNu |
|------|-------|------------------------------|-------|--------|--------|----------------|---------------|----------|
| 1197 | 5215 | TGI Fridays Long Island Iced | 12.99 | 1750mL | 1750 | 1 | 9.41 | |
| 1212 | 5255 | TGI Fridays Ultimte Mudslide | 12.99 | 1750mL | 1750 | 1 | 9.35 | |
| 9962 | 3140 | TGI Fridays Orange Dream | 14.99 | 1750mL | 1750 | 1 | 11.19 | |



In [28]: `vendor_invoice=pd.read_csv("C:\\\\Users\\\\Jagveer singh\\\\Desktop\\\\inventory\\\\vendor_in`

In [29]: `vendor_invoice.head()`

Out[29]:

| | VendorNumber | VendorName | InvoiceDate | PONumber | PODate | PayDate | Quantity |
|---|--------------|----------------------------|-------------|----------|------------|------------|----------|
| 0 | 105 | ALTAMAR BRANDS LLC | 2024-01-04 | 8124 | 2023-12-21 | 2024-02-16 | 6 |
| 1 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-01-07 | 8137 | 2023-12-22 | 2024-02-21 | 15 |
| 2 | 388 | ATLANTIC IMPORTING COMPANY | 2024-01-09 | 8169 | 2023-12-24 | 2024-02-16 | 5 |
| 3 | 480 | BACARDI USA INC | 2024-01-12 | 8106 | 2023-12-20 | 2024-02-05 | 10100 13 |
| 4 | 516 | BANFI PRODUCTS CORP | 2024-01-07 | 8170 | 2023-12-24 | 2024-02-12 | 1935 1 |



In [31]: vendor_invoice[vendor_invoice['VendorNumber'] == 4466]

Out[31]:

| | VendorNumber | VendorName | InvoiceDate | PONumber | PODate | PayDate | Quantity |
|------|---------------------|---------------------------|--------------------|-----------------|---------------|----------------|-----------------|
| 1 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-01-07 | 8137 | 2023-12-22 | 2024-02-21 | 15 |
| 90 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-01-19 | 8207 | 2023-12-27 | 2024-02-26 | 335 |
| 189 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-01-18 | 8307 | 2024-01-03 | 2024-02-18 | 41 |
| 284 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-01-27 | 8469 | 2024-01-14 | 2024-03-11 | 72 |
| 380 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-02-04 | 8532 | 2024-01-19 | 2024-03-15 | 79 |
| 486 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-02-09 | 8604 | 2024-01-24 | 2024-03-15 | 347 |
| 586 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-02-17 | 8793 | 2024-02-05 | 2024-04-02 | 72 |
| 691 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-03-01 | 8892 | 2024-02-12 | 2024-03-28 | 117 |
| 798 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-03-07 | 8995 | 2024-02-19 | 2024-04-02 | 129 |
| 899 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-03-12 | 9033 | 2024-02-22 | 2024-04-16 | 147 |
| 995 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-03-16 | 9180 | 2024-03-03 | 2024-04-19 | 211 |
| 1096 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-03-23 | 9244 | 2024-03-08 | 2024-04-21 | 161 |
| 1199 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-03-31 | 9371 | 2024-03-17 | 2024-05-13 | 176 |
| 1298 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-04-09 | 9491 | 2024-03-24 | 2024-05-08 | 215 |

| | VendorNumber | VendorName | InvoiceDate | PONumber | PODate | PayDate | Quantity |
|------|---------------------|---------------------------|--------------------|-----------------|---------------|----------------|-----------------|
| 1401 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-04-17 | 9583 | 2024-03-31 | 2024-05-12 | 110 |
| 1500 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-04-20 | 9639 | 2024-04-04 | 2024-06-04 | 515 |
| 1602 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-04-29 | 9800 | 2024-04-15 | 2024-06-07 | 275 |
| 1702 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-05-09 | 9886 | 2024-04-21 | 2024-06-12 | 312 |
| 1801 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-05-14 | 9999 | 2024-04-29 | 2024-06-26 | 310 |
| 1905 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-05-16 | 10095 | 2024-05-06 | 2024-06-27 | 215 |
| 1999 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-05-28 | 10169 | 2024-05-11 | 2024-07-04 | 327 |
| 2107 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-06-04 | 10257 | 2024-05-17 | 2024-07-08 | 376 |
| 2209 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-06-12 | 10346 | 2024-05-23 | 2024-07-20 | 640 |
| 2311 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-06-17 | 10445 | 2024-05-30 | 2024-07-19 | 288 |
| 2418 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-06-22 | 10600 | 2024-06-09 | 2024-08-01 | 308 |
| 2516 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-06-29 | 10695 | 2024-06-16 | 2024-08-12 | 143 |
| 2617 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-07-09 | 10777 | 2024-06-22 | 2024-08-15 | 18 |
| 2717 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-07-08 | 10836 | 2024-06-25 | 2024-08-14 | 8 |

| | VendorNumber | VendorName | InvoiceDate | PONumber | PODate | PayDate | Quantity |
|------|---------------------|---------------------------|--------------------|-----------------|---------------|----------------|-----------------|
| 2805 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-07-11 | 10969 | 2024-06-29 | 2024-08-18 | 94 |
| 2906 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-07-19 | 11085 | 2024-07-06 | 2024-09-04 | 601 |
| 3006 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-07-26 | 11187 | 2024-07-14 | 2024-08-26 | 1535 |
| 3109 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-08-03 | 11244 | 2024-07-18 | 2024-09-04 | 266 |
| 3210 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-08-11 | 11362 | 2024-07-26 | 2024-09-24 | 206 |
| 3308 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-08-18 | 11489 | 2024-08-04 | 2024-09-16 | 768 |
| 3408 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-08-24 | 11540 | 2024-08-08 | 2024-10-02 | 1207 |
| 3518 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-09-02 | 11716 | 2024-08-19 | 2024-09-29 | 433 |
| 3613 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-09-12 | 11771 | 2024-08-23 | 2024-10-11 | 370 |
| 3720 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-09-20 | 11901 | 2024-09-01 | 2024-10-30 | 358 |
| 3821 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-09-25 | 11993 | 2024-09-07 | 2024-10-23 | 233 |
| 3922 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-10-01 | 12125 | 2024-09-16 | 2024-11-07 | 284 |
| 4027 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-10-08 | 12235 | 2024-09-23 | 2024-11-20 | 258 |
| 4134 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-10-09 | 12253 | 2024-09-23 | 2024-11-14 | 1 |

| | VendorNumber | VendorName | InvoiceDate | PONumber | PODate | PayDate | Quantity |
|------|---------------------|---------------------------|--------------------|-----------------|---------------|----------------|-----------------|
| 4194 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-10-12 | 12321 | 2024-09-26 | 2024-11-19 | 172 |
| 4295 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-10-20 | 12466 | 2024-10-05 | 2024-11-26 | 280 |
| 4398 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-10-27 | 12515 | 2024-10-09 | 2024-11-30 | 178 |
| 4501 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-11-07 | 12702 | 2024-10-21 | 2024-12-11 | 183 |
| 4607 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-11-12 | 12752 | 2024-10-25 | 2024-12-11 | 216 |
| 4711 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-11-20 | 12828 | 2024-10-30 | 2024-12-18 | 262 |
| 4811 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-11-27 | 12929 | 2024-11-06 | 2025-01-04 | 270 |
| 4913 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-11-28 | 13092 | 2024-11-16 | 2024-12-30 | 209 |
| 5017 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-12-06 | 13134 | 2024-11-20 | 2025-01-18 | 305 |
| 5120 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-12-16 | 13254 | 2024-11-28 | 2025-01-13 | 262 |
| 5219 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-12-26 | 13432 | 2024-12-09 | 2025-01-27 | 231 |
| 5329 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-12-30 | 13483 | 2024-12-13 | 2025-02-11 | 221 |
| 5436 | 4466 | AMERICAN VINTAGE BEVERAGE | 2025-01-09 | 13627 | 2024-12-22 | 2025-02-05 | 413 |

In [32]: sales=pd.read_csv("C:\\Users\\Jagveer singh\\Desktop\\inventory\\sales.csv")

In [33]: `sales.head()`

Out[33]:

| | InventoryId | Store | Brand | Description | Size | SalesQuantity | SalesDollars | Sa |
|---|---------------------|-------|-------|----------------------------------|---------------|---------------|--------------|----|
| 0 | 1_HARDERSFIELD_1004 | 1 | 1004 | Jim Beam w/2 Rocks Glasses | 750mL | 1 | 16.49 | |
| 1 | 1_HARDERSFIELD_1004 | 1 | 1004 | Jim Beam w/2 Rocks Glasses | 750mL | 2 | 32.98 | |
| 2 | 1_HARDERSFIELD_1004 | 1 | 1004 | Jim Beam w/2 Rocks Glasses | 750mL | 1 | 16.49 | |
| 3 | 1_HARDERSFIELD_1004 | 1 | 1004 | Jim Beam w/2 Rocks Glasses | 750mL | 1 | 14.49 | |
| 4 | 1_HARDERSFIELD_1005 | 1 | 1005 | Maker's Mark Combo Pack | 375mL 2 Pk | 2 | 69.98 | |

In [35]: `sales[sales["VendorNo"] == 4466]`

Out[35]:

| | | InventoryId | Store | Brand | Description | Size | SalesQuantity | SalesDoll |
|-----------------|---------------------|-------------|-------|-------|------------------------------------|-------|---------------|-----------|
| 13370 | 1_HARDERSFIELD_5215 | | 1 | 5215 | TGI Fridays Long Island Iced | 1.75L | 1 | 12 |
| 13371 | 1_HARDERSFIELD_5215 | | 1 | 5215 | TGI Fridays Long Island Iced | 1.75L | 1 | 12 |
| 13372 | 1_HARDERSFIELD_5215 | | 1 | 5215 | TGI Fridays Long Island Iced | 1.75L | 1 | 12 |
| 13373 | 1_HARDERSFIELD_5215 | | 1 | 5215 | TGI Fridays Long Island Iced | 1.75L | 1 | 12 |
| 13374 | 1_HARDERSFIELD_5215 | | 1 | 5215 | TGI Fridays Long Island Iced | 1.75L | 1 | 12 |
| ... | | | | | | | | |
| 12819662 | 9_BLACKPOOL_5215 | | 9 | 5215 | TGI Fridays Long Island Iced | 1.75L | 1 | 12 |
| 12819801 | 9_BLACKPOOL_5255 | | 9 | 5255 | TGI Fridays Ultimte Mudslide | 1.75L | 1 | 12 |
| 12819802 | 9_BLACKPOOL_5255 | | 9 | 5255 | TGI Fridays Ultimte Mudslide | 1.75L | 1 | 12 |
| 12819803 | 9_BLACKPOOL_5255 | | 9 | 5255 | TGI Fridays Ultimte Mudslide | 1.75L | 1 | 12 |
| 12819804 | 9_BLACKPOOL_5255 | | 9 | 5255 | TGI Fridays Ultimte Mudslide | 1.75L | 1 | 12 |

9453 rows × 14 columns

In [39]: `purchase.groupby(['Brand', 'PurchasePrice'])[['Quantity', 'Dollars']].sum().head(3)`

Out[39]:

Quantity Dollars

| Brand | PurchasePrice | Quantity | Dollars |
|--------------|----------------------|-----------------|----------------|
| 58 | 9.28 | 3550 | 32944.0 |
| 60 | 7.40 | 1633 | 12084.2 |
| 61 | 10.60 | 312 | 3307.2 |

In [40]: `purchase_price.head()`

Out[40]:

| | Brand | Description | Price | Size | Volume | Classification | PurchasePrice | VendorNumber |
|---|--------------|-----------------------------|--------------|-------------|---------------|-----------------------|----------------------|---------------------|
| 0 | 58 | Gekkeikan Black & Gold Sake | 12.99 | 750mL | 750 | 1 | 9.28 | 8320 |
| 1 | 62 | Herradura Silver Tequila | 36.99 | 750mL | 750 | 1 | 28.67 | 1128 |
| 2 | 63 | Herradura Reposado Tequila | 38.99 | 750mL | 750 | 1 | 30.46 | 1128 |
| 3 | 72 | No. 3 London Dry Gin | 34.99 | 750mL | 750 | 1 | 26.11 | 9165 |
| 4 | 75 | Three Olives Tomato Vodka | 14.99 | 750mL | 750 | 1 | 10.94 | 7245 |

In [41]: `vendor_invoice.head()`

Out[41]:

| | VendorNumber | VendorName | InvoiceDate | PONumber | PODate | PayDate | Quantity |
|---|--------------|----------------------------|-------------|----------|------------|------------|----------|
| 0 | 105 | ALTAMAR BRANDS LLC | 2024-01-04 | 8124 | 2023-12-21 | 2024-02-16 | 6 |
| 1 | 4466 | AMERICAN VINTAGE BEVERAGE | 2024-01-07 | 8137 | 2023-12-22 | 2024-02-21 | 15 |
| 2 | 388 | ATLANTIC IMPORTING COMPANY | 2024-01-09 | 8169 | 2023-12-24 | 2024-02-16 | 5 |
| 3 | 480 | BACARDI USA INC | 2024-01-12 | 8106 | 2023-12-20 | 2024-02-05 | 10100 13 |
| 4 | 516 | BANFI PRODUCTS CORP | 2024-01-07 | 8170 | 2023-12-24 | 2024-02-12 | 1935 1 |



In [43]: `vendor_invoice['PONumber'].nunique()`

Out[43]: 5543

In [44]: `vendor_invoice.shape`

Out[44]: (5543, 10)

In [47]: `sales.groupby('Brand')[['SalesDollars', 'SalesPrice', 'SalesQuantity']].sum().head()`

Out[47]:

| Brand | SalesDollars | SalesPrice | SalesQuantity |
|-------|--------------|------------|---------------|
| 58 | 43341.54 | 28145.64 | 3446 |
| 60 | 18716.25 | 10720.79 | 1775 |
| 61 | 4364.88 | 363.74 | 312 |
| 62 | 119863.75 | 90154.51 | 3125 |
| 63 | 112249.22 | 88553.10 | 2778 |

| Brand | SalesDollars | SalesPrice | SalesQuantity |
|-------|--------------|------------|---------------|
| 58 | 43341.54 | 28145.64 | 3446 |
| 60 | 18716.25 | 10720.79 | 1775 |
| 61 | 4364.88 | 363.74 | 312 |
| 62 | 119863.75 | 90154.51 | 3125 |
| 63 | 112249.22 | 88553.10 | 2778 |

In [48]: `vendor_invoice.columns`

Out[48]: `Index(['VendorNumber', 'VendorName', 'InvoiceDate', 'PONumber', 'PODate', 'PayDate', 'Quantity', 'Dollars', 'Freight', 'Approval'], dtype='object')`

In [124...]: `freight_summary=vendor_invoice.groupby('VendorNumber')[['Freight']].sum().rename(columns={0: 'TotalFreight'})`

In [123...]: `freight_summary.head()`

Out[123...]

| | VendorNumber | FreightCost |
|---|--------------|-------------|
| 0 | 2 | 27.08 |
| 1 | 54 | 0.48 |
| 2 | 60 | 367.52 |
| 3 | 105 | 62.39 |
| 4 | 200 | 6.19 |

In [58]: `purchase.head()`

Out[58]:

| | InventoryId | Store | Brand | Description | Size | VendorNumber | VendorName |
|---|---------------------|-------|-------|------------------------------------|-------|--------------|----------------------------------|
| 0 | 69_MOUNTMEND_8412 | 69 | 8412 | Tequila Ocho Plata Fresno | 750mL | 105 | ALTAMAR BRANDS LLC |
| 1 | 30_CULCHETH_5255 | 30 | 5255 | TGI Fridays Ultimte Mudslide | 1.75L | 4466 | AMERICAN VINTAGE BEVERAGE |
| 2 | 34_PITMERDEN_5215 | 34 | 5215 | TGI Fridays Long Island Iced | 1.75L | 4466 | AMERICAN VINTAGE BEVERAGE |
| 3 | 1_HARDERSFIELD_5255 | 1 | 5255 | TGI Fridays Ultimte Mudslide | 1.75L | 4466 | AMERICAN VINTAGE BEVERAGE |
| 4 | 76_DONCASTER_2034 | 76 | 2034 | Glendalough Double Barrel | 750mL | 388 | ATLANTIC IMPORTING COMPANY |

In [60]: `purchase_price.head()`

Out[60]:

| | Brand | Description | Price | Size | Volume | Classification | PurchasePrice | VendorNumber |
|---|-------|-----------------------------|-------|-------|--------|----------------|---------------|--------------|
| 0 | 58 | Gekkeikan Black & Gold Sake | 12.99 | 750mL | 750 | 1 | 9.28 | 8320 |
| 1 | 62 | Herradura Silver Tequila | 36.99 | 750mL | 750 | 1 | 28.67 | 1128 |
| 2 | 63 | Herradura Reposado Tequila | 38.99 | 750mL | 750 | 1 | 30.46 | 1128 |
| 3 | 72 | No. 3 London Dry Gin | 34.99 | 750mL | 750 | 1 | 26.11 | 9165 |
| 4 | 75 | Three Olives Tomato Vodka | 14.99 | 750mL | 750 | 1 | 10.94 | 7245 |



In [64]:

```
filtered_purchase = purchase[purchase['PurchasePrice'] > 0]
```

In [79]:

```
merged_df = pd.merge(
    purchase,
    purchase_price,
    on='Brand',
    how='inner'
)
```

In [80]:

```
purchase[['VendorNumber', 'VendorName', 'Brand']]
```

Out[80]:

| | VendorNumber | VendorName | Brand |
|----------------|--------------|----------------------------|-------|
| 0 | 105 | ALTAMAR BRANDS LLC | 8412 |
| 1 | 4466 | AMERICAN VINTAGE BEVERAGE | 5255 |
| 2 | 4466 | AMERICAN VINTAGE BEVERAGE | 5215 |
| 3 | 4466 | AMERICAN VINTAGE BEVERAGE | 5255 |
| 4 | 388 | ATLANTIC IMPORTING COMPANY | 2034 |
| ... | ... | ... | ... |
| 2372469 | 90058 | ZORVINO VINEYARDS | 22298 |
| 2372470 | 90058 | ZORVINO VINEYARDS | 19556 |
| 2372471 | 90058 | ZORVINO VINEYARDS | 22297 |
| 2372472 | 90058 | ZORVINO VINEYARDS | 19557 |
| 2372473 | 90058 | ZORVINO VINEYARDS | 22298 |

2372474 rows × 3 columns

In [81]: `merged_df.head()`

Out[81]:

| | InventoryId | Store | Brand | Description_x | Size_x | VendorNumber_x | VendorNa |
|---|--------------------|-------|-------|---------------------------|--------|----------------|------------|
| 0 | 69_MOUNTMEND_8412 | 69 | 8412 | Tequila Ocho Plata Fresno | 750mL | 105 | ALT/ BRAND |
| 1 | 66_EANVERNESS_8412 | 66 | 8412 | Tequila Ocho Plata Fresno | 750mL | 105 | ALT/ BRAND |
| 2 | 15_WANBORNE_8412 | 15 | 8412 | Tequila Ocho Plata Fresno | 750mL | 105 | ALT/ BRAND |
| 3 | 66_EANVERNESS_8412 | 66 | 8412 | Tequila Ocho Plata Fresno | 750mL | 105 | ALT/ BRAND |
| 4 | 10_HORNSEY_8412 | 10 | 8412 | Tequila Ocho Plata Fresno | 750mL | 105 | ALT/ BRAND |

5 rows × 24 columns

In [105...]: `df = merged_df[['VendorNumber_x', 'VendorName_x', 'Brand', 'Price', 'Volume', 'Purc`In [106...]: `df.head()`

Out[106...]

| | VendorNumber_x | VendorName_x | Brand | Price | Volume | PurchasePrice_y |
|---|----------------|--------------------|-------|-------|--------|-----------------|
| 0 | 105 | ALTAMAR BRANDS LLC | 8412 | 49.99 | 750 | 35.71 |
| 1 | 105 | ALTAMAR BRANDS LLC | 8412 | 49.99 | 750 | 35.71 |
| 2 | 105 | ALTAMAR BRANDS LLC | 8412 | 49.99 | 750 | 35.71 |
| 3 | 105 | ALTAMAR BRANDS LLC | 8412 | 49.99 | 750 | 35.71 |
| 4 | 105 | ALTAMAR BRANDS LLC | 8412 | 49.99 | 750 | 35.71 |

In [107...]

```
dt = merged_df.groupby(['VendorNumber_x', 'VendorName_x', 'Brand'])[['Quantity', 'D'].rename(columns={'Quantity': 'TotalPurchaseQty', 'Dollars': 'TotalPurchaseDo'}
```

In [98]: dt

Out[98]:

| | VendorNumber_x | VendorName_x | Brand | TotalPurchaseQty | TotalPurchaseDo |
|-------|----------------|-------------------------------|-------|------------------|-----------------|
| 0 | 2 | IRA GOLDMAN AND WILLIAMS, LLP | 90085 | 8 | 190.88 |
| 1 | 2 | IRA GOLDMAN AND WILLIAMS, LLP | 90609 | 320 | 5440.00 |
| 2 | 54 | AAPER ALCOHOL & CHEMICAL CO | 990 | 1 | 105.07 |
| 3 | 60 | ADAMBA IMPORTS INTL INC | 771 | 39 | 446.16 |
| 4 | 60 | ADAMBA IMPORTS INTL INC | 3401 | 6 | 66.60 |
| ... | ... | ... | ... | ... | ... |
| 10688 | 173357 | TAMWORTH DISTILLING | 2804 | 210 | 6749.40 |
| 10689 | 173357 | TAMWORTH DISTILLING | 3666 | 520 | 9770.80 |
| 10690 | 173357 | TAMWORTH DISTILLING | 3848 | 28 | 652.40 |
| 10691 | 173357 | TAMWORTH DISTILLING | 3909 | 1232 | 23863.84 |
| 10692 | 201359 | FLAVOR ESSENCE INC | 90609 | 1 | 17.00 |

10693 rows × 5 columns

```
In [108... final_df = pd.merge(df, dt, on=['VendorNumber_x', 'VendorName_x', 'Brand'], how='le  
In [110... final_df = final_df[['VendorNumber_x', 'VendorName_x', 'Brand', 'Price', 'Volume',  
                           'TotalPurchaseQty', 'TotalPurchaseDol']].rename(columns={'Vend  
In [114... final_df.tail()
```

| | | VendorNumber | VendorName | Brand | Price | Volume | ActualPrice | TotalPurchaseC |
|--|--|--------------|------------|----------------------------|-------|--------|-------------|----------------|
| | | 2372469 | 9165 | ULTRA BEVERAGE COMPANY LLP | 27051 | 14.99 | 750 | 10.2 |
| | | 2372470 | 9165 | ULTRA BEVERAGE COMPANY LLP | 27051 | 14.99 | 750 | 10.2 |
| | | 2372471 | 9165 | ULTRA BEVERAGE COMPANY LLP | 27051 | 14.99 | 750 | 10.2 |
| | | 2372472 | 9165 | ULTRA BEVERAGE COMPANY LLP | 27051 | 14.99 | 750 | 10.2 |
| | | 2372473 | 9165 | ULTRA BEVERAGE COMPANY LLP | 27051 | 14.99 | 750 | 10.2 |



```
In [115... sales.head()
```

Out[115...]

| | InventoryId | Store | Brand | Description | Size | SalesQuantity | SalesDollars | Sa |
|---|---------------------|-------|-------|----------------------------------|---------------|---------------|--------------|----|
| 0 | 1_HARDERSFIELD_1004 | 1 | 1004 | Jim Beam w/2 Rocks Glasses | 750mL | 1 | 16.49 | |
| 1 | 1_HARDERSFIELD_1004 | 1 | 1004 | Jim Beam w/2 Rocks Glasses | 750mL | 2 | 32.98 | |
| 2 | 1_HARDERSFIELD_1004 | 1 | 1004 | Jim Beam w/2 Rocks Glasses | 750mL | 1 | 16.49 | |
| 3 | 1_HARDERSFIELD_1004 | 1 | 1004 | Jim Beam w/2 Rocks Glasses | 750mL | 1 | 14.49 | |
| 4 | 1_HARDERSFIELD_1005 | 1 | 1005 | Maker's Mark Combo Pack | 375mL 2 Pk | 2 | 69.98 | |



In [141...]

```
a=sales.groupby(['VendorNo','Brand'])[['SalesDollars','SalesPrice','SalesQuantity'],
    .rename(columns={'VendorNo':'VendorNumber','SalesDollars':'TotalSalesDol',
```

In [168...]

```
a['Description']=sales['Description']
```

In [169...]

```
data=pd.merge(final_df, freight_summary, on='VendorNumber',how='left')
```

In [127...]

Out[127...]

| | VendorNumber | VendorName | Brand | Price | Volume | ActualPrice | TotalPurchaseC |
|--|--------------|------------|-------|-------|--------|-------------|----------------|
|--|--------------|------------|-------|-------|--------|-------------|----------------|

| | | | | | | | |
|----------------|------|-------------------------------------|-------|-------|-----|------|----|
| 2372469 | 9165 | ULTRA BEVERAGE COMPANY LLP | 27051 | 14.99 | 750 | 10.2 | 10 |
| 2372470 | 9165 | ULTRA BEVERAGE COMPANY LLP | 27051 | 14.99 | 750 | 10.2 | 10 |
| 2372471 | 9165 | ULTRA BEVERAGE COMPANY LLP | 27051 | 14.99 | 750 | 10.2 | 10 |
| 2372472 | 9165 | ULTRA BEVERAGE COMPANY LLP | 27051 | 14.99 | 750 | 10.2 | 10 |
| 2372473 | 9165 | ULTRA BEVERAGE COMPANY LLP | 27051 | 14.99 | 750 | 10.2 | 10 |

In [170...]: `final_table=pd.merge(data,a,on='Brand',how='left')`In [171...]: `final_table.head()`

Out[171...]

| | VendorNumber | VendorName | Brand | Price | Volume | ActualPrice | TotalPurchaseQty | TotalCost |
|---|--------------|--------------------|-------|-------|--------|-------------|------------------|-----------|
| 0 | 105 | ALTAMAR BRANDS LLC | 8412 | 49.99 | 750 | 35.71 | | 320 |
| 1 | 105 | ALTAMAR BRANDS LLC | 8412 | 49.99 | 750 | 35.71 | | 320 |
| 2 | 105 | ALTAMAR BRANDS LLC | 8412 | 49.99 | 750 | 35.71 | | 320 |
| 3 | 105 | ALTAMAR BRANDS LLC | 8412 | 49.99 | 750 | 35.71 | | 320 |
| 4 | 105 | ALTAMAR BRANDS LLC | 8412 | 49.99 | 750 | 35.71 | | 320 |

In [172...]: `final_table = final_table.drop('VendorNo', axis=1)`In [173...]: `final_table.head()`

Out[173...]

| | VendorNumber | VendorName | Brand | Price | Volume | ActualPrice | TotalPurchaseQty | TotalSalesDol |
|---|--------------|--------------------|-------|-------|--------|-------------|------------------|---------------|
| 0 | 105 | ALTAMAR BRANDS LLC | 8412 | 49.99 | 750 | 35.71 | | 320 |
| 1 | 105 | ALTAMAR BRANDS LLC | 8412 | 49.99 | 750 | 35.71 | | 320 |
| 2 | 105 | ALTAMAR BRANDS LLC | 8412 | 49.99 | 750 | 35.71 | | 320 |
| 3 | 105 | ALTAMAR BRANDS LLC | 8412 | 49.99 | 750 | 35.71 | | 320 |
| 4 | 105 | ALTAMAR BRANDS LLC | 8412 | 49.99 | 750 | 35.71 | | 320 |



In [185...]

final_table.shape

Out[185...]

(2381916, 14)

In [186...]

final_table.unique()

Out[186...]

| | |
|------------------|-------|
| VendorNumber | 126 |
| VendorName | 129 |
| Brand | 10664 |
| Price | 343 |
| Volume | 31 |
| ActualPrice | 2151 |
| TotalPurchaseQty | 3876 |
| TotalPurchaseDol | 9799 |
| FreightCost | 126 |
| TotalSalesDol | 9457 |
| TotalSalesPrice | 8842 |
| TotalSalesQty | 3809 |
| TotalExciseTax | 7377 |
| Description | 1630 |
| dtype: | int64 |

In [187...]

final_table.columns

```
Out[187... Index(['VendorNumber', 'VendorName', 'Brand', 'Price', 'Volume', 'ActualPrice',  
   'TotalPurchaseQty', 'TotalPurchaseDol', 'FreightCost', 'TotalSalesDol',  
   'TotalSalesPrice', 'TotalSalesQty', 'TotalExciseTax', 'Description'],  
  dtype='object')
```

```
In [188... Vendor_sales_summary= final_table.groupby(['VendorNumber', 'VendorName', 'Brand'],
```

```
In [189... Vendor_sales_summary
```

Out[189...]

| | VendorNumber | VendorName | Brand | Price | Volume | ActualPrice | TotalPurchaseQt |
|-------|--------------|---|-------|--------|--------|-------------|-----------------|
| 0 | 2 | IRA GOLDMAN AND WILLIAMS, LLP | 90085 | 36.99 | 750 | 23.86 | |
| 1 | 2 | IRA GOLDMAN AND WILLIAMS, LLP | 90609 | 24.99 | 162.5 | 17.00 | 32 |
| 2 | 54 | AAPER ALCOHOL & CHEMICAL CO | 990 | 134.49 | 3750 | 105.07 | |
| 3 | 60 | ADAMBA IMPORTS INTL INC | 771 | 14.99 | 750 | 11.44 | 3 |
| 4 | 60 | ADAMBA IMPORTS INTL INC | 3401 | 14.99 | 1750 | 11.10 | |
| ... | ... | ... | ... | ... | ... | ... | |
| 10688 | 173357 | TAMWORTH DISTILLING | 2804 | 44.99 | 750 | 32.14 | 21 |
| 10689 | 173357 | TAMWORTH DISTILLING | 3666 | 24.99 | 375 | 18.79 | 52 |
| 10690 | 173357 | TAMWORTH DISTILLING | 3848 | 30.99 | 750 | 23.30 | 2 |
| 10691 | 173357 | TAMWORTH DISTILLING | 3909 | 24.99 | 750 | 19.37 | 123 |
| 10692 | 201359 | FLAVOR ESSENCE INC | 90609 | 24.99 | 162.5 | 17.00 | |

10693 rows × 14 columns



In [190...]

Vendor_sales_summary.dtypes

```
Out[190...]: VendorNumber      int64
VendorName        object
Brand            int64
Price           float64
Volume          object
ActualPrice     float64
TotalPurchaseQty   int64
TotalPurchaseDol  float64
FreightCost     float64
TotalSalesDol    float64
TotalSalesPrice   float64
TotalSalesQty    float64
TotalExciseTax   float64
Description      object
dtype: object
```

```
In [191...]: Vendor_sales_summary.isnull().sum()
```

```
Out[191...]: VendorNumber      0
VendorName        0
Brand            0
Price           0
Volume          0
ActualPrice     0
TotalPurchaseQty 0
TotalPurchaseDol 0
FreightCost     0
TotalSalesDol    178
TotalSalesPrice   178
TotalSalesQty    178
TotalExciseTax   178
Description      178
dtype: int64
```

```
In [192...]: Vendor_sales_summary['Volume']=Vendor_sales_summary['Volume'].astype('float64')
```

```
In [193...]: Vendor_sales_summary.fillna(0,inplace=True)
```

```
In [194...]: Vendor_sales_summary['VendorName']=Vendor_sales_summary['VendorName'].str.strip()
```

```
In [195...]: Vendor_sales_summary.isnull().sum()
```

```
Out[195...]: VendorNumber      0  
VendorName        0  
Brand            0  
Price            0  
Volume           0  
ActualPrice      0  
TotalPurchaseQty 0  
TotalPurchaseDol 0  
FreightCost      0  
TotalSalesDol    0  
TotalSalesPrice   0  
TotalSalesQty    0  
TotalExciseTax   0  
Description       0  
dtype: int64
```

```
In [196...]: Vendor_sales_summary.dtypes
```

```
Out[196...]: VendorNumber      int64  
VendorName        object  
Brand            int64  
Price            float64  
Volume           float64  
ActualPrice      float64  
TotalPurchaseQty int64  
TotalPurchaseDol float64  
FreightCost      float64  
TotalSalesDol    float64  
TotalSalesPrice   float64  
TotalSalesQty    float64  
TotalExciseTax   float64  
Description       object  
dtype: object
```

```
In [197...]: Vendor_sales_summary['Gross Profit']=Vendor_sales_summary['TotalSalesDol']-Vendor_s
```

```
In [200...]: Vendor_sales_summary['ProfitMargin']= (Vendor_sales_summary['Gross Profit']/Vendor_
```

```
In [201...]: Vendor_sales_summary
```

Out[201...]

| | VendorNumber | VendorName | Brand | Price | Volume | ActualPrice | TotalPurchaseQt |
|-------|--------------|---|-------|--------|--------|-------------|-----------------|
| 0 | 2 | IRA GOLDMAN AND WILLIAMS, LLP | 90085 | 36.99 | 750.0 | 23.86 | |
| 1 | 2 | IRA GOLDMAN AND WILLIAMS, LLP | 90609 | 24.99 | 162.5 | 17.00 | 32 |
| 2 | 54 | AAPER ALCOHOL & CHEMICAL CO | 990 | 134.49 | 3750.0 | 105.07 | |
| 3 | 60 | ADAMBA IMPORTS INTL INC | 771 | 14.99 | 750.0 | 11.44 | 3 |
| 4 | 60 | ADAMBA IMPORTS INTL INC | 3401 | 14.99 | 1750.0 | 11.10 | |
| ... | ... | ... | ... | ... | ... | ... | |
| 10688 | 173357 | TAMWORTH DISTILLING | 2804 | 44.99 | 750.0 | 32.14 | 21 |
| 10689 | 173357 | TAMWORTH DISTILLING | 3666 | 24.99 | 375.0 | 18.79 | 52 |
| 10690 | 173357 | TAMWORTH DISTILLING | 3848 | 30.99 | 750.0 | 23.30 | 2 |
| 10691 | 173357 | TAMWORTH DISTILLING | 3909 | 24.99 | 750.0 | 19.37 | 123 |
| 10692 | 201359 | FLAVOR ESSENCE INC | 90609 | 24.99 | 162.5 | 17.00 | |

10693 rows × 16 columns



In [202...]

Vendor_sales_summary['StockTurnover']=Vendor_sales_summary['TotalSalesQty']/Vendor_

In [204...]

Vendor_sales_summary['SalesPurchaseRatio']=Vendor_sales_summary['TotalSalesDol']/Ve

In [205... Vendor_sales_summary

Out[205...]

| | VendorNumber | VendorName | Brand | Price | Volume | ActualPrice | TotalPurchaseQty | TotalSalesValue |
|---|--------------|---|-------|--------|--------|-------------|------------------|-----------------|
| 0 | 2 | IRA GOLDMAN AND WILLIAMS, LLP | 90085 | 36.99 | 750.0 | 23.86 | | 8 |
| 1 | 2 | IRA GOLDMAN AND WILLIAMS, LLP | 90609 | 24.99 | 162.5 | 17.00 | | 320 |
| 2 | 54 | AAPER ALCOHOL & CHEMICAL CO | 990 | 134.49 | 3750.0 | 105.07 | | 1 |
| 3 | 60 | ADAMBA IMPORTS INTL INC | 771 | 14.99 | 750.0 | 11.44 | | 39 |
| 4 | 60 | ADAMBA IMPORTS INTL INC | 3401 | 14.99 | 1750.0 | 11.10 | | 6 |

| | | | | | | | | |
|---|----|---|-------|--------|--------|--------|--|-----|
| 0 | 2 | IRA GOLDMAN AND WILLIAMS, LLP | 90085 | 36.99 | 750.0 | 23.86 | | 8 |
| 1 | 2 | IRA GOLDMAN AND WILLIAMS, LLP | 90609 | 24.99 | 162.5 | 17.00 | | 320 |
| 2 | 54 | AAPER ALCOHOL & CHEMICAL CO | 990 | 134.49 | 3750.0 | 105.07 | | 1 |
| 3 | 60 | ADAMBA IMPORTS INTL INC | 771 | 14.99 | 750.0 | 11.44 | | 39 |
| 4 | 60 | ADAMBA IMPORTS INTL INC | 3401 | 14.99 | 1750.0 | 11.10 | | 6 |



Importing Libraries

In [215...]

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
import sqlite3
from scipy.stats import ttest_ind
import scipy.stats as stats
warnings.filterwarnings('ignore')
```

In [207...]

Vendor_sales_summary.head()

Out[207...]

| | | VendorNumber | VendorName | Brand | Price | Volume | ActualPrice | TotalPurchaseQty | TotalCost |
|---|----|--------------|---|-------|--------|--------|-------------|------------------|-----------|
| 0 | 2 | | IRA GOLDMAN AND WILLIAMS, LLP | 90085 | 36.99 | 750.0 | 23.86 | | 8 |
| 1 | 2 | | IRA GOLDMAN AND WILLIAMS, LLP | 90609 | 24.99 | 162.5 | 17.00 | | 320 |
| 2 | 54 | | AAPER ALCOHOL & CHEMICAL CO | 990 | 134.49 | 3750.0 | 105.07 | | 1 |
| 3 | 60 | | ADAMBA IMPORTS INTL INC | 771 | 14.99 | 750.0 | 11.44 | | 39 |
| 4 | 60 | | ADAMBA IMPORTS INTL INC | 3401 | 14.99 | 1750.0 | 11.10 | | 6 |

Exploratory Data Analysis

Previously, we examined the various tables in the database to identify key variables, understand their relationships, and determine which ones should be included in the final analysis.

In this phase of EDA, we will analyze the resultant table to gain insights into the distribution of each column. This will help us understand data patterns, identify anomalies, and ensure data quality before proceeding with further analysis.

Summary Statistics

In [209...]

Vendor_sales_summary.describe().T

Out[209...]

| | | count | mean | std | min | 25% |
|--|---------------------------|--------------|--------------|---------------|------------|----------------------|
| | VendorNumber | 10693.0 | 1.064989e+04 | 18752.805318 | 2.00 | 3951.000000 7153.0 |
| | Brand | 10693.0 | 1.803774e+04 | 12662.525384 | 58.00 | 5789.000000 18761.0 |
| | Price | 10693.0 | 3.564034e+01 | 148.239484 | 0.00 | 10.990000 15.9 |
| | Volume | 10693.0 | 8.473514e+02 | 664.278813 | 50.00 | 750.000000 750.0 |
| | ActualPrice | 10693.0 | 2.438302e+01 | 109.264519 | 0.00 | 6.840000 10.4 |
| | TotalPurchaseQty | 10693.0 | 3.140782e+03 | 11094.573249 | 1.00 | 36.000000 262.0 |
| | TotalPurchaseDol | 10693.0 | 3.010388e+04 | 123062.388765 | 0.00 | 453.180000 3655.2 |
| | FreightCost | 10693.0 | 6.142892e+04 | 60937.665572 | 0.09 | 14069.870000 50293.6 |
| | TotalSalesDol | 10693.0 | 4.222904e+04 | 167636.094802 | 0.00 | 729.270000 5299.4 |
| | TotalSalesPrice | 10693.0 | 1.879377e+04 | 44953.192308 | 0.00 | 289.710000 2859.4 |
| | TotalSalesQty | 10693.0 | 3.075344e+03 | 10950.377892 | 0.00 | 33.000000 262.0 |
| | TotalExciseTax | 10693.0 | 1.773027e+03 | 10972.786262 | 0.00 | 4.800000 46.6 |
| | Gross Profit | 10693.0 | 1.212516e+04 | 46225.273383 | -52002.78 | 52.590000 1390.4 |
| | ProfitMargin | 10693.0 | -inf | NaN | -inf | 13.233776 30.3 |
| | StockTurnover | 10693.0 | 1.711494e+00 | 6.040632 | 0.00 | 0.805889 0.9 |
| | SalesPurchaseRatio | 10693.0 | inf | NaN | 0.00 | 1.152522 1.4 |



In [210...]

df=Vendor_sales_summary

In [211...]

df

Out[211...]

| | | VendorNumber | VendorName | Brand | Price | Volume | ActualPrice | TotalPurchaseQt |
|-------|--|--------------|---|-------|--------|--------|-------------|-----------------|
| 0 | | 2 | IRA GOLDMAN AND WILLIAMS, LLP | 90085 | 36.99 | 750.0 | 23.86 | |
| 1 | | 2 | IRA GOLDMAN AND WILLIAMS, LLP | 90609 | 24.99 | 162.5 | 17.00 | 32 |
| 2 | | 54 | AAPER ALCOHOL & CHEMICAL CO | 990 | 134.49 | 3750.0 | 105.07 | |
| 3 | | 60 | ADAMBA IMPORTS INTL INC | 771 | 14.99 | 750.0 | 11.44 | 3 |
| 4 | | 60 | ADAMBA IMPORTS INTL INC | 3401 | 14.99 | 1750.0 | 11.10 | |
| ... | | ... | ... | ... | ... | ... | ... | |
| 10688 | | 173357 | TAMWORTH DISTILLING | 2804 | 44.99 | 750.0 | 32.14 | 21 |
| 10689 | | 173357 | TAMWORTH DISTILLING | 3666 | 24.99 | 375.0 | 18.79 | 52 |
| 10690 | | 173357 | TAMWORTH DISTILLING | 3848 | 30.99 | 750.0 | 23.30 | 2 |
| 10691 | | 173357 | TAMWORTH DISTILLING | 3909 | 24.99 | 750.0 | 19.37 | 123 |
| 10692 | | 201359 | FLAVOR ESSENCE INC | 90609 | 24.99 | 162.5 | 17.00 | |

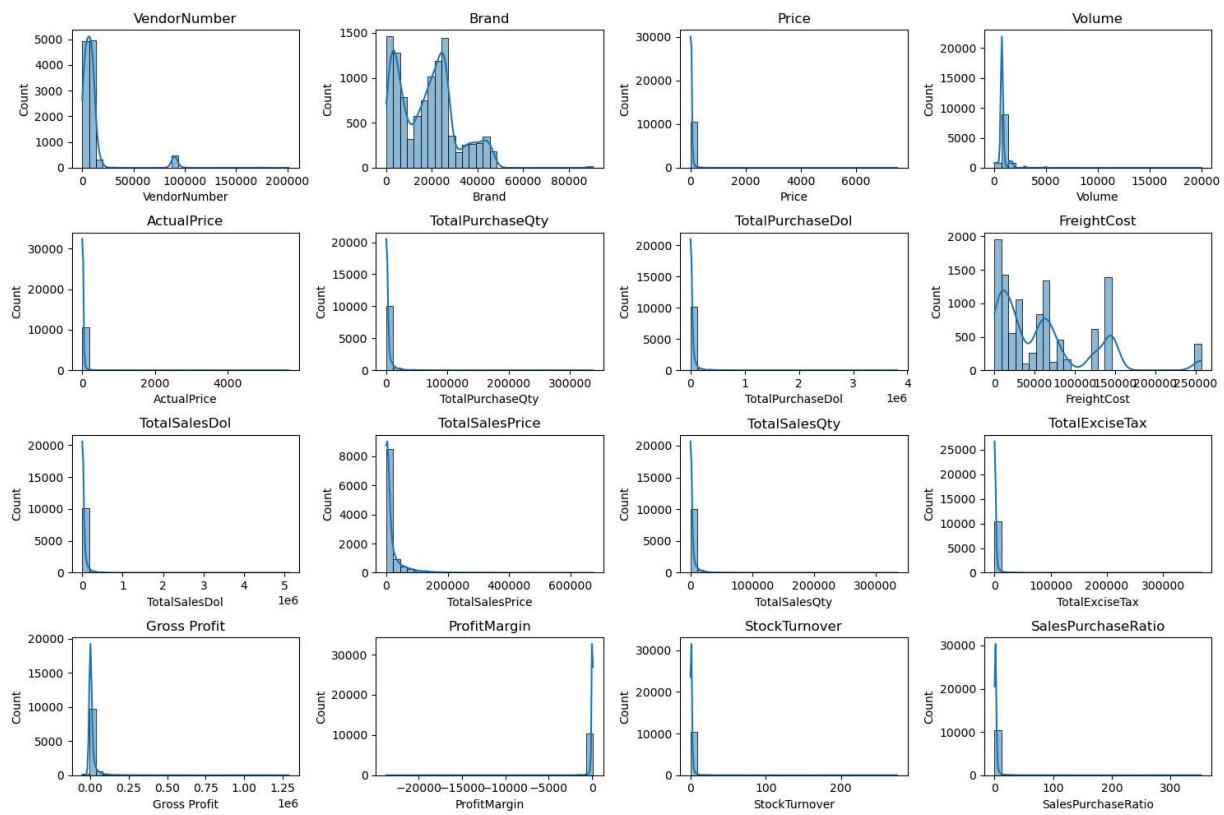
10693 rows × 18 columns



In [216...]

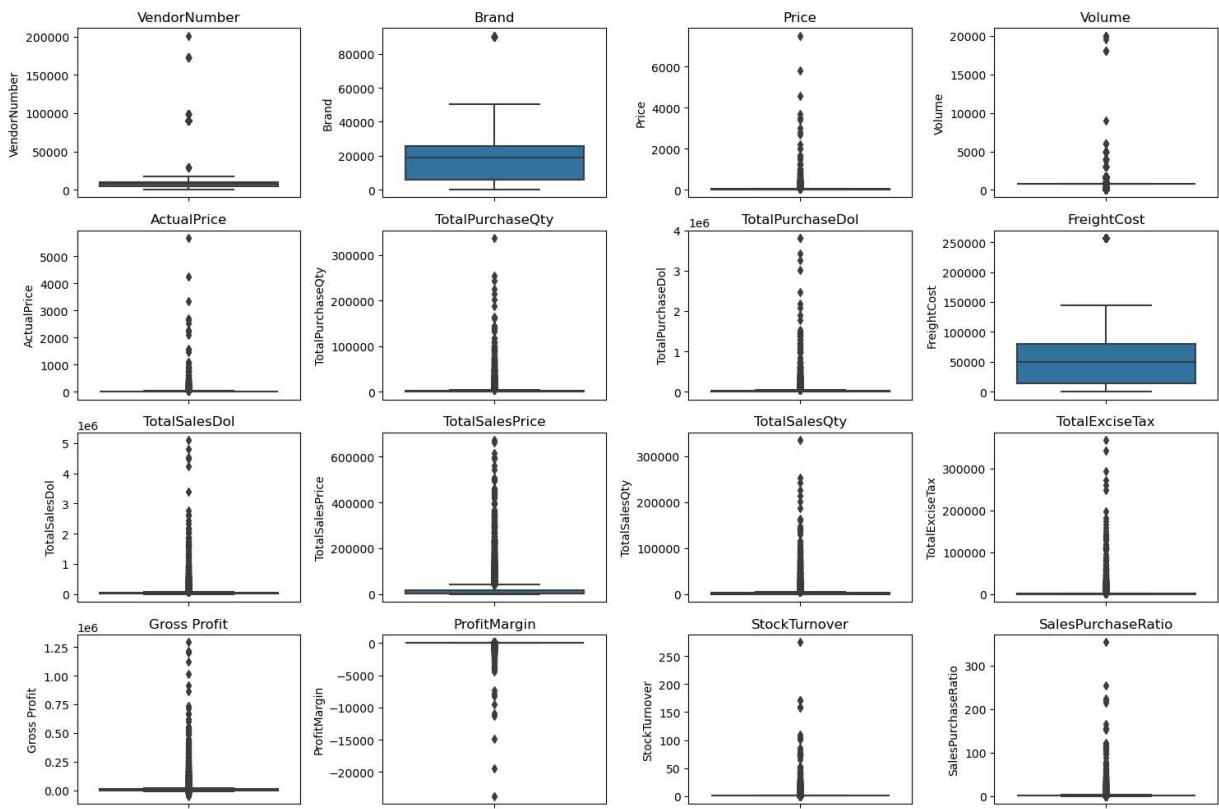
```
#Distribution Plots for Numerical Columns
numerical_cols = df.select_dtypes(include=np.number).columns
plt.figure(figsize=(15,10))
for i, col in enumerate(numerical_cols):
```

```
plt.subplot(4, 4, i+1)# Adjust grid layout as needed
sns.histplot(df[col], kde=True, bins=30)
plt.title(col)
plt.tight_layout()
plt.show()
```



In [217...]

```
# Outlier Detection with Boxplots
plt.figure(figsize=(15,10))
for i,col in enumerate(numerical_cols):
    plt.subplot(4,4,i+1)
    sns.boxplot(y=df[col])
    plt.title(col)
plt.tight_layout()
plt.show()
```



Summary Statistics Insights:

Negative & Zero Values:

Gross Profit: Minimum value is -52,002.78, indicating losses. Some products or transactions may be selling at a loss due to high costs or selling at discounts lower than the purchase price.

Total Sales Quantity & Sales Dollars: Minimum values are 0, meaning some products were purchased but never sold. These could be slow-moving or obsolete stock.

Profit Margin: Has a minimum of -infinite, which suggests cases where revenue is zero or even lower than costs.

Outliers Indicated by High Standard Deviations:

Purchase & Actual Prices: The max values (5,681.81 & 7,499.99) are significantly higher than the mean (24.39 & 35.64), indicating potential premium products.

Freight Cost: Huge variation, from 0.09 to 257,032.07, suggests logistics inefficiencies or bulk shipments.

Stock Turnover: Ranges from 0 to 274.5, implying some products sell extremely fast while others remain in stock indefinitely. Value more than 1 indicates that Sold quantity for that product is higher than purchased quantity due to either sales are being fulfilled from older stock.

In [218...]

```
df=df[  
    (df['Gross Profit'] > 0) &  
    (df['ProfitMargin'] > 0) &  
    (df['TotalSalesQty'] > 0)  
]
```

In [219...]

```
df
```

Out[219...]

| | | VendorNumber | VendorName | Brand | Price | Volume | ActualPrice | TotalPurchaseQty |
|-------|--------|------------------------|---|-------|-------|--------|-------------|------------------|
| 0 | 2 | | IRA GOLDMAN AND WILLIAMS, LLP | 90085 | 36.99 | 750.0 | 23.86 | 8 |
| 3 | 60 | | ADAMABA IMPORTS INTL INC | 771 | 14.99 | 750.0 | 11.44 | 39 |
| 6 | 105 | | ALTAMAR BRANDS LLC | 2529 | 29.99 | 750.0 | 23.25 | 12 |
| 7 | 105 | | ALTAMAR BRANDS LLC | 8412 | 49.99 | 750.0 | 35.71 | 320 |
| 9 | 200 | | AMERICAN SPIRITS EXCHANGE | 20789 | 14.99 | 750.0 | 9.73 | 96 |
| ... | ... | | ... | ... | ... | ... | ... | ... |
| 10683 | 172662 | SWEETWATER FARM | | 2191 | 19.99 | 375.0 | 14.70 | 138 |
| 10684 | 172662 | SWEETWATER FARM | | 2265 | 28.99 | 750.0 | 21.47 | 267 |
| 10686 | 172662 | SWEETWATER FARM | | 3632 | 28.49 | 750.0 | 21.58 | 554 |
| 10691 | 173357 | TAMWORTH DISTILLING | | 3909 | 24.99 | 750.0 | 19.37 | 1232 |
| 10692 | 201359 | FLAVOR ESSENCE INC | | 90609 | 24.99 | 162.5 | 17.00 | 1 |

8562 rows × 18 columns



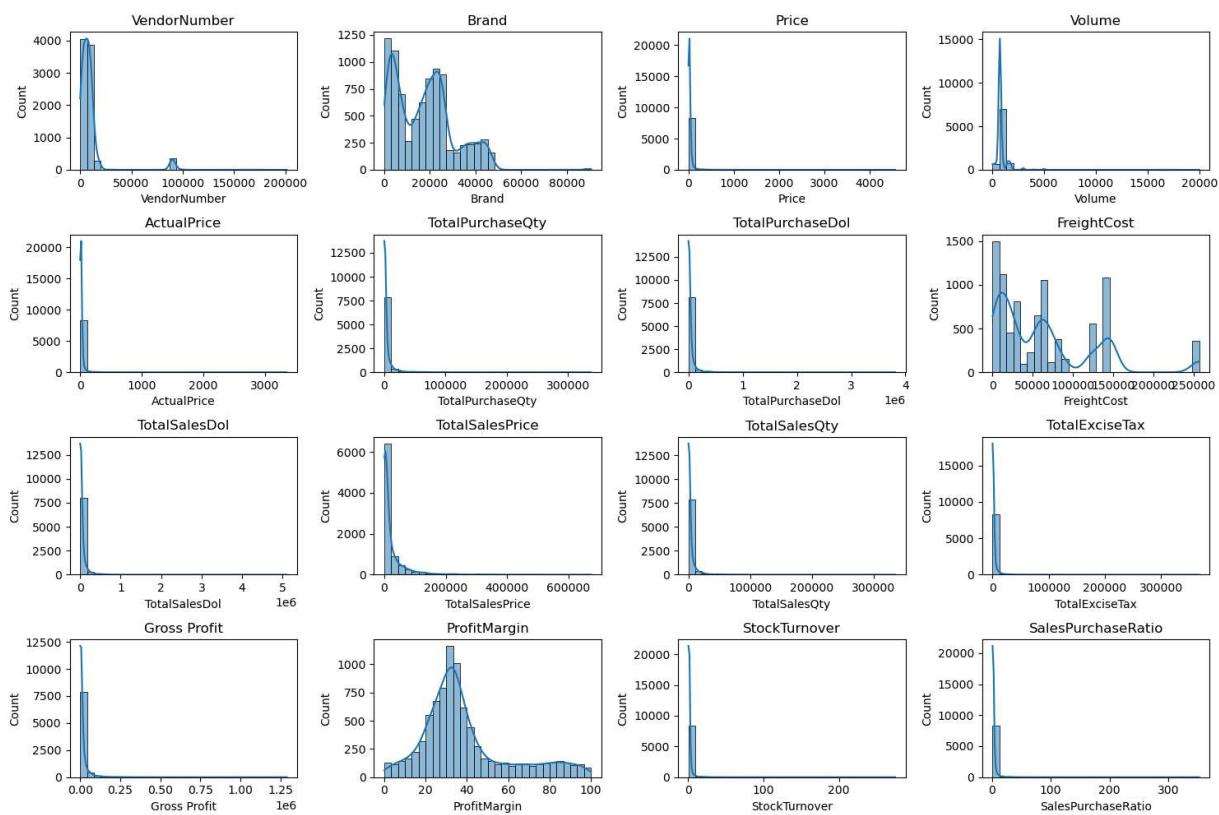
In [220...]

```
#Distribution Plots for Numerical Columns
numerical_cols = df.select_dtypes(include=np.number).columns
plt.figure(figsize=(15,10))
for i, col in enumerate(numerical_cols):
    plt.subplot(4, 4, i+1)# Adjust grid Layout as needed
```

```

sns.histplot(df[col], kde=True, bins=30)
plt.title(col)
plt.tight_layout()
plt.show()

```



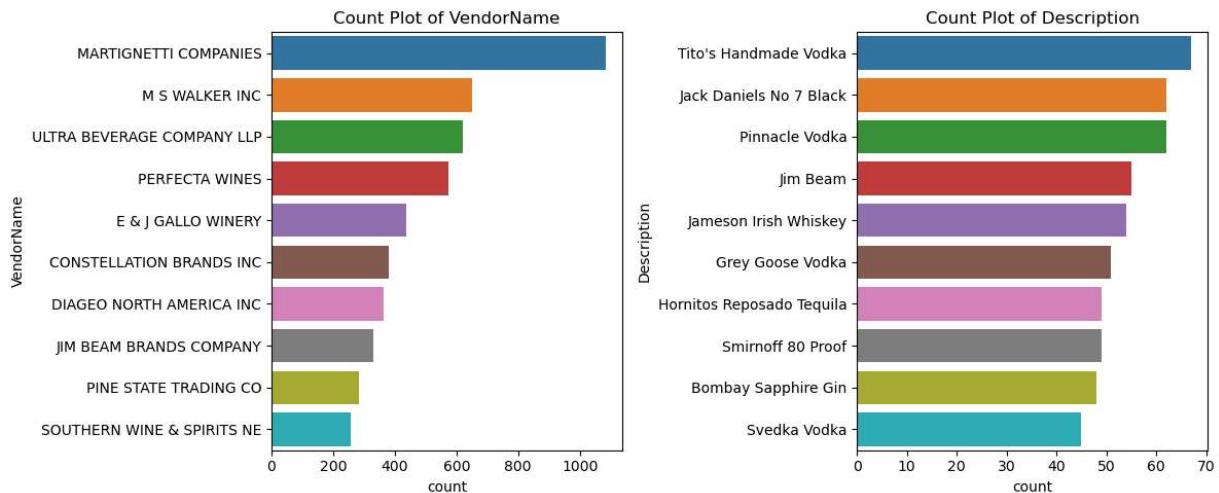
In [221...]

```

#Count Plots for categorical columns
categorical_cols=["VendorName", "Description"]

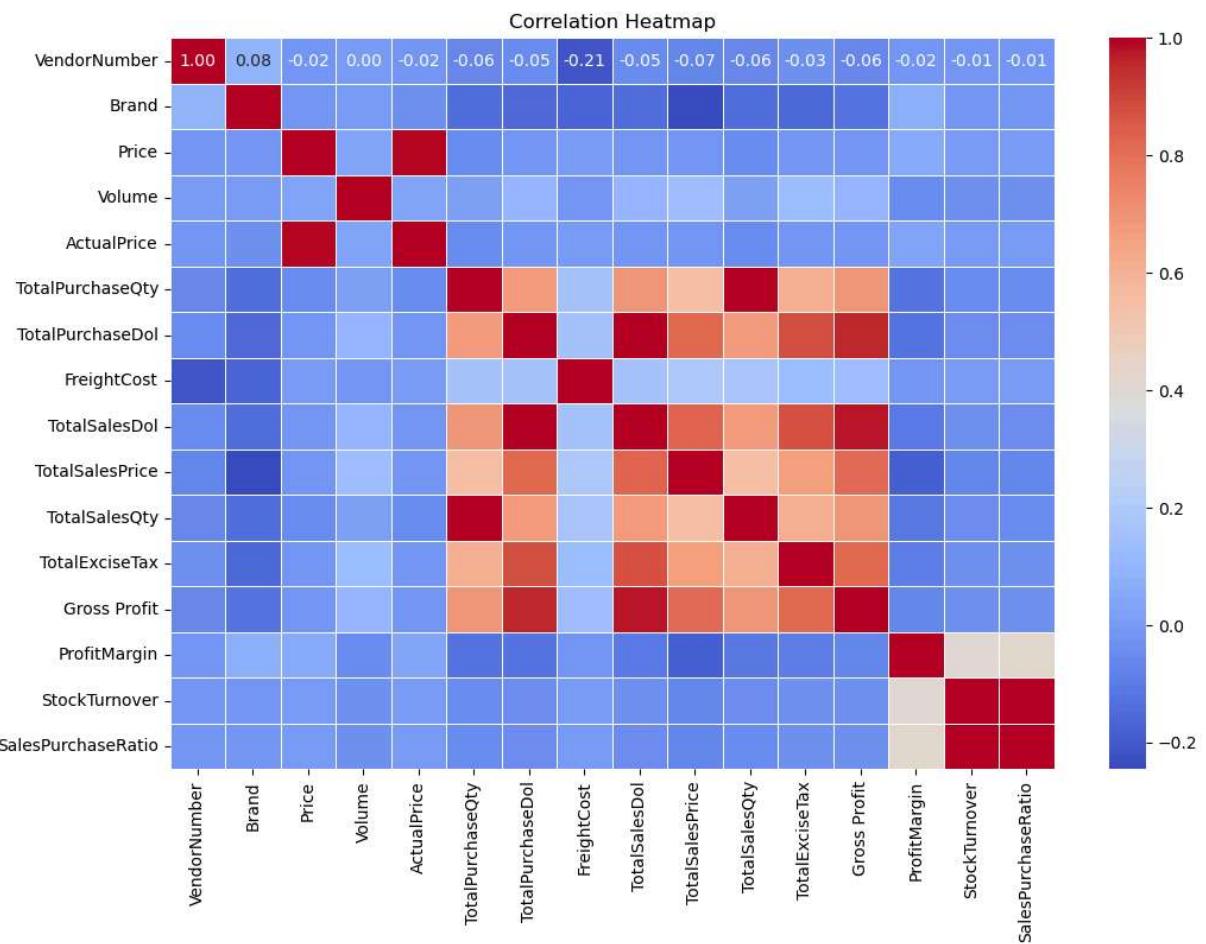
plt.figure(figsize=(12,5))
for i, col in enumerate(categorical_cols):
    plt.subplot(1,2,i+1)
    sns.countplot(y=df[col],order=df[col].value_counts().index[:10]) #Top 10 categorical values
    plt.title(f"Count Plot of {col}")
plt.tight_layout()
plt.show()

```



In [222...]

```
#Correlation Heatmap
plt.figure(figsize=(12,8))
correlation_matrix=df[numerical_cols].corr()
sns.heatmap(correlation_matrix, annot=True, fmt=".2f", cmap="coolwarm", linewidths=0.5)
plt.title("Correlation Heatmap")
plt.show()
```



Correlation Insights

PurchasePrice has weak correlations with TotalSalesDollar(-0.012) and GrossProfit(-0.016), suggesting that price variations do not significantly impact sales revenue or profit

Strong Correlation between total purchase quantity and total sales quantity(0.999), confirming efficient inventory turnover

Negative correlation between profit margin & total sales price (-0.179) suggests that as sales prices increases, margins decrease, possibly due to competitive pricing pressure.

StockTurnover has weak negative correlations with both GrossProfit(-0.038) and ProfitMargin(-0.055), indicating that faster turnover does not necessarily result in higher profitability

Data Analysis

Identify Brands that needs Promotional or pricing Adjustments which exhibit lower sales performance but highr profit margins.

```
In [230...]: brand_performance=df.groupby('Description').agg({'TotalSalesDol':'sum', 'ProfitMargin':'mean'})  
In [232...]: low_sales_threshold=brand_performance['TotalSalesDol'].quantile(0.15)  
high_margin_threshold=brand_performance['ProfitMargin'].quantile(0.85)  
In [233...]: low_sales_threshold  
Out[233...]: 4679.22  
In [235...]: high_margin_threshold  
Out[235...]: 53.76234666384892  
In [238...]: #Filter brands with Low sales but high profit margins  
target_brands=brand_performance[(brand_performance ['TotalSalesDol'] <= low_sales_threshold) & (brand_performance ['ProfitMargin'] >= high_margin_threshold)]  
print("Brands with Low Sales but High Profit Margins:")  
display (target_brands.sort_values( 'TotalSalesDol'))
```

Brands with Low Sales but High Profit Margins:

| | Description | TotalSalesDol | ProfitMargin |
|------|------------------------------|---------------|--------------|
| 120 | Belvedere Cytrus Vodka | 15.95 | 83.448276 |
| 219 | Bowmore Darkest Islay 15 Yr | 67.83 | 80.687012 |
| 294 | Candia Vyds Classic Cab Svgn | 80.97 | 54.649870 |
| 1239 | Seagrams VO Gold | 99.98 | 65.513103 |
| 984 | Nathanson Creek Chard | 103.95 | 80.444284 |
| ... | ... | ... | ... |
| 777 | Johnnie Walker Double Black | 4523.19 | 95.047964 |
| 1190 | Ruffino Aziano Chianti Clsc | 4526.89 | 63.070205 |
| 479 | Death's Door Gin | 4551.97 | 54.929844 |
| 1234 | Seagram's Red Berry Vodka | 4589.49 | 97.095102 |
| 937 | Michter's Small Batch Bourbn | 4605.21 | 61.750725 |

75 rows × 3 columns

```
In [254...]: brand_performance=brand_performance[brand_performance['TotalSalesDol']<10000]  
In [255...]: plt.figure(figsize=(10,6))  
sns.scatterplot(data=brand_performance,x='TotalSalesDol', y='ProfitMargin', color='Blue')  
sns.scatterplot(data=target_brands,x='TotalSalesDol', y='ProfitMargin', color='Red')
```

```

plt.axhline(high_margin_threshold, linestyle='--', color='black', label='High Margin')
plt.axhline(low_sales_threshold, linestyle='--', color='black', label='low sales Thre')
plt.xlabel("Total Sales($)")
plt.ylabel("Profit Margin (%)")
plt.title("Brands for Promotional or pricing Adjustments")
plt.legend()
plt.grid(True)
plt.show()

```



which vendor and brands demonstrate the highest sales performance?

```
In [256...]: top_vendors=df.groupby("VendorName")["TotalSalesDol"].sum().nlargest(10)
top_brands=df.groupby("Description")["TotalSalesDol"].sum().nlargest(10)
```

```
In [257...]: top_vendors
```

```
Out[257...]: VendorName
DIAGEO NORTH AMERICA INC      67990099.42
MARTIGNETTI COMPANIES        39330359.36
PERNOD RICARD USA            32063196.19
JIM BEAM BRANDS COMPANY      31423020.46
BACARDI USA INC              24854817.14
CONSTELLATION BRANDS INC     24218745.65
E & J GALLO WINERY           18399899.46
BROWN-FORMAN CORP            18247230.65
ULTRA BEVERAGE COMPANY LLP   16485663.87
M S WALKER INC                14703014.46
Name: TotalSalesDol, dtype: float64
```

```
In [258...]: top_brands
```

```
Out[258]: Description
Jack Daniels No 7 Black      11233575.51
Yukon Jack                   9754167.56
Hornitos Reposado Tequila   8620252.77
Liberty School Merlot CC    7718932.35
Zhenka 80 Proof              6520855.59
Velvet Crush Red Blend       6179041.86
Barefoot Moscato Cal        5459352.94
Alamos Slcn Malbec Mendoza 5207504.29
Jim Beam                      4961048.96
Smirnoff 80 Proof             4717091.16
Name: TotalSalesDol, dtype: float64
```

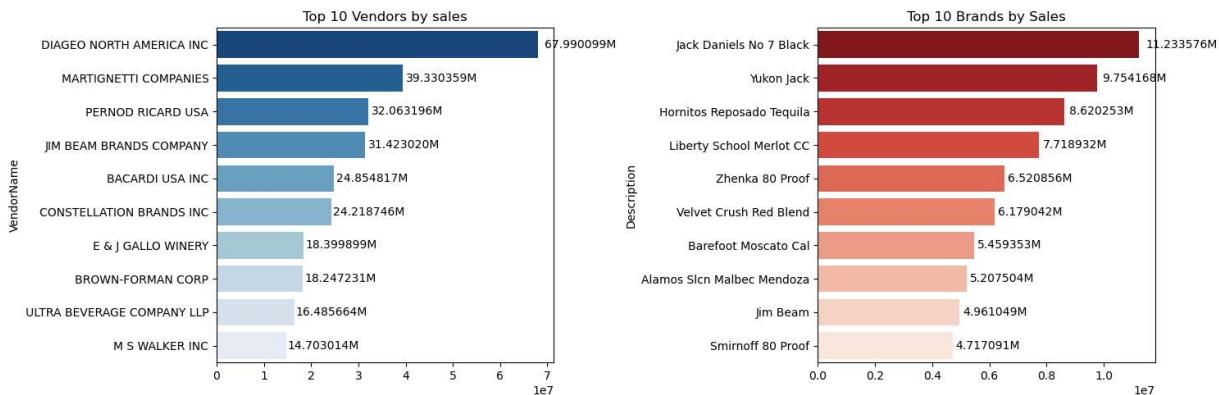
```
In [261]: def format_dollars(value):
    if value>=1_000_000:
        return f"{value/1_000_000:2f}M"
    elif value>=1_000:
        return f"{value/1_000:2f}K"
    else:
        return str(value)
```

```
In [265]: plt.figure(figsize=(15,5))

#Plot for Top Vendors
plt.subplot(1,2,1)
ax1=sns.barplot(y=top_vendors.index,x=top_vendors.values,palette="Blues_r")
plt.title("Top 10 Vendors by sales")

for bar in ax1.patches:
    ax1.text(bar.get_width()+(bar.get_width()*0.02),
            bar.get_y()+bar.get_height()/2,
            format_dollars(bar.get_width()),
            ha='left', va='center', fontsize=10, color='black')

#Plot for top brands
plt.subplot(1,2,2)
ax2=sns.barplot(y=top_brands.index.astype(str),x=top_brands.values,palette="Reds_r")
plt.title("Top 10 Brands by Sales")
for bar in ax2.patches:
    ax2.text(bar.get_width() + (bar.get_width()*0.02),
            bar.get_y()+ bar.get_height()/2,
            format_dollars(bar.get_width()),
            ha='left', va='center', fontsize=10, color='black')
plt.tight_layout()
plt.show()
```



Which Vendor contribute the most to total purchases dollars?

```
In [281... Vendor_Performance=df.groupby('VendorName').agg({
    'TotalPurchaseDol':'sum',
    'Gross Profit':'sum',
    'TotalSalesDol':'sum'}).reset_index()

In [296... Vendor_Performance['PurchaseContribution%']=Vendor_Performance['TotalPurchaseDol']/Vendor_Performance['TotalSalesDol']

In [297... Vendor_Performance=round(Vendor_Performance.sort_values('PurchaseContribution%', ascending=False).head(10))

In [298... #Display Top 10 Vendors
top_vendors=Vendor_Performance.head(10)
top_vendors['TotalSalesDol']=top_vendors['TotalSalesDol'].apply(format_dollars)
top_vendors['TotalPurchaseDol']=top_vendors['TotalPurchaseDol'].apply(format_dollars)
top_vendors['Gross Profit']=top_vendors['Gross Profit'].apply(format_dollars)
top_vendors
```

Out[298...]

| | VendorName | TotalPurchaseDol | Gross Profit | TotalSalesDol | PurchaseContribution% |
|-----|----------------------------|------------------|--------------|---------------|-----------------------|
| 25 | DIAGEO NORTH AMERICA INC | 50.097226M | 17.892873M | 67.990099M | 16.31 |
| 57 | MARTIGNETTI COMPANIES | 25.502096M | 13.828264M | 39.330359M | 8.30 |
| 68 | PERNOD RICARD USA | 23.851164M | 8.212032M | 32.063196M | 7.76 |
| 46 | JIM BEAM BRANDS COMPANY | 23.494304M | 7.928716M | 31.423020M | 7.65 |
| 6 | BACARDI USA INC | 17.432020M | 7.422797M | 24.854817M | 5.67 |
| 20 | CONSTELLATION BRANDS INC | 15.273708M | 8.945038M | 24.218746M | 4.97 |
| 11 | BROWN-FORMAN CORP | 13.238939M | 5.008291M | 18.247231M | 4.31 |
| 30 | E & J GALLO WINERY | 12.068539M | 6.331360M | 18.399899M | 3.93 |
| 106 | ULTRA BEVERAGE COMPANY LLP | 11.150990M | 5.334674M | 16.485664M | 3.63 |
| 53 | M S WALKER INC | 9.764400M | 4.938614M | 14.703014M | 3.18 |



In [299...]

`top_vendors['PurchaseContribution%'].sum()`

Out[299...]

65.71000000000001

In [300...]

`top_vendors['Cumulative_Contribution%']=top_vendors['PurchaseContribution%'].cumsum`

In [301...]

`top_vendors`

Out[301...]

| | VendorName | TotalPurchaseDol | Gross Profit | TotalSalesDol | PurchaseContribution% |
|-----|----------------------------|------------------|--------------|---------------|-----------------------|
| 25 | DIAGEO NORTH AMERICA INC | 50.097226M | 17.892873M | 67.990099M | 16.31 |
| 57 | MARTIGNETTI COMPANIES | 25.502096M | 13.828264M | 39.330359M | 8.30 |
| 68 | PERNOD RICARD USA | 23.851164M | 8.212032M | 32.063196M | 7.76 |
| 46 | JIM BEAM BRANDS COMPANY | 23.494304M | 7.928716M | 31.423020M | 7.65 |
| 6 | BACARDI USA INC | 17.432020M | 7.422797M | 24.854817M | 5.67 |
| 20 | CONSTELLATION BRANDS INC | 15.273708M | 8.945038M | 24.218746M | 4.97 |
| 11 | BROWN-FORMAN CORP | 13.238939M | 5.008291M | 18.247231M | 4.31 |
| 30 | E & J GALLO WINERY | 12.068539M | 6.331360M | 18.399899M | 3.93 |
| 106 | ULTRA BEVERAGE COMPANY LLP | 11.150990M | 5.334674M | 16.485664M | 3.63 |
| 53 | M S WALKER INC | 9.764400M | 4.938614M | 14.703014M | 3.18 |



In [302...]

```
fig, ax1 = plt.subplots(figsize=(10,6))

#Bar plot for Purchase Contribution%
sns.barplot(x=top_vendors['VendorName'], y=top_vendors['PurchaseContribution%'], palette='viridis')

for i, value in enumerate(top_vendors['PurchaseContribution%']):
    ax1.text(i, value - 1, str(value) + '%', ha='center', fontsize=10, color='white')

# Line Plot for Cumulative Contribution%

ax2 = ax1.twinx()
ax2.plot(top_vendors['VendorName'], top_vendors['Cumulative_Contribution%'], color='red')

ax1.set_xticklabels(top_vendors['VendorName'], rotation=90)

ax1.set_ylabel('PurchaseContribution %', color='blue')

ax2.set_ylabel('Cumulative Contribution %', color='red')
```

```

ax1.set_xlabel('Vendors')

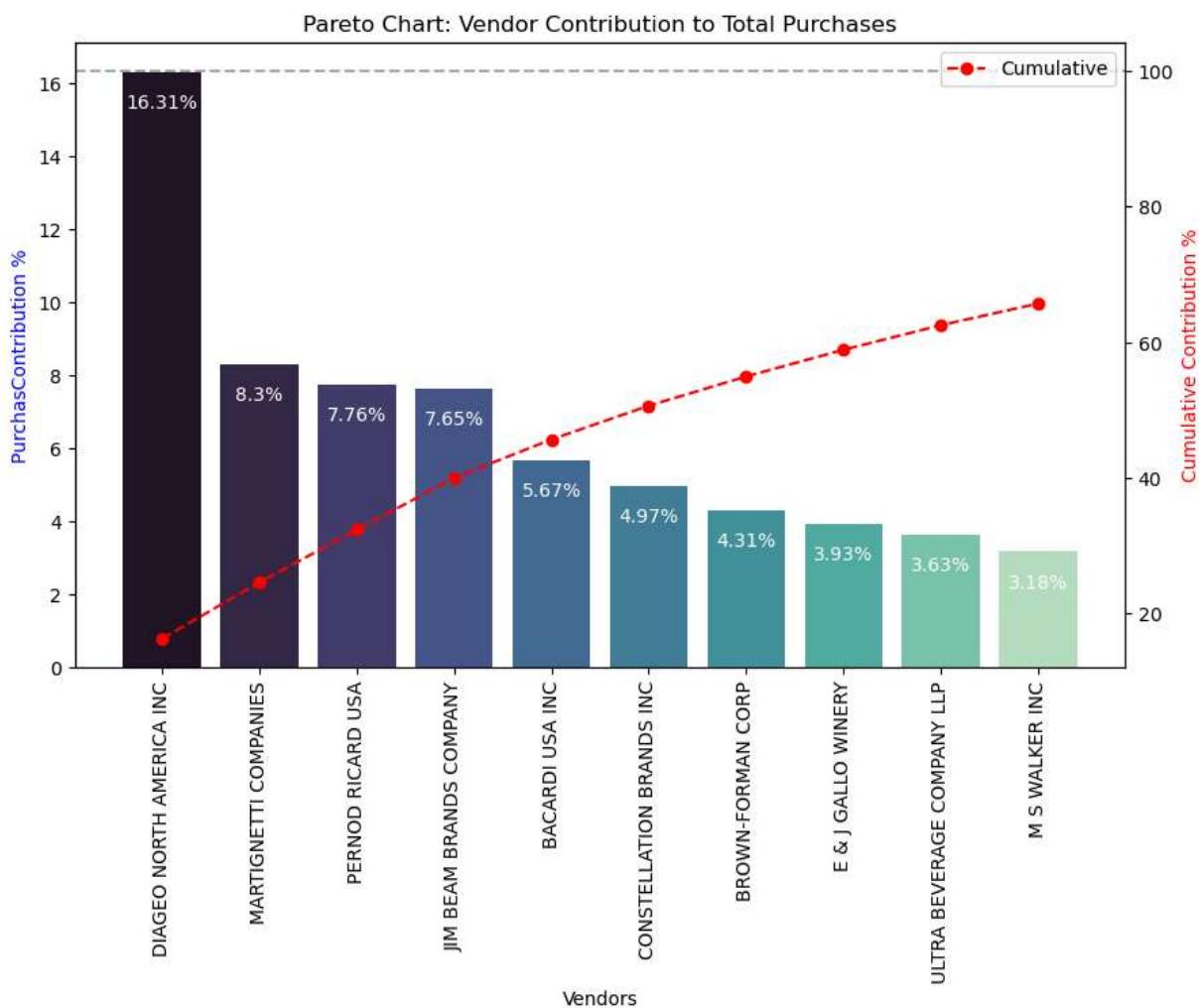
ax1.set_title('Pareto Chart: Vendor Contribution to Total Purchases')

ax2.axhline(y=100, color='gray', linestyle='dashed', alpha=0.7)

ax2.legend(loc='upper right')

plt.show()

```



How much of total procurement is dependent on the top Vendors?

```
In [305]: print(f"Total Purchase Contribution of total 10 vendor is {round(top_vendors['PurchaseContribution%'].sum(), 2)}%")
```

Total Purchase Contribution of total 10 vendor is 65.71%

```

In [310]: vendors = list(top_vendors['VendorName'].values)
purchase_contributions = list(top_vendors['PurchaseContribution%'].values)
total_contribution = sum(purchase_contributions)
remaining_contribution = 100 - total_contribution

# Append "Other Vendors" category

vendors.append("Other Vendors")
purchase_contributions.append(remaining_contribution)

```

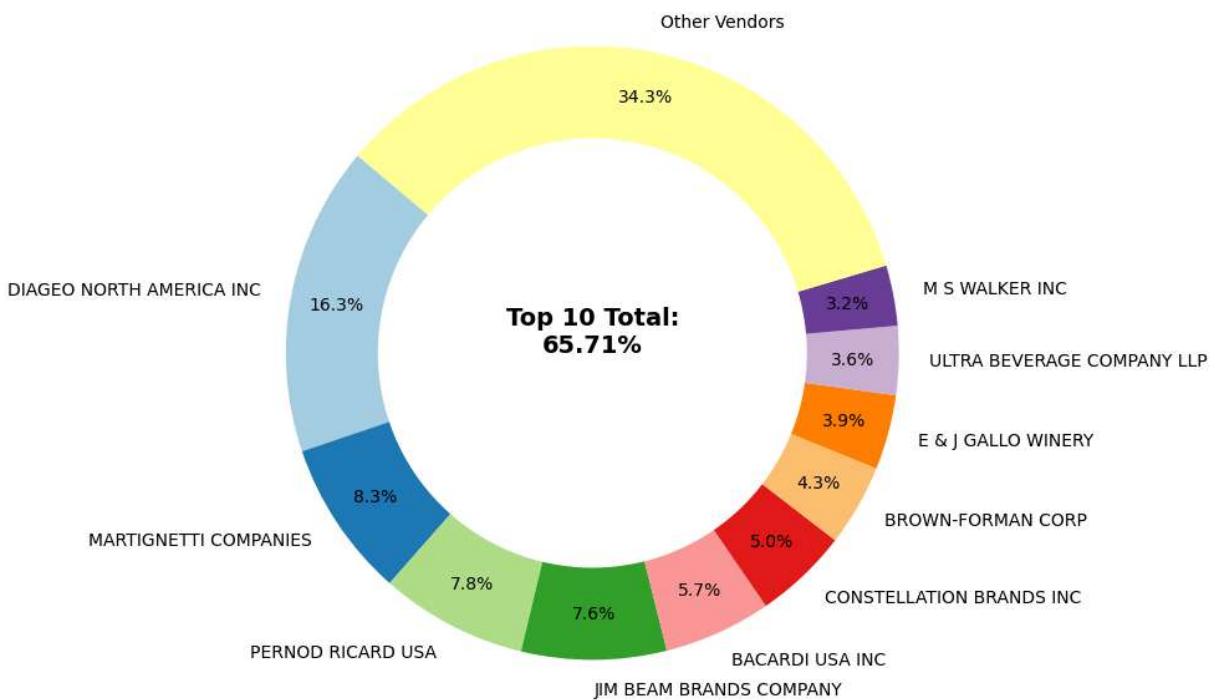
```
# Donut Chart
fig, ax = plt.subplots (figsize=(8, 8))
wedges, texts, autotexts = ax.pie(purchase_contributions, labels =vendors, autopct=
                                startangle=140, pctdistance=0.85, colors=plt.cm.P

# Draw a white circle in the center to create a "donut" effect
centre_circle = plt.Circle((0, 0), 0.70, fc='white')
fig.gca().add_artist(centre_circle)

# Add Total Contribution annotation in the center
plt.text(0, 0, f"Top 10 Total:\n{total_contribution:.2f}%", fontsize=14, fontweight='bold')

plt.title("Top 10 Vendor's Purchase Contribution (%)")
plt.show()
```

Top 10 Vendor's Purchase Contribution (%)



Does purchasing in bulk reduce the unit price, and what is the optimal purchase volume for cost savings?

```
In [311... df['UnitPurchasePrice']=df['TotalPurchaseDol']/df['TotalPurchaseQty']

In [313... df['OrderSize']=pd.qcut(df["TotalPurchaseQty"],q=3,labels=["Small","Medium","Large"])

In [316... df.groupby('OrderSize')[['UnitPurchasePrice']].mean()
```

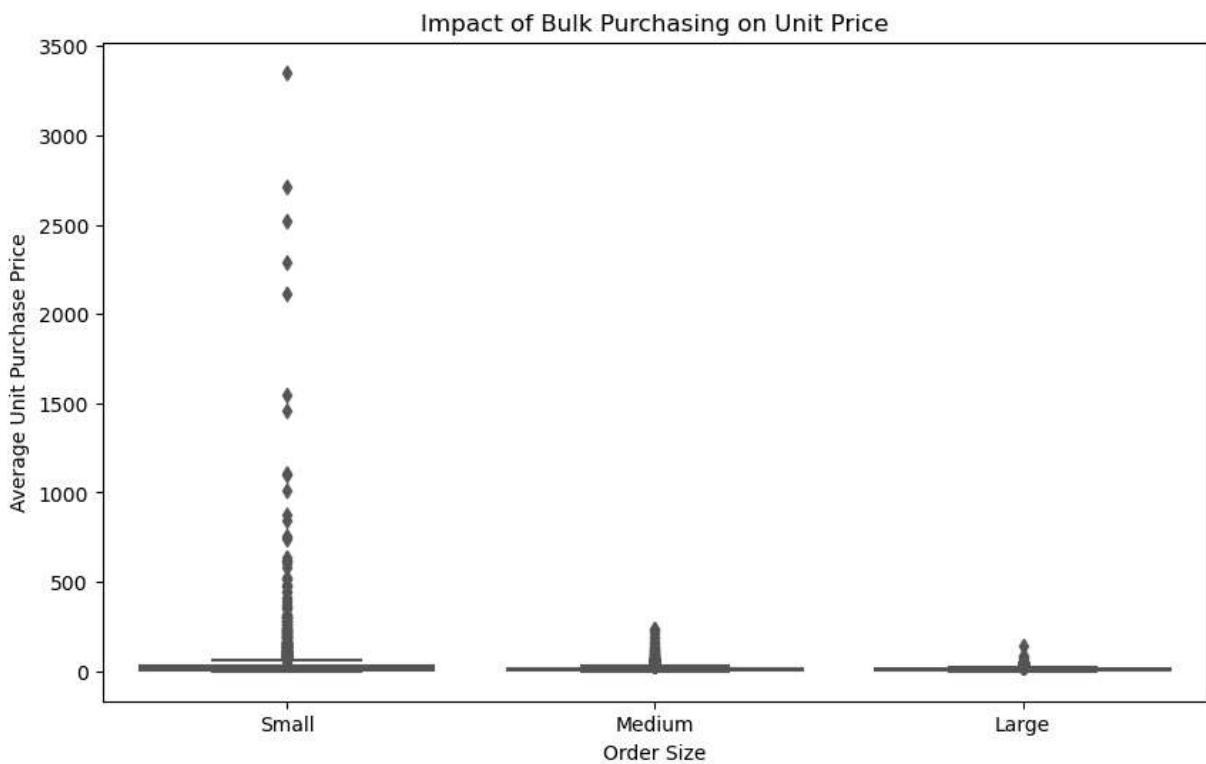
Out[316...]

UnitPurchasePrice**OrderSize**

| | |
|---------------|-----------|
| Small | 39.068606 |
| Medium | 15.479544 |
| Large | 10.771636 |

In [317...]

```
plt.figure(figsize=(10, 6))
sns.boxplot(data=df, x="OrderSize", y="UnitPurchasePrice", palette="Set2")
plt.xlabel("Order Size")
plt.title("Impact of Bulk Purchasing on Unit Price")
plt.ylabel("Average Unit Purchase Price")
plt.show()
```



Vendors buying in bulk (Large Order Size) get the lowest unit price (\$10.78 per unit), meaning higher margins if they can manage inventory efficiently.

The price difference between Small and Large orders is substantial (~72% reduction in unit cost)

This suggests that bulk pricing strategies successfully encourage vendors to purchase in larger volumes, leading to higher overall sales despite lower per-unit revenue.

Which vendor have low inventory turnover, indicating excess stock and slow-moving products?

In [320...]

```
df[df['StockTurnover'] < 1].groupby('VendorName')[['StockTurnover']].mean().sort_values
```

Out[320...]

StockTurnover

| VendorName | |
|------------------------------------|----------|
| ALISA CARR BEVERAGES | 0.615385 |
| PARK STREET IMPORTS LLC | 0.699412 |
| HIGHLAND WINE MERCHANTS LLC | 0.708333 |
| Circa Wines | 0.755676 |
| Dunn Wine Brokers | 0.766022 |
| CENTEUR IMPORTS LLC | 0.773953 |
| SMOKY QUARTZ DISTILLERY LLC | 0.783835 |
| TAMWORTH DISTILLING | 0.797078 |
| THE IMPORTED GRAPE LLC | 0.807569 |
| WALPOLE MTN VIEW WINERY | 0.820548 |

How much capital is locked in unsold inventory per vendor, and which vendors contribute the most to it?

In [324...]

```
df ["UnsoldInventoryValue"] = (df["TotalPurchaseQty"] - df ["TotalSalesQty"]) * df
print('Total Unsold Capital:', format_dollars (df ["UnsoldInventoryValue"].sum()))
```

Total Unsold Capital: 2.797645M

In [322...]

```
df.columns
```

Out[322...]

```
Index(['VendorNumber', 'VendorName', 'Brand', 'Price', 'Volume', 'ActualPrice',
       'TotalPurchaseQty', 'TotalPurchaseDol', 'FreightCost', 'TotalSalesDol',
       'TotalSalesPrice', 'TotalSalesQty', 'TotalExciseTax', 'Description',
       'Gross Profit', 'ProfitMargin', 'StockTurnover', 'SalesPurchaseRatio',
       'UnitPurchasePrice', 'OrderSize'],
      dtype='object')
```

In [326...]

```
# Aggregate Capital Locked per Vendor
inventory_value_per_vendor = df.groupby("VendorName") ["UnsoldInventoryValue"].sum()

# Sort Vendors with the Highest Locked Capital
inventory_value_per_vendor = inventory_value_per_vendor.sort_values(by="UnsoldInventoryValue")
inventory_value_per_vendor['UnsoldInventoryValue'] = inventory_value_per_vendor ['U
inventory_value_per_vendor.head(10)
```

Out[326...]

| | VendorName | UnsoldInventoryValue |
|-----|--------------------------|----------------------|
| 25 | DIAGEO NORTH AMERICA INC | 722.209050K |
| 46 | JIM BEAM BRANDS COMPANY | 554.665630K |
| 68 | PERNOD RICARD USA | 470.625610K |
| 116 | WILLIAM GRANT & SONS INC | 401.960830K |
| 79 | SAZERAC CO INC | 271.538910K |
| 30 | E & J GALLO WINERY | 228.282610K |
| 11 | BROWN-FORMAN CORP | 177.733740K |
| 20 | CONSTELLATION BRANDS INC | 133.617620K |
| 61 | MOET HENNESSY USA INC | 126.477700K |
| 77 | REMY COINTREAU USA INC | 118.598150K |

What is the 95% confidence intervals for profit margins of top-performing and low-performing vendors.

In [329...]

```
top_threshold = df ["TotalSalesDol"].quantile(0.75)
low_threshold = df ["TotalSalesDol"].quantile(0.25)
```

In [330...]

```
top_vendors = df [df ["TotalSalesDol"] >= top_threshold] ["ProfitMargin"].dropna()
low_vendors = df [df ["TotalSalesDol"] <= low_threshold] ["ProfitMargin"].dropna()
```

In [331...]

```
top_vendors
```

Out[331...]

```
12      16.999601
15      30.557678
20      30.827478
21      27.129385
24      24.329082
       ...
10655   25.173464
10656   22.886164
10657   22.652408
10673   27.103538
10675   28.056879
Name: ProfitMargin, Length: 2141, dtype: float64
```

In [343...]

```
def confidence_interval(data, confidence=0.95):
    mean_val = np.mean(data)
    std_err = np.std(data, ddof=1) / np.sqrt(len(data))
    t_critical= stats.t.ppf((1 + confidence) / 2, df=len(data) - 1)
    margin_of_error = t_critical * std_err
    return mean_val, mean_val - margin_of_error, mean_val + margin_of_error
```

In [344...]

```
top_mean,top_lower,top_upper = confidence_interval(top_vendors)
low_mean,low_lower,low_upper = confidence_interval(low_vendors)
```

```

print(f"Top Vendors 95% CI: ({top_lower:.2f}, {top_upper:.2f}), Mean: {top_mean:.2f}")
print(f"Low Vendors 95% CI: ({low_lower:.2f}, {low_upper:.2f}), Mean: {low_mean:.2f}")
plt.figure(figsize=(12, 6))

# Top Vendors Plot
sns.histplot(top_vendors, kde=True, color="blue", bins= 30, alpha=0.5, label="Top Vendors")
plt.axvline(top_lower, color="blue", linestyle="--", label=f"Top Lower: {top_lower:.2f}")
plt.axvline(top_upper, color="blue", linestyle="--", label=f"Top Upper: {top_upper:.2f}")
plt.axvline(top_mean, color="blue", linestyle="-", label=f"Top Mean: {top_mean:.2f}")

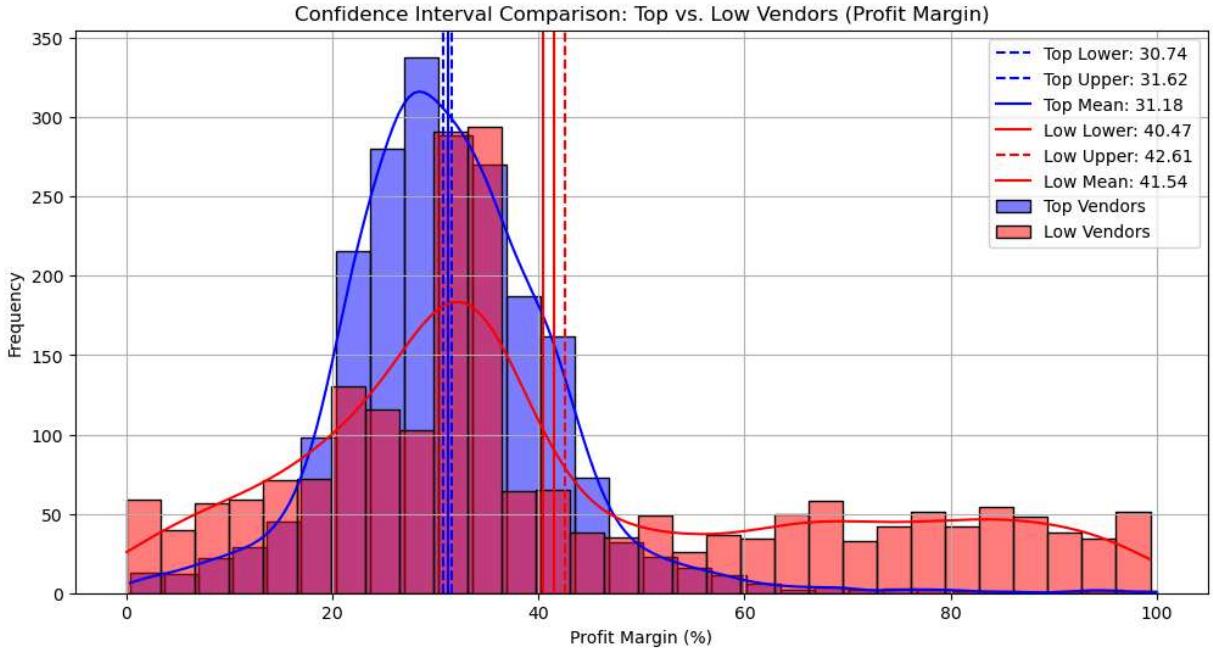
# Low Vendors Plot
sns.histplot(low_vendors, kde=True, color="red", bins= 30, alpha=0.5, label="Low Vendors")
plt.axvline(low_lower, color="red", linestyle="--", label=f"Low Lower: {low_lower:.2f}")
plt.axvline(low_upper, color="red", linestyle="--", label=f"Low Upper: {low_upper:.2f}")
plt.axvline(low_mean, color="red", linestyle="-", label=f"Low Mean: {low_mean:.2f}")

# Finalize Plot
plt.title("Confidence Interval Comparison: Top vs. Low Vendors (Profit Margin)")
plt.xlabel("Profit Margin (%)")
plt.ylabel("Frequency")
plt.legend()
plt.grid(True)
plt.show()

```

Top Vendors 95% CI: (30.74, 31.62), Mean: 31.18

Low Vendors 95% CI: (40.47, 42.61), Mean: 41.54



The confidence interval for low-performing vendors (40.48% to 42.62%) is significantly higher than that of top-performing vendors (30.74% to 31.61%).

This suggests that vendors with lower sales tend to maintain higher profit margins, potentially due to premium pricing or lower operational costs

For High-Performing Vendors: If they aim to improve profitability, they could explore selective price adjustments, cost optimization, or bundling strategies.

For Low-Performing Vendors: Despite higher margins, their low sales volume might indicate a need for better marketing, competitive pricing, or improved distribution strategies.

Is there a significant difference in profit margins between top-performing and low-performing vendors?

Hypothesis:

H₀ (Null Hypothesis): There is no significant difference in the mean profit margins of top-performing and low-performing vendors.

H₁ (Alternative Hypothesis): The mean profit margins of top-performing and low-performing vendors are significantly different.

```
In [348...]:  
top_threshold= df ["TotalSalesDol"].quantile(0.75)  
low_threshold = df ["TotalSalesDol"].quantile(0.25)  
  
top_vendors = df [df ["TotalSalesDol"] >= top_threshold] ["ProfitMargin"].dropna()  
low_vendors = df [df ["TotalSalesDol"] <= low_threshold] ["ProfitMargin"].dropna()  
  
# Perform Two-Sample T-Test  
t_stat, p_value = ttest_ind(top_vendors, low_vendors, equal_var=False)  
  
#Print results  
print(f"T-Statistic: {t_stat:.4f}, P-Value: {p_value:.4f}")  
  
if p_value < 0.05:  
    print("Reject H0: There is a significant difference in profit margins between top and low-performing vendors.  
else:  
    print("Fail to Reject H0: No significant difference in profit margins.")
```

T-Statistic: -17.5944, P-Value: 0.0000
Reject H₀: There is a significant difference in profit margins between top and low-performing vendors.

```
In [ ]:
```