```
#uploading the dataset
from google.colab import files
uploaded = files.upload()
```

Choose files Warehouse…ail_Sales.csv

- **Warehouse_and_Retail_Sales.csv**(text/csv) - 27451860 bytes, last modified: 26/06/2025 - 100% done

```
Saving Warehouse_and_Retail_Sales.csv to Warehouse_and_Retail_Sales.csv
```

```
#import python libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
#load and view the data
df = pd.read_csv("Warehouse_and_Retail_Sales.csv")
df.head()
```

	YEAR	MONTH	SUPPLIER	ITEM CODE	ITEM DESCRIPTION	ITEM TYPE	RETAIL SALES	RETAIL TRANSFERS	WAREHOUSE SALES
0	2020	1	REPUBLIC NATIONAL DISTRIBUTING CO	100009	BOOTLEG RED - 750ML	WINE	0.00	0.0	2.0
1	2020	1	PWSWN INC	100024	MOMENT DE PLAISIR - 750ML	WINE	0.00	1.0	4.0
2	2020	1	RELIABLE CHURCHILL LLLP	1001	S SMITH ORGANIC PEAR CIDER - 18.7OZ	BEER	0.00	0.0	1.0
3	2020	1	LANTERNA DISTRIBUTORS INC	100145	SCHLINK HAUS KABINETT - 750ML	WINE	0.00	0.0	1.0
4	2020	1	DIONYSOS IMPORTS INC	100293	SANTORINI GAVALA WHITE -	WINE	0.82	0.0	0.0

```
#understand shape and structure
print("Shape:", df.shape)
```

```
df.info()
```

```
Shape: (307645, 9)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 307645 entries, 0 to 307644
Data columns (total 9 columns):
 #   Column            Non-Null Count   Dtype
---  ------            --------------   -----
 0   YEAR              307645 non-null  int64
 1   MONTH             307645 non-null  int64
 2   SUPPLIER          307478 non-null  object
 3   ITEM CODE         307645 non-null  object
 4   ITEM DESCRIPTION  307645 non-null  object
 5   ITEM TYPE         307644 non-null  object
 6   RETAIL SALES      307642 non-null  float64
 7   RETAIL TRANSFERS  307645 non-null  float64
 8   WAREHOUSE SALES   307645 non-null  float64
dtypes: float64(3), int64(2), object(4)
memory usage: 21.1+ MB
```

```
#if any duplicates or missing rows
df.dropna(inplace=True)
df.drop_duplicates(inplace=True)
```

```
#rechecking after cleaning
print("Cleaned Shape:", df.shape)
print("Missing after cleaning:\n", df.isnull().sum())
print("Duplicates after cleaning:", df.duplicated().sum())
```

```
Cleaned Shape: (307477, 9)
Missing after cleaning:
 YEAR                0
MONTH               0
SUPPLIER            0
ITEM CODE           0
ITEM DESCRIPTION    0
ITEM TYPE           0
RETAIL SALES        0
RETAIL TRANSFERS    0
WAREHOUSE SALES     0
dtype: int64
Duplicates after cleaning: 0
```

```
#explore unique values
print("Unique ITEM TYPES:", df['ITEM TYPE'].unique())
print("Total SUPPLIERS:", df['SUPPLIER'].nunique())
```

```
Unique ITEM TYPES: ['WINE' 'BEER' 'LIQUOR' 'STR_SUPPLIES' 'KEGS' 'REF' 'NON-ALCOHOL'
 'DUNNAGE']
Total SUPPLIERS: 396
```

```
#reatil sales by item type
df.groupby('ITEM TYPE')['RETAIL SALES'].sum().sort_values(ascending=False)
```

ITEM TYPE	RETAIL SALES
LIQUOR	802691.43
WINE	746498.59
BEER	574220.53
NON-ALCOHOL	27150.31
STR_SUPPLIES	2234.90
REF	663.63
KEGS	0.00
DUNNAGE	0.00

dtype: float64

```
#warehouse sales by supplier
df.groupby('SUPPLIER')['WAREHOUSE SALES'].sum().sort_values(ascending=False)
```

WAREHOUSE SALES SUPPLIER	
CROWN IMPORTS	1651871.51
MILLER BREWING COMPANY	1425428.71
ANHEUSER BUSCH INC	1331170.84
HEINEKEN USA	829796.46
E & J GALLO WINERY	197463.78
...	...
AZABU DISTILLING CO LLC	0.00
A HARDY USA LTD	0.00
ZURENA LLC	0.00
ROBERT KACHER SELECTIONS LLC	-1.00
PREMIUM DISTRIBUTORS INC	-53782.00

396 rows × 1 columns

dtype: float64

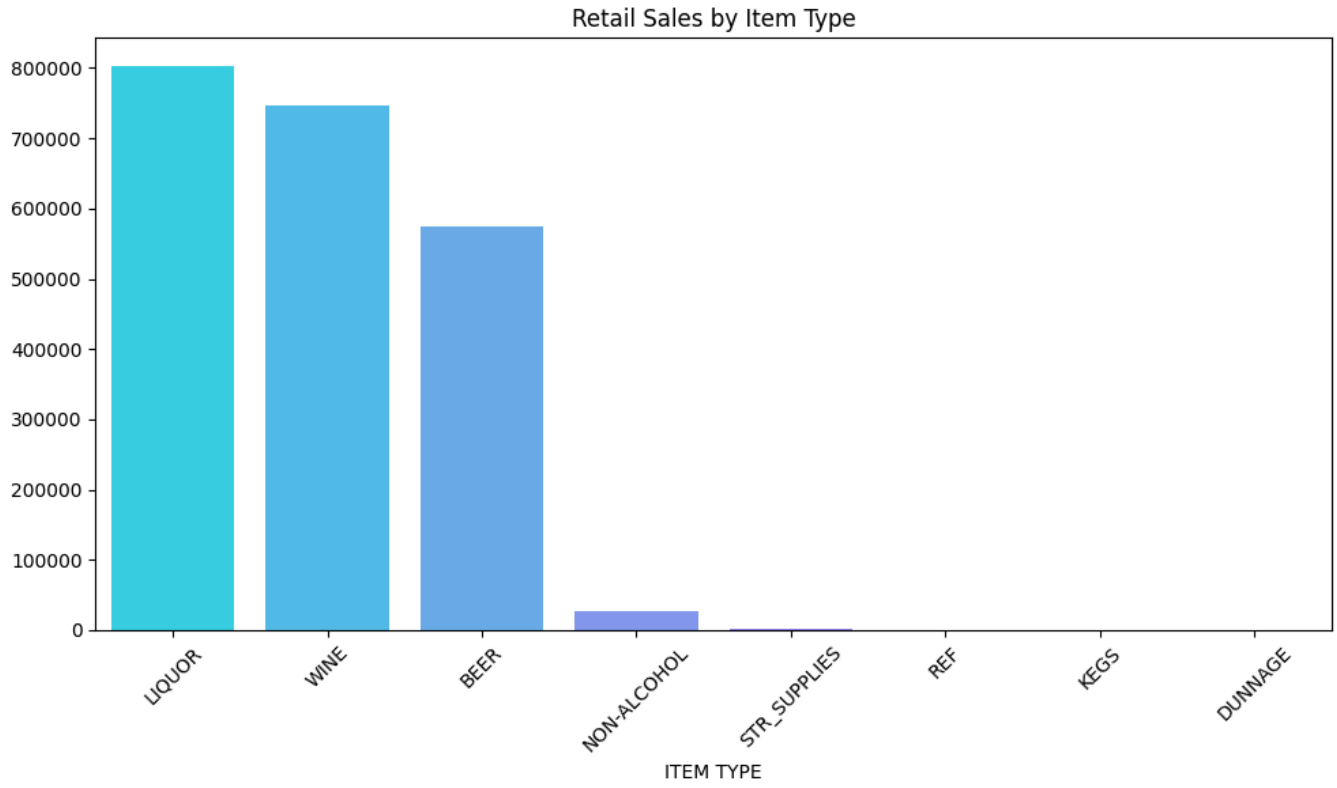
```
#graphs and visuals
item_sales = df.groupby('ITEM TYPE')['RETAIL SALES'].sum().sort_values(ascending=False)

plt.figure(figsize=(10,6))
sns.barplot(x=item_sales.index, y=item_sales.values, palette="cool")
plt.title("Retail Sales by Item Type")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

```
/tmp/ipython-input-11-2864504325.py:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0.

  sns.barplot(x=item_sales.index, y=item_sales.values, palette="cool")
```



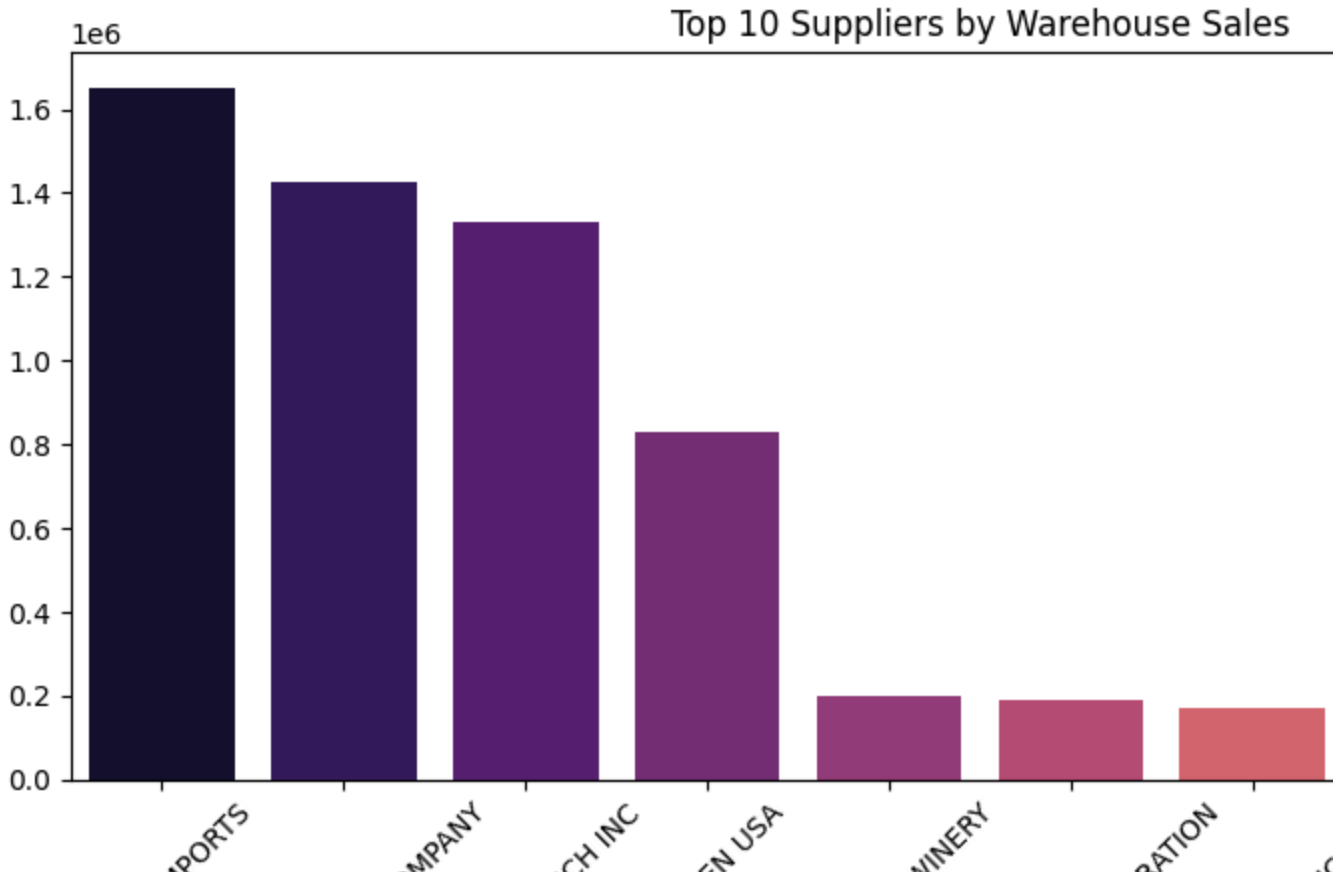

```
#warehouse  sales  by top suppliers
supplier_sales = df.groupby('SUPPLIER')['WAREHOUSE SALES'].sum().sort_values(ascending=False

plt.figure(figsize=(10,6))
sns.barplot(x=supplier_sales.index, y=supplier_sales.values, palette="magma")
plt.title("Top 10 Suppliers by Warehouse Sales")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

```
/tmp/ipython-input-12-2849460217.py:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0.

  sns.barplot(x=supplier_sales.index, y=supplier_sales.values, palette="magma")
```




```
# Step 11: Final Insights from Graph (Warehouse Sales by Top 10 Suppliers)

insights = [
    "📊 Insights from Warehouse Sales by Top 10 Suppliers:",
    "",
    "1️⃣ Crown Imports is the top-performing supplier, with warehouse sales crossing 1.6 mil
    "2️⃣ Miller Brewing Company and Anheuser Busch Inc follow closely in 2nd and 3rd positio
    "3️⃣ Heineken USA ranks 4th, but shows a noticeable drop from the top 3, indicating perf
    "4️⃣ Suppliers from rank 5 to 10 contribute much less, indicating that a small set of su
    "5️⃣ The sales distribution follows the 80/20 rule – majority of sales come from very fe
    "",
    "✅ Business Tip: Focus more on top 3 suppliers for inventory planning and explore perf
]

# Print each point line by line
for point in insights:
    print(point)
```