Northwind Traders

Exploratory Data Analysis (EDA)

shippers.csv	suriya	May 2, 2023	79 bytes
products.csv	suriya	May 2, 2023	3 KB
orders.csv	suriya	May 2, 2023	45 KB
order_details.csv	suriya	May 2, 2023	42 KB
employees.csv	suriya	May 2, 2023	512 bytes
customers.csv	suriya	May 2, 2023	7 KB
Copy of data_dictionary.csv	suriya	Feb 6, 2025	2 KB
categories.csv	suriya	May 2, 2023	406 bytes

Northwind Traders Provide total of 7 CSV Files

Step - 1: Merging a 7 datasets

```
[ ]  1 df14.columns
```

```
Index(['shipperID', 'employeeID', 'employeeName', 'title', 'reportsTo',
       'categoryID', 'categoryName', 'description', 'productID', 'productName',
       'quantityPerUnit', 'discontinued', 'customerID', 'contactName',
       'contactTitle', 'orderID', 'orderDate', 'requiredDate', 'shippedDate',
       'freight', 'quantity', 'discount', 'unitPrice', 'country', 'city',
       'companyName'],
      dtype='object')
```

Step 2: Handling Missing Values

to fill missing value - mean, median, mode, ffill, bfill and custom fill or delete column more than 70% missing values based on client suggestion or business analytics referred by client

i am filling missing values based on skew and outliers

title	0
reportsTo	241
categoryID	0
categoryName	0
description	0
productID	0
productName	0
quantityPerUnit	0
discontinued	0
customerID	0
contactName	0
contactTitle	0
orderID	0
orderDate	0
requiredDate	0
shippedDate	73
freight	0

Step 3: Outliers

Find Outlier using BoxPlot & IQR Method

```
1 outliers_table
```

	Column	Lower Outlier Count	Upper Outlier Count	Lower Outlier Values	Upper Outlier Values
0	shipperID	0	0	[]	[]
1	employeeID	0	0	[]	[]
2	reportsTo	0	0	[]	[]
3	categoryID	0	0	[]	[]
4	productID	0	0	[]	[]
5	discontinued	0	228	[]	[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
6	orderID	0	0	[]	[]
7	freight	0	155	[]	[348.14, 348.14, 297.18, 328.74, 297.18, 297.1...
8	quantity	0	87	[]	[120, 90, 65, 70, 77, 70, 77, 70, 110, 80, 120...
9	discount	0	0	[]	[]
10	unitPrice	0	77	[]	[263.5, 263.5, 263.5, 97.0, 123.79, 81.0, 97.0...

Step 4: EDA (Exploratory Data Analysis)

Differentiate Categorical and Numerical Columns

Exploratory Data Analysis (EDA)

```
[ ]  1 # Categorical and numerical column detection
     2 categorical_columns = df.select_dtypes(include=['object', 'category']).columns.tolist()
     3 numerical_columns = df.select_dtypes(include=['number']).columns.tolist()
     4
     5 print("Categorical Columns:", categorical_columns)
     6 print("Numerical Columns:", numerical_columns)
     7
```

```
', 'categoryName', 'description', 'productName', 'quantityPerUnit', 'customerID', 'contactName', 'contactTitle', 'orderDate',
', 'reportsTo', 'categoryID', 'productID', 'discontinued', 'orderID', 'freight', 'quantity', 'discount', 'unitPrice']
```

Check category column value check for Analysis

Columns have more 10 unique values i categorise as SQL Based EDA

Columns have less than 10 unique Values categorised as Chart Based EDA

Chart Based (EDA) - Count - 8

SQL Based (EDA) Count - 7

Chart Based (EDA) - Count - 8

1)employeeName,
2)title,
3)categoryName,
4)description,
5)country,
6)city,
7)companyName,
8)contactTitle

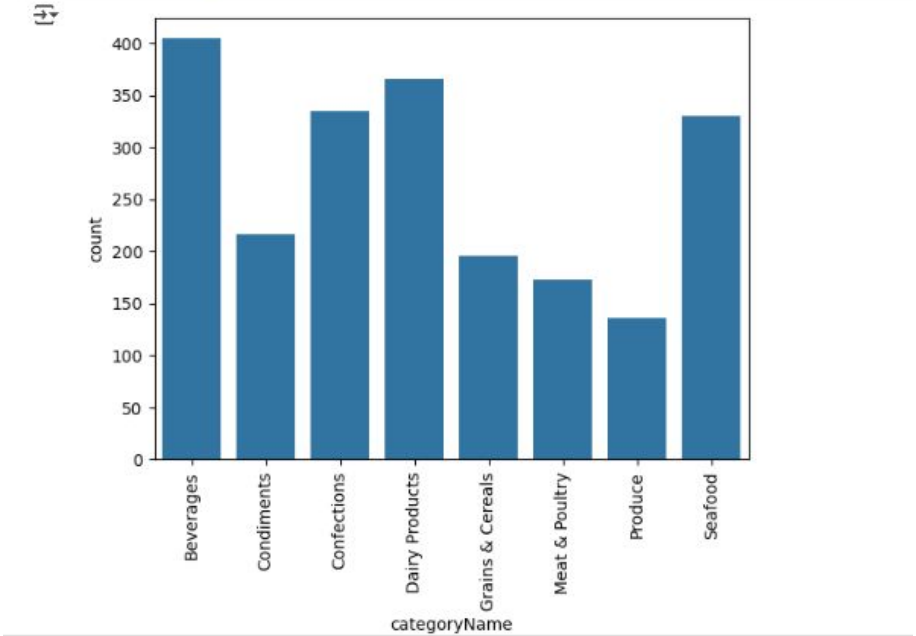
SQL Based (EDA) Count - 7

1) productName,
2)quantityPerUnit,
3)customerID,
4)orderDate,
5)requiredDate,
6)shippedDate,
7)contactName

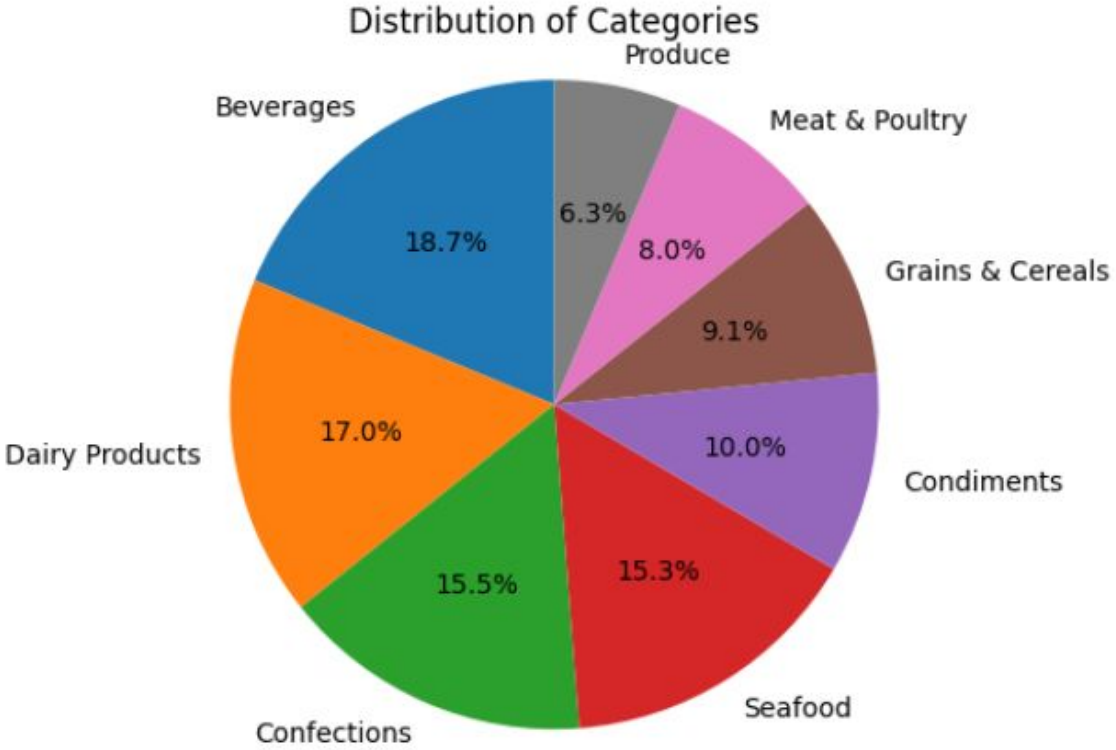
1) Univariate Analysis

Analysing which Category

```
1 sns.countplot(x='categoryName', data=df)
2 plt.xticks(rotation=90)  # Rotate x-axis labels for better readability
3 plt.show()
```



```
categoryName_counts = df['categoryName'].value_counts()
plt.pie(categoryName_counts, labels=categoryName_counts.index, autopct='%1.1f%%', startangle=90)
plt.title('Distribution of Categories') # Changed the title to reflect the data
plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.
plt.show()
```



Most Sales:-

1) Beverages Highest Sales Overall 18%

2) Dairy Product 2nd Most Highest Sales 17%

3) Condection 3rd Most Highest Sales 15.5%

4) Seafood 4th Most Highrst Sales 15.3%

#

#

#

1) Least Sales - Produce and Meet & Poultry

2) Bivariate Analysis

```
plt.figure(figsize=(12, 8))  # Adjust figure size if needed
sns.heatmap(cross_tab, annot=True, cmap="YlGnBu", fmt="d")
plt.title("Cross-Tabulation of Employee Name vs. Title")
plt.xlabel("category Name")
plt.ylabel("Country")
plt.xticks(rotation=45, ha='right')  # Rotate x-axis labels if needed
plt.tight_layout()
plt.show()
```

```
categoryName  Beverages  Condiments  Confections  Dairy Products  \
country
UK                  110          58           78             126
USA                 294         158          256             240

categoryName  Grains & Cereals  Meat & Poultry  Produce  Seafood
country
UK                          39              42       40       75
USA                         157             131       96      255
```


Cross-Tabulation of Employee Name vs. Title
Country
UK
USA
110
58
78
126
39
42
40
75
294
158
256
240
157
131
96
255
Beverages
Condiments
Confections
Dairy Products
Grains & Cereals
Meat & Poultry
Produce
Seafood
category Name
250
200
150
100
50

Insights

1) In USA Import Highest number of Beverages Import From Northwind Traders Compare to UK

2) Confections Category Second Mostly Imported

3) Seafood Category third Mostly Imported

4) Produce Category Least Mostly Imported

3) MultiVariate

in MultiVariate iam Analysing of Category - and anlysing each categories (descriptions) and which category is mostly exported and exported especially which Country or City, and Which Company Product are Mostly Exported from Northwind Traders

categoryName		Beverages	Condiments	Confections	Dairy Products	\
country	companyName					
UK	Federal Shipping	38	16	23	36	
	Speedy Express	31	18	19	44	
	United Package	41	24	36	46	
USA	Federal Shipping	80	47	78	72	
	Speedy Express	91	54	81	71	
	United Package	123	57	97	97	

categoryName		Grains & Cereals	Meat & Poultry	Produce	Seafood
country	companyName				
UK	Federal Shipping	14	15	12	24
	Speedy Express	16	10	13	23
	United Package	9	17	15	28
USA	Federal Shipping	42	42	28	78
	Speedy Express	43	34	26	72
	United Package	72	55	42	105


Cross-Tabulation: Country & Company vs Category Name
Country & Company
UK-Federal Shipping
UK-Speedy Express
UK-United Package
USA-Federal Shipping
USA-Speedy Express
USA-United Package
38 16 23 36 14 15 12 24
31 18 19 44 16 10 13 23
41 24 36 46 9 17 15 28
80 47 78 72 42 42 28 78
91 54 81 71 43 34 26 72
123 57 97 97 72 55 42 105
Beverages
Condiments
Confections
Dairy Products
Grains & Cereals
Meat & Poultry
Produce
Seafood
Category Name
120
100
80
60
40
20

Insights:

United Packages from USA are highly contributed to Northwind Traders, especially in Beverages and Seafood.

Speedy Express from USA is the second most highly contributed to Northwind Traders.

Conclusion:

1) Northwind Traders Export - USA & UK and New york From USA and London From UK

2) Especially USA is the Largest Importer from Northwind Traders
3) Northwind Traders Flagship Category was Beverage items, Confections and Seafoods
4) United Package From USA Largest Shipping Partner for Northwind Traders they highly shipped Beverages & Seafoods
5) Margaret Peacock - Sales Representative Cover highest number of orders

Thank You

1. Intro

Choose one approach to grab the audience's attention right from the start: unexpected, emotional, or simple.

➔ **Unexpected**
Highlight what's new, unusual, or surprising.

➔ **Emotional**
Give people a reason to care.

➔ **Simple**
Provide a simple unifying message for what is to come